Pacific Journal of Mathematics

Volume 327 No. 1 November 2023

PACIFIC JOURNAL OF MATHEMATICS

Founded in 1951 by E. F. Beckenbach (1906–1982) and F. Wolf (1904–1989)

msp.org/pjm

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

The subscription price for 2023 is US $605/year for the electronic version, and $820/year for print and electronic.
Subscriptions, requests for back issues and changes of subscriber address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 1945-5844 electronic, 0030-8730 printed) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFlow® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 327, No. 1, 2023
https://doi.org/10.2140/pjm.2023.327.1

THE HOMOLOGY OF THE PARTITION ALGEBRAS

RACHAEL BOYD, RICHARD HEPWORTH AND PETER PATZT

We show that the homology of the partition algebras, interpreted as appropriate Tor-groups, is isomorphic to that of the symmetric groups in a range of degrees that increases with the number of nodes. Further, we show that when the defining parameter δ of the partition algebra is invertible, the homology of the partition algebra is in fact isomorphic to the homology of the symmetric group in all degrees. These results parallel those obtained for the Brauer algebras in the authors' earlier work, but with significant differences and difficulties in the *inductive resolution* and *high acyclicity* arguments required to prove them. Our results join the growing literature on homological stability for algebras, which now encompasses the Temperley–Lieb, Brauer and partition algebras, as well as the Iwahori–Hecke algebras of types A and B.

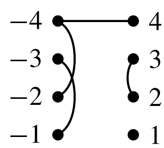
1. Introduction

In the last few years it has become increasingly apparent that the techniques of *homological stability*, which are most commonly applied to families of groups, can be successfully applied to families of *algebras*, where homology is interpreted as an appropriate Tor group. Indeed, Boyd and Hepworth [2020], Boyd, Hepworth and Patzt [2021], Hepworth [2022] and Moselle [2022] proved homological stability for Temperley–Lieb algebras, Brauer algebras, and Iwahori–Hecke algebras of types A and B respectively, and identified the stable homology in the first two cases. The Temperley–Lieb and Brauer algebras failed to satisfy a certain flatness condition that holds automatically for families of groups, necessitating the introduction of the new technique of *inductive resolutions*. Using related techniques, Sroka [2023] showed that the homology of the Temperley–Lieb algebra on an odd number of strands vanishes in positive degrees, in contrast to the known nonvanishing for an even number of strands. More recently, Boyde [2022] used a careful study of idempotents to unify and generalise the "invertible parameter" results from [Boyd and Hepworth 2020; Boyd et al. 2021], together with Sroka's vanishing result. In this paper, we prove homological stability for the partition algebras, and we identify their stable homology.

MSC2020: primary 16E40, 20J06; secondary 20B30.
Keywords: homology, homological stability, partition algebras.

The *partition algebras* were introduced independently by Jones [1994] and Martin [1994] for their relevance in studying Potts models in statistical mechanics. They are also important in representation theory as a Schur–Weyl dual to the symmetric group, as in the work of Halverson and Ram [2005, Theorems 5.4, 3.6] and Bowman, Doty and Martin [2022]. They contain a rich variety of subalgebras, including the planar partition, rook Brauer, rook, planar rook, Brauer, Motzkin and Temperley–Lieb algebras.

Given a commutative ring R, an element $\delta \in R$, and a nonnegative integer n, the partition algebra $\mathrm{P}_n(R, \delta)$ is defined to be the free module over R with basis given by the partitions of the set $\{-1, \dots, -n, 1, \dots, n\}$. These partitions can be drawn as diagrams with n nodes labelled $-1, \dots, -n$ on the left and n nodes labelled $1, \dots, n$ on the right. Nodes in the same block of a partition are then joined by edges. For ease of drawing, we do not include all edges but instead rely on transitivity. Disconnected nodes are allowed, corresponding to blocks of size one. For example, the following diagram shows the basis element $\{\{-1, -3\}, \{-2, -4, 4\}, \{2, 3\}, \{1\}\}$ of $\mathrm{P}_4(R, \delta)$:



Multiplication is given by placing the diagrams side by side, identifying the middle nodes, and replacing any blocks not connected to the right or left by a factor of δ.

Diagrams in which every node on the left is connected to a single node on the right, and nothing else, are called *permutation diagrams*, and are in bijection with elements of the symmetric group $\mathfrak{S}_n$. This gives rise to inclusion and projection maps

$$R\mathfrak{S}_n \xrightarrow{\iota} \mathrm{P}_n(R, \delta) \xrightarrow{\pi} R\mathfrak{S}_n$$

where ι sends permutations to permutation diagrams, and π does the reverse while sending all remaining diagrams to 0. In particular, $\pi \circ \iota$ is the identity map on $R\mathfrak{S}_n$.

We denote the trivial module of $R\mathfrak{S}_n$ by $\mathbb{1}$. Pulling back along π, we obtain the *trivial module* $\mathbb{1}$ of $\mathrm{P}_n(R, \delta)$. This gives us the homology groups $H_*(\mathfrak{S}_n, \mathbb{1}) = \mathrm{Tor}_*^{R\mathfrak{S}_n}(\mathbb{1}, \mathbb{1})$ of $\mathfrak{S}_n$ and $\mathrm{Tor}_*^{\mathrm{P}_n(R,\delta)}(\mathbb{1}, \mathbb{1})$ of $\mathrm{P}_n(R, \delta)$. There are induced homomorphisms ι_* and π_* on homology groups for which $\pi_* \circ \iota_*$ is again the identity, so that the homology of $\mathfrak{S}_n$ appears as a direct summand of the homology of $\mathrm{P}_n(R, \delta)$.

Theorem A. *Suppose that δ is invertible in R. Then the homology of the partition algebra is isomorphic to the homology of the symmetric group*:

$$\mathrm{Tor}_*^{\mathrm{P}_n(R,\delta)}(\mathbb{1}, \mathbb{1}) \cong H_*(\mathfrak{S}_n; \mathbb{1}).$$

Indeed, the inclusion and projection maps

$$R\mathfrak{S}_n \xrightarrow{\iota} \mathrm{P}_n(R,\delta) \xrightarrow{\pi} R\mathfrak{S}_n$$

induce inverse isomorphisms

$$\mathrm{Tor}_*^{R\mathfrak{S}_n}(\mathbb{1},\mathbb{1}) \xrightarrow[\cong]{\iota_*} \mathrm{Tor}_*^{\mathrm{P}_n(R,\delta)}(\mathbb{1},\mathbb{1}) \xrightarrow[\cong]{\pi_*} \mathrm{Tor}_*^{R\mathfrak{S}_n}(\mathbb{1},\mathbb{1}).$$

Our second result holds without any assumptions on the value of δ.

Theorem B. *The inclusion map* $\iota : R\mathfrak{S}_n \to \mathrm{P}_n(R,\delta)$ *induces a map in homology*

$$\iota_* : H_i(\mathfrak{S}_n;\mathbb{1}) \to \mathrm{Tor}_i^{\mathrm{P}_n(R,\delta)}(\mathbb{1},\mathbb{1})$$

that is an isomorphism in the range $n \geq 2i+1$.

An immediate consequence of Theorem B is the following corollary.

Corollary C. *The partition algebras satisfy homological stability, that is, the inclusion* $\mathrm{P}_{n-1}(R,\delta) \hookrightarrow \mathrm{P}_n(R,\delta)$ *induces a map*

$$\mathrm{Tor}_i^{\mathrm{P}_{n-1}(R,\delta)}(\mathbb{1},\mathbb{1}) \to \mathrm{Tor}_i^{\mathrm{P}_n(R,\delta)}(\mathbb{1},\mathbb{1})$$

that is an isomorphism in degrees $n \geq 2i+1$, *and this stable range is sharp. Furthermore,* $\mathrm{P}_n(R,\delta)$ *and* $\mathfrak{S}_n$ *have the same stable homology:*

$$\lim_{n\to\infty} H_*(\mathfrak{S}_n;\mathbb{1}) \cong \lim_{n\to\infty} \mathrm{Tor}_*^{\mathrm{P}_n(R,\delta)}(\mathbb{1},\mathbb{1}).$$

The first part of this corollary follows by combining Theorem B with Nakaoka's homological stability result for the symmetric groups, for which the stable range is sharp [Nakaoka 1960]. For the stable homology, the left-hand side of this isomorphism is well known by the Barratt–Priddy–Quillen theorem [Barratt and Priddy 1972; Friedlander and Mazur 1994]. The above results exactly parallel the situation for the Brauer algebras, and as discussed in [Boyd et al. 2021] are reminiscent of the relationship between $\mathfrak{S}_n$ and the automorphism groups of free groups $\mathrm{Aut}(F_n)$ (see Galatius [2011]).

1A. ***Outline, and comparison to previous work.*** In Section 2 we introduce partition algebras and provide the necessary background needed for the rest of the paper. In Section 3 we restate an abstract form of the principle that lies behind the technique of inductive resolutions that was introduced in [Boyd and Hepworth 2020], and was a crucial ingredient in [Boyd and Hepworth 2020] and [Boyd et al. 2021]. In Section 4 we establish the existence of inductive resolutions for the partition algebras. These are significantly more complicated than the Temperley–Lieb [Boyd and Hepworth 2020] and Brauer [Boyd et al. 2021] cases, and we find that we must consider several families of distinct modules in order to carry out our induction argument. In Section 5 we follow the argument of Boyd et al. [2021] to replace

Shapiro's lemma in the setting of partition algebras. The high connectivity result required for any new proof of homological stability is found in Section 6. Like our inductive resolutions argument, this is again more complicated than the analogous result in [Boyd et al. 2021], and heavily utilises the high connectivity of the complex of injective words with separators, introduced in that paper. We finish in Section 7 by giving an account of the proof of the main theorem, which follows the same general argument as in [Boyd et al. 2021].

It is common, in homological stability for families of groups, to find that proofs of different results have a very similar overall structure, yet the proofs that the relevant complexes are highly acyclic can differ radically. What we can now see in homological stability for algebras, comparing the work of this paper to that of [Boyd and Hepworth 2020] and [Boyd et al. 2021], is an analogous situation where the overall technique is used in multiple situations, but the details of the acyclicity proofs — and now also of the inductive resolutions proofs — are where the important differences and difficulties lie.

2. Partition algebras

In this section we introduce the partition algebra, together with some specific elements and modules that will be important later in the paper.

Definition 2.1 (the partition algebra [Jones 1994; Martin 1994]). As explained in the introduction, if R is a commutative ring, δ is a chosen element in R, and n is a nonnegative integer, then the partition algebra $\mathrm{P}_n(R, \delta)$ is defined to be the free module over R with basis given by the partitions of the set $\{-n, \dots, -1, 1, \dots, n\}$. These are drawn as diagrams with nodes $-1, \dots, -n$ on the left and nodes $1, \dots, n$ on the right, with arcs indicating which nodes lie in the same block of the partition. (We allow ourselves to omit some arcs and instead use transitivity to determine the blocks.) An example is shown in Figure 1. Multiplication is given by placing the diagrams side by side, identifying the middle nodes, and replacing any blocks not connected to the right or left by a factor of δ, as in Figure 2.

We will use the terms "partition" and "diagram" interchangeably to mean a basis element of $\mathrm{P}_n(R, \delta)$, and we will frequently abbreviate $\mathrm{P}_n(R, \delta)$ as P_n.

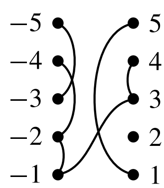


Figure 1. Visualization of the partition $\{\{-5, -3\}, \{-4, -2, -1, 3, 4\}, \{1, 5\}, \{2\}\}$.

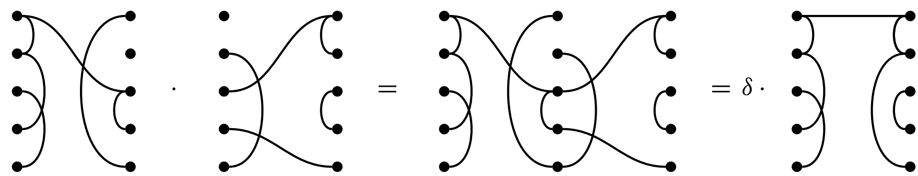


Figure 2. Multiplication in the partition algebra.

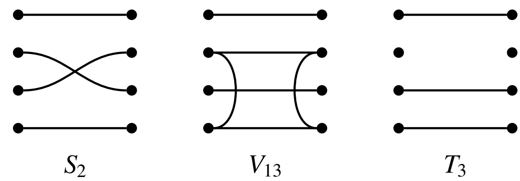


Figure 3. The elements S_2, V_{13}, $T_3 \in \mathrm{P}_4$.

The partition algebra is generated by three types of diagrams [Martin 1996], corresponding to the following partitions:

- For $1 \le i \le n-1$, S_i is the diagram corresponding to the partition with blocks of pairs $\{-j, j\}$ for $j \ne i, i+1$, together with $\{-(i+1), i\}$ and $\{-i, (i+1)\}$. These generate the group ring of the symmetric group, $\mathfrak{S}_n$, as a subalgebra of P_n.
- For $1 \le i \ne j \le n-1$, V_{ij} is the diagram corresponding to the partition with blocks of pairs $\{-k, k\}$ for $k \ne i, j$ and one block of size four $\{-j, -i, i, j\}$.
- For $1 \le i \le n$, T_i is the diagram corresponding to the partition with blocks of pairs $\{-j, j\}$ for $j \ne i$ and two singleton blocks $\{-i\}$ and $\{i\}$.

See Figure 3 for depictions of some of these.

We now introduce the modules we will be working with.

Recall that by a *permutation diagram* we mean a diagram in which each node on the left is joined to a single node on the right, and nothing else. Equivalently, permutation diagrams are ones that do not contain any singletons on the right *or* any blocks that contain ≥ 2 elements on the right.

Definition 2.2 (the trivial module $\mathbb{1}$). For any n, we define the *trivial* $R\mathfrak{S}_n$-bimodule $\mathbb{1}$ to be the module given by the ring R, upon which the permutations act as the identity.

For any n, we define the *trivial* P_n-bimodule $\mathbb{1}$ to be the module given by the ring R, upon which the permutation diagrams act as the identity, and all other diagrams act as 0. This is the same as acting with P_n on R via the projection $\pi : \mathrm{P}_n \to \mathfrak{S}_n$.

Definition 2.3. For $m \le n$, we can view P_m as a subalgebra of P_n. Given a partition of $\{-m, \dots, -1, 1, \dots, m\}$, the map which sends $(\pm 1, \dots, \pm m)$ to $(\pm(n-m+1), \dots, \pm n)$ induces a partition on $\{-n \dots, -(n-m+1), (n-m+1), \dots, n\}$.

We add the blocks $\{-i, i\}$ for all $i \in \{1, \dots, (n-m)\}$, resulting in a partition in P_n. Pictorially, we are taking diagrams in P_m and extending them to ones in P_n by adding new nodes *below* the existing ones, with horizontal connections between the new nodes. Then, under the action of this subalgebra, P_n can be viewed as a left P_n-module and a right P_m-module, and we obtain the induced left P_n-module $\mathrm{P}_n \otimes_{\mathrm{P}_m} \mathbb{1}$.

Proposition 2.4 [Patzt 2024, Proposition 2.5]. *The induced module* $\mathrm{P}_n \otimes_{\mathrm{P}_m} \mathbb{1}$ *is a free* R*-module and a quotient of* P_n.

In terms of diagrams, a basis for this module is the set of diagrams in which the top m *nodes on the right are placed under a box, satisfying the following condition:*

- *The box is connected to exactly* m *distinct blocks.*

Under this description, the action of P_n *on* $\mathrm{P}_n \otimes_{\mathrm{P}_m} \mathbb{1}$ *is given by pasting and simplifying the diagrams just as in the multiplication of* P_n, *and then identifying a diagram with* 0 *if it violates the condition above.*

Thus there are two ways that a diagram could be identified with 0 after left multiplication by a diagram in P_n: One of the blocks attached to the box could, after pasting, consist only of nodes in the centre (visually, that block is free to be retracted into the box, and then disappears). Alternatively, two or more distinct blocks that were attached to the box can become fused into a single block (visually, there is now a path of arcs with both ends attached to the box). These two possibilities correspond to the two ways in which a diagram in P_m can fail to be a permutation diagram, and therefore act as 0 on $\mathbb{1}$: It can have a singleton on the right, or it can have two nodes on the right belonging to the same block.

Example 2.5. Figure 4 depicts the module structure of $\mathrm{P}_5 \otimes_{\mathrm{P}_3} \mathbb{1}$. In the first example one of the blocks connected to the box consists entirely of nodes in the centre and therefore "vanishes" or "retracts into the box". In the second example the factor of δ arises due to a block that consists entirely of nodes in the centre and is not attached to the box.

3. The principle of inductive resolutions

In this brief section we state an abstract form of the principle that underlies the technique of inductive resolutions that appeared in [Boyd and Hepworth 2020] and [Boyd et al. 2021]. It allows us to identify modules that vanish under a fixed functor of the form $\mathrm{Tor}_i^A(M, -)$ by resolving them using modules that already have this property, hence the name "inductive resolutions". The theorem below is an abstraction of Section 3.3 of [Boyd and Hepworth 2020]. It can be regarded as an application of the general principle that a derived functor can be computed using resolutions by objects that are acyclic for that derived functor.

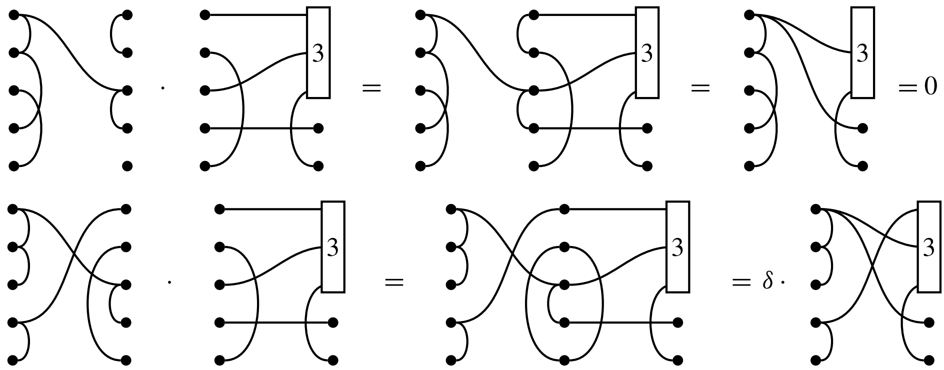


Figure 4. The module structure of $\mathrm{P}_5 \otimes_{\mathrm{P}_3} \mathbb{1}$.

Theorem 3.1. *Let A be an algebra over a ring R, and let M be a right A-module. Suppose that N is a left A-module equipped with a resolution $Q_* \to N$ with the following two properties*:

- $\mathrm{Tor}_*^A(M, Q_j)$ *vanishes in positive degrees for all* $j \geqslant 0$.
- $M \otimes_A Q_* \to M \otimes_A N$ *is a resolution.*

Then $\mathrm{Tor}_*^A(M, N)$ *vanishes in positive degrees.*

Proof. Let $P_* \to M$ be a projective resolution, so that for any left A-module B, the groups $\mathrm{Tor}_*^A(M, B)$ are computed by the complex $P_* \otimes_A B$. Consider the double complex $P_* \otimes_A Q_*$. There are two natural spectral sequences converging to the homology of the totalisation $\mathrm{Tot}(P_* \otimes_A Q_*)$. For more on these spectral sequences, see Section 5.6 of [Weibel 1994] and the summary in Section 3.2 of [Boyd and Hepworth 2020].

The first spectral sequence has E^1-term given by

$$^{I}E^1_{i,j} = H_j(P_i \otimes_A Q_*) \cong P_i \otimes_A H_j(Q_*),$$

with d^1 induced by the differential of P. The isomorphism holds because each P_i is projective and therefore flat. It follows that the E^2-term is

$$^{I}E^2_{i,j} = \mathrm{Tor}_i^A(M, H_j(Q_*)).$$

Since Q_* is a resolution of N, it follows that $^{I}E^2_{*,*}$ is simply $\mathrm{Tor}_*^A(M, N)$ concentrated on the x-axis, so that the same is true of $^{I}E^\infty_{*,*}$, and therefore we conclude that $H_*(\mathrm{Tot}(P_* \otimes_A Q_*)) \cong \mathrm{Tor}_*^A(M, N)$.

The second spectral sequence has E^1-term given by $^{II}E^1_{i,j} = H_j(P_* \otimes_A Q_i)$, i.e.,

$$^{II}E^1_{i,j} = \mathrm{Tor}_j^A(M, Q_i),$$

with d^1 induced by the boundary maps of Q_*. Our first assumption now shows that $^{II}E^1_{*,*}$ is concentrated on the x-axis, where it is given by $\mathrm{Tor}_0^A(M, Q_*) = M \otimes_A Q_*$.

Consequently ${}^{II}E^2_{*,*}$ is given by the homology of $M \otimes_A Q_*$, which by our second assumption is simply a copy of $M \otimes_A N$ at the origin. This shows that the homology of $\mathrm{Tot}(P_* \otimes_A Q_*)$ is simply a copy of $M \otimes_A N$ in degree 0.

Comparing the outcomes of the two spectral sequences, we see that $\mathrm{Tor}^A_*(M, N)$ vanishes in positive degrees, as required. □

4. Inductive resolutions

In this section, we will use the technique of *inductive resolutions*, which originated in [Boyd and Hepworth 2020] and was further used in [Boyd et al. 2021].

Definition 4.1. Suppose that X is a subset of the set $\{1, \dots, n\}$. Define J_X to be the left-ideal in P_n that is the R-span of all diagrams in which, among the nodes on the right labelled by elements of X, there is at least one singleton or one pair of nodes that are in the same block. For $m \le n$, let $J_{\{n-m+1,\dots,n\}}$ be denoted by J_m.

Observe that $J_n = J_{\{1,\dots,n\}}$ is the span of precisely the diagrams that are *not* permutation diagrams. It is therefore the kernel of the projection map $\pi : \mathrm{P}_n \to R\mathfrak{S}_n$.

Our aim in this section is to prove the following theorem, which will be used in the final section to understand the Tor groups $\mathrm{Tor}^{\mathrm{P}_n}_*(\mathbb{1}, \mathrm{P}_n \otimes_{\mathrm{P}_m} \mathbb{1})$.

Theorem 4.2. *Let $X \subseteq \{1, \dots, n\}$ and suppose that one of the following conditions holds*:

- *$|X| \le n$ and δ is invertible in R.*
- *$|X| < n$.*

Then the groups $\mathrm{Tor}^{\mathrm{P}_n}_*(\mathbb{1}, \mathrm{P}_n/J_X)$ vanish in positive degrees.

The proof of Theorem 4.2 will occupy the rest of the section. Aspects of the material are close to [Boyd and Hepworth 2020, Section 3] and [Boyd et al. 2021, Section 3], but overall the material here is significantly more complex.

Before we continue, we record the following lemma, which extends Theorem 4.2 to degree 0. We will need this lemma to prove the theorem.

Lemma 4.3. *Let J be a left ideal of P_n that is included in J_n. Then*

$$\mathbb{1} \otimes_{\mathrm{P}_n} \mathrm{P}_n/J \cong \mathbb{1}.$$

In particular,

$$\mathrm{Tor}^{\mathrm{P}_n}_0(\mathbb{1}, \mathrm{P}_n/J_X) \cong \mathbb{1}$$

for all $X \subset \{1, \dots, n\}$.

Proof. Due to the inclusions $0 \subset J \subset J_n$, we have the surjections

$$\mathbb{1} \otimes_{\mathrm{P}_n} \mathrm{P}_n \twoheadrightarrow \mathbb{1} \otimes_{\mathrm{P}_n} \mathrm{P}_n/J \twoheadrightarrow \mathbb{1} \otimes_{\mathrm{P}_n} \mathrm{P}_n/J_n.$$

Because

$$\mathbb{1} \otimes_{\mathrm{P}_n} \mathrm{P}_n \cong \mathbb{1} \quad \text{and} \quad \mathbb{1} \otimes_{\mathrm{P}_n} \mathrm{P}_n/J_n \cong \mathbb{1} \otimes_{\mathrm{P}_n} R\mathfrak{S}_n \cong \mathbb{1} \otimes_{R\mathfrak{S}_n} R\mathfrak{S}_n \cong \mathbb{1},$$

the above composition is an isomorphism and the first map must be also injective. □

4A. *Reducing to $\mathcal{A}_{X,x}$ and $\mathcal{B}_{X,x}$.* Our proof of Theorem 4.2 will be by induction on the cardinality of X. Ideally we would prove the inductive step by resolving P_n/J_X in terms of modules $\mathrm{P}_n/J_{X'}$ with $|X'| < |X|$. However, we were not able to find a straightforward argument along these lines. To organise the argument, in this subsection we introduce some intermediate modules, and later on we will build our resolutions with these.

Definition 4.4. Let $Y \subseteq X \subseteq \{1, \ldots, n\}$, and let $x \in X$ and $y \in Y$. We define three left P_n-submodules of P_n:

- A_x is the span of all diagrams in which x is a singleton.
- $B_{X,x}$ is the span of all diagrams in which x lies in the same block as some other element of X.
- M_Y is the span of all diagrams in which the elements of Y lie in the same block.

We define quotients of these as follows:

$$\mathcal{A}_{X,x} = \frac{A_x}{A_x \cap J_{X-\{x\}}}, \quad \mathcal{B}_{X,x} = \frac{B_{X,x}}{B_{X,x} \cap J_{X-\{x\}}}, \quad \mathcal{M}_{X,Y} = \frac{M_Y}{M_Y \cap J_{X-Y}}$$

The following result will be useful to verify the second condition of Theorem 3.1.

Lemma 4.5. *The modules $\mathcal{A}_{X,x}$, $\mathcal{B}_{X,x}$, and $\mathcal{M}_{X,Y}$ behave as follows under tensor product with $\mathbb{1}$.*

- *Let $x \in X \subseteq \{1, \ldots, n\}$, and suppose that $n \geq 2$. Then $\mathbb{1} \otimes_{\mathrm{P}_n} \mathcal{A}_{X,x} = 0$.*
- *Let $x \in X \subseteq \{1, \ldots, n\}$. Then $\mathbb{1} \otimes_{\mathrm{P}_n} \mathcal{B}_{X,x} = 0$.*
- *Let $Y \subseteq X \subseteq \{1, \ldots, n\}$ with $|Y| \geq 2$. Then $\mathbb{1} \otimes_{\mathrm{P}_n} \mathcal{M}_{X,Y} = 0$ and $\mathcal{M}_{X,Y}$ is a direct summand of P_n/J_{X-Y}.*

Proof. We will show that, under the relevant conditions, each of A_x, $B_{X,x}$ and M_Y vanishes under $\mathbb{1} \otimes_{\mathrm{P}_n} -$, and the same will then follow for $\mathcal{A}_{X,x}$, $\mathcal{B}_{X,x}$ and $\mathcal{M}_{X,Y}$.

To show that $\mathbb{1} \otimes_{\mathrm{P}_n} A_x = 0$, we take a diagram α in A_x, so that x is a singleton in α. Let β denote a diagram obtained from α by placing x into the same block as some other element on the right. (This is possible by the assumption that $n \geq 2$.) Then $\alpha = \beta \cdot T_x$, and β acts as 0 on $\mathbb{1}$, so that

$$1 \otimes \alpha = 1 \otimes \beta \cdot T_x = 1 \cdot \beta \otimes T_x = 0 \otimes T_x = 0,$$

noting that $T_x \in A_x$. Since A_x is the span of such diagrams α, this completes the proof.

The argument for the other two modules is similar. For $\mathbb{1} \otimes_{\mathrm{P}_n} B_{X,x}$, we take a diagram $\alpha \in B_{X,x}$, so that x is in the same block as some other element $w \in X$, and we use the factorisation $\alpha = \alpha \cdot V_{wx}$, noting that $V_{wx} \in B_{X,x}$.

For M_Y we take a diagram $\alpha \in M_Y$, so that all elements of Y lie in the same block, and factorise it as $\alpha = \alpha \cdot V_Y$ where $V_Y \in M_Y$ is the diagram with blocks $-Y \cup Y$ and $\{-p, p\}$ for $p \notin Y$; the assumption $|Y| \geq 2$ ensures that α acts as 0 on $\mathbb{1}$.

For the final claim about $\mathcal{M}_{X,Y}$, we use the fact that the element V_Y above is idempotent and sends J_{X-Y} into itself. $\square$

The following proposition breaks down the problem of resolving P_n/J_X into the analogous problem for $\mathcal{A}_{X,x}$ and $\mathcal{B}_{X,x}$.

Proposition 4.6. *Let $X \subseteq \{1, \dots, n\}$, let $x \in X$, and assume $n \geq 2$. The sequence below, in which all maps are induced by either an inclusion or an identity map, is a resolution of P_n/J_X:*

$$\begin{array}{ccccccccc} \dots & \longrightarrow & 0 & \longrightarrow & \mathcal{A}_{X,x} \oplus \mathcal{B}_{X,x} & \longrightarrow & \mathrm{P}_n/J_{X-\{x\}} & \longrightarrow & \mathrm{P}_n/J_X \\ & & 2 & & 1 & & 0 & & -1 \end{array}$$

Moreover, applying $\mathbb{1} \otimes_{\mathrm{P}_n} -$ to the sequence gives a resolution of $\mathbb{1} \otimes_{\mathrm{P}_n} \mathrm{P}_n/J_X$.

Proof. The map $\mathrm{P}_n/J_{X-\{x\}} \to \mathrm{P}_n/J_X$ is induced by the identity map on P_n and is well defined since $J_{X-\{x\}} \subset J_X$. The map $\mathcal{A}_{X,x} \to \mathrm{P}_n/J_{X-\{x\}}$ is induced by the inclusion $A_x \subset \mathrm{P}_n$ and is well defined since $(A_x \cap J_{X-\{x\}}) \subset J_{X-\{x\}}$, and a similar argument holds for the map $\mathcal{B}_{X,x} \to \mathrm{P}_n/J_{X-\{x\}}$.

Surjectivity of the right-hand map is immediate, giving exactness in degree -1.

To show exactness in degree 0, observe that the ideals $J_{X-\{x\}} \subseteq J_X$ are both spanned by certain diagrams, so that the kernel of

$$\mathrm{P}_n/J_{X-\{x\}} \to \mathrm{P}_n/J_X$$

is spanned by those diagrams that lie in J_X but not $J_{X-\{x\}}$. For a diagram to lie in J_X, some element of X must be a singleton, or two elements of X must lie in the same block. For it to not *also* be an element of $J_{X-\{x\}}$, there must only be one singleton, namely x, or only one pair of elements lying in the same block, of which one must be x. The diagrams with x a singleton are precisely the diagrams that span A_x, the diagrams in which x lies in the same block as some other element of X are precisely those that span $B_{X,x}$, and the proof follows.

To show exactness in degree 1, after unravelling the definitions of $\mathcal{A}_{X,x}$ and $\mathcal{B}_{X,x}$, it is sufficient to show that if we have $a \in A_x$ and $b \in B_{X,x}$ with $a + b \in J_{X-\{x\}}$, then $a, b \in J_{X-\{x\}}$. This follows quickly from fact that A_x and $B_{X,x}$ have no basis elements in common.

To prove the second claim, we will show that after applying $\mathbb{1}\otimes_{\mathrm{P}_n}-$, the resolution becomes

$$\begin{array}{ccccccccc} \dots & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & \mathbb{1} & \xrightarrow{\mathrm{Id}} & \mathbb{1} \\ & & 2 & & 1 & & 0 & & -1 \end{array}$$

so that the claim follows directly. The identification of the final two terms and the map between them follows from Lemma 4.3. The terms in degree 1 vanish by Lemma 4.5. (This is where we use the assumption $n \geq 2$.) $\square$

The last result, together with Theorem 3.1, shows that, in order to prove vanishing of higher Tor's for P_n/J_X by induction, we must first do the same for $\mathcal{A}_{X,x}$ and $\mathcal{B}_{X,x}$. In the next two subsections we will construct resolutions of these.

4B. *Resolving $\mathcal{A}_{X,x}$.* We now attempt to resolve $\mathcal{A}_{X,x}$. It will turn out that this requires different methods depending on which assumption from Theorem 4.2 we use: that δ is invertible, or that $|X| < n$. Under the first assumption we have:

Proposition 4.7. *Suppose that $X \subseteq \{1, \dots, n\}$ and that δ is invertible in R. Then the module $\mathcal{A}_{X,x}$ is a direct summand of $\mathrm{P}_n/J_{X-\{x\}}$.*

Proof. The element $\delta^{-1}T_x$ is an idempotent, thanks to the computation $T_x^2 = \delta T_x$. Right-multiplication by $\delta^{-1}T_x$ sends $J_{X-\{x\}}$ into itself, and therefore induces an idempotent endomorphism of $\mathrm{P}_n/J_{X-\{x\}}$. The image of this endomorphism consists of all left multiples of T_x, and this is precisely $\mathcal{A}_{X,x} = \frac{A_x}{A_x \cap J_{X-\{x\}}}$ as in the second paragraph of the proof of Lemma 4.5. $\square$

The above result shows that, if $\mathrm{P}_n/J_{X-\{x\}}$ has vanishing higher Tor's, then so does $\mathcal{A}_{X,x}$. When δ is not invertible, we need a more elaborate method using the following resolution.

Definition 4.8 (the resolution $C(X, x, y)$). Let $X \subset \{1, \dots, n\}$ with $|X| < n$, let $x \in X$, and let $y \in \{1, \dots, n\} - X$. We define $C(X, x, y) \to \mathcal{A}_{X,x}$ as in Figure 5.

Thus $C(X, x, y)$ is given by $\mathrm{P}_n/J_{X-\{x\}}$ in each degree. The maps are all given by right-multiplication by the indicated elements, so that the boundary maps alternate between $(1 - T_x V_{xy})$ and $T_x V_{xy}$, and the augmentation $\mathrm{P}_n/J_{X-\{x\}} \to \mathcal{A}_{X,x}$ is given by T_x. The maps are well defined thanks to the fact that $T_x V_{xy}$ and T_x send $J_{X-\{x\}}$ into itself; in the former case this follows from the fact that $y \notin X$.

To check that consecutive maps compose to 0, one uses $T_x V_{xy} T_x = T_x$ together with the resulting fact that $T_x V_{xy}$ is an idempotent. The fact that this really does define a resolution is given next.

Proposition 4.9. *Suppose that $X \subset \{1, \dots, n\}$ with $|X| < n$, let $x \in X$, and let $y \in \{1, \dots, n\} - X$. Assume that $n \geq 2$. Then*

$$C(X, x, y) \to \mathcal{A}_{X,x} \quad \textit{and} \quad \mathbb{1} \otimes_{\mathrm{P}_n} C(X, x, y) \to \mathbb{1} \otimes_{\mathrm{P}_n} \mathcal{A}_{X,x}$$

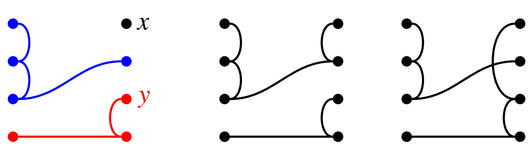
are both resolutions.

$$
\begin{tikzcd}
\vdots \arrow[d, "(1-T_x V_{xy})"'] & \\
\mathrm{P}_n/J_{X-\{x\}} \arrow[d, "T_x V_{xy}"'] & 2 \\
\mathrm{P}_n/J_{X-\{x\}} \arrow[d, "(1-T_x V_{xy})"'] & 1 \\
\mathrm{P}_n/J_{X-\{x\}} \arrow[d, "T_x"'] & 0 \\
\mathcal{A}_{X,x} & -1
\end{tikzcd}
$$

Figure 5. The resolution $C(X,x,y) \to \mathcal{A}_{X,x}$.

Proof. First, we must show that $C(X,x,y) \to \mathcal{A}_{X,x}$ is acyclic. In degree -1 this is clear since A_x consists of all left multiples of T_x. In degrees 1 and above, this is an immediate consequence of the fact that $T_x V_{xy}$ is an idempotent. In degree 0 we require a more complex argument, as follows.

Suppose α is a diagram in which x is a singleton. If B is a block of α other than $\{x\}$, then we write α_B for the diagram obtained from α by incorporating x into B. And we write α_y for the diagram α_{B_y}, where B_y is the block containing y. For example, the following diagrams show α with B_y and another block B, together with α_B and α_y.



α $\qquad$ α_B $\qquad$ α_y

Now observe that we have the relations

$$\alpha - \delta\alpha_y = \alpha(1 - T_x V_{xy}) \tag{4-1}$$

and, for each block B in α,

$$\alpha_B - \alpha_y = \alpha_B(1 - T_x V_{xy}). \tag{4-2}$$

Now consider an element $a \in \mathrm{P}_n/J_{X-\{x\}}$ in degree 0 that lies in the kernel of the augmentation $T_x : \mathrm{P}_n/J_{X-\{x\}} \to \mathcal{A}_{X,x}$. We wish to show that a is in the image of the differential, and we will do this by explaining how to adjust a by elements in the image of the differential (which does not change the fact that it lies in the kernel) in order to reduce it to 0. We can write a as a linear combination of diagrams in P_n that do not lie in $J_{X-\{x\}}$, and these can be divided into the following cases:

(1) Diagrams α in which x is a singleton. Using elements of the form (4-1), we may adjust a by elements in the image of the differential in order to replace all such diagrams α with ones of the form α_y.

(2) Diagrams in which x is connected to some element outwith $X-\{x\}$. These diagrams all have the form α_B, where α is the diagram obtained from the original by making x a singleton, and B is the block of α that originally contained x. Note that the assumption that the original diagram did not lie in $J_{X-\{x\}}$ means that α also does not lie in $J_{X-\{x\}}$. Using elements of the form (4-2), we may adjust a by elements in the image of the differential in order to replace all such diagrams α_B with ones of the form α_y.

(3) Diagrams β in which x is connected to exactly one element, say w, in $X-\{x\}$. Then in $\beta T_x V_{xy}$ the element $w\in X-\{x\}$ is a singleton, so that $\beta T_x V_{xy}\in J_{X-\{x\}}$. Consequently $\beta=\beta(1-T_xV_{xy})$ in $\mathrm{P}_n/J_{X-\{x\}}$, and in particular β lies in the image of the differential. We may therefore adjust a by elements in the image of the differential to remove all diagrams of this form.

After modifying a as explained in each item above, we may now write it as a linear combination $a=\sum_\alpha \lambda_\alpha\,\alpha_y$ where α ranges over all diagrams that are not in $J_{X-\{x\}}$ and in which x is a singleton. We know that a lies in the kernel of the differential, so that $a\cdot T_x=0$. However, we have $a\cdot T_x=\sum_\alpha\lambda_\alpha\,\alpha_y\cdot T_x=\sum_\alpha\lambda_\alpha\,\alpha$, and since the α are distinct diagrams not in $J_{X-\{x\}}$, we can conclude that $\lambda_\alpha=0$ for all α, or in other words that $a=0$. This completes the argument in degree 0, and so completes the proof that $C(X,x,y)\to\mathcal{A}_{X,x}$ is a resolution.

We now prove that $\mathbb{1}\otimes_{\mathrm{P}_n}C(X,x,y)\to\mathbb{1}\otimes_{\mathrm{P}_n}\mathcal{A}_{X,x}$ is a resolution. The target vanishes by Lemma 4.5, and $\mathbb{1}\otimes_{\mathrm{P}_n}\mathrm{P}_n/J_{X-\{x\}}=\mathbb{1}$ since $J_{X-\{x\}}$ acts as 0 on $\mathbb{1}$. Under the latter identification, the boundary maps, which used to be given by right-multiplication by the indicated elements, are now given by the action of those elements on $\mathbb{1}$, and therefore alternate between 0 and 1. The result follows. □

4C. *Resolving $\mathcal{B}_{X,x}$*. We now turn to the module $\mathcal{B}_{X,x}$, for which we build the following resolution.

Definition 4.10 (the resolution $D(X,x)\to\mathcal{B}_{X,x}$). Let $X\subseteq\{1,\dots,n\}$ and let $x\in X$. Define an augmented complex $D(X,x)\to\mathcal{B}_{X,x}$ as follows. In degree $i\geq 0$, $D(X,x)$ is given by

$$\bigoplus_{(x_0,\dots,x_i)}\mathcal{M}_{X,\{x,x_0,\dots,x_i\}},$$

$$\begin{array}{cc}
\vdots & \\
\Big\downarrow \delta_3 & \\
\bigoplus_{(x_0,x_1,x_2)} \mathcal{M}_{X,\{x,x_0,x_1,x_2\}} & 2 \\
\Big\downarrow \delta_2 & \\
\bigoplus_{(x_0,x_1)} \mathcal{M}_{X,\{x,x_0,x_1\}} & 1 \\
\Big\downarrow \delta_1 & \\
\bigoplus_{(x_0)} \mathcal{M}_{X,\{x,x_0\}} & 0 \\
\Big\downarrow \delta_0 & \\
\mathcal{B}_{X,x} & -1
\end{array}$$

Figure 6. The resolution $D(X, x) \to \mathcal{B}_{X,x}$. Summations are over tuples of distinct elements of $X - \{x\}$.

where the sum is over all tuples $(x_0, \dots, x_i)$ of *distinct* elements of $X - \{x\}$. And on the summand corresponding to a tuple $(x_0, \dots, x_i)$, the map δ_i is given by the map

$$\mathcal{M}_{X,\{x,x_0,\dots,x_i\}} \to \mathcal{M}_{X,\{x,x_0,\dots,x_{i-1}\}}$$

obtained from the inclusions

$$M_{\{x,x_0,\dots,x_i\}} \hookrightarrow M_{\{x,x_0,\dots,x_{i-1}\}}, \qquad J_{X-\{x,x_0,\dots,x_i\}} \hookrightarrow J_{X-\{x,x_0,\dots,x_{i-1}\}}.$$

The map δ_i is simply the sum of these individual maps. To it put briefly, δ_i is the map that forgets that x_i had to be in the same block as $x, x_0, \dots, x_{i-1}$. The complex is illustrated in Figure 6.

Each composite $\delta_{i-1} \circ \delta_i$ vanishes, because any element in its image is a sum of diagrams that each contain two elements $x_{i-1}, x_i \in X - \{x, x_1, \dots, x_{i-2}\}$ in the same block, and which therefore lie in $J_{X-\{x,x_1,\dots,x_{i-2}\}}$. We prove that this is indeed a resolution in Proposition 4.11.

Proposition 4.11. *$D(X, x) \to \mathcal{B}_{X,x}$ is indeed a resolution, and the same is true for $\mathbb{1} \otimes_{\mathrm{P}_n} D(X, x) \to \mathbb{1} \otimes_{\mathrm{P}_n} \mathcal{B}_{X,x}$.*

Proof. We first prove that $D(X, x) \to \mathcal{B}_{X,x}$ is acyclic.

In degree -1 we must show that δ_0 is surjective. A diagram in $B_{X,x}$ has x in the same block as some other element x_0 of $X-\{x\}$, and therefore lies in the image of the inclusion $M_{\{x,x_0\}} \hookrightarrow B_{X,x}$, and surjectivity follows.

In degree 0, we first observe that if we consider any two summands in degree 0, then their images under δ_0 have trivial intersection. Indeed, this follows quickly from the fact that if x_0 and x_0' are distinct elements of $X-\{x\}$, then

$$M_{\{x,x_0\}} \cap M_{\{x,x_0'\}} \subseteq J_{X-\{x\}},$$

which itself holds because a diagram in $M_{\{x,x_0\}} \cap M_{\{x,x_0'\}}$ has x_0 and x_0' in the same block as x, and therefore in the same block as one another. So to prove exactness in degree 0 we can look at just one x_0-summand at a time:

$$\begin{array}{cc}
\displaystyle\bigoplus_{x_1\in X-\{x,x_0\}} \mathcal{M}_{X,\{x,x_0,x_1\}} & 1 \\
\big\downarrow{\scriptstyle \delta_1} & \\
\mathcal{M}_{X,\{x,x_0\}} & 0 \\
\big\downarrow{\scriptstyle \delta_0} & \\
\mathcal{B}_{X,x} & -1
\end{array}$$

To prove that this sequence is exact at its middle term, observe that the kernel of δ_0 is spanned by diagrams in $M_{\{x,x_0\}}$ that lie in $J_{X-\{x\}} - J_{X-\{x,x_0\}}$. Pick such a diagram. For the diagram to lie in $J_{X-\{x\}}$, two elements of $X-\{x\}$ must be in the same block, or an element of $X-\{x\}$ must be a singleton. For the diagram to lie outwith $J_{X-\{x,x_0\}}$, since x_0 cannot be a singleton in $M_{\{x,x_0\}}$, we conclude that x_0 must be in the same block as some other element of $X-\{x\}$. So, x_0 lies in the same block as some element $x_1 \in X-\{x,x_0\}$, and since the diagram is in $M_{\{x,x_0\}}$ it follows that x, x_0, x_1 must all be in the same block. Thus the diagram is in $M_{\{x,x_0,x_1\}}$, and so lies in the image of δ_1.

To prove exactness in degree $i \geqslant 1$ and above, one first observes that in degrees $i-1, i, i+1$ the complex splits as a direct sum over $(x_0,\ldots,x_{i-1})$. It is therefore enough to concentrate on a single $(x_0,\ldots,x_{i-1})$-summand at a time. Having restricted to such a summand, one now proves exactness similarly to the degree 0 case, and we leave the details of this to the reader.

The fact that the resolution remains acyclic after applying $\mathbb{1}\otimes_{\mathrm{P}_n}-$ follows immediately from Lemma 4.5, which shows that in fact the resolution vanishes under this operation. $\square$

Proof of Theorem 4.2. We first tackle the cases $n=0,1$. When $n=0$ we have $\mathrm{P}_n = R$ and the claim follows immediately. When $n=1$, we either have $X=\varnothing$, or we have $X=\{1\}$ and δ invertible. When $X=\varnothing$ we have $J_X=0$, so that $\mathrm{P}_n/J_X=\mathrm{P}_n$

and the claim follows. Finally, when $X = \{1\}$ and δ is invertible, then J_X is the R-span, and indeed the P_n-span, of the idempotent $\delta^{-1}T_1$. Thus J_X and P_n/J_X are both direct summands of P_n, and in particular the latter is projective, so that the claim follows.

We now assume that $n \geq 2$, and prove the claim by strong induction on the cardinality of X. When $X = \varnothing$ we have $J_X = 0$, so that $\mathrm{P}_n/J_X = \mathrm{P}_n$ and the claim is immediate.

Suppose now that $|X| > 0$ and that the claim holds for all X' of a smaller cardinality. According to Proposition 4.6 and Theorem 3.1, it will be sufficient to show that the modules

$$\mathrm{P}_n/J_{X-\{x\}} \quad \mathcal{A}_{X,x} \quad \mathcal{B}_{X,x}$$

all vanish under $\mathrm{Tor}_i^{\mathrm{P}_n}(\mathbb{1}, -)$ for $i > 0$.

In the case of $\mathrm{P}_n/J_{X-\{x\}}$ we have $\mathrm{Tor}_i^{\mathrm{P}_n}(\mathbb{1}, \mathrm{P}_n/J_{X-\{x\}}) = 0$ by the inductive hypothesis.

For $\mathcal{A}_{X,x}$, we divide into the case where δ is invertible, and the case where $|X| < n$. When δ is invertible, Proposition 4.7 shows that $\mathcal{A}_{X,x}$ is a direct summand of $\mathrm{P}_n/J_{X-\{x\}}$, which vanishes under $\mathrm{Tor}_i^{\mathrm{P}_n}(\mathbb{1}, -)$ by the inductive hypothesis, so that $\mathcal{A}_{X,x}$ does as well. When $|X| < n$, Proposition 4.9 gives us resolutions

$$C(X, x, y) \to \mathcal{A}_{X,x} \quad \text{and} \quad \mathbb{1} \otimes_{\mathrm{P}_n} C(X, x, y) \to \mathbb{1} \otimes_{\mathrm{P}_n} \mathcal{A}_{X,x}.$$

The terms of $C(X, x, y)$ are all $\mathrm{P}_n/J_{X-\{x\}}$, which vanish under $\mathrm{Tor}_i^{\mathrm{P}_n}(\mathbb{1}, -)$ by the inductive hypothesis, so that Theorem 3.1 applies to tell us that the same is true for $\mathcal{A}_{X,x}$ itself.

For $\mathcal{B}_{X,x}$, Proposition 4.11 gives us the resolutions

$$D(X, x) \to \mathcal{B}_{X,x} \quad \text{and} \quad \mathbb{1} \otimes_{\mathrm{P}_n} D(X, x) \to \mathbb{1} \otimes_{\mathrm{P}_n} \mathcal{B}_{X,x}.$$

The terms of $D(X, x)$ are direct sums of modules of the form $\mathcal{M}_{X,\{x,x_0,\ldots,x_i\}}$. Each $\mathcal{M}_{X,\{x,x_0,\ldots,x_i\}}$ is a direct summand of $\mathrm{P}_n/J_{X-\{x,x_0,\ldots,x_i\}}$ by Lemma 4.5, and since $\mathrm{Tor}_i^{\mathrm{P}_n}(\mathbb{1}, -)$ vanishes on the latter, it also vanishes on the former. (Note that this is the only place in our argument where we have used *strong* induction.) We can now apply Theorem 3.1 to $D(X, x)$ to find that $\mathcal{B}_{X,x}$ vanishes under $\mathrm{Tor}_i^{\mathrm{P}_n}(\mathbb{1}, -)$ as required. $\square$

5. Replacing Shapiro's lemma

This section closely follows Section 4 of [Boyd et al. 2021]. We include all statements, and proofs of the lemmas which slightly differ in the case of partition algebras. The proof of Theorem 5.1 is identical to that in [Boyd et al. 2021], with adapted inputs.

As in the case for the Brauer algebras, we have inclusion and projection maps

$$R\mathfrak{S}_m \xrightarrow{\iota} \mathrm{P}_m \xrightarrow{\pi} R\mathfrak{S}_m.$$

These are compatible with the inclusions $\mathrm{P}_m \to \mathrm{P}_n$ and $R\mathfrak{S}_m \to R\mathfrak{S}_n$, and also respect the actions on the trivial module. They therefore induce the following maps of Tor-groups:

$$\mathrm{Tor}_*^{R\mathfrak{S}_n}(\mathbb{1}, R\mathfrak{S}_n \otimes_{R\mathfrak{S}_m} \mathbb{1}) \xrightarrow{\iota_*} \mathrm{Tor}_*^{\mathrm{P}_n}(\mathbb{1}, \mathrm{P}_n \otimes_{\mathrm{P}_m} \mathbb{1}) \xrightarrow{\pi_*} \mathrm{Tor}_*^{R\mathfrak{S}_n}(\mathbb{1}, R\mathfrak{S}_n \otimes_{R\mathfrak{S}_m} \mathbb{1}).$$

Then the main result of this section is the next theorem, which replaces Shapiro's lemma in the Quillen style proof of homological stability for groups.

Theorem 5.1. *Let $n \geqslant m \geqslant 0$. Suppose that δ is invertible in R, or that $m < n$. Then the maps*

$$\iota_* : \mathrm{Tor}_*^{R\mathfrak{S}_n}(\mathbb{1}, R\mathfrak{S}_n \otimes_{R\mathfrak{S}_m} \mathbb{1}) \to \mathrm{Tor}_*^{\mathrm{P}_n}(\mathbb{1}, \mathrm{P}_n \otimes_{\mathrm{P}_m} \mathbb{1})$$

and

$$\pi_* : \mathrm{Tor}_*^{\mathrm{P}_n}(\mathbb{1}, \mathrm{P}_n \otimes_{\mathrm{P}_m} \mathbb{1}) \to \mathrm{Tor}_*^{R\mathfrak{S}_n}(\mathbb{1}, R\mathfrak{S}_n \otimes_{R\mathfrak{S}_m} \mathbb{1})$$

are mutually inverse isomorphisms.

Theorem A follows immediately from Theorem 5.1 by taking δ invertible and $m = n$, using the identifications $R\mathfrak{S}_n \otimes_{R\mathfrak{S}_m} \mathbb{1} \cong \mathbb{1}$ and $\mathrm{P}_n \otimes_{\mathrm{P}_m} \mathbb{1} \cong \mathbb{1}$.

The remainder of this section is devoted to proving Theorem 5.1, which follows in exactly the same way as Theorem 4.1 of [Boyd et al. 2021] after some preparatory definitions and lemmas.

Recall from Definition 4.1 that $J_m \subseteq \mathrm{P}_n$ denotes the ideal consisting of all diagrams in which, among the nodes on the right labelled by $\{n-m+1, \ldots, n\}$, there is a least one singleton or one pair of nodes in the same block. Observe that P_n is a right $R\mathfrak{S}_m$-module, via the inclusions $R\mathfrak{S}_m \subseteq \mathrm{P}_m \subseteq \mathrm{P}_n$, and that this module structure preserves J_m, since right multiplying by a diagram which permutes the nodes $\{n-m+1, \ldots, n\}$ does not change whether there exists a singleton or two nodes in the same block in this set. Therefore we have that P_n/J_m is a right $R\mathfrak{S}_m$-module.

Lemma 5.2. *For $m \leq n$, P_n/J_m is free when regarded as a right $R\mathfrak{S}_m$-module.*

Proof. We have that P_n/J_m has basis consisting of the diagrams for which the nodes in $\{n-m+1, \ldots, n\}$ have no singleton, and no two nodes in the same block. This means that each node in $\{n-m+1, \ldots, n\}$ is attached to a distinct block in the diagram. Now, $\mathfrak{S}_m$ acts freely on this basis, since multiplying such a diagram with a permutation in $\mathfrak{S}_m$ results again in a diagram where the nodes in $\{n-m+1, \ldots, n\}$ are attached to distinct blocks. Under this action, the stabilizer of any such diagram is trivial. $\square$

Lemma 5.3. *For $m \leq n$, there is an isomorphism of left P_n-modules*

$$\mathrm{P}_n/J_m \otimes_{R\mathfrak{S}_m} \mathbb{1} \cong \mathrm{P}_n \otimes_{\mathrm{P}_m} \mathbb{1},$$

under which $(b + J_m) \otimes r \in \mathrm{P}_n/J_m \otimes_{R\mathfrak{S}_m} \mathbb{1}$ corresponds to $b \otimes r \in \mathrm{P}_n \otimes_{\mathrm{P}_m} \mathbb{1}$.

Proof. Throughout this proof we regard J_m as an ideal in P_n, and write $J_m \cap \mathrm{P}_m$ for the corresponding ideal in P_m.

Let us show that the maps

$$(b + J_m) \otimes r \mapsto b \otimes r \quad \text{and} \quad b \otimes r \mapsto (b + J_m) \otimes r$$

are well defined. It then immediately follows that they are inverses and thus isomorphisms.

For the first map, we need to show that $b\sigma \otimes r = b \otimes r$ for $\sigma \in \mathfrak{S}_m$ and that $j \otimes r \in \mathrm{P}_n \otimes_{\mathrm{P}_m} \mathbb{1}$ is zero if $j \in J_m$. The first equation follows immediately as $\sigma \in \mathfrak{S}_m \subset \mathrm{P}_m$ acts as the identity on $\mathbb{1}$. The second condition holds because if $j \in J_m$, then we can write j as a sum of products of the form $b \cdot j'$ where $b \in \mathrm{P}_n$ and $j' \in J_m \cap \mathrm{P}_m$, and for each such summand we have $b \cdot j' \otimes r = b \otimes j' \cdot b = b \otimes 0 = 0$.

For the second map, we let $b \in \mathrm{P}_n$, $b' \in \mathrm{P}_m$, and $r \in \mathbb{1}$, and show that

$$(bb' + J_m) \otimes r = (b + J_m) \otimes (b' \cdot r).$$

It is enough to prove this for $b' \in R\mathfrak{S}_m$ and $b' \in J_m \cap \mathrm{P}_m$ as $\mathrm{P}_m = R\mathfrak{S}_m \oplus (J_m \cap \mathrm{P}_m)$. For $b' \in R\mathfrak{S}_m$, we get the equation directly from the definition of the tensor product. For $b' \in J_m \cap \mathrm{P}_m$, we note that $bb' \in J_m = P_n \cdot J_m$ and thus $(bb' + J_m) \otimes r$ is zero, as is $(b + J_m) \otimes (b' \cdot r)$ since $b' \cdot r = 0$. □

Now recall from Theorem 4.2 that, under the hypotheses of Theorem 5.1,

$$\mathrm{Tor}_*^{\mathrm{P}_n}(\mathbb{1}, \mathrm{P}_n/J_m) = \begin{cases} \mathbb{1} & \text{if } * = 0, \\ 0 & \text{if } * > 0. \end{cases}$$

Proof of Theorem 5.1. The proof of Theorem 5.1 now follows exactly as in [Boyd et al. 2021, Proof of Theorem 4.1], replacing the occurrences of Br_n with P_n, and inputting Lemma 5.3 and Theorem 4.2 as appropriate. □

6. High connectivity

We build a complex similar to the one in [Hepworth 2022] and [Boyd et al. 2021].

Definition 6.1. For n a nonnegative integer, we define the chain complex $C_n = (C_n)_*$ of P_n-modules as follows. The degree p part $(C_n)_p$ is nonzero in degrees $-1 \leq p \leq n-1$, where it is given by

$$(C_n)_p = \mathrm{P}_n \otimes_{\mathrm{P}_{n-(p+1)}} \mathbb{1}.$$

So in degree -1 it follows that $(C_n)_{-1} = \mathrm{P}_n \otimes_{\mathrm{P}_n} \mathbb{1} \cong \mathbb{1}$. For $0 \leq p \leq n-1$ the degree p differential ∂^p is given by the alternating sum

$$\partial^p = \sum_{i=0}^{p} (-1)^i d_i^p : (C_n)_p \to (C_n)_{p-1},$$

where, algebraically, the map d_i^p for $1 \leq i \leq p$ is given by

$$d_i^p : \mathrm{P}_n \otimes_{\mathrm{P}_{n-(p+1)}} \mathbb{1} \to \mathrm{P}_n \otimes_{\mathrm{P}_{n-p}} \mathbb{1}, \quad x \otimes r \mapsto (x \cdot S_{n-p+i-1} \dots S_{n-p}) \otimes r$$

and

$$d_0^p : \mathrm{P}_n \otimes_{\mathrm{P}_{n-(p+1)}} \mathbb{1} \to \mathrm{P}_n \otimes_{\mathrm{P}_{n-p}} \mathbb{1}, \quad x \otimes r \mapsto x \otimes r.$$

In other words, when $i = 0$ the product $S_{n-p+i-1} \dots S_{n-p}$ is taken to be the empty product, i.e., the identity element.

In terms of diagrams, elements in degree p can be described as diagrams with an $(n-(p+1))$-box at the top right, as in Proposition 2.4 and the paragraph which follows it. If we label the nodes below the $(n-(p+1))$-box by $0, \dots, p$ from top to bottom, then d_i^p lifts up node i and plugs it into the box.

We now filter C_n. Note that in [Boyd et al. 2021] we first decomposed C_n based on the number of disjoint blocks on the left, and we could also do that here. However this is not necessary for the proof.

Definition 6.2. We define a filtration

$$F_0 C_n \subseteq F_1 C_n \subseteq \cdots \subseteq F_{\lfloor n/2 \rfloor} C_n = C_n$$

of C_n as follows. The j-th level $F_j C_n$ is generated by diagrams with at most j blocks that have at least 2 positive (right-hand) nodes and are not connected to the box. Note that this is indeed a filtration, since the boundary map can only decrease the number of blocks on the right not connected to the box.

We briefly recall the definition of the complex of injective words with separators.

Definition 6.3 (injective words with separators). Let X be a finite set and let $k \geqslant 0$. An *injective word on X with k separators* is a word with letters taken from the set $X \sqcup \{\,|\,\}$ consisting of X and the *separator* $|$, where each letter from X appears at most once, and where the separator appears exactly k times. When $k = 0$, then these are simply the injective words on X.

Definition 6.4 (the complex of injective words with separators). Let X be a finite set, let $s \geqslant 0$, and let R be a commutative ring. The *complex of injective words with s separators* is the R-chain complex $W_X^{(s)}$ concentrated in degrees $-1 \leq p \leq |X| - 1$, and defined as follows. In degree p, $(W_X^{(s)})_p$ has basis given by the injective words on X with s separators, and with $(p+1)$ letters from X. Thus such a word

$\boldsymbol{a} \in (W_X^{(s)})_p$ has length $s+p+1$. Let $r = s+p$ and $\boldsymbol{a} = a_0\, a_1 \ldots a_r$. The boundary operator $\partial^p : (W_X^{(s)})_p \to (W_X^{(s)})_{p-1}$ is defined by the rule

$$\partial^p(a_0\, a_1 \ldots a_r) = \sum_{i=0}^{r} (-1)^i\, a_0 \ldots \hat{a}_i \ldots a_r$$

subject to the condition that if the omitted letter is a separator, then the corresponding term is omitted (or identified with 0). In other words, the boundary is the signed sum of the words obtained by deleting the letters that come from X and *not deleting any separators*, but with signs determined by the position of the deleted letter among all letters *including the separator*:

$$\partial^p(a_0\, a_1 \ldots a_r) = \sum_{a_i \in X} (-1)^i\, a_0 \ldots \hat{a}_i \ldots a_r.$$

We will aim to identify the filtration quotients $F_j C_n / F_{j-1} C_n$ with a sum of shifted copies of the complex of injective words with separators, as in [Boyd et al. 2021]. (Note that in [Boyd et al. 2021] the argument for the Brauer algebras is somewhat simpler, and so the reader may wish to look at the Brauer proof first.)

To complete this identification, we exhibit a one-to-one correspondence between diagrams and tuples of data. This correspondence is complicated, so we start with the simple example of the tuple corresponding to a diagram with no box, and no restriction on the right-hand side blocks. Recall that a diagram is a pictorial way of representing a partition of the set $\{-n, \ldots, -1, 1, \ldots, n\}$.

A diagram D determines, and is determined by, a tuple (L, R, ϕ) consisting of:

- A partition L of $\{-1, \ldots, -n\}$.
- A partition R of $\{1, \ldots, n\}$.
- A labelling $\phi : R \to \{\varnothing\} \cup L$ with the property that $\phi(r) = \phi(r')$ only when $r = r'$ or $\phi(r) = \phi(r') = \varnothing$.

The correspondence sends a diagram D to the tuple (L, R, ϕ) for which:

- L is the induced partition on the left-hand nodes $-1, \ldots, -n$.
- R is the induced partition on the right-hand nodes $1, \ldots, n$.
- ϕ labels a block on the right by the (necessarily unique) block on the left to which it is attached, if any, and labels it by $\varnothing$ otherwise.

An example is shown in Figure 7. Here, the process of restricting the partition to the left and right sides of the diagram amounts to discarding all the left-to-right connections. Those left-to-right connections are instead recorded in the labelling ϕ. To see that ϕ satisfies the third property above because, observe that if it did not, then the diagram would have two distinct blocks r, r' on the right attached to the *same* block on the left. That would be a contradiction because then r and r' would in fact themselves be the same.

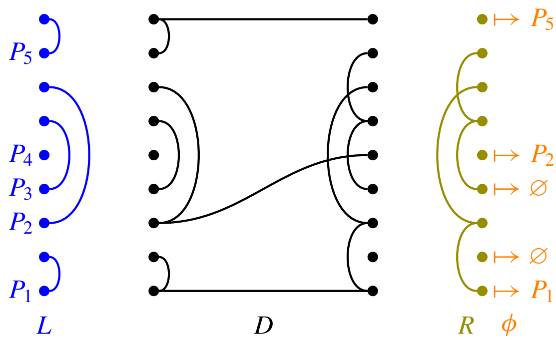


Figure 7. The process of extracting from a diagram D the tuple (L, R, ϕ). Blocks in L, and the labellings of R are indicated at their lowermost node.

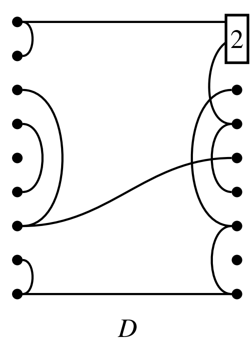


Figure 8. An example of a diagram D, when $n = 9$, $j = 1$ and $p = 6$.

We now observe that the filtration quotient $F_j C_n / F_{j-1} C_n$ has a basis in degree p consisting of diagrams which have an $(n - (p + 1))$-box on the right, and exactly j blocks with ≥ 2 nodes on the right that are not connected to the box. Here, the size of the box is determined by the degree as in Definition 6.1, and the condition on the j blocks follows from the definition of the filtration given in Definition 6.2. An example is given in Figure 8.

In the next definition, we explain how these basis diagrams determine a tuple of data, analogously to the discussion above. Once this data has been stripped from the diagram, we are left with the desired information of an injective word with separators. In this injective word, the letters encode left-to-right connections for which the block on the right has a single element; and the separators correspond to all other nodes below the box on the right. There are at least $2j$ of these separators, because there are precisely j blocks on the right that have 2 or more right-hand nodes and are not connected to the box. Later, in Lemma 6.8 we show how to conversely start with an injective word with separators and our tuple of data and rebuild the diagram.

Definition 6.5. A diagram in the basis of $(F_j C_n / F_{j-1} C_n)_p$ determines a tuple

$$(P, X, s, Y, f)$$

consisting of the following data:

- A partition P of $\{-1, \ldots, -n\}$.
- A subset X of the blocks of P.
- A number $2j \leq s \leq n - |X|$.
- A partition Y of $\{1, \ldots, s\}$, such that $\geq j$ blocks have size ≥ 2.
- A labelling $f : Y \to (\{\varnothing\} \cup P \setminus X) \times \{\Box, \neg\Box\}$ (where the symbols $\Box$, $\neg\Box$ represent "box" and "not-box" respectively) such that
  - singletons have first label $\varnothing$
  - no two blocks in Y can have the same first label in $P \setminus X$
  - exactly j blocks of size ≥ 2 have second label $\neg\Box$
  - exactly $n - s - |X|$ blocks have second label $\Box$.

The diagram D determines the tuple as follows (an example is shown in Figure 9):

- P is the partition of $\{-1, \ldots, -n\}$ given by restricting the blocks of D to the negative elements, i.e., to the nodes on the left-hand side of the diagram.
- X is the set of blocks in P which, when viewed in D, are connected to exactly one thing on the right (this can be a connection to the box, or to a single node).
- The number s is equal to the number of nodes on the right of D not connected to a block in X. These nodes are precisely those which are singletons, or are connected to another element on the right, or to the box. Therefore, every node in one of the j blocks of D that have at least 2 positive (right-hand) nodes and are not connected to the box (as in Definition 6.2) is included and so $s \geq 2j$. Also, none of the nodes that are connected to a block in X are included, so $s \leq n - |X|$. It follows that $n - |X| - s$ is the number of blocks connected to the box and to at least one node on the right.
- Y is the partition given by restricting D to the set of s nodes on the right that are not connected to the blocks of X (we relabel these $1, \ldots, s$, maintaining the order).
- The first entry of the labelling f, in $\varnothing \cup P \setminus X$, indicates whether the blocks of Y are disconnected from the rest of D (in which case the label is $\varnothing$), or connected to the left-hand side (in which case the label is the block in $P \setminus X$ that they are connected to). Singletons in Y cannot be connected to the left because otherwise

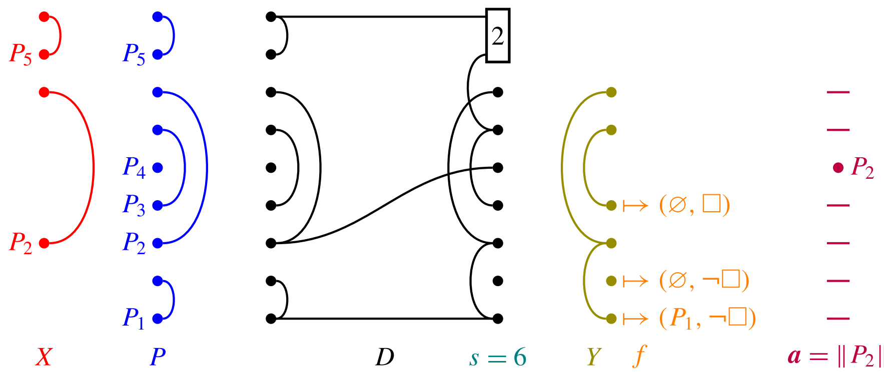


Figure 9. The process of extracting from the diagram D in Figure 8 the tuple (P, X, s, Y, f) and injective word $\boldsymbol{a}$ described in Definition 6.5. Blocks in P and the labelling f of Y are indicated at their lowermost node.

they would be connected to a block in X on the left. Thus their first label must be $\varnothing$. Two blocks in Y cannot be connected to the same block in $P \setminus X$, so two first labels can only be the same if they are both $\varnothing$.

The second entry of the labelling f is $\square$ if the block in Y is connected to the box in D and $\neg\square$ if it is not. The condition that there are exactly j blocks of size ≥ 2 with second label $\neg\square$ accounts for the diagram being in the filtration quotient $F_j C_n / F_{j-1} C_n$. The condition that there are exactly $n - s - |X|$ blocks with second label $\square$ follows from the above observation that this is the number of blocks connected to the box, containing at least one node on the right.

The remaining data in the diagram determines an injective word with s separators $\boldsymbol{a}$, of length $p + 1 - s$, on the set X, obtained as follows: If the i-th node (from the top) on the right is connected to a block in X, then the i-th letter of $\boldsymbol{a}$ is the corresponding element of X. Otherwise the i-th letter of $\boldsymbol{a}$ is a separator, and there are exactly s of these.

Definition 6.6. By the above discussion, we can define a map

$$\Phi_* : \frac{F_j C_n}{F_{j-1} C_n} \to \bigoplus_{P,X,s,Y,f} W_X^{(s)}[-s].$$

The direct sum is over all 5-tuples (P, X, s, Y, f) satisfying the properties listed at the start of Definition 6.5. A diagram D in $(F_j C_n / F_{j-1} C_n)_p$ is sent by Φ_p to the injective word with separators $\boldsymbol{a}$ in the degree p part of the summand $W_X^{(s)}[-s]$ corresponding to (P, X, s, Y, f), where (P, X, s, Y, f) and $\boldsymbol{a}$ are obtained as in Definition 6.5.

We now prove that Φ_* is a chain map and isomorphism. This will allow us to leverage the high connectivity of the complex of injective words with separators [Boyd et al. 2021, Proposition 5.14] to a high connectivity result for C_n, via the filtration.

Lemma 6.7. *Φ_* is a chain map.*

Proof. First, we claim that the 5-tuple (P, X, s, Y, f) associated, via Φ_*, to a basis diagram D in $(F_j C_n/F_{j-1} C_n)_p$ is preserved in all diagrams appearing in the boundary of D. Recall from Definition 6.1 that the boundary map ∂^p sends a diagram to the alternating sum of the diagrams obtained as follows: work through the nodes on the right of the diagram, and in each case move the node into the box. This clearly does not change the left-hand end of the diagram, and therefore all of the diagrams in the boundary have the same X and P associated to them. If the node that is moved into the box is a singleton, or was part of a block that was connected to the box, then these nodes are included in the count for s, but after moving it into the box, the resulting diagram either has a singleton in the box or has a loop at the box, and therefore again vanishes. The other nodes counting towards s are those that are part of a block with ≥ 2 elements from the right, and are not connected to the box. There are exactly j such blocks, and so moving any of their nodes into the box gives zero in the quotient $(F_j C_n/F_{j-1} C_n)_p$. Therefore the only nodes we can move into the box without getting zero, are those that are not counted by s, i.e., s remains constant under the boundary map. It follows that Y and f remain constant, since Y partitions these s nodes and f labels them.

The above paragraph demonstrates that $F_j C_n/F_{j-1} C_n$ splits as a direct sum indexed by the 5-tuples (P, X, s, Y, f). It now suffices to show that the assignment that sends a diagram with fixed (P, X, s, Y, f) to the corresponding injective word with separators $\boldsymbol{a}$ respects the boundary map. But this is clear: moving a node joined to a block in X into the box corresponds exactly to deleting one of the nonseparator letters from $\boldsymbol{a}$. $\square$

Lemma 6.8. *Φ_* is an isomorphism.*

Proof. We will prove that Φ_* is an isomorphism by showing that it is obtained from a bijection between the basis of $(F_j C_n/F_{j-1} C_n)$, which is given by diagrams, and the basis of $\bigoplus_{X,P,s,Y,f} W_X^{(s)}[-s]$, which is given by injective words with separators. To do this, we will explain how to (re)build a diagram in $(F_j C_n/F_{j-1} C_n)$ from a tuple (P, X, s, Y, f) and an injective word with separators $\boldsymbol{a}$.

We work in degree $s+k-1$ in the summand $W_X^{(s)}[-s]$ associated to a 5-tuple (P, X, s, Y, f). We therefore take an injective word $\boldsymbol{a}$ of length k with s separators, and we will build a diagram in $(F_j C_n/F_{j-1} C_n)_{s+k-1}$. We begin with an empty diagram with $s+k$ nodes on the right-hand side, and a box of size $n-s-k$; this is possible since $s+k \leq s+|X| \leq n$, where the latter inequality is one of the conditions imposed on the 5-tuple. Next, we build all the blocks on the left using P, and draw half-edges from the blocks in X to the right (don't connect these edges to anything yet). We place the injective word with separators vertically against the $s+k$ nodes on the right-hand side, and the word indicates connections from k of the nodes to half-edges from X. We connect the remaining half edges from X to the box. The separators indicate the positioning of the s nodes $\{1,\dots,s\}$ which are then partitioned by Y, and labelled by f. The first labels of Y indicate which blocks are connected to blocks on the left-hand side in $P \setminus X$. Finally, if the second label of a block in Y is $\square$ we connect the block to the box. Note that $|X|-k$ blocks of X are connected to the box, and $n-s-|X|$ blocks of Y are connected to the box, the latter property being another of our conditions on the 5-tuple. This means that exactly $n-s-k$ distinct blocks are connected to the box, and since this is the size of the box the diagram is nonzero in $(C_n)_{s+k}$. The diagram lies in $F_j\, C_n/F_{j-1}\, C_n$ since exactly j blocks in Y of size ≥ 2 have second label $\neg\square$ and are therefore not joined to the box, again by our conditions on the 5-tuple.

The last paragraph shows how to obtain, from a tuple (P, X, s, Y, f) and an injective word $\boldsymbol{a} \in W_X^{(s)}[-s]_{s+k-1}$, a diagram in the basis of $(F_j\, C_n/F_{j-1}\, C_n)_{s+k-1}$. It is now immediate to verify that this is inverse to the effect of Φ_* on bases. $\square$

Proposition 6.9. *For all $0 \leq j \leq \left\lfloor \frac{n}{2} \right\rfloor$, the filtration quotients $F_j\, C_n/F_{j-1}\, C_n$ satisfy $H_i(F_j\, C_n/F_{j-1}\, C_n) = 0$ for $i \leq \frac{n-3}{2}$.*

Proof. We first consider the case $n = 0$, where the only possibility is that $j = 0$ so that $F_0\, C_0 = C_0$. The claim is then that $H_i(C_0) = 0$ for $i \leq -\frac{3}{2}$, but since C_0 consists of a single copy of $\mathbb{1}$ in degree -1, this is immediate.

We now consider the case $n > 0$. Using Lemma 6.8 this is equivalent to the homology of $W_X^{(s)}[-s]$ vanishing in the desired range, for each 5-tuple (P, X, s, Y, f) satisfying the conditions of Definition 6.5. By [Boyd et al. 2021, Proposition 5.14], $H_i(W_X^{(s)}) = 0$ for $i \leq |X|-2$, so that $H_i(W_X^{(s)}[-s]) = 0$ for $i \leq |X|+s-2$. It will therefore suffice to show that $\left\lfloor \frac{n-3}{2} \right\rfloor \leq |X|+s-2$, or equivalently

$$\begin{cases} n \leq 2|X|+2s & \text{if } n \text{ even,} \\ n \leq 2|X|+2s-1 & \text{if } n \text{ odd.} \end{cases}$$

Let us first prove that we always have $n \leq 2|X|+2s$. Our conditions on the 5-tuple (P, X, s, Y, f) mean that $n-s-|X|$ is the number of blocks of Y with second f-label $\square$, so that in particular $n-s-|X| \leq |Y|$. And since Y is a partition of $\{1,\dots,s\}$ we have $|Y| \leq s$. Combining the last two inequalities and rearranging gives us $n \leq |X|+2s$. Because $|X| \geq 0$, we therefore have $n \leq 2|X|+2s$. In particular, this proves the proposition if n is even. If n is odd, it certainly cannot be equal to $2|X|+2s$ which is even. Therefore it can be at most one smaller: $n \leq 2|X|+2s-1$. $\square$

Theorem 6.10. *$H_i(C_n) = 0$ for $i \leq \frac{n-3}{2}$.*

Proof. By Proposition 6.9, the homology of the filtration quotient $F_j C_n / F_{j-1} C_n$ vanishes in degrees $i \leq \frac{n-3}{2}$ for all j. The same then holds for C_n itself by considering the long exact sequences associated to the short exact sequences

$$0 \to F_{j-1} C_n \to F_j C_n \to \frac{F_j C_n}{F_{j-1} C_n} \to 0. \qquad \square$$

7. Proof of Theorem B

The proof of Theorem B directly mirrors the proof of [Boyd et al. 2021, Theorem B], with the following substitutions:

- All instances of the Brauer algebra should be replaced with the partition algebra.
- The maps ι and π of [Boyd et al. 2021] should be replaced by the maps of the same name in the current paper. Similarly for the complex C_*.
- Theorem 5.4 of [Boyd et al. 2021] should be replaced with Theorem 6.10.
- Theorem 4.1 of [Boyd et al. 2021] should be replaced with Theorem 5.1.

We note that in the second paragraph of the proof of [Boyd et al. 2021, Theorem 6.3], there is an error, and the words "odd" and "even" should be transposed.

Acknowledgements

Boyd was supported by the United Kingdom Engineering and Physical Sciences Research Council grants EP/V043323/1 and EP/V043323/2. Patzt was supported by a Simons collaboration grant. Our thanks to the referee for detailed reading and helpful comments.

References

[Barratt and Priddy 1972] M. Barratt and S. Priddy, "On the homology of non-connected monoids and their associated groups", *Comment. Math. Helv.* **47** (1972), 1–14. MR Zbl

[Bowman et al. 2022] C. Bowman, S. Doty, and S. Martin, "Integral Schur–Weyl duality for partition algebras", *Algebr. Comb.* **5**:2 (2022), 371–399. MR Zbl

[Boyd and Hepworth 2020] R. Boyd and R. Hepworth, "The homology of the Temperley–Lieb algebras", preprint, 2020. To appear in *Geom. Topol.* Zbl arXiv 2006.04256

[Boyd et al. 2021] R. Boyd, R. Hepworth, and P. Patzt, "The homology of the Brauer algebras", *Selecta Math.* (*N.S.*) **27**:5 (2021), art. id. 85. MR Zbl

[Boyde 2022] G. Boyde, "Idempotents and homology of diagram algebras", preprint, 2022. Zbl arXiv 2212.01826

[Friedlander and Mazur 1994] E. M. Friedlander and B. Mazur, "Filtrations on the homology of algebraic varieties", Mem. Amer. Math. Soc. **529**, 1994. MR Zbl

[Galatius 2011] S. Galatius, "Stable homology of automorphism groups of free groups", *Ann. of Math.* (2) **173**:2 (2011), 705–768. MR Zbl

[Halverson and Ram 2005] T. Halverson and A. Ram, "Partition algebras", *European J. Combin.* **26**:6 (2005), 869–921. MR Zbl

[Hepworth 2022] R. Hepworth, "Homological stability for Iwahori–Hecke algebras", *J. Topol.* **15**:4 (2022), 2174–2215. MR Zbl

[Jones 1994] V. F. R. Jones, "The Potts model and the symmetric group", pp. 259–267 in *Subfactors* (Kyuzeso, 1993), World Sci. Publ., River Edge, NJ, 1994. MR Zbl

[Martin 1994] P. Martin, "Temperley–Lieb algebras for nonplanar statistical mechanics — the partition algebra construction", *J. Knot Theory Ramifications* **3**:1 (1994), 51–82. MR Zbl

[Martin 1996] P. Martin, "The structure of the partition algebras", *J. Algebra* **183**:2 (1996), 319–358. MR Zbl

[Moselle 2022] I. Moselle, "Homological stability for Iwahori–Hecke algebras of type B", preprint, 2022. Zbl arXiv 2211.06230

[Nakaoka 1960] M. Nakaoka, "Decomposition theorem for homology groups of symmetric groups", *Ann. of Math.* (2) **71** (1960), 16–42. MR Zbl

[Patzt 2024] P. Patzt, "Representation stability for diagram algebras", *J. Algebra* **638** (2024), 625–669. MR Zbl

[Sroka 2023] R. J. Sroka, "The homology of a Temperley–Lieb algebra on an odd number of strands", preprint, 2023. To appear in *Algebr. Geom. Topol.* Zbl arXiv 2202.08799

[Weibel 1994] C. A. Weibel, *An introduction to homological algebra*, Cambridge Studies in Advanced Mathematics **38**, Cambridge University Press, Cambridge, 1994. MR Zbl

Received June 20, 2023. Revised October 19, 2023.

RACHAEL BOYD
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF GLASGOW
GLASGOW
UNITED KINGDOM
rachael.boyd@glasgow.ac.uk

RICHARD HEPWORTH
INSTITUTE OF MATHEMATICS
UNIVERSITY OF ABERDEEN
ABERDEEN
UNITED KINGDOM
r.hepworth@abdn.ac.uk

PETER PATZT
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OKLAHOMA
NORMAN, OK
UNITED STATES
ppatzt@ou.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 327, No. 1, 2023
https://doi.org/10.2140/pjm.2023.327.29

REMARKS ON EIGENSPECTRA OF ISOLATED SINGULARITIES

BEN CASTOR, HAOHUA DENG, MATT KERR AND GREGORY PEARLSTEIN

We introduce a simple calculus, extending a variant of the Steenbrink spectrum, to describe Hodge-theoretic invariants for smoothings of isolated singularities with relative automorphisms. After computing these "eigenspectra" in the quasihomogeneous case, we give three applications to singularity bounding and monodromy of variations of Hodge structure (VHS).

Introduction

Recent work of M. Kerr and R. Laza on the Hodge theory of degenerations [Kerr et al. 2021; Kerr and Laza 2023] reexamined the mixed Hodge theory of the Clemens–Schmid and vanishing-cycle sequences, with an emphasis on applications to limits of period maps and compactifications of moduli. When a degenerating family of varieties has a finite group G acting on its fibers, these become exact sequences in the category of mixed Hodge structures with $G \times \mu_k$-action, where k is the order of T_{ss} (the semisimple part of monodromy). These kinds of situations often show up in generalized Prym or cyclic-cover constructions; for instance, instead of using the period map attached to a family of varieties, one may want to use the "exotic" period map arising from a cyclic cover branched along the family (e.g., [Allcock et al. 2002; 2011; Casalaina-Martin et al. 2012; Deligne and Mostow 1986; Dolgachev and Kondō 2007]).

In this note we explain how to encode the contributions of isolated singularities with G-action to the vanishing cohomology in terms of *G-spectra* (Definition 1.11). These are formal sums (with positive integer coefficients) of triples in $\mathbb{Z} \times \mathbb{Q} \times \mathfrak{R}$, where $\mathfrak{R}$ is the set of irreducible representations of G. The term *eigenspectrum* (Definition 1.12) refers to the specific case of a cyclic group $\langle g \rangle$ with fixed generator. (At the end of Section 3 and in most of Section 5 a larger group $\mathcal{G}$ nontrivially permutes the singularities; G always denotes a subgroup stabilizing them.)

In Section 1 this formalism emerges naturally from the general setting of a proper morphism of 1-parameter degenerations over a disk, by specializing the

MSC2020: 14D06, 14D07, 14J17, 32S25, 32S35.
Keywords: isolated singularity, nodes, spectrum, eigenspectrum, quasihomogeneous singularity, Calabi–Yau variety, variation of Hodge structure, monodromy.

morphism to an automorphism $g \in \mathrm{Aut}(\mathcal{X}/\Delta)$ fixing a singularity $x \in X_0$. The eigenspectrum $\sigma^g_{f,x}$ simply records the dimensions of simultaneous eigenspaces of g^* and T_{ss} in the (p, q)-subspaces of V_x (Definition 1.12). We give a general computation in Section 2 of $\sigma^g_{f,x}$ in the case of a quasihomogeneous singularity, in terms of a monomial basis for the associated Jacobian ring (Corollary 2.7).

In the remaining sections, we give three applications. The first, in Section 3, is to bounding the number of nodes on Calabi–Yau hypersurfaces in weighted projective spaces (Theorem 3.6) by passing to cyclic covers. There is already a large literature on node-bounding, including [Jaffe and Ruberman 1997; Kerr and Laza 2023; Miyaoka 1984; Schoen 1985; Varchenko 1983; van Straten 2020]. In the case of $\mathbb{P}^{n+1}$, our approach does not improve Varchenko's bound (e.g., 135 nodes for a quintic hypersurface in $\mathbb{P}^4$), but does yield a simpler proof. However, we do obtain the interesting result (in Theorem 3.11) that a CY hypersurface in $\mathbb{P}^{n+1}$ with isolated singularities and symmetric under $\mathfrak{S}_{n+2}$ cannot contain a node whose $\mathfrak{S}_{n+2}$-orbit has cardinality $(n+2)!$ (i.e., trivial stabilizer).

The other two applications concern codimension-one monodromy phenomena for VHSs over moduli of configurations of points and hyperplanes. In Section 4, the moduli space is $M_{0,2n}$, with the VHS arising from cyclic covers of $\mathbb{P}^1$ branched along the $2m$ ordered points. Propositions 4.5–4.6 and Example 4.7 describe the eigenspectra, LMHS and monodromy types along boundary strata of certain compactifications $\overline{M}^H_{0,2n}$ due to Hassett [2003], generalizing a computation of [Gallardo et al. 2021]. The cases $m = 2, 3, 4$, and 6 go back to work of Deligne and Mostow [1986] and feature a period map (isomorphism) to an arithmetic ball quotient. While the global/extended period map is not as elegant in the remaining cases, the point is that the codimension-one boundary behavior can be dealt with uniformly and efficiently using our calculus.

Our other main example, treated in Section 5, is the VHS $\mathcal{H} \to \mathcal{S}$ on the moduli space of general configurations of $(2n+2)$ hyperplanes in $\mathbb{P}^n$, arising from the middle (intersection) cohomology of a $2:1$ cover $X \to \mathbb{P}^n$ branched along these hyperplanes. These are singular Calabi–Yau n-folds admitting a crepant resolution, and have been studied in [Dolgachev and Kondō 2007; Gerkmann et al. 2007a; 2013; Sheng et al. 2015]. By passing to a smooth complete intersection 2^{2n}-cover of X and applying the Cayley trick (see [Kerr 2003, Section 4.5]), we replace X by a smooth hypersurface

$$Y \subset \mathbb{P}(\mathcal{O}_{\mathbb{P}^{2n+1}}(2)^{\oplus(n+1)})$$

with automorphisms by a group of order 2^{2n}. In codimension-one in moduli, Y acquires nodes, and a variant of Schoen's [1985] result ensures that these produce nontrivial symplectic transvections for $\mathcal{H}$ when n is odd. This gives an easy proof that the geometric monodromy group of $\mathcal{H}$ is maximal (for all n), and the period map “nonclassical”, a fact first proved by Gerkmann et al. [2013] for $n = 3$ and by Sheng et al. [2015] in general.

Notation. In this paper MHS stands for $\mathbb{Q}$-mixed Hodge structure. We shall make frequent use of the *Deligne bigrading* on a MHS V [Deligne 1971]. This is (by definition) the unique decomposition $V_{\mathbb{C}} = \bigoplus_{p,q\in\mathbb{Z}} V^{p,q}$ with the properties that

$$F^k V_{\mathbb{C}} = \bigoplus_{\substack{p,q\\ p\geq k}} V^{p,q}, \quad W_\ell V_{\mathbb{C}} = \bigoplus_{\substack{p,q\\ p+q\leq \ell}} V^{p,q}, \quad \text{and} \quad \overline{V^{q,p}} \equiv V^{p,q} \bmod \bigoplus_{\substack{a<p\\ b<q}} V^{a,b}.$$

We shall make free use of standard multiindex notation (for n-tuples of variables or field-elements) to simplify formulas, viz. $\underline{z} = (z_1, \ldots, z_n)$, $\mathbb{C}[\underline{z}] = \mathbb{C}[z_1, \ldots, z_n]$, $\underline{z}^{\underline{m}} = \prod_i z_i^{m_i}$, $\underline{m} \cdot \underline{w} = \sum_i m_i\, w_i$, $|\underline{m}| = \sum_i m_i$, $\underline{e}^{(i)} = i$-th standard basis vector, etc.

1. G-spectra and eigenspectra

Morphisms and mixed spectra. We begin in the general setting of a proper morphism

$$\begin{tikzcd} \mathcal{Y} \arrow[rr, "\pi"] \arrow[dr, "f'"'] & & \mathcal{X} \arrow[dl, "f"] \\ & \Delta & \end{tikzcd} \tag{1.1}$$

of complex analytic spaces over a disk, which we assume is the restriction to Δ of a proper morphism of quasiprojective varieties over an algebraic curve. (In particular, at the level of fibers we have that $\pi_t : Y_t \to X_t$ is a proper algebraic morphism of quasiprojective varieties.) Let $\mathcal{K}^\bullet \in D^b\,\mathrm{MHM}(\mathcal{X})$ and $\mathcal{L}^\bullet \in D^b\,\mathrm{MHM}(\mathcal{Y})$ be given, with a morphism $\rho : \mathcal{K}^\bullet \to R\pi_*\mathcal{L}^\bullet$. Writing $\imath : X_0 \hookrightarrow \mathcal{X}$ for the inclusion, the vanishing cycle triangle

$$\imath^* \xrightarrow{\text{sp}} \psi_f \xrightarrow{\text{can}} \phi_f \xrightarrow[{[+1]}]{\delta} \tag{1.2}$$

consists of functors from $D^b\,\mathrm{MHM}(\mathcal{X})$ to $D^b\,\mathrm{MHM}(X_0)$, with natural transformations between them; also, monodromy $T = T_{\text{ss}}\, e^N$ induces natural automorphisms of ψ_f and ϕ_f. By proper base-change and faithfulness of $\text{rat} : D^b\,\mathrm{MHM}(X_0) \to D^b_c(X_0)$, $R\pi_* : D^b\,\mathrm{MHM}(Y_0) \to D^b\,\mathrm{MHM}(X_0)$ intertwines the corresponding triangle (and monodromy actions) for $(\mathcal{Y}, f')$. Taking hypercohomology on X_0 yields:

1.3. Proposition. *We have the commutative diagram*

$$\begin{tikzcd} \to \mathbb{H}^k(X_0, \imath^*\mathcal{K}^\bullet) \arrow[r, "\text{sp}"] \arrow[d, "\rho"] & \mathbb{H}^k(X_0, \psi_f\mathcal{K}^\bullet) \arrow[r, "\text{can}"] \arrow[d, "\rho"] & \mathbb{H}^k(X_0, \phi_f\mathcal{K}^\bullet) \arrow[r, "\delta"] \arrow[d, "\rho"] & \mathbb{H}^{k+1}(X_0, \imath^*\mathcal{K}^\bullet) \to \arrow[d, "\rho"] \\ \to \mathbb{H}^k(Y_0, \imath^*\mathcal{L}^\bullet) \arrow[r, "\text{sp}"] & \mathbb{H}^k(Y_0, \psi_{f'}\mathcal{L}^\bullet) \arrow[r, "\text{can}"] & \mathbb{H}^k(Y_0, \phi_{f'}\mathcal{L}^\bullet) \arrow[r, "\delta"] & \mathbb{H}^{k+1}(Y_0, \imath^*\mathcal{L}^\bullet) \to \end{tikzcd}$$

with rows the vanishing-cycle (*long-exact*) *sequences*, *in which all arrows are morphisms of MHS. Moreover*, *the diagram intertwines the actions of* T_{ss} (*by automorphisms of MHS*) *and* N (*by nilpotent* $(-1,-1)$-*endomorphisms of MHS*), *which are trivial* (Id *resp.* 0) *on the end terms.*

1.4. Remark. If f, f' are themselves projective (hence proper), and $\mathcal{K}^\bullet$, $\mathcal{L}^\bullet$ semisimple with respect to the perverse t-structure (e.g., $\mathcal{K}^\bullet = \mathcal{IC}^\bullet_{\mathcal{X}},\ \mathcal{L}^\bullet = \mathcal{IC}^\bullet_{\mathcal{Y}}$), then the decomposition theorem applies, yielding Clemens–Schmid sequences (see [Kerr et al. 2021, Section 5]) which are then automatically compatible under ρ. The main consequence is that the local invariant cycle theorem holds, i.e., sp surjects onto the T-invariants.

Next, assume $\mathcal{X}$, $\mathcal{Y}$, $\{X_t\}_{t\neq 0}$, and $\{Y_t\}_{t\neq 0}$ are smooth, and take $\mathcal{L}^\bullet = \mathbb{Q}_{\mathcal{Y}}$ and $\mathcal{K}^\bullet = \mathbb{Q}_{\mathcal{X}}$; then the diagram in Proposition 1.3 becomes

$$\begin{array}{ccccccccc} \to H^k(X_0) & \xrightarrow{\mathrm{sp}} & H^k_{\mathrm{lim}}(X_t) & \xrightarrow{\mathrm{can}} & H^k_{\mathrm{van}}(X_t) & \xrightarrow{\delta} & H^{k+1}(X_0) \to \\ \downarrow{\scriptstyle \pi^*} & & \downarrow{\scriptstyle \pi^*} & & \downarrow{\scriptstyle \pi^*} & & \downarrow{\scriptstyle \pi^*} \\ \to H^k(Y_0) & \xrightarrow{\mathrm{sp}} & H^k_{\mathrm{lim}}(Y_t) & \xrightarrow{\mathrm{can}} & H^k_{\mathrm{van}}(Y_t) & \xrightarrow{\delta} & H^{k+1}(Y_0) \to \end{array} \tag{1.5}$$

Now if $n = \dim X_0$ and $\Sigma := \mathrm{sing}(X_0)$ is finite, then $H^k_{\mathrm{van}}(X_t) = \{0\}$ for $k \neq n$ and, defining $V_x := H^0 \imath_x^* \phi_f \mathbb{Q}_{\mathcal{X}}[n]$,

$$H^n_{\mathrm{van}}(X_t) \cong \bigoplus_{x\in\Sigma} V_x \tag{1.6}$$

as MHS. We have of course $\pi^{-1}(\Sigma) \subset \widetilde{\Sigma} := \mathrm{sing}(Y_0)$, and if $\dim Y_0 = n$ and $|\widetilde{\Sigma}| < \infty$ then, writing $\widetilde{V}_y := H^0 \imath_y^* \phi_{f'} \mathbb{Q}_{\mathcal{Y}}[n]$ for $y \in \widetilde{\Sigma}$, π^* restricts to morphisms

$$[\pi^*]_x : V_x \to \bigoplus_{y\in\pi^{-1}(x)} \widetilde{V}_y \tag{1.7}$$

of T-MHS — i.e., morphisms of MHS intertwining T (hence T_{ss} and N). These are local invariants.

Recall that T_{ss} acts through finite cyclic groups on each V_x (and $\widetilde{V}_y$), and let κ be the least common multiple of their orders. Write $\zeta_\kappa := e^{2\pi \boldsymbol{i}/\kappa}$ and $V^{p,q}_{x,\boldsymbol{e}(a/\kappa)}$ for the $\boldsymbol{e}(a/\kappa) := e^{2\pi\boldsymbol{i}(a/\kappa)} = \zeta_\kappa^a$-eigenspace of T_{ss} in $V_x^{p,q} \subset V_{x,\mathbb{C}}$. The explicit choice of $\zeta_\kappa \in \mathbb{C}$ is needed to make the following.

1.8. Definition. The *mixed spectrum* $\sigma_{f,x}$ of the isolated singularity $x \in \Sigma$ is the element $\sum_{\alpha,w} m^{f,x}_{\alpha,w}(\alpha, w)$ of the free abelian group $\mathbb{Z}\langle\mathbb{Q}\times\mathbb{Z}\rangle$, where we put $m^{f,x}_{\alpha,w} = \dim(V^{\lfloor\alpha\rfloor, w-\lfloor\alpha\rfloor}_{x,\boldsymbol{e}(\alpha)})$.[1]

[1]Here $\lfloor\cdot\rfloor$ is the greatest integer (floor) function; note also that $\boldsymbol{e}(\alpha)$ is equivalent to taking the fractional part $\{\alpha\} := \alpha - \lfloor\alpha\rfloor$ of α.

Evidently (1.7) must be compatible with the decompositions recorded by the mixed spectra.

Automorphisms and eigenspectra. Now let $\mathcal{G} \leq \mathrm{Aut}(\mathcal{X}/\Delta)$, with $\mathcal{X}$ and $\{X_t\}_{t\neq 0}$ smooth and $|\Sigma| < \infty$. Applying the foregoing results with $\mathcal{Y} = \mathcal{X}$, $f = f'$, and $\pi := g \in \mathcal{G}$, together with [Kerr et al. 2021, Proposition 5.5(i)], yields:

1.9. Corollary. *The vanishing-cycle sequence*

$$0 \to H^n(X_0) \xrightarrow{\mathrm{sp}} H^n_{\lim}(X_t) \xrightarrow{\mathrm{can}} \bigoplus_{x\in\Sigma} V_x \xrightarrow{\delta} H^{n+1}_{\mathrm{ph}}(X_0) \to 0 \tag{1.10}$$

is an exact sequence of $\mathcal{G} \times \mu_\kappa$*-MHS,*[2] *where the* $\langle T_{ss}\rangle \cong \mu_\kappa$*-action on the end terms is trivial. If* $\mathcal{X}/\Delta$ *is proper, then* $H^{n+1}_{\mathrm{ph}}(X_0) := \ker(\mathrm{sp}) \subseteq H^{n+1}(X_0)$ *is pure of weight* $n+1$.

The decomposition of terms in (1.10) into irreducible representations for $\mathcal{G} \times \mu_\kappa$ only becomes useful if we understand the action on the vanishing cohomology $\bigoplus_{x\in\Sigma} V_x$ for a given collection of singularities. In particular, if $gx = x$ then we need to further refine the spectrum under the resulting automorphism $g^* : V_x \to V_x$ of T-MHS.

1.11. Definition. Write $G \leq \mathrm{stab}(x) \leq \mathcal{G}$, and $\mathcal{R}_G$ for the set of complex irreducible representations of G. The *G-spectrum* $\sigma^G_{f,x}$ of x is the element

$$\sum_{(\alpha,w,U)} m^{f,x,G}_{\alpha,w,U}(\alpha, w, U)$$

of the free abelian group $\mathbb{Z}\langle \mathbb{Q} \times \mathbb{Z} \times \mathcal{R}_G\rangle$, where (for each (α, w))

$$V^{\lfloor\alpha\rfloor, w-\lfloor\alpha,\rfloor}_{x,\boldsymbol{e}(\alpha)} \cong \bigoplus_{U\in\mathcal{R}_G} U^{\oplus m^{f,x,G}_{\alpha,w,U}}$$

as G-representations.

In the special case where $G = \langle g\rangle \cong \mu_\ell$ is cyclic, the $\mathbb{C}$-irreps are characters indexed by the power $\zeta_\ell^c = e^{2\pi \boldsymbol{i}(c/\ell)}$ of ζ_ℓ to which g is sent.

1.12. Definition. The *eigenspectrum* of an isolated singularity x with automorphism g is the element

$$\sigma^g_{f,x} = \sum_{(\alpha,w,\gamma)} m^{f,x,g}_{\alpha,w,\gamma}(\alpha, w, \gamma) \in \mathbb{Z}\langle \mathbb{Q} \times \mathbb{Z} \times \mathbb{Q}/\mathbb{Z}\rangle,$$

where $m^{f,x,g}_{\alpha,w,\gamma}$ is the dimension of the eigenspace $(V^{\lfloor\alpha\rfloor,w-\lfloor\alpha,\rfloor}_{x,\boldsymbol{e}(\alpha)})^{\boldsymbol{e}(\gamma)} \subseteq V^{\lfloor\alpha\rfloor,w-\lfloor\alpha\rfloor}_{x,\boldsymbol{e}(\alpha)}$ for g^* with eigenvalue $\boldsymbol{e}(\gamma) = e^{2\pi\boldsymbol{i}\gamma}$.

[2]Again, this means that the action of $\mathcal{G}$ and T_{ss} on the MHSs (as automorphisms of MHS) commute with each other and with sp, can, and δ.

1.13. Remark. For $\mathcal{X}/\Delta$ proper (with hypotheses as in Corollary 1.9), $H^n(X_t)$ is a VHS on Δ^* whose automorphism group contains $\mathcal{G}$. For any field extension $K/\mathbb{Q}$, this decomposes as K-VHS into a direct sum of $\mathcal{G}$-isotypical components, corresponding to K-irreps of $\mathcal{G}$. The $\mathcal{G}$-action on and decomposition of $H^n_{\mathrm{lim}}(X_t)$ obtained by taking limits are the same as those arising from the $\mathcal{G}$-MHS structure on $H^n_{\mathrm{lim}}(X_t)$ in Corollary 1.9.

We now turn to the explicit computation of these eigenspectra in the simplest case.

2. Quasihomogeneous singularities with automorphism

Let $F \in \mathbb{C}[z_1, \ldots, z_{n+1}]$ (with $n > 0$) be a *quasihomogeneous polynomial* with an isolated singularity at the origin $\underline{0}$. That is to say, choosing a weight vector $\underline{w} = (w_1, \ldots, w_{n+1}) \in \mathbb{Q}^{n+1}_{>0}$ and setting

$$\mathfrak{M}_{\underline{w}} := \{\underline{m} \in \mathbb{Z}^{n+1}_{\geq 0} \mid \underline{m} \cdot \underline{w} = 1\},$$

we have

$$F = \sum_{\underline{m} \in \mathfrak{M}_{\underline{w}}} a_{\underline{m}}\, \underline{z}^{\underline{m}} \tag{2.1}$$

for some $a_{\underline{m}} \in \mathbb{C}$. We recall that the *degree* κ_F of F is the least integer such that $\kappa_F\, w_i \in \mathbb{N}$ for $i = 1, \ldots, n+1$; define $w_i := \kappa_F\, w_i$ and set $\underline{\kappa} := (\kappa_1, \ldots, \kappa_{n+1})$.

Next recall the setting of Definition 1.8, where $f : \mathcal{X} \to \Delta$ is a holomorphic map with quasiprojective fibers and smooth total space, with X_t smooth for $t \neq 0$ and $\mathrm{sing}(X_0) =: \Sigma$ finite. A singularity $x \in \Sigma \subset X_0$ is *quasihomogeneous* if f can be locally analytically identified with (2.1) for some $\underline{w}$. In that case, V_x and $\sigma_{f,x}$ identify with the vanishing cohomology

$$V_F := H^0\, \imath^*_{\underline{0}}\, \phi_F\, \mathbb{Q}_{\mathbb{C}^{n+1}} \tag{2.2}$$

of $F : \mathbb{C}^{n+1} \to \mathbb{C}$ at $\underline{0}$, and its mixed spectrum σ_F. These were first computed by Steenbrink [1977], and we briefly review the treatment from [Kerr and Laza 2023, Section 2] before passing to eigenspectra.

Writing

$$J_F := \left(\frac{\partial F}{\partial z_1}, \ldots, \frac{\partial F}{\partial z_{n+1}}\right) \subseteq \mathbb{C}[\underline{z}]$$

for the Jacobian ideal, let $\mathcal{B} \subset \mathbb{Z}^{n+1}_{\geq 0}$ be chosen so that the monomials $\{\underline{z}^{\underline{\beta}}\}_{\underline{\beta} \in \mathcal{B}}$ provide a basis of $\mathbb{C}[\underline{z}]/J_F$. Write $\mu_F := |\mathcal{B}|$ for the *Milnor number* of F, and $\ell(\underline{\beta}) := \frac{1}{\kappa_F} \sum_{i=1}^{n+1} \kappa_i(\beta_i + 1) = \sum_{i=1}^{n+1} w_i(\beta_i + 1)$.

2.3. Proposition. *We have* $\mu_F = \dim V_F$ *for the Milnor number and*

$$\sigma_F = \sum_{\underline{\beta} \in \mathcal{B}} (\alpha(\underline{\beta}), w(\underline{\beta})) \in \mathbb{Z}\langle \mathbb{Q} \times \mathbb{Z} \rangle$$

for the mixed spectrum, where $\alpha(\underline{\beta}) := n+1-\ell(\underline{\beta})$ *and* $w(\underline{\beta}) := n$ (*resp.* $n+1$) *if* $\alpha(\underline{\beta}) \notin \mathbb{Z}$ (*resp.* $\in \mathbb{Z}$).

Sketch. Perform a base-change followed by weighted blow-up at $\underline{0}$:

$$\begin{tikzcd} \mathbb{C}^{n+1} \arrow[d, "F"'] & \mathfrak{X} \arrow[l] \arrow[d, "\hat{F}"'] & \mathfrak{Y} \arrow[l, "\mathrm{Bl}_{\underline{\kappa}}"'] \arrow[dl, "\tilde{F}"] \\ \mathbb{C} & \Delta \arrow[l] & \\ t^{\kappa_F} & t \arrow[l, mapsto] & \end{tikzcd} \tag{2.4}$$

with exceptional divisor $\mathcal{E} = \{T^{\kappa_F} = F(\underline{Z})\} \subset \mathbb{WP}[1 : \underline{\kappa}] =: \boldsymbol{P}$ (in weighted homogeneous coordinates $T, Z_1, \ldots, Z_{n+1}$). The singular fiber $\mathbb{Y}_0 := \tilde{F}^{-1}(0)$ is the union of $\mathcal{E}$ and the proper transform $\widetilde{\mathbb{X}}_0$ of $\mathbb{X}_0 := F^{-1}(0) = \hat{F}^{-1}(0)$, meeting in

$$E := \mathcal{E} \cap \widetilde{\mathbb{X}}_0 = \{F(\underline{Z}) = 0\} \subset \boldsymbol{H} := \{T = 0\} \; (\cong \mathbb{WP}[\underline{\kappa}]) \subset \boldsymbol{P}.$$

The claim is then that $V_F \cong H^n(\mathcal{E} \setminus E)$, which can be checked using (1.5) with $\pi = \mathrm{Bl}_{\underline{\kappa}}$. Since E [resp. $\underline{0}$] is a deformation retract of $\mathbb{Y}_0$ [resp. $\mathbb{X}_0$], while $\mathbb{Y}_t = \mathbb{X}_t$ for $t \neq 0$, and $\phi_{\tilde{F}}\, \mathbb{Q}_{\mathcal{Y}} \simeq \imath_*^E \mathbb{Q}_E(-1)[-1]$ (see [Kerr et al. 2021, 6.3 and 8.3–8.4]), the diagram becomes

$$\begin{tikzcd} & 0 \arrow[r] & H^n_{\lim}(\mathbb{X}_t) \arrow[r, "\cong"] \arrow[d, equal] & V_F \arrow[r] \arrow[d, "\mathrm{Bl}^*"] & 0 \\ H^{n-2}(E)(-1) \arrow[r] & H^n(\mathcal{E}) \arrow[r] & H^n_{\lim}(\mathbb{Y}_t) \arrow[r] & H^{n-1}(E)(-1) \arrow[r] & H^{n+1}(\mathcal{E}) \end{tikzcd}$$

whence the result.

Next, one constructs a basis of $H^n(\mathcal{E} \setminus E)$ from $\mathcal{B}$, using residue theory. Writing (with $T := Z_0$)

$$\Omega_{\boldsymbol{P}} = \sum_{j=0}^{n+1} (-1)^j Z_j \, dZ_0 \wedge \cdots \wedge \widehat{dZ_j} \wedge \cdots \wedge dZ_{n+1},$$

for each $\underline{\beta} \in \mathcal{B}$ we set (with $\underline{Z}^{\underline{\beta}} = Z_1^{\beta_1} \cdots Z_{n+1}^{\beta_{n+1}}$)

$$\Omega_{\underline{\beta}} := \frac{T^{\kappa_F} \underline{Z}^{\underline{\beta}} \, \Omega_{\boldsymbol{P}}}{T(F(\underline{Z}) - T^{\kappa_F})^{\lceil \ell(\underline{\beta}) \rceil}} \in \Omega^{n+1}(\boldsymbol{P} \setminus \mathcal{E} \cap \boldsymbol{H}) \tag{2.5}$$

and $\omega_{\underline{\beta}} := \mathrm{Res}_{\mathcal{E} \setminus E}([\Omega_{\underline{\beta}}]) \in H^n(\mathcal{E} \setminus E)$. See [Kerr and Laza 2023, Theorem 2.2] for the proof that this has (p, q)-type $(\lfloor \alpha(\underline{\beta}) \rfloor, \lfloor \ell(\underline{\beta}) \rfloor)$, and [Steenbrink 1977, Theorem 1] for the assertion that the $\{\omega_{\underline{\beta}}\}$ give a basis. Note that $\lfloor \alpha(\underline{\beta}) \rfloor + \lfloor \ell(\underline{\beta}) \rfloor = w(\underline{\beta})$.

Finally, the action of T_{ss} is computed by $T \mapsto \zeta_{\kappa_F} T$, or equivalently (in weighted projective coordinates) by $Z_i \mapsto \zeta_{\kappa_F}^{-\kappa_i} Z_i = e^{-2\pi \boldsymbol{i} w_i} Z_i$. Clearly the effect of this on (2.5) is to multiply it by $e^{2\pi \boldsymbol{i} \sum w_i(\beta_i+1)} = e^{2\pi \boldsymbol{i} \alpha(\underline{\beta})}$, as desired. $\square$

Now given a finite group $G \leq \mathrm{Aut}(\mathcal{X}/\Delta)$ fixing $x \in \Sigma$, we can always choose local holomorphic coordinates on which the action is linear [Cartan 1954]. So for a given $g \in G$, we can choose coordinates to make the action diagonal, through roots of unity. Accordingly, we shall compute the eigenspectrum in the case where $g \in \mathrm{Aut}(\mathbb{C}^{n+1}, \underline{0})$ is given by

$$g(z_1, \ldots, z_{n+1}) := (\zeta_\ell^{c_1} z_1, \ldots, \zeta_\ell^{c_{n+1}} z_{n+1}) \tag{2.6}$$

and $F \in \mathbb{C}[\underline{z}]^{\langle g \rangle}$ is a g-invariant quasihomogeneous polynomial. In fact, taking $\mathcal{B} \subset \mathbb{Z}^{n+1}_{\geq 0}$ as above, we have:

2.7. Corollary. *The eigenspectrum σ_F^g is given by*

$$\sum_{\underline{\beta} \in \mathcal{B}} (\alpha(\underline{\beta}), w(\underline{\beta}), \gamma(\underline{\beta})) \in \mathbb{Z}\langle \mathbb{Q} \times \mathbb{Z} \times \mathbb{Q}/\mathbb{Z} \rangle,$$

where $\gamma(\underline{\beta}) := \frac{1}{\ell} \sum_{i=1}^{n+1} c_i(\beta_i + 1)$.

Proof. We only need to compute the action of g^* on $\omega_{\underline{\beta}}$, which is to say the effect of $Z_i \mapsto \zeta_\ell^{c_i} Z_i$ on $\underline{Z}^{\underline{\beta}}\, \Omega_{\underline{\beta}}$. This is just multiplication by $\zeta_\ell^{\sum c_i(\beta_i+1)} = e^{2\pi \mathbf{i} \gamma(\underline{\beta})}$. □

2.8. Example. For a Brieskorn–Pham singularity $F = \sum_{i=1}^{n+1} z_i^{\lambda_i}$, $\lambda_i = 1/w_i = \kappa_F/\kappa_i$, we have $\mathcal{B} = \times_{i=1}^{n+1} \{\mathbb{Z} \cap [0, d_i - 2]\}$. Hence, writing $\Gamma_m = \sum_{j=1}^{m-1} [j/m]$ in the group ring $\mathbb{Z}[\mathbb{Q}]$ (with product $*$), we have

$$\sum_{\underline{\beta} \in \mathcal{B}} [\alpha(\underline{\beta})] = \Gamma_{\lambda_1} * \cdots * \Gamma_{\lambda_{n+1}}.$$

This extends to

$$\sum_{\underline{\beta} \in \mathcal{B}} [(\alpha(\underline{\beta}), \gamma(\underline{\beta}))] = \widetilde{\Gamma}_{\lambda_1}\Big(\frac{c_1}{\ell}\Big) * \cdots * \widetilde{\Gamma}_{\lambda_{n+1}}\Big(\frac{c_{n+1}}{\ell}\Big)$$

in the group ring $\mathbb{Z}[\mathbb{Q} \times (\mathbb{Q}/\mathbb{Z})]$ if we write $\widetilde{\Gamma}_m\big(\frac{c}{\ell}\big) = \sum_{j=1}^{m-1} \big[\big(\frac{m-j}{m}, \frac{jc}{\ell}\big)\big]$.

2.9. Example. As a *specific* example, consider $F = z_1^2 + z_2^2 + z_3^{m+1} + z_4^3$, with $g(z_1, z_2, z_3, z_4) := (z_1, z_2, z_3, \zeta_3 z_4)$. Applying Example 2.8 to compute the eigenspectrum gives

$$\sum_{j=1}^{m} \Big[\Big(\tfrac{5}{3} + \frac{j}{m+1}, \tfrac{1}{3}\Big)\Big] + \sum_{j=1}^{m} \Big[\Big(\tfrac{4}{3} + \frac{j}{m+1}, \tfrac{2}{3}\Big)\Big].$$

We can interpret this scenario as a local snapshot of a $3:1$ cover of $\mathbb{P}^3$ branched over a cubic surface acquiring an A_m singularity. So the ζ_3-eigenspace of the $(1,2)$-part of vanishing cohomology has rank equal to the number of j's for which $\frac{5}{3} + j/(m+1) < 2$. Since the ζ_3-eigenspace of the general fiber ($=$ cubic 3-fold) has Hodge numbers $h^{1,2} = 1$ and $h^{2,1} = 4$, from $\frac{5}{3} + \frac{2}{7} < 2$ we see that m *cannot* be ≥ 6. This bound is sharp, since A_5 can occur on a cubic surface in the form $z_1^3 + z_2^3 - z_2 z_3^2$ (see, for example, [Sakamaki 2010]).

Applying the vanishing-cycle analysis directly on a cubic surface, without passing to a triple cover and using eigenspectra, does *not* rule out A_6. It was this sort of phenomenon that motivated this paper.

2.10. Remark. The eigenspectrum of a μ-constant (semiquasihomogeneous) deformation of (F, γ) remains constant. Even in the more general case of [Kerr and Laza 2023, Section 5.2], one can in principle still use the action of γ^* on the (local) Jacobian ring $\mathcal{O}_{n+1}/J_F$ to refine σ_F to σ_F^g. But Corollary 2.7 (and quasihomogeneous deformations of Example 2.8) will suffice for our purposes below.

3. Bounding nodes on Calabi–Yau hypersurfaces

It is a classical problem to bound the number of nodes (ordinary double points) on a projective hypersurface, especially for Calabi–Yau (CY) varieties. In this section, we use eigenspectra to produce such a bound for hypersurfaces in many weighted projective spaces (Example 3.8). Though our emphasis is on CY varieties for illustrative purposes, it is not limited to them. In the special case of projective space, our formula recovers the bound conjectured by Arnol'd [1981] and proved by Varchenko [1983] (see also [van Straten 2020]) by applying his semicontinuity theorem to the Bruce deformation. This includes the famous bound of 135 for a quintic threefold; see Examples 3.10.

Let $\mathbb{W} = \mathbb{WP}[e_0 : \cdots : e_{n+1}]$ be a weighted projective $(n+1)$-space with finitely many singularities.[3] Suppose we want to bound (numbers and types of) singularities on a hypersurface $X_0 = \{F_0(\underline{W}) = 0\} \subset \mathbb{W}$ of degree d, where a smooth such hypersurface would have Hodge numbers $\underline{h} = (h^{n,0}, h^{n-1,1}, \ldots, h^{0,n})$. Write $d_i = d/e_i$ for $i = 0, \ldots, n+1$.

We shall assume that the singularities of X_0 are all isolated. Taking a general deformation $F_t = F_0 + tG$ to produce a family of $f : \mathcal{X} \to \Delta$ with smooth total space, the vanishing-cycle sequence

$$0 \to H^n(X_0) \to H^n_{\lim}(X_t) \to \bigoplus_{x \in \Sigma} V_x \xrightarrow{\delta} H^{n+1}_{\mathrm{ph}}(X_0) \to 0 \tag{3.1}$$

offers a naive such bound: first, by Schmid's nilpotent orbit theorem, the rank of Gr_F^p remains constant in the limit, giving the second equality of

$$h^{p,n-p} = h^{p,n-p}(X_t) = \sum_q h^{p,q}_{\lim}(X_t) \geq \sum_q h^{p,q}(\ker(\delta)). \tag{3.2}$$

Moreover, the mixed spectrum $\sigma_{f,x}$ tells us the $h_\zeta^{p,q}(V_x) = \dim(V_{x,\zeta}^{p,q})$ (for each eigenvalue ζ of T_{ss}), and only the $V_{x,1}^{p,n+1-p}$ can map nontrivially under δ. Since

[3] We may assume (without loss of generality) that no $n+1$ of the e_i have a common factor.

the hyperplane class also has T_{ss}-eigenvalue 1, equation (3.2) forces

$$\sum_q \sum_{\zeta\neq 1} \dim(V_{x,\zeta}^{p,q}) \leq h_{\mathrm{pr}}^{p,n-p}.$$

When x is a node, i.e., $f \overset{\text{loc}}{\sim} \sum_{i=1}^{n+1} z_i^2$, Proposition 2.3 gives $V_{x,\mathbb{C}} = V_{x,-1}^{(n/2),(n/2)}$ for n even and $V_{x,1}^{(n+1)/2,(n+1)/2}$ for n odd. In the latter case, (3.2) yields no immediate bound on the number of nodes (though one does have results like [Kerr and Laza 2023, Theorem 2.9 and Corollary 2.11]). For $n = 2m$ even, (3.2) yields[4]

$$h_{\mathrm{pr}}^{(n/2),(n/2)}(X_t) = \text{coefficient of } \left[\frac{n}{2}+1\right] \text{ in } \Gamma_{d_0} * \Gamma_{d_1} * \cdots * \Gamma_{d_{n+1}} \tag{3.3}$$

as a bound, which while better than nothing is rather weak.

3.4. Example. The simplest nontrivial case is given by $\mathbb{W} = \mathbb{P}^3$ ($n = 2$) and ($d_0 = d_1 = d_2 = d_3$) $d = 4$, where

$$\begin{aligned} \Gamma_4^{*4} &= \left(\left[\tfrac{1}{4}\right]+\left[\tfrac{1}{2}\right]+\left[\tfrac{3}{4}\right]\right)^{*4} \\ &= [1]+4\left[\tfrac{5}{4}\right]+10\left[\tfrac{3}{2}\right]+16\left[\tfrac{7}{4}\right]+19[2]+16\left[\tfrac{9}{4}\right]+10\left[\tfrac{5}{2}\right]+4\left[\tfrac{11}{4}\right]+[3] \end{aligned} \tag{3.5}$$

correctly gives $19 = h_{\mathrm{pr}}^{1,1}(X_t)$. This is also a poor bound for the number of nodes on a quartic surface (see Example 3.8).

However, a simple trick can improve the bound while also giving one for odd n:

3.6. Theorem. *The number of nodes on X_0 is bounded by the coefficient, in $\Gamma_{d_0} * \Gamma_{d_1} * \cdots * \Gamma_{d_{n+1}}$, of $\left[\frac{n+1}{2}+\frac{1}{2d}\right]$ if n is even and d is odd, or of $\left[\frac{n+1}{2}+\frac{1}{d}\right]$ otherwise.*

Proof. Let $Y_t = \{F_t(\underline{W}) + W_{n+2}^d = 0\} \subset \mathbb{WP}[\underline{e}:1] =: \widetilde{\mathbb{W}}$ be the cyclic $d:1$-cover of $\mathbb{W}$ branched over X_t, with $g : W_{n+2} \mapsto \zeta_d W_{n+2}$ the cyclic automorphism. By Dolgachev's extension of the Griffiths residue theorem [Dolgachev 1982], a basis for the g^*-eigenspace $H_{\mathrm{pr}}^{n-q+1,q}(Y_t)^{\bar{\zeta}_d^j}$ ($t \neq 0$, $0 \leq j < d$) is given by the Poincaré residue classes

$$\mathrm{Res}_{Y_t}\left(\frac{\underline{W}^{\underline{k}-\underline{1}}\, W_{n+2}^{d-j-1}\, \Omega_{\widetilde{\mathbb{W}}}}{(F_t + W_{n+2}^d)^{q+1}}\right),$$

with $k_i \in \mathbb{Z} \cap (0, d_i)$ for $i = 0, \ldots, n+1$ and weights of numerator and denominator equal, that is, $\sum_{i=0}^{n+1} e_i k_i + (d-j) = (q+1)\,d$, or equivalently (dividing by d)

$$\sum_{i=0}^{n+1} \frac{k_i}{d_i} = q + \frac{j}{d}.$$

Hence $\dim \mathrm{Gr}_F^{n-q+1} H_{\lim}^{n+1}(Y_t)^{\bar{\zeta}_d^j} = h^{n-q+1,q}(Y_t)^{\bar{\zeta}_d^j}$ is given (for $0 < j < d$) by the coefficient of $[q + j/d]$ in $\Gamma_{d_0} * \cdots * \Gamma_{d_{n+1}}$.

[4]This is by the same residue theory as used in the proof of Theorem 3.6 below. The notation $*$ is from Example 2.8.

Each node $x \in X_0$ becomes an A_{d-1} singularity $y \in Y_0$, with eigenspectrum $\sum_{j=1}^{d-1}((n+1)/2+j/d, n+1, -j/d)$ unless n is even and d is even (in which case the middle entry is $n+2$ at $j=d/2$). If r is the number of nodes, applying equations (3.1)–(3.2) to $\mathcal{Y}$ and refining by g^*-eigenspaces therefore yields $h^{p_j,q_j}(Y_t)^{\bar{\zeta}_d^j} \geq r$ (for $0<j<d$), where $p_j = \lfloor (n+1)/2 + j/d \rfloor$ and $q_j = n+1-p_j$. Taking $j=1$ if n is odd and $j = \lceil (d+1)/2 \rceil$ if n is even (so that $p_j = (n+1)/2$ resp. $n/2+1$) yields $q_j + j/d = (n+1)/2 + 1/d$ resp. $n/2 + (1/d)\lceil (d+1)/2 \rceil$, hence the claimed bound. $\square$

3.7. Remark. As mentioned above, when $\mathbb{W} = \mathbb{P}^{n+1}$ this recovers Varchenko's [1983] bound. While Varchenko also uses the "cyclic-cover trick", our approach avoids the use of deformations and semicontinuity.

3.8. Example. For CY hypersurfaces in $\mathbb{P}^{n+1}$ ($d = n+2$), Theorem 3.6 yields the bounds 3, 16, 135, 1506, and 20993 for $n = 1, 2, 3, 4, 5$, the first two of which are sharp.[5] (This is also better than what (3.3) yields for $n=2$ and 4, namely 19 and 1751.) It is still not known whether 135 is sharp for quintic 3-folds. The well-known Fermat pencil has fiber $W_0^5 + \cdots + W_4^5 = 5W_0 \cdots W_4$, with $125 = |(\mathbb{Z}/5\mathbb{Z})^3|$ nodes, while the example of van Straten [1993] with 130 nodes remains the record.

3.9. Remark. For $n=2$, the following bound by Miyaoka [1984] sometimes yields better results. If X is any smooth projective surface which is smooth except at r nodes, and K_X is nef, then $r \leq 8\chi(\mathcal{O}_X) - \frac{8}{9}K_X^2$.

(a) For $X \subset \mathbb{P}^3$ a surface of degree d, this yields the bound

$$\tfrac{4}{3}(d-1)(d-2)(d-3) + 8 - \tfrac{8}{9}d(d-4)^2 = \tfrac{4}{9}d(d-1)^2,$$

which is better than Theorem 3.6 for $d \geq 6$ even or $d \geq 15$ odd. A case in point is $d=6$, where we get 85 by (3.3), 68 by Theorem 3.6, and 66 by [Miyaoka 1984]; this was further reduced to 65 (which is sharp) by a clever use of coding theory [Jaffe and Ruberman 1997]. Another is $d=8$, where we get $r \leq 174$.

(b) As a weighted projective example, one can consider surfaces X of degree 10 in $\mathbb{WP}[1:1:1:2]$. We have $\chi(\mathcal{O}_X) = 1 + h^2(\mathcal{O}_X) = 35$ and

$$(K_X \cdot K_X)_X = (X \cdot (X + K_{\mathbb{W}})^2)_{\mathbb{W}} = \frac{10(10-5)^2}{1 \cdot 1 \cdot 1 \cdot 2} = 125,$$

and hence $r \leq \left\lfloor \frac{1520}{9} \right\rfloor = 168$.

3.10. Examples. We consider some CY 3-fold hypersurfaces with r nodes in weighted projective 4-folds.

---

[5]The union of 3 lines in $\mathbb{P}^2$ has 3 nodes, and a Kummer quartic $K3$ in $\mathbb{P}^4$ has 16 nodes. The bounds here are the coefficients of $\left[\frac{n+1}{2} + \frac{1}{n+2}\right]$ in $\Gamma_{n+2}^{*(n+2)}$, e.g., 16 is the coefficient of $\left[\frac{7}{4}\right]$ in (3.5).

(i) $X_0 \subset \mathbb{WP}[1:1:1:1:2]$ of degree 6: Theorem 3.6 yields $r \leq 137$, while the "Fermat pencil" type example $W_0^6 + \cdots + W_3^6 + W_4^3 = 3 \cdot 2^{2/3} W_0 \cdots W_4$ has $|((\mathbb{Z}/6\mathbb{Z})^3 \times \mathbb{Z}/3\mathbb{Z})/(\mathbb{Z}/6\mathbb{Z})| = 108$ nodes.

(ii) $X_0 \subset \mathbb{WP}[1:1:1:1:4]$ of degree 8: the Theorem yields $r \leq 180$, while $W_0^8 + \cdots + W_3^8 + W_4^2 = 4W_0 \cdots W_4$ has $|((\mathbb{Z}/8\mathbb{Z})^3 \times \mathbb{Z}/2\mathbb{Z})/(\mathbb{Z}/8\mathbb{Z})| = 128$ nodes. Here we can improve both the bound and example, since X_0 is (by the quadratic formula) a double-cover of $\mathbb{P}^3$ branched along an r-nodal octic surface. So Remark 3.9(a) gives $r \leq 174$, while the Endrass [1997] example has $r = 168$.

(iii) $X_0 \subset \mathbb{WP}[1:1:1:2:5]$ of degree $d = 10$: Theorem 3.6 yields $r \leq 169$, but because these are double covers of $\mathbb{WP}[1:1:1:2]$ branched along an r-nodal dectic surface, Remark 3.9(b) reduces the bound to 168. The standard example is $W_0^{10} + W_1^{10} + W_2^{10} + W_3^5 + W_4^2 = 2^{4/5}\, 5^{1/2}\, W_0 \cdots W_4$, but this has only 100 nodes. One can do somewhat better by taking the preimage of a Togliatti quintic [Beauville 1980] (with 31 nodes avoiding the coordinate axes) under

$$\mathbb{WP}[1:1:1:2] \xrightarrow{1:2}\!\!\!\!\!\rightarrow \mathbb{WP}[1:1:2:2] \xrightarrow{1:2}\!\!\!\!\!\rightarrow \mathbb{WP}[1:2:2:2] \cong \mathbb{P}^3,$$

to get $4 \cdot 31 = 124$.

In the case of a symmetric hypersurface $X_0 \subset \mathbb{P}^{n+1}$, cut out by $F_0 \in \mathbb{C}[\underline{W}]^{\mathfrak{S}_{n+2}}$ (homogeneous of degree d), one can consider the family $\mathcal{Y} \to \Delta$ of d-fold cyclic covers branched along an $\mathfrak{S}_{n+2}$-invariant smoothing $\mathcal{X} \to \Delta$. A full accounting of this story gets into G-spectra ($G \cong \mu_d \times \mathrm{stab}_{\mathfrak{S}_{n+2}}(x)$) of the resulting A_{d-1} singularities of Y_0. This leads to constraints, via character theory of $\mathfrak{S}_{n+2}$, on how Σ can be built out of $\mathfrak{S}_{n+2}$-orbits. (However, it does not, for example, *rule out* the possibility of 135 nodes on an $\mathfrak{S}_5$-symmetric quintic threefold.) Here we shall only give the simplest result in this direction:

3.11. Theorem. *A symmetric CY hypersurface in* $\mathbb{P}^{n+1}$ (*of degree* $d = n+2$) *with isolated singularities cannot contain a node with trivial stabilizer in* $\mathfrak{S}_{n+2}$.

Proof. Suppose otherwise; then Y_0 has a set of $(n+2)!$ A_{n+1} singularities with eigenspectra

$$\sum_{j=1}^{n+1} \left(\frac{n+1}{2} + \frac{j}{n+2},\, n+1,\, \frac{-j}{n+2} \right),$$

contributing a subspace V of dimension $(n+2)!$ to the g^*-eigenspace[6] $H^{n+1}_{\mathrm{van}}(Y_t)^{\zeta_{n+2}}$. It is closed under the action of $\mathfrak{S}_{n+2}$, and the triviality of the stabilizers of these A_{n+1} singularities means that the trace of any $\sigma \in \mathfrak{S}_{n+2} \setminus \{1\}$ is zero. So V is a copy of the regular representation of $\mathfrak{S}_{n+2}$, which belongs to

$$\ker(\delta) \subseteq H^{(n+1)/2,(n+1)/2}_{\mathrm{van}}(Y_t)^{\zeta_{n+2}}.$$

[6]As before, $g : W_{n+2} \mapsto \zeta_{n+2} W_{n+2}$ denotes the cyclic automorphism of Y_t.

By the compatibility[7] of the vanishing-cycle sequence for $\mathcal{Y}$ with g^* and $\mathfrak{S}_{n+2}$, this forces a copy of the regular representation in $H_{\lim}^{(n+1)/2,(n+1)/2}(Y_t)^{\zeta_{n+2}}$, hence $H^{(n+1)/2,(n+1)/2}(Y_t)^{\zeta_{n+2}}$ for $t \neq 0$ (as $\mathfrak{S}_{n+2}$ acts on the VHS, compatibly with taking limits, see Remark 1.13).

Now $U := \mathcal{H}^{(n+1)/2,(n+1)/2}(Y_t)^{\zeta_{n+2}}$ has a basis of the form

$$\eta_{\underline{k}} := \mathrm{Res}_{Y_t}\left(\frac{\underline{W}^{\underline{k}-\underline{1}}\,\Omega_{\mathbb{P}^{n+2}}}{\left(F_0(\underline{W}) + W_{n+2}^{n+2}\right)^{(n+3)/2}}\right),$$

where $0 < k_i < n+2$ (for $i = 0, \ldots, n+1$) and (for equality of weights of numerator and denominator) $\left(\sum_{i=0}^{n+1} k_i\right) + 1 = \frac{n+3}{2}(n+2)$. Here $\mathfrak{S}_{n+2}$ acts trivially on the denominator, through the sign representation χ on $\Omega_{\mathbb{P}^{n+2}}$, and by permutations of the W_i on $\underline{W}^{\underline{k}-\underline{1}}$. We claim that U contains no copy of the trivial representation, *a fortiori* of the regular representation, furnishing the desired contradiction.

Clearly it is equivalent to show that the representation of $\mathfrak{S}_{n+2}$ on the $\mathbb{C}$-span $\tilde{U}$ ($\cong U \otimes \chi$) of the monomials $\{\underline{W}^{\underline{k}}\}_{\underline{k}\text{ as above}}$ contains no copy of χ. Suppose $o := \mathfrak{S}_{n+2}.\underline{W}^{\underline{k}}$ is an orbit and $\tilde{U}_o \subseteq \tilde{U}$ its span. By Burnside's lemma,

$$\frac{1}{(n+2)!} \sum_{g \in \mathfrak{S}_{n+2}} |o^g| = 1.$$

On the other hand, $\underline{k} = (k_0, \ldots, k_{n+1})$ contains a repeated entry since there are only $n+1$ choices for each k_i; hence for some transposition τ, $|o^\tau| \neq 0$. Since $\mathrm{sgn}(\tau) = -1$, this forces

$$\frac{1}{(n+2)!} \sum_{g \in \mathfrak{S}_{n+2}} \mathrm{sgn}(g)\,|o^g|,$$

which computes the number of copies of χ in $\tilde{U}_o$, to be zero. □

For $n = 1$ or 2 this result is obvious (since $6 > 3$ and $24 > 16$), but for $n = 3$, 4, or 5 it is less so (as $120 < 135$, $720 < 1506$, and $5040 < 20993$). In particular, since the examples of quintic 3-folds with 125 and 130 nodes are $\mathfrak{S}_5$-symmetric, and the latter has a 60-node orbit, it is interesting that a 120-node orbit is impossible.

4. Cyclic covers of $\mathbb{P}^1$

In the final two sections we turn to "codimension-one" monodromy phenomena for period maps arising from cyclic covers. We begin with a story that generalizes elliptic curves and goes back to Deligne and Mostow [1986] (see also [Moonen 2018]). Given distinct points $t_1, \ldots, t_{2m} \in \mathbb{P}^1$ (with projective coordinates $[S_i : T_i]$), define

$$C_{\underline{t}} := \left\{[Z_0 : Z_1 : Z_2] \in \mathbb{P}[1:1:2] \mid Z_2^m = \prod_{i=1}^{2m}(S_i Z_1 - T_i Z_0)\right\},$$

[7] This is nothing but Corollary 1.9 with $\mathcal{G} = \langle g \rangle \times \mathfrak{S}_{n+2}$.

with automorphism $g([Z_0 : Z_1 : Z_2]) := [Z_0 : Z_1 : \zeta_m Z_2]$. For $m = 2, 3, 4$, or 6, the sum of g^*-eigenspaces $H^1(C_{\underline{t}})^{\zeta_m} \oplus H^1(C_{\underline{t}})^{\bar{\zeta}_m}$ produces a $\mathbb{Q}$-VHS over $M_{0,2m}$,[8] and hence a period map to an arithmetic ball quotient $\Gamma \backslash \mathbb{B}_{2m-3}$. This turns out to be injective,[9] and extends to an isomorphism between GIT resp. Hassett/KSBA compactifications of $M_{0,2m}$ and Baily–Borel resp. toroidal compactifications of the ball quotient [Deligne and Mostow 1986; Gallardo et al. 2021].

So what if $m \neq 2, 3, 4$, or 6? In the discussion that ensues, we will not be concerned with ball quotients or even the period map per se, but only with

- the $\mathbb{Q}$-VHS $\mathcal{V}$ over $M_{0,2m}$ arising from $H^1(C_{\underline{x}})$,
- its sub-$\mathbb{C}$-VHSs $\mathcal{V}^{\zeta_m^j} := \ker(g^* - \zeta_m^j I)$ $(1 \leq j \leq m-1)$, and
- their limiting behavior along the boundary of the Hassett compactifications $\overline{M}_{0,[(1/m)+\epsilon]_{2m}}$ (see below).

The point is that *these can be considered uniformly for all $m \geq 2$*, not just $m = 2, 3, 4$, and 6. Moreover, using eigenspectra, *we can easily compute LMHS and monodromy types along the Hassett boundary strata*, as we demonstrate in Propositions 4.5–4.6 and Example 4.7. This is the first step toward a global study of the extended period map for this series of examples, which will necessarily go beyond the arithmetic ball quotient setting (see Remark 4.8). We also refer the reader to [Deng and Gallardo 2023], where global partial compactifications of the period maps for some other non-Deligne–Mostow cases are constructed.

To begin with, in affine coordinates $x = Z_1/Z_0,\ y = Z_2/Z_0,\ C_{\underline{t}}$ takes the form

$$y^m = f_{\underline{t}}(x) := \prod_{i=1}^{2m} (x - t_i)$$

[resp. $\prod_{i \neq j}(x - t_i)$ if $t_j = \infty$]. While there are three possibilities for the Newton polytope Δ, they all have the same interior integer points

$$(\Delta \setminus \partial\Delta) \cap \mathbb{Z}^2 = \{(i, j) \mid 1 \leq j \leq m-1,\ 1 \leq i \leq 2(m-j)-1\},$$

which provide a basis of $\Omega^1(C_{\underline{t}})$ via

$$\omega_{(i,j)} := \mathrm{Res}_{C_{\underline{t}}}\left(\frac{x^{i-1} y^{j-1}\, dx \wedge dy}{y^m - f_{\underline{t}}(x)}\right).$$

Since $g^*\omega_{(i,j)} = \zeta_m^j \omega_{(i,j)}$, we find that

$$\begin{cases} \mathrm{rk}(\mathcal{V}^{\zeta_m^j})^{1,0} = 2(m-j)-1, & \mathrm{rk}(\mathcal{V}^{\zeta_m^j})^{0,1} = 2j-1, \\ \mathrm{rk}\,\mathcal{V}^{\zeta_m^j} = 2m-2, & \mathrm{rk}\,\mathcal{V} = 2(m-1)^2. \end{cases} \tag{4.1}$$

---

[8] $M_{0,n}$ parametrizes *ordered* n-tuples of distinct points on $\mathbb{P}^1$ modulo the action of $\mathrm{PSL}_2(\mathbb{C})$.

[9] For $m = 6$ one has to quotient $M_{0,12}$ by $\mathfrak{S}_{12}$; see [Gallardo et al. 2021].

For example, if $m = 5$, then $C_{\underline{t}}$ has genus 12; and $\mathcal{V}_{\mathbb{C}}$ decomposes into four $\mathbb{C}$-VHSs $\{\mathcal{V}^{\zeta_5^j}\}_{j=1}^4$ with respective Hodge numbers (7, 1), (5, 3), (3, 5), and (1, 7).

4.2. Definition [Hassett 2003]. A *weighted stable rational curve* for the weight $\underline{\mu} := (\mu_1, \dots, \mu_n) \in \{(0, 1] \cap \mathbb{Q}\}^{\times n}$ is a pair[10] $(\mathcal{C}, \sum \mu_i p_i)$ with:

- $\mathcal{C}$ a nodal connected projective curve of arithmetic genus 0.
- Each p_i a smooth point of $\mathcal{C}$.
- If $p_{i_1} = \cdots = p_{i_r}$, then $\mu_{i_1} + \cdots + \mu_{i_r} \leq 1$.
- The $\mathbb{Q}$-divisor $K_{\mathcal{C}} + \sum_{i=1}^n \mu_i p_i$ is ample (i.e., on each irreducible component, the sum of weights plus number of nodes is > 2).

We will write $(\mu, \dots, \mu) =: [\mu]_n$ for repeated weights.

4.3. Theorem [Hassett 2003]. (i) *There exists a smooth projective fine moduli space $\overline{M}_{0,\underline{\mu}}$ parametrizing $\underline{\mu}$-weighted stable rational curves, and containing $M_{0,n}$ as a Zariski-open subset.*

(ii) *Given weights $\underline{\mu} = (\mu_1, \dots, \mu_n)$ and $\underline{\tilde{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_n)$ with $\mu_i \leq \tilde{\mu}_i$ $(\forall i)$, there exists a birational* ***reduction morphism*** *$\pi_{\underline{\tilde{\mu}},\underline{\mu}} : \overline{M}_{0,\underline{\tilde{\mu}}} \twoheadrightarrow \overline{M}_{0,\underline{\mu}}$ contracting all components which violate the ampleness property in Definition 4.2 for the weight $\underline{\tilde{\mu}}$.*

4.4. Remark. (a) $\overline{M}_{0,[1]_n}$ reproduces the Deligne–Mumford–Knudsen compactification $\overline{M}_{0,n}$.

(b) Although the ampleness property forces $\sum \mu_i > 2$, if for $|\underline{\mu}| = 2$ we *define* $\overline{M}_{0,\underline{\mu}}$ to be the GIT quotient $(\mathbb{P}^1)^n /\!\!/_{\underline{\mu}} \mathrm{SL}_2$, then Theorem 4.3(ii) extends to this case; and if we take $\tilde{\mu}_i = \mu_i + \epsilon$ $(\epsilon \in \mathbb{Q},\ 0 < \epsilon \ll 1)$ then $\pi_{\underline{\tilde{\mu}},\underline{\mu}}$ is Kirwan's partial desingularization which blows up the strictly semistable locus.

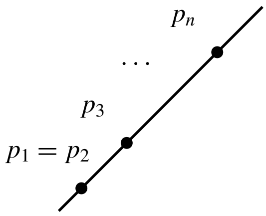
Our interest henceforth is in the equal-weight Hassett compactification

$$\overline{M}^H_{0,2m} := \overline{M}_{0,[(1/m)+\epsilon]_{2m}}$$

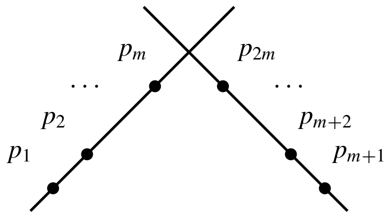
and its morphism π to $\overline{M}^{\mathrm{GIT}}_{0,2m} := \overline{M}_{0,[1/m]_{2m}}$. As the reader may check, the irreducible components of $\overline{M}^H_{0,2m} \setminus M_{0,2m}$ are of two types, parametrizing[11] stable weighted curves as shown (up to reordering of the $\{p_i\}$):

[10] Despite the sum notation, the order of points with equal weights is retained.

[11] More precisely, it is a dense open subset of each component that parametrizes the displayed objects.

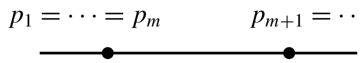


type (A)



type (B)

It is also clear that π preserves the type (A) strata whilst contracting the type (B) ones to a (strictly semistable) point parametrizing the object



The $\mathbb{C}$-VHSs $\mathcal{V}^{\zeta_m^j}$ admit canonical extensions across the smooth part of $\overline{M}^H_{0,2m}\backslash M_{0,2m}$, and we and we shall now compute the LMHS types there.

4.5. Proposition. *Along type* (*A*) *strata*:

• $\mathcal{V}^{\zeta_m^j}_{\mathrm{lim}}$ *is pure of weight* 1, *with* $h^{1,0} = 2m - 2j - 1$ *and* $h^{0,1} = 2j - 1$, *unless* $j = m/2$.

• *If* $j = m/2$, *then* $h^{1,1} = h^{0,0} = 1$, $h^{1,0} = h^{0,1} = m - 1$, *and* $T = e^N$ *(with* N *an isomorphism from the* (1, 1) *to* (0, 0 *part*).

• *If* $j > m/2$ (*resp.* $< m/2$), *then we have the decomposition*

$$\mathcal{V}^{\zeta_m^j}_{\mathrm{lim}} = \mathcal{V}^{\zeta_m^j}_{\mathrm{lim},1} \oplus \mathcal{V}^{\zeta_m^j}_{\mathrm{lim},\bar{\zeta}_m^{2j}}$$

into $T = T_{ss}$*-eigenspaces, where* $\mathcal{V}^{\zeta_m^j}_{\mathrm{lim},\bar{\zeta}_m^{2j}}$ *is* 1*-dimensional of type* (0, 1) (*resp.* (1, 0)).

Proof. Begin by locally modeling (the effect on $C_{\underline{t}}$ of) the collision of two points by $y^m + z^2 = s$, as $s \to 0$. This has eigenspectrum

$$\sum_{j=1}^{m-1} \left(\tfrac{3}{2} - \frac{j}{m}, w(j), \frac{j}{m}\right),$$

where $w(j) = 2$ if $j = m/2$ and 1 otherwise. Next, we apply the vanishing-cycle sequence (with $H^2_{\mathrm{ph}} = \{0\}$ since the degenerate curve remains irreducible) to compute the LMHS. Finally, we perform a base-change by $s \mapsto s^2$ to preserve ordering of points, which squares the eigenvalues of the T_{ss}-action; in other words, we replace $\frac{3}{2} - \frac{j}{m}$ by $\left\{2\left(\frac{3}{2} - \frac{j}{m}\right)\right\} + \left\lfloor \frac{3}{2} - \frac{j}{m} \right\rfloor$ ($\{\cdot\}$ denoting the fractional part), which gives the result. □

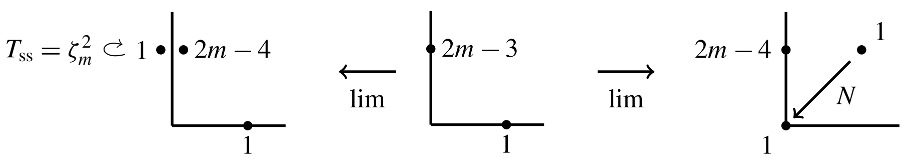
4.6. Proposition. *Along the type* (B) *strata*, *for each* $1 \leq j \leq m - 1$, $\mathcal{V}^{\zeta_m^j}_{\mathrm{lim}}$ *has Hodge numbers* $h^{1,1} = h^{0,0} = 1$, $h^{1,0} = 2m - 2j - 2$, *and* $h^{0,1} = 2j - 2$; N *is an isomorphism from the* (1, 1) *to* (0, 0) *part*, *and* $T = e^N$ *is unipotent.*

Proof. In the GIT compactification for *unordered* points, the degeneration is locally modeled by two copies of $y^m + x^m = s$, each with eigenspectrum

$$\sum_{j=1}^{m-1}\left(1, 2, \frac{j}{m}\right) + \sum_{j=2}^{m-1}\sum_{k=1}^{j-1}\left(\frac{k+m-j}{m}, 1, \frac{j}{m}\right) + \sum_{j=1}^{m-2}\sum_{k=j+1}^{m-1}\left(\frac{k+m-j}{m}, 1, \frac{j}{m}\right).$$

At this point one applies the vanishing-cycle sequence to deduce the form of the LMHS, noting that the degenerate curve is a union of m $\mathbb{P}^1$'s and $H^2_{\mathrm{ph}} \cong \mathbb{Q}(-1)^{\oplus m-1}$. For $\overline{M}^H_{0,2m}$, one then applies the base-change by $s \mapsto s^m$, which trivializes T_{ss}, allowing the extension class to vary along the type (B) stratum. $\square$

4.7. Example. Combining (4.1) with the two propositions, $\mathcal{V}^{\bar{\zeta}_m}$ has Hodge–Deligne diagrams



type (A) type (B)

For $m = 4$ (resp. 6), the monodromy in type (A) is thus given by a complex reflection (resp. "triflection").

4.8. Remark. For any m, we have that $\mathcal{V}^{\bar{\zeta}_m}$ $(\oplus\mathcal{V}^{\zeta_m})$ induces a map from the universal cover $\tilde{M}^{\mathrm{un}}_{0,2m}$ to a ball $\mathbb{B}_{2m-3}$. Moreover, both LMHS types have $2m-4$ complex moduli. However, for m different from 2, 3, 4, or 6, this does not lead to a tidy extended period map: as the projection of the monodromy to $U(1, 2m-3)$ is not discrete [Mostow 1988], the quotient of $\mathbb{B}_{2m-3}$ by this is not Hausdorff.

To circumvent this problem, we must replace $\mathbb{B}_{2m-3}$ by its product with other (nonball) symmetric domains, which receives the image of the period map for the $\mathbb{Q}$-VHS $\oplus_{(j,m)=1}\mathcal{V}^{\zeta_m^j}$. For instance, if $m = 5$ then the real points of the generic Mumford–Tate group of $\mathcal{V}$ take the form $U(1, 7) \times U(3, 5)$, and the full period map lands in a discrete quotient of the product $\mathbb{B}_7 \times \mathrm{I}_{3,5}$.

5. Hyperplane configurations and Dolgachev's conjecture

Both differential and asymptotic methods in Hodge theory can be used to establish that a VHS is "generic" in some sense. In [Gerkmann et al. 2013], differential methods (characteristic varieties and Yukawa couplings) were employed to show that the period map for the family of CY 3-folds $X \overset{2:1}{\twoheadrightarrow} \mathbb{P}^3$ branched over 8 planes does not factor through a locally symmetric variety of the form $\Gamma\backslash \mathrm{SU}(3, 3)/K$. Indeed, the geometric monodromy and Mumford–Tate groups of the corresponding VHS turn out to be as large as they can be (with both equal to the symplectic group Sp_{20}). This was later extended to similarly constructed CY n-fold families [Sheng et al. 2015], see below. Our goal here is to quickly deduce these results using eigenspectra and local monodromy, demonstrating the effectiveness of the asymptotic approach.

Let $L_0, \ldots, L_{2n+1} \subset \mathbb{P}^n$ be hyperplanes defined by linear forms ℓ_i, in general position in the sense that $\bigcup L_i$ is a normal crossing divisor. Consider the $2:1$ cover $X \overset{\pi}{\twoheadrightarrow} \mathbb{P}^n$ branched along $\bigcup L_i$, and the rank-1 $\mathbb{Q}$-local system $\mathbb{L}$ on

$$U = \mathbb{P}^n \setminus \left(\bigcup L_i\right) \overset{\jmath}{\hookrightarrow} \mathbb{P}^n,$$

with monodromy -1 about each L_i. Since X has finite quotient singularities, we have $\mathrm{IC}^\bullet_X = \mathbb{Q}_X[n]$ and[12]

$$H := H^n_{\mathrm{pr}}(X) := \frac{H^n(X)}{\pi^* H^n(\mathbb{P}^n)} \cong H^n(\mathbb{P}^n, \jmath_* \mathbb{L}) \cong \mathrm{IH}^n(\mathbb{P}^n, \mathbb{L}) \tag{5.1}$$

is a pure HS of weight n. By [Dolgachev and Kondō 2007, Lemma 8.2], it has Hodge numbers

$$h^{p,n-p}_{\mathrm{pr}}(X) = \binom{n}{p}^2 \Longrightarrow h^n_{\mathrm{pr}}(X) = \binom{2n}{n}. \tag{5.2}$$

It is polarized by the intersection form Q, which presents no difficulties as X has a smooth finite cover.

Taking $\mathcal{S} \subset (\check{\mathbb{P}}^n)^{2n+2}/\mathrm{PGL}_{n+1}(\mathbb{C}) =: \bar{\mathcal{S}}$ to be the (n^2-dimensional) moduli space of $2n+2$ ordered hyperplanes in $\mathbb{P}^n$ in general position, this construction yields a $\mathbb{Z}$-PVHS $\mathcal{H} \to \mathcal{S}$ of CY-n type with H as reference fiber. Let

$$\rho : \pi_1(\mathcal{S}) \to \mathrm{Aut}(H, Q)^\circ =: M_{\max}$$

be the monodromy representation of $\mathcal{H}$,[13] Π its geometric monodromy group, and M its Hodge (special Mumford–Tate) group. Here Π is the identity connected component of $\widetilde{\Pi} := \overline{\rho(\pi_1(\mathcal{S}))}^{\mathbb{Q}\text{-Zar}}$, and $\Pi \le M \le M_{\max}$. A conjecture attributed by [Sheng et al. 2015] to Dolgachev states that the period map for $\mathcal{H}$ factors through a locally symmetric variety (also n^2-dimensional) of type $I_{n,n}$,[14] which would imply that $\mathfrak{m}_{\mathbb{R}} \cong \mathfrak{su}(n,n)$. This is equivalent to saying that,

(5.3) *up to finite data* (i.e., after passing to a finite cover), *$\mathcal{H}$ is the n-th wedge power of a VHS of weight* 1 *and rank* $2n$.

[12] See [Hotta et al. 2008, Proposition 8.2.30] for the statement that $\mathrm{IC}^\bullet_{\mathbb{P}^n}\,\mathbb{L} = \jmath_*\mathbb{L}[n]$.

[13] Here $(\,\cdot\,)^\circ$ means the identity component as algebraic group (i.e., $\mathrm{SO}(H)$ instead of $\mathrm{O}(H)$ if n is even).

[14] Note that the "tautological VHS" over $I_{n,n}$ is already geometrically realized by the n-th primitive cohomology of a universal family of Weil abelian $2n$-folds.

The conjecture does hold for $n = 1$ and $n = 2$, but this merely reflects exceptional isomorphisms of Lie groups in low rank, namely

$$\mathrm{SU}(1, 1) \cong \mathrm{SL}_2(\mathbb{R}) \quad \text{and} \quad \mathrm{SU}(2, 2) \cong \mathrm{Spin}(2, 4)^+.$$

That is, in both of these cases we *also* have $\Pi \cong M_{\max}$ ($= \mathrm{SL}_2$ resp. $\mathrm{SO}(2, 4)$). For $n \geq 3$, in contrast, the conjecture would have $\Pi < M_{\max}$ a proper algebraic subgroup. In [Sheng et al. 2015, Proposition 8.2.30] (and earlier works [Gerkmann et al. 2007a; 2007b; 2013]), it was shown via quite computationally involved differential methods that in fact the monodromy is maximal for all n, and the conjecture fails for $n \geq 3$:

5.4. Theorem. $\Pi = M = M_{\max}$ *for all* $n \geq 1$.

In the remainder of this section, we explain how asymptotic methods provide a much simpler approach to these results. First we will give a careful argument disproving the conjecture for $n \geq 3$ odd, which *a priori* is a weaker statement than the Theorem in that case. (The relation to the main theme of his paper — specifically, to the setting of Corollary 1.9 — enters when we pass to the smooth finite cover $\hat{X}$ of X.) Then we sketch a proof of Theorem 5.4 using a more topological and monodromy-theoretic approach.

Disproof of* (5.3) *for n odd. Most of the analysis that follows works for all n, though the last step is inconclusive for even n.

To begin, consider a pencil $\mathbb{P}^1 \xhookrightarrow{\varepsilon} \overline{\mathcal{S}}$ of hyperplane configurations given by fixing $L_0, \ldots, L_{2n}$ (in general position) and letting $L_{2n+1} := L_s$ vary along a line in $\check{\mathbb{P}}^n$ (chosen to avoid linear spans of any $n-2$ L_i in $\check{\mathbb{P}}^n$).[15] Writing $\Sigma = \varepsilon^{-1}(\overline{\mathcal{S}} \setminus \mathcal{S})$, we have $|\Sigma| = \binom{2n+1}{n}$; and degenerations $\mathcal{X}_\sigma \to \Delta_\sigma$ of our double-covers at $\sigma \in \Sigma$ are locally modeled (with $t = s - \sigma$) by

$$w^2 \underset{\text{loc}}{=} x_1 \cdots x_n (t - x_1 - \cdots - x_n) \tag{5.5}$$

after a $\mathrm{PGL}_{n+1}(\mathbb{C})$-action. Accordingly, writing $X_0, \ldots, X_n$ for projective coordinates on $\mathbb{P}^n$, we take $\ell_i = X_i$ for $0 \leq i \leq n$ and $\ell_{n+1} = tX_0 - \sum_{i=1}^n X_i$, and $\ell_{n+2}, \ldots, \ell_{2n+1}$ "general".

Let $\underline{\ell} : \mathbb{P}^n \hookrightarrow \mathbb{P}^{2n+1}$ denote the linear embedding

$$[X_0 : \cdots : X_n] \mapsto [\ell_0(\underline{X}) : \cdots : \ell_{2n+1}(\underline{X})]$$

and $\phi : \mathbb{P}^{2n+1} \to \mathbb{P}^{2n+1}$ denote the map sending

$$[Z_0 : \cdots : Z_{2n+1}] \mapsto [Z_0^2 : \cdots : Z_{2n+1}^2].$$

[15] It already follows from Zariski's theorem [Voisin 2003, Theorem 3.22] that $\rho(\pi_1(\mathbb{P}^1 \setminus \Sigma)) = \rho(\pi_1(\mathcal{S}))$ but we won't need this.

Then the variety $\hat{X} := \phi^{-1}(\underline{\ell}(\mathbb{P}^n)) \subset \mathbb{P}^{2n+1}$ is a smooth complete intersection on which[16] $\mathcal{A} := (\mathbb{Z}/2\mathbb{Z})^{2n+2}/\mathbb{\Delta}(\mathbb{Z}/2\mathbb{Z})$ acts via $\underline{e}^{(i)} \mapsto \{Z_i \mapsto -Z_i\}$, with quotient $\mathbb{P}^n$; explicitly, we have

$$\hat{X} = \bigcap_{k=0}^{n} \{0 = F_k(\underline{Z}) := -Z_{n+k+1}^2 + \ell_{n+k+1}(Z_0^2, \ldots, Z_n^2)\}. \tag{5.6}$$

Write $\chi \in \mathrm{X}^*(\mathcal{A})$ for the character sending each $\underline{e}^{(i)} \mapsto -1$, $\mathcal{A}^\circ := \ker(\chi) \leq \mathcal{A}$, and $q : \hat{X} \twoheadrightarrow X$ for the quotient by $\mathcal{A}^\circ$; then $H \cong q^* H^n_{\mathrm{pr}}(X) \cong H^n(\hat{X})^\chi$. Since

$$F_0(\underline{Z}) = tZ_0^2 - \sum_{i=1}^{n+1} Z_i^2,$$

we have thus replaced our original *non*-isolated degeneration (5.5) by a nodal one.

Next, we use the "Cayley trick" to replace the complete intersection $\hat{X}$ by a hypersurface

$$Y := \left\{0 = F := \sum_{k=0}^{n} Y_k F_k(\underline{Z})\right\} \subset \mathbb{P}(\mathcal{O}_{\mathbb{P}^{2n+1}}(2)^{\oplus n+1}) =: \boldsymbol{P} \tag{5.7}$$

of dimension $3n$. We have an $\mathcal{A}$-equivariant isomorphism $H^{3n}(Y)(n) \cong H^n(\hat{X})$ of HSs, so that $H \cong H^{3n}(Y)^\chi(n)$. In affine coordinates $(z_1, \ldots, z_{2n+1}; y_1, \ldots, y_n)$, notice that $F = 0$ becomes[17]

$$0 = t - z_1^2 - \cdots - z_{n+1}^2 + \sum_{k=1}^{n} y_k(b_k - z_{n+k+1})(b_k + z_{n+k+1}) + \text{h.o.t.}, \tag{5.8}$$

where $b_k := \sqrt{F_k(1, 0, \ldots, 0)}$. So at $t = 0$, the singular fiber Y_σ has 2^n nodes at

$$\begin{aligned}(Z_0;\, &Z_1, \ldots, Z_{n+1}; Z_{n+2}, \ldots, Z_{2n+1}; Y_0; Y_1, \ldots, Y_n) \\ &= \big(1; 0, \ldots, 0; (-1)^{a_1} b_1, \ldots, (-1)^{a_n} b_n; 1; 0, \ldots, 0\big), \quad \underline{a} \in (\mathbb{Z}/2\mathbb{Z})^n,\end{aligned} \tag{5.9}$$

and the degeneration $\mathcal{Y}_\sigma \to \Delta_\sigma$ has smooth total space. The mixed spectrum of each node is $[((3n+1)/2, 3n+1)]$ for n odd and $[((3n+1)/2, 3n)]$ for n even; so T_σ acts through multiplication by $(-1)^{n+1}$ on

$$H^{3n}_{\mathrm{van}}(Y_t) \cong \mathbb{Q}\Big(-\Big\lfloor \frac{3n+1}{2} \Big\rfloor\Big)^{\oplus 2^n}. \tag{5.10}$$

Moreover, since the summands of (5.10) are represented by

$$\eta_{\underline{a}} = (-1)^{|\underline{a}|}(dz_1 \wedge \cdots \wedge dz_{2n+1} \wedge dy_1 \wedge \cdots \wedge dy_n)/F^{\lceil (3n+1)/2 \rceil}$$

near the nodes (5.9) (in the sense of [Kerr and Laza 2023, Section 2]), it has a 1-dimensional subspace (generated by $\eta_\chi := \sum (-1)^{|\underline{a}|}\eta_{\underline{a}}$) on which $\mathcal{A}$ acts through χ.

[16] Here $\mathbb{\Delta}$ denotes the diagonal embedding.

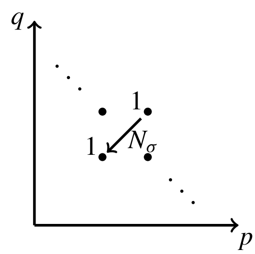
[17] Here "h.o.t." means terms vanishing to order 3 at the nodes.

Taking χ-eigenspaces of the vanishing-cycle sequence for $\mathcal{Y}_\sigma \to \Delta_\sigma$ and twisting by $\mathbb{Q}(n)$ now yields

$$(5.11)\quad 0 \to H^{3n}(Y_\sigma)^\chi(n) \xrightarrow{\mathrm{sp}^\chi} \underbrace{H^{3n}_{\lim}(Y_t)^\chi(n)}_{\cong H_{\lim}} \xrightarrow{\mathrm{can}^\chi} \mathbb{Q}\Big(-\Big\lfloor \frac{n+1}{2} \Big\rfloor\Big) \xrightarrow{\delta^\chi} H^{3n+1}_{\mathrm{ph}}(Y_\sigma)^\chi(n) \to 0.$$

We claim that $\delta = 0$. For n even, this is clear, since T_σ acts trivially on $H^{3n+1}_{\mathrm{ph}}(Y_\sigma)$ and by -1 on $\mathbb{Q}(-\lfloor (n+1)/2 \rfloor)$. So we conclude that T_σ acts on $H_{\lim}$ via an orthogonal reflection. This doesn't factor through $\bigwedge^n$ of any automorphism of $\mathbb{C}^{2n}$, but because it is finite (of order 2), this does not (yet) disprove the conjecture.

On the other hand, for n odd, it is not automatic that $\delta = 0$. (This is a well-known problem with nodal degenerations in odd dimensions, see [Kerr and Laza 2023, Section 2.2]; and as we saw in the proof of (5.5), our degenerations are finite quotients of nodal ones.) But if we can show $\delta = 0$, then the conjecture is *immediately* disproved (for odd $n \geq 3$). Here is why: by (5.6), $H_{\lim}$ then has a class of type $(n+1, n+1)$, which must go to an (n, n) class by N_σ,



forcing $\mathrm{rk}(N_\sigma) = 1$ (rather than 0). (In different terms, each T_σ is a nontrivial symplectic transvection.) But this is impossible for $\bigwedge^n$ of a nilpotent endomorphism of $\mathbb{C}^{2n}$.

To complete the (dis)proof, then, we apply [Kerr and Laza 2023, Theorem 2.9]: for a nodal degeneration $Y \rightsquigarrow Y_\sigma$ of an odd-dimensional hypersurface of a smooth projective variety $\boldsymbol{P}$ satisfying Bott vanishing, the rank of δ is the number m of nodes minus the rank of the map

$$\mathrm{ev} : H^0\Big(\boldsymbol{P}, K_{\boldsymbol{P}}\Big(\frac{3n+1}{2} Y_\sigma\Big)\Big) \to \mathbb{C}^m$$

given by evaluation at the nodes. The proof in [loc. cit.] is equivariant in $\mathcal{A}$, and so we find that $\delta^\chi = 0 \iff$ ev is nonzero on $H^0\big(\boldsymbol{P}, K_{\boldsymbol{P}}\big(\frac{3n+1}{2} Y_\sigma\big)\big)^\chi$, which can be checked at any node. Writing

$$\boldsymbol{e}_1 := \sum_{i=0}^{n} Y_i \frac{\partial}{\partial Y_i}, \quad \boldsymbol{e}_2 := \sum_{j=0}^{2n+1} Z_j \frac{\partial}{\partial Z_j} - 2\boldsymbol{e}_1, \quad \text{and} \quad \Omega := \langle \boldsymbol{e}_2, \langle \boldsymbol{e}_1, d\underline{Z} \wedge d\underline{Y} \rangle\rangle,$$

one checks that

$$Y_0 Z_0^2 \Omega / (F_{t=0})^{(3n+1)/2} \tag{5.12}$$

is a well-defined section of $K_{\mathbf{P}}\big(\frac{3n+1}{2} Y_\sigma\big)$ (see [Kerr 2003, Section 4.5]); and evidently $\mathcal{A}$ acts on it through χ. Clearly, it is nonzero on the fiber of $K_{\mathbf{P}}\big(\frac{3n+1}{2} Y_\sigma\big)$ at any of the nodes (5.9).

Sketch of proof of Theorem 5.4. Returning to the local picture (5.5), we now seek a more concrete topological description of the orthogonal reflections (n even) and symplectic transvections (n odd) through which T_σ acts on H. So let $U_0 \subset \mathbb{A}^n$ be the complement of the hyperplanes $x_1 = 0, \dots, x_n = 0$ and $x_1 + \cdots + x_n = 1$, and $\mathbb{L}_0$ the rank-1 local system on U_0 with monodromies -1 about each of them. While the singularity $x_\sigma \overset{\imath_\sigma}{\hookrightarrow} X_\sigma$ "at $\underline{0}$" in (5.5) isn't isolated, the vanishing-cycle complex $\phi_t\, \mathbb{Q}_{\mathcal{X}}$ is nothing but $\imath_*^\sigma V[-n]$, where $V := \mathrm{IH}^n(\mathbb{A}^n, \mathbb{L}_0)$ (as MHS). We begin with a local analogue of the covering argument just seen.

5.13. Lemma. (i) $\mathrm{IH}^n(\mathbb{A}^n, \mathbb{L}_0) \cong \mathbb{Q}(-\lfloor (n+1)/2 \rfloor)$.

(ii) *Local monodromy T_σ acts on V through multiplication by* $(-1)^{n+1}$.

(iii) *The canonical map* $\mathrm{can}_\sigma : H_{\lim} \to V$ *is onto.*

Proof. Define maps

- $f_0 : \mathbb{A}^n \hookrightarrow \mathbb{A}^{n+1}$ by $\underline{x} \mapsto \big(\underline{x}, 1 - \sum_{i=1}^n x_i\big)$ and
- $\phi_0 : \mathbb{A}^{n+1} \to \mathbb{A}^{n+1}$ by squaring all coordinates z_i.

Then $\hat{X}_0 := \phi_0^{-1}(f_0(\mathbb{A}^n)) \subset \mathbb{A}^{n+1}$ is the quadric hypersurface $\sum_{i=1}^{n+1} z_i^2 = 1$. The group $\mathcal{A}_0 := (\mathbb{Z}/2\mathbb{Z})^{n+1}$ acts on $\hat{X}_0$ (multiplying coordinates by ± 1), with quotient $\mathbb{A}^n$. The quotient $q_0 : \hat{X}_0 \twoheadrightarrow X_0$ by the augmentation subgroup $\mathcal{A}_0^\circ$ yields the obvious $2:1$ branched cover of $\mathbb{A}^n$, with $H^n(X_0) \cong \mathrm{IH}^n(\mathbb{A}^n, \mathbb{L}_0)$.

By the localization sequence for $\hat{X}_0$ (relative to its closure $\overline{\hat{X}}_0 \subset \mathbb{P}^{n+1}$) and weak Lefschetz, one easily shows that $H^j(\hat{X}_0) = 0$ for $j \neq n$,[18] and

$$H^n(\hat{X}_0) \cong \mathbb{Q}\Big(-\Big\lfloor \frac{n+1}{2} \Big\rfloor\Big).$$

(Writing $\partial \hat{X}_0 = \overline{\hat{X}}_0 \setminus \hat{X}_0$, this is $H^n(\overline{\hat{X}}_0)/H^{n-2}(\partial \hat{X}_0)(-1)$ for n even, and for n odd $\ker\{H^{n-1}(\partial \hat{X}_0)(-1) \to H^{n+1}(\overline{\hat{X}}_0)\}$.) A generator for the *dual* group $H_c^n(\hat{X}_0)$ is given by the real (vanishing) n-sphere $S_1^n := \big\{\sum z_i^2 = 1\big\} \cap \mathbb{R}^{n+1}$, whose class is invariant under $\mathcal{A}_0^\circ$ hence comes from $H_c^n(X_0)$. This gives (i).

The degeneration is modeled by replacing $\sum z_i^2 = 1$ by $\sum z_i^2 = t$; as the spectrum of $\sum z_i^2$ is $[(n+1)/2]$, the monodromy is as described in (ii). Finally, (iii) follows from the last subsection since can_σ identifies with can^χ in (5.11). □

[18] This simply recovers perversity of $\phi_f\, \mathbb{Q}_{\mathcal{X}}[n]$.

The vanishing sphere $S_t^n := \left\{\sum z_i^2 = t\right\} \cap \mathbb{R}^{n+1}$ in $\hat{X}_0$ has image in X_0 (by q_0) given by the double cover of $\left(\bigcap_{i=1}^n \{x_i \geq 0\}\right) \cap \left\{\sum x_i \leq t\right\}$. Let its image in X (essentially via $\mathrm{can}^{\mathcal{X}} : H_c^n(X_0) \to H^n(X)$) be denoted by ν_σ; this is the vanishing cycle at σ, a "double simplex" branched along $\mathcal{H}_s$ and n additional hyperplanes. It follows from (iii) that T_σ is a transvection/reflection in ν_σ. More precisely, rescaling Q to have $Q(\nu_\sigma, \nu_\sigma) = \frac{1}{2}(1 + (-1)^n)$,

$$T_\sigma(u) = u - 2Q(u, \nu_\sigma)\, \nu_\sigma \tag{5.14}$$

for $u \in H$.

Now consider the general setting where $L_{2n+1} = L_s$, $L_0 = \{X_0 = 0\}$, and the remaining L_i are in general position. An easy extension of (5.1) gives

$$H \cong \mathrm{IH}_c^n(\mathbb{A}^n, \mathbb{L}) \cong H_c^n(X \setminus L_0),$$

whence $H_{\mathrm{pr}}^n(X)$ is spanned by double simplices branched along $n+1$ of the $L_{i \geq 0}$. Obviously all of these can be rewritten as $\mathbb{Z}$-linear combinations of double simplices branched along L_s and n of the $\{L_i\}_{1 \leq i \leq 2n}$; call these ν_I, where $I \subset \{1, \dots, 2n\}$ with $|I| = n$. Since $\mathrm{rk}\, H = \binom{2n}{n}$ and there are $\binom{2n}{n}$ of these vanishing cycles, they form a $\mathbb{Q}$-basis of $H = H_{\mathrm{pr}}^n(X)$. Write T_I for the corresponding monodromies, and $\Gamma \leq \mathrm{Aut}(H_{\mathbb{C}}, Q)$ for the smallest $\mathbb{C}$-algebraic group containing them; clearly $\Gamma \leq \widetilde{\Pi}_{\mathbb{C}}$. Moreover, we note that if $|I \cap I'| = n-1$, then $Q(\nu_I, \nu_{I'}) = \pm 1$ (rescaling as above, compatibly with (5.14)).

Suppose then that $|I \cap I'| = n-1$. If n is odd, then $T_I(\nu_{I'}) = \nu_I \pm \nu_{I'} = \pm T_{I'}^{-1}(\nu_I)$, whence $\nu_{I'}$ is in the Γ-orbit of ν_I; so *all* the ν_J are in the Γ-orbit of ν_I. If n is even, then reasoning as in [Deligne 1980, Section 4.4] (see the paragraph after Lemme 4.4.3[s]), $T_I T_{I'}^{\pm 1}$ is a transvection and its Zariski closure a $\mathbb{G}_a$ including transformations which send $\nu_I \mapsto \nu_{I'}$ and vice versa; once again, all the ν_J are in the Γ-orbit of a single ν_I.

Let $R := \Gamma.\nu_I$ denote this orbit. Obviously it spans $H_{\mathbb{C}}$. Furthermore, for any $\delta \in R$, we have that Γ contains the transvection/reflection T_δ: writing $\delta = \gamma.\nu_I$ ($\gamma \in \Gamma$), we have $T_\delta = T_{\gamma.\nu_I} = \gamma T_I \gamma^{-1} \in \Gamma$. So Γ is in fact the $\mathbb{C}$-algebraic closure of the $\{T_\delta\}_{\delta \in R}$, and we are exactly in the situation of [Deligne 1980, Lemme 4.4.2]. Conclude that $\Gamma = \mathrm{Aut}(H_{\mathbb{C}}, Q)$, and hence $\widetilde{\Pi} = \mathrm{Aut}(H, Q)$, and thus $\Pi = \mathrm{Aut}(H, Q)^\circ$, proving Theorem 5.4.

5.15. Remark. After writing this paper we encountered the article [Xu 2018] which treats the more general setting of r-covers of $\mathbb{P}^n$ branched along hyperplanes by considering local monodromies (as we have just done). The argument is necessarily more complicated and technical than ours. However, in the case $r = 2$ (i.e., our setting) it appears to be incomplete.

If $r = 2$ and n is odd, Proposition 3.4 of [Xu 2018] does not actually establish that, in the notation of [loc. cit.], $e_{(1)}$ is nonzero; this is exactly the issue regarding possible nonvanishing of δ dealt with above. One could read [Xu 2018, Proposition 4.2] as confirming this in retrospect, but this makes the argument quite convoluted.

If $r = 2$ and n is even, the proof of [Xu 2018, Proposition 4.2] is wrong, as it makes use of the (false) statement that $\mathrm{Sp}_{2n}(\mathbb{R})$ "does not admit any nontrivial one-dimensional invariant subspace" in its action on $\bigwedge^n \mathbb{R}^{2n}$.

Acknowledgements

We thank P. Gallardo and R. Laza for valuable discussions related to this paper, and the referee for helpful expository suggestions. This work was partially supported by Simons Collaboration grant 634268 and NSF grant DMS-2101482 (Kerr). Pearlstein's work was partially supported by the following grants: MIUR Excellence Department Project awarded to the Department of Mathematics, University of Pisa, CUP I57G22000700001 and Progetti di Rilevante Interesse Nazionale (PRIN), Geometry of Algebraic Structures: Moduli, Invariants, Deformations. Kerr and Pearlstein would also like to thank the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the program "K-theory, algebraic cycles, and motivic homotopy theory" (supported by EPSRC grant number EP/R014604/1) when work on this paper was undertaken.

References

[Allcock et al. 2002] D. Allcock, J. A. Carlson, and D. Toledo, "The complex hyperbolic geometry of the moduli space of cubic surfaces", *J. Algebraic Geom.* **11**:4 (2002), 659–724. MR Zbl

[Allcock et al. 2011] D. Allcock, J. A. Carlson, and D. Toledo, *The moduli space of cubic threefolds as a ball quotient*, Mem. Amer. Math. Soc. **209**, 2011. MR Zbl

[Arnol'd 1981] V. Arnol'd, "On some problems in singularity theory", *Proc. Indian Acad. Sci. Math. Sci.* **90**:1 (1981), 1–9. MR Zbl

[Beauville 1980] A. Beauville, "Sur le nombre maximum de points doubles d'une surface dans $\boldsymbol{P}^3$ ($\mu(5) = 31$)", pp. 207–215 in *Journées de Géometrie Algébrique d'Angers* (Angers, 1979), Sijthoff and Noordhoff, 1980. MR Zbl

[Cartan 1954] H. Cartan, "Quotient d'une variété analytique par un groupe discret d'automorphismes", *Séminaire Henri Cartan* **6**:12 (1954), 1–13. Zbl

[Casalaina-Martin et al. 2012] S. Casalaina-Martin, D. Jensen, and R. Laza, "The geometry of the ball quotient model of the moduli space of genus four curves", pp. 107–136 in *Compact moduli spaces and vector bundles*, Contemp. Math. **564**, Amer. Math. Soc., Providence, RI, 2012. MR Zbl

[Deligne 1971] P. Deligne, "Théorie de Hodge, II", *Inst. Hautes Études Sci. Publ. Math.* **40** (1971), 5–57. MR Zbl

[Deligne 1980] P. Deligne, "La conjecture de Weil, II", *Inst. Hautes Études Sci. Publ. Math.* **52** (1980), 137–252. MR Zbl

[Deligne and Mostow 1986] P. Deligne and G. D. Mostow, "Monodromy of hypergeometric functions and nonlattice integral monodromy", *Inst. Hautes Études Sci. Publ. Math.* **63** (1986), 5–89. MR Zbl

[Deng and Gallardo 2023] H. Deng and P. Gallardo, "Eigenperiods and the moduli of points in the line", preprint, 2023. Zbl arXiv 2311.09352

[Dolgachev 1982] I. Dolgachev, "Weighted projective varieties", pp. 34–71 in *Group actions and vector fields* (Vancouver, B.C., 1981), Lecture Notes in Math. **956**, Springer, Berlin, 1982. MR Zbl

[Dolgachev and Kondō 2007] I. V. Dolgachev and S. Kondō, "Moduli of $K3$ surfaces and complex ball quotients", pp. 43–100 in *Arithmetic and geometry around hypergeometric functions*, Progr. Math. **260**, Birkhäuser, Basel, 2007. MR Zbl

[Endrass 1997] S. Endrass, "A projective surface of degree eight with 168 nodes", *J. Algebraic Geom.* **6**:2 (1997), 325–334. MR Zbl

[Gallardo et al. 2021] P. Gallardo, M. Kerr, and L. Schaffler, "Geometric interpretation of toroidal compactifications of moduli of points in the line and cubic surfaces", *Adv. Math.* **381** (2021), art. id. 107632. MR Zbl

[Gerkmann et al. 2007a] R. Gerkmann, S. Mao, and K. Zuo, "Disproof of modularity of moduli space of CY 3-folds of double covers of P3 ramified along eight planes in general positions", preprint, 2007. Zbl arXiv 0709.1051

[Gerkmann et al. 2007b] R. Gerkmann, M. Sheng, and K. Zuo, "Computational details on the disproof of modularity", preprint, 2007. arXiv 0709.1054

[Gerkmann et al. 2013] R. Gerkmann, M. Sheng, D. van Straten, and K. Zuo, "On the monodromy of the moduli space of Calabi–Yau threefolds coming from eight planes in $\mathbb{P}^3$", *Math. Ann.* **355**:1 (2013), 187–214. MR Zbl

[Hassett 2003] B. Hassett, "Moduli spaces of weighted pointed stable curves", *Adv. Math.* **173**:2 (2003), 316–352. MR Zbl

[Hotta et al. 2008] R. Hotta, K. Takeuchi, and T. Tanisaki, *D-modules, perverse sheaves, and representation theory*, Progress in Mathematics **236**, Birkhäuser, Boston, MA, 2008. MR Zbl

[Jaffe and Ruberman 1997] D. B. Jaffe and D. Ruberman, "A sextic surface cannot have 66 nodes", *J. Algebraic Geom.* **6**:1 (1997), 151–168. MR Zbl

[Kerr 2003] M. D. Kerr, *Geometric construction of regulator currents with applications to algebraic cycles*, Ph.D. thesis, Princeton University, 2003, available at https://www.proquest.com/docview/305317544. MR Zbl

[Kerr and Laza 2023] M. Kerr and R. Laza, "Hodge theory of degenerations, (II): vanishing cohomology and geometric applications", preprint, 2023. Zbl arXiv 2006.03953v2

[Kerr et al. 2021] M. Kerr, R. Laza, and M. Saito, "Hodge theory of degenerations, (I): consequences of the decomposition theorem", *Selecta Math.* (*N.S.*) **27**:4 (2021), art. id. 71. MR Zbl

[Miyaoka 1984] Y. Miyaoka, "The maximal number of quotient singularities on surfaces with given numerical invariants", *Math. Ann.* **268**:2 (1984), 159–171. MR Zbl

[Moonen 2018] B. Moonen, "The Deligne–Mostow list and special families of surfaces", *Int. Math. Res. Not.* **2018**:18 (2018), 5823–5855. MR Zbl

[Mostow 1988] G. D. Mostow, "On discontinuous action of monodromy groups on the complex n-ball", *J. Amer. Math. Soc.* **1**:3 (1988), 555–586. MR Zbl

[Sakamaki 2010] Y. Sakamaki, "Automorphism groups on normal singular cubic surfaces with no parameters", *Trans. Amer. Math. Soc.* **362**:5 (2010), 2641–2666. MR Zbl

[Schoen 1985] C. Schoen, "Algebraic cycles on certain desingularized nodal hypersurfaces", *Math. Ann.* **270**:1 (1985), 17–27. MR Zbl

[Sheng et al. 2015] M. Sheng, J. Xu, and K. Zuo, "The monodromy groups of Dolgachev's CY moduli spaces are Zariski dense", *Adv. Math.* **272** (2015), 699–742. MR Zbl

[Steenbrink 1977] J. Steenbrink, "Intersection form for quasi-homogeneous singularities", *Compositio Math.* **34**:2 (1977), 211–223. MR Zbl [van Straten 1993] D. van Straten, "A quintic hypersurface in $\boldsymbol{P}^4$ with 130 nodes", *Topology* **32**:4 (1993), 857–864. MR Zbl

[van Straten 2020] D. van Straten, "The Spectrum of Hypersurface Singularities", preprint, 2020. Zbl arXiv 2003.00519

[Varchenko 1983] A. N. Varchenko, "Semicontinuity of the spectrum and an upper bound for the number of singular points of the projective hypersurface", *Dokl. Akad. Nauk SSSR* **270**:6 (1983), 1294–1297. MR Zbl

[Voisin 2003] C. Voisin, *Hodge theory and complex algebraic geometry, II*, Cambridge Studies in Advanced Mathematics **77**, Cambridge University Press, Cambridge, 2003. MR Zbl

[Xu 2018] J. Xu, "Zariski density of monodromy groups via a Picard–Lefschetz type formula", *Int. Math. Res. Not.* **2018**:11 (2018), 3556–3586. MR Zbl

Received November 8, 2022. Revised December 21, 2023.

BEN CASTOR
DEPARTMENT OF MATHEMATICS
KENYON COLLEGE
GAMBIER, OH
UNITED STATES
castor1@kenyon.edu

HAOHUA DENG
MATHEMATICS DEPARTMENT
DUKE UNIVERSITY
DURHAM, NC
UNITED STATES
haohua.deng@duke.edu

MATT KERR
DEPARTMENT OF MATHEMATICS AND STATISTICS
WASHINGTON UNIVERSITY IN ST. LOUIS
ST. LOUIS, MO
UNITED STATES
matkerr@wustl.edu

GREGORY PEARLSTEIN
DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DI PISA
PISA
ITALY
greg.pearlstein@unipi.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 327, No. 1, 2023
https://doi.org/10.2140/pjm.2023.327.55

FOURIER BASES OF A CLASS OF PLANAR SELF-AFFINE MEASURES

MING-LIANG CHEN, JING-CHENG LIU AND ZHI-YONG WANG

Let $\mu_{M,D}$ be the planar self-affine measure generated by an expansive integer matrix $M \in M_2(\mathbb{Z})$ and a noncollinear integer digit set

$$D = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \begin{pmatrix} -\alpha_1-\beta_1 \\ -\alpha_2-\beta_2 \end{pmatrix} \right\}.$$

We show that $\mu_{M,D}$ is a spectral measure if and only if there exists a matrix $Q \in M_2(\mathbb{R})$ such that $(\tilde{M}, \tilde{D})$ is admissible, where $\tilde{M} = QMQ^{-1}$ and $\tilde{D} = QD$. In particular, when $\alpha_1\beta_2 - \alpha_2\beta_1 \notin 2\mathbb{Z}$, $\mu_{M,D}$ is a spectral measure if and only if $M \in M_2(2\mathbb{Z})$. This completely settles the spectrality of the self-affine measure $\mu_{M,D}$.

1. Introduction

Let μ be a Borel probability measure with compact support on $\mathbb{R}^n$, and let $\langle \cdot, \cdot \rangle$ denote the standard inner product on $\mathbb{R}^n$. We say that μ is a *spectral measure* if there exists a countable set $\Lambda \subset \mathbb{R}^n$ such that the exponential function system $E_\Lambda := \{e^{2\pi i \langle \lambda, x \rangle} : \lambda \in \Lambda\}$ forms an orthonormal basis for the Hilbert space $L^2(\mu)$. In this case, we call Λ a *spectrum* of μ and (μ, Λ) a *spectral pair*. In particular, if μ is the normalized Lebesgue measure supported on a Borel set Ω, then Ω is called a *spectral set*.

Spectral measure is a natural generalization of spectral set introduced by Fuglede [20], who proposed the famous conjecture that Ω is a spectral set if and only if Ω is a translational tile. It is known [22] that a spectral measure μ must be of pure type: μ is either discrete, or absolutely continuous or singularly continuous. The first singularly continuous spectral measure was constructed by Jorgensen and

The research is supported in part by the NNSF of China (Nos. 12071125, 12001183 and 12301104), the Hunan Provincial NSF (No. 2022JJ30369), the Jiangxi Provincial NSF (No. 20232BAB211003), the Doctoral Scientific Research Foundation of Gannan Normal University (No. BSJJ202241) and the Education Department Important Foundation of Hunan Province in China (No. 23A0095).
Liu is the corresponding author.
MSC2020: primary 28A25, 28A80; secondary 42C05, 46C05.
Keywords: self-affine measure, spectral measure, spectrum, admissible.

Pedersen in 1998 [24]. They proved that the middle-fourth Cantor measure is a spectral measure with a spectrum

$$\Lambda=\left\{\sum_{k=0}^{n}4^k\ell_k : \ell_k\in\{0,1\},\ n\in\mathbb{N}\right\}.$$

Following this discovery, there is a considerable number of papers on the spectrality of self-affine measures and the construction of their spectra; see [2; 3; 5; 6; 7; 8; 12; 13; 16; 18; 29]. These results are generalized further to some classes of Moran measures (see, e.g., [1; 9; 19]), and some surprising convergence properties of the associated Fourier series were discovered in [38; 39]. These fractal measures also have very close connections with the theory of multiresolution analysis in wavelet analysis; see [11].

In [14], Dutkay and Jorgensen summarized some known results regarding iterated function systems (IFS); see [23] for details. Two approaches to harmonic analysis on IFS have been popular: one based on a discrete version of the more familiar and classical second-order Laplace differential operator of potential theory; see [27; 28; 30]; and the other is based on Fourier series. The first model in turn is motivated by infinite discrete network of resistors, and the harmonic functions are defined by minimizing a global measure of resistance, but this approach does not rely on Fourier series. In contrast, the second approach begins with Fourier series, and it has its classical origins in lacunary Fourier series [26].

For an expansive real matrix $M\in M_n(\mathbb{R})$ and a finite digit set $D\subset\mathbb{R}^n$ with cardinality $\#D$, the *iterated function system* (IFS) $\{\phi_d(x)\}_{d\in D}$ is defined by $\phi_d(x)=M^{-1}(x+d)$ $(x\in\mathbb{R}^n,\ d\in D)$. By [23], there exists a unique probability measure $\mu_{M,D}$ satisfying

$$\mu_{M,D}=\frac{1}{\#D}\sum_{d\in D}\mu_{M,D}\circ\phi_d^{-1}. \tag{1-1}$$

It is supported on the unique nonempty compact set $T(M,D)=\bigcup_{d\in D}\phi_d(T(M,D))$. Hence

$$T(M,D)=\left\{\sum_{k=1}^{\infty}M^{-k}d_k : d_k\in D\right\}:=\sum_{k=1}^{\infty}M^{-k}D.$$

The measure $\mu_{M,D}$ and the set $T(M,D)$ are called *self-affine measure* and *self-affine set*, respectively. It is known that a self-affine measure $\mu_{M,D}$ can be expressed by the infinite convolution of discrete measures as

$$\mu_{M,D}=\delta_{M^{-1}D}*\delta_{M^{-2}D}*\delta_{M^{-3}D}*\cdots,$$

where $*$ is the convolution sign, $\delta_E=\frac{1}{\#E}\sum_{e\in E}\delta_e$ for a finite set E and δ_e is the Dirac measure at the point e.

Self-affine measures have the advantage that their Fourier transforms (see (2-1)) can be explicitly written down as an infinite product, which allows us to compute their zeros. The previous research on self-affine measures $\mu_{M,D}$ and their Fourier transform have revealed some surprising connections with a number of areas in mathematics such as harmonic analysis, dynamical systems, number theory and others (see, e.g., [21; 25; 37]).

In the previous works, the spectral self-affine measures are usually generated by compatible pairs (known also as Hadamard triples). The appearance of compatible pairs stems from the terminology of [38].

Definition 1.1. Let $M \in M_n(\mathbb{Z})$ be an expansive integer matrix, and let $D, S \subset \mathbb{Z}^n$ be two finite digit sets with $\#D = \#S = N$. We say that (M, D) is *admissible* (or $(M^{-1}D, S)$ forms a *compatible pair* or (M, D, S) forms a *Hadamard triple*) if the matrix

$$H = \frac{1}{\sqrt{N}} \left(e^{2\pi i \langle M^{-1} d, s\rangle} \right)_{d \in D, s \in S}$$

is unitary, i.e., $H^* H = I$, where I is a $n \times n$ identity matrix.

The well-known result of Jorgensen and Pedersen [24] shows that if (M, D) is admissible, then there are infinite families of orthogonal exponential functions in $L^2(\mu_{M,D})$. Dutkay and Jorgensen [13; 15] formulated the famous conjecture that if (M, D) is admissible, then $\mu_{M,D}$ is a spectral measure. It was first proved in one dimension by Łaba and Wang [29]. The conjecture is true in higher dimensions under some additional assumptions, introduced by Strichartz [38]. There are many other papers that investigated it in higher dimensional cases; see [12; 32]. In the end, Dutkay, Haussermann and Lai [16] proved that:

Theorem 1.2. *Let $M \in M_n(\mathbb{Z})$ be an expansive integer matrix, and let $D \subset \mathbb{Z}^n$ be a finite digit set. If (M, D) is admissible, then $\mu_{M,D}$ is a spectral measure.*

In [18], Fu, He and Lau gave an example to illustrate that the sufficient condition in Theorem 1.2 is not necessary in one dimension. For an expansive integer matrix $M \in M_2(\mathbb{Z})$ and the classic digit set $D = \left\{ \binom{0}{0}, \binom{1}{0}, \binom{0}{1} \right\}$, the spectrality and nonspectrality of the corresponding self-affine measure $\mu_{M,D}$ has been widely investigated by many researchers; see [12; 31; 32]. Eventually, An, He and Tao [2] completely settled the spectrality of $\mu_{M,D}$. More precisely, they showed that $\mu_{M,D}$ is a spectral measure if and only if (M, D) is admissible. For a more general integer digit set D with $0 \in D$ and $\#D = 3$, there is also a complete spectral characterization; see [4; 35; 36]. In addition to these, another important integer digit set is

$$D = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \begin{pmatrix} -\alpha_1 - \beta_1 \\ -\alpha_2 - \beta_2 \end{pmatrix} \right\}, \tag{1-2}$$

where $\alpha_1\beta_2 - \alpha_2\beta_1 \neq 0$. The existence of infinitely many orthogonal exponentials in $L^2(\mu_{M,D})$ has been fully studied in [33; 40; 41]. Recently, Fu and Tang [17] considered the special case where $\alpha_1 = 1$, $\alpha_2 = 0$, $\beta_1 = 0$ and $\beta_2 = 1$. They fully characterized the spectrality of the corresponding self-affine measures. However, to the best of our knowledge, the complete description of spectral properties of the general case (1-2) is not known yet. A natural subsequent question is:

Question 1. For an expansive integer matrix $M \in M_2(\mathbb{Z})$ and the digit set D given by (1-2), what is the sufficient and necessary condition for $\mu_{M,D}$ to be a spectral measure?

In the study of the spectrality of self-affine measures $\mu_{M,D}$ on $\mathbb{R}^n$, the finiteness and rationality of the set $\mathcal{Z}_D^n := \left\{x \in [0,1)^n : \sum_{d\in D} e^{2\pi i\langle d,x\rangle} = 0\right\}$ are pivotal. Many classic digit sets, such as $\{0, 1, \dots, N-1\}$, $\{(0,0)^t, (1,0)^t, (0,1)^t\}$ and the digit set D given by (1-2), exhibit the desired property. This has attracted a large number of researchers to study their spectrality of the corresponding self-affine measures. However, if $\mathcal{Z}_D^n$ is infinite or irrational, resolving the spectrality of the corresponding self-affine measure becomes a formidable challenge. For instance, consider $M \in M_2(\mathbb{Z})$ and $D = \{(0,0)^t, (1,0)^t, (0,1)^t, (1,1)^t\}$. It is easy to get that

$$\mathcal{Z}_D^2 = \left\{\begin{pmatrix}\frac{1}{2}\\ a\end{pmatrix} \cup \begin{pmatrix}a\\ \frac{1}{2}\end{pmatrix} : a \in [0,1)\right\}.$$

This means that $\mathcal{Z}_D^2$ encompasses a submanifold characterized by the free variable $a \in [0,1)$. For the more general digit set $D = \{0, u, v, u+v\} \subset \mathbb{Z}^2$, the set $\mathcal{Z}_D^2$ is infinite and includes free variables. The spectral properties of these self-affine measures have not been resolved.

The cardinality $\#D$ of a digit set D significantly influences the properties of $\mathcal{Z}_D^n$. In [3], An, He and Lai extensively classified four-element digit spectral self-similar measures on $\mathbb{R}$. They showed that if $\#D = 4$ and the corresponding self-similar measure is a spectral measure, then D is rational and $\mathcal{Z}_D^1$ is finite and rational. However, if D does not have any special structures and $\#D \geq 5$, the set $\mathcal{Z}_D^n$ is hard to calculate and may be irrational. For example, let $D = \{0, 1, 3, 5, 6\}$. Then $\mathcal{Z}_D^1 \subset \mathbb{R}\setminus\mathbb{Q}$ by [3, Example 5.2]. This makes it very difficult to study the spectrality of the corresponding self-similar measure.

Inspired by the above researches and due to the finiteness and rationality of the set $\mathcal{Z}_D^2$ corresponding to the digit set D given by (1-2), we can give an answer to Question 1. Before presenting our results, a reasonable assumption for the digit set D is necessary. Without loss of generality, we can assume that $\gcd(\alpha_1, \alpha_2, \beta_1, \beta_2) = 1$ by Lemma 2.2.

Our first main result is as follows:

Theorem 1.3. *Let $\mu_{M,D}$ be defined by* (1-1), *where $M \in M_2(\mathbb{Z})$ is an expansive integer matrix and D is given by* (1-2). *Then $\mu_{M,D}$ is a spectral measure if and only if there exists a matrix $Q \in M_2(\mathbb{R})$ such that $(\tilde{M}, \tilde{D})$ is admissible*, *where $\tilde{M} = QMQ^{-1}$ and $\tilde{D} = QD$.*

We remark that Theorem 1.3 gives a complete answer to the spectral Question 1. We now outline the strategy of the proof of Theorem 1.3. The sufficiency of Theorem 1.3 follows directly from Theorem 1.2 and Lemma 2.2. The more challenging part of the proof is the necessity. The key point is to construct a self-affine measure $\mu_{\tilde{M},\tilde{D}}$ so that it has the same spectrality as the measure $\mu_{M,D}$, and then the necessity follows immediately from Theorems 1.5 and 1.6. What is exciting is that the proof method of the necessity is new and completely different from the previous work proving spectral self-affine measures.

It is worth noting that if D satisfies $\alpha_1\beta_2 - \alpha_2\beta_1 \notin 2\mathbb{Z}$, we can give more explicit sufficient and necessary conditions for $\mu_{M,D}$ to be a spectral measure. Before presenting them, some notation is needed. For any integer $p \geq 2$, we define

$$\mathcal{F}_p^2 := \frac{1}{p}\left\{\begin{pmatrix} l_1 \\ l_2 \end{pmatrix} : 0 \leq l_1, l_2 \leq p-1,\ l_i \in \mathbb{Z}\right\} \quad \text{and} \quad \mathring{\mathcal{F}}_p^2 := \mathcal{F}_p^2 \setminus \{\mathbf{0}\}. \tag{1-3}$$

Under the above notation and the assumption of $\alpha_1\beta_2 - \alpha_2\beta_1 \notin 2\mathbb{Z}$, we give the second main result:

Theorem 1.4. *Let $\mu_{M,D}$ and $\mathring{\mathcal{F}}_p^2$ be defined by* (1-1) *and* (1-3), *respectively*, *where $M \in M_2(\mathbb{Z})$ is an expansive integer matrix and D is given by* (1-2). *If $\alpha_1\beta_2 - \alpha_2\beta_1 \notin 2\mathbb{Z}$, then the following statements are equivalent*:

(i) *$\mu_{M,D}$ is a spectral measure.*

(ii) *$M \in M_2(2\mathbb{Z})$.*

(iii) *$M\mathring{\mathcal{F}}_2^2 \subset \mathbb{Z}^2$.*

(iv) *(M, D) is admissible.*

We point out that the proofs of Theorems 1.3 and 1.4 are based on the precise form of the matrix $\tilde{M}$ in Theorem 1.3. Before giving the form, some technical work needs to be done. For an expansive integer matrix $M \in M_2(\mathbb{Z})$ and the digit set D given by (1-2), we can let $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $\alpha_1\beta_2 - \alpha_2\beta_1 = 2^\eta\gamma$ with $\eta \geq 0$ and $\gamma \notin 2\mathbb{Z}$. Without loss of generality, we assume $\gcd(\alpha_1, \alpha_2) = \alpha$ with $\alpha \notin 2\mathbb{Z}$ (otherwise, we can choose $\alpha = \gcd(\beta_1, \beta_2)$ with $\alpha \notin 2\mathbb{Z}$ since $\gcd(\alpha_1, \alpha_2, \beta_1, \beta_2) = 1$). Let $\alpha_1 = \alpha t_1$ and $\alpha_2 = \alpha t_2$ with $\gcd(t_1, t_2) = 1$. Then there exist $p, q \in \mathbb{Z}$ such that $pt_1 + qt_2 = 1$. Clearly, $\alpha = p\alpha_1 + q\alpha_2$ and $\alpha \mid \gamma$. For convenience, we define $\omega = p\beta_1 + q\beta_2$ and $\beta = \gamma/\alpha$. It is easy to check that $t_1\alpha_2 = t_2\alpha_1$ and $t_1\beta_2 - t_2\beta_1 = 2^\eta\beta$ with $\beta \notin 2\mathbb{Z}$.

Define $Q = \left(\begin{smallmatrix} p & q \\ -t_2 & t_1 \end{smallmatrix}\right)$. Then one has

$$\tilde{M} := QMQ^{-1} = \begin{pmatrix} (pa+qc)t_1+(pb+qd)t_2 & (pb+qd)p-(pa+qc)q \\ (ct_1-at_2)t_1+(dt_1-bt_2)t_2 & (dt_1-bt_2)p-(ct_1-at_2)q \end{pmatrix} \tag{1-4}$$

and

$$\tilde{D} := QD = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \begin{pmatrix} \omega \\ 2^\eta \beta \end{pmatrix}, \begin{pmatrix} -\alpha-\omega \\ -2^\eta \beta \end{pmatrix} \right\} \subset \mathbb{Z}^2. \tag{1-5}$$

Obviously, $\tilde{M}$ is an expansive integer matrix with $\det(\tilde{M}) = \det(M)$. Also, $\eta = 0$ and $\eta > 0$ are equivalent to $\alpha_1\beta_2 - \alpha_2\beta_1 \notin 2\mathbb{Z}$ and $\alpha_1\beta_2 - \alpha_2\beta_1 \in 2\mathbb{Z}$, respectively.

For $\eta = 0$ in $\tilde{D}$, we have the following conclusion, which is equivalent to Theorem 1.4 by using the property of similarity transformation.

Theorem 1.5. *Let $\mu_{\tilde{M},\tilde{D}}$ and $\mathring{\mathcal{F}}_p^2$ be defined by* (1-1) *and* (1-3), *respectively, where $\tilde{M}$ and $\tilde{D}$ are given by* (1-4) *and* (1-5), *respectively. If $\eta = 0$, then the following statements are equivalent*:

(i) *$\mu_{\tilde{M},\tilde{D}}$ is a spectral measure.*

(ii) $\tilde{M} \in M_2(2\mathbb{Z})$.

(iii) $\tilde{M}\mathring{\mathcal{F}}_2^2 \subset \mathbb{Z}^2$.

(iv) $(\tilde{M}, \tilde{D})$ *is admissible.*

On the other hand, if $\eta > 0$ in $\tilde{D}$, the form of $\tilde{M}$ is different from that in the case $\eta = 0$.

Theorem 1.6. *Let $\mu_{\tilde{M},\tilde{D}}$ be defined by* (1-1), *where $\tilde{M}$ and $\tilde{D}$ are given by* (1-4) *and* (1-5), *respectively. If $\eta > 0$, then $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure if and only if the matrix $\tilde{M} = \left(\begin{smallmatrix} \tilde{a} & \tilde{b} \\ \tilde{c} & \tilde{d} \end{smallmatrix}\right)$ satisfies $\tilde{a}, \tilde{d} \in 2\mathbb{Z}$ and $2^{\eta+1} \mid \tilde{c}$.*

We now give a brief explanation of the proofs of Theorems 1.5 and 1.6. The main technical difficulty in the proofs lies in "(i) $\Longrightarrow$ (ii)" of Theorem 1.5 and the necessity of Theorem 1.6. More precisely, the key point is to construct a Moran measure $\mu_{A,\tilde{M},\tilde{D}}$ (see (3-1)) so that it has the same spectrality as $\mu_{\tilde{M},\tilde{D}}$. For the matrix A, we need to cleverly describe its complete residue system (Proposition 3.3). We carefully investigate the structure of the spectrum of $\mu_{A,\tilde{M},\tilde{D}}$ (see (3-11)). And then we get a property of decomposition on the spectrum of $\mu_{\tilde{M},\tilde{D}}$ under the assumption that $\mu_{A,\tilde{M},\tilde{D}}$ is a spectral measure (Lemma 3.5). With their help, the proof becomes within reach.

The paper is organized as follows. In Section 2, we introduce some basic definitions and lemmas. In Section 3, we focus on proving Theorems 1.5 and 1.6. Finally, we prove Theorems 1.3 and 1.4, and give some concluding remarks in Section 4.

2. Preliminaries

For the self-affine measure $\mu_{M,D}$ defined by (1-1), the Fourier transform of $\mu_{M,D}$ is defined by

$$\hat{\mu}_{M,D}(\xi) = \int e^{2\pi i\langle x,\xi\rangle}\, d\mu_{M,D}(x) = \prod_{j=1}^{\infty} m_D(M^{*-j}\xi), \quad \xi \in \mathbb{R}^n, \tag{2-1}$$

where M^* denotes the transpose of M and $m_D(\cdot) = \frac{1}{\#D}\sum_{d\in D} e^{2\pi i\langle d,\cdot\rangle}$ is the *mask polynomial* of D. We denote the set of all the roots of $f(x)$ by $\mathcal{Z}(f)$, i.e., $\mathcal{Z}(f) = \{x : f(x) = 0\}$. Using (2-1), one has

$$\mathcal{Z}(\hat{\mu}_{M,D}) = \bigcup_{j=1}^{\infty} M^{*j}(\mathcal{Z}(m_D)). \tag{2-2}$$

For a countable set $\Lambda \subset \mathbb{R}^n$, $E_\Lambda = \{e^{2\pi i\langle\lambda,x\rangle} : \lambda \in \Lambda\}$ is an orthogonal family of $L^2(\mu_{M,D})$ if and only if $\hat{\mu}_{M,D}(\lambda_1 - \lambda_2) = 0$ for any $\lambda_1 \neq \lambda_2$, which is equivalent to

$$(\Lambda - \Lambda) \setminus \{\mathbf{0}\} \subset \mathcal{Z}(\hat{\mu}_{M,D}). \tag{2-3}$$

If E_Λ forms an orthogonal family of $L^2(\mu_{M,D})$, then Λ is called an *orthogonal set* of $\mu_{M,D}$. Note that the properties of spectra are invariant under a translation, so we can always assume that $\mathbf{0} \in \Lambda$.

In a number of applications, one encounters a measure μ and a subset Λ such that the functions $e^{2\pi i\langle\lambda,x\rangle}$ indexed by Λ are orthogonal in $L^2(\mu)$, but a separate argument is needed in order to show that the family is complete. Let

$$Q_{\mu,\Lambda}(\xi) = \sum_{\lambda\in\Lambda} |\hat{\mu}(\xi+\lambda)|^2, \quad \xi \in \mathbb{R}^n. \tag{2-4}$$

The following result is a basic criterion for the spectrality of μ.

Theorem 2.1 [24]. *Let μ be a Borel probability measure with compact support on $\mathbb{R}^n$, and let $\Lambda \subset \mathbb{R}^n$ be a countable set. Then*:

(i) Λ *is an orthogonal set of μ if and only if $Q_{\mu,\Lambda}(\xi) \le 1$ for $\xi \in \mathbb{R}^n$.*

(ii) Λ *is a spectrum of μ if and only if $Q_{\mu,\Lambda}(\xi) \equiv 1$ for $\xi \in \mathbb{R}^n$.*

The following lemma indicates that the spectrality of $\mu_{M,D}$ is invariant under a similarity transformation.

Lemma 2.2 [12]. *Let $D_1, D_2 \subset \mathbb{R}^n$ be two finite digit sets with the same cardinality, and let $M_1, M_2 \in M_n(\mathbb{R})$ be two expansive real matrices. If there exists a matrix $Q \in M_n(\mathbb{R})$ such that $M_2 = QM_1Q^{-1}$ and $D_2 = QD_1$, then μ_{M_1,D_1} is a spectral measure with spectrum Λ if and only if μ_{M_2,D_2} is a spectral measure with spectrum $Q^{*-1}\Lambda$.*

The following result is a known fact, which was proved in [16] and will be used in the proof of Proposition 3.3.

Lemma 2.3. *Let $M \in M_n(\mathbb{Z})$ be an expansive integer matrix, and let $D, S \subset \mathbb{Z}^n$ be two finite digit sets with the same cardinality. Then the following three statements are equivalent*:

(i) (M, D, S) *is a Hadamard triple.*

(ii) $m_D(M^{*-1}(s_1 - s_2)) = 0$ *for any distinct* $s_1, s_2 \in S$.

(iii) $(\delta_{M^{-1}D}, S)$ *is a spectral pair.*

Recalling that $\mu_{M,D}$ is defined by (1-1), we let A be a nonsingular matrix and define the Moran measure

$$\mu_{A,M,D} = \delta_{A^{-1}D} * \delta_{A^{-1}M^{-1}D} * \delta_{A^{-1}M^{-2}D} * \cdots . \tag{2-5}$$

It is clear that $\mu_{A,M,D} = \mu_{M,D}$ if $A = M$. The following lemma indicates the spectrality of $\mu_{M,D}$ is independent of A. The proof is the same as that of [9, Lemma 3.1; 10, Lemma 2.6]. For the convenience of readers, we include the proof here.

Lemma 2.4. *Let A be a nonsingular matrix, and let $\mu_{A,M,D}$ be defined by* (2-5). *Then*

$$\mu_{M,D} = \mu_{A,M,D} \circ (A^{-1}M).$$

Also, $(\mu_{M,D}, \Lambda)$ *is a spectral pair if and only if* $(\mu_{A,M,D}, A^*M^{*-1}\Lambda)$ *is a spectral pair.*

Proof. Applying (2-1) and (2-5), we have

$$\begin{aligned} \hat{\mu}_{A,M,D}(A^*M^{*-1}\xi) &= m_D(A^{*-1}A^*M^{*-1}\xi)\prod_{j=1}^{\infty} m_D(M^{*-j}A^{*-1}A^*M^{*-1}\xi) \\ &= \prod_{j=1}^{\infty} m_D(M^{*-j}\xi) = \hat{\mu}_{M,D}(\xi). \end{aligned} \tag{2-6}$$

Then $\mu_{M,D} = \mu_{A,M,D} \circ (A^{-1}M)$ by the uniqueness of Fourier transform.

Recall $Q_{\mu,\Lambda}(\xi)$ is defined by (2-4). Then, for $\xi \in \mathbb{R}^2$, it follows from (2-6) that

$$\begin{aligned} Q_{\mu_{M,D},\Lambda}(\xi) &= \sum_{\lambda\in\Lambda} |\hat{\mu}_{M,D}(\xi+\lambda)|^2 \\ &= \sum_{\lambda\in\Lambda} |\hat{\mu}_{A,M,D}(A^*M^{*-1}(\xi+\lambda))|^2 \\ &= \sum_{\lambda\in\Lambda} |\hat{\mu}_{A,M,D}(A^*M^{*-1}\xi + A^*M^{*-1}\lambda)|^2 \\ &= Q_{\mu_{A,M,D},A^*M^{*-1}\Lambda}(A^*M^{*-1}\xi). \end{aligned}$$

Hence the second assertion follows by Theorem 2.1. □

We conclude this section by recalling a useful lemma in our investigation, which was proved by Deng et al. in [9, Lemma 2.5].

Lemma 2.5. *Let $p_{i,j}$ be positive numbers such that $\sum_{j=1}^{n} p_{i,j} = 1$, and let $q_{i,j}$ be nonnegative numbers such that $\sum_{i=1}^{m} \max_{1\le j\le n} q_{i,j} \le 1$. Then*

$$\sum_{i=1}^{m}\sum_{j=1}^{n} p_{i,j}q_{i,j} = 1$$

if and only if $q_{i,1} = \cdots = q_{i,n}$ for $1 \le i \le m$ and $\sum_{i=1}^{m} q_{i,1} = 1$.

3. Proofs of Theorems 1.5 and 1.6

We focus on proving Theorems 1.5 and 1.6, that is, studying the spectrality of the measure $\mu_{\tilde{M},\tilde{D}}$, where $\tilde{M}$ and $\tilde{D}$ are given by (1-4) and (1-5), respectively. For this purpose, we first give some properties of $\mathcal{Z}(m_{\tilde{D}})$, and then investigate the structure of the spectrum of $\mu_{\tilde{M},\tilde{D}}$ under the assumption that $\mu_{A,\tilde{M},\tilde{D}}$ is a spectral measure, where $\mu_{A,\tilde{M},\tilde{D}}$ is defined by (2-5). With these preparations, we will achieve our goal.

By Lemma 2.4, without loss of generality, we assume in the rest of the paper that

$$A = \begin{pmatrix} 2^{\eta+1}\alpha\beta & 0 \\ 0 & 2^{\eta+1}\alpha\beta \end{pmatrix}.$$

The matrix A will be pivotal in constructing the spectrum of $\mu_{\tilde{M},\tilde{D}}$. Consequently,

$$\begin{aligned} \mu_{A,\tilde{M},\tilde{D}} &= \delta_{\frac{1}{2^{\eta+1}\alpha\beta}\tilde{D}} * (\mu_{\tilde{M},\tilde{D}} \circ 2^{\eta+1}\alpha\beta), \\ \hat{\mu}_{A,\tilde{M},\tilde{D}}(\xi) &= m_{\tilde{D}}\Big(\frac{\xi}{2^{\eta+1}\alpha\beta}\Big)\hat{\mu}_{\tilde{M},\tilde{D}}\Big(\frac{\xi}{2^{\eta+1}\alpha\beta}\Big). \end{aligned} \tag{3-1}$$

It is known that $m_{\tilde{D}}(x) = 0$ if and only if

$$\begin{cases} \alpha x_1 = \frac{1}{2} + k_1, \\ \omega x_1 + 2^\eta \beta x_2 = k_1', \end{cases} \quad \begin{cases} \alpha x_1 = k_2, \\ \omega x_1 + 2^\eta \beta x_2 = \frac{1}{2} + k_2', \end{cases} \quad \text{or} \quad \begin{cases} \alpha x_1 = \frac{1}{2} + k_3, \\ \omega x_1 + 2^\eta \beta x_2 = \frac{1}{2} + k_3', \end{cases} \tag{3-2}$$

where $k_1, k_2, k_3, k_1', k_2', k_3' \in \mathbb{Z}$. By a direct calculation, we have that

$$\mathcal{Z}(m_{\tilde{D}}) = \Theta_1 \cup \Theta_2 \cup \Theta_3, \tag{3-3}$$

where

$$\Theta_1 = \left\{ \frac{1}{2^{\eta+1}\alpha\beta} \begin{pmatrix} 2^\eta(2k_1\beta+\beta) \\ 2k_1'\alpha - 2k_1\omega - \omega \end{pmatrix} : k_1, k_1' \in \mathbb{Z} \right\},$$

$$\Theta_2 = \left\{ \frac{1}{2^{\eta+1}\alpha\beta} \begin{pmatrix} 2^{\eta+1}k_2\beta \\ 2k_2'\alpha - 2k_2\omega + \alpha \end{pmatrix} : k_2, k_2' \in \mathbb{Z} \right\},$$

$$\Theta_3 = \left\{ \frac{1}{2^{\eta+1}\alpha\beta} \begin{pmatrix} 2^\eta(2k_3\beta+\beta) \\ 2k_3'\alpha - 2k_3\omega + \alpha - \omega \end{pmatrix} : k_3, k_3' \in \mathbb{Z} \right\}.$$

Define

$$\Theta_0 = \left\{ \frac{1}{2^{\eta+1}\alpha\beta} \begin{pmatrix} 2^{\eta+1}k_0\beta \\ 2k_0'\alpha - 2k_0\omega \end{pmatrix} : k_0, k_0' \in \mathbb{Z} \right\}.$$

We now make a detailed analysis on the zero set $\mathcal{Z}(m_{\tilde{D}})$ of $m_{\tilde{D}}$.

Proposition 3.1. *With the above notation, the following statements hold*:

(i) $(\Theta_i - \Theta_i) \cap \mathcal{Z}(m_{\tilde{D}}) = \varnothing$ *for any* $i \in \{0, 1, 2, 3\}$.

(ii) $\Theta_i - \Theta_j \subset \mathcal{Z}(m_{\tilde{D}})$ *for any distinct* $i, j \in \{0, 1, 2, 3\}$.

(iii) *If* $\eta = 0$, *then* $\mathring{\mathcal{F}}_2^2 \subset \mathcal{Z}(m_{\tilde{D}})$, *where* $\mathring{\mathcal{F}}_2^2$ *is defined by* (1-3).

Proof. (i) Since $\alpha, \beta \in 2\mathbb{Z}+1$, from the definitions of $\mathcal{Z}(m_{\tilde{D}})$ and Θ_0, it can easily be seen that $\Theta_i - \Theta_i \subset \Theta_0$ for any $i \in \{0, 1, 2, 3\}$ and $\Theta_i \cap \Theta_0 = \varnothing$ for any $i \in \{1, 2, 3\}$. This yields $(\Theta_i - \Theta_i) \cap \mathcal{Z}(m_{\tilde{D}}) = \varnothing$ for all i, which proves (i).

(ii) For any $\theta_i \in \Theta_i$, it is easy to verify that

$$\pm(\theta_i - \theta_0) \in \Theta_i \quad (i \in \{1, 2, 3\}),$$
$$\pm(\theta_1 - \theta_2) \in \Theta_3, \quad \pm(\theta_1 - \theta_3) \in \Theta_2 \quad \text{and} \quad \pm(\theta_2 - \theta_3) \in \Theta_1.$$

Hence the assertion follows by using (3-3).

(iii) As $\eta = 0$ and $\alpha, \beta \in 2\mathbb{Z}+1$, it follows from (3-2) and (3-3) that

$$\left(\tfrac{1}{2}, 0\right)^t \in \Theta_1, \quad \left(0, \tfrac{1}{2}\right)^t \in \Theta_2 \quad \text{and} \quad \left(\tfrac{1}{2}, \tfrac{1}{2}\right)^t \in \Theta_3$$

if $\omega \in 2\mathbb{Z}$ and

$$\left(\tfrac{1}{2}, 0\right)^t \in \Theta_3, \quad \left(0, \tfrac{1}{2}\right)^t \in \Theta_2 \quad \text{and} \quad \left(\tfrac{1}{2}, \tfrac{1}{2}\right)^t \in \Theta_1$$

if $\omega \in 2\mathbb{Z}+1$. Therefore, $\mathring{\mathcal{F}}_2^2 \subset \Theta_1 \cup \Theta_2 \cup \Theta_3 = \mathcal{Z}(m_{\tilde{D}})$. ☐

Remark 3.2. Observing that $\alpha, \beta \in 2\mathbb{Z}+1$ in $\tilde{D}$, without loss of generality, we can further assume that $\alpha, \beta \geq 1$. In fact, if $\alpha < 0$ or $\beta < 0$, we take

$$Q = \begin{cases} \operatorname{diag}(-1, 1), & \text{if } \alpha < 0, \beta > 0, \\ \operatorname{diag}(1, -1), & \text{if } \alpha > 0, \beta < 0, \\ \operatorname{diag}(-1, -1), & \text{if } \alpha, \beta < 0. \end{cases}$$

Let $\bar{M} = Q\tilde{M}Q^{-1}$ and $\bar{D} = Q\tilde{D}$. By Lemma 2.2, we only need to consider the spectrality of $\mu_{\bar{M},\bar{D}}$. This implies that the assumption is reasonable.

To investigate the spectrality of $\mu_{\tilde{M},\tilde{D}}$, we need to construct a complete residue system of matrix A. In view of (3-1) and (3-3), one may easily get that

$$\mathcal{Z}(\hat{\mu}_{A,\tilde{M},\tilde{D}}) = \bigcup_{j=0}^{\infty} A^* \tilde{M}^{*j}(\mathcal{Z}(m_{\tilde{D}})) = \bigcup_{j=0}^{\infty} \tilde{M}^{*j}(2^{\eta+1}\alpha\beta(\Theta_1 \cup \Theta_2 \cup \Theta_3)) \subset \mathbb{Z}^2. \tag{3-4}$$

Throughout this paper, we set $\hbar_p = \{0, 1, \dots, p-1\}$ for an integer $p \geq 1$, and let

$$\mathcal{S}_q = \left\{ \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} : s_1 \in \hbar_{2^q\beta},\ s_2 \in \hbar_\alpha \right\} \quad \text{and} \quad \mathcal{T}_q = \bigcup_{i=0}^{3} \mathcal{T}_{q,i}, \tag{3-5}$$

where q is a nonnegative integer and

$$\mathcal{T}_{q,0} = \left\{ \frac{1}{2^{q+1}\alpha\beta} \begin{pmatrix} 2^{q+1}k_0\beta \\ 2k_0'\alpha - 2k_0\omega \end{pmatrix} : k_0 \in \hbar_\alpha,\ k_0' \in \hbar_{2^q\beta} \right\},$$
$$\mathcal{T}_{q,1} = \left\{ \frac{1}{2^{q+1}\alpha\beta} \begin{pmatrix} 2^q(2k_1\beta+\beta) \\ 2k_1'\alpha - 2k_1\omega - \omega \end{pmatrix} : k_1 \in \hbar_\alpha,\ k_1' \in \hbar_{2^q\beta} \right\},$$
$$\mathcal{T}_{q,2} = \left\{ \frac{1}{2^{q+1}\alpha\beta} \begin{pmatrix} 2^{q+1}k_2\beta \\ 2k_2'\alpha - 2k_2\omega + \alpha \end{pmatrix} : k_2 \in \hbar_\alpha,\ k_2' \in \hbar_{2^q\beta} \right\},$$
$$\mathcal{T}_{q,3} = \left\{ \frac{1}{2^{q+1}\alpha\beta} \begin{pmatrix} 2^q(2k_3\beta+\beta) \\ 2k_3'\alpha - 2k_3\omega + \alpha - \omega \end{pmatrix} : k_3 \in \hbar_\alpha,\ k_3' \in \hbar_{2^q\beta} \right\}.$$

Proposition 3.3. *With the above notation, the following statements hold*:

(i) $\mathcal{T}_{\eta,i} \subset \Theta_i$ *for any* $i \in \{0, 1, 2, 3\}$.

(ii) $(\delta_{A^{-1}\tilde{D}}, \mathcal{C})$ *is a spectral pair, where* $A = \operatorname{diag}(2^{\eta+1}\alpha\beta, 2^{\eta+1}\alpha\beta)$ *and* $\mathcal{C} = 2^{\eta+1}\alpha\beta\{\ell_0, \ell_1, \ell_2, \ell_3\}$ *for any* $\ell_i \in \mathcal{T}_{\eta,i}$.

(iii) $\mathcal{S}_\eta \oplus 2^{\eta+1}\alpha\beta\mathcal{T}_\eta$ *is a complete residue system of matrix* A *in* (ii).

Proof. According to the definitions of $\mathcal{T}_{\eta,i}$ and Θ_i, (i) is obvious. We now prove (ii). In view of Lemma 2.3, it suffices to prove that $m_{\tilde{D}}(A^{*-1}(c-c')) = 0$ for all distinct $c, c' \in \mathcal{C}$. Since $A = \operatorname{diag}(2^{\eta+1}\alpha\beta, 2^{\eta+1}\alpha\beta)$, it follows from Proposition 3.1(ii) and Proposition 3.3(i) that $A^{*-1}(c-c') \in \mathcal{Z}(m_{\tilde{D}})$. This implies $m_{\tilde{D}}(A^{*-1}(c-c')) = 0$, and the assertion (ii) follows.

Finally, we prove (iii). It is clear that the set $\mathcal{S}_\eta \oplus 2^{\eta+1}\alpha\beta\mathcal{T}_\eta$ can be written as

$$\begin{aligned} \mathcal{S}_\eta \oplus 2^{\eta+1}\alpha\beta\mathcal{T}_\eta &= \left\{ \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} : s_1 \in \hbar_{2^\eta\beta},\ s_2 \in \hbar_\alpha \right\} \oplus \begin{pmatrix} 2^\eta\beta & 0 \\ -\omega & \alpha \end{pmatrix} \left\{ \begin{pmatrix} k \\ k' \end{pmatrix} : k \in \hbar_{2\alpha},\ k' \in \hbar_{2^{\eta+1}\beta} \right\} \\ &:= \mathcal{S}_\eta \oplus \begin{pmatrix} 2^\eta\beta & 0 \\ -\omega & \alpha \end{pmatrix} \mathcal{Q}. \end{aligned} \tag{3-6}$$

To prove $\mathcal{S}_\eta \oplus 2^{\eta+1}\alpha\beta\mathcal{T}_\eta$ is a complete residue system of $A = \operatorname{diag}(2^{\eta+1}\alpha\beta, 2^{\eta+1}\alpha\beta)$, by using (3-6), it suffices to show that for any $(x, y)^t \in \mathbb{Z}^2$, there exist $(s_1, s_2)^t \in \mathcal{S}_\eta$, $(k, k')^t \in \mathcal{Q}$ and $(x', y')^t \in \mathbb{Z}^2$ such that

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \begin{pmatrix} 2^\eta\beta & 0 \\ -\omega & \alpha \end{pmatrix} \begin{pmatrix} k \\ k' \end{pmatrix} + 2^{\eta+1}\alpha\beta \begin{pmatrix} x' \\ y' \end{pmatrix}. \tag{3-7}$$

Since $\{0, 1, \dots, 2^\eta\beta - 1\} \oplus 2^\eta\beta\{0, 1, \dots, 2\alpha - 1\}$ is a complete residue system of $2^{\eta+1}\alpha\beta$, it follows that there exist $s_1 \in \{0, 1, \dots, 2^\eta\beta - 1\}$, $k \in \{0, 1, \dots, 2\alpha - 1\}$ and $x' \in \mathbb{Z}$ such that

$$x = s_1 + 2^\eta\beta k + 2^{\eta+1}\alpha\beta x'. \tag{3-8}$$

Also note that $\{0, 1, \dots, \alpha-1\}\oplus\alpha\{0, 1, \dots, 2^{\eta+1}\beta-1\}$ is another complete residue system of $2^{\eta+1}\alpha\beta$; thus there exist $s_2 \in \{0, 1, \dots, \alpha-1\}$, $k' \in \{0, 1, \dots, 2^{\eta+1}\beta-1\}$ and $y' \in \mathbb{Z}$ such that

$$y + \omega k = s_2 + \alpha k' + 2^{\eta+1}\alpha\beta y'. \tag{3-9}$$

The above equations (3-8) and (3-9) imply that (3-7) holds. □

Let Λ be a spectrum of $\mu_{A,\tilde{M},\tilde{D}}$ with $0 \in \Lambda$. By (2-3) and (3-4), we have $\Lambda \subset \mathbb{Z}^2$. This together with Proposition 3.3(iii) implies that for any $\lambda \in \Lambda$, there exist some $s \in \mathcal{S}_\eta$ and $\ell \in \mathcal{T}_\eta$ such that $\lambda = s + 2^{\eta+1}\alpha\beta\ell + 2^{\eta+1}\alpha\beta\gamma$ for some $\gamma \in \mathbb{Z}^2$. Then for $s \in \mathcal{S}_\eta$ and $\ell \in \mathcal{T}_\eta$, define

$$\Lambda_{s,\ell} = \{\gamma \in \mathbb{Z}^2 : s + 2^{\eta+1}\alpha\beta\ell + 2^{\eta+1}\alpha\beta\gamma \in \Lambda\}. \tag{3-10}$$

Then using (3-5), we have the decomposition

$$\Lambda = \bigcup_{s\in\mathcal{S}_\eta} \bigcup_{i\in\{0,1,2,3\}} \bigcup_{\ell\in\mathcal{T}_{\eta,i}} (s + 2^{\eta+1}\alpha\beta\ell + 2^{\eta+1}\alpha\beta\Lambda_{s,\ell}), \tag{3-11}$$

where $s + 2^{\eta+1}\alpha\beta\ell + 2^{\eta+1}\alpha\beta\Lambda_{s,\ell} = \varnothing$ if $\Lambda_{s,\ell} = \varnothing$. As $0 \in \Lambda$, it follows that

$$\Lambda_{0,0} \neq \varnothing. \tag{3-12}$$

Lemma 3.4. *Let Λ be a spectrum of $\mu_{A,\tilde{M},\tilde{D}}$ with $0 \in \Lambda$. If $\Lambda_{s,\ell}$ is a nonempty set, then $\Lambda_{s,\ell}$ is an orthogonal set of $\mu_{\tilde{M},\tilde{D}}$ for each $s \in \mathcal{S}_\eta$ and $\ell \in \mathcal{T}_\eta$.*

Proof. Suppose that $\Lambda_{s,\ell}$ is a nonempty set for $s \in \mathcal{S}_\eta$ and $\ell \in \mathcal{T}_\eta$. Then for any distinct $\lambda_1, \lambda_2 \in \Lambda_{s,\ell}$, it follows from (3-11) that

$$s + 2^{\eta+1}\alpha\beta\ell + 2^{\eta+1}\alpha\beta\lambda_1,\ s + 2^{\eta+1}\alpha\beta\ell + 2^{\eta+1}\alpha\beta\lambda_2 \in \Lambda.$$

Applying (2-3), we have $2^{\eta+1}\alpha\beta(\lambda_1 - \lambda_2) \in \mathcal{Z}(\hat{\mu}_{A,\tilde{M},\tilde{D}})$. Together with (3-1), $\lambda_1, \lambda_2 \in \mathbb{Z}^2$ and $m_{\tilde{D}}(\lambda_1 - \lambda_2) = 1$, we have

$$0 = \hat{\mu}_{A,\tilde{M},\tilde{D}}(2^{\eta+1}\alpha\beta(\lambda_1 - \lambda_2)) = m_{\tilde{D}}(\lambda_1 - \lambda_2)\hat{\mu}_{\tilde{M},\tilde{D}}(\lambda_1 - \lambda_2) = \hat{\mu}_{\tilde{M},\tilde{D}}(\lambda_1 - \lambda_2).$$

Thus $\lambda_1 - \lambda_2 \in \mathcal{Z}(\hat{\mu}_{\tilde{M},\tilde{D}})$, which means that $\Lambda_{s,\ell}$ is an orthogonal set of $\mu_{\tilde{M},\tilde{D}}$. □

The following lemma gives the structure of the spectrum of $\mu_{\tilde{M},\tilde{D}}$ under the assumption that $\mu_{A,\tilde{M},\tilde{D}}$ is a spectral measure.

Lemma 3.5. *Let Λ be a spectrum of $\mu_{A,\tilde{M},\tilde{D}}$ with $0 \in \Lambda$. For any $s \in \mathcal{S}_\eta$, choose a $i_s \in \{0, 1, 2, 3\}$ and write*

$$\Gamma = \bigcup_{s\in\mathcal{S}_\eta} \bigcup_{\ell\in\mathcal{T}_{\eta,i_s}} \left(\frac{s + 2^{\eta+1}\alpha\beta\ell}{2^{\eta+1}\alpha\beta} + \Lambda_{s,\ell} \right),$$

where $\Lambda_{s,\ell}$ is defined by (3-10). *Then Γ is a spectrum of $\mu_{\tilde{M},\tilde{D}}$ or an empty set.*

Proof. If Γ is a nonempty set, we will complete the proof in the following two steps.

Step 1. We prove that Γ is an orthogonal set of $\mu_{\tilde{M},\tilde{D}}$.

For any distinct $\varsigma_1, \varsigma_2 \in \Gamma$, we can write

$$\varsigma_k = \frac{s_k + 2^{\eta+1}\alpha\beta\ell_k}{2^{\eta+1}\alpha\beta} + \lambda_k,$$

where $s_k \in \mathcal{S}_\eta$, $\ell_k \in \mathcal{T}_{\eta,i_{s_k}}$, $\lambda_k \in \Lambda_{s_k,\ell_k}$ and $i_{s_k} \in \{0, 1, 2, 3\}$, $k = 1, 2$. Applying (3-1), the fact $\lambda_1, \lambda_2 \in \mathbb{Z}^2$ and the $\mathbb{Z}^2$-periodicity of $m_{\tilde{D}}$, one has

$$\begin{aligned} 0 &= \hat{\mu}_{A,\tilde{M},\tilde{D}}(2^{\eta+1}\alpha\beta(\varsigma_1 - \varsigma_2)) \\ &= m_{\tilde{D}}(\varsigma_1 - \varsigma_2)\hat{\mu}_{\tilde{M},\tilde{D}}(\varsigma_1 - \varsigma_2) \\ &= m_{\tilde{D}}\left(\frac{s_1 - s_2}{2^{\eta+1}\alpha\beta} + \ell_1 - \ell_2 + \lambda_1 - \lambda_2\right)\hat{\mu}_{\tilde{M},\tilde{D}}(\varsigma_1 - \varsigma_2) \\ &= m_{\tilde{D}}\left(\frac{s_1 - s_2}{2^{\eta+1}\alpha\beta} + \ell_1 - \ell_2\right)\hat{\mu}_{\tilde{M},\tilde{D}}(\varsigma_1 - \varsigma_2). \end{aligned} \tag{3-13}$$

We now claim that $m_{\tilde{D}}((s_1 - s_2)/(2^{\eta+1}\alpha\beta) + \ell_1 - \ell_2) \neq 0$. The proof will be divided into the following two cases.

Case 1: $s_1 = s_2$. In this case, it is clear that $\ell_1, \ell_2 \in \mathcal{T}_{\eta,i_{s_1}}$ by the definition of Γ. With Proposition 3.1(i) and Proposition 3.3(i), we derive that $\ell_1 - \ell_2 \notin \mathcal{Z}(m_{\tilde{D}})$. Thus the claim follows.

Case 2: $s_1 \neq s_2$. For this case, we prove the claim by contradiction. Suppose, on the contrary, that

$$\frac{s_1 - s_2}{2^{\eta+1}\alpha\beta} + \ell_1 - \ell_2 \in \mathcal{Z}(m_{\tilde{D}}). \tag{3-14}$$

By Proposition 3.1 and Proposition 3.3(i), one has $\ell_1 - \ell_2 \in \Theta_0 \cup \mathcal{Z}(m_{\tilde{D}})$. Combining this with (3-14), we conclude that

$$\frac{s_1 - s_2}{2^{\eta+1}\alpha\beta} \in \Theta_0 \cup \mathcal{Z}(m_{\tilde{D}}). \tag{3-15}$$

Using (3-5) and $s_1 \neq s_2$, it is easy to check that $s_1 - s_2 \in \mathfrak{B}$, where

$$\mathfrak{B} = \left\{ \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} : t_1 \in \{1 - 2^\eta\beta, \dots, 2^\eta\beta - 1\},\ t_2 \in \{1-\alpha, \dots, \alpha - 1\} \right\} \setminus \{\mathbf{0}\}.$$

Write $s_1 - s_2 = (t_1, t_2)^t \in \mathfrak{B}$. We first prove $t_1 = 0$. If $t_1 \neq 0$, it follows $t_1 \notin 2^\eta\beta\mathbb{Z}$. Then from the definitions of $\mathcal{Z}(m_{\tilde{D}})$ and Θ_0, it can easily be seen that

$$\frac{s_1 - s_2}{2^{\eta+1}\alpha\beta} = \frac{1}{2^{\eta+1}\alpha\beta} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \notin \Theta_0 \cup \mathcal{Z}(m_{\tilde{D}}).$$

This contradicts (3-15), which proves $t_1 = 0$.

Since $t_1 = 0$, it follows from $\beta \in 2\mathbb{Z} + 1$ that

$$\frac{s_1 - s_2}{2^{\eta+1}\alpha\beta} \notin \Theta_1 \cup \Theta_3.$$

Together with (3-15) and $t_1 = 0$, we have

$$\frac{s_1 - s_2}{2^{\eta+1}\alpha\beta} = \frac{1}{2^{\eta+1}\alpha\beta} \begin{pmatrix} 0 \\ t_2 \end{pmatrix} \in \Theta_0 \cup \Theta_2.$$

By a simple calculation, we deduce from $\beta \in 2\mathbb{Z} + 1$ that $t_2 \in \alpha\mathbb{Z}$. However, $(t_1, t_2)^t = (0, t_2)^t \in \mathfrak{B}$ means that $t_2 \notin \alpha\mathbb{Z}$, a contradiction. Hence the claim follows.

Applying the claim and (3-13), we obtain that

$$\hat{\mu}_{\tilde{M},\tilde{D}}(\varsigma_1 - \varsigma_2) = 0.$$

This implies that Γ is an orthogonal set of $\mu_{\tilde{M},\tilde{D}}$.

Step 2. We prove the completeness of the exponential function system $E_\Gamma = \{e^{2\pi i \langle \lambda, x\rangle} : \lambda \in \Gamma\}$.

Fix $s \in \mathcal{S}_\eta$. In view of Proposition 3.3(ii) and Theorem 2.1, one may get that, for any $\ell_{i_s} \in \mathcal{T}_{\eta,i_s}$,

$$\sum_{i_s=0}^{3} \left| m_{\tilde{D}}\left(\frac{s + 2^{\eta+1}\alpha\beta\ell_{i_s} + \xi}{2^{\eta+1}\alpha\beta} \right) \right|^2 \equiv 1. \tag{3-16}$$

In (3-16), let three of ℓ_0, ℓ_1, ℓ_2 and ℓ_3 be fixed, and the other be altered in $\mathcal{T}_{\eta,i_s}$. We can easily verify that, for all distinct $\ell, \ell' \in \mathcal{T}_{\eta,i_s}$,

$$\left| m_{\tilde{D}}\left(\frac{s + 2^{\eta+1}\alpha\beta\ell + \xi}{2^{\eta+1}\alpha\beta} \right) \right| = \left| m_{\tilde{D}}\left(\frac{s + 2^{\eta+1}\alpha\beta\ell' + \xi}{2^{\eta+1}\alpha\beta} \right) \right|. \tag{3-17}$$

Since $\Lambda_{s,\ell} \subset \mathbb{Z}^2$ and Λ is a spectrum of $\mu_{A,\tilde{M},\tilde{D}}$, it follows from the $\mathbb{Z}^2$-periodicity of $m_{\tilde{D}}(x)$ that

$$
\begin{aligned}
1 &\equiv \sum_{\lambda\in\Lambda} |\hat{\mu}_{A,\tilde{M},\tilde{D}}(\xi+\lambda)|^2 \\
&= \sum_{s\in\mathcal{S}_\eta}\sum_{i_s=0}^{3}\sum_{\ell\in\mathcal{T}_{\eta,i_s}}\sum_{\lambda'\in\Lambda_{s,\ell}} |\hat{\mu}_{A,\tilde{M},\tilde{D}}(\xi+s+2^{\eta+1}\alpha\beta\ell+2^{\eta+1}\alpha\beta\lambda')|^2 \\
&= \sum_{s\in\mathcal{S}_\eta}\sum_{i_s=0}^{3}\sum_{\ell\in\mathcal{T}_{\eta,i_s}} \Big|m_{\tilde{D}}\Big(\frac{s+2^{\eta+1}\alpha\beta\ell+\xi}{2^{\eta+1}\alpha\beta}\Big)\Big|^2 \\
&\qquad\qquad \sum_{\lambda'\in\Lambda_{s,\ell}} \Big|\hat{\mu}_{\tilde{M},\tilde{D}}\Big(\frac{s+2^{\eta+1}\alpha\beta\ell+\xi}{2^{\eta+1}\alpha\beta}+\lambda'\Big)\Big|^2 \\
&= \sum_{s\in\mathcal{S}_\eta}\sum_{i_s=0}^{3} \Big|m_{\tilde{D}}\Big(\frac{s+2^{\eta+1}\alpha\beta\ell_{i_s}+\xi}{2^{\eta+1}\alpha\beta}\Big)\Big|^2 \\
&\qquad\qquad \sum_{\ell\in\mathcal{T}_{\eta,i_s}}\sum_{\lambda'\in\Lambda_{s,\ell}} \Big|\hat{\mu}_{\tilde{M},\tilde{D}}\Big(\frac{s+2^{\eta+1}\alpha\beta\ell+\xi}{2^{\eta+1}\alpha\beta}+\lambda'\Big)\Big|^2,
\end{aligned}
\tag{3-18}
$$

where $\ell_{i_s} \in \mathcal{T}_{\eta,i_s}$, the first line follows from Theorem 2.1 and the second, third and fourth line follow from (3-11), (3-1) and (3-17), respectively.

We now choose $\xi \in \mathbb{R}^2 \setminus \mathbb{Q}^2$, and, for simplicity, write

$$
\begin{aligned}
p_{s,i_s} &= \Big|m_{\tilde{D}}\Big(\frac{s+2^{\eta+1}\alpha\beta\ell_{i_s}+\xi}{2^{\eta+1}\alpha\beta}\Big)\Big|^2, \\
q_{s,i_s} &= \sum_{\ell\in\mathcal{T}_{\eta,i_s}}\sum_{\lambda'\in\Lambda_{s,\ell}} \Big|\hat{\mu}_{\tilde{M},\tilde{D}}\Big(\frac{s+2^{\eta+1}\alpha\beta\ell+\xi}{2^{\eta+1}\alpha\beta}+\lambda'\Big)\Big|^2.
\end{aligned}
$$

Then one may derive from (3-3) that $p_{s,i_s} > 0$, and (3-18) becomes

$$
\sum_{s\in\mathcal{S}_\eta}\sum_{i_s=0}^{3} p_{s,i_s} q_{s,i_s} = 1. \tag{3-19}
$$

Note that Γ is an orthogonal set of $\mu_{\tilde{M},\tilde{D}}$; thus Theorem 2.1 implies that

$$
\sum_{s\in\mathcal{S}_\eta} \max\{q_{s,0}, q_{s,1}, q_{s,2}, q_{s,3}\} \le 1.
$$

Together with (3-16), (3-19) and Lemma 2.5, we conclude that

$$
\sum_{s\in\mathcal{S}_\eta}\sum_{\ell\in\mathcal{T}_{\eta,i_s}}\sum_{\lambda'\in\Lambda_{s,\ell}} \Big|\hat{\mu}_{\tilde{M},\tilde{D}}\Big(\frac{s+2^{\eta+1}\alpha\beta\ell+\xi}{2^{\eta+1}\alpha\beta}+\lambda'\Big)\Big|^2 = 1, \quad i_s = 0,1,2,3, \tag{3-20}
$$

and

$$\sum_{\ell\in\mathcal{T}_{\eta,0}}\sum_{\lambda'\in\Lambda_{s,\ell}}\left|\hat{\mu}_{\tilde{M},\tilde{D}}\left(\frac{s+2^{\eta+1}\alpha\beta\ell+\xi}{2^{\eta+1}\alpha\beta}+\lambda'\right)\right|^2 \tag{3-21}$$
$$=\sum_{\ell\in\mathcal{T}_{\eta,1}}\sum_{\lambda'\in\Lambda_{s,\ell}}\left|\hat{\mu}_{\tilde{M},\tilde{D}}\left(\frac{s+2^{\eta+1}\alpha\beta\ell+\xi}{2^{\eta+1}\alpha\beta}+\lambda'\right)\right|^2$$
$$=\sum_{\ell\in\mathcal{T}_{\eta,2}}\sum_{\lambda'\in\Lambda_{s,\ell}}\left|\hat{\mu}_{\tilde{M},\tilde{D}}\left(\frac{s+2^{\eta+1}\alpha\beta\ell+\xi}{2^{\eta+1}\alpha\beta}+\lambda'\right)\right|^2$$
$$=\sum_{\ell\in\mathcal{T}_{\eta,3}}\sum_{\lambda'\in\Lambda_{s,\ell}}\left|\hat{\mu}_{\tilde{M},\tilde{D}}\left(\frac{s+2^{\eta+1}\alpha\beta\ell+\xi}{2^{\eta+1}\alpha\beta}+\lambda'\right)\right|^2$$

for any $s\in\mathcal{S}_\eta$.

By continuity, we conclude that (3-20) and (3-21) hold for all $\xi\in\mathbb{R}^2$. Therefore, Theorem 2.1 shows that Γ is a spectrum of $\mu_{\tilde{M},\tilde{D}}$ for any group $\{i_s\}_{s\in\mathcal{S}_\eta}$ with $i_s\in\{0,1,2,3\}$. The proof is complete. □

Remark 3.6. Suppose $\Lambda=\bigcup_{s\in\mathcal{S}_\eta}\bigcup_{i\in\{0,1,2,3\}}\bigcup_{\ell\in\mathcal{T}_{\eta,i}}(s+2^{\eta+1}\alpha\beta\ell+2^{\eta+1}\alpha\beta\Lambda_{s,\ell})$ is a spectrum of $\mu_{A,\tilde{M},\tilde{D}}$ with $0\in\Lambda$. Then we can conclude from (3-21) that for any $s\in\mathcal{S}_\eta$, one of the following two statements holds:

(i) There exist some $\ell_{i_s}\in\mathcal{T}_{\eta,i_s}$ such that $\Lambda_{s,\ell_{i_s}}\neq\varnothing$ for all $0\le i_s\le 3$.

(ii) $\Lambda_{s,\ell}=\varnothing$ for any $\ell\in\mathcal{T}_\eta=\bigcup_{i=0}^{3}\mathcal{T}_{\eta,i}$.

In particular, the assumption $0\in\Lambda$ implies $\Lambda_{0,0}\neq\varnothing$. Therefore, (i) always holds for $s=0$, which illustrates that there must exist $\ell_{i_0}\in\mathcal{T}_{\eta,i_0}$ such that $\Lambda_{0,\ell_{i_0}}\neq\varnothing$ for all $1\le i_0\le 3$.

In order to prove Theorems 1.5 and 1.6 more conveniently, we define

$$\Phi_0=\{\upsilon\in\mathbb{Z}^2:\upsilon=(0,0)^t \pmod{2\mathbb{Z}^2}\},$$
$$\Phi_1=\{\upsilon\in\mathbb{Z}^2:\upsilon=(1,0)^t \pmod{2\mathbb{Z}^2}\},$$
$$\Phi_2=\{\upsilon\in\mathbb{Z}^2:\upsilon=(0,1)^t \pmod{2\mathbb{Z}^2}\},$$
$$\Phi_3=\{\upsilon\in\mathbb{Z}^2:\upsilon=(1,1)^t \pmod{2\mathbb{Z}^2}\}.$$

Then

$$\mathbb{Z}^2=\bigcup_{i=0}^{3}\Phi_i. \tag{3-22}$$

We have all ingredients for the proof of Theorem 1.5.

Proof of Theorem 1.5. We will prove this theorem by the circle (ii) $\Longrightarrow$ (iii) $\Longrightarrow$ (iv) $\Longrightarrow$ (i) $\Longrightarrow$ (ii).

(ii) $\Longrightarrow$ (iii): If $\tilde{M} \in M_2(2\mathbb{Z})$, we can write $\tilde{M} = \left(\begin{smallmatrix} 2\tilde{a} & 2\tilde{b} \\ 2\tilde{c} & 2\tilde{d} \end{smallmatrix}\right)$ with $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d} \in \mathbb{Z}$. Then with (1-3), it is easy to verify that

$$\tilde{M}\mathring{\mathcal{F}}_2^2 = \left\{ \begin{pmatrix} \tilde{a} \\ \tilde{c} \end{pmatrix}, \begin{pmatrix} \tilde{b} \\ \tilde{d} \end{pmatrix}, \begin{pmatrix} \tilde{a}+\tilde{b} \\ \tilde{c}+\tilde{d} \end{pmatrix} \right\} \subset \mathbb{Z}^2.$$

Hence the assertion follows.

(iii) $\Longrightarrow$ (iv): Suppose $\tilde{M}\mathring{\mathcal{F}}_2^2 \subset \mathbb{Z}^2$, which implies $\tilde{C} := \tilde{M}^*\mathcal{F}_2^2 \subset \mathbb{Z}^2$. Then using Lemma 2.3 and Proposition 3.1(iii), we obtain that $(\tilde{M}, \tilde{D}, \tilde{C})$ is a Hadamard triple. Therefore, $(\tilde{M}, \tilde{D})$ is admissible.

(iv) $\Longrightarrow$ (i): If $(\tilde{M}, \tilde{D})$ is admissible, $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure by Theorem 1.2.

(i) $\Longrightarrow$ (ii): Suppose that $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure, and let $A = \operatorname{diag}(2\alpha\beta, 2\alpha\beta)$. In view of Lemma 2.4, one may derive that $\mu_{A,\tilde{M},\tilde{D}}$ is also a spectral measure. Let Λ be a spectrum of $\mu_{A,\tilde{M},\tilde{D}}$ with $0 \in \Lambda$. First, we construct a spectrum of $\mu_{\tilde{M},\tilde{D}}$. Recall that $\mathcal{T}_{\eta,i}$ and Φ_i are defined by (3-5) and (3-22), respectively. By $\eta = 0$ and a simple calculation, one has $2\alpha\beta\tilde{M}^*\mathcal{T}_{\eta,0} \subset \Phi_0$. For $i \in \{1, 2, 3\}$, we can suppose that $2\alpha\beta\tilde{M}^*\mathcal{T}_{\eta,i} \subset \Phi_{j_i}$ for some $j_i \in \{0, 1, 2, 3\}$. Consequently,

$$\bigcup_{i=1}^{3} 2\alpha\beta\tilde{M}^*\mathcal{T}_{\eta,i} \subset \bigcup_{i=1}^{3} \Phi_{j_i}.$$

This means that for any $s \in \mathcal{S}_\eta \setminus \{0\}$, there exists $i_s \in \{0, 1, 2, 3\}$ such that $s + 2\alpha\beta\ell_s \notin \bigcup_{j=1}^{3} 2\alpha\beta\tilde{M}^*\mathcal{T}_{\eta,j} + 2\mathbb{Z}^2$ for any $\ell_s \in \mathcal{T}_{\eta,i_s}$. Define

$$\Gamma = \Delta_{0,0} \cup \bigcup_{s\in\mathcal{S}_\eta\setminus\{0\}} \Delta_{s,i_s}, \tag{3-23}$$

where $\Delta_{0,0} = \bigcup_{\ell_0\in\mathcal{T}_{\eta,0}} (\ell_0 + \Lambda_{0,\ell_0})$, $\Delta_{s,i_s} = \bigcup_{\ell_s\in\mathcal{T}_{\eta,i_s}} ((s + 2\alpha\beta\ell_s)/(2\alpha\beta) + \Lambda_{s,\ell_s})$ with

$$(s + 2\alpha\beta\mathcal{T}_{\eta,i_s}) \cap \left(\bigcup_{j=1}^{3} 2\alpha\beta\tilde{M}^*\mathcal{T}_{\eta,j} + 2\mathbb{Z}^2 \right) = \varnothing, \tag{3-24}$$

and Λ_{s,ℓ_s} is defined by (3-10). In view of Lemma 3.5, we get that Γ is a spectrum of $\mu_{\tilde{M},\tilde{D}}$. Moreover, it follows from $0 \in \Lambda$ and Lemma 2.4 that $0 \in \Gamma$.

Second, we prove that for any $i \in \{1, 2, 3\}$, there must exist $\ell_i \in \mathcal{T}_{\eta,i}$ such that $2\alpha\beta\tilde{M}^*\ell_i \in 2\mathbb{Z}^2$. Since Γ is a spectrum of $\mu_{\tilde{M},\tilde{D}}$ with $0 \in \Gamma$, it follows from Lemma 2.4 that $2\alpha\beta\tilde{M}^{*-1}\Gamma$ is a spectrum of $\mu_{A,\tilde{M},\tilde{D}}$ with $0 \in 2\alpha\beta\tilde{M}^{*-1}\Gamma$. Using (3-11), one has

$$2\alpha\beta\tilde{M}^{*-1}\Gamma = \bigcup_{s'\in\mathcal{S}_\eta} \bigcup_{i\in\{0,1,2,3\}} \bigcup_{\ell_i'\in\mathcal{T}_{\eta,i}} (s' + 2\alpha\beta\ell_i' + 2\alpha\beta\Lambda'_{s',\ell_i'}), \tag{3-25}$$

where

$$\Lambda'_{s',\ell'_i}=\{\gamma\in\mathbb{Z}^2 : s'+2\alpha\beta\ell'_i+2\alpha\beta\gamma\in 2\alpha\beta\tilde{M}^{*-1}\Gamma\}.$$

For $s'=0$ and $\ell'_i=0\in\mathcal{T}_{\eta,0}$, we have $\Lambda'_{0,0}\neq\varnothing$ since $0\in 2\alpha\beta\tilde{M}^{*-1}\Gamma$. By Remark 3.6, there must exist $\ell'_i\in\mathcal{T}_{\eta,i}$ such that $\Lambda'_{0,\ell'_i}\neq\varnothing$ for all $1\le i\le 3$. Let $\lambda'_i\in\Lambda'_{0,\ell'_i}$, where $i=1,2,3$. Therefore, (3-23) and (3-25) imply that there exist $s_i\in\mathcal{S}_\eta$, $\ell_i\in\bigcup_{j=0}^3\mathcal{T}_{\eta,j}$ and $\lambda_i\in\Lambda_{s_i,\ell_i}$ such that $(s_i+2\alpha\beta\ell_i)/(2\alpha\beta)+\lambda_i\in\Gamma$ and

$$2\alpha\beta\tilde{M}^*\ell'_i+2\alpha\beta\tilde{M}^*\lambda'_i=s_i+2\alpha\beta\ell_i+2\alpha\beta\lambda_i\quad\text{for } i=1,2,3. \tag{3-26}$$

Moreover, it follows from (3-24) that $s_i+2\alpha\beta\ell_i\notin\bigcup_{j=1}^3 2\alpha\beta\tilde{M}^*\mathcal{T}_{\eta,j}+2\mathbb{Z}^2$ if $s_i\neq 0$ for $i=1,2,3$. However, by noting that $\lambda_i,\lambda'_i\in\mathbb{Z}^2$, (3-26) implies that

$$s_i+2\alpha\beta\ell_i\in 2\alpha\beta\tilde{M}^*\ell'_i+2\mathbb{Z}^2\subset 2\alpha\beta\tilde{M}^*\mathcal{T}_{\eta,i}+2\mathbb{Z}^2\subset\bigcup_{j=1}^3 2\alpha\beta\tilde{M}^*\mathcal{T}_{\eta,j}+2\mathbb{Z}^2$$

for $i=1,2,3$. Therefore, the above discussion shows that $s_i=0$ for $i=1,2,3$, and hence $\ell_i\in\mathcal{T}_{\eta,0}$ by the definition of Γ. This implies $2\alpha\beta\ell_i\in 2\mathbb{Z}^2$ for $i=1,2,3$. Combining this with $\tilde{M}\in M_2(\mathbb{Z})$, $s_i=0$ and $\lambda_i,\lambda'_i\in\mathbb{Z}^2$, one may infer from (3-26) that

$$2\alpha\beta\tilde{M}^*\ell'_i=2\alpha\beta\ell_i+2\alpha\beta(\lambda_i-\tilde{M}^*\lambda'_i)\in 2\mathbb{Z}^2\quad\text{for } i=1,2,3.$$

Therefore, $2\alpha\beta\tilde{M}^*\ell'_i\in 2\mathbb{Z}^2$ for some $\ell'_i\in\mathcal{T}_{\eta,i}$, where $i=1,2,3$.

It remains to prove $\tilde{M}\in M_2(2\mathbb{Z})$. For any $i\in\{1,2,3\}$, the above conclusion shows that there must exist $\ell_i\in\mathcal{T}_{\eta,i}$ such that $2\alpha\beta\tilde{M}^*\ell_i\in 2\mathbb{Z}^2$. For these $\ell_i\in\mathcal{T}_{\eta,i}$, $i=1,2,3$, by the definition of $\mathcal{T}_{\eta,i}$ and the fact $\alpha,\beta\in 2\mathbb{Z}+1$, it can easily be checked that

$$\{2\alpha\beta\ell_i : i=1,2,3\}=\left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix},\begin{pmatrix}1\\1\end{pmatrix}\right\}\pmod{2\mathbb{Z}^2}.$$

This together with $2\alpha\beta\tilde{M}^*\ell_i\in 2\mathbb{Z}^2$ and a simple calculation gives that $\tilde{M}^*\in M_2(2\mathbb{Z})$, which is equivalent to $\tilde{M}\in M_2(2\mathbb{Z})$. This finishes the proof of Theorem 1.5. □

The following lemma plays an important role in the proof of Theorem 1.6.

Lemma 3.7. *Let $\mu_{\tilde{M},\tilde{D}}$ be a spectral measure, where $\tilde{M}$ and $\tilde{D}$ are given by* (1-4) *and* (1-5), *respectively. If $\eta>0$ in $\tilde{D}$, then $\tilde{M}=\left(\begin{smallmatrix}\tilde{a}&\tilde{b}\\\tilde{c}&\tilde{d}\end{smallmatrix}\right)$ satisfies $2^{\eta+1}\mid\tilde{c}$.*

Proof. Suppose, on the contrary, that $2^{\eta+1}\nmid\tilde{c}$. Then one may write $\tilde{c}=2^\tau c'$ for some integer $\tau\le\eta$ and $c'\in 2\mathbb{Z}+1$. Let $Q_1=\operatorname{diag}(1,1/2^\tau)$. A simple calculation gives

$$M_1:=Q_1\tilde{M}Q_1^{-1}=\begin{pmatrix}\tilde{a}&2^\tau\tilde{b}\\c'&\tilde{d}\end{pmatrix}\in M_2(\mathbb{Z})$$

and

$$D_1:=Q_1\tilde{D}=\left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\alpha\\0\end{pmatrix},\begin{pmatrix}\omega\\2^{\eta-\tau}\beta\end{pmatrix},\begin{pmatrix}-\alpha-\omega\\-2^{\eta-\tau}\beta\end{pmatrix}\right\}\subset\mathbb{Z}^2,$$

where $\alpha, \beta \in 2\mathbb{Z}+1$. Since $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure, it follows from Lemmas 2.2 and 2.4 that μ_{M_1,D_1} and μ_{A_1,M_1,D_1} are also spectral measures, where $A_1 = \mathrm{diag}(2^{\eta-\tau+1}\alpha\beta, 2^{\eta-\tau+1}\alpha\beta)$ and μ_{A_1,M_1,D_1} is defined by (2-5).

If $\tau = \eta$, it follows from Theorem 1.5 that $M_1 \in M_2(2\mathbb{Z})$. This means that $c' \in 2\mathbb{Z}$, a contradiction. Hence the assertion follows.

If $\tau < \eta$, we derive the contradiction by constructing a spectrum of μ_{M_1,D_1}. Recall that $\mathcal{S}_{\eta-\tau}$ and $\mathcal{T}_{\eta-\tau} = \bigcup_{i=0}^{3} \mathcal{T}_{\eta-\tau,i}$ are defined by (3-5). We first prove the following two claims.

Claim 1. *Let Φ_1 and Φ_3 be given by* (3-22). *Then*

$$2^{\eta-\tau+1}\alpha\beta M_1^* \mathcal{T}_{\eta-\tau,2} \subset \begin{cases} \Phi_1, & \text{if } \tilde{d} \in 2\mathbb{Z}, \\ \Phi_3, & \text{if } \tilde{d} \in 2\mathbb{Z}+1. \end{cases}$$

Proof of Claim 1. For any $\ell \in \mathcal{T}_{\eta-\tau,2}$, there exist $k \in \hbar_\alpha$ and $k' \in \hbar_{2^{\eta-\tau}\beta}$ such that

$$\ell = \frac{1}{2^{\eta-\tau+1}\alpha\beta} \begin{pmatrix} 2^{\eta-\tau+1}k\beta \\ 2k'\alpha - 2k\omega + \alpha \end{pmatrix}. \tag{3-27}$$

Since $M_1 = \left(\begin{smallmatrix} \tilde{a} & 2^\tau \tilde{b} \\ c' & \tilde{d} \end{smallmatrix}\right)$, $\tau < \eta$ and $\alpha, c' \in 2\mathbb{Z}+1$, it follows from (3-27) that

$$2^{\eta-\tau+1}\alpha\beta M_1^* \ell = \begin{pmatrix} 2(2^{\eta-\tau}k\tilde{a}\beta + (k'\alpha - k\omega)c') + c'\alpha \\ 2(2^\eta k\tilde{b}\beta + (k'\alpha - k\omega)\tilde{d}) + \tilde{d}\alpha \end{pmatrix} = \begin{pmatrix} 1 \\ \tilde{d} \end{pmatrix} \pmod{2\mathbb{Z}^2}.$$

Consequently, $2^{\eta-\tau+1}\alpha\beta M_1^* \ell \in \Phi_1$ if $\tilde{d} \in 2\mathbb{Z}$, and $2^{\eta-\tau+1}\alpha\beta M_1^* \ell \in \Phi_3$ if $\tilde{d} \in 2\mathbb{Z}+1$. So the claims follows. □

Claim 2. *Let Φ_1 and Φ_3 be given by* (3-22). *Then for any $s \in \mathcal{S}_{\eta-\tau} \setminus \{0\}$, the following two statements hold*:

(i) *There exist some $i_s \in \{0, 1, 2, 3\}$ such that $s + 2^{\eta-\tau+1}\alpha\beta\ell_s \notin \Phi_1$ for any $\ell_s \in \mathcal{T}_{\eta-\tau,i_s}$.*

(ii) *There exist some $i_s \in \{0, 1, 2, 3\}$ such that $s + 2^{\eta-\tau+1}\alpha\beta\ell_s \notin \Phi_3$ for any $\ell_s \in \mathcal{T}_{\eta-\tau,i_s}$.*

Proof of Claim 2. Begin by observing that if $\alpha \in 2\mathbb{Z}+1$ and $\tau < \eta$, then for any $\ell_i \in \mathcal{T}_{\eta-\tau,i}$, $i = 0, 1, 2, 3$, we have

$$2^{\eta-\tau+1}\alpha\beta\ell_0 = \begin{pmatrix} 2^{\eta-\tau+1}k_0\beta \\ 2k_0'\alpha - 2k_0\omega \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \pmod{2\mathbb{Z}^2},$$

$$2^{\eta-\tau+1}\alpha\beta\ell_1 = \begin{pmatrix} 2^{\eta-\tau}(2k_1\beta + \beta) \\ 2k_1'\alpha - 2k_1\omega - \omega \end{pmatrix} = \begin{pmatrix} 0 \\ \omega \end{pmatrix} \pmod{2\mathbb{Z}^2},$$

$$2^{\eta-\tau+1}\alpha\beta\ell_2 = \begin{pmatrix} 2^{\eta-\tau+1}k_2\beta \\ 2k_2'\alpha - 2k_2\omega + \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \pmod{2\mathbb{Z}^2}$$

and

$$2^{\eta-\tau+1}\alpha\beta\ell_3=\begin{pmatrix}2^{\eta-\tau}(2k_3\beta+\beta)\\2k_3'\alpha-2k_3\omega+\alpha-\omega\end{pmatrix}=\begin{pmatrix}0\\\omega-1\end{pmatrix}\pmod{2\mathbb{Z}^2}$$

for some $k_i\in\hbar_\alpha$ and $k_i'\in\hbar_{2^{\eta-\tau}\beta}$. Without loss of generality, we assume that $\omega\in 2\mathbb{Z}$ (the case $\omega\in 2\mathbb{Z}+1$ can be similarly proved). Then a simple calculation gives

$$2^{\eta-\tau+1}\alpha\beta\ell_0,\,2^{\eta-\tau+1}\alpha\beta\ell_1\in\Phi_0\quad\text{and}\quad 2^{\eta-\tau+1}\alpha\beta\ell_2,\,2^{\eta-\tau+1}\alpha\beta\ell_3\in\Phi_2.\tag{3-28}$$

Recall that $\mathcal{T}_{\eta-\tau}=\bigcup_{i=0}^{3}\mathcal{T}_{\eta-\tau,i}$. Then for any $s=(s_1,s_2)^t\in\mathcal{S}_{\eta-\tau}\setminus\{\mathbf{0}\}$, we take

$$\ell_s\in\begin{cases}\mathcal{T}_{\eta-\tau}, & \text{if } s_1\in 2\mathbb{Z},\\ \mathcal{T}_{\eta-\tau,2}\cup\mathcal{T}_{\eta-\tau,3}, & \text{if } s_1\in 2\mathbb{Z}+1,\, s_2\in 2\mathbb{Z},\\ \mathcal{T}_{\eta-\tau,0}\cup\mathcal{T}_{\eta-\tau,1}, & \text{if } s_1,s_2\in 2\mathbb{Z}+1.\end{cases}$$

This together with (3-28) yields that $s+2^{\eta-\tau+1}\alpha\beta\ell_s\notin\Phi_1$, which proves (i). For (ii), we take

$$\ell_s\in\begin{cases}\mathcal{T}_{\eta-\tau}, & \text{if } s_1\in 2\mathbb{Z},\\ \mathcal{T}_{\eta-\tau,0}\cup\mathcal{T}_{\eta-\tau,1}, & \text{if } s_1\in 2\mathbb{Z}+1,\, s_2\in 2\mathbb{Z},\\ \mathcal{T}_{\eta-\tau,2}\cup\mathcal{T}_{\eta-\tau,3}, & \text{if } s_1,s_2\in 2\mathbb{Z}+1.\end{cases}$$

Consequently, $s+2^{\eta-\tau+1}\alpha\beta\ell_s\notin\Phi_3$ by (3-28). Thus Claim 2 follows. □

We now continue with the proof of the case $\tau<\eta$. In the following proof, we might as well assume $\tilde{d}\in 2\mathbb{Z}$ in M_1. If $\tilde{d}\in 2\mathbb{Z}+1$, we only need to replace Claim 2(i) with Claim 2(ii).

Since $\tau<\eta$ and $\tilde{d}\in 2\mathbb{Z}$, it follows from Claim 2(i) that for any $s\in\mathcal{S}_{\eta-\tau}\setminus\{\mathbf{0}\}$, there must exist some $i_s\in\{0,1,2,3\}$ such that $s+2^{\eta-\tau+1}\alpha\beta\ell_s\notin\Phi_1$ for any $\ell_s\in\mathcal{T}_{\eta-\tau,i_s}$. Let $\tilde{\Lambda}$ be a spectrum of μ_{A_1,M_1,D_1} with $0\in\tilde{\Lambda}$. Define

$$\tilde{\Gamma}=\tilde{\Delta}_{0,0}\cup\bigcup_{s\in\mathcal{S}_{\eta-\tau}\setminus\{0\}}\tilde{\Delta}_{s,i_s},$$

where

$$\tilde{\Delta}_{0,0}=\bigcup_{\ell_0\in\mathcal{T}_{\eta-\tau,0}}(\ell_0+\tilde{\Lambda}_{0,\ell_0}),\qquad \tilde{\Delta}_{s,i_s}=\bigcup_{\ell_s\in\mathcal{T}_{\eta-\tau,i_s}}\left(\frac{s+2^{\eta-\tau+1}\alpha\beta\ell_s}{2^{\eta-\tau+1}\alpha\beta}+\tilde{\Lambda}_{s,\ell_s}\right)$$

with

$$(s+2^{\eta-\tau+1}\alpha\beta\mathcal{T}_{\eta-\tau,i_s})\cap\Phi_1=\varnothing,$$

and

$$\tilde{\Lambda}_{s,\ell_s}=\{\gamma\in\mathbb{Z}^2: s+2^{\eta-\tau+1}\alpha\beta\ell_s+2^{\eta-\tau+1}\alpha\beta\gamma\in\tilde{\Lambda}\}.$$

Using the similar argument as in the proof of Lemma 3.5, we can show that $\tilde{\Gamma}$ is a spectrum of μ_{M_1,D_1} with $0\in\tilde{\Gamma}$.

Next, we prove that there must exist $\ell \in \mathcal{T}_{\eta-\tau,2}$ such that $2^{\eta-\tau+1}\alpha\beta M_1^{*}\ell \in 2\mathbb{Z}^2$. Since $\tilde{\Gamma}$ is a spectrum of μ_{M_1,D_1} with $0 \in \tilde{\Gamma}$, it follows from Lemma 2.4 that $2^{\eta-\tau+1}\alpha\beta M_1^{*-1}\tilde{\Gamma}$ is a spectrum of μ_{A_1,M_1,D_1} with $0 \in 2^{\eta-\tau+1}\alpha\beta M_1^{*-1}\tilde{\Gamma}$. Similar to (3-25), we have that

$$2^{\eta-\tau+1}\alpha\beta M_1^{*-1}\tilde{\Gamma} = \bigcup_{s'\in\mathcal{S}_{\eta-\tau}} \bigcup_{i\in\{0,1,2,3\}} \bigcup_{\ell_i'\in\mathcal{T}_{\eta-\tau,i}} (s' + 2^{\eta-\tau+1}\alpha\beta\ell_i' + 2^{\eta-\tau+1}\alpha\beta\tilde{\Lambda}'_{s',\ell_i'}),$$

where

$$\tilde{\Lambda}'_{s',\ell_i'} = \{\gamma \in \mathbb{Z}^2 : s' + 2^{\eta-\tau+1}\alpha\beta\ell_i' + 2^{\eta-\tau+1}\alpha\beta\gamma \in 2^{\eta-\tau+1}\alpha\beta M_1^{*-1}\tilde{\Gamma}\}.$$

For $s' = 0$ and $\ell_i' = 0 \in \mathcal{T}_{\eta-\tau,0}$, it follows from $0 \in 2^{\eta-\tau+1}\alpha\beta M_1^{*-1}\tilde{\Gamma}$ that $\tilde{\Lambda}'_{0,0} \neq \varnothing$. Similar to Remark 3.6, one may infer that there exists $\ell_2' \in \mathcal{T}_{\eta-\tau,2}$ such that $\tilde{\Lambda}'_{0,\ell_2'} \neq \varnothing$. Therefore, applying Claim 1 and the similar argument as in the proof of Theorem 1.5, we can easily conclude that $2^{\eta-\tau+1}\alpha\beta M_1^{*}\ell_2' \in 2\mathbb{Z}^2$. Thus the assertion follows.

Finally, we prove $2^{\eta+1} \mid \tilde{c}$. The above discussion means that there exist some $\ell \in \mathcal{T}_{\eta-\tau,2}$ such that $2^{\eta-\tau+1}\alpha\beta M_1^{*}\ell \in 2\mathbb{Z}^2$. For these $\ell \in \mathcal{T}_{\eta-\tau,2}$, it follows from (3-27) that

$$2^{\eta-\tau+1}\alpha\beta M_1^{*}\ell = \begin{pmatrix} 2(2^{\eta-\tau}k\tilde{a}\beta+(k'\alpha-k\omega)c')+c'\alpha \\ 2(2^{\eta}k\tilde{b}\beta+(k'\alpha-k\omega)\tilde{d})+\tilde{d}\alpha \end{pmatrix}$$

for some $k \in \hbar_\alpha$ and $k' \in \hbar_{2^{\eta-\tau}\beta}$. Together with $2^{\eta-\tau+1}\alpha\beta M_1^{*}\ell \in 2\mathbb{Z}^2$, it yields that $c'\alpha \in 2\mathbb{Z}$. This contradicts the fact $c', \alpha \in 2\mathbb{Z}+1$, and hence the assumption $2^{\eta+1} \nmid \tilde{c}$ does not hold. Therefore, we obtain $2^{\eta+1} \mid \tilde{c}$, and complete the proof. □

Having established the above preparation, now we are in a position to prove Theorem 1.6.

Proof of Theorem 1.6. We first prove the necessity. Suppose $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure. In view of Lemma 3.7, we have that $\tilde{M} = \left(\begin{smallmatrix} \tilde{a} & \tilde{b} \\ \tilde{c} & \tilde{d} \end{smallmatrix}\right)$ satisfies $2^{\eta+1} \mid \tilde{c}$. Thus one may write $\tilde{c} = 2^{\eta+1}\kappa$ with $\kappa \in \mathbb{Z}$. Let $\tilde{Q} = \mathrm{diag}(1, 1/2^{\eta})$. By a simple calculation, we get

$$\bar{M} := \tilde{Q}\tilde{M}\tilde{Q}^{-1} = \begin{pmatrix} \tilde{a} & 2^{\eta}\tilde{b} \\ 2\kappa & \tilde{d} \end{pmatrix} \tag{3-29}$$

and

$$\bar{D} := \tilde{Q}\tilde{D} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \begin{pmatrix} \omega \\ \beta \end{pmatrix}, \begin{pmatrix} -\alpha-\omega \\ -\beta \end{pmatrix} \right\}. \tag{3-30}$$

Since $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure, it follows from Lemma 2.2 that $\mu_{\bar{M},\bar{D}}$ is also a spectral measure. Then with Theorem 1.5, we have $\bar{M} \in M_2(2\mathbb{Z})$. This together with (3-29) gives that $\tilde{a}, \tilde{d} \in 2\mathbb{Z}$. Hence the necessity follows.

Now we are devoted to proving the sufficiency. Suppose $\tilde{M} = \left(\begin{smallmatrix}\tilde{a} & \tilde{b}\\ \tilde{c} & \tilde{d}\end{smallmatrix}\right)$, where $\tilde{a}, \tilde{d} \in 2\mathbb{Z}$ and $2^{\eta+1} \mid \tilde{c}$. Then there exist $a^*, c^*, d^* \in \mathbb{Z}$ such that $\tilde{a} = 2a^*$, $\tilde{c} = 2^{\eta+1}c^*$ and $\tilde{d} = 2d^*$. Let $\tilde{Q} = \operatorname{diag}(1, 1/2^\eta)$. A simple calculation gives

$$M' := \tilde{Q}\tilde{M}\tilde{Q}^{-1} = \begin{pmatrix} 2a^* & 2^\eta \tilde{b} \\ 2c^* & 2d^* \end{pmatrix},$$

and $\bar{D} = \tilde{Q}\tilde{D}$ is given by (3-30). Since $\eta > 0$, it follows from Theorem 1.5 that $\mu_{M',\bar{D}}$ is a spectral measure. Therefore, $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure by Lemma 2.2.

This completes the proof of Theorem 1.6. □

4. Proofs of Theorems 1.3 and 1.4

We are committed to investigating the spectrality of the measure $\mu_{M,D}$, where $M \in M_2(\mathbb{Z})$ is an expansive integer matrix and D is given by (1-2). We first prove Theorem 1.3 by using Theorems 1.5 and 1.6, and then prove Theorem 1.4. Finally, we provide some concluding remarks.

Proof of Theorem 1.3. The sufficiency follows directly from Theorem 1.2 and Lemma 2.2. Now we are devoted to proving the necessity. Suppose that $\mu_{M,D}$ is a spectral measure. Let $\eta = \max\{r : 2^r \mid (\alpha_1\beta_2 - \alpha_2\beta_1)\}$, and let $\tilde{M}$ and $\tilde{D}$ be given by (1-4) and (1-5), respectively. That is, $\tilde{M} = QMQ^{-1}$ and $\tilde{D} = QD$. In view of Lemma 2.2, $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure. It suffices to prove that there exists a matrix $\tilde{Q} \in M_2(\mathbb{R})$ such that $(\bar{M}, \bar{D})$ is admissible, where $\bar{M} = \tilde{Q}\tilde{M}\tilde{Q}^{-1}$ and $\bar{D} = \tilde{Q}\tilde{D}$. The proof will be divided into the following two cases.

Case 1: $\eta = 0$. Since $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure, it follows from $\eta = 0$ and Theorem 1.5 that $(\tilde{M}, \tilde{D})$ is admissible. Thus the assertion follows by taking $\tilde{Q} = \operatorname{diag}(1, 1)$.

Case 2: $\eta > 0$. Since $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure, Theorem 1.6 implies that one may write $\tilde{M} = \left(\begin{smallmatrix} 2a' & b' \\ 2^{\eta+1}c' & 2d' \end{smallmatrix}\right)$, where $a', b', c', d' \in \mathbb{Z}$. We take $\tilde{Q} = \operatorname{diag}(1, 1/2^\eta)$. Then

$$\bar{M} = \tilde{Q}\tilde{M}\tilde{Q}^{-1} = \begin{pmatrix} 2a' & 2^\eta b' \\ 2c' & 2d' \end{pmatrix} \quad \text{and} \quad \bar{D} = \tilde{Q}\tilde{D} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \begin{pmatrix} \omega \\ \beta \end{pmatrix}, \begin{pmatrix} -\alpha-\omega \\ -\beta \end{pmatrix} \right\}.$$

Using $\eta > 0$, it is clear that $\bar{M} \in M_2(2\mathbb{Z})$. Hence $(\bar{M}, \bar{D})$ is admissible by Theorem 1.5.

This completes the proof of Theorem 1.3. □

Next, we focus on proving Theorem 1.4.

Proof of Theorem 1.4. Let $\tilde{M}$ and $\tilde{D}$ be given by (1-4) and (1-5), respectively. That is,

$$\tilde{M} = QMQ^{-1} \quad \text{and} \quad \tilde{D} = QD, \tag{4-1}$$

where the matrix $Q \in M_2(\mathbb{Z})$ satisfies $\det(Q) = 1$. In view of Lemma 2.2, $\mu_{M,D}$ is a spectral measure if and only if $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure. This implies that Theorem 1.4(i) is equivalent to Theorem 1.5(i). Note that $\det(Q) = 1$; hence, by a simple calculation, one has that

$$M \in M_2(2\mathbb{Z}) \iff \tilde{M} \in M_2(2\mathbb{Z}).$$

Thus Theorem 1.4(ii) and (iii) are equivalent to Theorem 1.5(ii) and (iii), respectively. Finally, from the Definition 1.1 and (4-1), it is easy to see that $(\tilde{M}, \tilde{D})$ is admissible $\iff$ there exists a set $\tilde{C} \subset \mathbb{Z}^2$ such that $(\tilde{M}, \tilde{D}, \tilde{C})$ is a Hadamard triple $\iff$ $(M, D, Q^*\tilde{C})$ is a Hadamard triple $\iff$ (M, D) is admissible. Consequently, Theorem 1.4(iv) is equivalent to Theorem 1.5(iv).

Therefore, the desired result now is obtained by appeal to Theorem 1.5. □

At the end of this paper, we give some further remarks and list an open question which is related to our main results. The following example is specifically used to display our results, which are convenient to judge whether the measure $\mu_{M,D}$ in Question 1 is a spectral measure.

Example 4.1. Let $M_1 = \left(\begin{smallmatrix} 2 & b \\ 2 & 2 \end{smallmatrix}\right)$ and $M_2 = \left(\begin{smallmatrix} 2 & b \\ 4 & 2 \end{smallmatrix}\right)$ be two expansive integer matrices, and let

$$D_1 = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\} \quad \text{and} \quad D_2 = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ -2 \end{pmatrix} \right\}.$$

Then the following statements hold:

(i) μ_{M_1,D_1} and μ_{M_2,D_1} are spectral measures if and only if $b \in 2\mathbb{Z}$.

(ii) μ_{M_1,D_2} is a nonspectral measure, while μ_{M_2,D_2} is a spectral measure.

Proof. By a simple calculation, this follows directly from Theorems 1.5 and 1.6. □

It is worth noting that if $\alpha_1\beta_2 - \alpha_2\beta_1 \in 2\mathbb{Z}$ in Theorem 1.3, we cannot give the specific form of matrix M. However, if α_1, α_2, β_1 and β_2 are fixed, we can describe the specific form by applying Theorem 1.6. The following simple but interesting example is devoted to illustrating this fact.

Example 4.2. Let $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ be an expansive integer matrix, and let

$$D = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 8 \end{pmatrix}, \begin{pmatrix} -4 \\ -10 \end{pmatrix} \right\}.$$

Then $\mu_{M,D}$ is a spectral measure if and only if $a, d \in 2\mathbb{Z}$ and $c \in 4\mathbb{Z}$.

Proof. Write $Q = \left(\begin{smallmatrix} 3 & -1 \\ -2 & 1 \end{smallmatrix}\right)$. Then it is direct to compute that

$$\tilde{M} := QMQ^{-1} = \begin{pmatrix} 3a-c+2(3b-d) & 3a-c+3(3b-d) \\ c-2a+2(d-2b) & c-2a+3(d-2b) \end{pmatrix}$$

and

$$\tilde{D} := QD = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix} \right\}.$$

By Lemma 2.2, $\mu_{M,D}$ is a spectral measure if and only if $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure.

For the sufficiency, it follows from $a, d \in 2\mathbb{Z}$ and $c \in 4\mathbb{Z}$ that there exist $\tilde{a}, \tilde{c}, \tilde{d} \in \mathbb{Z}$ such that $a = 2\tilde{a}$, $d = 2\tilde{d}$ and $c = 4\tilde{c}$. Thus $\tilde{M}$ becomes

$$\tilde{M} = \begin{pmatrix} 2(3\tilde{a}-2\tilde{c}+3b-d) & 3a-c+3(3b-d) \\ 4(\tilde{c}-\tilde{a}+\tilde{d}-b) & 2(2\tilde{c}-a+3\tilde{d}-b) \end{pmatrix}.$$

This together with Theorem 1.6 yields that $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure, and hence the sufficiency follows.

Conversely, suppose $\mu_{\tilde{M},\tilde{D}}$ is a spectral measure. Applying Theorem 1.6, we have

$$\begin{aligned} &3a - c + 2(3b - d) \in 2\mathbb{Z}, \\ &c - 2a + 2(d - 2b) \in 4\mathbb{Z}, \\ &c - 2a + 3(d - 2b) \in 2\mathbb{Z}. \end{aligned}$$

Consequently, $3a - c$, $c + 3d \in 2\mathbb{Z}$ and $c - 2a + 2d \in 4\mathbb{Z}$. By a simple calculation, we infer that $a, d \in 2\mathbb{Z}$ and $c \in 4\mathbb{Z}$. This proves the necessity. $\square$

We remark here that the digit set D in (1-2) satisfies $\alpha_1\beta_2 - \alpha_2\beta_1 \neq 0$, and so it is of interest to consider the following question:

Question 2. For an expansive matrix $M \in M_2(\mathbb{Z})$ and the digit set

$$D = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \begin{pmatrix} -\alpha_1-\beta_1 \\ -\alpha_2-\beta_2 \end{pmatrix} \right\}$$

with $\alpha_1\beta_2 - \alpha_2\beta_1 = 0$, what is the sufficient and necessary condition for $\mu_{M,D}$ to be a spectral measure?

In fact, for the matrix M and the digit set D given in the above question, using the methods of [34], we can find an integer matrix Q such that $\bar{M} := QMQ^{-1}$ and

$$\bar{D} := QD = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \begin{pmatrix} \beta \\ 0 \end{pmatrix}, \begin{pmatrix} -\alpha-\beta \\ 0 \end{pmatrix} \right\},$$

where $\alpha, \beta \in \mathbb{Z}$ and $\bar{M}$ is an expansive integer matrix with $\det(\bar{M}) = \det(M)$. Lemma 2.2 indicates that to consider the spectrality of $\mu_{M,D}$, we only need to consider the measure $\mu_{\bar{M},\bar{D}}$. However, it is apparent that the set $\mathcal{Z}(m_{\bar{D}})$ includes free variables since the root of

$$m_{\bar{D}}(\xi) = \frac{1}{\#\bar{D}} \sum_{d \in \bar{D}} e^{2\pi i \langle d, \xi \rangle} = \frac{1}{\#\bar{D}} (1 + e^{2\pi\alpha\xi_1} + e^{2\pi\beta\xi_1} + e^{2\pi(-\alpha-\beta)\xi_1}) = 0$$

is independent of ξ_2, where $\xi = (\xi_1, \xi_2)^t$. We have not yet discovered an effective method to address this situation. An answer to Question 2 may provide insights into the study of the spectrality of fractal measures.

Acknowledgement

The authors would like to thank the anonymous referees for valuable suggestions.

References

[1] L.-X. An and X.-G. He, "A class of spectral Moran measures", *J. Funct. Anal.* **266**:1 (2014), 343–354. MR Zbl

[2] L.-X. An, X.-G. He, and L. Tao, "Spectrality of the planar Sierpinski family", *J. Math. Anal. Appl.* **432**:2 (2015), 725–732. MR Zbl

[3] L.-X. An, X. He, and C.-K. Lai, "Classification of spectral self-similar measures with four-digit elements", preprint, 2022. arXiv 2209.05619

[4] M.-L. Chen and J.-C. Liu, "The cardinality of orthogonal exponentials of planar self-affine measures with three-element digit sets", *J. Funct. Anal.* **277**:1 (2019), 135–156. MR Zbl

[5] M.-L. Chen, J.-C. Liu, and X.-Y. Wang, "Spectrality of a class of self-affine measures on $\mathbb{R}^2$", *Nonlinearity* **34**:11 (2021), 7446–7469. MR Zbl

[6] X.-R. Dai, "When does a Bernoulli convolution admit a spectrum?", *Adv. Math.* **231**:3-4 (2012), 1681–1693. MR Zbl

[7] X.-R. Dai, X.-G. He, and K.-S. Lau, "On spectral N-Bernoulli measures", *Adv. Math.* **259** (2014), 511–531. MR Zbl

[8] Q.-R. Deng and K.-S. Lau, "Sierpinski-type spectral self-similar measures", *J. Funct. Anal.* **269**:5 (2015), 1310–1326. MR Zbl

[9] Q.-R. Deng and M.-T. Li, "Spectrality of Moran-type self-similar measures on $\mathbb{R}$", *J. Math. Anal. Appl.* **506**:1 (2022), art. id. 125547. MR Zbl

[10] Q.-R. Deng, X.-G. He, M.-T. Li, and Y.-L. Ye, "Spectrality of Moran–Sierpinski measures", To appear in *Acta Math. Sin.* (*Engl. Ser.*).

[11] D. E. Dutkay and P. E. T. Jorgensen, "Wavelets on fractals", *Rev. Mat. Iberoam.* **22**:1 (2006), 131–180. MR Zbl

[12] D. E. Dutkay and P. E. T. Jorgensen, "Analysis of orthogonality and of orbits in affine iterated function systems", *Math. Z.* **256**:4 (2007), 801–823. MR Zbl

[13] D. E. Dutkay and P. E. T. Jorgensen, "Fourier frequencies in affine iterated function systems", *J. Funct. Anal.* **247**:1 (2007), 110–137. MR Zbl

[14] D. E. Dutkay and P. E. T. Jorgensen, "Duality questions for operators, spectrum and measures", *Acta Appl. Math.* **108**:3 (2009), 515–528. MR Zbl

[15] D. E. Dutkay and P. E. T. Jorgensen, "Probability and Fourier duality for affine iterated function systems", *Acta Appl. Math.* **107**:1-3 (2009), 293–311. MR Zbl

[16] D. E. Dutkay, J. Haussermann, and C.-K. Lai, "Hadamard triples generate self-affine spectral measures", *Trans. Amer. Math. Soc.* **371**:2 (2019), 1439–1481. MR Zbl

[17] Y.-S. Fu and M.-W. Tang, "Spectrality of homogeneous Moran measures on $\mathbb{R}^n$", *Forum Math.* **35**:1 (2023), 201–219. MR Zbl

[18] X. Fu, X. He, and K.-S. Lau, "Spectrality of self-similar tiles", *Constr. Approx.* **42**:3 (2015), 519–541. MR Zbl

[19] Y.-S. Fu, X.-G. He, and Z.-X. Wen, "Spectra of Bernoulli convolutions and random convolutions", *J. Math. Pures Appl.* (9) **116** (2018), 105–131. MR Zbl

[20] B. Fuglede, "Commuting self-adjoint partial differential operators and a group theoretic problem", *J. Functional Analysis* **16** (1974), 101–121. MR Zbl

[21] K. Gröchenig and W. R. Madych, "Multiresolution analysis, Haar bases, and self-similar tilings of $\mathbf{R}^n$", *IEEE Trans. Inform. Theory* **38**:2, part 2 (1992), 556–568. MR Zbl

[22] X.-G. He, C.-K. Lai, and K.-S. Lau, "Exponential spectra in $L^2(\mu)$", *Appl. Comput. Harmon. Anal.* **34**:3 (2013), 327–338. MR Zbl

[23] J. E. Hutchinson, "Fractals and self-similarity", *Indiana Univ. Math. J.* **30**:5 (1981), 713–747. MR Zbl

[24] P. E. T. Jorgensen and S. Pedersen, "Dense analytic subspaces in fractal L^2-spaces", *J. Anal. Math.* **75** (1998), 185–228. MR Zbl

[25] P. E. T. Jorgensen, K. A. Kornelson, and K. L. Shuman, "Affine systems: asymptotics at infinity for fractal measures", *Acta Appl. Math.* **98**:3 (2007), 181–222. MR Zbl

[26] J.-P. Kahane, "Géza Freud and lacunary Fourier series", *J. Approx. Theory* **46**:1 (1986), 51–57. MR Zbl

[27] J. Kigami, "Local Nash inequality and inhomogeneity of heat kernels", *Proc. London Math. Soc.* (3) **89**:2 (2004), 525–544. MR Zbl

[28] J. Kigami, R. S. Strichartz, and K. C. Walker, "Constructing a Laplacian on the diamond fractal", *Experiment. Math.* **10**:3 (2001), 437–448. MR Zbl

[29] I. Ł aba and Y. Wang, "On spectral Cantor measures", *J. Funct. Anal.* **193**:2 (2002), 409–420. MR Zbl

[30] M. L. Lapidus, J. W. Neuberger, R. J. Renka, and C. A. Griffith, "Snowflake harmonics and computer graphics: numerical computation of spectra on fractal drums", *Internat. J. Bifur. Chaos Appl. Sci. Engrg.* **6**:7 (1996), 1185–1210. MR Zbl

[31] J.-L. Li, "Non-spectral problem for a class of planar self-affine measures", *J. Funct. Anal.* **255**:11 (2008), 3125–3148. MR Zbl

[32] J. Li, "Spectral self-affine measures on the planar Sierpinski family", *Sci. China Math.* **56**:8 (2013), 1619–1628. MR Zbl

[33] J.-L. Li, "Analysis of μ_M, D-orthogonal exponentials for the planar four-element digit sets", *Math. Nachr.* **287**:2-3 (2014), 297–312. MR Zbl

[34] J.-C. Liu and J. J. Luo, "Spectral property of self-affine measures on $\mathbb{R}^n$", *J. Funct. Anal.* **272**:2 (2017), 599–612. MR Zbl

[35] J.-C. Liu and Z.-Y. Wang, "The spectrality of self-affine measure under the similar transformation of $GL_n(p)$", *Constr. Approx.* **58**:3 (2023), 687–712. MR Zbl

[36] J.-C. Liu, Y. Zhang, Z.-Y. Wang, and M.-L. Chen, "Spectrality of generalized Sierpinski-type self-affine measures", *Appl. Comput. Harmon. Anal.* **55** (2021), 129–148. MR Zbl

[37] R. S. Strichartz, "Self-similarity in harmonic analysis", *J. Fourier Anal. Appl.* **1**:1 (1994), 1–37. MR Zbl

[38] R. S. Strichartz, "Mock Fourier series and transforms associated with certain Cantor measures", *J. Anal. Math.* **81** (2000), 209–238. MR Zbl

[39] R. S. Strichartz, "Convergence of mock Fourier series", *J. Anal. Math.* **99** (2006), 333–353. MR Zbl

[40] J. Su, Y. Liu, and J.-C. Liu, "Non-spectrality of the planar self-affine measures with four-element digit sets", *Fractals* **27**:7 (2019), art. id. 1950115. MR Zbl

[41] J. Su, Z.-Y. Wang, and M.-L. Chen, "Orthogonal exponential functions of the planar self-affine measures with four digits", *Fractals* **28**:1 (2020), art. id. 2050016. Zbl

Received September 17, 2023. Revised December 8, 2023.

MING-LIANG CHEN
SCHOOL OF MATHEMATICS AND COMPUTER SCIENCE
GANNAN NORMAL UNIVERSITY
GUANGZHOU
CHINA
mathcml@163.com

JING-CHENG LIU
KEY LABORATORY OF COMPUTING AND STOCHASTIC MATHEMATICS (MINISTRY OF EDUCATION)
HUNAN NORMAL UNIVERSITY
CHANGSHA
CHINA
jcliu@hunnu.edu.cn

ZHI-YONG WANG
COLLEGE OF MATHEMATICS AND COMPUTATIONAL SCIENCE
HUNAN FIRST NORMAL UNIVERSITY
CHANGSHA
CHINA
wzyzzql@163.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 327, No. 1, 2023
https://doi.org/10.2140/pjm.2023.327.83

GROUP TOPOLOGIES ON AUTOMORPHISM GROUPS OF HOMOGENEOUS STRUCTURES

ZANIAR GHADERNEZHAD AND JAVIER DE LA NUEZ GONZÁLEZ

We provide sufficient conditions for the standard topology (generated by stabilizers of finite sets) on the automorphism group of a countable homogeneous structure to be minimal among all Hausdorff group topologies on the group. Under certain assumptions, such as when the structure is the Fraïssé limit of a relational class with the free amalgamation property, we are able to classify all the group topologies on the automorphism group coarser than the standard topology even when the latter is not minimal.

1. Introduction

Minimality. A topological group (G, τ) consists of a group $(G, \cdot)$ and a topology τ on G such that the map $\rho : G \times G \to G$, where $\rho(g, h) = gh^{-1}$, is jointly continuous.

Definition 1.1. A Hausdorff topological group (G, τ) is called *minimal* if G does not admit a Hausdorff group topology strictly coarser than τ or, equivalently, if every bijective continuous homomorphism from G to another Hausdorff topological group is a homeomorphism. The topological group (G, τ) is *totally minimal* if every continuous surjective homomorphism to a Hausdorff topological group is open.

Clearly, every totally minimal group is minimal. Also, for a topological group (G, τ), if the only strictly coarser topology is $\{\varnothing, G\}$ then (G, τ) is totally minimal. Indeed, in that case for any continuous surjective homomorphism $\phi : (G, \tau) \to (H, \sigma)$ the pullback $\phi^*(\sigma)$ of σ by ϕ satisfies $\phi^*(\sigma) \subseteq \tau$ and thus $\phi^*(\sigma) \in \{\tau, \{\varnothing, G\}\}$, so the map σ is either a homeomorphism or the trivial map. For a group topology $\tau' \subsetneq \tau$, by considering the closure of the identity in τ', one easily sees that this applies, in particular, to the case in which (G, τ) is minimal and has no nontrivial normal closed subgroups.

Ghadernezhad supported by the Leverhulme Grant RPG-2017-179. González supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreements no. 291111 and no. 336983 and from the Basque Government Grant IT974-16.

MSC2020: primary 22A05, 22F50; secondary 03C15, 03C45, 20B27.

Keywords: minimal topological groups, automorphism groups, oligomorphic groups.

The notion of minimality for topological groups was introduced as early as 1971 as a generalization of compactness. In fact it is easy to see that any compact Hausdorff topological group is minimal. For more information about minimality, we refer the reader to the survey by Dikranjan and Megrelishvili [2014].

Given a group G of permutations of some set Ω and $A \subseteq \Omega$, let

$$G_A = \{g \in G \mid ga = a \text{ for all } a \in A\}.$$

Let $[\Omega]^{<\omega}$ be the set of all finite subsets of Ω. The collection $\{G_A \mid A \in [\Omega]^{<\omega}\}$ is a base of neighbourhoods at the identity of a group topology which we call the *standard topology* and denote by τ_{st}. More generally for each G-invariant $X \subseteq \Omega$ there is an associated group topology τ_{st}^X generated by $\{G_A \mid A \in [X]^{<\omega}\}$.

One of the earliest results on minimality due to Gaughan [1967] states that $(S_\infty, \tau_{\mathrm{st}})$ is totally minimal, where S_∞ denotes the group of all permutations of a countable set Ω.

Given a countable first-order structure $\mathcal{M}$ with universe M, the automorphism group of $\mathcal{M}$ is a τ_{st}-closed subgroup of $S_\infty = S(M)$ and vice versa: any closed subgroup of $S(M)$ is the automorphism group of some countable structure on M. The interplay between the dynamical properties of $\mathrm{Aut}(\mathcal{M})$ and the logical and combinatorial properties of $\mathcal{M}$ has been widely studied in the literature, beginning with the characterization due to Engeler, Ryll-Nardzewski, Svenonius and others of oligomorphic subgroups of S_∞ as the automorphism groups of ω-categorical countable structures. Recall that an oligomorphic group is a closed subgroup of S_∞ whose diagonal action on M^n has finitely many orbits, for each $n \in \mathbb{N}$.

In this context τ_{st} is often referred to in the literature as the pointwise convergence topology.

In light of the above the following is thus a natural question, already asked in [Dikranjan and Megrelishvili 2014].

Problem 1. *Let $\mathcal{M}$ be a countable ω-categorical* (ω-*saturated*, *sufficiently nice*) *first-order structure and $G = \mathrm{Aut}(\mathcal{M})$. When is (G, τ_{st})* (*totally*) *minimal*?

A deep result in this direction appeared in recent work by Ben Yaacov and Tsankov [2016], where the authors show that automorphism groups of countable ω-categorical, stable continuous structures are totally minimal with respect to the pointwise convergence topology. This specializes to the result that the automorphism groups of classical ω-categorical stable structures are totally minimal with respect to τ_{st}.

Not all oligomorphic groups are minimal with respect to τ_{st}. As pointed out in [Ben Yaacov and Tsankov 2016], an example of this is $\mathrm{Aut}(\mathbb{Q}, <)$ (see Corollary C for a generalization). However even in those cases it is possible to formulate the following more general question:

Problem 2. *Let $\mathcal{M}$ be a countable ω-categorical* (*or sufficiently nice*) *first-order structure and $G = \mathrm{Aut}(\mathcal{M})$. Describe the lattice of all Hausdorff group topologies on G coarser than τ_{st}.*

This work was mainly motivated by [Ben Yaacov and Tsankov 2016] and is meant as a preliminary exploration of Problems 1 and 2 in the classical setting outside the stability constraint.

In its broadest lines the strategy followed by [Ben Yaacov and Tsankov 2016] goes back to [Uspenskij 2008], where the author shows that the isometry group of the Urysohn sphere is totally minimal with the pointwise convergence topology. Both proofs rely on the assumption that the group in question is Roelcke precompact and use a well-behaved independence relation among (small) subsets of the structure to endow the Roelcke precompletion of the group with a topological semigroup structure. Information on the topological quotients of the original group is then recovered from the latter via the functoriality of Roelcke compactification and the Ellis lemma. Recall that a topological group (G, τ) is *Roelcke precompact* if for any neighbourhood W of 1 there exists a finite $F \subset G$ such that $WFW = G$. For closed subgroups of S_∞ this is equivalent to being oligomorphic.

In contrast, our methods for obtaining (partial) minimality results are completely elementary. There are drawbacks to this lack of sophistication: for instance, we are not able to recover the result in [Ben Yaacov and Tsankov 2016] for classical structures. On the other hand we do not rely on assumptions of Roelcke precompactness (except for certain residual assumptions in some cases). Although we are not discussing metric structures or Urysohn spaces in this paper, we would like to mention that a refinement of the approached presented here has enabled us to answer in the positive the question about the minimality of the isometry group of the (unbounded) Urysohn space posed in [Uspenskij 2008].

Problems 1 and 2 could be also formulated for semigroup topologies on the endomorphism monoid of a countable relational structure. Some general techniques for characterising minimal and maximal semigroup topologies on the endomorphism monoid of a countable relational structure have been recently introduced in [Elliott et al. 2023].

Main results. Generally speaking, an *independence relation* is a ternary relation $\mathop{\smile\!\!\!\!\!\!|}$ defined on some collection of sets of elements of the structure such that $A \mathop{\smile\!\!\!\!\!\!|}_C B$ is meant to capture the intuitive idea that B does not contain any information about A not already contained in C. The paradigmatic example is that of forking independence in model theory. The study of the connections between the existence of a well-behaved independence relation on a homogeneous structure (see Definition 2.1) and the properties of the automorphism group goes back to [Tent and Ziegler 2013] (see also [Evans et al. 2016]).

We provide a simple technical criterion (Proposition 2.12) for (relative) minimality for τ_{st} in a relatively general setting. We derive from this general minimality results stated in terms of the existence of an independence relation satisfying certain axioms and in turn derive from this two main theorems. The first applies to Fraïssé limits of free amalgamation classes, i.e., Fraïssé classes closed under free amalgamation (more details in Section 3). Some well-known examples of Fraïssé limits of free amalgamation classes are the random graph, random hypergraph, homogeneous K_n-free graphs for $n \geqslant 3$, etc.

Theorem A. *Let $\mathcal{M}$ be the Fraïssé limit of a free amalgamation class in a countable relational language. Let $G = \mathrm{Aut}(\mathcal{M})$. Then any group topology $\tau \subseteq \tau_{\mathrm{st}}$ on G is of the form τ_{st}^X, where $X \subseteq M$ is some G-invariant set. In particular, if the action of G on M is transitive, then there are no nontrivial group topologies on G strictly coarser than τ_{st} and thus (G, τ_{st}) is totally minimal.*

Rather than the free amalgamation property directly, the proof of Theorem A uses the freedom axiom, a more abstract property introduced in [Conant 2017].

The second application of the Proposition 2.12 is in the context of *simple* theories. Simple structures (i.e., theories) occupy an important place in classification theory. We refer the reader to [Tent and Ziegler 2012], [Wagner 2000] and [Kim 2014] for the definition of simple theories, forking and canonical bases.

A simple theory T is called *one-based* if $\mathrm{Cb}(a/A) \subseteq \mathrm{bdd}(a)$ for any hyperimaginary element a and a small subset A of the monster model. Our second main result is the following:

Theorem B. *Let $\mathcal{M}$ be a simple, ω-saturated countable structure with locally finite algebraic closure and weak elimination of imaginaries. Assume furthermore that* $\mathrm{Th}(\mathcal{M})$ *is one-based. Let $G = \mathrm{Aut}(\mathcal{M})$. Then:*

(1) *If G acts transitively on M, then (G, τ_{st}) is minimal.*

(2) *If all singletons are algebraically closed, then any group topology τ on G coarser than τ_{st} is of the form τ_{st}^X for some G-invariant $X \subseteq M$.*

Technically speaking, the use of the freedom axiom and stationarity in Theorem A is replaced in Theorem B by that of one-basedness and the independence property for forking independence in simple theories.

One important class of structures that fall under the assumptions of Theorem B are Lie geometries and their affine spaces as described in [Cherlin and Hrushovski 2003] and [Kantor et al. 1989]. Another class of examples of structures to which Theorem B applies can be obtained using the general techniques in [Chatzidakis and Pillay 1998].

Finally, we present a natural variant of ideas of [Uspenskij 2008] and [Ben Yaacov and Tsankov 2016] in the context of automorphism groups of first-order structures.

Given a structure $\mathcal{M}$ with group of automorphisms G, we describe a semigroup of partial types $R^{\mathrm{pa}}(\mathcal{M})$ containing G consisting of partial infinitary types encoding the relationships between two copies of $\mathcal{M}$, and show that any idempotent in $R^{\mathrm{pa}}(\mathcal{M})$ which is invariant under the involution given by exchanging the blocks of coordinates corresponding to the two models and the action of G can be associated to a group topology on G coarser than τ_{st}.

We show that under certain mild conditions, the topology τ_{st} on the automorphism group of any distal Fraïssé limit is not minimal.

Corollary C. *Let $\mathcal{M}$ be any distal Fraïssé limit in a finite relational language with trivial algebraic closure. Then the type q_{inf} defines a group topology on $G = \mathrm{Aut}(\mathcal{M})$ strictly coarser than τ_{st}.*

Layout. The paper is organized as follows. In Section 2 we prove our main technical criterion, Proposition 2.12, of (relative) minimality for τ_{st}.

Section 3 contains some preliminary discussion on independence relations and Fraïssé constructions, along with the proofs of Theorems A and B. In Section 3D we have provided an example where we show total minimality is not preserved under taking open finite-index subgroups. Finally in Section 3E we have shown that τ_{st} in certain simple ω-categorical structures built using the Hrushovski construction method are minimal (Corollary 3.11). Structures that are built using this method and predimension functions are not one-based.

Section 4 is dedicated to the systematic connection between group topologies below the standard topology and types described above as well as the proof of Corollary C.

2. A relative minimality criterion for τ_{st}

Given a topological group (G, τ) and $g \in G$ we denote by $\mathcal{N}_\tau(g)$ the filter of (not necessarily open) neighbourhoods of g in τ. Since $\mathcal{N}_\tau(g) = g\mathcal{N}_\tau(1_G) = \mathcal{N}_\tau(1_G)g$ for any $g \in G$, any group topology τ is uniquely determined by $\mathcal{N}_\tau(1_G)$. Given a filter $\mathcal{V}$ on G at 1_G such that

- for every $U \in \mathcal{V}$ there is $V \in \mathcal{V}$ such that $V^{-1} \subseteq U$,
- for every $U \in \mathcal{V}$ there is $V \in \mathcal{V}$ such that $VV \subseteq U$, and
- $U^g \in \mathcal{V}$ for every $U \in \mathcal{V}$ and $g \in G$,

there is a unique group topology τ on G such that $\mathcal{V} = \mathcal{N}_\tau(1_G)$. Given a family $\mathcal{Y}$ of subsets of G containing 1_G, we say that $\mathcal{Y}$ generates a group topology τ at the identity if $\mathcal{Y}$ generates $\mathcal{N}_\tau(1_G)$ as a filter.

Given a set X we let $[X]^{<\omega}$ stand for the collection of all finite subsets of X. Our setting consists of an infinite set Ω and some $G \leqslant S(\Omega)$, where $S(\Omega)$ is the group of permutations of Ω. It is easy to see using the criterion above that the collection $\{G_A \mid A \in [\Omega]^{<\omega}\}$ is a base of neighbourhoods of the identity of a unique group topology τ_{st}, which we will refer to as the standard topology. We are mainly interested in the case in which Ω is countable, in which case $S(\Omega)$, abbreviated as S_∞, is a Polish group.

By a closure operator on $[\Omega]^{<\omega}$ we mean a map $\mathrm{cl} : [\Omega]^{<\omega} \to [\Omega]^{<\omega}$ that preserves inclusion and satisfies $A \subseteq \mathrm{cl}(A) = \mathrm{cl}(\mathrm{cl}(A))$, for each $A \in [\Omega]^{<\omega}$. There is a bijective correspondence between (G-equivariant) closure operators cl and (G-invariant) families $\mathfrak{X} \subseteq [\Omega]^{<\omega}$ closed under intersections. Each $\mathfrak{X}$ gives a closure operator $\mathrm{cl}(-)$ by taking as $\mathrm{cl}(A)$ for any finite A the smallest set in $\mathfrak{X}$ containing A. In the opposite direction we associate cl with the class of cl-closed sets: $\mathfrak{X} = \{A \in [\Omega]^{<\omega} \mid \mathrm{cl}(A) = A\}$.

Given a family $\mathfrak{X}$ of subsets of a set Ω, denote by $(\mathfrak{X})$ the collection of all (finite) tuples of elements whose coordinates enumerate some member of $\mathfrak{X}$. As is customary, the same letter will be used to refer to either a tuple or the corresponding set depending on the context. In particular we might use an expression such as BC to denote the union of the ranges of B and C.

Given tuples A, B, C of elements from Ω we write $A \cong^G B$ if there exists some $g \in G$ such that $gA = B$ and given an additional C we write $A \cong^G_C B$ if there is $g \in G_C$ such that $gA = B$. Given $A \subset \Omega$ we let $\mathrm{acl}^G(A)$ stand for the union of all elements of Ω whose orbit under G_{A_0} is finite for some finite subset A_0 of A. We say $\mathrm{acl}^G(-)$ is *locally finite* if $\mathrm{acl}^G(A)$ is finite whenever A is. In that case the restriction of acl^G to $[\Omega]^{<\omega}$ is a closure operator on $[\Omega]^{<\omega}$. We write $\mathfrak{X}^G = \{A \in [\Omega]^{<\omega} \mid \mathrm{acl}^G(A) = A\}$ and we say that acl^G is *trivial* if $\mathfrak{X}^G = [\Omega]^{<\omega}$.

Definition 2.1. Let $\mathcal{M}$ be a structure with universe M.

- The structure $\mathcal{M}$ is called *homogeneous* if for every $A, B \subseteq M$ such that $|A| = |B| < |M|$ and $\mathrm{tp}(A) = \mathrm{tp}(B)$ there is an automorphism of $\mathcal{M}$ which sends A to B.
- The structure $\mathcal{M}$ is called *ω-saturated* if for every $A \in [M]^{<\omega}$ any type over A is realised in $\mathcal{M}$.
- A relational structure $\mathcal{M}$ is called *ultrahomogeneous* if any isomorphism between finite substructures of $\mathcal{M}$ extends to an automorphism of $\mathcal{M}$.

Let G be the group of automorphisms of some structure $\mathcal{M}$ with universe M. Recall that if $\mathcal{M}$ is countable and ω-saturated, then for finite A we have that $\mathrm{acl}^G(A)$ coincides with the algebraic closure of A. If $\mathcal{M}$ is ω-saturated and countable, then it is homogeneous. In particular a relational structure $\mathcal{M}$ if ω-saturated and countable, then it is ultrahomogeneous. Typical examples of countable ultrahomogeneous structures are structures obtained from the Fraïssé construction method in a relational language (see Section 3B).

The proof of the following statements contains two auxiliary observations. As usual in such cases we mark the end of the proof of the subordinate results with a shaded (as opposed to white) square.

Proposition 2.2. *Let G be a group of permutations of a set Ω for which* $\mathrm{acl}^G(-)$ *is locally finite. Suppose we are given some G-invariant $X \subseteq \Omega$ and another group topology $\tau^* \subset \tau_{\mathrm{st}}^X$ such that for some constant $K \in \mathbb{N}$ the following property holds*:

($\diamond$) *For any $A, B \in \mathfrak{X}^G$ and $U \in \mathcal{N}_{\tau^*}(1_G)$ there exists $U' \in \mathcal{N}_{\tau^*}(1_G)$ such that $((G_A \cap U)G_B)^K = G_{A\cap B} \cap U'$.*

Then any group topology $\tau \subseteq \tau_{\mathrm{st}}^X$ must satisfy at least one of the following two conditions:

(1) *Given $x \in X$ there exists $W \in \mathcal{N}_\tau(1_G)$ such that $gx \in \mathrm{acl}^G(x)$ for each $g \in W$.*

(2) *There exists some G-invariant $X' \subsetneq X$ such that for all $W \in \mathcal{N}_\tau(1_G)$ there is $U' \in \mathcal{N}_{\tau^*}(1_G)$ and $U'' \in \mathcal{N}_{\tau_{\mathrm{st}}^{X'}}(1_G)$ such that $U' \cap U'' \subseteq W$.*

Proof. Assume the first alternative does not hold. Then there is $x_0 \in X$ such that for any $W \in \mathcal{N}_\tau(1_G)$ there exists $g \in W$ such that $g(x_0) \notin \mathrm{acl}^G(x_0)$. Let $X' = X \setminus G\cdot x_0$. Our goal is to show point (2), that is, that any neighbourhood W of 1_G in τ is also a neighbourhood of the identity in any topology containing τ^* and $\tau_{\mathrm{st}}^{X'}$. We prove this via two observations.

Observation 2.3. *For any $a \in G\cdot x_0$, any finite $B \subset \Omega$ and any $W \in \mathcal{N}_\tau(1_G)$ there exists some $g \in W$ such that $ga \notin B$.*

Proof. Suppose the condition above fails for some a, B, and W. By Neumann's lemma there exists some $h \in G_a$ such that $h(B) \cap B \subseteq \mathrm{acl}^G(a)$. This means that any g in $W \cap W^{h^{-1}} \in \mathcal{N}_\tau(1_G)$ must take a to a point in $\mathrm{acl}^G(a)$. This contradicts the choice of x_0 and the fact that any $a \in G\cdot x_0$ must have the same property, by invariance of $\mathcal{N}_\tau(1_G)$ under conjugation. ■

The following observation follows from ($\diamond$) by an induction argument and we leave the proof to the reader.

Observation 2.4. *There is a function $\mu : \mathbb{N} \to \mathbb{N}$ such that given any finite collection $\{B_j\}_{j=1}^r \subset \mathfrak{X}^G$, $U \in \mathcal{N}_{\tau^*}(1_G)$ and $W \subseteq G$ containing $U \cap \bigcup_{j=1}^r G_{B_j}$ there exists $U' \in \mathcal{N}_{\tau^*}(1_G)$ such that $G_{\bigcap_{j=1}^r B_j} \cap U' \subseteq W^{\mu(r)}$.*

Fix some arbitrary $W \in \mathcal{N}_\tau(1_G)$. Pick $W_0 = W_0^{-1} \in \mathcal{N}_\tau(1_G)$ such that $W_0^{2K} \subseteq W$. Since $\tau \subseteq \tau_{\mathrm{st}}^X$, there exists some finite $A \subset X$ such that $G_A \subseteq W_0$. By local finiteness we may assume $A = \mathrm{acl}^G(A)$. Let $\{a_j\}_{j=1}^r := A \cap (G\cdot x_0)$.

Pick $W_1 = W_1^{-1} \in \mathcal{N}_\tau(1_G)$ such that $W_1^{3\mu(r)} \subseteq W_0$, where μ is the function given by Observation 2.4. Let $B \subset \Omega$ be a finite subset such that $G_B \subset W_1$. We may assume again $B \in \mathfrak{X}^G$. By Observation 2.3 for any $1 \leqslant j \leqslant r$ there exists some $g_j \in W_1$ such that $g_j a_j \notin B$ or, equivalently, $a_j \notin B_j := g_j^{-1} B$. Notice that $G_{B_j} = G_B^{g_j} \subseteq W_1^3$.

Let $C = \bigcap_{j=1}^r B_j$. According to Observation 2.4 (for $U = G$) there is $U' \in \mathcal{N}_{\tau^*}(1_G)$ such that $G_C \cap U' \subset (W_1^3)^{\mu(r)} \subseteq W_0$. A final direct application of ($\diamond$) (again with $U = G$) yields some $U_0' \in \mathcal{N}_{\tau^*}(1_G)$ such that

$$U_0' \cap G_{C \cap A} \subseteq (G_C G_A)^K \subseteq W_0^{2K} \subseteq W.$$

By construction $C \cap A \subseteq X'$ and thus $U_0'' := G_{A \cap C} \in \tau_{\mathrm{st}}^{X'}$. As W is an arbitrary neighbourhood of 1_G in τ we conclude that case (2) of the statement holds and so we are done. □

We elaborate further on the same idea:

Lemma 2.5. *Let G be a group of permutations of a set Ω, $\{X_j\}_{j \in J}$ some collection of G-invariant subsets of Ω and $Z = \bigcap_{j \in J} X_j$. Assume that $\mathrm{acl}^G(x) = x$ for any $x \in \Omega$ and that there exists $K > 0$ such that for any finite $A, B \subset \Omega$ we have $(G_A G_B)^K = G_{A \cap B}$. Then $\tau_{\mathrm{st}}^Z = \bigcap_{j \in J} \tau_{\mathrm{st}}^{X_j}$.*

Proof. We begin by noting that just as in Observation 2.4 one can show by induction:

Observation 2.6. *There exists a function $\mu : \mathbb{N} \to \mathbb{N}$ such that for any finite collection $\{B_l\}_{l=1}^r \subseteq [\Omega]^{<\omega}$ and any $V \subseteq G$ containing G_{B_l} for all $1 \leqslant l \leqslant r$ we have $G_{\bigcap_{l=1}^r B_l} \subseteq V^{\mu(r)}$.*

Let $\tau_0 = \bigcap_{j \in J} \tau_{\mathrm{st}}^{X_j}$. The inclusion $\tau_{\mathrm{st}}^Z \subseteq \tau_0$ is clear. Take now any $W \in \mathcal{N}_{\tau_0}(1_G)$. Fix $j_0 \in J$. Since $W \in \tau_{\mathrm{st}}^{X_{j_0}}$, there exists some finite $A \subseteq X_{j_0}$ such that $G_A \subseteq W$. Let $\{a_j\}_{j=1}^r := A \setminus Z$. Pick $W_0 = W_0^{-1} \in \mathcal{N}_{\tau_0}(1_G)$ such that $W_0^{\mu(r+1)} \subseteq W$. For each $1 \leqslant l \leqslant r$ choose some $j_l \in J$ such that $a_l \notin X_{j_l}$ and then some finite $B_l \subseteq X_{j_l}$ such that $G_{B_l} \subseteq W_0$. Observation 2.6 and the choice of W_0 implies $G_C \subseteq W$, where $C = A \cap \bigcap_{l=1}^r B_l$. Since $C \subseteq Z$ we have shown $U \subseteq W$ for some $U \in \tau_{\mathrm{st}}^Z$. Since $W \in \mathcal{N}_{\tau_0}(1_G)$ was arbitrary we have $\tau_0 \subseteq \tau_{\mathrm{st}}^Z$ and we are done. □

Lemma 2.7. *Let G be the automorphism group of some structure $\mathcal{M}$ endowed with a G-invariant locally finite closure operator $\mathrm{cl}(-)$ on M and a group topology τ coarser than τ_{st}. Assume that the action of G is transitive and there is some $W \in \mathcal{N}_\tau(1_G)$ and $a \in M$ such that $ga \in \mathrm{cl}(a)$, for each $g \in W$. Then either τ is not Hausdorff or $\tau = \tau_{\mathrm{st}}$.*

Proof. Notice that by the transitivity of the action of G on M and continuity of the inverse operation for every $a \in M$ there are $U_a, W_a \in \mathcal{N}_\tau(1_G)$ such that $f(a) \in \mathrm{cl}(a)$ for any $f \in W_a$ and $g^{-1}(a) \in \mathrm{cl}(a)$ for any $g \in U_a$. For a finite tuple A in M we write $W_A = \bigcap_{a \in A} W_a$. Given $a, b \in M$, we say that $a \sim b$ if $a \in \mathrm{cl}(b)$ and $b \in \mathrm{cl}(a)$. This is clearly an equivalence relation. If we let $W_a' = W_a \cap \bigcap_{z \in \mathrm{cl}(a)} U_z$, then any $f \in W_a'$ must preserve the class $[a] \in M/{\sim}$ setwise, that is, $W_a' \subset G_{[a]}$. Indeed, if $g \in W'_a$, then $ga \in \mathrm{cl}(a)$. On the other hand, since $g \in U_{ga}$ we must have $a = g^{-1}ga \in \mathrm{cl}(ga)$, so $a \sim ga$.

For any $V \in \mathcal{N}_\tau(1_G)$ and any finite $\sim$-closed $A \subset M$ consider the set

$$Y_V^A = \{f : A \to A \mid \exists g \in V \text{ such that } g\restriction_A = f \text{ and } g([a]) = [a] \text{ for all } a \in A\}.$$

Notice that this set is finite, and that given $\sim$-closed $A \subset B \subset M$ and $f \in Y_V^B$ we have $f\restriction_A \in Y_V^A$. Invariance should be clear from the fact that A is $\sim$-closed and the definition of Y_V^A.

Lemma 2.8. *Either $Y_V^A = \{\mathrm{id}_A\}$ for some $V \in \mathcal{N}_\tau(1_G)$ and finite $\sim$-closed A or there exists $f \in G \setminus \{1_G\}$ such that for all $\sim$-closed $A \subset M$ and all $V \in \mathcal{N}_\tau(1_G)$ we have $f\restriction_A \in Y_V^A$.*

Proof. Recall that according to the assumption the closure is locally finite. If the first alternative is not the case, then from Observation 2.4 and König's lemma it follows that there is a function $f : M \to M$ such that $f\restriction_A \in Y_V^A$ for any $\sim$-closed A and $V \in \mathcal{N}_\tau(1_G)$. The fact that $f\restriction_A$ is a type-preserving bijection of A for any such A implies $f \in G$. ■

If the first possibility in Lemma 2.8 holds, then G_A contains $W'_A \cap V$ and is thus a neighbourhood of the identity in τ, which implies that $\tau = \tau_{\mathrm{st}}$. We claim that if the second possibility is satisfied the resulting $f \in G \setminus \{1_G\}$ satisfies $f \in \bigcap_{V \in \mathcal{N}_\tau(1_G)} V$, and therefore τ is not Hausdorff. Given any $V \in \mathcal{N}_\tau(1_G)$, the closure in τ of any $W \in \mathcal{N}_\tau(1_G) \cap \tau_{\mathrm{st}}$ satisfying $W = W^{-1}$ and $W^2 \subset V$ is itself contained in V. Indeed, if $h \in \overline{W}$ then there is $h' \in hW \cap W$ and thus $h = h'(h'')^{-1} \in W^2$ for some $h'' \in W$. Hence, $\mathcal{N}_\tau(1_G)$ admits a basis consisting entirely of τ_{st}-closed neighbourhoods of the identity. It is thus enough to show that f belongs to the closure of V in τ_{st} for any $V \in \mathcal{N}_\tau(1_G)$, which is immediate from the definition of Y_V^A. □

The following ubiquitous observation is crucial for the application of the results above. We provide a proof for the sake of completeness.

Lemma 2.9. *Let G be a group of permutations of a set Ω and A, B tuples of elements from Ω for which there is a chain $A = A_0, B_0, \ldots, B_{n-1}, A_n = g(A)$ such that $A_iB_i \cong^G A_{i+1}B_i \cong^G AB$ for $0 \leqslant i < n$. Then $g \in (G_AG_B)^nG_A$.*

Proof. The proof is by induction on n. In the base case $n = 0$ we have $A = g(A)$, that is, $g \in G_A$. Assume now $n > 0$. Since $AB_0 \cong^G AB$, there exists $h \in G_A$ such that $h(B_0) = B$. Now $A_1B_0 \cong^G AB$ implies $h(A_1)B = h(A_1)h(B_0) \cong^G A_1B_0 \cong^G AB$, which implies that there exists $h' \in G_B$ such that $h'(h(A_1)) = A$. Applying induction to the sequence $(A'_i, B'_i)_{i=0}^{n-1}$ given by $A'_i = h'h(A_{i+1})$, $B'_i = h'h(B_{i+1})$ yields that $h'hg \in (G_AG_B)^{n-1}G_A$, from which it follows that $g \in (G_AG_B)^nG_A$, as desired. □

Definition 2.10. Suppose we are given a group G of permutations of a set Ω, and $\mathfrak{X}$ a G-invariant family of subsets of Ω closed under intersection. We say $\mathfrak{X}$ has the *n-zigzag* property (with respect to the action of G) if for every $A, B \in (\mathfrak{X})$ and any A' with $A \cong^G_{A\cap B} A'$ there are $A_0, \dots, A_n$ and $B_0, \dots, B_{n-1}$ such that

(1) $A_0 := A$, and $A_n = A'$;

(2) $A_i B_i \cong^G A_{i+1} B_i \cong^G AB$ for $0 \leqslant i \leqslant n-1$.

We will refer to the sequence $A_0, B_0, A_1, \dots, A_n$ above as an *(n, B)-zigzag path* from A to A'.

Observation 2.11. *Given an n-zigzag path as above if we write $C = A \cap B$ then $C \subseteq A_i B_i \cong^G_C A_{i+1} B_i \cong^G_C AB$ for all $0 \leqslant i \leqslant n-1$. In particular, $A_i \cap B_i = A_{i+1} \cap B_i = C$.*

Notice that for fixed A, B and n, the existence of a (n, B)-zigzag path from A to A' depends only on the orbit of A' under G_A.

Proposition 2.12. *Suppose $\mathcal{M}$ is a countable first-order structure and $G = \mathrm{Aut}(\mathcal{M})$. Assume $\mathrm{acl}^G(-)$ is locally finite and $\mathfrak{X}^G$ corresponding to acl^G has the n-zigzag property for some n. Then*:

(1) *If the action of G on M is transitive*, *then (G, τ_{st}) is minimal.*

(2) *If $\mathrm{acl}^G(x) = x$ for any $x \in M$, then any group topology $\tau \subseteq \tau_{\mathrm{st}}$ is of the form τ^X_{st} for some G-invariant $X \subseteq M$.*

Proof. For any $A, B \in \mathfrak{X}$ and any $g \in G_{A\cap B}$ the n-zigzag property applied to A, B and $A' = gA$, together with Lemma 2.9, implies that $g \in (G_A G_B)^n G_A$. Therefore $G_{A\cap B} = (G_A G_B)^n G_A$ and we can apply Proposition 2.2 with $\tau^* = \{\varnothing, G\}$ under the common assumptions of (1) and (2). By the same reason we can also apply Lemma 2.5 under the assumptions of (2).

Let us show (1) first. Let τ be a group topology on G coarser than τ_{st}. If the first alternative in Proposition 2.2 holds, then by Lemma 2.7 either τ is not Hausdorff or $\tau = \tau_{\mathrm{st}}$. Since by assumption the only invariant subsets of M are $\varnothing$ and M, the second alternative implies that $\tau = \{\varnothing, G\}$.

Let us now show (2). Let τ be a group topology on G coarser than τ_{st}. By Lemma 2.5 (see the discussion in the first paragraph) there exists some unique minimal G-invariant set X such that $\tau \subseteq \tau^X_{\mathrm{st}}$. Apply Proposition 2.2 with $\tau^* = \{\varnothing, G\}$. The second alternative produces some G-invariant $X' \subsetneq X$ such that $\tau \subseteq \tau^{X'}_{\mathrm{st}}$, in contradiction with the choice of X. Since we assume acl^G to be trivial, the first alternative implies $\tau = \tau^X_{\mathrm{st}}$. □

3. Minimality and independence

3A. ***Independence.*** Throughout this section we work in the following setting: Ω is a set, G is a permutation group of Ω, $\mathrm{cl}(-)$ a G-equivariant closure operator on $[\Omega]^{<\omega}$ and $\mathfrak{X} = \{\mathrm{cl}(A) \mid A \in [\Omega]^{<\omega}\}$ the associated family of closed sets. Our goal is to derive concrete applications from the results of the previous section to the case where Ω is the underlying set of a first-order structure $\mathcal{M}$ and $G = \mathrm{Aut}(\mathcal{M})$.

Definition 3.1. Given $\mathrm{cl}(-)$ and $\mathfrak{X}$ as above and a ternary relation $\mathop{\smile\!\!\!\!\!\!\mid}$ between members of $[\Omega]^{<\omega}$ we say that $(\mathrm{cl}, \mathop{\smile\!\!\!\!\!\!\mid})$ (alternatively, $(\mathfrak{X}, \mathop{\smile\!\!\!\!\!\!\mid})$) is a *compatible pair* if for all $A, B, C, D \in [\Omega]^{<\omega}$ the following properties are satisfied:

- (compatibility) $A \mathop{\smile\!\!\!\!\!\!\mid}_C B$ if and only if $A \mathop{\smile\!\!\!\!\!\!\mid}_{\mathrm{cl}(C)} B$ if and only if $\mathrm{cl}(AC) \mathop{\smile\!\!\!\!\!\!\mid}_C \mathrm{cl}(BC)$.
- (invariance) If $g \in G$ and $A \mathop{\smile\!\!\!\!\!\!\mid}_B C$ then $gA \mathop{\smile\!\!\!\!\!\!\mid}_{gB} gC$.
- (weak monotonicity) If $A \mathop{\smile\!\!\!\!\!\!\mid}_B CD$ or $AD \mathop{\smile\!\!\!\!\!\!\mid}_B C$ then $A \mathop{\smile\!\!\!\!\!\!\mid}_B C$.
- (antireflexivity) If $A \mathop{\smile\!\!\!\!\!\!\mid}_C B$, then $A \cap B \subseteq \mathrm{cl}(C)$.

We write $A \mathop{\smile\!\!\!\!\!\!\mid} B$ as an abbreviation of $A \mathop{\smile\!\!\!\!\!\!\mid}_\varnothing B$.

Definition 3.2. We define some additional properties for a compatible pair $(\mathfrak{X}, \mathop{\smile\!\!\!\!\!\!\mid})$:

- (transitivity) If $A \mathop{\smile\!\!\!\!\!\!\mid}_B C$ and $A \mathop{\smile\!\!\!\!\!\!\mid}_{BC} D$, then $A \mathop{\smile\!\!\!\!\!\!\mid}_B CD$.
- (symmetry) If $A \mathop{\smile\!\!\!\!\!\!\mid}_B C$ then $C \mathop{\smile\!\!\!\!\!\!\mid}_B A$.
- (existence) For any A, B, C there is $g \in G_B$ such that $gA \mathop{\smile\!\!\!\!\!\!\mid}_B C$.
- (independence) Suppose we are given $A, B_1, B_2, C_1, C_2 \in (\mathfrak{X})$ such that $B_1 \mathop{\smile\!\!\!\!\!\!\mid}_A B_2$, $A \subseteq B_i$ and $C_i \mathop{\smile\!\!\!\!\!\!\mid}_A B_i$ for $i = 1, 2$ and $C_1 \cong^G_A C_2$. Then there exists $D \in \mathfrak{X}$ such that $D \cong^G_{B_i} C_i$ for $i = 1, 2$ and $D \mathop{\smile\!\!\!\!\!\!\mid}_A B_1B_2$.
- (stationarity) If $B \in \mathfrak{X}$ and $A_i \mathop{\smile\!\!\!\!\!\!\mid}_B C$ for $i = 1, 2$, then $A_1 \cong^G_B A_2$ implies $A_1 \cong^G_{BC} A_2$.

We also consider these properties:

- (freedom) $\mathfrak{X} = [\Omega]^{<\omega}$ and if $A \mathop{\smile\!\!\!\!\!\!\mid}_C B$ and $C \cap AB \subseteq D \subseteq C$, then $A \mathop{\smile\!\!\!\!\!\!\mid}_D B$.
- (one-basedness) $A \mathop{\smile\!\!\!\!\!\!\mid}_{A\cap B} B$ for every $A, B \in \mathfrak{X}$.

The one-basedness property admits the following generalization:

Definition 3.3. Given $k \geqslant 1$, we say that $(\mathfrak{X}, \mathop{\smile\!\!\!\!\!\!\mid})$ satisfies *k-narrowness* if, for any $C, A_0, A_1, \ldots, A_k$ in $\mathfrak{X}$, the conditions

- $A_i \cap A_{i+1} = C$ for each $0 \leqslant i \leqslant k-1$,
- $A_{i+1} \mathop{\smile\!\!\!\!\!\!\mid}_{A_i} A_{i-1} \cdots A_0$ for each $1 \leqslant i \leqslant k-1$

imply that $A_0 \mathop{\smile\!\!\!\!\!\!\mid}_C A_k$ (notice that for $k = 1$ we recover the one-basedness property).

Lemma 3.4. *Let $(\mathfrak{X}, \mathop{\smile\!\!\!\!\!\!\mid})$ be a compatible pair that satisfies existence. Then*:

(1) *If it satisfies freedom or one-basedness, then for any $A, B \in \mathfrak{X}$ there is $A' \in \mathfrak{X}$ such that $A' \cong^G_B A$, $A' \cap A = A \cap B$ and $A \mathop{\smile\!\!\!\!\!\!\mid}_{A\cap B} A'$.*

(2) *If it satisfies transitivity, symmetry and* $2m$*-narrowness, then for any* $A, B \in \mathfrak{X}$ *there is* $A' \in \mathfrak{X}$ *such that an* (m, B)*-zigzag path from* A *to* A' *exists,* $A' \cap A = A' \cap B$ *and* $A \mathop{\smile\!\!\!\!|}_{A\cap B} A'$.

Proof. Existence yields $A' \in \mathfrak{X}$ such that $A' \cong^G_B A$ and $A' \mathop{\smile\!\!\!\!|}_B A$. Antireflexivity implies that $A' \cap A \subseteq B$, i.e., $A' \cap A \subseteq A \cap B$. On the other hand $A' \cong^G_B A$ implies $A \cap B = A' \cap B$.

If we assume the freedom axiom, then $A' \mathop{\smile\!\!\!\!|}_{A\cap B} A$ follows from $A' \mathop{\smile\!\!\!\!|}_B A$ and $B \cap (A' \cup A) = (B \cap A') \cup (B \cap A) = B \cap A$. Alternatively, the same conclusion follows directly from one-basedness.

Let $C = A \cap B$. For (2) construct sequences $B_0 = B, B_1, \dots, B_{m-1}$ and $A_0 = A, A_1, \dots, A_m$ as follows. Assuming we have already taken $(A_i, B_i)_{i=0}^k$, existence provides $A_{k+1} \cong^G_{B_k} A_k$ with $A_{k+1} \mathop{\smile\!\!\!\!|}_{B_k} A_0B_0 \cdots A_kB_k$. By the same token, for $k \leqslant m$ we can choose $B_{k+1} \cong^G_{A_{k+1}} B_k$ with $B_{k+1} \mathop{\smile\!\!\!\!|}_{A_k} A_0B_0 \cdots A_{k+1}$. It is clear that this yields an (m, B)-zigzag path from A to A_m.

By transitivity, $A_j \mathop{\smile\!\!\!\!|}_{B_{j-1}} A_l$ for any $0 \leqslant l \leqslant j-1$, so that $A_j \cap A_l \subseteq A_j \cap B_{j-1}$ by antireflexivity. Since $A_j \cap B_{j-1} = C$ and $C \subset A_j \cap A_l$ by Observation 2.11 we conclude that $A_j \cap A_l = C$. Arguing in a similar manner one can show that $A_j \cap B_l = C$ for any $0 \leqslant j \leqslant m$ and $0 \leqslant l \leqslant m-1$. This establishes that the sequence $A_0, B_0, \dots, B_{m-1}, A_m$ satisfies the first property of the condition in the definition of $2m$-narrowness, while the second follows by transitivity and construction. If we let $A' = A_m$ we then get $A' \mathop{\smile\!\!\!\!|}_C A$ and $A \mathop{\smile\!\!\!\!|}_C A'$ by symmetry, while the sequence above is an (m, B)-zigzag path from A to A'. □

Lemma 3.5. *Let* $(\mathfrak{X}, \mathop{\smile\!\!\!\!|})$ *be a compatible pair satisfying symmetry, existence and transitivity and assume that for any* $A, B \in \mathfrak{X}$ *there exists an* (m, B)*-zigzag path from* A *to some* A_1 *such that* $A_1 \mathop{\smile\!\!\!\!|}_{A\cap B} A$. *Then:*

(1) *If stationarity holds, then* $\mathfrak{X}$ *has the* $2m$*-zigzag property.*

(2) *If independence holds, then* $\mathfrak{X}$ *has the* $4m$*-zigzag property.*

Proof. Let $A, A', B \in \mathfrak{X}$ with $A' \cong^G_{A\cap B} A$. Let $C := A \cap B$. In both cases using the assumption we start by choosing $A_1 \in \mathfrak{X}$ for which there is an m-zigzag path from A to A_1 and $A_1 \mathop{\smile\!\!\!\!|}_C A$.

Consider (1) first. By extension there is A_2 such that $A_2 \cong^G_A A_1$ and $A_2 \mathop{\smile\!\!\!\!|}_A A'A$. The first implies the existence of an (m, B)-zigzag path from A to A_2. The second, together with $A_2 \mathop{\smile\!\!\!\!|}_C A$, implies $A_2 \mathop{\smile\!\!\!\!|}_C A'A$ by right transitivity. By weak monotonicity we get $A_2 \mathop{\smile\!\!\!\!|}_C A'$ and by symmetry $A \mathop{\smile\!\!\!\!|}_C A_2$ and $A' \mathop{\smile\!\!\!\!|}_C A_2$. Stationarity yields $A \cong^G_{A_2} A'$. Thus, there is also an (m, B')-zigzag path from A_2 to A', where $A'B' \cong^G AB$ and combining both paths we get a $(2m, B)$-zigzag path from A to A'.

We move on to case (2). By invariance and existence there is A_1' such that $A_1'A' \cong^G A_1A$ (so that by invariance $A_1' \mathop{\smile\!\!\!\!|}_C A'$) and $A_1' \mathop{\smile\!\!\!\!|}_{A'} A'A_1$. Transitivity and monotonicity then imply $A_1' \mathop{\smile\!\!\!\!|}_C A_1$.

Independence applied to the tuple C, A_1, A'_1, A, A' in place of the A, B_1, B_2, C_1, C_2 of the definition implies the existence of some D such that $DA_1 \cong^G AA_1$ and $DA'_1 \cong^G AA_1$. This witnesses the existence of a $(4m, B)$-zigzag path from A to A'. Notice that symmetry is required in order to get $A' \mathop{\smile\!\!\!\!\!\!|}_C A'_1$. □

3B. ***Review of Fraïssé construction.*** Let us briefly review the Fraïssé construction method in a relational language. For a more detailed discussion we refer the reader to the survey by Macpherson [2011].

Let $\mathfrak{L}$ be a relational signature and $\mathcal{K}$ be a countable class of finite $\mathfrak{L}$-structures closed under isomorphism. Suppose $A, B \in \mathcal{K}$. By $A \subseteq B$ we mean A is an $\mathfrak{L}$-substructure of B. We say $\mathcal{K}$ is a *Fraïssé class* if it satisfies the following properties:

- (HP) $\mathcal{K}$ is closed under substructures.
- (JEP) For any $A, B \in \mathcal{K}$ there is C in $\mathcal{K}$ such that $A, B \subseteq C$.
- (AP) Given $A_1, A_2, B \in \mathcal{K}$ and isometric embeddings $g_i : B \to A_i$ for $i = 1, 2$ there exists $C \in \mathcal{K}$ and isometric embeddings $h_i : A_i \to C$ such that $h_1 \circ g_1 = h_2 \circ g_2$.

According to a theorem of Fraïssé, for any Fraïssé class $\mathcal{K}$ there is a unique countable structure $\mathcal{M}$ called the *Fraïssé limit* of $\mathcal{K}$, denoted by $\mathrm{Flim}(\mathcal{K})$, such that

- $\mathcal{M}$ is ultrahomogeneous (see Definition 2.1);
- $\mathrm{Age}(\mathcal{M})$, the collection of all finite substructures of $\mathcal{M}$, coincides with $\mathcal{K}$.

Classical examples of Fraïssé limit structures are $(\mathbb{Q}, <)$ and the random graph. If $\mathfrak{L}$ is empty, then $\mathcal{K}$ is the class of finite sets and $\mathrm{Flim}(\mathcal{K})$ an infinite countable set. More generally, we say $\mathcal{K}$ is *trivial* if the equality type of a finite tuple of elements from M determines its type (equivalently, if $\mathrm{Aut}(\mathcal{M})$ is the full permutation group of M).

Suppose A, B and C are structures in some relational language $\mathfrak{L}$ with $A \subseteq B, C$. By the *free-amalgam* of B and C over A, denoted by $B \otimes_A C$, we mean the structure with domain $B \amalg_A C$ in which a relation holds for a tuple a if and only if it already did in either B or C.

By a *free amalgamation class* we mean a class $\mathcal{K}$ of finite structures in a relational language satisfying (HP) and such that $B \otimes_A C \in \mathcal{K}$ for any $A, B, C \in \mathcal{K}$ such that $A \subseteq B, C$. We write $B \mathop{\smile\!\!\!\!\!\!|}_A^{\mathrm{fr}} C$ if and only if the structure generated by ABC is isomorphic (with the right identifications) with the free amalgam $B \otimes_A C$. If $B \mathop{\smile\!\!\!\!\!\!|}^{\mathrm{fr}} C$ we say B and C are *free* from each other.

Theorem A. *Let $\mathcal{M}$ be the Fraïssé limit of a free amalgamation class in a countable relational language. Let $G = \mathrm{Aut}(\mathcal{M})$. Then any group topology $\tau \subseteq \tau_{\mathrm{st}}$ on G is of the form τ_{st}^X, where $X \subseteq M$ is some G-invariant set. In particular, if the action of G on M is transitive, then there are no nontrivial group topologies on G strictly coarser than τ_{st} and thus (G, τ_{st}) is totally minimal.*

Proof. First note that the algebraic closure in any Fraïssé limit of a free amalgamation class is trivial (follows from Lemma 2.1.4 in [Macpherson 2011]). If we let $\mathfrak{X} = [M]^{<\omega}$, where M is the underlying set of $\mathcal{M}$ and $\mathop{\smile\!\!\!\!\!\!\!\mid}\ = \mathop{\smile\!\!\!\!\!\!\!\mid}^{\mathrm{fr}}$, then part (1) of Lemma 3.4 and part (1) of Lemma 3.5 apply to the pair $(\mathfrak{X}, \mathop{\smile\!\!\!\!\!\!\!\mid})$. Together, they imply $\mathfrak{X}$ has the 2-zigzag property with respect to the action of G. The result then follows from an application of Proposition 2.12. □

3C. *Small, one-based simple theories.* Recall that given an $\mathfrak{L}$-structure $\mathcal{M}$ and $A \subseteq \mathcal{M}$, a subset X of $\mathcal{M}^n$ is called *definable over A* it is the solution set of some $\mathfrak{L}$-formula with parameters in A. For a model $\mathcal{M}$ of a complete theory and any definable equivalence relation E over $\varnothing$ on n-tuples one can consider the equivalence classes of $\mathcal{M}^n/E$ as elements of a new sort in an extended multisorted language. These classes are referred to as *imaginary* elements. A theory T is said to have *weak elimination of imaginaries* if for any $n \geqslant 1$ and any imaginary element $e = a/E$, where E is a definable equivalence relation on $\mathcal{M}^n$ over the empty set, there is a finite tuple c in $\mathcal{M}$ such that e is definable over c (i.e., the single solution of some formula over c) and c is algebraic over e (i.e., every element of c is a solution of some formula over e which has only finitely many solutions); see [Tent and Ziegler 2012]. Within a saturated model of the theory an element a is definable (algebraic) over B if its orbit under the stabilizer of B is a singleton (finite). Roughly speaking in theories with weak elimination of imaginaries, the imaginary elements are coded (in a weak sense) in the original structure.

Understanding simple theories requires dealing with hyperimaginaries. A *hyperimaginary* is an equivalence class of a type definable equivalence relation of a possibly infinite tuple over the empty set, where a type is an infinite conjunction of finitely consistent formulas. Recall that a theory *eliminates hyperimaginaries* if any hyperimaginary element is interdefinable with a sequence of imaginaries. See [Wagner 2000] or [Kim 2014] for details on these concepts.

Theorem B. *Let $\mathcal{M}$ be a simple, ω-saturated countable structure with locally finite algebraic closure and weak elimination of imaginaries. Assume furthermore that* $\mathrm{Th}(\mathcal{M})$ *is one-based. Let $G = \mathrm{Aut}(\mathcal{M})$. Then:*

(1) *If G acts transitively on M, then (G, τ_{st}) is minimal.*

(2) *If all singletons are algebraically closed, then any group topology τ on G coarser than τ_{st} is of the form τ_{st}^X for some G-invariant $X \subseteq M$.*

Proof. As cl we take the algebraic closure acl and $\mathop{\smile\!\!\!\!\!\!\!\mid}$ the forking independence. We claim part (1) of Lemma 3.4 and part (2) of Lemma 3.5 both apply to $(\mathfrak{X}, \mathop{\smile\!\!\!\!\!\!\!\mid})$.

The pair clearly satisfies invariance, weak monotonicity, transitivity and symmetry. Existence follows from the fact that M is ω-saturated, so it is left to check one-basedness and independence in the sense of Definition 3.2.

It is known that *small* simple theories which admit *finite coding* have elimination of hyperimaginaries (for definitions and details, see [Wagner 2000, Section 6 and Proposition 6.1.21]). Furthermore, one-based simple theories admit the finite coding property. These all imply in our setting that we have elimination of hyperimaginaries.

Take $A, B \in \mathfrak{X}$. The fact that the theory is one-based and has elimination of hyperimaginaries implies $A \perp_{\mathrm{acl}^{\mathrm{eq}}(A) \cap \mathrm{acl}^{\mathrm{eq}}(B)} B$. The relation $A \perp_{A \cap B} B$ follows then from weak elimination of imaginaries.

Lastly, elimination of hyperimaginaries and weak elimination of imaginaries imply that the type of a tuple over a finite acl-closed set determines its Lascar strong type over that same set. Hence, Kim and Pillay's independence theorem [1998] (see also Chapter 2.3 and Theorem 2.3.1 in [Kim 2014]) translates into abstract independence (amalgamation of types) for (acl, $\perp$) in that case. □

For stable theories the notion of being k-ample (for some $k \geqslant 1$) generalizes the negation of one-basedness. See [Evans 2003] for details. When algebraic closure is trivial, not k-ampleness translates into (acl, $\perp^f$) being k-narrow, where $\perp^f$ is the forking independence. From an argument similar to the one in the two theorems above we can deduce the following result:

Theorem 3.6. *Let $\mathcal{M}$ be a countable ω-saturated stable structure such that* $\mathrm{Th}(\mathcal{M})$ *has trivial algebraic closure, has weak elimination of imaginaries, and is not k-ample for some $k \geqslant 1$. Then any group topology on $G = \mathrm{Aut}(\mathcal{M})$ coarser than τ_{st} is of the form τ_{st}^X for some G-invariant $X \subseteq M$.*

3D. ***An example that shows total minimality is not preserved under taking open finite-index subgroups.*** Consider the relational language $\mathfrak{L}_1 = \{E^{(2)}, P^{(1)}\}$, and let $\mathcal{K}_1$ be the class of all finite $\mathfrak{L}_1$-structures in which E is interpreted as the edge relation of a bipartite graph with edges only between the domain of the unary predicate P and its complement. Consider also the class $\mathcal{K}_2$ in the language $\mathfrak{L}_2 = \{E^{(2)}, F^{(2)}\}$ consisting of all finite $\mathfrak{L}$-structures in which F is interpreted as an equivalence relation with at most 2 classes and E as the edge relation of a bipartite graph with edges only among vertices that belong to distinct F-classes.

Let $\mathcal{M}_i = \mathrm{Flim}(\mathcal{K}_i)$ and $G_i = \mathrm{Aut}(\mathcal{M}_i)$. Clearly $\mathcal{M}_2$ is a reduct of $\mathcal{M}_1$, so that $G_1 \lhd G_2$ and in fact $[G_2 : G_1] = 2$. It is easy to check that $\mathcal{K}_1$ has free amalgamation and then by Theorem A there are exactly two group topologies on G_1 strictly coarser than τ_{st}, namely $\tau_{\mathrm{st}}^{P(\mathcal{M}_1)}$ and $\tau_{\mathrm{st}}^{\neg P(\mathcal{M}_1)}$. Notice that both are Hausdorff, since no automorphism of $\mathcal{M}_1$ can fix $P(\mathcal{M}_1)$ or its complement (given any two points a, b, there exists c in P (resp. $\neg P$) such that $\mathrm{tp}(c, a) \neq \mathrm{tp}(c, b)$), so $(G_1, \tau_{\mathrm{st}})$ is not minimal.

In this case we have an additional non-Hausdorff group topology, $\tau^* = \{\varnothing, G_1\}$. Apply Proposition 2.2 to conclude that any group topology on G_1 strictly contained in τ_{st} is contained in τ^*.

On the other hand, it follows from Theorem B that $(G_2, \tau_{\mathrm{st}})$ is minimal.

3E. ***Simple nonmodular predimension Hrushovski construction.*** Hrushovski's predimension construction was introduced as a means of producing countable structures with a certain combinatorial property of the algebraic closure. This method was used by Hrushovski to build strongly minimal structures which are not field-like or vector space-like, as well as a stable ω-categorical pseudoplane. There are many variants of the method, but to fix notation, we consider the following basic case and later focus on a version that produces ω-categorical structures. We refer readers to [Wagner 1994; Baldwin and Shi 1996; Evans et al. 2016] for most of the properties that are mentioned here about Hrushovski constructions and some of their variations.

Suppose $s \geqslant 2$ and $\eta \in (0, 1]$. We work with the class $\mathcal{C}$ of finite s-uniform hypergraphs, that is, structures in a language with a single s-ary relation symbol $R(x_1, \dots, x_s)$ whose interpretation is invariant under permutation of coordinates and satisfies $R(x_1, \dots, x_s) \to \bigwedge_{i<j}(x_i \neq x_j)$.

To each $B \in \mathcal{C}$ we assign the predimension

$$\delta(B) = |B| - \eta|R[B]|,$$

where $R[B]$ denotes the set of hyperedges on B. For $A \subseteq B$, we define $A \leqslant B$ if and only if for all B' with $A \subseteq B' \subseteq B$ we have $\delta(A) \leqslant \delta(B')$, and let $\mathcal{C}_\eta := \{B \in \mathcal{C} \mid \varnothing \leqslant B\}$. The following is standard.

Lemma 3.7. *Suppose $A, B \subseteq C \in \mathcal{C}_\eta$. Then*:

(1) $\delta(AB) \leqslant \delta(A) + \delta(B) - \delta(A \cap B)$.

(2) *If $A \leqslant B$ and $X \subseteq B$, then $A \cap X \leqslant X$.*

(3) *If $A \leqslant B \leqslant C$, then $A \leqslant C$.*

If $A, B \subseteq C \in \mathcal{C}_\eta$ then we define $\delta(A/B) = \delta(AB) - \delta(B)$. Note that this is equal to $|A \setminus B| - \eta|R[AB] \setminus R[B]|$. Then $B \leqslant AB$ if and only if $\delta(A'/B) \geqslant 0$ for all $A' \subseteq A$. Moreover, if N is an infinite $\mathfrak{L}$-structure such that $A \subseteq N$, we write $A \leqslant N$ whenever $A \leqslant B$ for every finite substructure B of N that contains A. For $\mathfrak{L}$-structures A and X, where A is finite and X is of any cardinality, if $A \leqslant X$ then we say A is $\leqslant$-*closed* in X. One can show $\mathcal{C}_\eta$ has the $\leqslant$-free amalgamation property (see Lemma 4.8 in [Baldwin and Shi 1996]), by which we mean free amalgamation with respect to $\leqslant$ inclusions. An analogue of Fraïssé's theorem holds in this situation:

Proposition 3.8. *There is a unique countable structure $\mathcal{M}^\eta$, up to isomorphism, satisfying*:

(1) *The set of all finite substructures of $\mathcal{M}^\eta$, up to isomorphism, is precisely $\mathcal{C}_\eta$.*

(2) *$\mathcal{M}^\eta = \bigcup_{i \in \omega} A_i$, where $(A_i : i \in \omega)$ is a chain of $\leqslant$-closed finite sets.*

(3) *If $A \leqslant \mathcal{M}^\eta$ and $A \leqslant B \in \mathcal{C}_\eta$, then there is an embedding $f : B \to \mathcal{M}^\eta$ with $f\restriction_A = \mathrm{id}_A$ and $f(B) \leqslant \mathcal{M}^\eta$.*

The structure $\mathcal{M}^\eta$ that is obtained in the above proposition is called the *Hrushovski generic* structure.

Here we briefly discuss a variation on the Hrushovski's predimension construction method as a way to generate ω-categorical structures. The original version of this is used to provide a counterexample to Lachlan's conjecture, where it is used to construct a stable ω-categorical pseudoplane (see Section 5 in [Wagner 1994]). Here we follow a similar setting to that used in Section 5.2 of [Evans et al. 2016].

Suppose $\eta = m/n \in (0, 1]$, where $\gcd(m, n) = 1$. Consider the same setting of the previous subsection for $\mathfrak{L}$ and $\mathcal{C}_\eta$. For $A, B \in \mathcal{C}_\eta$, where $A \subset B$, define $A \leqslant_d B$ when $\delta(A'/A) > 0$, for all A' with $A \subsetneq A' \subseteq B$. For a suitable choice of an unbounded convex increasing function $f : \mathbb{R}^{\geqslant 0} \to \mathbb{R}^{\geqslant 0}$ and restricting $\mathcal{C}_\eta$ to

$$\mathcal{C}_\eta^f := \{A \in \mathcal{C}_\eta \mid \delta(X) \geqslant f(|X|) \text{ for all } X \subseteq A\},$$

one can show $(\mathcal{C}_\eta^f, \leqslant_d)$ has the $\leqslant_d$-free amalgamation property. We call these f *good* and denote the associated countable *generic structure* by $\mathcal{M}_\eta^f$, which is going to be ω-categorical.

Remark 3.9. To obtain a good function, we can take some piecewise smooth f whose right derivative f' satisfies $f'(x) \leqslant 1/x$ and is nonincreasing for $x \geqslant 1$. The latter condition implies that $f(x + y) \leqslant f(x) + yf'(x)$ (for $y \geqslant 0$). It can be shown that, under these conditions, $\mathcal{C}_\eta^f$ has the free $\leqslant_d$-amalgamation property. Details can be found in Section 6.2 and Example 6.2.27 in [Wagner 2000].

We assume that f is a good function. We will assume that $f(0) = 0$ and $f(1) > 0$, and in this case the $\leqslant$-closure of the empty set is empty. We shall also assume that $f(1) = n$ and one can show $\mathrm{Aut}(\mathcal{M}_\eta^f)$ acts transitively on M_η^f. See Examples 5.11 and 5.12 in Section 5.2 of [Evans et al. 2016] for details.

Given any finite subset X of $\mathcal{M}_\eta^f$, one can show there is a smallest finite subset Y with $X \subseteq Y \leqslant_d \mathcal{M}_\eta^f$, for which we use the notation $\mathrm{cl}^d(X)$. Let $\mathfrak{X}^d := \{\mathrm{cl}^d(A) \mid A \in [\mathcal{M}_\eta^f]^{<\omega}\}$. Given $A, B, C \in \mathfrak{X}^d$ one can define $A \mathop{\smile\!\!\!\!\!\mid}^d_B C$ if and only if $\mathrm{cl}^d(AB) \cup \mathrm{cl}^d(BC) = \mathrm{cl}^d(ABC)$ and $\mathrm{cl}^d(AB) \cap \mathrm{cl}^d(BC) = B$. Note that in this case, $\mathrm{cl}^d(ABC)$ is the free amalgam of $\mathrm{cl}^d(AB)$ and $\mathrm{cl}^d(BC)$ over B.

Lemma 3.10. $(\mathfrak{X}^d, \mathop{\smile\!\!\!\!\!\mid}^d)$ *satisfies* 3*-narrowness.*

Proof. Suppose C, A_0, A_1, A_2, A_3 are d-closed sets in $\mathfrak{X}^d$ with $A_i \cap A_{i+1} = C$ for $0 \leqslant i \leqslant 2$, where $A_3 \mathop{\smile\!\!\!\!\!\mid}^d_{A_2} A_1 A_0$ and $A_2 \mathop{\smile\!\!\!\!\!\mid}^d_{A_1} A_0$. We want to show $A_3 \mathop{\smile\!\!\!\!\!\mid}^d_C A_0$. First we claim $A_3 \cap A_0 = C$. By the assumption $C \subseteq A_0 \cap A_3$. From $A_3 \mathop{\smile\!\!\!\!\!\mid}^d_{A_2} A_1 A_0$ we know $\mathrm{cl}^d(A_3 A_2) \cap \mathrm{cl}^d(A_2 A_1 A_0) = A_2$, which implies $A_3 \cap A_0 \subseteq A_3 \cap A_2 = C$.

It remains to show $A_0 A_3$ is d-closed. If not, then there is $e \in \mathrm{cl}^d(A_0 A_3) \setminus A_0 A_3$ such that e is R-related to some elements in $\hat{A}_3 \subseteq A_3 \setminus C$ and to some elements in $\hat{A}_0 \subseteq A_0 \setminus C$, where $\delta(E/\hat{A}_0 \hat{A}_3 C) \leqslant 0$ for some $E \subseteq \mathrm{cl}^d(A_0 A_3)$, where $e \in E$ (see Section 4.2 in [Evans et al. 2016] for details of properties of minimally simply algebraic extensions). From $A_3 \mathop{\smile\kern-0.5em\raise0.3ex\hbox{$\mid$}}^d_{A_2} A_1A_0$ we know $\mathrm{cl}^d(A_0A_1A_2A_3)$ is the free amalgam of $\mathrm{cl}^d(A_3A_2)$ and $\mathrm{cl}^d(A_2A_1A_0)$ over A_2. Since $e \in \mathrm{cl}^d(A_0A_1A_2A_3)$, then this implies $e \in A_2$. Because $A_2 \mathop{\smile\kern-0.5em\raise0.3ex\hbox{$\mid$}}^d_{A_1} A_0$, we have $\hat{A}_0 \subseteq A_1$. This contradicts the fact that $A_0 \cap A_1 = C$. □

Then, combining Lemma 5.7 in [Evans et al. 2016] with Lemma 3.4(2), by Lemma 3.10, one can see $(\mathfrak{X}^d, \mathop{\smile\kern-0.5em\raise0.3ex\hbox{$\mid$}}^d)$ satisfies all the properties of Lemma 3.5(1). Then using Proposition 2.12 we conclude the following.

Corollary 3.11. *Suppose that f is a good function and let $\mathcal{M}_\eta^f$ be an ω-categorical Hrushovski generic structure such that $G = \mathrm{Aut}(\mathcal{M}_\eta^f)$ acts transitively on M_η^f. Then (G, τ_{st}) is a minimal topological group.*

4. Topologies and types

In this section we describe a general way of constructing group topologies below the standard topology on the automorphism group of a first-order structure. Our ideas are inspired by [Ben Yaacov and Tsankov 2016] and [Uspenskij 2008]. In fact, when $\mathcal{M}$ is an ω-categorical structure the space $R^{\mathrm{pa}}(\mathcal{M})$ as defined below consisting of complete types can be identified with the Roelcke compactification of $\mathrm{Aut}(\mathcal{M})$ as described in [Ben Yaacov and Tsankov 2016]. However the goal here is to establish a way of parametrizing topologies that does not depend on the existence of a well-behaved independence relation. We prove Corollary C at the end of the section as an application.

Let $\mathcal{M}$ be a first-order structure and $T = \mathrm{Th}(\mathcal{M})$. Consider two tuples of variables $x = (x_m)_{m\in M}$ and $y = (y_m)_{m\in M}$ indexed by the elements of M. Given some finite tuple $a = (a_1, a_2, \ldots, a_k) \subset M$ we write x_a in lieu of $(x_{a_1}, x_{a_2}, \ldots, x_{a_k})$. Let $p_M(x) = \mathrm{tp}(M)$, where the variable x_m is made to correspond with $m \in M$. Let $R(\mathcal{M})$ stand for the collection of all T-complete types in variables x, y containing $p_M(x) \cup p_M(y)$ and write $R^{\mathrm{pa}}(\mathcal{M})$ for the collection of partial types in variables x, y in T containing $p_M(x) \cup p_M(y)$. Here we assume types are deduction closed. Given any partial type $p(x, y)$ we will denote the deduction closure of $p(x, y) \cup p_M(x) \cup p_M(y)$ in T as $\langle p\rangle$. The set $R^{\mathrm{pa}}(\mathcal{M})$ can be endowed with the so-called *logic topology*, which we denote by τ_L, generated by neighbourhoods of the form $[\phi] = \{p \in R^{\mathrm{pa}}(\mathcal{M}) \mid \phi \in p\}$, where ϕ is any formula in (x, y). The result is a Stone space.

Given $p_1, p_2 \in R^{\mathrm{pa}}(\mathcal{M})$ we let $(p_1 * p_2)(x, y) \in R^{\mathrm{pa}}(\mathcal{M})$ denote the collection of all formulas $\psi(x, y)$ such that there exist $\phi_i(x, y) \in p_i(x, y)$ for $i = 1, 2$ such that

$$\phi_1(x, z) \wedge \phi_2(z, y) \vdash \psi(x, y).$$

Given $p \in R^{\mathrm{pa}}$, let $\bar{p} \in R^{\mathrm{pa}}$ be defined by $\theta(x, y) \in \bar{p} \leftrightarrow \theta(y, x) \in p$. It can be checked that $*$ endows $R^{\mathrm{pa}}(\mathcal{M})$ with a semigroup structure. Furthermore, one can show that $*$ is a continuous map $R^{\mathrm{pa}}(\mathcal{M}) \times R^{\mathrm{pa}}(\mathcal{M}) \to R^{\mathrm{pa}}(\mathcal{M})$ and $p \mapsto \bar{p}$ is also continuous with respect to τ_L. For the first, assume $p_1, p_2 \in R^{\mathrm{pa}}(\mathcal{M})$ and $\psi(x, y)$ is a formula with $p_1 * p_2 \in [\psi(x, y)]$. Then the definition of $*$, together with compactness, implies the existence of $\phi_1(x, z) \in p_1$ and $\phi_2(z, y) \in p_2$ such that $T \cup \{\phi_1(x, z), \phi_2(z, y)\} \vdash \psi(x, y)$, which implies that $[\phi_1] * [\phi_2] \subseteq [\psi]$. If we let $0 = \langle \varnothing \rangle \in R^{\mathrm{pa}}$ then clearly $p * 0 = 0$ for any $p \in R^{\mathrm{pa}}$. We write $p \leqslant q$ for $p \vdash q$.

Every $g \in \mathrm{Aut}(\mathcal{M})$ is associated to some type $\iota(g) = \langle x_{gm} = y_m \rangle_{m \in M} \in R^{\mathrm{pa}}$. It can be easily checked that ι is a continuous homomorphic embedding of (G, τ_{st}) into $(R^{\mathrm{pa}}(\mathcal{M}), \tau_L)$ whose image is contained in $R(\mathcal{M})$. We will write simply g instead of $\iota(g)$. Notice that $p^g := g^{-1} * p * g = \{\phi(x_a, y_b) \mid \phi(x_{g \cdot a}, y_{g \cdot b}) \in p\}$ for any $p \in R^{\mathrm{pa}}$ and $g \in G$.

Definition 4.1. Suppose $\mathcal{M}$ is an $\mathfrak{L}$-structure and $G = \mathrm{Aut}(\mathcal{M})$. We say that $q \in R^{\mathrm{pa}}$ is an *invariant idempotent* if the following conditions are satisfied:

(1) $1_G \leqslant q$;

(2) $q = \bar{q}$;

(3) $q * q = q$; and

(4) $q = q^g$ for any $g \in G$.

Notice that (1) implies $q = 1_G * q \leqslant q * q$, so that item (3) could be replaced by the a priori weaker condition $q * q \leqslant q$.

Given a formula $\phi(x, y)$, let $N_\phi := \iota^{-1}([\phi]) = \{g \in G \mid \mathcal{M} \models \phi(ga, b)\}$. Given an invariant idempotent $q \in R^{\mathrm{pa}}(\mathcal{M})$, let $\mathcal{N}_q = \{N_\phi \mid \phi(x, y) \in q\}$.

Lemma 4.2. *Given any structure $\mathcal{M}$ the following statements hold, where $G =$* $\mathrm{Aut}(\mathcal{M})$:

(1) *Given any invariant idempotent $q \in R^{\mathrm{pa}}(\mathcal{M})$ the family $\mathcal{N}_q$ forms a basis of neighbourhoods of a group topology τ_q on G (necessarily unique by invariance under translations).*

(2) *The closure of 1_G in τ_q coincides with the collection of all $g \in G$ such that $g \leqslant q$.*

(3) *Given invariant idempotents $p, q \in R^{\mathrm{pa}}(\mathcal{M})$ such that $p \leqslant q$ we have $\tau_p \supseteq \tau_q$. Conversely, if $\mathcal{M}$ is countable and ω-saturated then $\tau_p \supseteq \tau_q$ implies $p \leqslant q$.*

Proof. On the one hand, for any $\phi(x_A, y_B) \in q$, we have

$$\begin{aligned} N_{\phi(x,y)}^{-1} &= \{g \in G \mid \mathcal{M} \models \phi(g^{-1}a, b)\} \\ &= \{g \in G \mid \mathcal{M} \models \phi(a, gb)\} = N_{\phi(y,x)} \in \mathcal{N}_{\bar{q}} = \mathcal{N}_q. \end{aligned}$$

On the other hand, the condition $q * q = q$ is equivalent to the following: for any ϕ and finite A and B there is $C \subset M$ and formulas $\psi(x_A, z_C), \psi'(z_C, y_B) \in q$ such that modulo T we have

$$p_M(x) \cup p_M(y) \cup p_M(z) \cup \{\psi(x_A, z_C) \wedge \psi'(z_C, y_B)\} \vdash \phi(x_A, y_B). \tag{1}$$

Let $N = N_{\psi(x_A, y_C) \wedge \psi(x_C, y_B)}$. Given $h, g \in N$ we have $\mathcal{M} \models \psi(gA, C) \wedge \psi'(hC, B)$. The formulas are of course h-invariant, and hence $\mathcal{M} \models \psi(hgA, hC)$. Likewise, $hgA \models p_A$ and $hC \models p_C$ and thus by (1) we conclude that $\mathcal{M} \models \phi(hgA, B)$ and therefore $hg \in N_\phi$. This settles part (1). Part (2) follows easily from the fact that $\iota(g)$ is a complete type for $g \in G$ and is left to the reader. As for (3), the implication from left to right is trivial. Assume now $\mathcal{M}$ is countable and ω-saturated and we are given p, q such that $p \not\leq q$. Then there exists some $\phi(x_a, y_a) \in q$ for $a \in [M]^{<\omega}$ such that $p \notin [\phi]$.

Consider the type $r(x) \in S^{|a|}(a)$ given by

$$r(x) = \mathrm{tp}^x(a) \cup \{\neg\phi(x, a)\} \cup \{\psi(x, a) \mid \psi(x, y) \in p\}.$$

It follows from the discussion above that $r(x)$ is consistent and thus, by our assumption on $\mathcal{M}$, realized by some $a' \in M^{<\omega}$. Since $\mathcal{M}$ is homogeneous, there is $g \in G$ such that $a' = ga$. Since $\mathcal{M} \models \psi(ga, a)$, for each $\psi(x, y) \in p$ but $\mathcal{M} \models \neg\phi(ga, a)$ we conclude that $g \in N_\psi \setminus N_\phi$ for any $\psi \in p$ and thus that $\tau_p \not\supseteq \tau_q$. □

Remark 4.3. The element $1_G \in G$ seen as an element in R^{pa} is an invariant idempotent. The associated topology τ_{1_G} is just the standard topology. It can be checked by inspection that all topologies on automorphism groups that feature in this paper are of the form τ_q for some invariant idempotent q. In particular, any topology of the form τ_{st}^X for some $\mathrm{Aut}(\mathcal{M})$-invariant set X is of the form τ_p, where p is the type generated by all formulas of the form $x_a = y_a,\ a \in X$.

The following question arises naturally.

Question 3. *Let $\mathcal{M}$ be a countable w-categorical (homogeneous) structure. Is it true that any group topology on* $\mathrm{Aut}(\mathcal{M})$ *is of the form τ_q for some invariant idempotent $q \in R^{\mathrm{pa}}$?*

4A. ***Nonminimality in the trivial* acl *case.*** To conclude in this final subsection we show minimality fails for the automorphism groups of certain Fraïssé limits. Fix some structure $\mathcal{M}$ in a finite relational language in which acl is *trivial*, i.e., $\mathrm{acl}(A) = A$ for any finite $A \subset M$. Consider the type $q_{\inf} \in R^{\mathrm{pa}}(\mathcal{M})$ generated by all the formulas of the form $\phi(x_A, y_B)$, where $\phi \in \mathrm{tp}(A, B)$, for finite $A, B \subseteq M$ with $A \cap B = \varnothing$. Notice that $q_{\inf}$ is clearly invariant under the action of $\mathrm{Aut}(\mathcal{M})$ on x_M and y_M.

Definition 4.4. We say that $\mathcal{M}$ has the *separation property* if for any two disjoint finite tuples $a, b \in [M]^{<\omega}$ there exists $c \in [M]^{<\omega}$ disjoint from both a and b such that $\mathrm{tp}^{x,z}(a, c) \cup \mathrm{tp}^{z,y}(c, b) \vdash \mathrm{tp}^{x,y}(a, b)$.

Lemma 4.5. *Assuming* acl *is trivial in $\mathcal{M}$, the type q_{inf} is an invariant idempotent in $R^{\mathrm{pa}}(\mathcal{M})$ if and only if $\mathcal{M}$ has the separation property. If in addition to this $\mathcal{M}$ is countable and ω-saturated, then $q_{\mathrm{inf}} \not\leqslant 1_G$ and thus $\tau_{q_{\mathrm{inf}}}$ is strictly coarser than $\tau_{\mathrm{st}} = \tau_{1_G}$.*

Proof. Properties (1), (2) and (4) of Definition 4.1 are immediate from the definition of q_{inf}. For property (3) all we need to check is that $q * q \leqslant q$, as remarked after Definition 4.1, but this is precisely the content of the separation property, as in its definition, $\mathrm{tp}^{x,z}(a, c) \cup \mathrm{tp}^{z,y}(c, b) \vdash \mathrm{tp}^{x,y}(a, b)$, we have $\mathrm{tp}^{x,y}(a, c) \cup \mathrm{tp}^{x,y}(c, b) \subseteq q_{\mathrm{inf}}$ and thus $\mathrm{tp}^{x,y}(a, b) \subseteq q_{\mathrm{inf}} * q_{\mathrm{inf}}$ for the arbitrary fragment $\mathrm{tp}^{x,y}(a, b) \subseteq q_{\mathrm{inf}}$ we started with.

If $q_{\mathrm{inf}} = 1_G$, then for any $b \in M$ there must be some finite $A \subseteq M \setminus \{b\}$ such that $\mathrm{tp}^{x_A,y_b}(A, b) \vdash y_b = x_b$, which can only be the case if $b \in \mathrm{dcl}(A)$. The final claim then follows from last point of Lemma 4.2. Namely, from (3) of Lemma 4.2, if $1_G \leqslant q_{\mathrm{inf}}$ then $\tau_{1_G} = \tau_{\mathrm{st}} \supseteq \tau_{q_{\mathrm{inf}}}$. Using the second part of (3), if $\tau_{\mathrm{st}} = \tau_{q_{\mathrm{inf}}}$ then $q \leqslant 1_G$, which contradicts the fact that $q_{\mathrm{inf}} \not\leqslant 1_G$. □

Distal theories are a particular class of NIP theories introduced in [Simon 2013]. One main feature is the following fact [Chernikov and Simon 2015, Theorem 21]:

Fact 4.6. *Let T be distal. Then for any formula $\phi(x, y)$ there is a formula $\theta(x, z)$ such that for any $\mathrm{tp}^{\phi}(a/C)$ over a finite set of parameters C there is a tuple $d \subset C$ such that $\theta(a, d)$ holds, and $\theta(x, d) \vdash \mathrm{tp}^{\phi}(a/C)$, i.e., $\theta(x, y) \cup \mathrm{tp}^{y}(d, C) \vdash \mathrm{tp}^{\phi}(x/C)$, where $|y| = |d|$.*

Lemma 4.7. *Let $\mathcal{M}$ be any distal Fraïssé limit in a finite relational language with trivial algebraic closure. Then $\mathcal{M}$ has the separation property.*

Proof. Consider any two disjoint finite tuples $a, b \in M$. Since $\mathcal{M}$ has quantifier elimination, there exists some formula $\phi(x, y)$ such that for any $C \subset M$ the full type $\mathrm{tp}(a/C)$ is equivalent to the ϕ-type $\mathrm{tp}^{\phi}(a/C)$ ($|a| = |x|$). Let $\theta(x, z)$ be the formula provided by Fact 4.6 and let $s = |z|$. Take a sequence $b_{-s}, b_{-s+1}, \ldots, b_0 = b$, $b_1, \ldots, b_s$ of instances of $\mathrm{tp}(b/a)$ indiscernible over a, where b_i and b_j are disjoint for $i \neq j$. Let $C = b_{-s} b_{-s+1} \cdots b_s$, and let d be the tuple obtained from applying Fact 4.6 to $\mathrm{tp}(a/C)$. Let J be the set of indices $j \in \{-s, -s+1, \ldots, s\}$ such that $d \cap b_j \neq \varnothing$. Now, there must be some $j_0 \in \{-s, -s+1, \ldots, s\} \setminus J$ and some order-preserving bijection $\phi : J \cup \{j_0\} \to J' \subseteq \mathbb{Z}$ sending j_0 to 0. Since $(b_i)_i$ is indiscernible, the fact that $\mathrm{tp}(a/b_l)_{l \in J}$ isolates $\mathrm{tp}(a/b_l)_{l=-s}^{s}$ implies that $\mathrm{tp}(a/b_l)_{l \in J' \setminus \{0\}}$ isolates $\mathrm{tp}(a/b_l)_{l \in J'}$, so that the tuple $C = (b_l)_{l \in J' \setminus \{0\}}$ witnesses the separation property for the pair (a, b). □

Corollary C. *Let $\mathcal{M}$ be any distal Fraïssé limit in a finite relational language with trivial algebraic closure. Then the type q_{inf} defines a group topology on $G = \mathrm{Aut}(\mathcal{M})$ strictly coarser than τ_{st}.*

Many Fraïssé structures, such as nontrivial reducts of $(\mathbb{Q}, \leqslant)$ and ω-categorical finitely ramified ordered trees, satisfy the assumptions of Corollary C.

Acknowledgements

The authors would like to thank Dugald Macpherson for many helpful comments and discussions. We would also like to thank David M. Evans, Itay Kaplan and Todor Tsankov for the encouraging comments and thoughtful suggestions on early versions of this paper. We also appreciate the referee's thoughtful and detailed comments, which led to significant improvement of the exposition.

References

[Baldwin and Shi 1996] J. T. Baldwin and N. Shi, "Stable generic structures", *Ann. Pure Appl. Logic* **79**:1 (1996), 1–35. MR Zbl

[Ben Yaacov and Tsankov 2016] I. Ben Yaacov and T. Tsankov, "Weakly almost periodic functions, model-theoretic stability, and minimality of topological groups", *Trans. Amer. Math. Soc.* **368**:11 (2016), 8267–8294. MR Zbl

[Chatzidakis and Pillay 1998] Z. Chatzidakis and A. Pillay, "Generic structures and simple theories", *Ann. Pure Appl. Logic* **95**:1-3 (1998), 71–92. MR Zbl

[Cherlin and Hrushovski 2003] G. Cherlin and E. Hrushovski, *Finite structures with few types*, Annals of Mathematics Studies **152**, Princeton Univ. Press, 2003. MR Zbl

[Chernikov and Simon 2015] A. Chernikov and P. Simon, "Externally definable sets and dependent pairs, II", *Trans. Amer. Math. Soc.* **367**:7 (2015), 5217–5235. MR Zbl

[Conant 2017] G. Conant, "An axiomatic approach to free amalgamation", *J. Symb. Log.* **82**:2 (2017), 648–671. MR Zbl

[Dikranjan and Megrelishvili 2014] D. Dikranjan and M. Megrelishvili, "Minimality conditions in topological groups", pp. 229–327 in *Recent progress in general topology*, *III*, edited by K. P. Hart et al., Atlantis, Paris, 2014. MR Zbl

[Elliott et al. 2023] L. Elliott, J. Jonušas, J. D. Mitchell, Y. Péresse, and M. Pinsker, "Polish topologies on endomorphism monoids of relational structures", *Adv. Math.* **431** (2023), art. id. 109214. MR Zbl

[Evans 2003] D. M. Evans, "Ample dividing", *J. Symbolic Logic* **68**:4 (2003), 1385–1402. MR Zbl

[Evans et al. 2016] D. M. Evans, Z. Ghadernezhad, and K. Tent, "Simplicity of the automorphism groups of some Hrushovski constructions", *Ann. Pure Appl. Logic* **167**:1 (2016), 22–48. MR Zbl

[Gaughan 1967] E. D. Gaughan, "Topological group structures of infinite symmetric groups", *Proc. Nat. Acad. Sci. U.S.A.* **58** (1967), 907–910. MR Zbl

[Kantor et al. 1989] W. M. Kantor, M. W. Liebeck, and H. D. Macpherson, "$\aleph_0$-categorical structures smoothly approximated by finite substructures", *Proc. London Math. Soc.* (3) **59**:3 (1989), 439–463. MR Zbl

[Kim 2014] B. Kim, *Simplicity theory*, Oxford Logic Guides **53**, Oxford Univ. Press, 2014. MR Zbl

[Kim and Pillay 1998] B. Kim and A. Pillay, "From stability to simplicity", *Bull. Symbolic Logic* **4**:1 (1998), 17–36. MR Zbl

[Macpherson 2011] D. Macpherson, "A survey of homogeneous structures", *Discrete Math.* **311**:15 (2011), 1599–1634. MR Zbl

[Simon 2013] P. Simon, "Distal and non-distal NIP theories", *Ann. Pure Appl. Logic* **164**:3 (2013), 294–318. MR Zbl

[Tent and Ziegler 2012] K. Tent and M. Ziegler, *A course in model theory*, Lecture Notes in Logic **40**, Cambridge Univ. Press, 2012. MR Zbl

[Tent and Ziegler 2013] K. Tent and M. Ziegler, "On the isometry group of the Urysohn space", *J. Lond. Math. Soc.* (2) **87**:1 (2013), 289–303. MR Zbl

[Uspenskij 2008] V. V. Uspenskij, "On subgroups of minimal topological groups", *Topology Appl.* **155**:14 (2008), 1580–1606. MR Zbl

[Wagner 1994] F. O. Wagner, "Relational structures and dimensions", pp. 153–180 in *Automorphisms of first-order structures*, edited by R. Kaye and D. Macpherson, Oxford Univ. Press, 1994. MR Zbl

[Wagner 2000] F. O. Wagner, *Simple theories*, Mathematics and its Applications **503**, Kluwer Academic, Dordrecht, 2000. MR Zbl

Received May 2, 2023. Revised November 6, 2023.

ZANIAR GHADERNEZHAD
SCHOOL OF COMPUTING
UNIVERSITY OF BUCKINGHAM
UNITED KINGDOM
zaniar.ghadernezhad@buckingham.ac.uk

JAVIER DE LA NUEZ GONZÁLEZ
KOREA INSTITUTE FOR ADVANCED STUDY (KIAS)
SEOUL
SOUTH KOREA
javier.delanuez@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 327, No. 1, 2023
https://doi.org/10.2140/pjm.2023.327.107

PRIME SPECTRUM AND DYNAMICS FOR NILPOTENT CANTOR ACTIONS

STEVEN HURDER AND OLGA LUKINA

A minimal equicontinuous action by homeomorphisms of a discrete group Γ on a Cantor set $\mathfrak{X}$ is locally quasianalytic if each homeomorphism has a unique extension from small open sets to open sets of uniform diameter on $\mathfrak{X}$. A minimal action is stable if the action on $\mathfrak{X}$ of the closure of Γ in the group of homeomorphisms of $\mathfrak{X}$ is locally quasianalytic.

When Γ is virtually nilpotent, we say that $\Phi : \Gamma \times \mathfrak{X} \to \mathfrak{X}$ is a nilpotent Cantor action. We show that a nilpotent Cantor action with finite prime spectrum must be stable. We also prove there exist uncountably many distinct Cantor actions of the Heisenberg group, necessarily with infinite prime spectrum, which are not stable.

1. Introduction

A minimal equicontinuous action $\Phi : \Gamma \times \mathfrak{X} \to \mathfrak{X}$ of a countable group Γ on a Cantor space $\mathfrak{X}$ is called a *generalized odometer* [9; 14]. When $\Gamma = \mathbb{Z}$, this is just the abstract form of a traditional odometer action of the integers. For $\Gamma = \mathbb{Z}^n$ with $n \geq 2$, one obtains a more complex class of actions, whose classification becomes increasingly intractable as n increases [27], even while the dynamical properties of minimal equicontinuous Cantor actions by $\mathbb{Z}^n$ are well behaved. For Γ in general, we simply refer to these as *Cantor actions*, which will always be assumed *minimal* and *equicontinuous*.

It is a classical result that a $\mathbb{Z}$-odometer is classified by its Steinitz order, which is calculated using a representation of the action as an inverse limit of actions on finite cyclic groups. One can also associate to a Cantor action by $\mathbb{Z}^n$ its Steinitz order and also a collection of types, called its *typeset*, which consists of equivalence classes of Steinitz orders of individual elements of $\mathbb{Z}^n$. As discussed by Thomas [28, Section 4], the additional data of the typeset is still not sufficient to reduce the classification problem for Cantor actions by $\mathbb{Z}^n$ to a standard Borel equivalence relation.

MSC2020: primary 20E18, 37B05, 37B45; secondary 57S10.
Keywords: odometers, Cantor actions, profinite groups, Steinitz numbers, Heisenberg group.

In the authors' work [20], we associate the *type* and *typeset* invariants to a Cantor action $(\mathfrak{X}, \Gamma, \Phi)$ for an arbitrary countable group Γ. The type $\tau[\mathfrak{X}, \Gamma, \Phi]$ is the asymptotic equivalence class of the Steinitz order $\xi(\mathfrak{X}, \Gamma, \Phi)$ of a presentation of the action as an inverse limit of actions of Γ on finite sets.

Associated to the type $\tau[\mathfrak{X}, \Gamma, \Phi]$ is an even more basic invariant, the *prime spectrum* $\pi[\mathfrak{X}, \Gamma, \Phi]$, which consists of the set of primes which appear in a Steinitz order $\xi(\mathfrak{X}, \Gamma, \Phi)$ representing the type $\tau[\mathfrak{X}, \Gamma, \Phi]$; see Definition 2.14. The prime spectrum decomposes into two parts,

$$\pi[\mathfrak{X}, \Gamma, \Phi] = \pi_\infty[\mathfrak{X}, \Gamma, \Phi] \cup \pi_f[\mathfrak{X}, \Gamma, \Phi],$$

where the *infinite prime spectrum* $\pi_\infty[\mathfrak{X}, \Gamma, \Phi]$ consists of the primes that occur with infinite multiplicity in $\xi(\mathfrak{X}, \Gamma, \Phi)$ and the *finite prime spectrum* $\pi_f[\mathfrak{X}, \Gamma, \Phi]$ consists of the primes that occur with finite multiplicity. The prime spectrum and the finite prime spectrum are only well defined modulo finite subsets of $\pi_f[\mathfrak{X}, \Gamma, \Phi]$.

Definition 1.1. A Cantor action $(\mathfrak{X}, \Gamma, \Phi)$ has *finite spectrum* if the prime spectrum $\pi[\mathfrak{X}, \Gamma, \Phi]$ is a finite set and is said to have *infinite spectrum* otherwise.

The classification of Cantor actions for Γ is, in general, intractable and one seeks invariants for Cantor actions which at least distinguish between particular classes of actions. The authors' works [15; 16; 17; 18] study dynamical properties which yield invariants of Cantor actions. In particular, one of the most basic invariants is the property that the action is either *stable* or *wild*. The purpose of this note is to give a relation between the prime spectrum of a Cantor action and the wild property.

As explained in detail in Section 2E below, the property that the action $(\mathfrak{X}, \Gamma, \Phi)$ is stable is a property of the action of the completion $\mathfrak{G}(\Phi) = \overline{\Phi(\Gamma)} \subset \mathrm{Homeo}(\mathfrak{X})$, which is a profinite group naturally acting on $\mathfrak{X}$. The property that the action $(\mathfrak{X}, \Gamma, \Phi)$ is locally quasianalytic is defined in Definition 2.10, and $(\mathfrak{X}, \Gamma, \Phi)$ is *stable* if the action of $\mathfrak{G}(\Phi)$ on $\mathfrak{X}$ is also locally quasianalytic. If $(\mathfrak{X}, \Gamma, \Phi)$ is stable, then $(\mathfrak{X}, \Gamma, \Phi)$ is locally quasianalytic. The converse need not hold even for actions of nilpotent groups, as we show later.

A Cantor action $(\mathfrak{X}, \Gamma, \Phi)$ is said to be *nilpotent* if Γ contains a finitely generated nilpotent subgroup with finite index. This class of group actions is particularly interesting, as it has the natural next level of complexity after the abelian Cantor actions. We show the following three results for nilpotent Cantor actions.

Theorem 1.2. *Let* $(\mathfrak{X}, \Gamma, \Phi)$ *be a nilpotent Cantor action. If the prime spectrum* $\pi[\mathfrak{X}, \Gamma, \Phi]$ *is finite, then the action is stable.*

Theorem 1.2 does not have a converse. We show that every collection of primes, finite or infinite, can be realized as the prime spectrum of a stable nilpotent Cantor action.

Theorem 1.3. *Let π_f and π_∞ be two distinct sets of primes, where π_f is a finite set and π_∞ is a nonempty finite or infinite set. Then there exists a stable nilpotent Cantor action $(\mathfrak{X}, \Gamma, \Phi)$ such that $\pi_\infty[\mathfrak{X}, \Gamma, \Phi] = \pi_\infty$ and $\pi_f[\mathfrak{X}, \Gamma, \Phi] = \pi_f$.*

Let $(\mathfrak{X}, \Gamma, \Phi)$ be an abelian Cantor action. If the action is effective, then it is free, and the action of the closure $\mathfrak{G}(\Phi)$ is also free, which implies that the action is stable. An effective nilpotent Cantor action need not be free and may even have elements which fix every point in a clopen subset of the Cantor set $\mathfrak{X}$. The authors showed in their work [16] that nilpotent Cantor actions are locally quasianalytic, which means that such subsets of fixed points cannot be arbitrarily small, as their diameter has lower bound which is uniform over the Cantor set $\mathfrak{X}$. It is then surprising to discover that if one allows $\mathfrak{G}(\Phi)$ to have infinite prime spectrum then one can construct wild nilpotent actions, for which the action of the closure $\mathfrak{G}(\Phi)$ is not locally quasianalytic, as shown in Theorem 1.4. In addition, Theorem 1.4 is a realization result, which shows that every infinite set of primes can be realized as the prime spectrum of a wild nilpotent Cantor action.

Theorem 1.4. *Given any two distinct sets π_f and π_∞ of primes, where π_f is infinite and π_∞ is any (possibly empty) set, there is a minimal equicontinuous action $(\mathfrak{X}, \Gamma, \Phi)$ of the Heisenberg group such that $\pi_f[\mathfrak{X}, \Gamma, \Phi] = \pi_f$ and $\pi_\infty[\mathfrak{X}, \Gamma, \Phi] = \pi_\infty$.*

Moreover, there exists an uncountable number of nilpotent Cantor actions $(\mathfrak{X}, \Gamma, \Phi)$ of the Heisenberg group Γ with infinite prime spectra such that

(1) *each $(\mathfrak{X}, \Gamma, \Phi)$ is topologically free,*

(2) *each $(\mathfrak{X}, \Gamma, \Phi)$ is wild,*

(3) *the prime spectra of such actions are pairwise distinct.*

The notion of return equivalence for Cantor actions and its relationship with conjugacy of action is explained in Section 2D. The result of Corollary 1.5 below follows from the result that the prime spectrum of the action is an invariant of its return equivalence class; see Theorem 2.16.

Corollary 1.5. *There exists an uncountable number of nilpotent Cantor actions $(\mathfrak{X}, \Gamma, \Phi)$ of the Heisenberg group Γ which are not return equivalent and therefore not conjugate.*

The conclusion of Theorem 1.4 is used in [19] for the calculation of the mapping class groups of solenoidal manifolds whose base is a nil-manifold.

We note that for more general groups Γ, an analog of Theorem 1.2 need not hold. For example, a weakly branch group, as studied in [3; 5; 6; 25], acts on the boundary of a d-regular rooted tree, and so has finite prime spectrum $\{d\}$, but the dynamics of the action on the Cantor boundary are wild.

Question 1.6. Let $(\mathfrak{X}, \Gamma, \Phi)$ be a Cantor action. For which classes of groups Γ does the finiteness of the prime spectrum of the action imply that the action is stable?

The paper is organized as follows. In Section 2A we recall basic properties of minimal equicontinuous group actions on Cantor sets. In particular, the definition of the prime spectrum of a minimal equicontinuous action is given in Definition 2.14. We prove Theorem 1.2 in Section 3, and give basic examples of nilpotent Cantor actions in Section 4. In Section 5 we construct examples of stable and wild actions of the Heisenberg group with prescribed prime spectrum, proving Theorems 1.3 and 1.4, from which we deduce Corollary 1.5.

2. Cantor actions

We recall some of the basic properties of Cantor actions, as required for the proofs of Theorems 1.2 and 1.4. More complete discussions of the properties of equicontinuous Cantor actions are given in the text by Auslander [1], the papers by Cortez and Petite [9], Cortez and Medynets [8], and the authors' works, in particular [10; 11; 17, Section 3].

2A. ***Basic concepts.*** Let $(\mathfrak{X}, \Gamma, \Phi)$ denote an action $\Phi : \Gamma \times \mathfrak{X} \to \mathfrak{X}$. We write $g \cdot x$ for $\Phi(g)(x)$ when appropriate. The orbit of $x \in \mathfrak{X}$ is the subset $\mathcal{O}(x) = \{g \cdot x \mid g \in \Gamma\}$. The action is *minimal* if for all $x \in \mathfrak{X}$, its orbit $\mathcal{O}(x)$ is dense in $\mathfrak{X}$.

An action $(\mathfrak{X}, \Gamma, \Phi)$ is *equicontinuous* with respect to a metric $d_{\mathfrak{X}}$ on $\mathfrak{X}$ if for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in \mathfrak{X}$ and $g \in \Gamma$ we have that $d_{\mathfrak{X}}(x, y) < \delta$ implies $d_{\mathfrak{X}}(g \cdot x, g \cdot y) < \varepsilon$. The property of being equicontinuous is independent of the choice of the metric on $\mathfrak{X}$ which is compatible with the topology of $\mathfrak{X}$.

Now assume that $\mathfrak{X}$ is a Cantor space. Let $\mathrm{CO}(\mathfrak{X})$ denote the collection of all clopen (closed and open) subsets of $\mathfrak{X}$, which forms a basis for the topology of $\mathfrak{X}$. For $\phi \in \mathrm{Homeo}(\mathfrak{X})$ and $U \in \mathrm{CO}(\mathfrak{X})$, the image $\phi(U)$ belongs to $\mathrm{CO}(\mathfrak{X})$. The next result is folklore, and a proof is given in [16, Proposition 3.1].

Proposition 2.1. *For $\mathfrak{X}$ a Cantor space, a minimal action $\Phi : \Gamma \times \mathfrak{X} \to \mathfrak{X}$ is equicontinuous if and only if the Γ-orbit of every $U \in \mathrm{CO}(\mathfrak{X})$ is finite for the induced action $\Phi_* : \Gamma \times \mathrm{CO}(\mathfrak{X}) \to \mathrm{CO}(\mathfrak{X})$.*

Definition 2.2. We say that $U \subset \mathfrak{X}$ is *adapted* to the action $(\mathfrak{X}, \Gamma, \Phi)$ if U is a *nonempty clopen* subset, and for any $g \in \Gamma$, $g \cdot U \cap U \neq \varnothing$ implies $g \cdot U = U$.

The proof of [16, Proposition 3.1] shows that given $x \in \mathfrak{X}$ and clopen set $x \in W$, there is an adapted clopen set U with $x \in U \subset W$.

For an adapted set U, the set of "return times" to U,

$$\Gamma_U = \{g \in \Gamma \mid g \cdot U \cap U \neq \varnothing\}, \tag{1}$$

is a subgroup of Γ, called the *stabilizer* of U. Then for $g, g' \in \Gamma$ with $g \cdot U \cap g' \cdot U \neq \varnothing$ we have $g^{-1} g' \cdot U = U$, and hence $g^{-1} g' \in \Gamma_U$. Thus, the translates $\{g \cdot U \mid g \in \Gamma\}$ form a finite clopen partition of $\mathfrak{X}$ and are in one-to-one correspondence with the quotient space $X_U = \Gamma / \Gamma_U$. Then Γ acts by permutations of the finite set X_U and so the stabilizer group $\Gamma_U \subset G$ has finite index. Note that this implies that if $V \subset U$ is a proper inclusion of adapted sets, then the inclusion $\Gamma_V \subset \Gamma_U$ is also proper.

Definition 2.3. Let $(\mathfrak{X}, \Gamma, \Phi)$ be a Cantor action. A properly descending chain of clopen sets $\mathcal{U} = \{U_\ell \subset \mathfrak{X} \mid \ell \geq 0\}$ is said to be an *adapted neighborhood basis* at $x \in \mathfrak{X}$ for the action Φ if $x \in U_{\ell+1} \subset U_\ell$ is a proper inclusion for all $\ell > 0$, with $U_0 = \mathfrak{X}$, $\bigcap_{\ell > 0} U_\ell = \{x\}$, and each U_ℓ is adapted to the action Φ.

Given $x \in \mathfrak{X}$ and $\varepsilon > 0$, Proposition 2.1 implies there exists an adapted clopen set $U \in \mathrm{CO}(\mathfrak{X})$ with $x \in U$ and $\mathrm{diam}(U) < \varepsilon$. Thus, one can choose a descending chain $\mathcal{U}$ of adapted sets in $\mathrm{CO}(\mathfrak{X})$ whose intersection is x, from which the next result follows:

Proposition 2.4. *Let $(\mathfrak{X}, \Gamma, \Phi)$ be a Cantor action. Given $x \in \mathfrak{X}$, there exists an adapted neighborhood basis $\mathcal{U}$ at x for the action Φ.*

Combining the above remarks, we have:

Corollary 2.5. *Let $(\mathfrak{X}, \Gamma, \Phi)$ be a Cantor action and $\mathcal{U}$ be an adapted neighborhood basis. Set $\Gamma_\ell = \Gamma_{U_\ell}$, with $\Gamma_0 = \Gamma$. Then there is a nested chain of finite index subgroups $\mathcal{G}_{\mathcal{U}} = \{\Gamma_0 \supset \Gamma_1 \supset \cdots\}$.*

2B. *Profinite completion.* Let $\Phi(\Gamma) \subset \mathrm{Homeo}(\mathfrak{X})$ denote the image subgroup for an action $(\mathfrak{X}, \Gamma, \Phi)$. When the action is equicontinuous, the closure $\overline{\Phi(\Gamma)} \subset \mathrm{Homeo}(\mathfrak{X})$ in the *uniform topology of maps* is a separable profinite group. We adopt the notation $\mathfrak{G}(\Phi) \equiv \overline{\Phi(\Gamma)}$.

Let $\widehat{\Phi} : \mathfrak{G}(\Phi) \times \mathfrak{X} \to \mathfrak{X}$ denote the induced action of $\mathfrak{G}(\Phi)$ on $\mathfrak{X}$. For $\hat{g} \in \mathfrak{G}(\Phi)$, we write its action on $\mathfrak{X}$ by $\hat{g} \cdot x = \widehat{\Phi}(\hat{g})(x)$. Since the action $(\mathfrak{X}, \Gamma, \Phi)$ is minimal, the action of $\widehat{\Phi}$ on $\mathfrak{X}$ is transitive; that is, for all $x \in \mathfrak{X}$, the orbit $\{\hat{g} \cdot x \mid \hat{g} \in \mathfrak{G}(\Phi)\} = \mathfrak{X}$. Given $x \in \mathfrak{X}$, introduce the isotropy group

$$\mathfrak{D}(\Phi, x) = \{\hat{g} \in \mathfrak{G}(\Phi) \mid \hat{g} \cdot x = x\} \subset \mathrm{Homeo}(\mathfrak{X}), \tag{2}$$

which is a closed subgroup of $\mathfrak{G}(\Phi)$, and thus is either finite or is an infinite profinite group. As the action $\widehat{\Phi} : \mathfrak{G}(\Phi) \times \mathfrak{X} \to \mathfrak{X}$ is transitive, the conjugacy class of $\mathfrak{D}(\Phi, x)$ in $\mathfrak{G}(\Phi)$ is independent of the choice of x, and by abuse of notation we omit the subscript x. The group $\mathfrak{D}(\Phi)$ is called the *discriminant* of the action $(\mathfrak{X}, \Gamma, \Phi)$ in [11; 15; 17] and is called a *parabolic* subgroup (of the profinite completion of a countable group) in the works by Bartholdi and Grigorchuk [4; 5].

2C. ***Algebraic Cantor actions.*** We next describe the algebraic construction of Cantor actions, starting with a group chain in a given group Γ, and then deriving the Cantor action from this data. This is often the most versatile method of constructing examples of Cantor actions with specific properties.

Let $\mathcal{G} = \{\Gamma = \Gamma_0 \supset \Gamma_1 \supset \Gamma_2 \supset \cdots\}$ be a properly descending chain of finite index subgroups. Let $X_\ell = \Gamma/\Gamma_\ell$ and note that Γ acts transitively on the left on the finite set X_ℓ. The inclusion $\Gamma_{\ell+1} \subset \Gamma_\ell$ induces a natural Γ-invariant quotient map $p_{\ell+1} : X_{\ell+1} \to X_\ell$. Introduce the inverse limit

$$
\begin{aligned}
(3)\qquad X_\infty &\equiv \varprojlim \{p_{\ell+1} : X_{\ell+1} \to X_\ell \mid \ell \geq 0\} \\
&= \{(x_0, x_1, \ldots) \in X_\infty \mid p_{\ell+1}(x_{\ell+1}) = x_\ell \text{ for all } \ell \geq 0\} \subset \prod_{\ell \geq 0} X_\ell.
\end{aligned}
$$

Then X_∞ is a Cantor space with the Tychonoff topology, where the actions of Γ on the factors X_ℓ induce a minimal equicontinuous action $\Phi_\infty : \Gamma \times X_\infty \to X_\infty$. There is a natural basepoint $x_\infty \in X_\infty$ given by the cosets of the identity element $e \in \Gamma$, so $x_\infty = (e\Gamma_\ell)$. An adapted neighborhood basis of x_∞ is given by the clopen sets

$$
(4)\qquad V_\ell = \{x = (x_i) \in X_\infty \mid x_i = e\Gamma_i \in X_i,\ 0 \leq i \leq \ell\} \subset X_\infty.
$$

There is a tautological identity $\Gamma_\ell = \Gamma_{V_\ell}$ where Γ_{V_ℓ} is the isotropy group as defined by Corollary 2.5.

Given a minimal equicontinuous Cantor action $\Phi : \Gamma \times \mathfrak{X} \to \mathfrak{X}$ and an adapted neighborhood basis $\mathcal{U} = \{U_\ell \subset \mathfrak{X} \mid \ell \geq 0\}$ at $x \in \mathfrak{X}$, Corollary 2.5 yields a group chain $\mathcal{G}_{\mathcal{U}} = \{\Gamma_0 \supset \Gamma_1 \supset \cdots\}$. We can then associate to this group chain an algebraic action $\Phi_\infty : \Gamma \times X_\infty \to X_\infty$ as above.

For each $\ell \geq 0$, we have the "partition coding map" $\Theta_\ell : \mathfrak{X} \to X_\ell$ which is Γ-equivariant. The maps $\{\Theta_\ell\}$ are compatible with the map on quotients in (3), and so they induce a limit map $\Theta_x : \mathfrak{X} \to X_\infty$. The fact that the diameters of the clopen sets $\{V_\ell\}$ tend to zero implies that Θ_x is a homeomorphism. Also, $\Theta_x(x) = x_\infty \in X_\infty$. The following is folklore:

Theorem 2.6 [10, Appendix A]. *The map $\Theta_x : \mathfrak{X} \to X_\infty$ induces an isomorphism of the Cantor actions $(\mathfrak{X}, \Gamma, \Phi)$ and $(X_\infty, \Gamma, \Phi_\infty)$.*

The action $(X_\infty, \Gamma, \Phi_\infty)$ is called the *odometer model* centered at x for the action $(\mathfrak{X}, \Gamma, \Phi)$. The dependence of the model on the choices of a base point $x \in \mathfrak{X}$ and adapted neighborhood basis $\mathcal{U}$ is discussed in detail in the works [10; 12; 15; 17].

Next, we develop the algebraic model for the profinite action $\widehat{\Phi} : \mathfrak{G}(\Phi) \times \mathfrak{X} \to \mathfrak{X}$ of the completion $\mathfrak{G}(\Phi) \equiv \overline{\Phi(\Gamma)} \subset \mathrm{Homeo}(\mathfrak{X})$. Choose a group chain $\{\Gamma_\ell \mid \ell \geq 0\}$ as above, which provides an algebraic model for the action $(\mathfrak{X}, \Gamma, \Phi)$.

For each $\ell \geq 1$, let $C_\ell \subset \Gamma_\ell$ denote the *core* of Γ_ℓ, i.e., the largest normal subgroup of Γ_ℓ in Γ. So

$$C_\ell = \mathrm{Core}(\Gamma_\ell) = \bigcap_{g \in \Gamma} g\Gamma_\ell g^{-1} \subset \Gamma_\ell. \tag{5}$$

As Γ_ℓ has finite index in Γ, the same holds for C_ℓ. Observe that for all $\ell \geq 0$, we have $C_{\ell+1} \subset C_\ell$.

Introduce the quotient group $Q_\ell = \Gamma/C_\ell$ with identity element $e_\ell \in Q_\ell$. There are natural quotient maps $q_{\ell+1} : Q_{\ell+1} \to Q_\ell$, and we can form the inverse limit group

$$\widehat{\Gamma}_\infty \equiv \varprojlim \{q_{\ell+1} : Q_{\ell+1} \to Q_\ell \mid \ell \geq 0\} \tag{6}$$

$$= \{(g_\ell) = (g_0, g_1, \ldots) \mid g_\ell \in Q_\ell,\ q_{\ell+1}(g_{\ell+1}) = g_\ell \ \text{ for all } \ell \geq 0\} \subset \prod_{\ell \geq 0} \Gamma_\ell, \tag{7}$$

which is a Cantor space with the Tychonoff topology. The left actions of Γ on the spaces $X_\ell = \Gamma/\Gamma_\ell$ induce a minimal equicontinuous action of $\widehat{\Gamma}_\infty$ on X_∞, again denoted by $\widehat{\Phi} : \widehat{\Gamma}_\infty \times X_\infty \to X_\infty$. Note that the isotropy group of the action of $Q_\ell = \Gamma_\ell/C_\ell$ at the identity coset in $X_\ell = \Gamma/\Gamma_\ell$ is the subgroup $D_\ell = \Gamma_\ell/C_\ell$.

Denote the points in $\widehat{\Gamma}_\infty$ by $\hat{g} = (g_\ell) \in \widehat{\Gamma}_\infty$ where $g_\ell \in Q_\ell$. There is a natural basepoint $\hat{e}_\infty \in \widehat{\Gamma}_\infty$ given by the cosets of the identity element $e \in \Gamma$, so $\hat{e}_\infty = (e_\ell)$ where $e_\ell = eC_\ell \in Q_\ell$ is the identity element in Q_ℓ.

For each $\ell \geq 0$, let $\Pi_\ell : \widehat{\Gamma}_\infty \to Q_\ell$ denote the projection onto the ℓ-th factor in (6), so in the coordinates of (7), we have $\Pi_\ell(\hat{g}) = g_\ell \in Q_\ell$. The maps Π_ℓ are continuous for the profinite topology on $\widehat{\Gamma}_\infty$, so the preimages of points in Q_ℓ are clopen subsets. In particular, the fiber of $\Pi_\ell : \widehat{\Gamma}_\infty \to Q_\ell$ over e_ℓ is the normal subgroup

$$\widehat{C}_\ell = \Pi_\ell^{-1}(e_\ell) = \{(g_i) \in \widehat{\Gamma}_\infty \mid g_i \in C_i,\ 0 \leq i \leq \ell\}. \tag{8}$$

The collection $\{\widehat{C}_\ell \mid \ell \geq 1\}$ forms a basis of clopen neighborhoods of $\hat{e}_\infty \in \widehat{\Gamma}_\infty$. That is, for each clopen set $\widehat{U} \subset \widehat{\Gamma}_\infty$ with $\hat{e}_\infty \in \widehat{U}$, there exists $\ell_0 > 0$ such that $\widehat{C}_\ell \subset \widehat{U}$ for all $\ell \geq \ell_0$.

Theorem 2.7 [10, Theorem 4.4]. *There is an isomorphism $\hat{\tau} : \mathfrak{G}(\Phi) \to \widehat{\Gamma}_\infty$ such that $\hat{\tau}$ conjugates the profinite action $(\mathfrak{X}, \mathfrak{G}(\Phi), \widehat{\Phi})$ with the profinite action $(X_\infty, \widehat{\Gamma}_\infty, \widehat{\Phi}_\infty)$. In particular, $\hat{\tau}$ identifies the isotropy group $\mathfrak{D}(\Phi)$ with the inverse limit subgroup*

$$D_\infty = \varprojlim \{q_{\ell+1} : \Gamma_{\ell+1}/C_{\ell+1} \to \Gamma_\ell/C_\ell \mid \ell \geq 0\} \subset \widehat{\Gamma}_\infty. \tag{9}$$

The maps $q_{\ell+1}$ in the formula (9) need not be surjections, and thus the calculation of the inverse limit D_∞ can involve some subtleties. For example, it is possible that each group Q_ℓ is nontrivial for $\ell > 0$, and yet D_∞ is the trivial group.

2D. ***Equivalence of Cantor actions.*** We next recall the notions of equivalence of Cantor actions. The first and strongest is that of *isomorphism*, which is a generalization of the notion of conjugacy of topological actions. For $\Gamma = \mathbb{Z}$, isomorphism corresponds to the notion of "flip conjugacy" introduced in the work of Boyle and Tomiyama [7]. The definition below also appears in the papers [8; 15; 23].

Definition 2.8. Cantor actions $(\mathfrak{X}_1, \Gamma_1, \Phi_1)$ and $(\mathfrak{X}_2, \Gamma_2, \Phi_2)$ are said to be *isomorphic* if there is a homeomorphism $h : \mathfrak{X}_1 \to \mathfrak{X}_2$ and a group isomorphism $\Theta : \Gamma_1 \to \Gamma_2$ such that

$$\Phi_1(g) = h^{-1} \circ \Phi_2(\Theta(g)) \circ h \in \mathrm{Homeo}(\mathfrak{X}_1) \quad \text{for all } g \in \Gamma_1. \tag{10}$$

The notion of *return equivalence* for Cantor actions is weaker than isomorphism and is natural when considering the dynamical properties of Cantor systems which should be independent of the restriction of the action to a clopen cross-section.

Given a minimal equicontinuous Cantor action $(\mathfrak{X}, \Gamma, \Phi)$ and an adapted set $U \subset \mathfrak{X}$, recall that Γ_U denotes the isotropy group for U, as in (1). By a small abuse of notation, we use Φ_U to denote both the restricted action $\Phi_U : \Gamma_U \times U \to U$ and the induced quotient action $\Phi_U : \mathcal{H}_U \times U \to U$ for $\mathcal{H}_U = \Phi(\Gamma_U) \subset \mathrm{Homeo}(U)$. Then $(U, \mathcal{H}_U, \Phi_U)$ is called the *holonomy action* for Φ.

Definition 2.9. Two minimal equicontinuous Cantor actions $(\mathfrak{X}_1, \Gamma_1, \Phi_1)$ and $(\mathfrak{X}_2, \Gamma_2, \Phi_2)$ are *return equivalent* if there exists an adapted set $U_1 \subset \mathfrak{X}_1$ for the action Φ_1 and an adapted set $U_2 \subset \mathfrak{X}_2$ for the action Φ_2, such that the holonomy actions $(U_1, \mathcal{H}_{1,U_1}, \Phi_{1,U_1})$ and $(U_2, \mathcal{H}_{2,U_2}, \Phi_{2,U_2})$ are isomorphic.

If the actions Φ_1 and Φ_2 are isomorphic in the sense of Definition 2.8, then they are return equivalent with $U_1 = \mathfrak{X}_1$ and $U_2 = \mathfrak{X}_2$. However, the notion of return equivalence is weaker even for this case, as the conjugacy is between the holonomy groups $\mathcal{H}_{1,\mathfrak{X}_1}$ and $\mathcal{H}_{2,\mathfrak{X}_2}$, and not the groups Γ_1 and Γ_2.

2E. ***Locally quasianalytic.*** The quasianalytic property for Cantor actions was introduced by Álvarez López and Candel in [24, Definition 9.4] as a generalization of the notion of a *quasianalytic action* studied by Haefliger for actions of pseudogroups of real-analytic diffeomorphisms. The authors introduced a local form of the quasianalytic property in [11; 15]:

Definition 2.10 [15, Definition 2.1]. A topological action $(\mathfrak{X}, \Gamma, \Phi)$ on a metric Cantor space $\mathfrak{X}$ is *locally quasianalytic* (LQA) if there exists $\varepsilon > 0$ such that for any nonempty open set $U \subset \mathfrak{X}$ with $\mathrm{diam}(U) < \varepsilon$, and for any nonempty open subset $V \subset U$, and elements $g_1, g_2 \in \Gamma$,

$$\text{if } \Phi(g_1)|V = \Phi(g_2)|V, \text{ then } \Phi(g_1)|U = \Phi(g_2)|U. \tag{11}$$

The action is said to be *quasianalytic* if (11) holds for $U = \mathfrak{X}$.

In other words, $(\mathfrak{X}, \Gamma, \Phi)$ is locally quasianalytic if for every $g \in \Gamma$, the homeomorphism $\Phi(g)$ has unique extensions on the sets of diameter $\varepsilon > 0$ in $\mathfrak{X}$, with ε uniform over $\mathfrak{X}$. We note that for a countable group Γ, an effective action $(\mathfrak{X}, \Gamma, \Phi)$ is topologically free if and only if it is quasianalytic.

Recall that a group Γ is *Noetherian* [2] if every increasing chain of subgroups has a maximal element. Equivalently, a group is Noetherian if every subgroup of Γ is finitely generated. A group is *topologically Noetherian* if every increasing chain of closed subgroups has a maximal element; see Section 3 for details.

Theorem 2.11 [16, Theorem 1.6]. *Let Γ be a Noetherian group. Then a minimal equicontinuous Cantor action $(\mathfrak{X}, \Gamma, \Phi)$ is locally quasianalytic.*

A finitely generated nilpotent group is Noetherian, so as a corollary we obtain that all Cantor actions by finitely generated nilpotent groups are locally quasianalytic.

The notion of a locally quasianalytic Cantor action extends to the case of a profinite group action $\widehat{\Phi} : \mathfrak{G} \times \mathfrak{X} \to \mathfrak{X}$.

Definition 2.12. Let $(\mathfrak{X}, \Gamma, \Phi)$ be a Cantor action and $\widehat{\Phi} : \mathfrak{G} \times \mathfrak{X} \to \mathfrak{X}$ the induced profinite action. We say that the action is *stable* if the induced profinite action $(\mathfrak{X}, \mathfrak{G}(\Phi), \widehat{\Phi})$ is locally quasianalytic, and we say it is *wild* otherwise.

A profinite completion $\mathfrak{G}$ of a Noetherian group Γ need not be Noetherian, as can be seen for the example of $\Gamma = \mathbb{Z}$, where $\mathfrak{G}$ is the full profinite completion of $\mathbb{Z}$. More generally, a finitely generated nilpotent group Γ is always Noetherian, while Proposition 3.4 gives an "if and only if" condition for a profinite completion $\mathfrak{G}$ of Γ to be topologically Noetherian.

2F. ***Type and typeset for Cantor actions.*** A Steinitz number ξ can be written uniquely as the formal product over the set of primes Π:

$$\xi = \prod_{p \in \Pi} p^{\chi_\xi(p)}, \tag{12}$$

where the *characteristic function* $\chi_\xi : \Pi \to \{0, 1, \dots, \infty\}$ counts the multiplicity with which a prime p appears in the infinite product ξ.

Definition 2.13. Two Steinitz numbers ξ and ξ' are said to be *asymptotically equivalent* if there exists finite integers $m, m' \geq 1$ such that $m \cdot \xi = m' \cdot \xi'$, and we then write $\xi \overset{a}{\sim} \xi'$.

A *type* is an asymptotic equivalence class of Steinitz numbers. The type associated to a Steinitz number ξ is denoted by $\tau[\xi]$.

In terms of their characteristic functions χ_1, χ_2, we have $\xi \overset{a}{\sim} \xi'$ if and only if the following conditions are satisfied:

- $\chi_1(p) = \chi_2(p)$ for all but finitely many primes $p \in \Pi$.

- $\chi_1(p) = \infty$ if and only if $\chi_1(p) = \infty$ for all primes $p \in \Pi$.

Given two types, τ and τ', we write $\tau \leq \tau'$ if there exists representatives $\xi \in \tau$ and $\xi' \in \tau'$ such that their characteristic functions satisfy $\chi_\xi(p) \leq \chi_{\xi'}(p)$ for all primes $p \in \Pi$.

Definition 2.14. Let π denote the set of primes. Given $\xi = \prod_{p\in\pi} p^{\chi_\xi(p)}$, define

- $\pi(\xi) = \{p \in \pi \mid \chi_\xi(p) > 0\}$, *the prime spectrum of* ξ,
- $\pi_f(\xi) = \{p \in \pi \mid 0 < \chi_\xi(p) < \infty\}$, *the finite prime spectrum of* ξ,
- $\pi_\infty(\xi) = \{p \in \pi \mid \chi_\xi(p) = \infty\}$, *the infinite prime spectrum of* ξ.

Note that if $\xi \overset{a}{\sim} \xi'$, then $\pi_\infty(\xi) = \pi_\infty(\xi')$. The property that $\pi_f(\xi)$ is an *infinite* set is also preserved by asymptotic equivalence of Steinitz numbers.

Next, we define the type of a Cantor action $(X_\infty, \Gamma, \Phi_\infty)$ defined by a chain of finite index subgroups, $\mathcal{G} = \{\Gamma = \Gamma_0 \supset \Gamma_1 \supset \cdots\}$. Let $C_\ell \subset \Gamma_\ell$ denote the normal core of Γ_ℓ.

Definition 2.15. Let (X_∞, Γ, Φ) be a minimal equicontinuous Cantor action defined by a group chain $\mathcal{G}$. The type $\tau[X_\infty, \Gamma, \Phi_\infty]$ of the action is the equivalence class of the Steinitz order

$$\xi(X_\infty, \Gamma, \Phi_\infty) = \mathrm{lcm}\{\#X_\ell = \#(\Gamma/\Gamma_\ell) \mid \ell > 0\}. \tag{13}$$

Finally, we note the following result:

Theorem 2.16 [20, Theorem 1.9]. *Let $(\mathfrak{X}, \Gamma, \Phi)$ be a Cantor action. The Steinitz order $\xi(\mathfrak{X}, \Gamma, \Phi)$ is defined as the Steinitz order for an algebraic model $(X_\infty, \Gamma, \Phi_\infty)$ of the action, which does not depend upon the choice of an algebraic model. The type $\tau[\mathfrak{X}, \Gamma, \Phi]$ depends only on the return equivalence class of the action.*

2G. ***Type for profinite groups.*** The *Steinitz order* $\Pi[\mathfrak{G}]$ of a profinite group $\mathfrak{G}$ is defined by the supernatural number associated to a presentation of $\mathfrak{G}$ as an inverse limit of finite groups (see [26, Chapter 2.3; 29, Chapter 2]). The Steinitz order appears in the study of analytic representations of profinite groups associated to groups acting on rooted trees; see, for example, [22].

Recall that for a profinite group $\mathfrak{G}$, an open subgroup $\mathfrak{U} \subset \mathfrak{G}$ has finite index [26, Lemma 2.1.2].

Definition 2.17. Let $(\mathfrak{X}, \Gamma, \Phi)$ be a minimal equicontinuous Cantor action, with choice of a basepoint $x \in \mathfrak{X}$. The *Steinitz orders* of the action are defined as

(1) $\xi(\mathfrak{G}(\Phi)) = \mathrm{lcm}\{\#\,\mathfrak{G}(\Phi)/\mathfrak{N} \mid \mathfrak{N} \subset \mathfrak{G}(\Phi)$ open normal subgroup$\}$,

(2) $\xi(\mathfrak{D}(\Phi)) = \mathrm{lcm}\{\#\,\mathfrak{D}(\Phi)/(\mathfrak{N} \cap \mathfrak{D}(\Phi)) \mid \mathfrak{N} \subset \mathfrak{G}(\Phi)$ open normal subgroup$\}$,

(3) $\xi(\mathfrak{G}(\Phi) : \mathfrak{D}(\Phi)) = \mathrm{lcm}\{\#\,\mathfrak{G}(\Phi)/(\mathfrak{N} \cdot \mathfrak{D}(\Phi)) \mid \mathfrak{N} \subset \mathfrak{G}(\Phi)$ open normal subgroup$\}$.

The Steinitz orders satisfy the Lagrange identity, where the multiplication is taken in the sense of supernatural numbers,

$$\xi(\mathfrak{G}(\Phi)) = \xi(\mathfrak{G}(\Phi) : \mathfrak{D}(\Phi)) \cdot \xi(\mathfrak{D}(\Phi)). \tag{14}$$

Thus, we always have $\tau[\mathfrak{D}(\Phi)] \leq \tau[\mathfrak{G}(\Phi)]$. The following is a direct consequence of the definitions:

Theorem 2.18. *Let $(\mathfrak{X}, \Gamma, \Phi)$ be a Cantor action. Then there is equality of Steinitz orders*, $\xi(\mathfrak{X}, \Gamma, \Phi) = \xi(\mathfrak{G}(\Phi) : \mathfrak{D}(\Phi))$.

3. Nilpotent actions

We apply the notion of the Steinitz order of a nilpotent Cantor action to the study of its dynamical properties. The proof of Theorem 1.2 is based on the special properties of the profinite completions of nilpotent groups, in particular the uniqueness of their Sylow p-subgroups, and on the relation of this algebraic property with the dynamics of the action.

3A. ***Noetherian groups.*** A countable group Γ is said to be *Noetherian* [2] if every increasing chain of subgroups $\{H_i \mid i \geq 1\}$ of Γ has a maximal element H_{i_0}. The group $\mathbb{Z}$ is Noetherian; a finite product of Noetherian groups is Noetherian; and a subgroup and quotient group of a Noetherian group is Noetherian. Thus, a finitely generated nilpotent group is Noetherian.

The notion of a Noetherian group has a generalization which is useful for the study of actions of profinite groups.

Definition 3.1 [29, page 153]. A profinite group $\mathfrak{G}$ is said to be *topologically Noetherian* if every increasing chain of *closed* subgroups $\{\mathfrak{H}_i \mid i \geq 1\}$ of $\mathfrak{G}$ has a maximal element $\mathfrak{H}_{i_0}$.

We illustrate this concept with two canonical examples of profinite completions of $\mathbb{Z}$.

Example 3.2. Let $\widehat{\mathbb{Z}}_p$ denote the p-adic integers, for p a prime. That is, $\widehat{\mathbb{Z}}_p$ is the completion of $\mathbb{Z}$ with respect to the chain of subgroups $\mathcal{G} = \{\Gamma_\ell = p^\ell \mathbb{Z} \mid \ell \geq 1\}$. The closed subgroups of $\widehat{\mathbb{Z}}_p$ are given by $p^i \cdot \widehat{\mathbb{Z}}_p$ for some fixed $i > 0$, and hence satisfy the ascending chain property in Definition 3.1.

Example 3.3. Let $\widehat{\pi} = \{p_i \mid i \geq 1\}$ be an infinite collection of distinct primes. Define an increasing chain of subgroups of $\mathbb{Z}$ as $\mathcal{G}_{\widehat{\pi}} = \{\Gamma_\ell = p_1 p_2 \cdots p_\ell \mathbb{Z} \mid \ell \geq 1\}$. Let $\widehat{\mathbb{Z}}_{\widehat{\pi}}$ be the completion of $\mathbb{Z}$ with respect to the chain $\mathcal{G}_{\widehat{\pi}}$. Then we have a topological isomorphism

$$\widehat{\mathbb{Z}}_{\widehat{\pi}} \cong \prod_{i \geq 1} \mathbb{Z}/p_i \mathbb{Z}. \tag{15}$$

Let $H_\ell = \mathbb{Z}/p_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/p_\ell\mathbb{Z}$ be the direct sum of the first ℓ factors. Then $\{H_\ell \mid \ell \geq 1\}$ is an increasing chain of subgroups of $\widehat{\mathbb{Z}}_{\widehat{\pi}}$ which does not stabilize, so $\widehat{\mathbb{Z}}_{\widehat{\pi}}$ is not topologically Noetherian.

These two examples illustrate the idea behind the proof of the following result.

Proposition 3.4. *Let Γ be a finitely generated nilpotent group, and let $\widehat{\Gamma}$ be a profinite completion of Γ. Then $\widehat{\Gamma}$ is topologically Noetherian if and only if the prime spectrum $\pi(\xi(\widehat{\Gamma}))$ is finite.*

Proof. First, recall some basic facts about profinite groups. (See, for example, [29, Chapter 2].) For a prime p, a finite group H is a p-group if every element of H has order a power of p. A profinite group $\mathfrak{H}$ is a pro-p-group if $\mathfrak{H}$ is the inverse limit of finite p-groups. A Sylow p-subgroup $\mathfrak{H} \subset \mathfrak{G}$ is a maximal pro-p-subgroup [29, Definition 2.2.1].

A profinite group $\mathfrak{G}$ is *pro-nilpotent* if it is the inverse limit of finite nilpotent groups. For example, if $\mathfrak{G}$ is a profinite completion of a nilpotent group Γ, then $\mathfrak{G}$ is pro-nilpotent.

The group $\mathfrak{G}$ is topologically finitely generated if it contains a dense subgroup $\Gamma \subset \mathfrak{G}$ where Γ is finitely generated. The completion $\mathfrak{G}(\Phi)$ associated to a Cantor action $(\mathfrak{X}, \Gamma, \Phi)$ with Γ finitely generated is topologically finitely generated.

Assume that $\mathfrak{G}$ is pro-nilpotent. Then for each prime p, there is a unique Sylow p-subgroup of $\mathfrak{G}$, which is normal in $\mathfrak{G}$ (see [29, Proposition 2.4.3]). Denote this group by $\mathfrak{G}_{(p)}$. Also, $\mathfrak{G}_{(p)}$ is nontrivial if and only if $p \in \pi(\xi(\mathfrak{G}))$. We use the following result for pro-nilpotent groups, which is a consequence of [29, Proposition 2.4.3].

Proposition 3.5. *Let $\mathfrak{G}$ be a profinite completion of a finitely generated nilpotent group Γ. Then there is a topological isomorphism*

$$\mathfrak{G} \cong \prod_{p \in \pi(\xi(\mathfrak{G}))} \mathfrak{G}_{(p)}. \tag{16}$$

From the isomorphism (16) it follows immediately that if the prime spectrum $\pi(\xi(\mathfrak{G}))$ is infinite, then $\mathfrak{G}$ is not topologically Noetherian. To see this, list $\pi(\xi(\mathfrak{G})) = \{p_i \mid i = 1, 2, \ldots\}$. Then we obtain an infinite strictly increasing chain of closed subgroups

$$\mathfrak{H}_\ell = \prod_{i=1}^{\ell} \mathfrak{G}_{(p_i)}.$$

If the prime spectrum $\pi(\xi(\mathfrak{G}))$ is finite, then the isomorphism (16) reduces the proof that $\mathfrak{G}$ is topologically Noetherian to the case of showing that if $\mathfrak{G}$ is topologically finitely generated, then each of its Sylow p-subgroups is Noetherian. The group $\mathfrak{G}_{(p)}$ is nilpotent and topologically finitely generated, so we can use the lower central series for $\mathfrak{G}_{(p)}$ and induction to reduce to the case where $\mathfrak{H}$ is a topologically finitely generated abelian pro-p-group, and so is isomorphic to a finite product of p-completions of $\mathbb{Z}$, which are topologically Noetherian.

Observe that a profinite completion $\mathfrak{G}$ of a finitely generated nilpotent group Γ is a topologically finitely generated nilpotent group, and we apply the above remarks. □

Corollary 3.6. *Let Γ be a virtually nilpotent group; that is, there exists a finitely generated nilpotent subgroup $\Gamma_0 \subset \Gamma$ of finite index. Then a profinite completion $\mathfrak{G}$ of Γ is topologically Noetherian if and only if its prime spectrum $\pi(\xi(\mathfrak{G}))$ is finite.*

Proof. We can assume that Γ_0 is a normal subgroup of Γ. Thus, its closure $\mathfrak{G}_0 \subset \mathfrak{G}$ satisfies the hypotheses of Proposition 3.4, and the Steinitz orders satisfy $\xi(\mathfrak{G}_0) \overset{\mathrm{a}}{\sim} \xi(\mathfrak{G})$. As $\mathfrak{G}_0$ is topologically Noetherian if and only if $\mathfrak{G}$ is topologically Noetherian, the claim follows. □

3B. ***Dynamics of Noetherian groups.*** We relate the topologically Noetherian property of a profinite group with the dynamics of a Cantor action of the group to obtain the proof of Theorem 1.2. We first give the profinite analog of [16, Theorem 1.6]. We follow the outline of its proof in [16].

Proposition 3.7. *Let $\mathfrak{G}$ be a topologically Noetherian group. Then a minimal equicontinuous action $(\mathfrak{X}, \mathfrak{G}, \widehat{\Phi})$ on a Cantor space $\mathfrak{X}$ is locally quasianalytic.*

Proof. The closure $\mathfrak{G}(\Phi)$ is contained in Homeo$(\mathfrak{X})$, so the action $\widehat{\Phi}$ of $\mathfrak{G}(\Phi)$ is effective. Suppose that the action $\widehat{\Phi}$ is not locally quasianalytic. Then there exists an infinite properly decreasing chain of clopen subsets of $\mathfrak{X}$, $\{U_1 \supset U_2 \supset \cdots\}$, which satisfy, for all $\ell \geq 1$, the properties

- U_ℓ is adapted to the action $\widehat{\Phi}$ with isotropy subgroup $\mathfrak{G}_{U_\ell} \subset \mathfrak{G}$;
- there is a closed subgroup $K_\ell \subset \mathfrak{G}_{U_{\ell+1}}$ whose restricted action to $U_{\ell+1}$ is trivial, but the restricted action of K_ℓ to U_ℓ is effective.

Hence, we obtain a properly increasing chain of closed subgroups $\{K_1 \subset K_2 \subset \cdots\}$ in $\mathfrak{G}$, which contradicts the assumption that $\mathfrak{G}$ is topologically Noetherian. □

Proof of Theorem 1.2. Let $(\mathfrak{X}, \Gamma, \Phi)$ be a nilpotent Cantor action, and we are given that the prime spectrum $\pi(\xi(\mathfrak{G}(\Phi)))$ is finite. Then there exists a finitely generated nilpotent subgroup $\Gamma_0 \subset \Gamma$ of finite index, and we can assume without loss of generality that Γ_0 is normal. Let $\mathfrak{G}(\Phi)_0$ be the closure of Γ_0 in $\mathfrak{G}(\Phi)$. The group $\mathfrak{G}(\Phi)$ has finite prime spectrum implies that the group $\mathfrak{G}(\Phi)_0$ has finite prime spectrum, and thus by Proposition 3.4 the group $\mathfrak{G}(\Phi)_0$ is topologically Noetherian. Let $x \in \mathfrak{X}$. Then it suffices to show that the action of Γ_0 on the orbit $\mathfrak{X}_0 = \mathfrak{G}(\Phi)_0 \cdot x$ is stable. This reduces the proof to showing the claim when Γ is nilpotent. Then the profinite closure $\mathfrak{G}(\Phi)$ is also nilpotent, and we have a profinite action $(\mathfrak{X}, \mathfrak{G}(\Phi), \widehat{\Phi})$.

Suppose that the action $\widehat{\Phi}$ is not locally quasianalytic. Then there exists an increasing chain of closed subgroups $K_\ell \subset \mathfrak{D}(\Phi)$ where K_ℓ acts trivially on the clopen subset $U_\ell \subset \mathfrak{X}$. As $\mathfrak{D}(\Phi)$ is a closed subgroup of $\mathfrak{G}(\Phi)$, the increasing chain $\{K_\ell \mid \ell > 0\}$ consists of closed subgroups of $\mathfrak{G}(\Phi)$. This contradicts the fact that $\mathfrak{G}(\Phi)$ is topologically Noetherian. Hence, the action $\widehat{\Phi}$ must be locally quasianalytic. That is, the action $(\mathfrak{X}, \Gamma, \Phi)$ is stable. □

4. Basic examples

We construct two basic examples of nilpotent Cantor actions. These examples illustrate the principles behind the subsequent more complex constructions in Section 5, which are used to prove Theorems 1.3 and 1.4.

The integer Heisenberg group is the simplest nonabelian nilpotent group, and it can be represented as the upper triangular matrices in $\mathrm{GL}(3, \mathbb{Z})$. That is,

$$\Gamma = \left\{ \begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} \mid a, b, c \in \mathbb{Z} \right\}. \tag{17}$$

We denote a 3×3 matrix in Γ by the coordinates as (a, b, c).

Example 4.1. A *renormalizable Cantor action*, as defined in [21], can be constructed from the group chain defined by a proper self-embedding of a group Γ into itself.

For a prime $p \geq 2$, define the self-embedding $\varphi_p : \Gamma \to \Gamma$ by $\varphi(a, b, c) = (pa, pb, p^2c)$. Then define a group chain in Γ by setting

$$\Gamma_\ell = \varphi_p^\ell(\Gamma) = \{(p^\ell a, p^\ell b, p^{2\ell}c) \mid a, b, c \in \mathbb{Z}\}, \qquad \bigcap_{\ell>0} \Gamma_\ell = \{e\}.$$

For $\ell > 0$, the normal core for Γ_ℓ is given by

$$C_\ell = \mathrm{core}(\Gamma_\ell) = \{(p^{2\ell}a, p^{2\ell}b, p^{2\ell}c) \mid a, b, c \in \mathbb{Z}\},$$

and so the quotient group is given by $Q_\ell = \Gamma / C_\ell \cong \{(\bar{a}, \bar{b}, \bar{c}) \mid \bar{a}, \bar{b}, \bar{c} \in \mathbb{Z}/p^{2\ell}\mathbb{Z}\}$. The profinite group $\widehat{\Gamma}_\infty$ is the inverse limit of the quotient groups Q_ℓ so we have $\widehat{\Gamma}_\infty = \{(\hat{a}, \hat{b}, \hat{c}) \mid \hat{a}, \hat{b}, \hat{c} \in \widehat{\mathbb{Z}}_{p^2}\}$. Thus, $\xi(\widehat{\Gamma}) = \{p^\infty\}$. Even though the quotient groups Γ_ℓ / C_ℓ are all nontrivial, for this action the inverse limit D_∞ is the trivial group. This follows from the fact that there are inclusions

$$\Gamma_{2\ell} = \{(p^{2\ell}a, p^{2\ell}b, p^{4\ell}c) \mid a, b, c \in \mathbb{Z}\} \subset C_\ell = \{(p^{2\ell}a, p^{2\ell}b, p^{2\ell}c) \mid a, b, c \in \mathbb{Z}\}.$$

The triviality of D_∞ implies that there is an equivalent group chain for the action [10] which can be chosen so that every subgroup in the chain is normal in Γ.

Example 4.2. For distinct primes $p, q \geq 2$, define the self-embedding $\varphi_{p,q} : \Gamma \to \Gamma$ by $\varphi(a, b, c) = (pa, qb, pqc)$. Then define a group chain in Γ by setting

$$\Gamma_\ell = \varphi_{p,q}^\ell(\Gamma) = \{(p^\ell a, q^\ell b, (pq)^\ell c) \mid a, b, c \in \mathbb{Z}\}, \qquad \bigcap_{\ell>0} \Gamma_\ell = \{e\}.$$

For $\ell > 0$, the normal core for Γ_ℓ is given by

$$C_\ell = \mathrm{core}(\Gamma_\ell) = \{((pq)^\ell a, (pq)^\ell b, (pq)^\ell c) \mid a, b, c \in \mathbb{Z}\},$$

and so the quotient group is given by $Q_\ell = \Gamma/C_\ell \cong \{(\bar{a}, \bar{b}, \bar{c}) \mid \bar{a}, \bar{b}, \bar{c} \in \mathbb{Z}/(pq)^\ell\mathbb{Z}\}$. The profinite group $\widehat{\Gamma}_\infty$ is the inverse limit of the quotient groups Q_ℓ, so we have $\widehat{\Gamma}_\infty = \{(\hat{a}, \hat{b}, \hat{c}) \mid \hat{a}, \hat{b}, \hat{c} \in \widehat{\mathbb{Z}}_{pq}\}$. Thus, $\xi(\widehat{\Gamma}_\infty) = \{p^\infty, q^\infty\}$, and D_∞ is the inverse limit of the finite groups Γ_ℓ/C_ℓ by (9), so $D_\infty \cong \widehat{\mathbb{Z}}_q \times \widehat{\mathbb{Z}}_p$.

5. Nilpotent actions with prescribed spectrum

We construct stable actions of the discrete Heisenberg group with prescribed prime spectrum, proving Theorem 1.3. Then we construct examples of wild nilpotent Cantor actions, proving Theorem 1.4, from which we deduce Corollary 1.5. For simplicity, our examples all use the Heisenberg group represented by 3×3 matrices. Of course, these examples can be generalized to the integer upper triangular matrices in all dimensions, where there is much more freedom in the choices made in the construction. The calculations become correspondingly more tedious, but yield analogous results. It seems reasonable to expect that similar constructions can be made for any finitely generated torsion-free nilpotent (nonabelian) group Γ. That is, there are always group chains in Γ which yield wild nilpotent Cantor actions.

Let $\Gamma \subset \mathrm{GL}(3, \mathbb{Z})$ denote the discrete Heisenberg group, given by formula (17). The basis for the constructions below is the structure theory for nilpotent group completions in Proposition 3.5, in particular the formula (16). Given sets of primes π_f and π_∞, we embed an infinite product of finite actions, as in Section 5A, into a profinite completion $\widehat{\Gamma}_\infty$ of Γ, and thus obtain a nilpotent Cantor action $(X_\infty, \Gamma, \Phi_\infty)$ on the quotient space $X_\infty = \widehat{\Gamma}_\infty/D_\infty$.

5A. ***Basic components of the construction.*** Fix a prime $p \geq 2$.

For $n \geq 1$ and $0 \leq k < n$, we have the finite groups

$$G_{p,n} = \left\{ \begin{bmatrix} 1 & \bar{a} & \bar{c} \\ 0 & 1 & \bar{b} \\ 0 & 0 & 1 \end{bmatrix} \mid \bar{a}, \bar{b}, \bar{c} \in \mathbb{Z}/p^n\mathbb{Z} \right\}, \quad H_{p,n,k} = \left\{ \begin{bmatrix} 1 & p^k\bar{a} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mid \bar{a} \in \mathbb{Z}/p^n\mathbb{Z} \right\}$$

Note that $\#[G_{p,n}] = p^{3n}$ and $\#[H_{p,n,k}] = p^{n-k}$.

Let $\bar{x} = (1, 0, 0)$, $\bar{y} = (0, 1, 0)$, $\bar{z} = (0, 0, 1) \in G_{p,n}$. Then $\bar{x} \cdot \bar{y} \cdot \bar{x}^{-1} = \bar{y}\bar{z}$ and $\bar{x} \cdot \bar{z} \cdot \bar{x}^{-1} = \bar{z}$. That is, the adjoint action of $\bar{x}$ on the "plane" in the $(\bar{y}, \bar{z})$-coordinates is a "shear" action along the $\bar{z}$-axis, and the adjoint action of $\bar{x}$ on the $\bar{z}$-axis fixes all points on the $\bar{z}$-axis.

Set $X_{p,n,k} = G_{p,n}/H_{p,n,k}$. Then the isotropy group of the action of $G_{p,n}$ on $X_{p,n,k}$ at the coset $H_{p,n,k}$ of the identity element is $H_{p,n,k}$. The core subgroup $C_{p,n,k} \subset H_{p,n,k}$ contains elements in $H_{p,n,k}$ which fix every point in $X_{p,n,k}$. The action of $\bar{x} \in H_{p,n,k}$ on the coset space $X_{p,n,k}$ satisfies

$$\Phi_\infty(\bar{x})(\bar{y}\, H_{p,n,k}) = \bar{y}\bar{z}\, H_{p,n,k}, \tag{18}$$

so the identity is the only element in $G_{p,n}$ which acts trivially on every coset in $X_{p,n,k}$, so $C_{p,n,k}$ is the trivial group. Then $D_{p,n,k} = H_{p,n,k}/C_{p,n,k} = H_{p,n,k}$, and for each $g \in H_{p,n,k}$ its action fixes the cosets of the multiples of $\bar{z}$.

5B. ***Stable nilpotent actions with finite or infinite prime spectrum.*** We now prove Theorem 1.3 by constructing a family of stable examples with prescribed prime spectra.

Let π_f and π_∞ be two disjoint collections of primes, with π_f a finite set and π_∞ a nonempty set.

Enumerate $\pi_f = \{q_1, q_2, \ldots, q_m\}$, and then choose integers $1 \leq r_i \leq n_i$ for $1 \leq i \leq m$.

Enumerate $\pi_\infty = \{p_1, p_2, \ldots\}$ with the convention (for notational convenience) that if ℓ is greater than the number of primes in π_∞ then we set $p_\ell = 1$. For each $\ell \geq 1$, define the integers

$$M_\ell = q_1^{r_1} q_2^{r_2} \cdots q_m^{r_m} \cdot p_1^\ell p_2^\ell \cdots p_\ell^\ell, \tag{19}$$

$$N_\ell = q_1^{n_1} q_2^{n_2} \cdots q_m^{n_m} \cdot p_1^\ell p_2^\ell \cdots p_\ell^\ell. \tag{20}$$

For all $\ell \geq 1$, observe that M_ℓ divides N_ℓ.

Define a subgroup of the Heisenberg group Γ, in the coordinates in (17),

$$\Gamma_\ell = \{(aM_\ell, bN_\ell, cN_\ell) \mid a, b, c \in \mathbb{Z}\}.$$

Its core subgroup is given by $C_\ell = \{(aN_\ell, bN_\ell, cN_\ell) \mid a, b, c \in \mathbb{Z}\}$. Observe that

$$\mathbb{Z}/N_\ell\mathbb{Z} \cong \mathbb{Z}/q_1^{n_1}\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/q_m^{n_m}\mathbb{Z} \oplus \mathbb{Z}/p_1^\ell\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/p_\ell^\ell\mathbb{Z}.$$

By Proposition 3.5, and in the notation of Section 5A, we have for $k_i = n_i - r_i$ that

$$\widehat{\Gamma}_\infty = \varprojlim\{\Gamma/C_\ell \to \Gamma/C_{\ell-1} \mid \ell \geq 1\} \cong \prod_{i=1}^m G_{q_i,n_i} \cdot \prod_{j=1}^\infty \widehat{\Gamma}_{(p_j)}, \tag{21}$$

$$D_\infty = \varprojlim\{\Gamma_\ell/C_\ell \to \Gamma_{\ell-1}/C_{\ell-1} \mid \ell \geq 1\} \cong \prod_{i=1}^m H_{q_i,n_i,k_i}. \tag{22}$$

Then the Cantor space $X_\infty = \widehat{\Gamma}_\infty / D_\infty$ associated to the group chain $\{\Gamma_\ell \mid \ell \geq 1\}$ is given by

$$X_\infty \cong \prod_{i=1}^{m} X_{q_i,n_i,k_i} \times \prod_{j=1}^{\infty} \widehat{\Gamma}_{(p_j)}. \tag{23}$$

In particular, as the first factor in (23) is a finite product of finite sets, the second factor defines an open neighborhood

$$U = \prod_{i=1}^{m} \{x_i\} \times \prod_{j=1}^{\infty} \widehat{\Gamma}_{(p_j)},$$

where $x_i \in X_{q_i,n_i,k_i}$ is the basepoint given by the coset of the identity element. That is, U is a clopen neighborhood of the basepoint in X_∞. The isotropy group of U is given by

$$\widehat{\Gamma}_\infty | U = \prod_{i=1}^{m} H_{q_i,n_i,k_i} \times \prod_{j=1}^{\infty} \widehat{\Gamma}_{(p_j)}. \tag{24}$$

The restriction of $\widehat{\Gamma}_\infty | U$ to U is isomorphic to the subgroup

$$K|U = \prod_{i=1}^{m} \{\bar{e}_i\} \times \prod_{j=1}^{\infty} \widehat{\Gamma}_{(p_j)} \subset \mathrm{Homeo}(U), \tag{25}$$

where $\bar{e}_i \in G_{q_i,n_i}$ is the identity element. The group $K|U$ acts freely on U, and thus the action of $\widehat{\Gamma}_\infty$ on X_∞ is locally quasianalytic. The prime spectrum of the action of Γ on X_∞ is the union $\widehat{\pi} = \pi_f \cup \pi_\infty = \pi(\xi(\widehat{\Gamma}_\infty))$. If π_∞ is infinite, then the prime spectrum of the action is infinite. Note that the group Γ embeds into $\widehat{\Gamma}_\infty$, since the integers M_ℓ and N_ℓ tend to infinity with ℓ. This completes the proof of Theorem 1.3.

5C. ***Wild nilpotent actions with infinite prime spectrum.*** We prove Theorem 1.4. We must show that every infinite set of primes can be realized as the prime spectrum of a wild action of the Heisenberg group Γ, as defined by (17). Let π_f and π_∞ be disjoint collections of primes, with π_f an infinite set and π_∞ arbitrary, possibly empty.

Enumerate $\pi_f = \{q_1, q_2, \ldots\}$ and choose integers $1 \leq r_i < n_i$ for $1 \leq i < \infty$.

Enumerate $\pi_\infty = \{p_1, p_2, \ldots\}$, again with the convention that if ℓ is greater than the number of primes in π_∞ then we set $p_\ell = 1$.

As in Section 5B, for each $\ell \geq 1$, define the integers

$$M_\ell = q_1^{r_1} q_2^{r_2} \cdots q_\ell^{r_\ell} \cdot p_1^\ell p_2^\ell \cdots p_\ell^\ell, \quad N_\ell = q_1^{n_1} q_2^{n_2} \cdots q_\ell^{n_\ell} \cdot p_1^\ell p_2^\ell \cdots p_\ell^\ell.$$

For $\ell \geq 1$, define a subgroup of the Heisenberg group Γ, in the coordinates in (17),

$$\Gamma_\ell = \{(aM_\ell, bN_\ell, cN_\ell) \mid a, b, c \in \mathbb{Z}\}. \tag{26}$$

Its core subgroup is given by $C_\ell = \{(aN_\ell, bN_\ell, cN_\ell) \mid a, b, c \in \mathbb{Z}\}$. For $k_i = n_i - r_i$ we then have

$$\widehat{\Gamma}_\infty \cong \prod_{i=1}^{\infty} G_{q_i,n_i} \cdot \prod_{j=1}^{\infty} \widehat{\Gamma}_{(p_j)}, \qquad D_\infty \cong \prod_{i=1}^{\infty} H_{q_i,n_i,k_i}. \tag{27}$$

The Cantor space $X_\infty = \widehat{\Gamma}_\infty / D_\infty$ associated to the group chain $\{\Gamma_\ell \mid \ell \geq 1\}$ is given by

$$X_\infty \cong \prod_{i=1}^{\infty} X_{q_i,n_i,k_i} \times \prod_{j=1}^{\infty} \widehat{\Gamma}_{(p_j)}. \tag{28}$$

The first factor in (23) is an infinite product of finite sets, so fixing the first ℓ coordinates in this product determines a clopen subset of X_∞. Let $x_i \in X_{q_i,n_i,k_i}$ denote the coset of the identity element, which is the basepoint in X_{q_i,n_i,k_i}. Then for each $\ell \geq 1$, we define a clopen set in X_∞ by

$$U_\ell = \prod_{i=1}^{\ell} \{x_i\} \times \prod_{i=\ell+1}^{\infty} X_{q_i,n_i,k_i} \times \prod_{j=1}^{\infty} \widehat{\Gamma}_{(p_j)}. \tag{29}$$

By calculations in Section 5A, the subgroup H_{q_i,n_i,k_i} is the isotropy group of the basepoint $x_i \in X_{q_i,n_i,k_i}$. Thus, the isotropy subgroup of U_ℓ for the $\widehat{\Gamma}_\infty$-action is given by

$$\widehat{\Gamma}_\infty|_{U_\ell} = \prod_{i=1}^{\ell} H_{q_i,n_i,k_i} \times \prod_{i=\ell+1}^{\infty} G_{q_i,n_i} \times \prod_{j=1}^{\infty} \widehat{\Gamma}_{(p_j)}. \tag{30}$$

For $j \neq i$, the subgroup H_{q_i,n_i,k_i} acts as the identity on the factors X_{q_j,n_j,k_j} in (28). Thus, the image of $\widehat{\Gamma}_\infty|_{U_\ell}$ in $\mathrm{Homeo}(U_\ell)$ is isomorphic to the subgroup

$$Z_\ell = \widehat{\Gamma}_\infty|U_\ell = \prod_{i=1}^{\ell} \{\bar{e}_i\} \times \prod_{i=\ell+1}^{\infty} G_{q_i,n_i} \times \prod_{j=1}^{\infty} \widehat{\Gamma}_{(p_j)} \subset \mathrm{Homeo}(U_\ell), \tag{31}$$

where $\bar{e}_i \in G_{q_i,n_i}$ is the identity element.

We next show that this action is not stable; that is, for any $\ell > 0$ there exists a clopen subset $V \subset U_\ell$ and nontrivial $\hat{g} \in Z_\ell$ so that the action of $\widehat{\Gamma}_\infty$ restricts to the identity map on V.

We can assume without loss of generality that $V = U_{\ell'}$ for some $\ell' > \ell$. Consider the restriction map for the isotropy subgroup of Z_ℓ to $U_{\ell'}$ which is given by

$$\rho_{\ell,\ell'} : Z_\ell|_{U_{\ell'}} \to Z_{\ell'} \subset \mathrm{Homeo}(U_{\ell'}).$$

We must show that there exists $\ell' > \ell$ such that this map has a nontrivial kernel. Calculate this map in terms of the product representations above:

$$Z_\ell|_{U_{\ell'}} = \prod_{i=1}^{\ell} \{\bar{e}_i\} \times \prod_{i=\ell+1}^{\ell'} H_{q_i,n_i,k_i} \times \prod_{i=\ell'+1}^{\infty} G_{q_i,n_i} \times \prod_{j=1}^{\infty} \widehat{\Gamma}_{(p_j)}. \tag{32}$$

For $\ell < i \leq \ell'$, the group H_{q_i,n_i,k_i} fixes the point $\prod_{i=1}^{\ell'}\{x_i\}$, and acts trivially on $\prod_{i=\ell'+1}^{\infty} X_{q_i,n_i,k_i}$. Thus, the kernel of the restriction map contains the second factor in (32):

$$\prod_{i=\ell+1}^{\ell'} H_{q_i,n_i,k_i} \subset \ker\{\rho_{\ell,\ell'} : Z_\ell|_{U_{\ell'}} \to \mathrm{Homeo}(U_{\ell'})\}. \tag{33}$$

As this group is nontrivial for all $\ell' > \ell$, the action of $\widehat{\Gamma}_\infty$ on X_∞ is not locally quasianalytic, and hence the action of Γ on X_∞ is wild. Also, the prime spectrum of the action of Γ on X_∞ equals the union $\widehat{\pi} = \pi_f \cup \pi_\infty$.

We now prove the second part of Theorem 1.4, showing that choices in the construction above can be made in such a way that the action of Γ on a Cantor set is topologically free while the action of $\widehat{\Gamma}_\infty$ is wild, and the prime spectrum is prescribed.

Choose an infinite set of distinct primes $\pi_f = \{q_1, q_2, \ldots\}$, and let π_∞ be empty. Choose the constants as in Section 5A, with $n_i = 2$ and $k_i = 1$ for all $i \geq 1$.

Define the Cantor space X_∞ by (28), where the second factor is trivial; that is, a point. The action of $\widehat{\Gamma}_\infty$ is wild by the calculations in formulas (30) to (33).

We claim that the action of Γ on X_∞ is topologically free. If not, then there exists an open set $U \subset X_\infty$ and $g \in \Gamma$ such that the action of $\Phi_\infty(g)$ is nontrivial on X_∞ but leaves the set U invariant and restricts to the identity action on U. The action of Γ on X_∞ is minimal, so there exists $h \in \Gamma$ with $h \cdot x_\infty \in U$. Then $\Phi_\infty(h^{-1}gh)(x_\infty) = x_\infty$ and the action $\Phi_\infty(h^{-1}gh)$ fixes an open neighborhood of x_∞. Replacing g with $h^{-1}gh$ we can assume that $\Phi_\infty(g)(x_\infty) = x_\infty \in U$. From the definition (29), the clopen sets

$$U_\ell = \prod_{i=1}^{\ell}\{x_i\} \times \prod_{i=\ell+1}^{\infty} X_{q_i,2,1} \tag{34}$$

form a neighborhood basis at x_∞, and thus there exists $\ell > 0$ such that $U_\ell \subset U$.

The group Γ embeds into $\widehat{\Gamma}_\infty$ along the diagonal in the product (16). That is, we can write $g = (g, g, \ldots) \in \prod_{i=1}^{\infty} G_{q_i,2}$. The action of $\Phi_\infty(g)$ is factorwise, and $\Phi_\infty(g)(x_\infty) = x_\infty$ implies that $g \in D_\infty \cong \prod_{i=1}^{\infty} H_{q_i,n_i,k_i}$. The assumption that $\Phi_\infty(g)$ fixes the points in U implies that it acts trivially on each factor $X_{q_i,2,1}$ for $i > \ell$. As each factor $H_{q_i,2,1}$ acts effectively on $X_{q_i,2,1}$ this implies that the projection of g to the i-th factor group $H_{q_i,2,1}$ is the identity for $i > \ell$. This implies that every entry above the diagonal in the matrix representation of g in (17) is divisible by an infinite number of distinct primes $\{q_i \mid i \geq \ell\}$, so by the prime factorization theorem the matrix g is the identity.

Alternatively, observe that we have $g \in \prod_{i=1}^{\ell} H_{q_i,2,1}$. This is a finite product of finite groups, which implies that $g \in \Gamma$ is a torsion element. However, the Heisenberg group Γ is torsion-free, and hence g must be the identity. Thus, the action of Γ on X_∞ must be topologically free.

Finally, the above construction allows the choice of any infinite subset π_f of distinct primes, and there are an uncountable number such choices which are distinct. Thus, by Theorem 1.9 in [20] there are an uncountable number of topologically free, wild nilpotent Cantor actions with distinct prime spectrum. This completes the proof of Theorem 1.4.

5D. ***Proof of Corollary 1.5.*** Consider the family of wild topologically free actions on the Heisenberg group Γ with infinite distinct prime spectrum, as constructed at the end of Section 5C. We show that the uncountable number of infinite choices of π_f in this family can be made so that the actions have pairwise disjoint types.

By Definition 2.13, for two Steinitz numbers ξ and ξ' we have that their types are equal, $\tau(\xi) = \tau(\xi')$, if and only if there exist integers m, m' such that $m \cdot \xi = m' \cdot \xi'$. Thus two actions with prime spectra π_f and π'_f have distinct types if and only if π_f and π'_f differ by an infinite number of entries. This happens, for instance, if π_f and π'_f are *almost disjoint infinite sets*, i.e., they are infinite sets with finite intersection.

The set of prime numbers is countable, so the family of infinite almost disjoint subsets of prime numbers is uncountable if and only if the family of infinite almost disjoint subsets of natural numbers is uncountable. The family of almost disjoint subsets of natural numbers is uncountable by [13, Corollary 2.3]. Since the set of finite subsets of natural numbers is countable, the set of almost disjoint infinite subsets of natural numbers is uncountable.

It follows that the prime spectra of the uncountable family of actions of the Heisenberg group in Theorem 1.4 can be chosen so that they form a family of almost disjoint infinite sets. Then their types are pairwise distinct, and by Theorem 2.16 these actions of the Heisenberg group are pairwise not return equivalent. Therefore, they are pairwise not conjugate.

References

[1] J. Auslander, *Minimal flows and their extensions*, North-Holland Mathematics Studies **153**, North-Holland, Amsterdam, 1988. MR

[2] R. Baer, "Noethersche Gruppen", *Math. Z.* **66** (1956), 269–288. MR Zbl

[3] L. Bartholdi, "Endomorphic presentations of branch groups", *J. Algebra* **268**:2 (2003), 419–443. MR Zbl

[4] L. Bartholdi and R. I. Grigorchuk, "On the spectrum of Hecke type operators related to some fractal groups", pp. 5–45 in *Din. Sist.*, *Avtom. i Beskon. Gruppy*, vol. 231, 2000. MR Zbl

[5] L. Bartholdi and R. I. Grigorchuk, "On parabolic subgroups and Hecke algebras of some fractal groups", *Serdica Math. J.* **28**:1 (2002), 47–90. MR Zbl

[6] L. Bartholdi, R. I. Grigorchuk, and Z. Šunić, "Branch groups", pp. 989–1112 in *Handbook of algebra*, vol. 3, edited by M. Hazewinkel, Handb. Algebr. **3**, North-Holland, Amsterdam, 2003. MR Zbl

[7] M. Boyle and J. Tomiyama, "Bounded topological orbit equivalence and C^*-algebras", *J. Math. Soc. Japan* **50**:2 (1998), 317–329. MR Zbl

[8] M. I. Cortez and K. Medynets, “Orbit equivalence rigidity of equicontinuous systems”, *J. Lond. Math. Soc.* (2) **94**:2 (2016), 545–556. MR Zbl

[9] M. I. Cortez and S. Petite, “G-odometers and their almost one-to-one extensions”, *J. Lond. Math. Soc.* (2) **78**:1 (2008), 1–20. MR Zbl

[10] J. Dyer, S. Hurder, and O. Lukina, “The discriminant invariant of Cantor group actions”, *Topology Appl.* **208** (2016), 64–92. MR Zbl

[11] J. Dyer, S. Hurder, and O. Lukina, “Molino theory for matchbox manifolds”, *Pacific J. Math.* **289**:1 (2017), 91–151. MR Zbl

[12] R. Fokkink and L. Oversteegen, “Homogeneous weak solenoids”, *Trans. Amer. Math. Soc.* **354**:9 (2002), 3743–3755. MR Zbl

[13] S. Geschke, “Almost disjoint and independent families”, pp. 1–9 in *Aspects of Descriptive Set Theory* (Kyoto, 2011), edited by H. Fujita, Sūrikaisekikenkyūsho Kōkyūroku **1790**, RIMS, Kyoto, 2012.

[14] T. Giordano, I. F. Putnam, and C. F. Skau, “$\mathbb{Z}^d$-odometers and cohomology”, *Groups Geom. Dyn.* **13**:3 (2019), 909–938. MR Zbl

[15] S. Hurder and O. Lukina, “Wild solenoids”, *Trans. Amer. Math. Soc.* **371**:7 (2019), 4493–4533. MR Zbl

[16] S. Hurder and O. Lukina, “Orbit equivalence and classification of weak solenoids”, *Indiana Univ. Math. J.* **69**:7 (2020), 2339–2363. MR Zbl

[17] S. Hurder and O. Lukina, “Limit group invariants for non-free Cantor actions”, *Ergodic Theory Dynam. Systems* **41**:6 (2021), 1751–1794. MR Zbl

[18] S. Hurder and O. Lukina, “Nilpotent Cantor actions”, *Proc. Amer. Math. Soc.* **150**:1 (2022), 289–304. MR Zbl

[19] S. Hurder and O. Lukina, “Mapping class groups of solenoidal manifolds”, preprint, 2023.

[20] S. Hurder and O. Lukina, “Type invariants for solenoidal manifolds”, preprint, 2023. arXiv 2305.00863

[21] S. Hurder, O. Lukina, and W. van Limbeek, “Cantor dynamics of renormalizable groups”, *Groups Geom. Dyn.* **15**:4 (2021), 1449–1487. MR Zbl

[22] S. Kionke, “Groups acting on rooted trees and their representations on the boundary”, *J. Algebra* **528** (2019), 260–284. MR Zbl

[23] X. Li, “Continuous orbit equivalence rigidity”, *Ergodic Theory Dynam. Systems* **38**:4 (2018), 1543–1563. MR Zbl

[24] J. A. Álvarez López and A. Candel, “Equicontinuous foliated spaces”, *Math. Z.* **263**:4 (2009), 725–774. MR

[25] V. Nekrashevych, *Self-similar groups*, Mathematical Surveys and Monographs **117**, American Mathematical Society, Providence, RI, 2005. MR Zbl

[26] L. Ribes and P. Zalesskii, *Profinite groups*, Ergeb. Math. Grenzgeb. **40**, Springer, 2000. MR Zbl

[27] S. Thomas, “The classification problem for torsion-free abelian groups of finite rank”, *J. Amer. Math. Soc.* **16**:1 (2003), 233–258. MR Zbl

[28] S. Thomas, “The classification problem for finite rank Butler groups”, pp. 329–338 in *Models, modules and abelian groups*, edited by R. Göbel and B. Goldsmith, Walter de Gruyter, Berlin, 2008. MR Zbl

[29] J. S. Wilson, *Profinite groups*, London Mathematical Society Monographs. New Series **19**, Oxford University Press, 1998. MR Zbl

Received May 6, 2023. Revised November 28, 2023.

STEVEN HURDER
DEPARTMENT OF MATHEMATICS, STATISTICS & COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS AT CHICAGO
CHICAGO, IL
UNITED STATES
hurder@uic.edu

OLGA LUKINA
MATHEMATICAL INSTITUTE
LEIDEN UNIVERSITY
LEIDEN
NETHERLANDS
o.lukina@math.leidenuniv.nl

PACIFIC JOURNAL OF MATHEMATICS
Vol. 327, No. 1, 2023
https://doi.org/10.2140/pjm.2023.327.129

A NOTE ON THE DISTINCT DISTANCES PROBLEM IN THE HYPERBOLIC PLANE

ZHIPENG LU AND XIANCHANG MENG

We provide a proof of a Guth–Katz-type lower bound for the distinct distances problem in the hyperbolic plane. Our construction follows the framework of Guth and Katz to deal with $\mathrm{PSL}_2(\mathbb{R})$ and the corresponding incidence structure in projective geometry. In addition, we deduce a new sum-product estimate in the form of a hyperbolic metric formula based on this lower bound.

1. Introduction

The distinct distances problem was first proposed by Erdős [3] in the Euclidean plane. He conjectured the lower bound $\gtrsim N/\sqrt{\log N}$ for the number of distinct distances between pairs of points among N points in the plane. (Here $A \gtrsim B$ means $A \geq cB$ for some absolute constant $c > 0$.) After a half-century of progression with partial results, there came the major breakthrough by Guth and Katz [4] who proved the nearly optimal bound $\gtrsim N/\log N$. Foremostly they invented the tool of polynomial partitioning and promoted profound applications in incidence geometry and other areas, later developed by themselves and many other authors; for instances, see [1; 6].

In this paper, we deal with the distinct distances problem in the hyperbolic plane $\mathbb{H}^2$ and prove the nearly optimal bound in equivalent strength with [4]. Following an idea of Tao's blog [11], Rudnev and Selig [9] described a proof using the Klein quadric in Plüker coordinates without exploiting symmetries in the hyperbolic plane. By contrast, following the framework of Elekes and Sharir, as in [4], we give an independent proof by carefully studying isometries of $\mathbb{H}^2$ in a more Guth–Katz ethnic language. More specifically, we prove:

Theorem 1.1. *For any set $P \subset \mathbb{H}^2$ of N points, we have*

$$|\{d_{\mathbb{H}^2}(p,q),\ p,q \in P\}| \gtrsim N/\log N,$$

where $|A|$ denotes the cardinality of a set A and $d_{\mathbb{H}^2}$ denotes the hyperbolic metric on $\mathbb{H}^2$.

MSC2020: 11P70, 52C10.
Keywords: Erdős distinct distances, hyperbolic geometry, sum-product.

In the case of the Euclidean plane, Guth and Katz [4] used the framework of Elekes and Sharir [2] to reduce the distinct distances problem to an incidence problems of lines, then derived the lower bound resorting to ruled surface theory and polynomial partitioning. Elekes and Sharir's framework serves as a realization of the Erlangen program (see [7] for historical background) for the distinct distances problem in the Euclidean plane. However, this framework cannot apply directly to the case of the hyperbolic plane. For the hyperbolic plane $\mathbb{H}^2$, we consider its isometry group $\mathrm{PSL}_2(\mathbb{R})$. Distinguished from Guth and Katz's coordinate of lines, our lines lie in $\mathbb{P}^3$ rather than $\mathbb{R}^3$. We need further linearizations to reduce our coordinate of lines to $\mathbb{R}^3$. Subsequently we need to overcome the difficulty of constructing vector fields in order to use ruled surface theory. See Section 2 for details.

In addition, we deduce a new sum-product-type result using Theorem 1.1. For any finite sets $A \subset \mathbb{R}\backslash\{0\}$, $B \subset \mathbb{R}$, define $P = \{b + i|a| : a \in A,\ b \in B\}$, and $P' = \{-b + i|a| : a \in A,\ b \in B\}$. Note that explicitly we have the hyperbolic distance formula

$$2\cosh d_{\mathbb{H}^2}(x_1 + iy_1, x_2 + iy_2) = \frac{(x_1 - x_2)^2 + y_1^2 + y_2^2}{y_1 y_2}$$

and $|\{|x| : x \in E\}| \geq \frac{1}{2}|E|$ for any finite set $E \subset \mathbb{R}$. By applying Theorem 1.1 to P and P', we get:

Theorem 1.2. *Let $A \subset \mathbb{R}\backslash\{0\}$, $B \subset \mathbb{R}$ be finite sets. Then we have*

$$\left|\left\{\frac{a_1^2 + a_2^2 + (b_1 - b_2)^2}{a_1 a_2} : a_1, a_2 \in A,\ b_1, b_2 \in B\right\}\right| \gtrsim \frac{|A|\,|B|}{\log(|A|) + \log(|B|)},$$

and

$$\left|\left\{\frac{a_1^2 + a_2^2 + (b_1 + b_2)^2}{a_1 a_2} : a_1, a_2 \in A,\ b_1, b_2 \in B\right\}\right| \gtrsim \frac{|A|\,|B|}{\log(|A|) + \log(|B|)}.$$

By adding or subtracting 2 on the elements in the above sets, the factor $a_1^2 + a_2^2$ can be replaced by $(a_1 + a_2)^2$ or $(a_1 - a_2)^2$.

Remark 1. In particular, if $|A|$ and $|B|$ are all about the size $\asymp N$, the above lower bounds become $\gtrsim N^2/\log N$.

A variant of the distinct distances problem has been previously used by Roche-Newton and Rudnev [8] to study sum-product-type estimates. See also the work of Jones [5] for estimates of other sum-product-types using incidence geometry. Very recently, Sheffer and Zahl [10] derived a sum-product-type estimate for complex numbers.

2. Proof of Theorem 1.1

We use Elekes and Sharir's framework to reduce the counting of distinct distances to an incidence problem of lines in the real projective space $\mathbb{P}^3$. To overcome the difficulty of linearizing projective lines in $\mathbb{P}^3$, we turn the incidence of lines in $\mathbb{P}^3$ into that of lines in $\mathbb{R}^3$ by certain conjugation. Then fulfilling the requirements for our lines in $\mathbb{R}^3$ as Guth and Katz in Proposition 2.8 of [4] amounts to a more concrete proof of the lower bound $\gtrsim N/\log N$ of distinct distances among N points in $\mathbb{H}^2$.

Framework. Let $\mathbb{H}^2$ be the hyperbolic plane and $G = \mathrm{PSL}_2(\mathbb{R})$ be its isometry group which acts on $\mathbb{H}^2$ by Möbius transformation:

$$z \mapsto \gamma \cdot z = \frac{az+b}{cz+d} \quad \text{for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{R}),\ z \in \mathbb{H}^2.$$

Let $P \subset \mathbb{H}^2$ be a set of N points and define the set of *distance quadruples*

$$Q(P) := \{(p_1, p_2, p_3, p_4) \in P^4 : d(p_1, p_2) = d(p_3, p_4) \neq 0\}, \tag{1}$$

where $d(\cdot\,,\cdot)$ denotes the hyperbolic metric. Denote the distance set by

$$d(P) := \{d(p_1, p_2) : p_1 \neq p_2 \in P\}.$$

Then we have a close relation between $d(P)$ and $Q(P)$ as follows. Suppose $d(P) = \{d_i : 1 \le i \le m\}$ and n_i is the number of pairs of points in P with distance d_i. So $|Q(P)| = \sum_{i=1}^m n_i^2$. Since $\sum_{i=1}^m n_i = 2\binom{N}{2} = N^2 - N$, by Cauchy–Schwarz inequality we get

$$(N^2-N)^2 = \left(\sum_{i=1}^m n_i\right)^2 \le \left(\sum_{i=1}^m n_i^2\right) m = |Q(P)|\,|d(P)|.$$

Rearranging the inequality gives

$$|d(P)| \ge \frac{N^4 - 2N^3}{|Q(P)|}. \tag{2}$$

Any quadruple $(p_1, p_2, p_3, p_4) \in Q(P)$ uniquely determines an isometry $g \in G$ such that $g(p_1) = p_3$, $g(p_2) = p_4$. Suppose $p_1 = x + iy$, $p_3 = x' + iy' \in \mathbb{H}^2$ $(y, y' > 0)$ and there is some $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G$ such that

$$A \cdot (x+iy) = \frac{a(x+iy)+b}{c(x+iy)+d} = x' + iy',$$

for $i = \sqrt{-1}$. Rearranging terms we get

$$ax + b + iay = cxx' + dx' - cyy' + i(cxy' + dy' + cx'y),$$

or equivalently the system of linear equations

$$
\begin{aligned}
xa + b + (yy' - xx')c - x'd &= 0,\\
ya - (xy' + x'y)c - y'd &= 0.
\end{aligned}
\tag{3}
$$

Its solution set in $\mathbb{R}^4$ is the intersection of two distinct hyperplanes, which turns out to be a two-dimensional plane passing through the origin. If, in addition, $A \cdot p_2 = p_4$, the point (a, b, c, d) also lies in another distinct two-dimensional plane intersecting the above plane at a line since $p_1 \neq p_2$, $p_3 \neq p_4$ as follows.

Lemma 2.1. *The equations of* (3) *determine a unique dimension-*2 *hyperplane in* $\mathbb{R}^4$ *for each distinct pair of points in* $\mathbb{H}^2$. *In particular, any quadruple* $(p_1, p_2, p_3, p_4) \in Q(P)$ *determines a unique isometry.*

Proof. A fairly complicated elementary computation on 4×4 matrices derived from (3) allows us to see this, but here we prove it by geometric arguments.

First, a nonidentity real Möbius transformation can have at most one fixed point in $\mathbb{H}^2$, since $\frac{az+b}{cz+d} = z$ implies $cz^2 + (d-a)z - b = 0$ which has 1 or no roots in $\mathbb{H}^2$ for real coefficients. If two isometries $\gamma_1, \gamma_2 \in \mathrm{PSL}_2(\mathbb{R})$ satisfy $\gamma_i \cdot p_1 = p_3$ and $\gamma_i \cdot p_2 = p_4$, then $\gamma_1^{-1}\gamma_2$ fixes both p_1 and p_2, a contradiction ($p_1 \neq p_2$). This is to say a quadruple in $Q(P)$ determines at most one isometry, or equivalently, two systems of equations for two pairs of points as in (3) define different planes that intersect on at most one line.

Then we verify the existence of solution. Since $\mathrm{PSL}_2(\mathbb{R})$ acts on $\mathbb{H}^2$ transitively (which can also be seen from (3)), let $\gamma_j \cdot i = p_j$, $j = 1, \dots, 4$. Then

$$\gamma \cdot p_1 = p_3, \;\; \gamma \cdot p_2 = p_4 \quad \Longleftrightarrow \quad \gamma_3^{-1}\gamma\gamma_1 \cdot i = i, \;\; \gamma_4^{-1}\gamma\gamma_2 \cdot i = i.$$

For $i = (0, 1)$, (3) simply becomes

$$
\begin{aligned}
b + c &= 0,\\
a - d &= 0.
\end{aligned}
$$

Let its solution plane be π; then the desired solution set of γ is $\gamma_3\pi\gamma_1^{-1} \cap \gamma_4\pi\gamma_2^{-1} = \gamma_3(\pi \cap \gamma_3^{-1}\gamma_4\pi\gamma_2^{-1}\gamma_1)\gamma_1^{-1}$. Note that $d(i, \gamma_2^{-1}\gamma_1 \cdot i) = d(\gamma_2 \cdot i, \gamma_1 \cdot i) = d(p_2, p_1) = d(p_4, p_3) = d(\gamma_4 \cdot i, \gamma_3 \cdot i) = d(i, \gamma_4^{-1}\gamma_3 \cdot i)$. Hence there exists a rotation $\gamma \in \pi$ about i that transfers $\gamma_2^{-1}\gamma_1 \cdot i$ to $\gamma_4^{-1}\gamma_3 \cdot i$, that is, $\gamma\gamma_2^{-1}\gamma_1 \cdot i = \gamma_4^{-1}\gamma_3 \cdot i$, or $\gamma_3^{-1}\gamma_4\gamma\gamma_2^{-1}\gamma_1 \cdot i = i$. This is to say

$$\gamma \in \pi \cap \gamma_3^{-1}\gamma_4\pi\gamma_2^{-1}\gamma_1,$$

so that $\gamma_3^{-1}\gamma_4\pi\gamma_2^{-1}\gamma_1 \neq \varnothing$ and then the desired solution set $\gamma_3\pi\gamma_1^{-1} \cap \gamma_4\pi\gamma_2^{-1}$ is not empty. □

Thus all (a, b, c, d) lying in the intersection line of two planes defined by (3) in $\mathbb{R}^4$ project to a single point as $[a : b : c : d] \in \mathbb{P}^3$. This gives a map $E : Q(P) \to G$. Define, for any $p, q \in \mathbb{H}^2$,

$$S_{pq} := \{g \in G : g(p) = q\},$$

which are one-dimensional curves in G. Similar to [4, Lemmas 2.4 and 2.6], we have

(i) if $|P \cap gP| = k$, then $|E^{-1}(g)| = 2\binom{k}{2}$;

(ii) and $|P \cap gP| \geq k$ if and only if g lies in at least k of the curves $\{S_{pq}\}_{p,q \in P}$.

Thus we derive that

$$|Q(P)| = \sum_{k=2}^{N} 2\binom{k}{2} |\{g : |P \cap gP| = k\}| \lesssim \sum_{k=2}^{N} k\, |G_k(P)|, \tag{4}$$

where $G_k(P) \subset G$ consists of $g \in G$ with $|P \cap gP| \geq k$. Henceforth we focus on estimating $|G_k(P)|$ for $k = 2$ and $k \geq 3$ as in Sections 3 and 4 of [4].

Incidence of projective lines in $\mathbb{P}^3$. For any $g \in G$, we have $d(gp, gq) = d(p, q)$ so that shifting P to gP does not affect counting of distinct distances. Now for a quadruple $(p_1, p_2, p_3, p_4) \in Q(P)$, suppose $E((p_1, p_2, p_3, p_4)) = h$, i.e., $hp_1 = p_3$, $hp_3 = p_4$. After shifting we get

$$E((gp_1, gp_2, gp_3, gp_4)) = ghg^{-1}.$$

In the matrix form of G, we manage to reshape the distance quadruples as follows.

Proposition 2.2. *For any finite set of points $P \subset \mathbb{H}^2$, there is an isometry $g \in$ $\mathrm{PSL}_2(\mathbb{R})$ such that all matrices in $E(Q(gP))$ have nonvanishing upper-left corners.*

Proof. We use translations $T_x = \left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right)$ with $x \in \mathbb{R}$. For any $h = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{R})$ we calculate that

$$T_x h T_x^{-1} = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a+cx & -cx^2+(d-a)x+b \\ c & d-cx \end{pmatrix}.$$

Suppose $E(Q(P))$ consists of $\left(\begin{smallmatrix} a_i & b_i \\ c_i & d_i \end{smallmatrix}\right) \in \mathrm{PSL}_2(\mathbb{R})$, $1 \leq i \leq K$. Note that a_i and c_i cannot be both zero, we choose nonzero x such that $a_i + c_i x \neq 0$ for all $i = 1, \ldots, K$. For such x we have $E(Q(T_x P)) = T_x E(Q(P)) T_x^{-1}$ consisting of matrices with nonvanishing upper-left corners. $\square$

Remark 2. For any finite set of points in the upper-half plane, we may also dilate points by hyperbolic isometries so that they all have sufficiently large absolute values. Note that a Möbius transformation $\left(\begin{smallmatrix} 0 & b \\ c & d \end{smallmatrix}\right) \cdot z = \frac{b}{cz+d}$ basically inverts the absolute value of z, so that it cannot map z with large absolute values to points with large absolute values. Thus after dilation, Möbius transformations with vanishing upper-left corners do not occur as isometries in consideration.

Hence without loss of generality, we assume $p_x,\ p_y,\ q_x,\ q_y \gg 1$ for points $p = px + ip_y$, $q = q_x + q_y$ in consideration, that is, far away in the first quadrant. We have the following observation through (3). First, each S_{pq} is a projective line in $\mathbb{P}^3 \supset G = \mathrm{PSL}_2(\mathbb{R})$. We use the natural manifold atlas

$$\mathbb{P}^3 = \mathbb{R}^3_1 \cup \mathbb{R}^3_2 \cup \mathbb{R}^3_3 \cup \mathbb{R}^3_4,$$

with $\mathbb{R}^3_1 = \{[1 : b : c : d] \mid b, c, d \in \mathbb{R}\} \simeq \mathbb{R}^3$ and $\mathbb{R}^3_i \simeq \mathbb{R}^3$, $i = 2, 3, 4$, similarly defined with i-th entry equal to 1 in the projective coordinate. Analogously we use

$$G = \bigcup_{i=1}^{4} G_i, \quad G_i = \mathrm{PSL}_2(\mathbb{R}) \cap \mathbb{R}^3_i.$$

In particular, G_1 consists of matrices with nonvanishing upper-left corners. Then the restriction $S_{pq} \cap G_i$ becomes a real line in $\mathbb{R}^3_i$, and by Proposition 2.2, there exists $g \in G$ such that $G_k(gP) \subset G_1$ for each $k \geq 2$. Abusing notation, we always denote by L_{pq} the real line $S_{(gp)(gq)} \cap \mathbb{R}^3_1$ in the manifold atlas of $\mathbb{P}^3$. The incidences among curves S_{pq} are now equivalent to that of lines L_{pq} in $\mathbb{R}^3$ ($\mathbb{R}^3_1$). Explicitly L_{pq} has the following linear parametrization.

Proposition 2.3. *For any $p = p_x + ip_y$, $q = q_x + iq_y \in \mathbb{H}^2$, the line L_{pq} can be parametrized as*

$$(5) \quad \left(-\frac{q_y(p_x^2 + p_y^2) + p_y(q_x^2 + q_y^2)}{p_x q_y + q_x p_y}, \frac{p_y + q_y}{p_x q_y + q_x p_y}, 0\right) + t\left(\frac{p_y(q_x^2 + q_y^2)}{p_x q_y + q_x p_y}, -\frac{q_y}{p_x q_y + q_x p_y}, 1\right),$$

for $t \in \mathbb{R}$.

Proof. For any $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \cdot p = q$ with $a = 1$ and $t = d + 1$ as parameter, we get from (3),

$$(6) \quad \begin{aligned} b &= -\frac{q_y(p_x^2 + p_y^2) + p_y(q_x^2 + q_y^2)}{p_x q_y + q_x p_y} + \frac{p_y(q_x^2 + q_y^2)}{p_x q_y + q_x p_y} t, \\ c &= \frac{p_y + q_y}{p_x q_y + q_x p_y} - \frac{q_y}{p_x q_y + q_x p_y} t, \end{aligned}$$

which gives us the parametrization of points $(b, c, t) \in L_{pq}$. $\square$

Remark 3. There are other parametrizations of L_{pq}, say for $b = t$ as the parameter. Here the roles of p and q are symmetric in that the intersection of L_{pq} and L_{qp} is on the plane $t = 0$.

Since there are nonlinear terms in our parametrization, which is not a problem for Guth and Katz [4], we have to consider different families of lines that rule surfaces and the vector fields on reguli to get the following.

Proposition 2.4. *For any set of N points $P \subset \mathbb{H}^2_{>0} := \{x + iy : x, y > 0\}$ and $\mathcal{L} = \{L_{pq} : p, q \in P\}$, no more than N lines of $\mathcal{L}$ lie in a common plane and no more than $O(N)$ lines of $\mathcal{L}$ lie in a common regulus.*

Proof. We consider the families $L_q := \{L_{pq}\}_{p\in\mathbb{H}^2_{>0}}$ of lines targeting at q. First, for any $p' \neq p$, the line $L_{p'q}$ does not intersect L_{pq}. Note that $L_{pq} \subset S_{pq}$, and suppose $L_{pq} \cap L_{p'q} \neq \varnothing$. Then there would be some $g \in G$ such that $gp' = gp = q$, a contradiction. Moreover by (5), the directions of L_{pq} and $L_{p'q}$ are different:

$$\left(\frac{p_y(q_x^2+q_y^2)}{p_xq_y+q_xp_y}, -\frac{q_y}{p_xq_y+q_xp_y}, 1\right) = (\xi_1, \xi_2, 1)$$

has a unique solution for fixed q and ξ_1, ξ_2. Thus different L_q's have no lines in common and belong to different rulings of a ruled surface if any. Note that ξ_1, ξ_2 cannot be zero since $p_x, p_y, q_x, q_y > 0$. Indeed, equivalently we have

$$\begin{pmatrix} -\xi_1 q_y & q_x^2+q_y^2-q_x\xi_1 \\ \xi_2 q_y & \xi_2 q_x \end{pmatrix} \begin{pmatrix} p_x \\ p_y \end{pmatrix} = \begin{pmatrix} 0 \\ -q_y \end{pmatrix},$$

whose associate matrix has determinant $-(q_x^2+q_y^2)\xi_2 q_y \neq 0$. Hence lines of L_q are pairwise skew and no two of its lines lie in a common plane. Therefore any plane intersects each L_q at most one line and intersects $\mathcal{L}$ at most N lines.

To prove the second part, we construct a vector field $V = (V_1, V_2, V_3)$ on $\mathbb{R}^3$ tangent to lines of L_q for any fixed $q = q_x + iq_y \in \mathbb{H}^2_{>0}$. By (3) we locate p such that L_{pq} passes through any given $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ as follows ($a = 1$, $x_1 = b$, $x_2 = c$, $x_3 = d$):

$$\begin{aligned} p_x + x_1 + (p_yq_y - p_xq_x)x_2 - q_xx_3 &= 0, \\ p_y - (p_xq_y + q_xp_y)x_2 - q_yx_3 &= 0, \end{aligned}$$

or equivalently,

$$\begin{aligned} (1 - q_xx_2)p_x + (q_yx_2)p_y &= q_xx_3 - x_1, \\ (-q_yx_2)p_x + (1 - q_xx_2)p_y &= q_yx_3, \end{aligned}$$

which has solution

$$\begin{pmatrix} p_x \\ p_y \end{pmatrix} = \frac{1}{(1-q_xx_2)^2 + q_y^2x_2^2} \begin{pmatrix} q_xx_1x_2 - (q_x^2+q_y^2)x_2x_3 - x_1 + q_xx_3 \\ -q_yx_1x_2 + q_yx_3 \end{pmatrix}.$$

By (5), we set the direction of L_{pq} as

$$((q_x^2+q_y^2)p_y, -q_y, q_yp_x + q_xp_y) = \frac{1}{(1-q_xx_2)^2 + q_y^2x_2^2}(V_1, V_2, V_3),$$

where

$$\begin{aligned}
V_1 &= -q_y(q_x^2+q_y^2)(x_1x_2-x_3),\\
V_2 &= -q_y[(1-q_xx_2)^2+q_y^2x_2^2],\\
V_3 &= -q_y(q_x^2+q_y^2)x_2x_3-q_yx_1+2\,q_xq_yx_3.
\end{aligned}$$

Let $V=(V_1, V_2, V_3)$; then V has degree 2. Note that $p\in\mathbb{H}^2_{>0}$, the vector field is defined over the open subset

$$U_q := \{(x_1,x_2,x_3)\in\mathbb{R}^3 \mid q_xx_1x_2-(q_x^2+q_y^2)x_2x_3-x_1+q_xx_3>0,\ -q_yx_1x_2+q_yx_3>0\},$$

and we always consider the pieces of reguli restricted in U_q.

Now suppose a line L_{pq} lies in a regulus R defined by a degree-2 irreducible polynomial f in $\mathbb{R}^3$. Then at any point $x\in L_{pq}$ we have the Taylor expansion

$$f(x+tV(x)) = f(x)+\nabla(f)\cdot V(x)t+\tfrac{1}{2}V^TH(f)Vt^2,$$

where $\nabla(f)$ is the gradient of f and $H(f)$ is the Hessian matrix of f.

By Bezout's lemma (Lemma 3.1 of [4]), if more than 9 lines of L_q are contained in R, f would have a common factor with both $\nabla(f)\cdot V$ and $V^TH(f)V$, which have degree 3 and 4, respectively. By irreducibility, f must be the common factor so that f vanishes on each line of L_q with direction $V(x)$ for any $x\in R$ by the Taylor expansion above, that is, L_q is a ruling of R. Since a regulus has only two rulings, R can only contain at most 8 lines from $N-2$ families L_q which are not rulings of R and $2N$ lines of L_{q_1}, L_{q_2} if they are rulings of R. In total, there are at most $2N+8(N-1)=10N-8$ lines of $\mathcal{L}$ lying in R. □

Now we already reduced the problem to incidence geometry in the Euclidean space. Applying ruled surface theory and polynomial partitioning to reproduce Guth and Katz's Theorem 2.10 and 2.11 of [4], we get the following lower bound for the distinct distances problem in the hyperbolic plane. It has the same strength as the result of Guth and Katz for the Euclidean plane.

Theorem 2.5. *For $P\subset\mathbb{H}^2$ any set of N points and $\mathcal{L}=\{L_{pq}\mid p,q\in P\}$, let G_k be the set of points where at least k lines of $\mathcal{L}$ meet for $2\le k\le N$. Then*

$$|G_k| \lesssim N^3k^{-2}.$$

Consequently, by (4), $|Q(P)|\lesssim N^3\log N$, *and by* (2), *we have* $|d(p)|\gtrsim N/\log N$.

Acknowledgements

We are thankful for helpful discussions with Harald Helfgott. We are also thankful for the comments from Misha Rudnev. Both authors were partially supported by the Humboldt Professorship of Professor Helfgott while completing the work.

Currently, Lu is supported by Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515110206) and 2023 Shenzhen National Science Foundation (No. 20231128220938001). Meng is supported by the National Natural Science Foundation of China (Grant No. 12201346) and Shandong Provincial Foundation (No. 2022HWYQ-046 & No. ZR2022QA001).

References

[1] X. Du, L. Guth, and X. Li, "A sharp Schrödinger maximal estimate in $\mathbb{R}^2$", *Ann. of Math.* (2) **186**:2 (2017), 607–640. MR Zbl

[2] G. Elekes and M. Sharir, "Incidences in three dimensions and distinct distances in the plane", *Combin. Probab. Comput.* **20**:4 (2011), 571–608. MR Zbl

[3] P. Erdős, "On sets of distances of n points", *Amer. Math. Monthly* **53**:5 (1946), 248–250. Zbl

[4] L. Guth and N. H. Katz, "On the Erdős distinct distances problem in the plane", *Ann. of Math.* (2) **181**:1 (2015), 155–190. MR Zbl

[5] T. G. F. Jones, "New results for the growth of sets of real numbers", *Discrete Comput. Geom.* **54**:4 (2015), 759–770. MR Zbl

[6] H. Kaplan, J. Matoušek, and M. Sharir, "Simple proofs of classical theorems in discrete geometry via the Guth–Katz polynomial partitioning technique", *Discrete Comput. Geom.* **48**:3 (2012), 499–517. MR Zbl

[7] F. Klein, "Vergleichende betrachtungen über neuere geometrische forschungen", *Math. Ann.* **43**:1 (1893), 63–100. Translated in *Bull. Amer. Math. Soc.* **2**:10 (1893), 215–249. MR Zbl

[8] O. Roche-Newton and M. Rudnev, "On the Minkowski distances and products of sum sets", *Israel J. Math.* **209**:2 (2015), 507–526. MR Zbl

[9] M. Rudnev and J. M. Selig, "On the use of the Klein quadric for geometric incidence problems in two dimensions", *SIAM J. Discrete Math.* **30**:2 (2016), 934–954. MR Zbl

[10] A. Sheffer and J. Zahl, "Distinct distances in the complex plane", *Trans. Am. Math. Soc.* **374**:9 (2021), 6691–6725. MR Zbl

[11] T. Tao, "Lines in the Euclidean group SE(2)", blog post, 2011, http://terrytao.wordpress.com/2011/03/05/lines-in-the-euclidean-group-se2/.

Received September 4, 2023. Revised January 19, 2024.

ZHIPENG LU
GUANGDONG LABORATORY OF MACHINE PERCEPTION AND INTELLIGENT COMPUTING
DEPARTMENT OF ENGINEERING
SHENZHEN MSU-BIT UNIVERSITY
SHENZHEN
CHINA
zhipeng.lu@hotmail.com

XIANCHANG MENG
SCHOOL OF MATHEMATICS
SHANDONG UNIVERSITY
JINAN
CHINA
xianchang.meng@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 327, No. 1, 2023
https://doi.org/10.2140/pjm.2023.327.139

THE ALGEBRAIC TOPOLOGY OF 4-MANIFOLD MULTISECTIONS

Delphine Moussard and Trenton Schirmer

A multisection of a 4-manifold is a decomposition into 1-handlebodies intersecting pairwise along 3-dimensional handlebodies or along a central closed surface; this generalizes the Gay–Kirby trisections. We show how to compute the twisted absolute and relative homology, the torsion and the equivariant intersection form of a 4-manifold from a multisection diagram. The homology and torsion are given by a complex of free modules defined by the diagram and the intersection form is expressed in terms of the intersection form on the central surface. We give efficient proofs, with very few computations, thanks to a retraction of the (possibly punctured) 4-manifold onto a CW-complex determined by the multisection diagram. Further, a multisection induces an open book decomposition on the boundary of the 4-manifold; we describe the action of the monodromy on the homology of the page from the multisection diagram.

1. Introduction and main results

A trisection is a type of combinatorial structure on 4-manifolds which was discovered by Gay and Kirby [2016] via Morse 2-functions. They proved that any smooth 4-manifold, possibly with boundary, can be decomposed as the union of three 4-dimensional 1-handlebodies, with 3-dimensional 1-handlebodies as pairwise intersections and a compact surface as global intersection. Such a trisection can be described by a diagram, namely the central surface with collections of curves that define the 3-dimensional pieces. A trisection diagram determines a smooth 4-manifold up to diffeomorphism, so that one should be able to read topological invariants of the manifold on the diagram. In the setting of closed 4-manifolds, Feller, Klug, Schirmer and Zemke [Feller et al. 2018] provided a computation of the homology and intersection form of the manifold from a trisection diagram, and Florens and Moussard [2022] derived the twisted homology and torsion, and the equivariant intersection form. Following these papers, Tanimoto [2023] computed the homology of 4-manifolds with connected boundary. Here we recover and

MSC2020: primary 57K41; secondary 57K40.
Keywords: trisection, trisection diagram, homology of 4-manifolds, intersection form.

generalize these results, computing from a diagram the twisted absolute and relative homology and torsion and the equivariant intersection form for any trisected 4-manifold with boundary. Moreover, we work with “multisections” in the sense of Islambouli and Naylor [2024], namely a cyclic decomposition of the manifold into any number of 4-dimensional 1-handlebodies, where successive pieces meet along 3-dimensional 1-handlebodies while nonsuccessive ones meet along the central surface. We propose a more efficient approach. While Feller, Klug, Schirmer and Zemke worked with a handle decomposition of the manifold underlying the trisection, Florens and Moussard directly used the datum of the trisection. This last method reduced the homological computations, but the computation of torsion was quite intricate. Here we consider a deformation-retraction of the (possibly punctured) manifold onto a CW-complex associated with the multisection diagram. This simplifies the computations and provides the torsion “for free”. This retraction could be useful for further computations of homological or homotopical invariants.

A multisection of a 4-manifold X with boundary induces an open book decomposition on the boundary. The monodromy of this open book has been described algorithmically by Castro, Gay and Pinzón-Caicedo [Castro et al. 2018a] from a diagram. Here we derive the action of the monodromy on the homology of the page from which can be derived a computation of the homology of ∂X as well as the Alexander module of the binding determined by the monodromy.

For 4-manifolds with boundary, the handlebodies of a multisection inherit (hyper) compression bodies structures related to the way they intersect the boundary of the manifold.

Definition 1.1. A *compression body* C is a cobordism from a compact orientable surface $\partial_- C$ to a connected compact orientable surface $\partial_+ C$ which is constructed using only 1-handles. Likewise a *hyper compression body* V is a cobordism from a compact orientable 3-manifold $\partial_- V$ to a connected compact orientable 3-manifold $\partial_+ V$ constructed using only 1-handles. A *lensed* (hyper) compression body is then obtained by collapsing the vertical boundary of the cobordism so that the boundary of $\partial_+ C$ ($\partial_+ V$) becomes identified with the boundary of $\partial_- C$ ($\partial_- V$).

In the case that $\partial_- C = \varnothing$, it is understood at C is built using only 1-handles attached to a single 0-handle. A (lensed) compression body is *trivial* if $\partial_- C \cong \partial_+ C$. This means it is just a thickened surface $S \times I$, or if lensed, it is obtained from $S \times I$ by collapsing the I-fibers of $\partial S \times I$.

Definition 1.2. A *multisection* of a compact orientable 4-manifold X is a decomposition $X = X_1 \cup \cdots \cup X_n$ into 4-dimensional 1-handlebodies X_i with the following properties (all arithmetic involving indices is mod n):

(1) Each X_i has a lensed hyper compression body structure such that $\partial_- X_i = X_i \cap \partial X$, and if $\partial X \neq \varnothing$, there is a fixed surface Σ_∂ such that, for all $1 \leq i \leq n$,

$\partial_- X_i$ is diffeomorphic to the trivial lensed compression body obtained by pinching the vertical boundary of $\Sigma_\partial \times I$.

(2) $\Sigma = \bigcap_{i=1}^n X_i$ is a compact connected orientable surface.

(3) $C_i = X_i \cap X_{i+1}$ is a 3-dimensional 1-handlebody with a lensed compression body structure satisfying $\partial_+ C_i = \Sigma$ and $\partial_- C_i = C_i \cap \partial X \cong \Sigma_\partial$ for all i.

(4) $X_i \cap X_j = \Sigma$ when $|i-j| > 1$.

A multisection is called a *trisection* when $n = 3$.

The condition that the C_i are 1-handlebodies implies that Σ is closed if and only if Σ_∂ is closed. We shall consider the case when Σ_∂ contains no closed components, and within this context Σ will be closed if and only if X itself is closed. This is the framework of most of the literature on trisections and within this framework a unified calculation of the algebraic topology is possible. The specific case where more general compression bodies are allowed, i.e., the case in which Σ_∂ has closed components, was considered in the original paper of Gay and Kirby [2016]; however, the calculations become more delicate and require special treatment. Moreover in the case that Σ_∂ contains components that are spheres, diagrams no longer determine a unique 4-manifolds up to diffeomorphism. We postpone the homology computations in this case to a forthcoming publication in order to avoid the extra complications here.

In the case that $\partial X \neq \varnothing$, it is also to be understood that for all $i \mod n$, $\partial_- X_i$ is parametrized as $\Sigma_\partial \times I/\sim$ in such a way that $\partial_- C_{i-1} = \Sigma_\partial \times \{0\}$ and $\partial_- C_i = \Sigma_\partial \times \{1\}$. Thus, the multisection induces an open book decomposition on ∂X with page Σ_∂.

We fix once and for all a multisected manifold $X = \bigcup_{1 \le i \le n} X_i$, and set $C_i = X_i \cap X_{i+1}$ and $\Sigma = \bigcap_i X_i$.

Definition 1.3. Let C be a compression body. A *defining collection of disks* for C is a collection $\mathcal{D}$ of disks properly embedded in C such that $C \setminus \eta(\mathcal{D})$ is a thickening of $\partial_- C$ (for instance the cocore disks of the 1-handles in the definition). The boundary $\partial\mathcal{D} \subset \partial_+ C$ is a *defining collection of curves* for C.

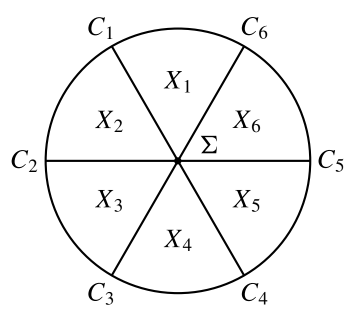


Figure 1. Schematic of a multisection.

Definition 1.4. A *diagram* of the multisection $X=\bigcup_{1\le i\le n} X_i$ is a tuple $(\Sigma; c_1,\dots,c_n)$ where c_i is a defining collection of curves for C_i.

A multisection diagram determines a unique smooth 4-manifold [Castro et al. 2018b]. The structure of the X_i gives some constraints on the curves of a multisection diagram. For each i, X_i is obtained from a thickened $\partial_- X_i$ by attaching 1-handles, so that $\partial_+ X_i \cong (S^2\times S^1)^{\sharp k} \# \big(\#\,\partial_- X_i\big)$, where k is the number of 1-handles in excess of the minimum required to connect $\partial_- X_i$, and $\#\,\partial_- X_i$ is the connected sum of all components of $\partial_- X_i$. Now Definition 1.2 implies that $C_{i-1}\cup_\Sigma C_i$ is a sutured Heegaard splitting of $\partial_+ X_i$, so that the Heegaard diagram $(\Sigma; c_{i-1}, c_i)$ is always handleslide-diffeomorphic to a standard diagram as represented in Figure 2.

Fix a homomorphism $\varphi : \mathbb{Z}[\pi_1(X)] \to R$, where R is a commutative ring. We shall express the absolute and relative homology of X, twisted by φ, in terms of the multisection diagram. Fix a point $*\in \mathrm{Int}(\Sigma)$ and let L_i^φ be the submodule of $H_1^\varphi(\Sigma, *)$ generated by the homology classes of the curves in c_i. In Section 3, we obtain the following result (Theorem 3.8, Remark 3.9 and Lemma 3.11).

Theorem 1.5. *The homology of X is given by the chain complex of free R-modules*

$$(\mathcal{C}) \qquad 0\to \bigoplus_{i=1}^{n} (L_{i-1}^\varphi \cap L_i^\varphi) \xrightarrow{\partial_2} \bigoplus_{i=1}^{n} L_i^\varphi \xrightarrow{\partial_1} H_1^\varphi(\Sigma, *) \xrightarrow{\partial_0} H_0^\varphi(*),$$

where

$$\partial_2((x_i)_{1\le i\le n}) = (x_i - x_{i+1})_{1\le i\le n} \quad \textit{and} \quad \partial_1((x_i)_{1\le i\le n}) = \sum_{i=1}^{n} x_i.$$

Moreover, if R is a field, an explicit complex basis of $(\mathcal{C})$ can be given such that $\tau^\varphi(X;h) = \tau(\mathcal{C}; b, h)$.

Let Σ' be the surface Σ with a small open disk removed, such that the point $*$ lies on the boundary of the removed disk. For $1\le i\le n$, let $\mathcal{J}_i^\varphi$ be the orthogonal complement in $H_1^\varphi(\Sigma', \partial\Sigma)$ of L_i^φ with respect to the equivariant intersection pairing on $H_1^\varphi(\Sigma, *) \times H_1^\varphi(\Sigma', \partial\Sigma)$. We prove the following in Section 4 (Theorem 4.9, Lemma 4.6 and Remark 4.10).

Theorem 1.6. *If $\partial X \ne \varnothing$, the twisted homology of $(X, \partial X)$ is given by the chain complex of free R-modules*

$$(\mathcal{C}_\partial) \qquad H_2^\varphi(\Sigma, \Sigma') \xrightarrow{\partial_3} \bigoplus_{i} (\mathcal{J}_{i-1}^\varphi \cap \mathcal{J}_i^\varphi) \xrightarrow{\partial_2} \bigoplus_{i} \mathcal{J}_i^\varphi \xrightarrow{\partial_1} H_1^\varphi(\Sigma', \partial\Sigma) \to 0,$$

where

$$\partial_3([\Sigma]) = [\partial(\Sigma\setminus\Sigma')], \quad \partial_2((x_i)_{1\le i\le n}) = ((x_i - x_{i+1})_{1\le i\le n}), \quad \partial_1((x_i)_{1\le i\le n}) = \sum_{i=1}^{n} x_i.$$

Moreover, if R is a field, an explicit complex basis of $(\mathcal{C}_\partial)$ can be given such that $\tau^\varphi(X, \partial X; h) = \tau(\mathcal{C}_\partial; b, h)$.

When Σ is closed, we define the L_i^φ in $H_1^\varphi(\Sigma', *)$. In this closed case, these are lagrangians, namely they are their own orthogonal complement with respect to the intersection form. The next result is obtained in Section 6 (Theorem 6.4, Remark 6.5 and Lemma 6.2).

Theorem 1.7. *If X is closed, the twisted homology of X is given by the chain complex of free R-modules*

$$(\bar{\mathcal{C}})\qquad H_2^\varphi(\Sigma,\Sigma') \xrightarrow{\partial_3} \bigoplus_i (L_{i-1}^\varphi \cap L_i^\varphi) \xrightarrow{\partial_2} \bigoplus_i L_i^\varphi \xrightarrow{\partial_1} H_1^\varphi(\Sigma', *) \to H_0^\varphi(*),$$

where

$$\partial_3([\Sigma]) = [\partial\Sigma'], \quad \partial_2((x_i)_{1\le i\le n}) = ((x_i - x_{i+1})_{1\le i\le n}), \quad \partial_1((x_i)_{1\le i\le n}) = \sum_{i=1}^n x_i.$$

Moreover, if R is a field, an explicit complex basis of $(\bar{\mathcal{C}})$ can be given such that $\tau^\varphi(X;h) = \tau(\bar{\mathcal{C}}; b, h)$.

These three theorems allow us to represent homology classes by mainly explicit chains in the multisected manifold which meet transversely along copies of the central surface Σ. This provides a simple description of the intersection form on X (Theorems 5.1 and 6.6).

Theorem 1.8. *Suppose $h_1 = [(x_i)_{1\le i\le n}]$ and $h_2 = [(y_i)_{1\le i\le n}]$ in $H_2^\varphi(X)$, where $(x_i)_{1\le i\le n}, (y_i)_{1\le i\le n} \in \bigoplus_i L_i^\varphi$. Then*

$$\langle h_1, h_2\rangle_X^\varphi = \sum_{1\le i<j\le n} \langle x_i, y_j\rangle_\Sigma^\varphi,$$

where $\langle\,\cdot\,,\cdot\,\rangle_X^\varphi$ and $\langle\,\cdot\,,\cdot\,\rangle_\Sigma^\varphi$ are the equivariant intersection forms on $H_2^\varphi(X)$ and $H_1^\varphi(\Sigma, *)$ respectively ($H_1^\varphi(\Sigma', *)$ if X is closed).

The intersection pairing on $H_2^\varphi(X) \times H_2^\varphi(X, \partial X)$ is similar (Theorem 5.3). In odd dimensions, the intersection pairings are especially simple (Theorems 5.4 and 6.6).

Theorem 1.9. *Suppose that either $h_1 \in H_1^\varphi(X)$ corresponds to the element $a \in H_1^\varphi(\Sigma, *)$ and $h_2 \in H_3^\varphi(X, \partial X)$ corresponds to the element $b \in \bigcap_i \mathcal{J}_i^\varphi$, or $h_1 \in H_1^\varphi(X, \partial X)$ corresponds to the element $a \in H_1^\varphi(\Sigma', \partial\Sigma)$ and $h_2 \in H_3^\varphi(X)$ corresponds to the element $b \in \bigcap_i L_i^\varphi$ ($a \in H_1^\varphi(\Sigma', \partial\Sigma)$ and $b \in \bigcap_i L_i^\varphi$ if X is closed). Then*

$$\langle h_1, h_2\rangle_X^\varphi = \langle a, b\rangle_\Sigma^\varphi.$$

Plan of the paper. In Section 2, we recall the definitions of twisted homology, torsion and equivariant intersection pairing. Our discussion is somewhat discursive to help readers build intuition. Sections 3 and 4 are devoted to the twisted homology and torsion of a 4-manifold with nonempty boundary, respectively absolute and relative. In Section 5, we describe the intersection forms. Section 6 treats the case of a closed 4-manifold. Section 7 deals with the boundary: action in homology

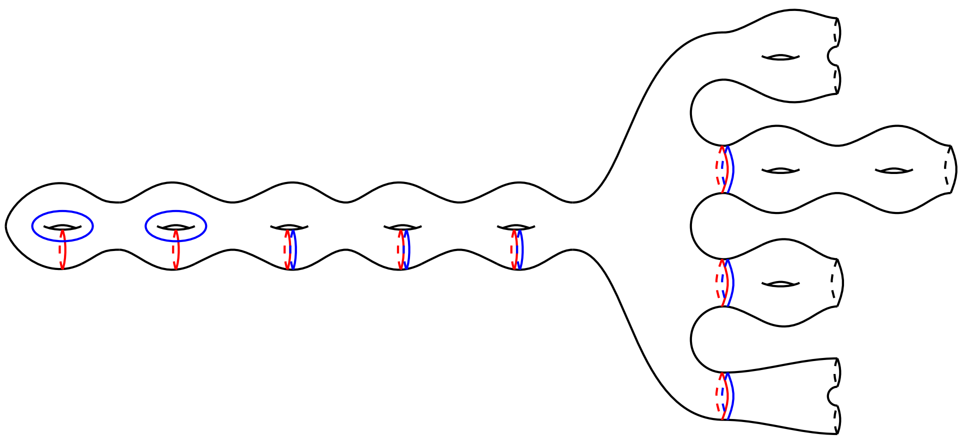


Figure 2. Heegaard diagram for $C_{i-1} \cup C_i$. In this example, C_{i-1} and C_i are constructed with eight 1-handles and X_i with six 1-handles. The manifold X has four boundary components. The components of the page Σ_∂ have a pair (genus, number of boundary components) equal to $(1, 2)$, $(2, 1)$, $(1, 1)$ and $(0, 2)$.

of the monodromy of the open book and homology of the boundary. Finally, in Section 8, we treat some examples.

Conventions. The notation we set above is assumed to be fixed for the remainder of the paper. That is, X is always multisected by n hyper compression bodies X_i which meet in compression bodies C_i, all of which are attached radially about the central fiber Σ. Additionally, $Y = C_1 \cup \cdots \cup C_n$ shall be referred to as the *spine* of the multisection. Also, $\varphi : \mathbb{Z}[\pi_1(X)] \to R$ is a homomorphism to a commutative ring R. Throughout the paper, if Z is a subset of a manifold M, $\eta(Z)$ denotes a regular neighborhood of Z in M.

2. Algebraic preliminaries

2A. ***Twisted homology.*** Let $\pi = \pi_1(X)$ and let R be a ring. A group homomorphism $\varphi : \pi \to R^*$ induces a ring homomorphism $\mathbb{Z}[\pi] \to R$. Throughout, both of these homomorphisms shall be denoted by φ and called the "twisting map."

Let $\widetilde{X}$ denote the universal cover of X, and for any $Z \subset X$, let $\widetilde{Z}$ denote the inverse image of Z under the covering map $\widetilde{X} \to X$ ($\widetilde{Z}$ will usually not be the universal cover of Z). Then π acts on both $\widetilde{X}$ and $\widetilde{Z}$ by deck transformations, which induces a left $\mathbb{Z}[\pi]$-module structure on $C_*(\widetilde{X}, \widetilde{Z})$. This allows us to define a chain complex of R-modules

$$C_i^\varphi(X, Z) = R \otimes_\varphi C_i(\widetilde{X}, \widetilde{Z}).$$

The usual boundary maps on $C_*(\widetilde{X}, \widetilde{Z})$ induce $\mathbb{Z}[\pi]$-module morphisms, and the boundary maps of $C_*^{\varphi}(X, Z)$ are then obtained by tensoring with the identity map on R. The resulting homology groups are denoted by $H_*^{\varphi}(X, Z)$.

To understand the structure of the twisted chain groups, observe first that by definition, for any $g \in \pi_1(X)$ and choice of lift $\tilde{e}$ of an i-cell e of X, $1 \otimes (g \cdot [\tilde{e}]) = \varphi(g) \otimes [\tilde{e}]$. It follows from the transitivity of the π action on $\widetilde{X}$ that a choice of lift for every i-cell in X determines an R-basis of $C_i^{\varphi}(X, Z)$, and thus $C_i^{\varphi}(X, Z)$ is always a freely generated R-module of the same rank as the $\mathbb{Z}$-rank of $C_i(X, Z)$, for any twisting map.

The effect of the twisting map is to be found in how the boundary maps are changed, and thereby also the resulting homology groups. Intuitively, one thinks of R as something like a tangent space to each point of X, and multiplication by $\varphi(g)$ corresponds to the monodromy action of g. For example, if a 1-chain e corresponds to the element $g \in \pi$ with its endpoints on the 0-chain v, then in untwisted homology we would have $\partial e = v - v = 0$, but with twisted homology we have $\partial e = \varphi(g)v - v = (\varphi(g) - 1)v$. The choice of lift does not affect the homology because different choices of lift amount to scalar multiplication of a basis element by a unit in R.

For example, if φ is the trivial map $\pi \to R^*$, so that $\varphi(g) = 1$ for every $g \in \pi$, then in this case for all lifts $\tilde{e}_1, \tilde{e}_2$ of a given cell e of X, we have

$$1 \otimes [\tilde{e}_1] = 1 \otimes (g \cdot [\tilde{e}_2]) = \varphi(g) \otimes [\tilde{e}_2] = 1 \otimes [\tilde{e}_2].$$

In other words, all the lifts of e determine the same chain in $C_i^{\varphi}(X, Z)$, and the projection map $\widetilde{X} \to X$ thus induces a chain isomorphism $C_*^{\varphi}(X, Z) \to C_*(X, Z; R)$, where $C_*(X, Z; R)$ is the usual chain complex for the (untwisted) homology with R coefficients.

On the other extreme, if φ is the inclusion $\pi \hookrightarrow \mathbb{Z}[\pi]^*$, then *all* distinct lifts of a cell e to $\widetilde{X}$ determine chains which differ by multiplication by a unit in $\pi \subset \mathbb{Z}[\pi]$. This example where $\varphi = \iota : \pi \to \mathbb{Z}[\pi]^*$ is in a sense universal. For if one can compute matrices which describe the boundary maps of $C_*^{\iota}(X, Z)$ in terms of some fixed cellular basis, then for any other map $\varphi : \pi \to R^*$, one simply substitutes $\varphi(g)$ for every g in the matrices of $C_*^{\iota}(X, Z)$ to obtain matrices of the boundary maps for $C_*^{\varphi}(X, Z)$ with respect to the same basis.

As a simple but instructive example, if $X = S^1$ and $Z = \varnothing$, then we may identify π with the cyclic group generated by t, and

$$C_0^{\iota}(S^1) \cong C_1^{\iota}(S^1) \cong \mathbb{Z}[t, t^{-1}].$$

All other chain groups are trivial as with the untwisted case, and the one nontrivial boundary map is multiplication by $t^n(t-1)$ (where n depends only on the choices of lifts). Therefore $H_1^{\iota}(S^1) \cong 0$ and $H_0^{\iota}(S^1)$ is $\mathbb{Z}$, considered as a $\mathbb{Z}[t, t^{-1}]$-module whose action is given by $P(t) \cdot a = P(1)a$. More generally, given a homomorphism

$\varphi : \pi \to R^*$, where R is a ring,

$$C_0^{\varphi}(S^1) \cong C_1^{\varphi}(S^1) \cong R,$$

and the boundary map is multiplication by $\varphi(t)^n(\varphi(t)-1)$. If $\varphi(t)=1$ (the untwisted case), then

$$H_0^{\varphi}(S^1) \cong H_1^{\varphi}(S^1) = R,$$

but if $(\varphi(t)-1)$ is a unit then the entire twisted homology becomes trivial.

Returning to generalities, if we let $\widehat{X}$ denote the cover of X associated with a normal subgroup $N \subset \ker(\varphi)$, then φ factors through a map $\pi_1(X)/N \to R$, which we also call φ. If $\widehat{Z}$ is the inverse image of Z under the covering map $\widehat{X} \to X$, then the chain complex $C_*(\widehat{X}, \widehat{Z})$ becomes a $\mathbb{Z}[\pi]$-module and if we define the chain groups $R \otimes_{\varphi} C_*(\widehat{X}, \widehat{Z})$, the resulting homology will be the same as $H_*^{\varphi}(X, Z)$ through an isomorphism induced by the subcovering $\widetilde{X} \to \widehat{X}$, similar to how twisted homology reduces to ordinary homology when φ is the trivial map. This observation is important for making geometric sense of long exact sequences in the twisted context. Consider, for example, the long exact sequence associated to the pair (X, Z), which looks like this:

$$\cdots \to H_i^{\varphi}(X) \to H_i^{\varphi}(X, Z) \xrightarrow{f} H_{i-1}^{\varphi\circ\iota}(Z) \xrightarrow{g} H_{i-1}^{\varphi}(X) \to \cdots.$$

Considering $H_{i-1}^{\varphi\circ\iota}(Z)$ in isolation, observe the twisting map is $\varphi \circ \iota$, where $\iota : \pi_1(Z) \to \pi_1(X)$ is induced by inclusion. If the map ι is a surjection, then the inverse image $p^{-1}(Z)$ of Z under the universal cover $p : \widetilde{X} \to X$ will be a subcover of the universal cover $\widetilde{Z} \to Z$. We cannot easily give a geometric interpretation of f if we use the universal cover of Z to define its twisted homology, but if we instead use the cover $p^{-1}(Z)$ of Z to recover $H_{i-1}^{\varphi\circ\iota}(Z)$, then f will be induced on the chain level by taking the boundary of a relative i-cycle in $R \otimes_{\varphi} C_i(\widetilde{X}, \widehat{Z})$, and g will be induced by inclusion as usual. More generally, $p^{-1}(Z)$ may correspond to a disjoint union of copies of the subcover $\widehat{Z} \to Z$ associated with $\ker(\varphi \circ \iota)$. In this case, f first takes the boundary of a relative i-cycle in $(\widetilde{X}, p^{-1}(Z))$ and these, after multiplication by an appropriate element of $\mathbb{Z}[\pi]$, will all be identified with cells in one preferred component of $p^{-1}(Z)$, which we identify with $\widehat{Z}$.

For the reader who would like to build intuition with twisted coefficients, we strongly recommend working out the geometric details of the long exact sequence of the pairs (S^1, v), where v is a point in S^1, and (T^2, G), where T^2 is the 2-torus and G is its standard 1-skeleton with two edges and one vertex. It is similar to working with integral homology; the only complication is that one must understand the topology of appropriate covers to carry out calculations.

2B. ***Torsion.*** We recall the algebraic setup; see [Milnor 1966] and [Turaev 2001] for further details. Let $\mathbb{K}$ be a field. If V is a finite-dimensional $\mathbb{K}$-vector space and b and c are two bases of V, we denote by $[b/c]$ the determinant of the matrix expressing the basis change from b to c. The bases b and c are *equivalent* if $[b/c]=1$. Let $\mathcal{C}$ be a finite complex of finite-dimensional $\mathbb{K}$-vector spaces:

$$\mathcal{C}=(\mathcal{C}_m \xrightarrow{\partial_m} \mathcal{C}_{m-1} \longrightarrow \cdots \xrightarrow{\partial_1} \mathcal{C}_0).$$

A *complex basis* of $\mathcal{C}$ is a family $c=(c_m,\ldots,c_0)$ where c_i is a basis of $\mathcal{C}_i$ for all $i \in \{0,\ldots,m\}$. A *homology basis* of $\mathcal{C}$ is a family $h=(h_m,\ldots,h_0)$ where h_i is a basis of the homology group $H_i(\mathcal{C})$ for all $i \in \{0,\ldots,m\}$. If we have chosen a basis b_j of the space of j-dimensional boundaries $B_j(\mathcal{C})=\operatorname{Im}\partial_{j+1}$ for all $j \in \{0,\ldots,m-1\}$, and a homology basis h of $\mathcal{C}$, this induces a class of bases $(b_i h_i)\bar{b}_{i-1}$ which consists of the elements of b_i, a choice of representatives for h_i, and the image $\bar{b}_{i-1}$ of b_{i-1} under some section of ∂_i. Neither the choice of h_i-representatives nor the choice of section used to define $\bar{b}_{i-1}$ affects the equivalence class of the resulting basis of $\mathcal{C}_i$, because they differ from one another by linear combinations of b_i.

The *torsion* of the $\mathbb{K}$-complex $\mathcal{C}$, equipped with a complex basis c and a homology basis h, is the scalar

$$\tau(\mathcal{C};c,h)=\prod_{i=0}^{m}[(b_i h_i)\bar{b}_{i-1}/c_i]^{(-1)^{i+1}} \in \mathbb{K}^*,$$

where $[(b_i h_i)\bar{b}_{i-1}/c_i]$ denotes the determinant of the change of basis matrix from c_i to $(b_i h_i)\bar{b}_{i-1}$.

This definition does not depend even on the choice of $b_0,\ldots,b_m$, because of the alternating exponent. The value depends only on the choice of c and h. Of course, by making appropriate choices of c and h, we can make the torsion equal to whatever we want (indeed, just multiply one element of c or h by a scalar and you will multiply or divide the entire torsion by that scalar). In practice, $\mathcal{C}$ will be the twisted cellular chain complex associated with the CW-space X, and c_i will be a *geometric* base of the chain groups $R\otimes_\varphi C_i^\varphi(\widetilde{X})$ that is represented by the cells in a lift of the i-skeleton of X to $\widetilde{X}$. Different choices of lift can change the final value of the torsion by an element of $\pm\varphi(\pi)$, so we mod out by this ambiguity and obtain a torsion function $\tau:\mathcal{H}(X)\to\mathbb{K}^*/(\pm\varphi(\pi))$, where $\mathcal{H}(X)$ is the set of all homology bases of the associated (twisted) cellular chain complex of X.

Specifically, assume R is a field, and (X,Z) is a CW-pair. Let $\tilde{c}$ be a basis of the complex of free $\mathbb{Z}[\pi_1(X)]$-modules $C(\widetilde{X},\widetilde{Z})$ obtained by lifting each relative cell of (X,Z) to $\widetilde{X}$. Then $c=\widetilde{c}\otimes 1$ is a *geometric basis* of $C^\varphi(X,Z)$. We need such bases to define the torsion.

Definition 2.1. Given a homology basis h of $H^\varphi(X,Z)$ and a geometric basis c of $C^\varphi(X,Z)$, the torsion of $(X,Z;\varphi)$ is

$$\tau^\varphi(X,Z;h)=\tau(C^\varphi(X,Z);c,h)\in R/\pm\varphi(\pi_1(X)).$$

The remarkable fact is that τ^φ is a topological invariant when $Z=\varnothing$: any two choices of CW-decomposition for X will result in the same torsion. In the case of a true pair ($Z\neq\varnothing$) the torsion remains invariant under CW-subdivision.

Our results below will describe how to pick out geometric bases given the curves on a multisection diagram. In particular, we will describe the curves on the central surface Σ that explicitly correspond to 1-, 2- and 3-cells in the multisected 4-manifold.

2C. ***The equivariant intersection form.*** Let W be a compact oriented m-manifold, R be a commutative ring and $\varphi:\mathbb{Z}[\pi_1(W)]\to R$ be a morphism. Let A and B be disjoint subsets of ∂W. For $q\in\{0,\dots,m\}$, the *equivariant intersection pairing of W relative to A and B* with coefficient in R, introduced by Reidemeister [1939], is the sesquilinear map

$$\langle\,\cdot\,,\,\cdot\,\rangle_W^\varphi : H_q^\varphi(W,A)\times H_{m-q}^\varphi(W,B)\to R$$

defined by

$$\langle[x\otimes r],[x'\otimes r']\rangle_W^\varphi=\sum_{h\in H_1(W)/\ker(\varphi)}\langle x,h.x'\rangle_{\overline{W}}\,\varphi(h)rr',$$

where we are abusing notation slightly by letting φ denote the group homomorphism from $H_1(W)$ into R^*, $\overline{W}\twoheadrightarrow W$ is the covering associated with $\ker\varphi$ and $\langle\,\cdot\,,\,\cdot\,\rangle_{\overline{W}}$ stands for the algebraic intersection in $\overline{W}$. By Blanchfield's duality theorem [1957, Theorem 2.6], if W is smooth, $\varphi(H_1(W))$ is a free multiplicative subgroup of R, and $\partial W=A\sqcup B$, this pairing is nondegenerate on $(H_q^\varphi(W,A)/\mathrm{Tor})\times(H_{m-q}^\varphi(W,B)/\mathrm{Tor})$. The standard (i.e., nonequivariant) intersection pairing is recovered with a trivial twisting map (i.e., $R=\mathbb{Z}$ and $\varphi(\pi_1(W))=\{1\}$).

When $A=B=\varnothing$ and $q=m-q$, the equivariant intersection pairing defines a nondegenerate *equivariant intersection form* on $H_q^\varphi(W)/\mathrm{Tor}$. (In general, if the modules H and K are identified by a canonical isomorphism, a pairing on the product $H\times K$ defines a form on $H\cong K$. A pairing may be considered up to isomorphism of either H or K, while for a form, one may restrict to applying the same isomorphism on both factors. Therefore a form carries more information.)

For a 4-manifold X, the intersection form is standardly defined as a bilinear form on $H^2(X,\partial X)\times H^2(X,\partial X)$ by applying the cup product of two cochains to the fundamental form $[X]\in H_4(X,\partial X)$. Via Poincaré duality, this form can be defined on $H_2(X)\times H_2(X)$ and, as such, it coincides with the above intersection form for a trivial twisting map.

3. The twisted absolute homology groups and torsion

In this section we derive chain complexes for the twisted homology groups $H_*^\varphi(X)$, assuming that $\partial X\neq\varnothing$. Recall that, in this case, $\partial\Sigma\neq\varnothing$.

Definition 3.1. Let V be a hyper compression body. A *defining collection of balls* for V is a collection $\mathcal{B}$ of 3-balls properly embedded in V such that $V \setminus \eta(\mathcal{B})$ is a thickening of $\partial_- V$.

We assume for the remainder of this section that a fixed choice of defining collections $\mathcal{D}_i$ and $\mathcal{B}_i$ of disks and balls has been made for all $1 \leq i \leq n$.

The following lemmas provide a 3-complex onto which X deformation retracts. Our calculations hinge on a careful understanding of how the cells of this complex are mirrored by simple closed curves on Σ. Recall that $Y = C_1 \cup \cdots \cup C_n$ denotes the spine of the multisection.

Lemma 3.2. *The manifold X retracts onto $\Sigma \cup \bigcup_{i=1}^n (\mathcal{D}_i \cup \mathcal{B}_i)$. Further, the quad $(X, Y, \Sigma, *)$ deformation retracts on a CW-complex (Z_3, Z_2, Z_1, Z_0), where $Z_0 = *$, Z_1 is a bouquet of loops defining a basis of $H_1(\Sigma)$, $Z_2 = Z_1 \cup \mathcal{D}$, and $Z_3 = Z_2 \cup \mathcal{B}$.*

Proof. By definition, $X_i \setminus \eta(\mathcal{B}_i) \cong \partial_- X_i \times I$, so each X_i retracts onto $\partial_+ X_i \cup \mathcal{B}_i$, and hence X retracts onto $Y \cup \bigcup_i \mathcal{B}_i$. Likewise each C_i retracts onto $\partial_+ C_i \cup \mathcal{D}_i = \Sigma \cup \mathcal{D}_i$, so that Y further retracts down to $\Sigma \cup \bigcup_i \mathcal{D}_i$. This gives the first assertion. For the second assertion, we get $Z_1 = \Sigma$ and $Z_2 = \Sigma \cup \mathcal{D}$ and we conclude by further retracting Σ. $\square$

Corollary 3.3. *The twisted homology of X is the homology of the following complex*:

$$(\mathcal{C}') \qquad 0 \to H_3^\varphi(X, Y) \to H_2^\varphi(Y, \Sigma) \to H_1^\varphi(\Sigma, *) \to H_0^\varphi(*).$$

Proof. Lemma 3.2 shows that the complex above is isomorphic to the cellular homology complex

$$0 \to H_3^\varphi(Z_3, Z_2) \to H_2^\varphi(Z_2, Z_1) \to H_1^\varphi(Z_1, Z_0) \to H_0^\varphi(Z_0)$$

via the map induced by the inclusion $Z_3 \hookrightarrow X$, which is a simple homotopy. $\square$

Definition 3.4. Let L_i^φ denote the submodule of $H_1^\varphi(\Sigma, *)$ generated by the twisted homology classes of the components of c_i.

Following the approach of [Florens and Moussard 2022], we shall express the complex $(\mathcal{C}')$ in terms of these submodules. They have the following homological interpretation.

Lemma 3.5. *The module L_i^φ naturally identifies with the kernel of the inclusion map $\iota_* : H_1^\varphi(\Sigma) \to H_1^\varphi(C_i)$.*

Proof. Since the components of c_i bound disks in C_i, it is clear that $L_i^\varphi \subset \ker(\iota_*)$; since these disks cut C_i into a thickened $\partial_- C_i$, the reverse inclusion follows. $\square$

Lemma 3.6. $$H_2^\varphi(C_i, \Sigma) \cong L_i^\varphi \quad \textit{for all } i.$$

Proof. $H_1^\varphi(C_i, \Sigma) = 0$ because C_i deformation-retracts onto $\Sigma \cup \mathcal{D}_i$, and thus the exact sequence of the pair (C_i, Σ) gives $H_2^\varphi(C_i, \Sigma) \cong \ker(H_1^\varphi(\Sigma) \to H_1^\varphi(C_i))$. $\square$

Lemma 3.7. $H_3^\varphi(X_i, C_{i-1} \cup C_i) \cong L_{i-1}^\varphi \cap L_i^\varphi$ *for all i.*

Proof. Since X_i is a 4-dimensional 1-handlebody, its order 2 and 3 homology is trivial, and the exact sequence of the pair $(X_i, C_{i-1} \cup C_i)$ gives $H_3^\varphi(X_i, C_{i-1} \cup C_i) \cong H_2^\varphi(C_{i-1} \cup C_i)$. Now the exact sequence of the pair $(C_{i-1} \cup C_i, \Sigma)$ gives

$$0 \to H_2^\varphi(C_{i-1} \cup C_i) \xrightarrow{\iota} H_2^\varphi(C_{i-1}, \Sigma) \oplus H_2^\varphi(C_i, \Sigma) \xrightarrow{\pi} H_1^\varphi(\Sigma),$$

where the identification $H_2^\varphi(C_{i-1} \cup C_i, \Sigma) \cong H_2^\varphi(C_{i-1}, \Sigma) \oplus H_2^\varphi(C_i, \Sigma)$ follows from the Mayer–Vietoris sequence associated to the decomposition of the pair $(C_{i-1} \cup C_i, \Sigma)$ into (C_{i-1}, Σ) and (C_i, Σ). Now, the map π is the difference of the maps $H_2^\varphi(C_{i-1}, \Sigma) \to H_1^\varphi(\Sigma)$ and $H_2^\varphi(C_i, \Sigma) \to H_1^\varphi(\Sigma)$ given by the exact sequences of the pairs (C_{i-1}, Σ) and (C_i, Σ), which give the identifications $H_2^\varphi(C_{i-1}, \Sigma) \cong L_{i-1}^\varphi$ and $H_2^\varphi(C_i, \Sigma) \cong L_i^\varphi$ of Lemma 3.6. It follows that the kernel of π, and thus the image of ι, identifies with the intersection $L_{i-1}^\varphi \cap L_i^\varphi$. □

Theorem 3.8. *The homology of X is given by the chain complex*

$$(\mathcal{C}) \qquad 0 \to \bigoplus_{i=1}^{n} (L_{i-1}^\varphi \cap L_i^\varphi) \xrightarrow{\partial_2} \bigoplus_{i=1}^{n} L_i^\varphi \xrightarrow{\partial_1} H_1^\varphi(\Sigma, *) \xrightarrow{\partial_0} H_0^\varphi(*),$$

where $\partial_2((x_i)_{1 \le i \le n}) = (x_i - x_{i+1})_{1 \le i \le n}$ and $\partial_1((x_i)_{1 \le i \le n}) = \sum_{i=1}^n x_i$. Moreover, if R is a field, the complex basis b of $(\mathcal{C})$ described in Remark 3.9 forms a geometric basis for the torsion of X, meaning that $\tau^\varphi(X; h) = \tau(\mathcal{C}; b, h)$.

Proof. Since $H_2^\varphi(Y, \Sigma) \cong \bigoplus_i H_2^\varphi(C_i, \Sigma)$ and $H_3^\varphi(X, Y) \cong \bigoplus_i H_3^\varphi(X_i, C_{i-1} \cup C_i)$, we can conclude with Corollary 3.3 and Lemmas 3.6 and 3.7. See Remark 3.9 for the explication of the geometric bases. □

Remark 3.9. The geometric bases of $\mathcal{C}'$ are images of cellular bases under the map induced by inclusion $Z_3 \hookrightarrow X$. The maps described in Lemmas 3.6 and 3.7 then define an isomorphism from $\mathcal{C}'$ to $\mathcal{C}$, and the geometric bases of $\mathcal{C}$ are then the images of geometric bases of $\mathcal{C}'$ under this map. This yields the following more concrete description of what the geometric bases b look like for $\mathcal{C}$:

- b_0 is given by the basepoint $*$,
- b_1 is defined by any set of loops on which Σ retracts,
- b_2 is any basis corresponding to a tuple of defining curves $(c_i)_{1 \le i \le n}$,
- b_3 is any basis corresponding to a tuple of "double curves" for the pairs (c_{i-1}, c_i).

By a "double curve" for a pair (c_{i-1}, c_i), we mean any curve on Σ which simultaneously bounds disks in C_{i-1} and C_i. The constraints on multisection diagrams imply that $L_{i-1}^\varphi \cap L_i^\varphi$ admits bases represented by double curves; see Figure 2.

It might not be easy to find a system of double curves from a diagram, since it implies some possibly unobvious handleslides. It is not necessary in this algebraic computation; see Remark 3.12.

Corollary 3.10. *We have the expressions*

$$H_1^{\varphi}(X) \cong H_1^{\varphi}(\Sigma)/\Big(\bigoplus_i L_i^{\varphi}\Big), \quad H_3^{\varphi}(X) \cong \bigcap_i L_i^{\varphi},$$

where we slightly abuse notation by viewing $L_i^{\varphi} \subset H_1^{\varphi}(\Sigma) \subset H_1^{\varphi}(\Sigma, *)$.

Proof. For H_1, the pair $(\Sigma, *)$ gives $H_1^{\varphi}(\Sigma) \cong \ker(H_1^{\varphi}(\Sigma, *) \to H_0^{\varphi}(*))$. $\square$

A satisfying point in Theorem 3.8 is that the modules of the complex $(\mathcal{C})$ are free.

Lemma 3.11. *The modules* $H_1^{\varphi}(\Sigma, *)$ *and* L_i^{φ} *are free* R*-modules of respective ranks* $2g+b-1$ *and* p, *where* g *is the genus of* Σ, b *is its number of boundary components and* p *is the number of curves in each collection* c_i. *The modules* $L_{i-1}^{\varphi} \cap L_i^{\varphi}$ *are also free, and their rank does not depend on* R *and* φ.

Proof. Since $\partial\Sigma \neq \varnothing$, Σ deformation retracts onto a bouquet of $2g+b-1$ loops with central vertex $*$. Hence $C_1^{\varphi}(\Sigma, *) \cong R^{2g+b-1}$ is the only nontrivial twisted chain module of $(\Sigma, *)$ and $H_1^{\varphi}(\Sigma, *) \cong R^{2g+b-1}$. The retraction can be chosen so that the components of c_i are loops of the bouquet, and hence L_i^{φ} is a free submodule of $H_1^{\varphi}(\Sigma, *)$ with basis given by the classes of these components. Moreover, up to handleslide, we can assume the components of c_{i-1} and c_i are in standard position (see Figure 2), so that a basis of $L_{i-1}^{\varphi} \cap L_i^{\varphi}$ is given by the parallel curves in these collections. $\square$

Remark 3.12. We can now explain how to simplify the computation of torsion, avoiding the explicit exhibition of systems of double curves. Consider the subring $R_0 = \varphi(\mathbb{Z}[\pi_1(X)])$ of the field R; note that $R_0 = \mathbb{Z}[\varphi(\pi_1(X))]$ and $R_0^* = \pm\varphi(\pi_1(X))$. To avoid confusion, we denote by $L_i^{R_0}$ the module associated to the map φ viewed with values in R_0. The natural map from $H_1^{\varphi}(\Sigma, *; R_0)$ to $H_1^{\varphi}(\Sigma, *; R)$ sends $L_i^{R_0}$ onto L_i^{φ}. The submodules of $H_1^{\varphi}(\Sigma, *; R)$ that appear in the complex $(\mathcal{C})$ of Theorem 3.8 are free and are images of the similar submodules of $H_1^{\varphi}(\Sigma, *; R_0)$, which have the same rank. An R_0-*basis* of such a submodule of $H_1^{\varphi}(\Sigma, *; R)$ is a basis that is the image of a basis of the corresponding submodule of $H_1^{\varphi}(\Sigma, *; R_0)$. A tuple of double curves defines such an R_0-basis. Any other R_0-basis can be used to compute the torsion. Actually, changing the basis of a homology module in the computation of the torsion multiplies the torsion by the determinant of the change of basis. For R_0-bases, this determinant is the same as the determinant of the corresponding change of basis of the corresponding submodule of $H_1^{\varphi}(\Sigma, *; R_0)$, so that it is an element of $R_0^* = \pm\varphi(\pi_1(X))$.

4. The twisted relative homology groups and torsion

In this section, we compute the twisted relative homology and torsion of X. The computation of the homology of $(X, \partial X)$ ends up being formally similar to that of X: it involves a retraction onto a 3-complex in X. However, in order to make the relative cellular structure clearly apparent, we leave ∂X fixed throughout the retraction. To carry out such a retraction, X has to be punctured, but once the homology of the punctured version of X is computed rel ∂X, it is easy to recover the homology of X itself rel ∂X.

Definition 4.1. Let C be a lensed compression body. An *r-defining collection of disks* for C is a disjoint union $\mathcal{D}^r$ of disks, with boundary in $\partial_+ C$ or made of an arc in $\partial_+ C$ and an arc in $\partial_- C$, such that $C \setminus \eta(\mathcal{D}^r)$ is a 3-ball. The intersection with $\partial_+ C$ of an r-defining collection of disks for C is a *complete collection of arcs and curves* for C.

Likewise if V is a hyper compression body then an *r-defining collection of balls* for V is a union of 3-balls $\mathcal{B}^r$ such that $V \setminus \eta(\mathcal{B}^r)$ is a 4-ball.

Remark 4.2. The r in these definitions stands for "relative". Note that an r-defining collection of disks can be chosen to contain a defining collection of disks, and similarly for collections of balls.

Such r-defining collections of disks do exist. First take a subcollection of a defining collection of disks for C, dropping those that do not carry homology relative to boundary. Then add the products with the interval in $\partial_- C \times I$ of arcs that cut $\partial_- C$ into a disjoint union of disks. A similar argument shows existence of r-defining collections of balls for the hyper compression bodies under consideration here.

Fix r-defining collections $\mathcal{D}_i^r$ and $\mathcal{B}_i^r$ of disks and balls for C_i and X_i respectively. Set $\mathcal{D}^r = \bigcup_{i=1}^n \mathcal{D}_i^r$ and $\mathcal{B}^r = \bigcup_{i=1}^n \mathcal{B}_i^r$. For all $Z \subset X$, let $Z' = Z \setminus \eta(*)$.

Lemma 4.3. *The manifold X' deformation retracts onto $\Sigma' \cup \mathcal{D}^r \cup \mathcal{B}^r \cup \partial X$. Further, the quad $(X', Y' \cup \partial X, \Sigma' \cup \partial X, \partial X)$ deformation retracts rel ∂X onto a CW-complex $(Z_3^\partial \cup \partial X, Z_2^\partial \cup \partial X, Z_1^\partial \cup \partial X, \partial X)$, where Z_1^∂ is made of arcs and loops on Σ', $Z_2^\partial = Z_1^\partial \cup \mathcal{D}^r$, $Z_3^\partial = Z_2^\partial \cup \mathcal{B}^r$.*

Proof. The proof is similar to that of Lemma 3.2, but instead of retracting from the boundary, we retract "inside out" from the puncture $*$. Because $X_i \setminus \eta(\mathcal{B}_i^r)$ is a ball and meets $\eta(*)$ in a small 4-ball that has been "scooped out" of the boundary, we obtain a retraction of X_i' onto $(\partial X_i)' \cup \mathcal{B}_i^r$. Carrying this retraction out for each i yields a retraction of X' onto $Y' \cup \mathcal{B}^r \cup \partial X$ — recall that $\partial X_i = (X_i \cap X_{i-1}) \cup (X_i \cap X_{i+1}) \cup (X_i \cap \partial X)$. Since each $C_i \setminus \eta(\mathcal{D}_i^r)$ is also a ball which intersects $\eta(*)$ along a scooped out 3-ball, Y' can further be retracted onto $\Sigma' \cup \mathcal{D}^r$. This gives the first assertion, and the second one follows. $\square$

Corollary 4.4. *The homology of* $(X, \partial X)$ *is given by the chain complex*

$$(\mathcal{C}'_\partial)\quad H_4^\varphi(X, X') \to H_3^\varphi(X', Y' \cup \partial X) \to H_2^\varphi(Y', \Sigma' \cup \partial Y) \to H_1^\varphi(\Sigma', \partial\Sigma) \to 0.$$

Proof. Lemma 4.3 immediately gives the following cellular chain complex for $(X', \partial X)$:

$$0 \to H_3^\varphi(X', Y' \cup \partial X) \to H_2^\varphi(Y' \cup \partial X, \Sigma' \cup \partial X) \to H_1^\varphi(\Sigma' \cup \partial X, \partial X) \to 0,$$

or equivalently,

$$0 \to H_3^\varphi(X', Y' \cup \partial X) \to H_2^\varphi(Y', \Sigma' \cup \partial Y) \to H_1^\varphi(\Sigma', \partial\Sigma) \to 0.$$

Now, the long exact sequence of the triple $(X, X', \partial X)$ shows that $H_k^\varphi(X, \partial X) \cong H_k^\varphi(X', \partial X')$ for $k = 1, 2$ and $H_3^\varphi(X, \partial X) \cong H_3^\varphi(X', \partial X)/\mathrm{Im}(H_4^\varphi(X, X'))$. Finally, the long exact sequence of the triple $(X, Y' \cup \partial X, \partial X)$ identifies $H_4^\varphi(X, \partial X)$ with $H_4^\varphi(X, Y' \cup \partial X)$ and the long exact sequence of the triple $(X, X', Y' \cup \partial X)$ identifies $H_4^\varphi(X, Y' \cup \partial X)$ with the kernel of $H_4^\varphi(X, X') \to H_3^\varphi(X', Y' \cup \partial X)$. □

Definition 4.5. Let $\mathcal{J}_i^\varphi$ denote the subgroup of $H_1^\varphi(\Sigma', \partial\Sigma)$ generated by any complete collection of arcs and curves for C_i on Σ'.

Lemma 4.6 gives an alternative interpretation of $\mathcal{J}_i^\varphi$. Identifying $H_1^\varphi(\Sigma, *)$ with $H_1(\Sigma', \partial\eta(*))$ via the excision equivalence, and using the decomposition $\partial\Sigma' = \partial\Sigma \cup \partial\eta(*)$, we have an equivariant intersection form on $H_1^\varphi(\Sigma, *) \times H_1^\varphi(\Sigma', \partial\Sigma)$.

Lemma 4.6. *The modules* $H_1^\varphi(\Sigma', \partial\Sigma)$ *and* $\mathcal{J}_i^\varphi$ *are free* R*-modules of respective ranks* $2g+b-1$ *and* $2g+b-1-n$*. The modules* $\mathcal{J}_{i-1}^\varphi \cap \mathcal{J}_i^\varphi$ *are also free. Moreover,* $\mathcal{J}_i^\varphi$ *is the orthogonal complement of* L_i^φ *with respect to the equivariant intersection pairing on* $H_1^\varphi(\Sigma, *) \times H_1^\varphi(\Sigma', \partial\Sigma)$.

Proof. Let Z_1^∂ be any collection of $2g + b - 1$ arcs properly embedded in Σ' which are pairwise disjoint and cut Σ' into a disk. Then Σ' retracts onto $Z_1^\partial \cup \partial\Sigma$, showing that $H_1^\varphi(\Sigma', \partial\Sigma) \cong R^{2g+b-1}$. The 1-complex Z_1^∂ can be chosen so that $2g - p + b - 1$ of the arcs form a complete collection of arcs and curves for C_i (start with a complete collection of arcs and curves, replace closed curves by arcs, and add as many arcs as needed) whose twisted homology classes generate $\mathcal{J}_i^\varphi$. A basis of $\mathcal{J}_{i-1}^\varphi \cap \mathcal{J}_i^\varphi$ is provided by a subcollection of these.

Now, a curve c_i^0 in the family c_i bounds a disk in C_i, while an arc γ in a complete collection of arcs and curves for C_i cobounds a disk in C_i with an arc in $\partial_- C_i$. Assuming transversality of the two disks, it follows that the intersection of c_i^0 and γ is the boundary of a union of embedded intervals and hence contains as many positive as negative intersection points. Hence $\mathcal{J}_i^\varphi$ is contained in the orthogonal complement of L_i^φ, and the equality follows by a dimension argument, using the nondegeneracy of the intersection form. □

Lemma 4.7. $H_2^\varphi(C_i', (\partial C_i)') \cong \mathcal{J}_i^\varphi$ *for all* i.

Proof. The long exact sequence of the triple $(C_i', (\partial C_i)', \partial_- C_i)$, together with the excision equivalence $((\partial C_i)', \partial_- C_i) \sim (\Sigma', \partial\Sigma)$ give the short exact sequence

$$0 \to H_2(C_i', (\partial C_i)') \to H_1(\Sigma', \partial\Sigma) \xrightarrow{\zeta} H_1(C_i', \partial_- C_i).$$

Now C_i' is obtained from a thickened $\partial_- C_i$ by adding only 1-handles, so that the kernel of ζ contains the homology classes of curves in Σ' that have trivial algebraic intersection with the cocores of these 1-handles, cocores whose boundaries generate L_i^φ. We conclude that $H_2(C_i', (\partial C_i)') \cong \ker(\zeta) \cong \mathcal{J}_i^\varphi$. □

Lemma 4.8. $H_3^\varphi(X_i', (\partial X_i)') \cong \mathcal{J}_{i-1}^\varphi \cap \mathcal{J}_i^\varphi$ *for all* i.

Proof. Since X_i' is obtained from a thickened $\partial_- X_i$ by adding 1-handles, the exact sequence of the triple $(X_i', (\partial X_i)', \partial_- X_i)$ gives an isomorphism $H_3^\varphi(X_i', (\partial X_i)') \cong H_2^\varphi((\partial X_i)', \partial_- X_i)$; this last module is isomorphic to $H_2(C_{i-1}' \cup C_i', \partial_- C_{i-1} \cup \partial_- C_i)$. The long exact sequence of the triple $\big(C_{i-1}' \cup C_i', (\partial C_{i-1})' \cup (\partial C_i)', \partial_- C_{i-1} \cup \partial_- C_i\big)$ and the excision equivalence $\big((\partial C_{i-1})' \cup (\partial C_i)', \partial_- C_{i-1} \cup \partial_- C_i\big) \sim (\Sigma', \partial\Sigma)$ give

$$0 \to H_2^\varphi(C_{i-1}' \cup C_i', \partial_- C_{i-1} \cup \partial_- C_i) \xrightarrow{\iota} H_2^\varphi(C_{i-1}', (\partial C_{i-1})') \oplus H_2^\varphi(C_i', (\partial C_i)') \xrightarrow{\pi} H_1^\varphi(\Sigma', \partial\Sigma).$$

The conclusion follows from Lemma 4.7 with an argument analogous to that of Lemma 3.7. □

Theorem 4.9. *If* $\partial X \neq \varnothing$, *the twisted homology of* $(X, \partial X)$ *is given by the chain complex*

$$(\mathcal{C}_\partial) \qquad H_2^\varphi(\Sigma, \Sigma') \xrightarrow{\partial_3} \bigoplus_i (\mathcal{J}_{i-1}^\varphi \cap \mathcal{J}_i^\varphi) \xrightarrow{\partial_2} \bigoplus_i \mathcal{J}_i^\varphi \xrightarrow{\partial_1} H_1^\varphi(\Sigma', \partial\Sigma) \to 0,$$

where $\partial_3([\Sigma]) = [\partial(\Sigma \setminus \Sigma')]$, $\partial_2((x_i)_{1\le i\le n}) = ((x_i - x_{i+1})_{1\le i\le n})$ *and* $\partial_1((x_i)_{1\le i\le n}) = \sum_{i=1}^n x_i$. *Moreover, if R is a field, the complex basis b of $(\mathcal{C}_\partial)$ described in Remark 4.10 forms a geometric basis for the relative torsion of X, meaning that* $\tau^\varphi(X, \partial X; h) = \tau(\mathcal{C}_\partial; b, h)$.

Proof. Start with the complex $(\mathcal{C}_\partial')$ of Corollary 4.4. For the order 2 and 3 terms, use Mayer–Vietoris sequences to get the identifications $H_2^\varphi(Y', \Sigma' \cup \partial Y) \cong \bigoplus_i H_2^\varphi(C_i', (\partial C_i)')$ and $H_3^\varphi(X', Y' \cup \partial X) \cong \bigoplus_i H_3^\varphi(X_i', (\partial X_i)')$ and conclude with Lemmas 4.7 and 4.8. A generator of $H_4^\varphi(X, X')$ is sent onto the class of $\partial\eta(*)$ in $H_3^\varphi(X', Y' \cup \partial X)$. Following the isomorphisms in Lemma 4.8, we see that this class corresponds to the class in $\bigoplus_i \mathcal{J}_{i-1}^\varphi \cap \mathcal{J}_i^\varphi$ of the curve $\partial\eta(*)$, where the neighborhood is now understood in Σ, which is the boundary of a generator of $H_2(\Sigma, \Sigma')$. □

Remark 4.10. As with the absolute case, we can obtain a concrete description of what the geometric bases b look like for $\mathcal{C}_\partial$:

- $b_1 = \{[e_1], [e_2], \dots, [e_n]\}$, where each e_i is an edge of Z_1^∂ (i.e., any set of arcs which cut Σ into a disk),
- $b_2 =$ any basis corresponding to a tuple of complete collections of arcs and curves for C_i,
- $b_3 =$ any basis corresponding to a tuple of "double arcs and curves" for the pairs (C_{i-1}, C_i), or any other R_0-basis with $R_0 = \varphi(\mathbb{Z}[\pi_1(X)])$ (see Remark 3.12),
- $b_4 =$ the fundamental class of $H_2(\Sigma, \Sigma')$.

Corollary 4.11. *We have the following expressions for the twisted homology of* $(X, \partial X)$:

$$H_1^\varphi(X, \partial X) \cong H_1^\varphi(\Sigma', \partial\Sigma) / \bigoplus_i \mathcal{J}_i^\varphi, \quad H_3^\varphi(X, \partial X) \cong \bigcap_i \overline{\mathcal{J}}_i^\varphi,$$

where $\overline{\mathcal{J}}_i^\varphi$ *denotes the image of* $\mathcal{J}_i^\varphi$ *under the inclusion map* $H_1^\varphi(\Sigma', \partial\Sigma) \to H_1^\varphi(\Sigma, \partial\Sigma)$.

Proof. For H_3, the long exact sequence of the triple $(\Sigma, \Sigma', \partial\Sigma)$ gives the exact sequence

$$H_2^\varphi(\Sigma, \Sigma') \xrightarrow{\zeta} H_1^\varphi(\Sigma', \partial\Sigma) \to H_1^\varphi(\Sigma, \partial\Sigma) \to 0,$$

and we have $H_3^\varphi(X, \partial X) \cong \left(\bigcap_i \mathcal{J}_i^\varphi\right)/\mathrm{Im}(\zeta)$. □

5. Intersection forms

We keep in this section the assumption that $\partial X \neq \varnothing$. The intersection forms are formally identical to the closed case treated in [Florens and Moussard 2022]. The upshot is that the intersections between various cycles in X can all be made to coincide with intersections in Σ. Below we assume that $\Sigma' = \Sigma \setminus \eta(*)$, so that there is a natural isomorphism $H_1^\varphi(\Sigma, *) \cong H_1^\varphi(\Sigma', \partial\eta(*))$, and we identify each L_i^φ with its image under this map below. Note that, in the nontwisted case, $H_1(\Sigma, *)$ naturally identifies with $H_1(\Sigma)$.

Theorem 5.1. *Suppose* $h_1 = [(x_i)_{1\le i\le n}]$ *and* $h_2 = [(y_i)_{1\le i\le n}]$ *in* $H_2^\varphi(X)$, *where* $(x_i)_{1\le i\le n}, (y_i)_{1\le i\le n} \in \bigoplus_i L_i^\varphi$. *Then*

$$\langle h_1, h_2\rangle_X^\varphi = \sum_{1\le i<j\le n} \langle x_i, y_j\rangle_\Sigma^\varphi,$$

where $\langle\,\cdot\,,\cdot\,\rangle_X^\varphi$ *and* $\langle\,\cdot\,,\cdot\,\rangle_\Sigma^\varphi$ *are the equivariant intersection forms on* $H_2^\varphi(X)$ *and* $H_1^\varphi(\Sigma, *)$ *respectively.*

Proof. It suffices to show that the analogous claim holds true in the untwisted integral homology groups of $\overline{X}$, which denotes the cover of X associated to $\ker(\varphi)$. For any $Z \subset X$ let $\overline{Z}$ denote the inverse image of Z under the cover $\overline{X} \to X$. Because $\pi_1(\Sigma)$, $\pi_1(C_i)$, and $\pi_1(X_i)$ all surject onto $\pi_1(X)$ via the inclusion map,

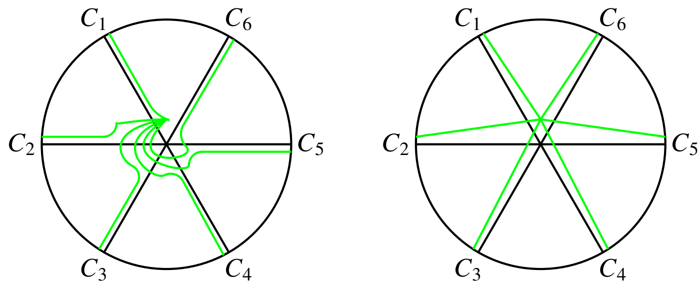


Figure 3. Pushing the relative 2-skeleton. The 2-skeleton appears in black, while the pushed relative 2-skeleton appears in green. At each intersection of the black lines and the green lines lies a copy of Σ.

$\overline{\Sigma}$, $\overline{C}_i$, and $\overline{X}_i$ are connected as well. In the finite case these lifts combine to form a multisection of $\overline{X}$, and in the case of an infinite-sheeted cover they form what is essentially a multisection, except the pieces involved have infinite genus. In particular, just as in the finite case, $\eta(\overline{\Sigma})$ is a trivial disk bundle and the lifted compression bodies $\overline{C}_i$ meet each disk in the bundle along rays which are disjoint except at the center point.

There is a cellular structure on $\overline{X}$ obtained by lifting the cell structures of X and $(X, \partial X)$ described in Lemmas 3.2 and 4.3 to $\overline{X}$. If Z_2 is the 2-skeleton of X described in Lemma 3.2, then $\overline{Z}_2$ is a 2-skeleton for $\overline{X}$ which lies in $\overline{\bigcup_i C_i}$. As observed in [Florens and Moussard 2022], we may push each $\overline{Z}_2 \cap \overline{C}_i$ slightly into its collar so that it is pushed into $\overline{\bigcup_{1\le j\le i} X_j}$. This being done, the intersections between 2-chains in $\overline{Z}_2$ and 2-chains in the pushed $\overline{Z}_2$ will coincide with intersections between the boundaries of the subchains lying just in $\overline{Z_2 \cap C_i}$, and these intersections occur on diverse copies of Σ; see the left-hand side of Figure 3. □

Remark 5.2. Different expressions can be given for the intersection form by diversely pushing the relative 2-skeleton. The right-hand side of Figure 3 suggests another possibility with fewer terms.

Similarly one can compute the intersection pairings on $H_k^\varphi(X) \times H_{4-k}^\varphi(X, \partial X)$. For $k = 2$, the expression is analogous to that of Theorem 5.1.

Theorem 5.3. *If $h_1 = [(x_i)_{1\le i\le n}] \in H_2^\varphi(X)$ and $h_2 = [(y_i)_{1\le i\le n}] \in H_2^\varphi(X, \partial X)$, where $(x_i)_{1\le i\le n} \in \bigoplus_i L_i^\varphi$, and $(y_i)_{1\le i\le n} \in \bigoplus_i \mathcal{J}_i^\varphi$, then*

$$\langle h_1, h_2\rangle_X^\varphi = \sum_{1\le i<j\le n} \langle x_i, y_j\rangle_\Sigma^\varphi,$$

where $\langle\cdot,\cdot\rangle_X^\varphi$ and $\langle\cdot,\cdot\rangle_\Sigma^\varphi$ are the equivariant intersection pairings on $H_2^\varphi(X) \times H_2^\varphi(X, \partial X)$ and $H_1^\varphi(\Sigma, *) \times H_1^\varphi(\Sigma', \partial\Sigma)$ respectively.

Proof. The proof of Theorem 5.1 applies with the following adaptation: we consider the relative 2-skeleton Z_2^∂ of Lemma 4.3 and we look at intersections between 2-chains in $\bar{Z}_2$ and 2-chains in the pushed $\overline{Z_2^\partial}$. □

The intersection pairings on the odd-dimensional homology groups are described even more simply.

Theorem 5.4. *Suppose that either $h_1 \in H_1^\varphi(X)$ corresponds to the element $a \in H_1^\varphi(\Sigma, *)$ and $h_2 \in H_3^\varphi(X, \partial X)$ corresponds to the element $b \in \bigcap_i \mathcal{J}_i^\varphi$, or $h_1 \in H_1^\varphi(X, \partial X)$ corresponds to the element $a \in H_1^\varphi(\Sigma', \partial\Sigma)$ and $h_2 \in H_3^\varphi(X)$ corresponds to the element $b \in \bigcap_i L_i^\varphi$. Then*

$$\langle h_1, h_2 \rangle_X^\varphi = \langle a, b \rangle_\Sigma^\varphi.$$

Proof. The proof is similar in structure to the proof of Theorem 5.1, except that now we observe that every chain in $H_1(\overline{X})$ or $H_1(\overline{X}, \partial\overline{X})$ is geometrically represented by linear combinations of curves $c \subset \overline{\Sigma} \subset \overline{X}$, and the chains in $H_1(\overline{X}, \partial\overline{X})$ or $H_3(\overline{X})$ can be geometrically represented by linear combinations of balls which meet $\overline{\Sigma}$ only in linear combinations of double curves. Thus no isotopy is needed, the intersections between the 1-chains and the 3-chains already correspond exactly to the intersections of their representatives in $H_1(\overline{\Sigma})$. □

6. The case of closed 4-manifolds

In this section, we compute the twisted homology, torsion and intersection forms when X is closed. It mainly follows the lines of the computation of relative homology, since we need again to puncture X. However, it mixes some features of the absolute and relative cases. For instance, when X is closed, r-defining collections of disks and balls are the same as ordinary defining collections. Since there is no additive difficulty with regards to the nonclosed case, we skip the details.

We fix $\star \in \Sigma$; for $Z \subset X$, we set $Z' = Z \setminus \eta(\star)$ and we fix $* \in \partial\Sigma'$. Let $\mathcal{D}$ and $\mathcal{B}$ be unions of defining collections of disks and balls for the C_i and the X_i respectively. Lemma 4.3 still holds, and provides the following corollary.

Lemma 6.1. *The quad $(X', Y', \Sigma', *)$ deformation retracts onto a CW-complex (Z_3, Z_2, Z_1, Z_0), where $Z_0 = *$, Z_1 is made of loops on Σ', $Z_2 = Z_1^\partial \cup \mathcal{D}$, $Z_3 = Z_2 \cup \mathcal{B}$. Subsequently, the homology of X is given by the chain complex*

$$(\bar{\mathcal{C}}') \qquad H_4^\varphi(X, X') \to H_3^\varphi(X', Y') \to H_2^\varphi(Y', \Sigma') \to H_1^\varphi(\Sigma', *) \to H_0^\varphi(*).$$

Now, L_i^φ denotes the submodule of $H_1^\varphi(\Sigma', *)$ generated by the homology classes of the curves in c_i.

Lemma 6.2. *The modules $H_1^\varphi(\Sigma', *)$ and L_i^φ are free R-modules of ranks $2g$ and g respectively. The modules $L_{i-1}^\varphi \cap L_i^\varphi$ are also free. Moreover, L_i^φ is a lagrangian for the equivariant intersection form on $H_1^\varphi(\Sigma', *)$.*

Lemma 6.3. *$H_2^\varphi(C_i', (\partial C_i)') \cong L_i^\varphi$ and $H_3^\varphi(X_i', (\partial X_i)') \cong L_{i-1}^\varphi \cap L_i^\varphi$ for all i.*

Theorem 6.4. *If X is closed, the twisted homology of X is given by the chain complex*

$$(\bar{C}) \qquad H_2^\varphi(\Sigma, \Sigma') \xrightarrow{\partial_3} \bigoplus_i (L_{i-1}^\varphi \cap L_i^\varphi) \xrightarrow{\partial_2} \bigoplus_i L_i^\varphi \xrightarrow{\partial_1} H_1^\varphi(\Sigma', *) \to H_0^\varphi(*),$$

where

$$\partial_3([\Sigma]) = [\partial \Sigma'], \quad \partial_2((x_i)_{1\le i\le n}) = ((x_i - x_{i+1})_{1\le i\le n}), \quad \partial_1((x_i)_{1\le i\le n}) = \sum_{i=1}^n x_i.$$

Moreover, if R is a field, the complex basis b of $(\bar{C})$ described in Remark 6.5 forms a geometric basis for the torsion of X, meaning that $\tau^\varphi(X; h) = \tau(\bar{C}; b, h)$.

Remark 6.5. Once again, we can describe geometric torsion bases b for $\bar{C}$:

- $b_0 = [*]$,
- $b_1 =$ any set of loops based at $*$ which cut Σ into a disk,
- $b_2 =$ any basis corresponding to a tuple of defining collections of curves for C_i,
- $b_3 =$ any basis corresponding to a tuple of "double curves" for the pairs (C_{i-1}, C_i), or any other R_0-basis with $R_0 = \varphi(\mathbb{Z}[\pi_1(X)])$ (see Remark 3.12),
- $b_4 =$ the fundamental class of $H_2(\Sigma, \Sigma')$.

Expressions for the intersection forms on $H_2^\varphi(X)$ and on $H_1^\varphi(X) \times H_3^\varphi(X)$ are again obtained in terms of the intersection form on $H_1^\varphi(\Sigma', *)$. Strictly speaking, this intersection form is defined on $H_1^\varphi(\Sigma', *_1) \times H_1^\varphi(\Sigma', *_2)$ for two distinct basepoints $*_1$ and $*_2$ on $\partial\Sigma'$. Again, in the nontwisted case, $H_1(\Sigma', *)$ identifies with $H_1(\Sigma)$.

Theorem 6.6. *Suppose that $h_1 = [(x_i)_{1\le i\le n}]$ and $h_2 = [(y_i)_{1\le i\le n}] \in H_2^\varphi(X)$, where $(x_i)_{1\le i\le n}, (y_i)_{1\le i\le n} \in \bigoplus_i L_i^\varphi$. Then*

$$\langle h_1, h_2\rangle_X^\varphi = \sum_{1\le i<j\le n} \langle x_i, y_j\rangle_\Sigma^\varphi,$$

where $\langle\cdot,\cdot\rangle_X^\varphi$ and $\langle\cdot,\cdot\rangle_\Sigma^\varphi$ are the equivariant intersection forms on $H_2^\varphi(X)$ and $H_1^\varphi(\Sigma', *)$.

Suppose that $h_1 \in H_1^\varphi(X)$ corresponds to the element $a \in H_1^\varphi(\Sigma', *)$ and that $h_2 \in H_3^\varphi(X)$ corresponds to the element $b \in \bigcap_i L_i^\varphi$. Then $\langle h_1, h_2\rangle_X^\varphi = \langle a, b\rangle_\Sigma^\varphi$.

7. The boundary: monodromy and homology

In this section, we assume $\partial X \neq \varnothing$ and we compute the action of the monodromy of the open book induced on ∂X on the homology of the page Σ_∂; we then deduce the homology of ∂X. All homology groups are considered with coefficients in $\mathbb{Z}$. We denote by Σ_i the result of compressing Σ along c_i, which is a copy of Σ_∂.

Given a compact surface S with no closed component, a *cut system for S* is a family of arcs on S that cuts each component of S into a disk.

Our main tool is the algorithm of Castro, Gay and Pinzón-Caicedo which describes the monodromy of the open book from a trisection diagram [Castro et al. 2018a]. Although they work with trisections in the case of a connected page, their result extends directly to the setting of multisections with multiple boundary components.

Proposition 7.1 (Castro, Gay and Pinzón-Caicedo). *Let e be any choice of arcs in Σ, disjoint from c_1, that forms a cut system for Σ_1. The monodromy $\phi : \Sigma_1 \to \Sigma_1$ which defines the open book decomposition of ∂X is encoded by its action on e, which in turn is described by the following algorithm. For i running from 1 to n, slide the curves c_{i+1} over one another and slide the arcs e over the curves c_i, until e is disjoint from the curves c_{i+1}. The family $c_1' \cup e'$ which results from these n steps will generally be distinct from the original family $c_1 \cup e$. Perform one final sequence of handleslides of the arcs and curves $c_1' \cup e'$ which sends c_1' to c_1. The resulting cut system e' for Σ_1 is $\phi(e)$.*

It is necessary to explicitly index the arcs e and keep track of this index throughout the algorithm, but the simple closed curves c_i need not be indexed.

We denote by L_i the subgroup of $H_1(\Sigma)$ generated by the homology classes of the curves in c_i, and we let $\mathcal{J}_i$ denote its orthogonal in $H_1(\Sigma, \partial\Sigma)$ with respect to the intersection pairing on $H_1(\Sigma) \times H_1(\Sigma, \partial\Sigma)$ (see Section 2C). Similarly, we denote by $\mathcal{L}_i$ the subgroup of $H_1(\Sigma, \partial\Sigma)$ generated by the homology classes of the curves in c_i, and we let J_i denote its orthogonal in $H_1(\Sigma)$.

Lemma 7.2. *The groups L_i, J_i, $\mathcal{L}_i$ and $\mathcal{J}_i$ are free abelian groups of ranks p, $g+h+b-1$, $g-h$ and $2g+b-p-1$ respectively, where g is the genus of Σ, h the genus of Σ_∂ and b the number of boundary components of both. Moreover, L_i and $\mathcal{L}_i$ are primitive subgroups of J_i and $\mathcal{J}_i$ respectively, so that the quotients J_i/L_i and $\mathcal{J}_i/\mathcal{L}_i$ both are free abelian groups of rank $g+h+b-p-1$.*

Proof. Up to diffeomorphism and handleslides, the curves of the collection c_i can be put in standard position; see Figure 2. From this standard position, one can draw curves providing bases for the different groups under study; see Figure 4.

The group L_i is generated by the homology classes of the curves in c_i, which are p homologically independent curves on Σ, and thus $\operatorname{rk} L_i = p$. The group $\mathcal{L}_i$ is generated by the same curves, but some of them are trivial in $H_1(\Sigma, \partial\Sigma)$. There are $g-h$ nontrivial ones, which are p homologically independent, so that $\operatorname{rk} \mathcal{L}_i = g-h$. Bases for J_i and $\mathcal{J}_i$ can be obtained by completing the given bases for L_i and $\mathcal{L}_i$ respectively, giving the remainder of the statement. Note that the ranks of J_i and $\mathcal{J}_i$ can be recovered from the fact that these groups are orthogonal complements of $\mathcal{L}_i$ and L_i respectively. □

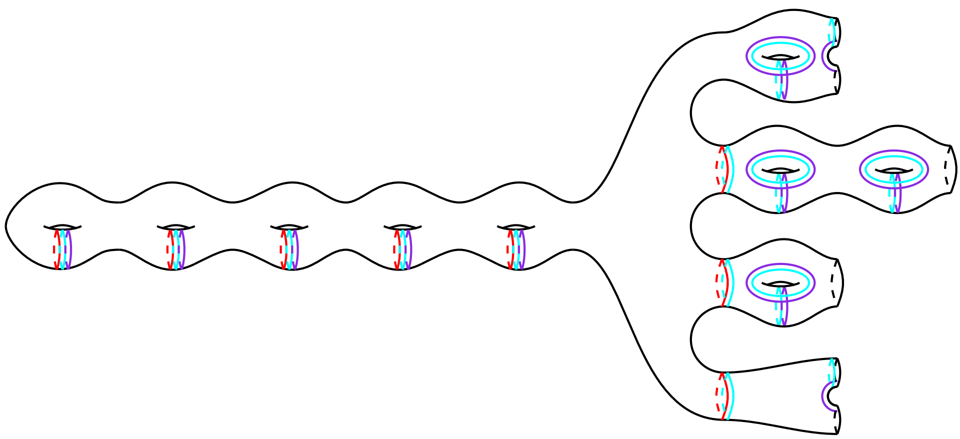


Figure 4. Curves on Σ for the compression body C_i. The curves of c_i are in red; their homology classes in $H_1(\Sigma)$ form a basis of L_i and the five leftmost ones provide a basis of $\mathcal{L}_i \subset H_1(\Sigma, \partial\Sigma)$. The homology classes of the blue and violet curves form bases of J_i and $\mathcal{J}_i$ respectively.

Lemma 7.3. *There are natural identifications $H_1(\Sigma_i, \partial\Sigma) \cong \mathcal{J}_i/\mathcal{L}_i$ and $H_1(\Sigma_i) \cong J_i/L_i$.*

Proof. Recall C_i is a lensed compression body, so that we can write its boundary as $\partial C_i = \Sigma \cup_{\partial\Sigma} \Sigma_i$. In particular, we have excision equivalences $(\partial C_i, \Sigma_i) \sim (\Sigma, \partial\Sigma)$ and $(\partial C_i, \Sigma) \sim (\Sigma_i, \partial\Sigma)$.

We first view C_i as a thickened Σ_i with 1-handles attached on the positive boundary whose cocores are the curves in c_i. This shows that $H_2(C_i, \Sigma_i) = 0$ and $H_1(C_i, \Sigma_i)$ is generated by the classes of the cores of the 1-handles. Hence the long exact sequence of the triple $(C_i, \partial C_i, \Sigma_i)$ gives

$$0 \to H_2(C_i, \partial C_i) \to H_1(\Sigma, \partial\Sigma) \to H_1(C_i, \Sigma_i) \to 0.$$

Now the image of an element of $H_1(\Sigma, \partial\Sigma)$ is determined by its algebraic intersection with the curves in c_i, and thus $H_2(C_i, \partial C_i) \cong \mathcal{J}_i$.

Likewise, viewing the compression body C_i as a thickened Σ with 2-handles glued along the curves in c_i on the negative boundary shows that $H_1(C_i, \Sigma) = 0$ and $H_2(C_i, \Sigma)$ is generated by the classes of the cores of the 2-handles. Now the long exact sequence of the triple $(C_i, \partial C_i, \Sigma)$ gives

$$H_2(C_i, \Sigma) \to H_2(C_i, \partial C_i) \to H_1(\Sigma_i, \partial\Sigma) \to 0.$$

Since the 2-handles are glued along the curves in c_i, the image of $H_2(C_i, \Sigma)$ corresponds to $\mathcal{L}_i$ in the above identification of $H_2(C_i, \partial C_i)$ with the subgroup $\mathcal{J}_i$ of $H_1(\Sigma, \partial\Sigma)$. This gives the identification $H_1(\Sigma_i, \partial\Sigma) \cong \mathcal{J}_i/\mathcal{L}_i$.

We now repeat the whole argument replacing ∂C_i by $\Sigma \sqcup \overline{\Sigma}_i$, where $\overline{\Sigma}_i$ is obtained from Σ_i by removing an open collar neighborhood of its boundary. The first step gives

$$0 \to H_2(C_i, \Sigma \sqcup \overline{\Sigma}_i) \to H_1(\Sigma) \to H_1(C_i, \Sigma_i)$$

and $H_2(C_i, \Sigma \sqcup \overline{\Sigma}_i) \cong J_i$, and the second step gives $H_2(C_i, \Sigma) \to H_2(C_i, \Sigma \sqcup \overline{\Sigma}_i) \to H_1(\Sigma_i) \to 0$ and $H_2(C_i, \Sigma) \cong L_i$ in $H_2(C_i, \Sigma \sqcup \overline{\Sigma}_i) \cong J_i$. □

Let e be a family of arcs in Σ, disjoint from c_1, that forms a cut system for Σ_1; note that it defines a basis of $H_1(\Sigma_1, \partial\Sigma)$. Let a_i be a family of simple closed curves on Σ that defines a basis of $\mathcal{L}_i/(\mathcal{L}_i \cap \mathcal{L}_{i+1})$ or $L_i/(L_i \cap L_{i+1})$ (in the sequel, we may consider their homology classes in $H_1(\Sigma)$ or $H_1(\Sigma, \partial\Sigma)$). For $\mu = (\mu_i)_{1\le i\le s}$ and $\mu' = (\mu'_i)_{1\le i\le t}$ families of $H_1(\Sigma, \partial\Sigma)$ and $H_1(\Sigma)$, define the matrix $\mu \cdot \mu' = (\langle \mu_i, \mu'_j\rangle_\Sigma)_{1\le i\le s, 1\le j\le t}$.

Proposition 7.4. *Let $\phi : \Sigma_1 \to \Sigma_1$ be the monodromy which defines the open book on ∂X. Define matrices R_i and families e_i in $H_1(\Sigma, \partial\Sigma)$ recursively as follows*:

- $R_0 = 0$ *and* $e_1 = e$,
- $R_i = -(e_i \cdot a_{i+1})(a_i \cdot a_{i+1})^{-1}$ *and* $e_{i+1} = e_i + R_i a_i$.

Fix a basis of the free $\mathbb{Z}$-module $\mathcal{J}_1$ which admits e as a subfamily and write the families e and e_{n+1} in this basis. Then the action of the monodromy of the open book of ∂X on $H_1(\Sigma_1, \partial\Sigma) \cong \mathcal{J}_1/(\mathcal{L}_1)$ is given in the basis e by the matrix of $R = e^t e_{n+1}$, where e^t is the transpose of e.

Proof. Following the algorithm of Proposition 7.1, we define families of arcs and curves e_i on Σ, disjoint from c_i, that define bases of $H_1(\Sigma_i, \partial\Sigma)$, by $e_1 = e$ and $e_{i+1} = e_i + r_i a_i$, where the r_i are matrices to compute. Since e_i is disjoint from c_i, we have $0 = e_{i+1} \cdot a_{i+1} = e_i \cdot a_{i+1} + r_i(a_i \cdot a_{i+1})$, so that $r_i = R_i$. Now e_{n+1} expresses $\phi(e)$ in the fixed basis of $\mathcal{J}_1$. Multiply by e^t to get it in the basis e of $H_1(\Sigma_1, \partial\Sigma)$. □

The following lemma gives the homology of a 3-manifold from an open book decomposition. A similar computation can be found in [Etnyre and Ozbagci 2008, Section 2.1].

Lemma 7.5. *Let M be a 3-manifold with an open book (S, ϕ). The homology of M is the homology of the complex*

$$0 \to \mathbb{Z}^s \xrightarrow{0} H_1(S, \partial S) \xrightarrow{\xi} H_1(S) \xrightarrow{0} \mathbb{Z}^s \to 0,$$

where $\xi([\mu]) = [-\mu \cup \phi(\mu)]$ and s is the number of components of S.

Proof. First note that S and M necessarily have the same number of connected components, so that s is also the number of components of M.

Consider the triple $\big(S\times[0,1], \partial(S\times[0,1]), S\times\{0\}\big)$. Since $S\times[0,1]$ deformation retracts on $S\times\{0\}$, the homology of the corresponding pair is trivial. Also, the open book structure gives a map $\phi : S\times[0,1]\to M$, injective on the interior, such that the $S_t=\phi(S\times\{t\})$ are the pages, with $S_0=S_1=S$. The map ϕ induces an isomorphism in homology: $H_*\big(S\times[0,1], \partial(S\times[0,1])\big)\cong H_*(M,S)$. Further, the inclusion of S as $S\times\{1\}$ in $\partial(S\times[0,1])$ induces an isomorphism $H_*\big(\partial(S\times[0,1]), S\times\{0\}\big)\cong H_*(S,\partial S)$. Finally $H_*(M,S)\cong H_{*-1}(S,\partial S)$. Hence the long exact sequence of the pair (M,S) gives

$$0\to H_2(M)\to H_1(S,\partial S)\xrightarrow{\xi} H_1(S)\to H_1(M)\to 0.$$

Finally, given an arc a properly embedded in $(S,\partial S)$, $a\times[0,1]$ is a relative 2-cycle for the pair

$$\big(S\times[0,1], \partial(S\times[0,1])\big)\sim(M,S),$$

whose boundary is $-a\cup\phi(a)$. □

To compute the homology of ∂X, we need to understand the homology classes $\phi(\mu)-\mu$ in $H_1(\Sigma_1)$. We keep the notations defined before and in Proposition 7.4.

Proposition 7.6. *Define families ε_i in $H_1(\Sigma)$ as follows: $\varepsilon_1=0$ and $\varepsilon_{i+1}=\varepsilon_i+R_ia_i$. Fix a basis b_L of $\mathcal{L}_1\cong L_1$ and complete it into a basis (b_L,b) of J_1. Write e in the basis (b_L,e) of $\mathcal{J}_1$ and ε_{n+1} in the basis (b_L,b) of J_1. The homology of ∂X is the homology of the complex*

$$0\to\mathbb{Z}^s\xrightarrow{0}\frac{\mathcal{J}_1}{\mathcal{L}_1}\xrightarrow{\xi}\frac{J_1}{L_1}\xrightarrow{0}\mathbb{Z}^s\to 0,$$

where s is the number of components of Σ and ξ is given in the bases e and b by the matrix $S=e^t\varepsilon_{n+1}$.

Proof. The ε_i represent the homology classes in $H_1(\Sigma)$ of the e_i-e. Throughout the algorithm of Proposition 7.1, they are added curves as in Proposition 7.4, but we now view the result in $H_1(\Sigma)$ at each step. □

8. Sample calculations

Example 1. The trisection diagram $(\Sigma;\alpha,\beta,\gamma)$ in Figure 5 is a diagram for a disk bundle X over S^2 with Euler number -2 obtained by Castro, Gay and Pinzón-Caicedo in [Castro et al. 2018a, Section 5.1]. In this example, all homology groups have coefficients in $\mathbb{Z}$. We first compute the (relative) homology and intersection form of X from this diagram.

In $H_1(\Sigma)=\langle\alpha_1,\beta_1,\alpha_2,\beta_2,\gamma_1\rangle$, we have $L_\alpha=\langle\alpha_1,\alpha_2\rangle$, $L_\beta=\langle\beta_1,\beta_2\rangle$, $L_\gamma=\langle\gamma_1,\alpha_2-2\beta_1+\beta_2\rangle$. All pairwise intersections of these subgroups are trivial. The

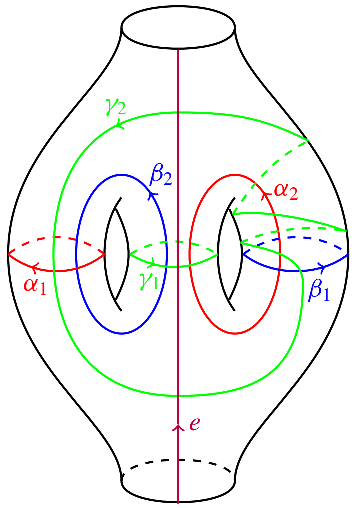


Figure 5. A trisection diagram of a disk bundle over S^2 with Euler number -2.

homology of X is the homology of the complex

$$0 \to L_\alpha \oplus L_\beta \oplus L_\gamma \to H_1(\Sigma) \xrightarrow{0} \mathbb{Z},$$

giving $H_1(X) = 0$, $H_2(X) \cong \mathbb{Z}$ and $H_3(X) = 0$. Note that the rightmost differential in always zero when working with coefficients in $\mathbb{Z}$.

In $H_1(\Sigma, \partial\Sigma) = \langle \alpha_1, \beta_1, \alpha_2, \beta_2, e\rangle$, we have $\mathcal{J}_\alpha = \langle \alpha_1, \alpha_2, e\rangle$, $\mathcal{J}_\beta = \langle \beta_1, \beta_2, e\rangle$, $\mathcal{J}_\gamma = \langle \alpha_1 - \beta_1, \alpha_2 - 2\beta_1 + \beta_2, e - \beta_2\rangle$; for $\mathcal{J}_\gamma$, we obtain these expressions by considering a complete collection of arcs and curves for C_γ made of γ_1, γ_2 and an arc joining the two boundary components avoiding the γ_i. Pairwise intersections are $\mathcal{J}_\alpha \cap \mathcal{J}_\beta = \langle e\rangle$, $\mathcal{J}_\beta \cap \mathcal{J}_\gamma = \langle e - \beta_2\rangle$, $\mathcal{J}_\gamma \cap \mathcal{J}_\alpha = \langle 2\alpha_1 - \alpha_2 - e\rangle$. The relative homology of X is the homology of the complex

$$\mathbb{Z} \xrightarrow{0} \bigoplus_{\nu \neq \nu'} \mathcal{J}_\nu \cap \mathcal{J}_{\nu'} \to \bigoplus_{\nu} \mathcal{J}_\nu \to H_1(\Sigma, \partial\Sigma) \to 0,$$

where $\nu, \nu' \in \{\alpha, \beta, \gamma\}$, giving $H_1(X, \partial X) = 0$, $H_2(X, \partial X) \cong \mathbb{Z}$ and $H_3(X, \partial X) = 0$. Note that the leftmost differential in always zero when working with coefficients in $\mathbb{Z}$.

A generator of $H_2(X)$ is given by $(\alpha_2, \beta_2 - 2\beta_1, -\gamma_2) \in L_\alpha \oplus L_\beta \oplus L_\gamma$. Using this generator, we can compute the intersection form of X:

$$\langle \alpha_2, \beta_2 - 2\beta_1\rangle_\Sigma + \langle \alpha_2, -\gamma_2\rangle_\Sigma + \langle \beta_2 - 2\beta_1, -\gamma_2\rangle_\Sigma = 2.$$

We now consider the monodromy of the open book on ∂X. We set $a_1 = (\alpha_1, \alpha_2)$, $a_2 = (\beta_1, \beta_2)$ and $a_3 = (\gamma_1, \gamma_2)$. Starting with $R_0 = 0$ and $e_1 = e$, we compute

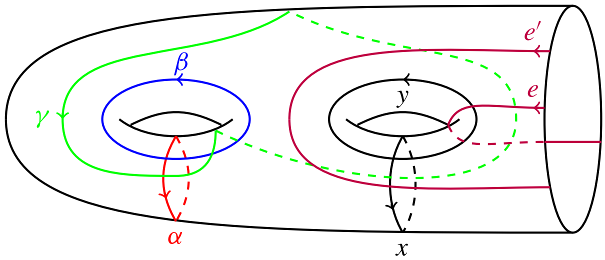


Figure 6. A trisection diagram.

$R_1 = (0\ \ 0)$, so that $e_2 = e$, then $R_2 = (0\ -1)$ and $e_3 = e - \beta_2$, and finally $R_3 = (-2\ \ 1)$ and $e_4 = e - 2\alpha_1 + \alpha_2$. Utilizing the basis (α_1, α_2, e) of $\mathcal{J}_\alpha$, this gives $R = (1)$ and shows that the action of the monodromy on $H_1(\Sigma_1, \partial\Sigma_1)$ is trivial. Now starting with $\varepsilon_1 = 0$, we get $\varepsilon_2 = 0$, $\varepsilon_3 = -\beta_2$ and $\varepsilon_4 = \alpha_2 - 2\alpha_1 - 2\zeta$, where $\zeta = \gamma_1 + \beta_1 - \alpha_1$ is a generator of J_α/L_α that we use as basis. Hence $S = (-2)$. Finally, the homology of ∂X is the homology of the complex

$$0 \to \mathbb{Z} \xrightarrow{0} \langle e\rangle \xrightarrow{-2} \langle\zeta\rangle \xrightarrow{0} \mathbb{Z} \to 0,$$

giving $H_1(\partial X) = \mathbb{Z}/2\mathbb{Z}$ and $H_2(\partial X) = 0$.

Example 2. Let X be the 4-manifold defined by the trisection diagram $(\Sigma; \alpha, \beta, \gamma)$ in Figure 6.

In $H_1(\Sigma) = \langle\alpha, \beta, x, y\rangle$, we have $L_\alpha = \langle\alpha\rangle$, $L_\beta = \langle\beta\rangle$, $L_\gamma = \langle-\alpha + \beta + y\rangle$. Pairwise intersections are trivial and we get $H_1(X; \mathbb{Z}) = \langle x\rangle \cong \mathbb{Z}$, $H_2(X; \mathbb{Z}) = 0$ and $H_3(X; \mathbb{Z}) = 0$.

In $H_1(\Sigma, \partial\Sigma) = \langle\alpha, \beta, e, e'\rangle$, we have $\mathcal{J}_\alpha = \langle\alpha, e, e'\rangle$, $\mathcal{J}_\beta = \langle\beta, e, e'\rangle$, $\mathcal{J}_\gamma = \langle\beta - \alpha, \alpha + e, e'\rangle$. This gives $H_1(X, \partial X) = 0$, $H_2(X, \partial X) = 0$ and $H_3(X, \partial X) = \mathcal{J}_\alpha \cap \mathcal{J}_\beta \cap \mathcal{J}_\gamma \cong \mathbb{Z}$.

We define $\varphi : \mathbb{Z}[\pi_1(X, *)] \to \mathbb{Z}[t^{\pm 1}]$ by $\varphi(x) = t$. Let us compute the associated twisted homology and torsion. Fix a lift $\tilde{*}$ of the basepoint $*$. For $\zeta \in \pi_1(\Sigma, *)$, we denote by $\tilde{\zeta}$ the lift of ζ starting at $\tilde{*}$. Since $\gamma = \alpha^{-1}xyx^{-1}\alpha\beta\alpha^{-1}$ in $\pi_1(\Sigma, *)$, we have $\tilde{\gamma} = -\tilde{\alpha} + \tilde{\beta} + t\tilde{y}$ in $H_1^\varphi(\Sigma, *)$. Hence, in $H_1^\varphi(\Sigma, *) = \langle\tilde{\alpha}, \tilde{\beta}, \tilde{x}, \tilde{y}\rangle$, we have $L_\alpha^\varphi = \langle\tilde{\alpha}\rangle$, $L_\beta^\varphi = \langle\tilde{\beta}\rangle$, $L_\gamma^\varphi = \langle-\tilde{\alpha} + \tilde{\beta} + t\tilde{y}\rangle$. From the complex

$$0 \to L_\alpha^\varphi \oplus L_\beta^\varphi \oplus L_\gamma^\varphi \to H_1^\varphi(\Sigma, *) \to H_0^\varphi(*) \to 0,$$

we get $H_0^\varphi(X; \mathbb{Z}[t^{\pm 1}]) \cong \mathbb{Z}[t^{\pm 1}]/(t-1) \cong \mathbb{Z}$ and $H_i^\varphi(X; \mathbb{Z}[t^{\pm 1}]) = 0$ for $i > 0$. This implies that the homology of X with coefficients in $\mathbb{Q}(t)$ is trivial, so that the torsion won't depend on the choice of a homology basis. Set $c_2 = (\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$, $c_1 = (\tilde{\alpha}, \tilde{\beta}, \tilde{x}, \tilde{y})$ and $c_0 = (\tilde{*})$ as complex bases for the above complex and $b_1 = (\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$ and $b_0 = ((t-1)\tilde{*})$ as bases of the images of the boundary map.

Then the torsion is given by

$$\tau^{\varphi}(X) = \left[\frac{b_1}{c_2}\right]^{-1}\left[\frac{b_1 b_0}{c_1}\right]\left[\frac{b_0}{c_0}\right]^{-1} = -t(t-1)^{-1} \in \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}].$$

Finally, we consider the monodromy of the open book on ∂X. We set $a_1 = \alpha$, $a_2 = \beta$ and $a_3 = \gamma$; note that $a_3 = e' - \alpha + \beta$ in $H_1(\Sigma, \partial\Sigma)$. Starting with $R_0 = 0$ and $e_1 = \binom{e}{e'}$, we get $R_1 = \binom{0}{0}$ and $e_2 = e_1$, then $R_2 = \binom{1}{0}$ and $e_3 = \binom{e+\beta}{e'}$, and finally $R_3 = \binom{-1}{0}$ and $e_4 = \binom{e-e'+\alpha}{e'}$. Utilizing the basis (e, e', α) of $\mathcal{J}_1$, we obtain

$$e^t e_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

as the matrix giving the action of the monodromy in the basis (e, e') of $H_1(\Sigma_1, \partial\Sigma)$.

To get the homology of ∂X, we start with $\varepsilon_1 = (0,0)$ and the computation gives $\varepsilon_2 = (0,0)$, $\varepsilon_3 = (\beta, 0)$, $\varepsilon_4 = (\alpha - y, 0)$. It follows that the homology of ∂X is the homology of the complex

$$0 \to \mathbb{Z} \xrightarrow{0} \langle e, e' \rangle \xrightarrow{\xi} \langle x, y \rangle \xrightarrow{0} \mathbb{Z} \to 0,$$

where $\xi(e) = -y$ and $\xi(e') = 0$. Thus $H_1(\partial X) \cong H_2(\partial X) \cong \mathbb{Z}$.

Example 3. The quadrisection diagram $(\Sigma; (c_i)_{1\le i\le 4})$ represents the manifolds $S^2 \times S^2$ (see for instance [Islambouli and Naylor 2024], or decompose each factor S^2 into two disks and recover this quadrisection). We shall use it to recover the homology with coefficients in $\mathbb{Z}$ and the intersection form of $S^2 \times S^2$.

In $H_1(\Sigma) = \langle c_1, c_2 \rangle$, we have $L_1 = L_3 = \langle c_1 \rangle$ and $L_2 = L_4 = \langle c_2 \rangle$. All pairwise intersections are trivial. The homology of $S^2 \times S^2$ is the homology of the complex

$$\mathbb{Z} \to 0 \to \bigoplus_{1\le i\le 4} L_i \to H_1(\Sigma) \xrightarrow{0} \mathbb{Z},$$

giving $H_1(X) = 0$, $H_2(X) \cong \mathbb{Z}^2$ and $H_3(X) = 0$.

A basis of $H_2(S^2 \times S^2)$ is given by $(c_1, 0, -c_1, 0)$ and $(0, c_2, 0, -c_2)$ in $\bigoplus_{1\le i\le 4} L_i$. In this basis, we obtain the intersection form as $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$.

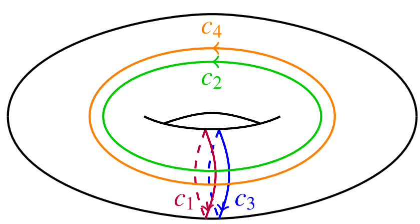


Figure 7. A quadrisection diagram of $S^2 \times S^2$.

Acknowledgements

This work has its source in the Spring Trisectors Workshop organized in 2019 at Georgia University by David Gay, Jeffrey Meier and Alexander Zupan; we are deeply grateful to the organizers. We also thank the participants, especially David Gay and Peter Feller, for motivating conversations.

References

[Blanchfield 1957] R. C. Blanchfield, "Intersection theory of manifolds with operators with applications to knot theory", *Ann. of Math.* (2) **65** (1957), 340–356. MR Zbl

[Castro et al. 2018a] N. A. Castro, D. T. Gay, and J. Pinzón-Caicedo, "Diagrams for relative trisections", *Pacific J. Math.* **294**:2 (2018), 275–305. MR Zbl

[Castro et al. 2018b] N. A. Castro, D. T. Gay, and J. Pinzón-Caicedo, "Trisections of 4-manifolds with boundary", *Proc. Natl. Acad. Sci. USA* **115**:43 (2018), 10861–10868. MR Zbl

[Etnyre and Ozbagci 2008] J. B. Etnyre and B. Ozbagci, "Invariants of contact structures from open books", *Trans. Amer. Math. Soc.* **360**:6 (2008), 3133–3151. MR Zbl

[Feller et al. 2018] P. Feller, M. Klug, T. Schirmer, and D. Zemke, "Calculating the homology and intersection form of a 4-manifold from a trisection diagram", *Proc. Natl. Acad. Sci. USA* **115**:43 (2018), 10869–10874. MR Zbl

[Florens and Moussard 2022] V. Florens and D. Moussard, "Torsions and intersection forms of 4-manifolds from trisection diagrams", *Canad. J. Math.* **74**:2 (2022), 527–549. MR Zbl

[Gay and Kirby 2016] D. Gay and R. Kirby, "Trisecting 4-manifolds", *Geom. Topol.* **20**:6 (2016), 3097–3132. MR Zbl

[Islambouli and Naylor 2024] G. Islambouli and P. Naylor, "Multisections of 4-manifolds", *Trans. Amer. Math. Soc.* **377**:2 (2024), 1033–1068. MR Zbl

[Milnor 1966] J. Milnor, "Whitehead torsion", *Bull. Amer. Math. Soc.* **72** (1966), 358–426. MR Zbl

[Reidemeister 1939] K. Reidemeister, "Durchschnitt und Schnitt von Homotopieketten", *Monatsh. Math. Phys.* **48** (1939), 226–239. MR Zbl

[Tanimoto 2023] H. Tanimoto, "Homology of relative trisection and its applications", *J. Knot Theory Ramifications* **32**:4 (2023), art. id. 2350024. MR Zbl

[Turaev 2001] V. Turaev, *Introduction to combinatorial torsions*, Birkhäuser, 2001. MR Zbl

Received December 29, 2021. Revised January 11, 2024.

DELPHINE MOUSSARD
INSTITUT DE MATHÉMATIQUES DE MARSEILLE
AIX-MARSEILLE UNIVERSITY
MARSEILLE
FRANCE
delphine.moussard@univ-amu.fr

TRENTON SCHIRMER
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF GEORGIA
ATHENS, GA
UNITED STATES
trenton.schirmer@uga.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 327, No. 1, 2023
https://doi.org/10.2140/pjm.2023.327.167

APPROXIMATION OF REGULAR SASAKIAN MANIFOLDS

GIOVANNI PLACINI

We investigate the problem of approximating a regular Sasakian structure by CR immersions in a standard sphere. Namely, we show that this is always possible for compact Sasakian manifolds. We prove an approximation result for noncompact η-Einstein manifolds via immersions in the infinite-dimensional sphere and complement this with several examples.

1. Introduction and statements of the main results

Sasakian geometry is often considered the odd-dimensional analogue of Kähler geometry. This is due to the fact that a Sasakian manifold sits in a so-called "Kähler sandwich". Namely, a $(2n+1)$-dimensional Sasakian manifold comes with a Kähler $(2n+2)$-dimensional cone and a transverse Kähler geometry of dimension $2n$. This interplay translates to the fact that the solution of some problems in Sasakian geometry is equivalent to that of others in its older even-dimensional analogue. The problem considered in this paper falls into this case. Namely, we ask whether a given regular Sasakian structure can be approximated by CR immersions in a standard sphere. In analogy with a celebrated result of Tian, Ruan and Zelditch [14; 15; 16], it was proven in [9] that any compact Sasakian manifold is approximated by CR embeddings in a weighted sphere. Here we investigate two related questions. Firstly, when the Sasakian structure is regular, it is natural to ask whether one can get a similar result to [9, Theorem 1] under the requirement that the model space is a *standard* Sasakian sphere. Our first result shows that one can trade the injectivity of the embeddings for regularity in order to obtain immersions into the standard Sasakian sphere.

Theorem 1. *Let (M, η, g) be a compact regular Sasakian manifold. Then there exist a sequence of CR immersions $\varphi_k : M \to S^{2N+1}$ into standard Sasakian spheres such that suitable transverse homotheties of the induced structures converge to (η, g) in the C^∞-norm.*

The author is supported by INdAM and GNSAGA - Gruppo Nazionale per le Strutture Algebriche, Geometriche e le loro Applicazioni and by GOACT - Funded by Fondazione di Sardegna and funded by PNRR e.INS Ecosystem of Innovation for Next Generation Sardinia (CUP F53C22000430001, codice MUR ECS00000038).

MSC2020: primary 53C25, 53C42; secondary 53C55.

Keywords: Sasakian structure approximation, CR immersions, η-Einstein manifolds, Bergman kernel.

The solution to this problem is related to Kähler geometry in the following way. In the regular case the Kähler cone of a Sasakian manifold M is the total space of a line bundle L over a Kähler manifold X (without the zero section). It turns out that one can construct such immersions by means of an orthonormal basis for the space of sections of L with respect to a certain scalar product. Finding such a basis is a classical problem in Kähler geometry deeply connected with the computation of Bergman kernels, special metrics and approximation of metrics; see, for instance, [6; 7; 8; 12; 15]. Notice that one cannot avoid transverse homotheties because the transverse Kähler metric induced by an immersion in the standard sphere is an integral basic class while this is not necessarily true for the Sasakian structures we want to approximate.

Theorem 1 above heavily relies on the compactness of the Sasakian manifold M. Our second question asks which conditions are sufficient for the existence of such approximation results in the noncompact case. This is clearly a very broad question so we focus on the case of η-Einstein manifolds. In the compact case Cappelletti-Montano and Loi [5] studied immersions of compact regular η-Einstein manifolds into spheres with codimension 2. Here we prove an approximation result for (possibly noncompact) regular complete η-Einstein manifolds.

Theorem 2. *Let M be a complete regular η-Einstein manifold. Then the Sasakian structure on M can be approximated by suitable $\mathcal{D}$-homotheties of a sequence of Sasakian structures induced by CR embeddings in S^∞.*

Also in this case, the immersions are constructed from a basis of the space of sections of a certain line bundle L over a Kähler manifold X. The main difference with Theorem 1 is the fact that X is not compact so that the space of sections of L could be infinite-dimensional.

As a particular case, all homogeneous Sasakian manifolds can be endowed with homogeneous η-Einstein metrics. One should compare our result with [11, Theorem 1.5] where the authors classify homogeneous Sasakian manifolds which admit an immersion in S^∞. In fact, the approximation in Theorem 2 is constant for those homogeneous Sasakian manifolds which can be immersed in S^∞. In the last section of the paper we exhibit two examples of genuine approximations. Namely, we consider a Sasakian structure on $\mathbb{C}^* \times S^1$ and $\mathbb{D}^* \times S^1$ where $\mathbb{C}^*$ and $\mathbb{D}^*$ are the punctured plane and disc, respectively. We provide a sequence of embeddings of $\mathbb{C}^* \times S^1$ and $\mathbb{D}^* \times S^1$ into S^∞ which approximate the given structures. In terms of Kähler geometry, we compute the orthonormal basis of the space of sections of the trivial bundle over $\mathbb{C}^*$ and $\mathbb{D}^*$.

Structure of the paper. The paper is organized as follows. In Section 2 we review the basics of Sasakian geometry with particular focus on Sasakian immersions and regular Sasakian structures. The remainder of the paper is divided into three sections. Namely, in Sections 3 and 4 we prove Theorems 1 and 2, respectively. Finally, Section 5 contains the computation of some explicit CR immersions of noncompact η-Einstein manifolds into S^∞ approximating the given structure.

2. Sasakian manifolds

Sasakian geometry can be understood in terms of contact metric geometry and via the associated Kähler cone (see the monograph of Boyer and Galicki [4]). We will present both formulations for the reader's convenience, but we will focus mostly on the regular case for it is central in this paper. In the following all manifolds and orbifolds are assumed to be connected.

A *K-contact structure* (η, Φ, R, g) on a manifold M consists of a contact form η and an endomorphism Φ of the tangent bundle $\mathrm{T}M$, satisfying the properties

- $\Phi^2 = -\mathrm{Id} + R \otimes \eta$ where R is the Reeb vector field of η,
- $\Phi_{|\mathcal{D}}$ is an almost complex structure compatible with the symplectic form $\mathrm{d}\eta$ on $\mathcal{D} = \ker \eta$,
- the Reeb vector field R is Killing with respect to the metric

$$g(\,\cdot\,,\cdot\,) = \tfrac{1}{2}\mathrm{d}\eta(\,\cdot\,, \Phi(\,\cdot\,)) + \eta(\,\cdot\,)\eta(\,\cdot\,).$$

Given such a structure one can consider the almost complex structure I on the Riemannian cone $(M \times \mathbb{R}^+, t^2 g + \mathrm{d}t^2)$ given by

- $I = \Phi$ on $\mathcal{D} = \ker \eta$, and
- $R = I(t\partial_t)_{|\{t=1\}}$.

A *Sasakian structure* is a K-contact structure (η, Φ, R, g) such that the associated almost complex structure J is integrable, and therefore $(M \times \mathbb{R}^+, t^2 g + \mathrm{d}t^2, J)$ is Kähler. A *Sasakian* manifold is a manifold M, equipped with a Sasakian structure (η, Φ, R, g).

Equivalently, one can define Sasakian manifolds in terms of Kähler cones. Namely, a Sasakian structure on a smooth manifold M is defined to be a Kähler cone structure on $M \times \mathbb{R}^+ = Y$, that is, a Kähler structure (g_Y, J) on Y of the form $g_Y = t^2 g + \mathrm{d}t^2$ where t is the coordinate on $\mathbb{R}^+$ and g a metric on M. Then (Y, g_Y, J) is called the Kähler cone of M which, in turn, is identified with the submanifold $\{t = 1\}$. The Kähler form on Y is then given by

$$\Omega_Y = \frac{i}{2}\partial\bar{\partial} t^2.$$

The Reeb vector field on Y is defined as

$$R = J(t\partial_t).$$

This defines a holomorphic Killing vector field with metric dual 1-form

$$\eta = \frac{g_Y(R, \cdot)}{t^2} = \mathrm{d}^c \log t = i(\bar\partial - \partial)\log t,$$

where $d^c = i(\bar\partial - \partial)$. Notice that J induces an endomorphism Φ of $\mathrm{T}M$ by setting

- $\Phi = J$ on $\mathcal{D} = \ker \eta_{|\mathrm{T}M}$, and
- $\Phi(R) = 0$.

Equivalently, the endomorphism Φ is determined by g and η by setting

$$g(X, Z) = \tfrac{1}{2}\mathrm{d}\eta(X, \Phi Z) \quad \text{for } X, Z \in \mathcal{D}.$$

It is easy to see that, when restricted to $M = \{t = 1\}$, (η, Φ, R, g) is a Sasakian structure in the contact metric sense whose Kähler cone is (Y, g_Y, J) itself. When this does not lead to confusion, we will use R and η to indicate both the objects on Y and on M.

Since g and η are invariant for R, the Reeb foliation is transversally Kähler in the sense that the distribution $\mathcal{D}$ admits a Kähler structure (g^T, ω^T, J^T) which is invariant under R. Explicitly, we have

$$\omega^T = \tfrac{1}{2}\mathrm{d}\eta, \quad J^T = \Phi_{|\mathcal{D}} \quad \text{and} \quad g^T(X, Z) = \tfrac{1}{2}\mathrm{d}\eta(X, J^T Z) = g_{|\mathcal{D}}.$$

In particular, one can see that

$$\omega^T = \tfrac{1}{2}\mathrm{d}\eta = \frac{i}{2}\mathrm{d}(\bar\partial - \partial)\log t = i\partial\bar\partial \log t. \tag{1}$$

The Reeb vector field defines a foliation on M, called the Reeb foliation. A very important dichotomy of Sasakian structures is given by the regularity of the leaves of the Reeb foliation. Namely, if there exist foliated charts such that each leaf intersects a chart finitely many times, the Sasakian structure is called *quasiregular*. Other wise it is called *irregular*. If every leaf intersects every chart only once, the Sasakian structure is said to be *regular*. Compact regular and quasiregular Sasakian manifold are fairly well understood due to the following result.

Theorem 3 [4]. *Let (M, η, Φ, R, g) be a quasiregular compact Sasakian manifold. Then the space of leaves of the Reeb foliation (X, ω, g_ω) is a compact Kähler cyclic orbifold with integral Kähler form $\frac{1}{2\pi}\omega$ so that the projection $\pi : (M, g) \to (X, g_\omega)$ is a Riemannian submersion. Also, X is a smooth manifold if and only if the Sasakian structure on M is regular.*

Any principal S^1-orbibundle M with Euler class $-\frac{1}{2\pi}[\omega] \in H^2_{\mathrm{orb}}(X, \mathbb{Z})$ over a compact Kähler cyclic orbifold (X, ω) admits a Sasakian structure.

This result allows us to reformulate the geometry of a compact regular Sasakian manifold M in terms of the algebraic geometry of the Kähler manifold X. We will illustrate in detail this correspondence for its importance in the remainder of the paper. Let us first introduce the concept of a $\mathcal{D}$-homothetic transformation of a Sasakian structure.

Definition 4 ($\mathcal{D}$-homothety or a transverse homothety). Let (M, η, Φ, R, g) be a (not necessarily compact) Sasakian manifold and $a \in \mathbb{R}$ a positive number. One can define the Sasakian structure $(\eta_a, \Phi_a, R_a, g_a)$ from (η, Φ, R, g) as

$$\eta_a = a\eta, \quad \Phi_a = \Phi, \quad R_a = \frac{R}{a}, \quad g_a = ag + (a^2 - a)\eta \otimes \eta = ag^T + \eta_a \otimes \eta_a.$$

Equivalently, we can define the same structure on M by setting a new coordinate on the Kähler cone as $\tilde{t} = t^a$. It is clear from the formulation above that this induces on $M = \{\tilde{t} = 1\} = \{t = 1\}$ the same Sasakian structure $(\eta_a, \Phi_a, R_a, g_a)$. We will call this structure the $\mathcal{D}_a$-homothety of (η, Φ, R, g).

Now let the compact regular Sasakian manifold (M, η, Φ, R, g) be given and consider the projection $\pi : (M, g) \to (X, \omega)$ given above. Notice that π locally identifies the contact distribution $\mathcal{D}$ with the tangent space of X. Therefore, up to $\mathcal{D}$-homothety, we have that $\pi^*(\omega) = \frac{1}{2}\mathrm{d}\eta$. The endomorphism Φ determines the complex structure on X and g induces the Kähler metric g_ω, i.e., $g^T = \pi^* g_\omega$.

In this case the class $\frac{1}{2\pi}[\omega] \in H^2(X, \mathbb{Z})$ defines an ample line bundle L over X. The cone $Y = M \times \mathbb{R}^+$ is identified with the complement of the zero section in $L^{-1} = L^*$ in the following way. Let h be a Hermitian metric on L such that

$$\omega = -i\partial\bar{\partial} \log h.$$

Then its dual h^{-1} on L^{-1} defines the second coordinate of $(p, t) \in M \times \mathbb{R}^+ = L^{-1} \setminus \{0\}$ by

$$\begin{aligned} t : L^{-1} \setminus \{0\} &\to \mathbb{R}^+, \\ (p, v) &\mapsto |v|_{h_p^{-1}}, \end{aligned} \tag{2}$$

where v is a vector of L^{-1} in the fiber over p. With this notation the Kähler form on the Kähler cone $(M \times \mathbb{R}^+, t^2 g + \mathrm{d}t^2, I)$ is given by

$$\Omega = \frac{i}{2}\partial\bar{\partial} t^2. \tag{3}$$

The Sasakian structure can be read from this data as

$$\omega^T = -i\partial\bar{\partial} \log h, \quad \eta = i(\bar{\partial} - \partial) \log t. \tag{4}$$

Therefore, the choice of a Hermitian metric h on an ample line bundle L over a compact Kähler manifold X completely determines a Sasakian structure on the $U(1)$-orbibundle associated to L^{-1}. The Sasakian manifold (M, η, R, g, Φ) so obtained is called a *Boothby–Wang bundle* over (X, ω). Observe that, although the differentiable manifold is uniquely determined by $\frac{1}{2\pi}[\omega]$, the Sasakian structure does depend on, and is in fact determined by, the choice of h.

The most basic example is the standard Sasakian structure on S^{2n+1}, that is, the Boothby–Wang bundle determined by the Fubini–Study metric $h = h_{\mathrm{FS}}$ on $\mathcal{O}(1)$ over $\mathbb{C}\mathrm{P}^n$. We give the details of this construction to further illustrate the formulation above.

Example 5 (standard Sasakian sphere). Let $h = h_{\mathrm{FS}}$ be the Fubini–Study Hermitian metric on the holomorphic line bundle $\mathcal{O}(1)$ over $\mathbb{C}\mathrm{P}^n$. Recall that its dual metric h^{-1} on $\mathcal{O}(-1)\setminus\{0\} = \mathbb{C}^{n+1}\setminus\{0\}$ is given by the Euclidean norm. This defines a coordinate t on the Kähler cone $\mathcal{O}(-1)\setminus\{0\} = \mathbb{C}^{n+1}\setminus\{0\} = S^{2n+1}\times\mathbb{R}^+$. Namely, for coordinates $z = (z_0, z_1, \ldots, z_n)$ on $\mathbb{C}^{n+1}$ we have

$$t : \mathbb{C}^{n+1} \to \mathbb{R}^+, \qquad z \mapsto |z| = \sqrt{\sum_{j=0}^{n} z_j \bar{z}_j}.$$

Now the Kähler metric on the cone is nothing but the flat metric

$$\Omega_{\mathrm{flat}} = \frac{i}{2}\partial\bar{\partial}t^2 = \frac{i}{2}\sum \mathrm{d}z_j \wedge \mathrm{d}\bar{z}_j.$$

The Reeb vector field R_0 and the contact form η_0 read

$$R_0 = J(t\partial_t) = i\sum z_j\partial_{z_j} - \bar{z}_j\partial_{\bar{z}_j}, \quad \eta_0 = i(\bar{\partial}-\partial)\log t = \frac{i}{2t^2}\sum z_j \mathrm{d}\bar{z}_j - \bar{z}_j \mathrm{d}z_j.$$

It is clear that, when restricted to S^{2n+1}, η_0 and R_0, together with the round metric g_0 and the restriction Φ_0 of J to $\ker\eta_0$, give a Sasakian structure on S^{2n+1}. This corresponds exactly to the Hopf bundle $S^{2n+1} \to \mathbb{C}\mathrm{P}^n$. We have

$$\pi^*\omega_{\mathrm{FS}} = \omega^T = \tfrac{1}{2}\mathrm{d}\eta_0 = \frac{i}{2\,|z|^4}\sum_j |z_j|^2 \mathrm{d}z_j \wedge \mathrm{d}\bar{z}_j - \sum_{j,k} \bar{z}_j z_k \mathrm{d}z_j \wedge \mathrm{d}\bar{z}_k,$$

where $\pi : \mathbb{C}^{n+1}\setminus\{0\} \to \mathbb{C}\mathrm{P}^n$ is the standard projection.

Analogously, one can define the standard Sasakian structure on the infinite dimensional sphere $S^\infty = \left\{(z_o, z_1, \ldots) \in \ell^2(\mathbb{C}) : \sum |z|^2 = 1\right\}$ (all sums are now infinite). In this case the Kähler cone $S^\infty \times \mathbb{R}^+$ is the complex space $\ell^2(\mathbb{C})\setminus\{0\}$ with the flat Kähler metric and the space of Reeb leaves is $\mathbb{C}\mathrm{P}^\infty$.

In general the space of leaves of the Reeb foliations X is not even an orbifold. Nevertheless, when the Sasakian structure is regular and complete, X is a Kähler manifold; see, for example, [13].

We now switch our attention back to not necessarily compact Sasakian manifolds and recall another well known class of deformations of Sasakian structures, the so-called transverse Kähler transformations. Namely, given a Kähler cone $Y = M \times \mathbb{R}^+$, we consider all Kähler metrics on (Y, J) that are compatible with the Reeb field R. In other terms, these are potentials $\tilde{t}^2$ such that $t\partial_t = \tilde{t}\partial_{\tilde{t}}$. This means that the corresponding Kähler and contact forms satisfy

$$\widetilde{\Omega} = \Omega + i\partial\bar{\partial}e^{2f}, \quad \tilde{\eta} = \eta + d^c f$$

for a function f invariant under ∂_t and R. Such functions are called *basic functions*. We still need to identify the manifolds $\{\tilde{t} = 1\}$ and $\{t = 1\}$. This is done by means of the diffeomorphism

$$\begin{aligned} F &: Y \to Y, \\ (p, t) &\mapsto (p, te^{-f(p)}), \end{aligned}$$

which maps $\{t = 1\}$ to $\{t = e^{-f(p)}\} = \{\tilde{t} = 1\}$. It is elementary to check that η, R and $d^c f$ are invariant under F so that $\tilde{\eta} = \eta + d^c f$ holds on M. Furthermore, the transverse Kähler forms are related by $\widetilde{\omega}^T = \omega^T + i\partial\bar{\partial} f$. Notice that when the Sasaki structure is quasiregular, basic functions correspond to functions on the base orbifold X. Thus, if t comes from a Hermitian metric h^{-1} on L^{-1}, such a transformation is given by replacing h^{-1} with $e^f h^{-1}$ for a function $f : X \to \mathbb{C}$ such that $\omega + i\partial\bar{\partial} f > 0$. This is equivalent to picking a different Kähler form $\widetilde{\omega}$ in the same class as ω. We summarize the above discussion in the following definition.

Definition 6 (transverse Kähler deformations). Let (M, η, R, g, Φ) be a Sasakian manifold with Kähler cone (Y, J) and Kähler potential t^2. A transverse Kähler transformation is given by replacing t with $\tilde{t} = e^f t$ for a basic function f and leaving (Y, J, R) unchanged. When the Sasaki structure is quasiregular and given as in (4), a transverse Kähler transformation is given by replacing h^{-1} with $e^f h^{-1}$.

We are mostly interested with immersions and embeddings of Sasakian manifolds. We recall the relevant definitions. Two Sasakian manifolds $(M_1, \eta_1, R_1, g_1, \phi_1)$ and $(M_2, \eta_2, R_2, g_2, \phi_2)$ are equivalent if there exists a diffeomorphism $\varphi \colon M_1 \to M_2$ such that

$$\varphi^*\eta_2 = \eta_1 \quad \text{and} \quad \varphi^* g_2 = g_1.$$

If this holds, then φ also satisfies $\varphi_* \circ \phi_1 = \phi_2 \circ \varphi_*$ and $\varphi_* R_1 = R_2$. As implicitly intended in the definitions above, a Sasakian equivalence from a Sasakian manifold (M, η, R, g, ϕ) to itself is often called a *Sasakian transformation* of (M, η, R, g, ϕ).

One can relax the condition on Sasakian equivalences to define Sasakian embedding and immersions. Namely, one does not request the map between Sasakian manifolds to be a diffeomorphism while requiring that it preserves the Sasakian structures. In particular, given two Sasakian manifolds $(M_1, \eta_1, R_1, g_1, \phi_1)$ and $(M_2, \eta_2, R_2, g_2, \phi_2)$, a *Sasakian immersion* (*resp. embedding*) of M_1 in M_2 is an immersion (resp. embedding) $\varphi\colon M_1 \to M_2$ such that

$$\varphi^*\eta_2 = \eta_1, \quad \varphi^*g_2 = g_1, \quad \varphi_*R_1 = R_2 \quad \text{and} \quad \varphi_* \circ \phi_1 = \phi_2 \circ \varphi_*.$$

We can rephrase this definition in terms of the Kähler cone of the Sasakian manifolds M_1 and M_2. Namely, the map φ satisfying the conditions above clearly extends to a map

$$\begin{aligned} \widetilde{\varphi}\colon M_1 \times \mathbb{R} &\to M_2 \times \mathbb{R}, \\ (p, t) &\mapsto (\varphi(p), t). \end{aligned}$$

It is clear that if φ is a Sasakian immersion (resp. embedding), then $\widetilde{\varphi}$ is a Kähler immersion (resp. embedding).

If, conversely, Y_1 and Y_2 are the Kähler cones of M_1 and M_2 with coordinates t_1 and t_2, then a Kähler immersion (resp. embedding) $\widetilde{\varphi}\colon Y_1 \to Y_2$ such that $\widetilde{\varphi}^*(t_2) = t_1$ restricts to a Sasakian immersion (resp. embedding) $\varphi : M_1 \to M_2$. Since it is often more useful to our purposes, we give the following definition.

Definition 7 (Sasakian immersion and embedding). Let M_1 and M_2 be two Sasakian manifolds with Kähler cones Y_1 and Y_2 and coordinates t_1 and t_2 respectively. A *Sasakian immersion* (*resp. embedding*) of M_1 in M_2 is a Kähler immersion (resp. embedding) $\varphi\colon Y_1 \to Y_2$ such that $\varphi^*(t_2) = t_1$.

Remark 8. Given the equivalence between a Sasakian immersion $M_1 \to M_2$ and a Kähler immersion of the Kähler cones, with an abuse of notation, we will often denote both maps with the same letter.

A special class among Sasakian structures is that of η-Einstein structures. These are the Sasakian analogues of Kähler–Einstein metrics. Namely, using the canonical splitting $TM = \mathcal{D} \oplus T_{\mathcal{F}}$ where $\mathcal{D} = \ker \eta$ and $T_{\mathcal{F}}$ denotes the tangent bundle to the Reeb foliation $\mathcal{F}$, write the metric as

$$g = g^T + \eta \otimes \eta. \tag{5}$$

With an abuse of notation we write g^T for both the transverse metric and the metric on X in the quasiregular case. It follows from (5) that the Riemannian properties of M can be expressed in terms of those of the transverse Kähler geometry and of the contact form η. For instance, the Ricci tensor of g is given by

$$\mathrm{Ric}_g = \mathrm{Ric}_{g^T} - 2g. \tag{6}$$

A Sasakian manifold (M, η, ϕ, R, g) is said to be *η-Einstein* if the Ricci tensor satisfies

$$\mathrm{Ric}_g = \lambda g + \nu \eta \otimes \eta \tag{7}$$

for some constants λ, $\nu \in \mathbb{R}$. It follows from (6) and (7) that a Sasakian manifold is η-Einstein with constants (λ, ν) if and only if, its transverse geometry is Kähler–Einstein with Einstein constant $\lambda + 2$ (see, e.g., [4] for details).

2.1. ***CR immersions of regular and complete Sasakian manifolds into spheres.*** We recall now some facts about CR immersions of Sasakian manifolds into finite- and infinite-dimensional standard spheres. We only set the notation and report some useful results for us; the interested reader can refer to [11, Section 5]

Let M be a compact regular Sasakian manifold. By Theorem 3, M is a $U(1)$-principal bundle $\pi : M \to X$ over a compact Kähler manifold (X, ω) with $2\pi^*\omega = d\eta$. Furthermore, M is the unitary bundle associated to the line bundle L^{-1} where $c_1(L) = [\omega]$. This last condition implies that L is ample. In other terms, (X, L) is a polarized Kähler manifold. Therefore, for $k \in \mathbb{N}$ large enough, the bundle $L^{\otimes k} = L^k$ is very ample, and we can define the Kodaira embedding $\psi_k : X \to \mathbb{C}\mathrm{P}^{N_k}$ where $\dim(H^0(L)) = N_k + 1$. Then there exists a CR embedding $\varphi_k : M \to S^{2N_k+1}$ of M into the standard sphere covering the Kodaira embedding ψ_k or, equivalently, a holomorphic embedding of $\varphi_k : Y \to \mathbb{C}^{N_k+1} \setminus \{0\}$ of the Kähler cone $Y = M \times \mathbb{R}^+$ into the Kähler cone $S^{2N_k+1} \times \mathbb{R}^+$. In fact we have:

Proposition 9 [11, Proposition 5.1]. *Let M be the compact regular Sasakian manifold determined by the Hermitian bundle (L, h) over a compact projective manifold X. For every integer $k \gg 0$ there exists a holomorphic embedding $\varphi_k : M \times \mathbb{R}^+ \to S^{2N_k+1} \times \mathbb{R}^+$ such that $\varphi_k^*(\tau) = B_k t^k$ where B_k is the Bergman kernel of L^k, τ and t are the coordinates on the second factor of $S^{2N_k+1} \times \mathbb{R}^+$ and $M \times \mathbb{R}^+$, respectively.*

The same construction can be performed when the Sasakian manifold M is the unitary bundle associated to the positive Hermitian bundle (L, h) on a noncompact Kähler manifold (X, ω) with $\omega = -i\partial\bar{\partial}\log h$. In this case we cannot immerse M into a finite-dimensional sphere because the space of sections $H^0(L)$ is replaced by the Hilbert space $\mathcal{H}_{k,h}$ of integrable sections; see [11] for details. Nevertheless one gets the following noncompact analogue.

Proposition 10. *Let M be the regular Sasakian manifold determined by the Hermitian bundle (L, h) over a noncompact Kähler manifold X and assume the space $\mathcal{H}_{k,h}$ is nontrivial. Then there exists a holomorphic immersion $\varphi_k : M \times \mathbb{R}^+ \to S^\infty \times \mathbb{R}^+$ such that $\varphi_k^*(\tau) = \varepsilon_k t^k$ where ε_k is the ε-function of $\mathcal{H}_{k,h}$, τ and t are the coordinates on the second factor of $S^\infty \times \mathbb{R}^+$ and $M \times \mathbb{R}^+$, respectively.*

Remark 11. Although $\varphi_k^*(h_{\mathrm{FS}}^{-1})$ is not a Hermitian metric on the line bundle L^{-1} (it does not scale correctly under the $\mathbb{C}^*$-action), it defines a change of coordinate $(p, t) \mapsto (p, B_k t^k)$ (or $(p, t) \mapsto (p, \varepsilon_k t^k)$ in the noncompact case) on $M \times \mathbb{R}^+$

corresponding to the composition of the $\mathcal{D}_k$-homothetic transformation ($t \mapsto t^k$) with a transverse Kähler deformation ($t^k \mapsto B_k t^k$).

3. Approximation of compact regular structures via immersions into spheres

Proof of Theorem 1. Assume (M, η, R, g, Φ) to be a compact regular Sasakian manifold. Suppose we have performed a $\mathcal{D}$-homothetic transformation so that M is the unit bundle $\pi : M \to X$ associated to a holomorphic line bundle L^{-1} over a projective manifold (X, ω) with $\pi^*\omega = \frac{1}{2}\mathrm{d}\eta$.

We can then apply Proposition 9 to get a sequence of holomorphic immersions $\varphi_k : M \times \mathbb{R}^+ \to S^{2N_k+1} \times \mathbb{R}^+$ such that $\varphi_k^*(\tau) = B_k t^k$ where B_k is the Bergman kernel of L^k, τ and t are the coordinates on the second factor of $S^{2N_k+1} \times \mathbb{R}^+$ and $M \times \mathbb{R}^+$, respectively. Notice that τ is the coordinate induced by the flat metric on $\mathbb{C}^{N_k+1} \setminus \{0\} = S^{2N_k+1} \times \mathbb{R}^+$ or, equivalently, by the Hermitian metric h_{FS} on $\mathcal{O}(-1)$ whose curvature is $-\omega_{\mathrm{FS}}$.

Now the $\frac{1}{k}$-transverse homothety of the structure induced on M by the immersion into S^{2N_k+1} is a transverse Kähler deformation of the original Sasakian structure determined by the Bergman kernel B_k, (compare Definition 6 and Remark 11). By [16, Corollary 2] the first coefficient of the asymptotic expansion of the Bergman kernel B_k smoothly converges to 1 when k goes to infinity. Therefore, the $\mathcal{D}_{1/k}$-homotheties of the structures determined by pullback coordinates $\varphi_k^*(\tau) = B_k t^k$ converge smoothly to (η, R, g, Φ).

We can resume the maps involved in the proof, with the notation of Section 2.1, in the diagram

$$\begin{tikzcd}
(M, \eta_k, g_k) \arrow[d, "p_k"] \arrow[dr, "\varphi_k"] \arrow[dd, bend right, "\pi"'] & \\
(M_k, \bar\eta_k, \bar g_k) \arrow[r, "\widetilde{\psi}_k"] \arrow[d, "\pi_k"] & (S^{2N_k+1}, \eta_0, g_0) \arrow[d, "\pi_{\mathrm{FS}}"] \\
(X, \omega_k) \arrow[r, "\psi_k"] & \left(\mathbb{CP}^{N_k}, \omega_{\mathrm{FS}}\right)
\end{tikzcd}$$

where $(M_k, \bar\eta_k, \bar g_k)$ is the unit bundle associated to L^{-k} endowed with the Sasakian structure pulled back via ψ_k and (M, η_k, g_k) is the Sasakian structure determined by the coordinate $\varphi_k^*(\tau) = B_k t^k$. □

Remark 12. Notice that we used a $\mathcal{D}$-homothety as the first step of the proof to get an actual Boothby–Wang bundle $\pi : M \to X$. To avoid this and obtain the convergence to the original Sasakian metric, one can compose the $\mathcal{D}_{1/k}$-homothety in our proof with the inverse of the homothetic transformation considered in the beginning.

4. Approximation of η-Einstein regular structures

Proof of Theorem 2. We cannot deduce that M is an S^1-bundle over a Kähler manifold because M is not necessarily compact. Nevertheless, the Reeb foliation still defines a fibration $\pi : M \to X$ over a Kähler manifold (X, ω) because M is regular and complete; see [13]. Now the fiber of this fibration is either $\mathbb{R}$ or S^1.

Let us deal first with the case where the fiber is S^1. Regardless of whether or not M is compact, since the Sasakian structure on M is regular and the fiber is S^1, it is the unit bundle of a line bundle L^{-1} over X such that $c_1(L) = [\omega]$. Choose a Hermitian metric h on L whose Ricci curvature form is ω. Notice that (X, ω) is Kähler–Einstein because M is η-Einstein.

We now invoke a result of Ma and Marinescu on the Bergman kernel of noncompact manifolds. Namely, we apply [12, Theorem 6.1.1] to the line bundle L over X. The hypotheses of this theorem are satisfied as (X, ω) is a Kähler–Einstein manifold so that there exists a positive constant C such that $i\mathrm{Ric}(\omega) > C\omega$. In our case this implies that the space of sections $\mathcal{H}_{k,h}$ is nontrivial so that Proposition 10 provides a sequence of holomorphic immersions $\varphi_k : M \times \mathbb{R}^+ \to S^\infty \times \mathbb{R}^+$ such that $\varphi_k^*(\tau) = \varepsilon_k t^k$ where ε_k is the ε-function of $\mathcal{H}_{k,h}$, τ and t are the coordinates on the second factor of $S^\infty \times \mathbb{R}^+$ and $M \times \mathbb{R}^+$, respectively.

Again by [12, Theorem 6.1.1] (see also [1]) the ε-function ε_k admits an asymptotic expansion whose first coefficient is 1. Therefore, taking the $\frac{1}{k}$-homothety of the Sasakian structure on M defined by the pullback coordinate $\varphi_k^*(\tau) = \varepsilon_k t^k$ we get a sequence of structures which converge to the given η-Einstein one for $k \to \infty$.

Now the argument when the fiber is $\mathbb{R}$ easily follows from the previous one. Namely, in this case the fibration is trivial, i.e., $M \cong X \times \mathbb{R}$. Since $\mathbb{Z}$ acts on $X \times \mathbb{R}$ by Sasakian isometries via the flow of the Reeb vector field, the quotient is the η-Einstein manifold $N = X \times S^1$ and the $\mathbb{Z}$-covering map $\tilde{\pi} : M \to N$ is a Sasakian immersion. Now N is an η-Einstein manifold fibering over a Kähler–Einstein manifold X with fiber S^1. By the previous case, there exists a sequence of CR immersions $\varphi_k : N \to S^\infty$ such that suitable $\mathcal{D}$-homotheties of the induced structures converge on N to the original η-Einstein structure. Therefore, the pullback to M of such structures under $\tilde{\pi}$ converge to the η-Einstein structure we began with. Notice that these structures are transverse homotheties of the ones induced via the CR immersions $\tilde{\pi} \circ \varphi_k : M \to S^\infty$. That is, we can perform the transverse homotheties on N or on M interchangeably. This concludes the proof. □

5. Explicit examples of approximations of η-Einstein structures

We exploit the equivalence between polarizations (L, h) of a Kähler manifold X and Sasakian structures on a Boothby–Wang bundle over X to describe explicitly some embeddings of noncompact inhomogeneous η-Einstein manifolds into S^∞.

Namely, we compute an orthonormal basis for the Hilbert space $\mathcal{H}_{k,h}$ of sections of a line bundle L over a noncompact inhomogeneous Kähler manifold. This provides instances of approximations of inhomogeneous η-Einstein metrics which cannot be isometrically CR immersed in a sphere.

Example 13 (fibring on the punctured plane). Consider the punctured plane $\mathbb{C}^* = \mathbb{C}\setminus\{0\}$ endowed with the complete Calabi–Yau metric g_0^* induced by the Kähler form

$$\omega_0^* = \frac{i}{2}\frac{\mathrm{d}z\wedge \mathrm{d}\bar{z}}{|z|^2}, \tag{8}$$

where z is the coordinate on $\mathbb{C}^*$. Since this Kähler form admits a global potential $F = \frac{1}{2}\log^2|z|^2$, it is exact. Therefore, we can endow $\mathbb{C}^*\times S^1$, i.e., the unit bundle of the trivial bundle $L = \mathbb{C}^*\times\mathbb{C}$, with an η-Einstein structure. Namely, denoting the standard volume form on S^1 by α, the contact form η on $\mathbb{C}^*\times S^1$ is given by $\eta = \alpha + i(\bar\partial - \partial)\log F$. The Sasakian metric is $g = g_0^* + \eta\otimes\eta$ and the endomorphism ϕ is given by the lift of the complex structure of $\mathbb{C}^*$ to the contact distribution. We want to give an explicit expression of the embeddings of the η-Einstein manifolds $\mathbb{C}^*\times S^1$ just described into S^∞.

The Kähler space $(\mathbb{C}^*, \omega_0^*)$ and its polarizations were studied by Loi and Zuddas in [10]. We report here the essential points which are relevant to our discussion. For any positive integer k

$$h^k(f(z), f(z)) = e^{\frac{-k}{2}\log^2|z|^2}|f(z)|^2 \tag{9}$$

is a Hermitian metric on L^k whose curvature is $k\omega_0^*$. By the discussion in the previous section, it is enough to compute an orthonormal basis of $\mathcal{H}_{k,h}$ to get the components of the embedding φ_k of $L^{-k}\setminus\{0\}$ into $\ell^2(\mathbb{C})$; see also [11, Section 5]. Namely, we need sections s_j such that

$$\langle s_j, s_j\rangle_k = \int_{\mathbb{C}^*} h^k(s_j(z), s_j(z))\omega_0^* = \int_{\mathbb{C}^*} e^{\frac{-k}{2}\log^2|z|^2}|s_j(z)|^2\frac{i}{2}\frac{\mathrm{d}z\wedge\mathrm{d}\bar{z}}{|z|^2} = 1 \tag{10}$$

and such that $\langle s_j, s_l\rangle_k = 0$ for $j\neq l$. It is easy to check that the functions z^j for $j\in\mathbb{Z}$ are orthogonal and they form a basis of $\mathcal{H}_{k,h}$ for all k since holomorphic functions are determined by their Laurent series. A simple computation shows that

$$\langle z^j, z^j\rangle_k = \int_{\mathbb{C}^*} e^{\frac{-k}{2}\log^2|z|^2}|z|^{2j}\frac{i}{2}\frac{\mathrm{d}z\wedge\mathrm{d}\bar{z}}{|z|^2} = \sqrt{\frac{2}{k}}\pi^{\frac{3}{2}}e^{\frac{j^2}{2k}}. \tag{11}$$

Hence an orthonormal basis for the Hilbert space $\mathcal{H}_{k,h}$ consists of the sections

$$s_{k,j} = \left(\frac{\sqrt{k}e^{\frac{-j^2}{2k}}}{\sqrt{2}\pi^{\frac{3}{2}}}\right)^{\frac{1}{2}} z^j$$

for $j \in \mathbb{Z}$. In other words, the sections $s_{k,j}$ are the components of a holomorphic immersion φ_k of $\mathbb{C}^* \times \mathbb{C}^*$ into $\ell^2(\mathbb{C})$ and the potential of the induced transverse metric is

$$F_k := \varphi_k^*(|\cdot|^2) = \sum_{j \in \mathbb{Z}} \frac{\sqrt{k} e^{\frac{-j^2}{2k}}}{\sqrt{2}\pi^{\frac{3}{2}}} |z|^{2j}$$

so that the induced Hermitian metric on L^{-k} is

$$\varphi_k^*(h_{\mathrm{FS}}^{-1})(f(z), f(z)) = e^{F_k} |f(z)|^2.$$

One can check that the k-th root of this Hermitian metric converges to (a multiple of) the metric $h = e^F |\cdot|^2$ without invoking [12, Theorem 6.1.1]; see [10, Theorem 3.6] for a direct proof.

Example 14 (fibring on the punctured disc). As in the previous example we will construct noncompact Sasakian manifolds fibring over a noncompact inhomogeneous Kähler manifold X with a global Kähler potential. Here we take $X = \mathbb{D}^* = \{z \in \mathbb{C} : 0 < |z|^2 < 1\}$ equipped with the hyperbolic Kähler–Einstein metric

$$\omega_{\mathrm{hyp}}^* = \frac{i}{2} \frac{\mathrm{d}z \wedge \mathrm{d}\bar{z}}{|z|^2 \log^2(|z|^2)} \tag{12}$$

whose potential is $F = -\log(-\log |z|^2)$. By Theorem 2, in analogy with the previous example, we can endow $\mathbb{D}^* \times S^1$ with an η-Einstein structure with contact structure $\eta = \alpha + i(\bar{\partial} - \partial) \log F$. By Theorem 2 this Sasakian structure can be approximated by (suitable transverse homotheties) of structures induced by immersions of $\mathbb{C}^* \times S^1$ into S^∞. We want to give here the explicit expression of these immersions.

The Kähler space $(\mathbb{C}^*, \omega_{\mathrm{hyp}}^*)$ was studied in [2; 3] in relation to Bergman kernels of punctured surfaces. The polarization we are interested in is the k-th powers of the trivial line bundle endowed with the Hermitian metric

$$h^k(f(z), f(z)) = e^{k \log(-\log |z|^2)} |f(z)|^2. \tag{13}$$

We compute an orthonormal basis of $\mathcal{H}_{k,h}$ to get the components of the embedding φ_k of $L^{-k} \setminus \{0\}$ into $\ell^2(\mathbb{C})$; see also [11, Section 5]. Namely, we need sections s_j such that

$$\langle s_j, s_j \rangle_k = \int_{\mathbb{D}^*} h^k(s_j(z), s_j(z)) \omega_{\mathrm{hyp}}^* = \int_{\mathbb{D}^*} e^{k \log(-\log |z|^2)} |s_j(z)|^2 \frac{i}{2} \frac{\mathrm{d}z \wedge \mathrm{d}\bar{z}}{|z|^2 \log^2 |z|^2} = 1$$

and such that $\langle s_j, s_l \rangle_k = 0$ for $j \neq l$. It is easy to check that if a holomorphic function on $\mathbb{D}^*$ has finite norm, then its Laurent expansion involves only the terms z^j for positive $j \in \mathbb{Z}$. The functions z^j for $j > 0$ are orthogonal and they form a basis of $\mathcal{H}_{k,h}$ for all k. We can then compute

$$\begin{aligned}\langle z^j, z^j\rangle_k &= \int_{\mathbb{D}^*} e^{k\log(-\log|z|^2)}|z|^{2j}\frac{i}{2}\frac{\mathrm{d}z\wedge\mathrm{d}\bar{z}}{|z|^2\log^2|z|^2}\\ &= \frac{i}{2}\int_{\mathbb{D}^*}(-\log|z|^2)^{k-2}|z|^{2j-2}\,\mathrm{d}z\wedge\mathrm{d}\bar{z}\\ &= 2\pi\int_0^1(-\log\rho^2)^{k-2}\rho^{2j-1}\,\mathrm{d}\rho,\end{aligned}$$

where the last equality is obtained passing to polar coordinates. Substituting $e^x=\rho^2$ first and $-jx=w$ one gets

$$\begin{aligned}\langle z^j, z^j\rangle_k &= 2\pi\int_0^1(-\log\rho^2)^{k-2}\rho^{2j-1}\,\mathrm{d}\rho\\ &= \pi\int_{-\infty}^0(-x)^{k-2}e^{jx}\,\mathrm{d}x\\ &= \frac{\pi}{j^{k-1}}\int_0^\infty w^{k-2}e^{-w}\,\mathrm{d}w = \frac{\pi(k-2)!}{j^{k-1}}.\end{aligned}$$

Hence an orthonormal basis for the Hilbert space $\mathcal{H}_{k,h}$ consists of the sections

$$s_{k,j} = \left(\frac{j^{k-1}}{\pi(k-2)!}\right)^{\frac{1}{2}} z^j$$

for $j>0$ and these are the components of the holomorphic immersion φ_k of $\mathbb{D}^*\times\mathbb{C}^*$ into $\ell^2(\mathbb{C})$. In particular the potential of the induced transverse metric is

$$F_k := \varphi_k^*(|\cdot|^2) = \sum_{j>0}\frac{j^{k-1}|z|^{2j}}{\pi(k-2)!}$$

so that the induced Hermitian metric on L^{-k} is

$$\varphi_k^*(h_{\mathrm{FS}}^{-1})\,(f(z), f(z)) = e^{F_k}|f(z)|^2.$$

The k-th root of this Hermitian metric converges to (a multiple of) the metric $h=e^F|\cdot|^2$ by [12, Theorem 6.1.1].

Remark 15. Notice that we can lift the η-Einstein structure of $\mathbb{C}^*\times S^1$ (resp. $\mathbb{D}^*\times S^1$) to $\mathbb{C}^*\times\mathbb{R}$ (resp. $\mathbb{D}^*\times\mathbb{R}$). As in the proof of Theorem 2, by composing with the covering map, we can lift the immersions into S^∞ too.

Observe that none of these Sasakian manifolds are homogeneous Sasakian so that we provided explicit Sasakian immersions φ_k of regular inhomogeneous η-Einstein manifolds into S^∞ (when considered with the induced structure). This should be compared with [11, Theorem 1.5] where it is proven that a homogeneous Sasakian manifold can be immersed into S^∞ if and only if its fundamental group is cyclic.

Our examples show that, if the manifold is not assumed to be homogeneous, there is no such restriction on the fundamental group.

References

[1] C. Arezzo, A. Loi, and F. Zuddas, "Szegö kernel, regular quantizations and spherical CR-structures", *Math. Z.* **275**:3-4 (2013), 1207–1216. MR Zbl

[2] H. Auvray, X. Ma, and G. Marinescu, "Bergman kernels on punctured Riemann surfaces", *Math. Ann.* **379**:3-4 (2021), 951–1002. MR Zbl

[3] H. Auvray, X. Ma, and G. Marinescu, "Quotient of Bergman kernels on punctured Riemann surfaces", *Math. Z.* **301**:3 (2022), 2339–2367. MR Zbl

[4] C. P. Boyer and K. Galicki, *Sasakian geometry*, Oxford University Press, 2008. MR Zbl

[5] B. Cappelletti-Montano and A. Loi, "Einstein and η-Einstein Sasakian submanifolds in spheres", *Ann. Mat. Pura Appl.* (4) **198**:6 (2019), 2195–2205. MR Zbl

[6] S. K. Donaldson, "Scalar curvature and projective embeddings, I", *J. Differential Geom.* **59**:3 (2001), 479–522. MR Zbl

[7] S. K. Donaldson, "Scalar curvature and projective embeddings, II", *Q. J. Math.* **56**:3 (2005), 345–356. MR Zbl

[8] A. Loi and R. Mossa, "Uniqueness of balanced metrics on holomorphic vector bundles", *J. Geom. Phys.* **61**:1 (2011), 312–316. MR Zbl

[9] A. Loi and G. Placini, "Any Sasakian structure is approximated by embeddings into spheres", preprint, 2022. arXiv 2210.00790

[10] A. Loi and D. Zuddas, "Some remarks on Bergmann metrics", *Riv. Mat. Univ. Parma* (6) **4** (2001), 71–86. MR Zbl

[11] A. Loi, G. Placini, and M. Zedda, "Immersions into Sasakian space forms", preprint, 2023. arXiv 2305.05509

[12] X. Ma and G. Marinescu, *Holomorphic Morse inequalities and Bergman kernels*, Progress in Mathematics **254**, Birkhäuser, Basel, 2007. MR Zbl

[13] B. L. Reinhart, "Foliated manifolds with bundle-like metrics", *Ann. of Math.* (2) **69** (1959), 119–132. MR Zbl

[14] W.-D. Ruan, "Canonical coordinates and Bergmann metrics", *Comm. Anal. Geom.* **6**:3 (1998), 589–631. MR Zbl

[15] G. Tian, "On a set of polarized Kähler metrics on algebraic manifolds", *J. Differential Geom.* **32**:1 (1990), 99–130. MR Zbl

[16] S. Zelditch, "Szegö kernels and a theorem of Tian", *Internat. Math. Res. Notices* 6 (1998), 317–331. MR Zbl

Received May 10, 2023. Revised December 20, 2023.

GIOVANNI PLACINI
DIPARTIMENTO DI MATEMATICA E INFORMATICA
UNIVERSITÀ DEGLI STUDI DI CAGLIARI
CAGLIARI
ITALY

giovanni.placini@unica.it

Guidelines for Authors

Authors may submit articles at msp.org/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 327 No. 1 November 2023