Pacific Journal of Mathematics

Volume 285 No. 2 December 2016

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 2016 is US $440/year for the electronic version, and $600/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 0030-8730) 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. 285, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.285.257

THE SU(N) CASSON–LIN INVARIANTS FOR LINKS

HANS U. BODEN AND ERIC HARPER

We introduce the SU(N) Casson–Lin invariants for links L in S^3 with more than one component. Writing $L = \ell_1 \cup \cdots \cup \ell_n$, we require as input an n-tuple $(a_1, \ldots, a_n) \in \mathbb{Z}^n$ of *labels*, where a_j is associated with ℓ_j. The SU(N) Casson–Lin invariant, denoted $h_{N,a}(L)$, gives an algebraic count of certain projective SU(N) representations of the link group $\pi_1(S^3 \smallsetminus L)$, and the family $h_{N,a}$ of link invariants gives a natural extension of the SU(2) Casson–Lin invariant, which was defined for knots by X.-S. Lin and for 2-component links by Harper and Saveliev. We compute the invariants for the Hopf link and more generally for chain links, and we show that, under mild conditions on the labels $(a_1, \ldots, a_n)$, the invariants $h_{N,a}(L)$ vanish whenever L is a split link.

Introduction

The goal of this paper is to construct SU(N) Casson–Lin invariants $h_{N,a}(L)$ for oriented links L in S^3. These invariants are defined as a signed count of conjugacy classes of certain irreducible projective SU(N) representations of $\pi_1(S^3 \smallsetminus L)$ with a nontrivial 2-cocycle. Given an oriented link L with n components, the 2-cocycle is determined by an n-tuple $a = (a_1, \ldots, a_n) \in \mathbb{Z}^n$ of labels, and the choice of labels is made so that the resulting 2-cocycle is nontrivial. This is critical in what follows because it prohibits the existence of reducibles; see Proposition 2.2. We denote the resulting algebraic count as $h_{N,a}(L)$, and the following theorem is the main result of this paper.

Main theorem. *Suppose $L \subset S^3$ is an oriented n-component link with $n \geq 2$ and $a = (a_1, \ldots, a_n)$ is an allowable n-tuple of labels. Then the integer $h_{N,a}(L)$ is a well-defined invariant of L.*

We briefly outline how the above theorem is established. By Alexander's theorem [1923], every link $L \subset S^3$ can be realized as the closure $L = \hat{\sigma}$ for some braid $\sigma \in B_k$.

Boden was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. Harper was supported by CIRGET and McMaster postdoctoral fellowships.
MSC2010: 20C15, 57M25.
Keywords: braids, links, representation spaces, Casson–Lin invariant.

The braid group B_k acts naturally on the free group F_k, and this induces an action on the space of $\mathrm{SU}(N)$ representations of F_k, which we denote as

$$R_k = \mathrm{Hom}(F_k, \mathrm{SU}(N)) = \mathrm{SU}(N) \times \cdots \times \mathrm{SU}(N).$$

We extend this action to the wreath product $\mathbb{Z}_N \wr B_k$ as follows. Identifying $\mathbb{Z}_N$ with the center of $\mathrm{SU}(N)$, for $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_k) \in (\mathbb{Z}_N)^k$ and $X = (X_1, \ldots, X_k) \in R_k$, we set $(\varepsilon, \sigma)(X) = (\varepsilon_1 \sigma(X)_1, \ldots, \varepsilon_k \sigma(X)_k)$. This extends the braid group action on R_k to an action of $\mathbb{Z}_N \wr B_k$, and in fact every fixed point $\mathrm{Fix}(\varepsilon\sigma) \subseteq R_k$ can be identified with a projective representation of the link group $G_L = \pi_1(S^3 \smallsetminus L)$.

The key result is Proposition 2.2, which shows that every element $X \in \mathrm{Fix}(\varepsilon\sigma)$ is irreducible. Consequently, writing $R_k^* \subset R_k$ for the subspace of irreducible $\mathrm{SU}(N)$ representations, Proposition 2.2 implies that the graph $\Gamma_{\varepsilon\sigma}^*$ and the diagonal Δ_k^* intersect in a compact subset of $R_k^* \times R_k^*$. It follows that one can arrange transversality of the intersection $\Gamma_{\varepsilon\sigma}^* \cap \Delta_k^*$ by a compactly supported isotopy, and using natural orientations on the quotients $\hat{\Gamma}_{\varepsilon\sigma} = \Gamma_{\varepsilon\sigma}^* / \mathrm{PU}(N)$ and $\hat{\Delta}_k = \Delta_k^* / \mathrm{PU}(N)$, we define $h_{N,a}(\varepsilon\sigma)$ as the oriented intersection number of $\hat{\Gamma}_{\varepsilon\sigma}$ and $\hat{\Delta}_k$. Our main result is then established by showing that $h_{N,a}(\varepsilon\sigma)$ is independent of the choice of compatible k-tuple $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_k)$ (Proposition 3.4), and that it is invariant under the two Markov moves (Propositions 3.5 and 3.6). It follows that $h_{N,a}(L)$ gives a well-defined invariant of the link $L \subset S^3$.

One of the virtues of this approach is that it leads to a direct method for computing the invariants, and we illustrate this by computing $h_{N,a}(L)$ for the Hopf link and for chain links (Propositions 4.3 and 4.5) and by showing that the invariants vanish for split links (Proposition 4.6).

Gauge Theory. One motivation for defining link invariants in terms of the $\mathrm{SU}(N)$ representations of the link group is that these representations can be identified with flat connections on a principal $\mathrm{SU}(N)$ bundle over the link exterior, which allows for a gauge theoretic interpretation. This approach was originally used by Taubes [1990] to interpret Casson's invariant $\lambda(\Sigma)$ of homology 3-spheres Σ in terms of flat $\mathrm{SU}(2)$ connections, and using similar ideas, Floer [1988] defined $\mathbb{Z}_8$-graded groups $\mathrm{HF}_*(\Sigma)$ called the instanton Floer homology and whose Euler characteristic equals the Casson invariant.

The Casson–Lin invariants can also be interpreted gauge theoretically, as we now explain. X.-S. Lin [1992] originally defined the invariant $h(K)$ of knots $K \subset S^3$ as an algebraic count of conjugacy classes of *tracefree* irreducible $\mathrm{SU}(2)$ representations of $\pi_1(S^3 \smallsetminus K)$ and proved that $h(K) = \mathrm{sign}(K)/2$, half the signature of K. More recently, C. Herald [1997] used gauge theory to define an extended Casson–Lin invariant $h_\alpha(K)$ for knots $K \subset \Sigma^3$ in homology 3-spheres which allows for more general meridional trace conditions, and he generalized Lin's formula by showing that $h_\alpha(K) = \operatorname{sign}_\alpha(K)/2$, half the Tristram–Levine α-twisted signature of K. (Similar results were obtained by M. Heusener and J. Kroll [1998].) O. Collin and B. Steer [1999] then used moduli spaces of orbifold connections to define an associated Floer homology theory for knots whose Euler characteristic equals $h_\alpha(K)$ and P. Kronheimer and T. Mrowka [2011b] further developed the instanton Floer homology theory of knots in, and they used it to prove a strong nontriviality result for Khovanov homology [Kronheimer and Mrowka 2011a].

Harper and N. Saveliev [2010] used projective SU(2) representations to extend the Casson–Lin invariant to 2-component links L in S^3, and they showed that $h(L) = \pm\operatorname{lk}(\ell_1, \ell_2)$, where $\operatorname{lk}(\ell_1, \ell_2)$ is the linking number of $L = \ell_1 \cup \ell_2$. They gave a gauge theoretic description of the invariant $h(L)$ in [2012], where they also described Floer homology groups with Euler characteristic equal to $h(L)$.

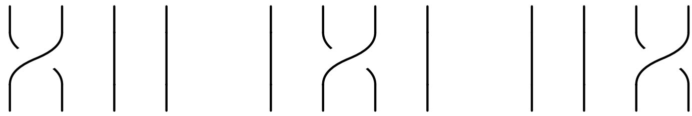
In view of all of these results, it is natural to ask whether the SU(N) Casson–Lin invariants introduced here can also be interpreted gauge theoretically. We plan to address this question in a future article using moduli spaces of projective SU(N) representations; see [Ruberman and Saveliev 2004]. We hope to use this approach to extend the Casson–Lin invariants $h_{N,a}(L)$ to links $L \subset \Sigma^3$ in homology 3-spheres and to describe corresponding Floer homology groups. In particular, we expect this approach will help clarify the relationship between the invariants $h_{N,a}(L)$ studied here and the SU(N) instanton Floer groups constructed by Kronheimer and Mrowka [2011b]. It is possible that this approach will also shed light on other interesting questions, such as whether and how the Casson–Lin invariants are related to classical link invariants, such as the higher linking numbers.

We give a brief outline of the contents of this paper. In Section 1, we introduce the notation for braids $\sigma \in B_k$, links $L \subset S^3$, and SU(N) representations that is used throughout the article. In Section 2, we introduce allowable labels $(a_1, \dots, a_n)$ for a given n-component link $L \subset S^3$, and a compatible k-tuple $(\varepsilon_1, \dots, \varepsilon_k)$ for a braid $\sigma \in B_k$ with closure L. We also introduce projective SU(N) representations of the link group G_L and establish irreducibility of elements of $\operatorname{Fix}(\varepsilon\sigma)$. In Section 3, we define the invariant $h_{N,a}(L)$ as an oriented intersection number and prove it is independent of the various choices involved. In Section 4, we calculate the invariants $h_{N,a}(L)$ for the Hopf link and the n-component chain link, and we prove a general vanishing result for the invariants for split links.

1. Braids and representations

In this section, we introduce the results for braids, links, and SU(N) representations that will be used throughout the article.

1A. ***The braid group.*** We denote by B_k the group of geometric braids on k strands with standard generators $\sigma_1, \dots, \sigma_{k-1}$ and relations $\sigma_i \sigma_j = \sigma_j \sigma_i$ for $|j - i| > 1$ and



σ_1 σ_2 σ_3

Figure 1. The three generators of B_4.

$\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}$. The generators are depicted in Figure 1. Note that in this paper we follow Convention 1.13 of [Kassel and Turaev 2008].

Each braid $\sigma \in B_k$ determines a permutation, and the resulting map $B_k \to S_k$, which sends the generator σ_i to the transposition $\overline{\sigma_i} = (i, i+1)$, is a surjection. Given $\sigma \in B_k$, we let $\bar{\sigma} \in S_k$ denote the corresponding permutation. Under this map, the symmetric group S_k acts on the set $\{1, \dots, k\}$ on the right. For $i \in \{1, \dots, k\}$, we write $(i)^{\bar{\sigma}}$ for the image of i under $\bar{\sigma} \in S_k$.

Let F_k be the free group with free generating set $x_1, \dots, x_k$. There is a natural right action of B_k on F_k defined by setting $\sigma_i : F_k \to F_k$ to be the map

$$\begin{cases} x_i \mapsto x_{i+1}, & \\ x_{i+1} \mapsto (x_{i+1})^{-1}\, x_i\, x_{i+1}, & \\ x_j \mapsto x_j, & j \neq i, i+1. \end{cases}$$

This action defines a faithful representation $\varrho : B_k \to \mathrm{Aut}(F_k)$, and we use it to identify B_k with its image in $\mathrm{Aut}(F_k)$ under ϱ. As this is a right action, we will use x_i^{σ} to denote the image of x_i under $\sigma \in B_k$.

Example 1.1. We explain how to read the action of a braid, which is explained in [Fenn et al. 1997, Section 2.4] for the left action, and we present the details for the right action.

The basic idea is to view the free group F_k as the fundamental group of a 2-disk with k punctures and keep track of basepoints as you move the disk vertically, letting the punctures move along the braid. Specifically, label the top strands $x_1, \dots, x_k$ from left to right. Then push the labels down, inserting a Wirtinger relation at each crossing. At the bottom of the braid the strands will be labeled by words $w_1, \dots, w_k$ in $x_1, \dots, x_k$, and the right action of σ is given by the automorphism sending x_i to $x_i^{\sigma} := w_i$.

Using Figure 2, we determine the actions of $\sigma_1\sigma_2$ and $\sigma_2\sigma_1$ on $F_3 = \langle x_1, x_2, x_3 \rangle$ to be given by

$$\begin{cases} x_1^{\sigma_1\sigma_2} = x_2, \\ x_2^{\sigma_1\sigma_2} = x_3, \\ x_3^{\sigma_1\sigma_2} = x_3^{-1}x_2^{-1}x_1x_2x_3, \end{cases} \qquad \text{and} \qquad \begin{cases} x_1^{\sigma_2\sigma_1} = x_3, \\ x_2^{\sigma_2\sigma_1} = x_3^{-1}x_1x_3, \\ x_3^{\sigma_2\sigma_1} = x_3^{-1}x_2x_3. \end{cases}$$

We point out two facts about the action of B_k on F_k, both of which are easily verified for each generator. Firstly, for any $\sigma \in B_k$, the permutation σ acts by

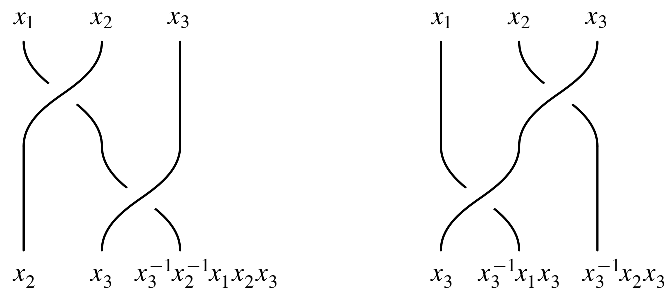


Figure 2. Reading the action of the braids $\sigma_1\sigma_2$ (left) and $\sigma_2\sigma_1$ (right).

conjugation and permutation on the generating set $x_1, \dots, x_k$ for F_k. Indeed,

$$x_i^\sigma = w\, x_{(i)^{\bar\sigma}}\, w^{-1}, \tag{1}$$

where $w \in F_k$ is some word depending on σ and i. Secondly, every braid $\sigma \in B_k$ preserves the product $x_1 \cdots x_k$,

$$(x_1 \cdots x_k)^\sigma = x_1 \cdots x_k. \tag{2}$$

1B. ***The group of a link.*** Every link L in S^3 can be realized as the closure $L = \hat\sigma$ of a braid σ. We regard L as an oriented link, where the strands of the braid σ are oriented in the downward direction. The link group $G_L = \pi_1(S^3 \smallsetminus L)$ admits a standard presentation

$$G_L = \pi_1(S^3 \smallsetminus \hat\sigma) = \langle x_1, \dots, x_k \mid x_i = x_i^\sigma, \text{ and } i = 1, \dots, k\rangle. \tag{3}$$

The number of components of the link $L = \hat\sigma$ is the number of disjoint cycles in the permutation $\bar\sigma$. We will be interested in n-component links, that is, the closures of braids σ with

$$\bar\sigma = (i_1, \dots, i_{k_1})(i_{k_1+1}, \dots, i_{k_2}) \cdots (i_{k_{n-1}+1}, \dots, i_{k_n}), \tag{4}$$

where $1 \le k_1 < k_2 < \cdots < k_n = k$. We define multi-indices $I_1, I_2, \dots, I_n$ by setting $I_j = \{i_{k_{j-1}+1}, \dots, i_{k_j}\}$ for $j = 1, \dots, n$, and we denote $\bar\sigma = (I_1) \cdots (I_n)$. If $L = \ell_1 \cup \cdots \cup \ell_n$ is the closure of a braid σ, we will assume that the cycles in the permutation $\bar\sigma = (I_1) \cdots (I_n)$ are written correspondingly, so that the component ℓ_j of L corresponds to the braid closure of the strands in I_j.

1C. ***The special unitary group.*** Consider the Lie group $\mathrm{SU}(N)$ of unitary $N \times N$ matrices with determinant one. Recall that $\mathrm{SU}(N)$ has real dimension $N^2 - 1$ and has center isomorphic to $\mathbb{Z}_N = \{\omega^d \mid d \in \mathbb{Z}\}$, where $\omega = e^{2\pi i/N}$. Notice that we are viewing $\mathbb{Z}_N$ as the subgroup of $U(1)$ consisting of N-th roots of unity, and for this reason we view it as a multiplicative group and identify it with the center of $\mathrm{SU}(N)$ via the map defined by sending $\omega^d \mapsto \omega^d I$.

Since every matrix in $\mathrm{SU}(N)$ is diagonalizable, conjugacy classes in $\mathrm{SU}(N)$ are completely determined by their eigenvalues when considered with multiplicities. Given $A \in \mathrm{SU}(N)$ we denote its conjugacy class by C_A. There is a unique conjugacy class C_A which is preserved under multiplication by $\omega = e^{2\pi i/N}$, and this is the conjugacy class of the diagonal matrix A whose eigenvalues are the N distinct N-th roots of $(-1)^{N-1}$. Setting $\xi = e^{2\pi i/2N}$, then A is given by the diagonal matrix $\mathrm{diag}(1, \omega, \dots, \omega^{N-1})$ when N is odd and by $\mathrm{diag}(\xi, \omega\xi, \dots, \omega^{N-1}\xi)$ when N is even. In either case, since the eigenvalues of A are all distinct, we see that the stabilizer of A is the standard maximal torus T^{N-1} in $\mathrm{SU}(N)$ and that $C_A \cong \mathrm{SU}(N)/T^{N-1}$ is the variety of full flags in $\mathbb{C}^N$ and has real dimension $N^2 - N$.

1D. SU(N) *representations.* For a discrete group G, let $\mathrm{R}(G) = \mathrm{Hom}(G, \mathrm{SU}(N))$ denote the variety of $\mathrm{SU}(N)$ representations of G. For convenience, we set $\mathrm{R}_k = \mathrm{R}(F_k) = \mathrm{SU}(N) \times \cdots \times \mathrm{SU}(N)$ to be the variety of $\mathrm{SU}(N)$ representations of the free group F_k. The faithful representation $\varrho : B_k \to \mathrm{Aut}(F_k)$ induces a representation

$$\tilde{\varrho} : B_k \to \mathrm{Diff}(\mathrm{R}_k) \tag{5}$$

given by $\tilde{\varrho}(\sigma)(\alpha) = \alpha \circ \sigma$. We will often abuse notation and simply denote $\tilde{\varrho}(\sigma)$ by σ.

Remark 1.2. [Long 1989, Theorem 2.1] implies that $\tilde{\varrho}$ is defined on $\mathrm{Aut}(F_k)$ and is faithful.

Example 1.3. Consider σ_1 and $\sigma_2 \in B_3$. For $X = (X_1, X_2, X_3) \in \mathrm{R}_3$, we have

$$\sigma_1(X) = (X_2,\ X_2^{-1}X_1X_2,\ X_3) \quad \text{and} \quad \sigma_2(X) = (X_1,\ X_3,\ X_3^{-1}X_2X_3).$$

Using this, one can easily compute that

$$\sigma_1^2(X) = (X_2^{-1}X_1X_2,\ X_2^{-1}X_1^{-1}X_2X_1X_2,\ X_3)$$

and further that

$$\sigma_1\sigma_2(X) = (X_2,\ X_3,\ X_3^{-1}X_2^{-1}X_1X_2X_3)$$

and

$$\sigma_2\sigma_1(X) = (X_3,\ X_3^{-1}X_1X_3,\ X_3^{-1}X_2X_3).$$

Using the standard presentation (3) of the link group, we notice that $\mathrm{R}(G_L)$ can be identified with $\mathrm{Fix}(\sigma) \subset \mathrm{R}_k$,

$$\mathrm{R}(G_L) = \{(X_1, \dots, X_k) \in \mathrm{R}_k \mid X_i = \sigma(X)_i\}.$$

A k-tuple $(X_1, \dots X_k) \in \mathrm{R}_k$ is called *reducible* if it can be simultaneously conjugated by an element of $\mathrm{SU}(N)$ such that each X_i has the form

$$X_i = \begin{pmatrix} A_i & 0 \\ 0 & B_i \end{pmatrix}, \tag{6}$$

where A_i is a block of size N_1 and B_i is a block of size N_2, and $N_1 + N_2 = N$. A k-tuple $(X_1, \dots X_k) \in \mathrm{R}_k$ is *irreducible* if it is not reducible.

1E. ***The wreath product*** $\mathbb{Z}_N \wr B_k$**.** The wreath product $\mathbb{Z}_N \wr B_k$ is the semidirect product of B_k with $(\mathbb{Z}_N)^k$, where B_k acts on $(\mathbb{Z}_N)^k$ by permutation. In other words, $\mathbb{Z}_N \wr B_k$ consists of pairs $(\varepsilon, \sigma) \in (\mathbb{Z}_N)^k \times B_k$, and the group structure is given by

$$(\varepsilon, \sigma) \cdot (\varepsilon', \sigma') = (\varepsilon\bar{\sigma}(\varepsilon'),\ \sigma\sigma').$$

Here, $\bar{\sigma}$ acts on $\varepsilon' = (\varepsilon'_1, \dots, \varepsilon'_k)$ by permutation, i.e., $\bar{\sigma}(\varepsilon') = (\varepsilon'_{(1)^{\bar{\sigma}}}, \dots, \varepsilon'_{(k)^{\bar{\sigma}}})$ In particular, it follows that $\sigma(\varepsilon X) = \bar{\sigma}(\varepsilon)\sigma(X)$.

We extend the representation (5) to the representation

$$\tilde{\varrho} : \mathbb{Z}_N \wr B_k \to \mathrm{Diff}(\mathrm{R}_k) \tag{7}$$

defined by sending the pair (ε, σ) to the diffeomorphism $\varepsilon\sigma : \mathrm{R}_k \to \mathrm{R}_k$, where

$$\varepsilon\sigma(X) = (\varepsilon_1\sigma(X)_1, \dots, \varepsilon_k\sigma(X)_k).$$

Thus ε twists the coordinates of $\sigma(X)$ by elements of the center $\mathbb{Z}_N$.

Example 1.4. For $X = (X_1, X_2, X_3) \in \mathrm{R}_3$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3) \in (\mathbb{Z}_N)^3$, we have

$$(\varepsilon\sigma_1)(X_1, X_2, X_3) = (\varepsilon_1 X_2,\ \varepsilon_2 X_2^{-1} X_1 X_2,\ \varepsilon_3 X_3)$$

and

$$\sigma_1(\varepsilon X) = \bar{\sigma}_1(\varepsilon)\sigma_1(X) = (\varepsilon_2 X_2,\ \varepsilon_1 X_2^{-1} X_1 X_2,\ \varepsilon_1 X_1,\ \varepsilon_3 X_3).$$

2. Projective representations of the link group

Our goal in this paper is to define invariants of L, and we will do so by performing a signed count of certain irreducible projective SU(N) representations.

2A. ***Projective representations.*** Suppose $\sigma \in B_k$ is a braid whose closure $\hat{\sigma}$ is a link L in S^3. For any k-tuple $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k) \in (\mathbb{Z}_N)^k$, an element $X = (X_1, \dots, X_k) \in \mathrm{SU}(N)^k$ in $\mathrm{Fix}(\varepsilon\sigma)$ determines a PU(N) representation of the link group G_L, i.e., a homomorphism $\tilde{\alpha} : G_L \to \mathrm{PU}(N)$. To see this, note that for any $X \in \mathrm{Fix}(\varepsilon\sigma)$, since $\varepsilon_i\sigma(X)_i = X_i$ holds in SU(N) and $\varepsilon_i \in \mathbb{Z}_N$ is central, the equation $\sigma(\tilde{X})_i = \tilde{X}_i$ holds for the k-tuple $\tilde{X} \in \mathrm{PU}(N)^k$, which shows that $\tilde{X}$ determines a representation $\tilde{\alpha} : G_L \to \mathrm{PU}(N)$.

Given a discrete group G, we define a *projective representation* of G to be a *function* (not a homomorphism!) $\alpha : G \to \mathrm{SU}(N)$ such that $\alpha(gh)\alpha(h)^{-1}\alpha(g)^{-1} \in \mathbb{Z}_N$ for all $g, h \in G$. For any projective representation $\alpha : G \to \mathrm{SU}(N)$, its composition with the surjection $\mathrm{Ad} : \mathrm{SU}(N) \to \mathrm{PU}(N)$ gives rise to a representation $\tilde{\alpha} = \mathrm{Ad}\,\alpha : G \to \mathrm{PU}(N)$, and thus every projective representation $\alpha : G \to \mathrm{SU}(N)$ is the lift of an honest representation $\tilde{\alpha} : G \to \mathrm{PU}(N)$. Alternatively, any representation

$\tilde{\alpha} : G \to \mathrm{PU}(N)$ can be lifted to a projective representation $\alpha : G \to \mathrm{SU}(N)$, though the lift is generally not unique.

Given a projective representation $\alpha : G \to \mathrm{SU}(N)$, we can associate a map $c : G \times G \to \mathbb{Z}_N$ defined by $c(g,h) = \alpha(gh)\alpha(h)^{-1}\alpha(g)^{-1}$. Notice that the map c satisfies the condition that $c(gh,k)c(g,h) = c(g,hk)c(h,k)$ for all $g,h,k \in G$, and hence c is a 2-cocycle of G.

For a fixed 2-cocycle $c : G \times G \to \mathbb{Z}_N$, let $\mathrm{PR}_c(G)$ denote the set of projective representations $\alpha : G \to \mathrm{SU}(N)$ whose associated 2-cocycle is c. If G is finitely generated with generating set $\{g_1, \dots, g_k\}$, then any projective representation $\alpha \in \mathrm{PR}_c(G)$ is completely determined by the 2-cocycle c and the elements $\alpha(g_1), \dots, \alpha(g_k) \in \mathrm{SU}(N)$, and in this way one can realize $\mathrm{PR}_c(G)$ as a subset of $\mathrm{SU}(N)^k$. It is a compact real algebraic variety.

2B. ***Allowable labels and compatible k-tuples.*** Given a link L in S^3 with n components, we can write $L = \ell_1 \cup \dots \cup \ell_n$. An n-tuple $a = (a_1, \dots, a_n) \in \mathbb{Z}^n$ of integers is called *allowable* if the following three conditions are satisfied:

(i) $0 \le a_i < N$ for $i = 1, \dots, n$,

(ii) $d = \gcd(a_1, \dots, a_n)$ is relatively prime to N,

(iii) $a_1 + \dots + a_n$ is a multiple of N.

An allowable n-tuple $(a_1, \dots, a_n)$ is called an n-tuple of *labels* for L, and a_j is the label corresponding to the j-th component ℓ_j of L.

Suppose now that L is the closure of a braid $\sigma \in B_k$, and write the permutation $\bar{\sigma}$ as a product $(I_1) \cdots (I_n)$ of disjoint cycles in such a way that I_j corresponds to the j-th component ℓ_j of L.

Recall that $\omega = e^{2\pi i/N}$. A k-tuple $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k) \in (\mathbb{Z}_N)^k$ for σ is said to be compatible with the choice of labels $(a_1, \dots, a_n)$ of L if it satisfies the conditions

$$\prod_{i \in I_j} \varepsilon_i = \omega^{a_j}, \tag{8}$$

for $j = 1, \dots, n$. This effectively labels each strand of the braid σ so that, upon closure of the braid, the j-th component ℓ_j of L is assigned the number a_j for its label. Note that with this choice $\varepsilon\sigma$ also preserves condition (2) since, by (8) and condition (iii), we have

$$\begin{aligned}(\varepsilon\sigma)(X)_1 \cdots (\varepsilon\sigma)(X)_k &= (\varepsilon_1 \cdots \varepsilon_k) X_1 \cdots X_k \\ &= (\omega^{a_1} \cdots \omega^{a_n}) X_1 \cdots X_k = X_1 \cdots X_k.\end{aligned} \tag{9}$$

2C. ***An obstruction to lifting.*** For $X \in \mathrm{Fix}(\varepsilon\sigma)$ we will show that the associated representation $\tilde{\alpha} : G_L \to \mathrm{PU}(N)$ does not lift to an $\mathrm{SU}(N)$ representation. Essential for this conclusion is that the k-tuple ε is compatible with a, the choice of labels for L. In particular, we use (8) and condition (ii) to give a nonzero obstruction to lifting $\tilde{\alpha}$ to an SU(N) representation.

Proposition 2.1. *The representation* $\tilde{\alpha} : G_L \to \mathrm{PU}(N)$ *does not lift to an* $\mathrm{SU}(N)$ *representation.*

Proof. Lift $\tilde{\alpha}$ arbitrarily to a map $\alpha : G_L \to \mathrm{SU}(N)$. Since α is a lift of $\tilde{\alpha}$, for each i we see that $\alpha(x_i) = \eta_i X_i$ for some $\eta_i \in \mathbb{Z}_N$. Let $\eta = (\eta_1, \dots, \eta_k) \in (\mathbb{Z}_N)^k$ be the corresponding k-tuple.

We assume that α is a representation. This implies that $\eta X \in \mathrm{Fix}(\sigma)$. Since X is also a fixed point of $\varepsilon\sigma$,

$$\eta_i X_i = \sigma(\eta_i X_i) = \eta_{(i)^{\bar{\sigma}}} \sigma(X)_i = (\varepsilon_i)^{-1} \eta_{(i)^{\bar{\sigma}}} X_i.$$

By condition (ii), some $a_j \neq 0$, and we assume without loss of generality that $a_1 \neq 0$. Consider the component ℓ_1 associated with the multi-index $I_1 = (i_1, \dots, i_{k_1})$; then (8) implies that

$$\eta_{i_1} = (\varepsilon_{i_1})^{-1} \eta_{i_2} = (\varepsilon_{i_1})^{-1} (\varepsilon_{i_2})^{-1} \eta_{i_3} = \cdots = (\varepsilon_{i_1})^{-1} \cdots (\varepsilon_{i_k})^{-1} \eta_{i_1} = \omega^{-a_1} \eta_{i_1},$$

which is a contradiction since $\omega^{-a_1} \neq 1$. □

2D. *Irreducibility for elements in* Fix($\varepsilon\sigma$). We now show that for any allowable n-tuple $a = (a_1, \dots, a_n) \in \mathbb{Z}^n$ of labels and compatible k-tuple $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k) \in (\mathbb{Z}_N)^k$, every $X \in \mathrm{Fix}(\varepsilon\sigma)$ is irreducible. The key to the proof is condition (ii) on the labels.

Proposition 2.2. *If* $X \in \mathrm{Fix}(\varepsilon\sigma)$, *then* X *is irreducible.*

Proof. Suppose to the contrary that $X \in \mathrm{Fix}(\varepsilon\sigma)$ is reducible, which means that up to conjugation, we can assume

$$X_i = \begin{pmatrix} A_i & 0 \\ 0 & B_i \end{pmatrix},$$

where A_i has size N_1 and B_i has size N_2.

The first step is to consider the component ℓ_1 of L. It is obtained by closing the strands of σ associated with the cycle $I_1 = (i_1, \dots, i_{k_1})$ of $\bar{\sigma}$. By (1), there are words $W_1, \dots, W_{k_1}$ in $X_1, \dots, X_k$ such that

$$\begin{aligned} X_{i_1} &= \varepsilon_{i_1} W_1 X_{i_2} W_1^{-1} \\ &= (\varepsilon_{i_1}\varepsilon_{i_2}) W_1 W_2 X_{i_3} W_2^{-1} W_1^{-1} \\ &\vdots \\ &= (\varepsilon_{i_1} \cdots \varepsilon_{i_{k_1}}) W_1 \cdots W_{k_1} X_{i_1} W_{k_1}^{-1} \cdots W_1^{-1} \\ &= \omega^{a_1} W X_{i_1} W^{-1}, \end{aligned}$$

where the last step follows by setting $W = W_1 \cdots W_{k_1}$ and applying (8).

Since W is a word in the X_i, and each X_i is block diagonal, it follows that W is also block diagonal so we can write

$$W = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}.$$

Applying this to the equation above, we see that the same relationship must hold for the blocks, so

$$A_{i_1} = \omega^{a_1} P A_{i_1} P^{-1}, \tag{10}$$

and taking the determinant of both sides of (10), we see that

$$\det(A_{i_1}) = \omega^{a_1 N_1} \det(A_{i_1}).$$

Since $\det(A_{i_1}) \neq 0$, this implies $\omega^{a_1 N_1} = 1$.

Now repeat the argument for the other components of the link L. For the component ℓ_j, which is the one obtained by closing the strands of σ associated with the cycle I_j, (8) implies that $\omega^{a_j N_1} = 1$, and we see this holds for each $j = 1, \ldots, n$. However, since $\omega = e^{2\pi i/N}$ is a primitive N-th root of unity, this can only happen if N divides $a_j N_1$ for each $j = 1, \ldots, n$. This contradicts condition (ii) on the labels, and we conclude that each $X \in \mathrm{Fix}(\varepsilon\sigma)$ is in fact irreducible. □

Remark 2.3. We would like to thank the referee for the following observation. Suppose $L = \ell_1 \cup \cdots \cup \ell_n$ is a link and let $\Lambda_i \cong \langle \mu_i \rangle \times \langle \lambda_i \rangle \cong \mathbb{Z} \times \mathbb{Z}$ denote the i-th peripheral subgroup of G_L, where μ_i and λ_i denote the meridian and longitude, respectively, of ℓ_i. Given a representation $\tilde{\alpha} : G_L \to \mathrm{PU}(N)$, let $\omega(\tilde{\alpha}) \in H^2(G_L, \mathbb{Z}_N)$ denote the obstruction cocycle, which is related to the commutator pairing of the restriction $\tilde{\alpha}|_{\Lambda_i}$ as follows. If $\alpha : G_L \to \mathrm{SU}(N)$ is a set-theoretic lift of $\tilde{\alpha}$, then the *commutator pairing* of Λ_i is the map $c_i : \Lambda_i \times \Lambda_i \to \mathbb{Z}_N$, given by $c_i(x, y) = [\alpha(x), \alpha(y)]$. Since Λ_i is free abelian of rank two,

$$\theta : H^2(\Lambda_i, \mathbb{Z}_N) \xrightarrow{\sim} \mathrm{Hom}(\Lambda_i \wedge \Lambda_i, \mathbb{Z}_N) \cong \mathbb{Z}_N,$$

and one can show that $\theta(\omega(\tilde{\alpha}|_{\Lambda_i})) = c_i$.

In the previous proof, the element $W = W_1 \ldots W_{k_1}$ is the image of the longitude λ_1 of ℓ_1, thus our computation that $[X_{i_1}, W] = \omega^{a_1}$ determined the commutator pairing $c_1 = \theta(\omega(\tilde{\alpha}|_{\Lambda_1}))$ by showing that $c_1(\mu_1, \lambda_1) = \omega^{a_1}$. The labels $a_1, \ldots, a_n$ thus determine the commutator pairings associated to the peripheral subgroups $\Lambda_1, \ldots, \Lambda_n$.

3. The link invariants

Throughout this section, we assume that σ is a braid with closure $\hat{\sigma} = L$, a link with n components $L = \ell_1 \cup \cdots \cup \ell_n$, and that $a = (a_1, \ldots, a_n)$ is an n-tuple of allowable labels, with a_j the label for the component ℓ_j.

In this section we define $h_{N,a}(\varepsilon\sigma)$ for compatible k-tuples ε, and we show that it gives rise to an invariant of n-component links in S^3.

We define $h_{N,a}(\varepsilon\sigma)$ as an algebraic count of certain projective $\mathrm{SU}(N)$ representations in $\mathrm{Fix}(\varepsilon\sigma)$, namely those that satisfy the monodromy condition $X_i \in C_A$. In other words, we require each X_i to be in the conjugacy class of matrices with characteristic polynomial $p_A(t) = t^N + (-1)^N$.

We will first show that $h_{N,a}(\varepsilon\sigma)$ is independent of choice of ε, and then we prove that $h_{N,a}(\varepsilon\sigma)$ gives rise to a well-defined invariant of the underlying link L by showing that it is invariant under the Markov moves.

3A. ***The definition of $h_{N,a}(\varepsilon\sigma)$.*** Recall that A is the diagonal matrix consisting of the N-th roots of $(-1)^{N-1}$, i.e.,

$$A = \begin{cases} \mathrm{diag}(1, \omega, \ldots, \omega^{N-1}) & \text{if } N \text{ is odd,} \\ \mathrm{diag}(\xi, \omega\xi, \ldots, \omega^{N-1}\xi) & \text{if } N \text{ is even.} \end{cases}$$

We impose the following monodromy condition and restrict to k-tuples lying in the subset $Q_k \subset \mathrm{R}_k$ given by

$$Q_k = \{(X_1, \ldots, X_k) \in \mathrm{R}_k \mid X_i \in C_A\}.$$

Since $Q_k = (C_A)^k$ is a just a k-fold product of C_A, we see that Q_k is a manifold of dimension $k(N^2 - N)$.

Let $\Delta_k = \{(X, X)\} \subset Q_k \times Q_k$ be the diagonal and $\Gamma_{\varepsilon\sigma} = \{(X, \varepsilon\sigma(X))\} \subset Q_k \times Q_k$ be the graph of $\varepsilon\sigma$. Notice that we can identify points in the intersection $\Delta_k \cap \Gamma_{\varepsilon\sigma}$ with elements in $Q_k \cap \mathrm{Fix}(\varepsilon\sigma)$.

For certain choices of labels, it will follow that $\mathrm{Fix}(\varepsilon\sigma) \subset Q_k$, i.e., that these monodromy conditions are automatically satisfied. This will occur whenever the labels have the property that each a_i is relatively prime to N. For a simple example, suppose N is prime, n is a positive multiple of N, and d is any positive integer less than N. Then one can easily verify that $a = (d, d, \ldots, d)$ is an allowable n-tuple of labels, and the next result implies that $\mathrm{Fix}(\varepsilon\sigma) \subset Q_k$ for any k-tuple $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_k)$ compatible with these labels.

Proposition 3.1. *Suppose $(a_1, \ldots, a_n)$ is an allowable n-tuple of labels such that each a_j is relatively prime to N, and suppose $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_k) \in (\mathbb{Z}_N)^k$ is compatible with a. Then* $\mathrm{Fix}(\varepsilon\sigma) \subset Q_k$, *i.e., if* $X \in \mathrm{Fix}(\varepsilon\sigma)$, *then each X_j is conjugate to A.*

Proof. The condition on a_j ensures that ω^{a_j} generates $\mathbb{Z}_N$ for each $j = 1, \ldots, n$, where $\omega = e^{2\pi i/N}$. Write the induced permutation $\bar{\sigma} = (I_1) \cdots (I_n)$ as a product of disjoint cycles, where I_j corresponds to the j-th component of $L = \hat{\sigma}$. Then for any $i \in I_j$, we can apply the same argument used to prove Proposition 2.2 to see that

$$X_i = \omega^{a_j} W X_i W^{-1}.$$

Thus, the set of eigenvalues of X_i is invariant under multiplication by ω^{a_j}, and this shows the eigenvalues of X are given by the set

$$\{\xi, \omega^{a_j}\xi, \ldots, \omega^{(N-1)a_j}\xi\} = \{\xi, \omega\xi, \ldots, \omega^{N-1}\xi\}$$

for some ξ satisfying $\xi^N = (-1)^{N-1}$. When N is odd, one can take $\xi = 1$, and when N is even, one can take $\xi = e^{2\pi i/2N}$. This shows that X_i is conjugate to A.

Alternative argument: consider the characteristic polynomial of both sides of the above equation; we see that

$$p_{X_i}(t) = p_{\omega^{a_j} X_i}(t) = p_{X_i}(\omega^{-a_j} t).$$

Since ω^{-a_j} has order N, $p_{X_i}(t)$ must be a polynomial in t^N, and indeed the only possibility is that $p_{X_i}(t) = t^N + (-1)^N$. $\square$

We define

$$H_k = \{(X, Y) \in Q_k \times Q_k \mid X_1 \cdots X_k = Y_1 \cdots Y_k\},$$

and we note that H_k is not a manifold because of the presence of reducibles. Recall that $(X, Y) \in Q_k \times Q_k$ is called *reducible* if all X_i and Y_i can be simultaneously conjugated into block diagonal form as in (6). We note that the subset $S_k \subset Q_k \times Q_k$ of reducibles is closed, and that $(Q_k \times Q_k)^* = (Q_k \times Q_k) \setminus S_k$ is an open manifold of dimension $2k(N^2 - N)$.

Proposition 3.2. *The subset $H_k^* = H_k \setminus S_k$ of irreducible representations is an open manifold of dimension $2k(N^2 - N) - (N^2 - 1)$.*

Proof. Clearly $H_k^* = f^{-1}(I)$, where $f : (Q_k \times Q_k)^* \to \mathrm{SU}(N)$ is the map defined by $f(X, Y) = X_1 \cdots X_k Y_k^{-1} \cdots Y_1^{-1}$. We will show that I is a regular value of f, i.e., that $df_{(X,Y)}$ is surjective for all $(X, Y) \in f^{-1}(I)$. It is enough to prove this statement for the map $f : Q_\ell^* \to \mathrm{SU}(N)$ given by $f(X_1, \ldots, X_\ell) = X_1 \cdots X_\ell$.

Clearly the matrix A, since it is diagonal, lies on the standard maximal torus $T^{N-1} \subset \mathrm{SU}(N)$ with Lie algebra

$$\mathfrak{t} = \left\{ \begin{pmatrix} ia_1 & & 0 \\ & \ddots & \\ 0 & & ia_N \end{pmatrix} \;\middle|\; a_1 + \cdots + a_N = 0 \right\}.$$

Since A has the standard maximal torus as its stabilizer group, we can identify the tangent space $T_A(C_A)$ with the orthogonal complement $\mathfrak{t}^\perp$ in $\mathfrak{su}(N)$, which is the subspace

$$\mathfrak{t}^{\perp} = \left\{ \begin{pmatrix} 0 & z_{12} & \cdots & z_{1,N} \\ -\bar{z}_{12} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & z_{N-1,N} \\ -\bar{z}_{1,N} & \cdots & -\bar{z}_{N-1,N} & 0 \end{pmatrix} \;\middle|\; z_{ij} \in \mathbb{C} \right\}.$$

There is a similar decomposition of $\mathfrak{su}(N)$ at each X_i in the ℓ-tuple $(X_1, \dots, X_\ell)$. Because each X_i has N distinct eigenvalues, it lies on a unique maximal torus $T_i \cong T^{N-1}$ in $\mathrm{SU}(N)$. We let $\mathfrak{t}_i \subset \mathfrak{su}(N)$ denote the corresponding Lie subalgebra, which is the Lie algebra of the stabilizer group of X_i. Using the decomposition $\mathfrak{su}(N) = \mathfrak{t}_i \oplus \mathfrak{t}_i^{\perp}$, we can identify the tangent space $T_{X_i}(C_A)$ with $\mathfrak{t}_i^{\perp} X_i$, the right translation of the subspace $\mathfrak{t}_i^{\perp} \subset \mathfrak{su}(N)$ by X_i. It is helpful to note that, in terms of the specific subspaces identified above, we have $\mathfrak{t}_i = \mathrm{Ad}_{P_i} \mathfrak{t}$ and $\mathfrak{t}_i^{\perp} = \mathrm{Ad}_{P_i} \mathfrak{t}^{\perp}$ for any matrix $P_i \in \mathrm{SU}(N)$ such that $X_i = P_i A P_i^{-1}$.

Using the fact that $\mathfrak{t}_i$ is the Lie algebra of the stabilizer subgroup of X_i, one can see that irreducibility of the ℓ-tuple $(X_1, \dots, X_\ell)$ is equivalent to the condition that $\mathfrak{t}_1 \cap \cdots \cap \mathfrak{t}_\ell = \{0\}$.

For $u_i \in \mathfrak{t}_i^{\perp}$, we set $x_i = u_i X_i \in \mathfrak{t}_i^{\perp} X_i = T_{X_i}(C_A)$. Differentiating and using the fact that $X_1 \cdots X_\ell = I$, we obtain

$$\begin{aligned} \frac{d}{dt}(X_1 + t x_1)(X_2 + t x_2) \cdots (X_\ell + t x_\ell)\Big|_{t=0} & \\ &= x_1 X_2 \cdots X_\ell + X_1 x_2 X_2 \cdots X_\ell + \cdots + X_1 \cdots X_{\ell-1} x_\ell \\ &= u_1 + X_2 u_2 X_2^{-1} + \cdots + (X_2 \cdots X_\ell) u_\ell (X_\ell^{-1} \cdots X_2^{-1}). \end{aligned}$$

In order to show that the map df_X is onto, we claim that, given any $v \in \mathfrak{su}(N)$, we can find $u_i \in \mathfrak{t}_i^{\perp}$ for $i = 1, \dots, \ell$ such that

$$v = u_1 + X_2 u_2 X_2^{-1} + \cdots + (X_2 \cdots X_\ell) u_\ell (X_\ell^{-1} \cdots X_2^{-1}). \tag{11}$$

Notice that we can solve (11) for any

$$v \in \mathfrak{t}_1^{\perp} \cap (X_1 \mathfrak{t}_2^{\perp} X_1^{-1}) \cap \cdots \cap (X_1 \cdots X_{\ell-1}) \mathfrak{t}_\ell^{\perp} (X_{\ell-1}^{-1} \cdots X_1^{-1}).$$

Notice further that since $\mathfrak{t}_i$ is the Lie algebra of the maximal torus containing X_i, we have $\mathfrak{t}_i \cap \mathfrak{t}_{i+1} = t_i \cap (X_i \mathfrak{t}_{i+1} X_i^{-1})$. More generally, for any subspace $V \subset \mathfrak{su}(N)$, we have $\mathfrak{t}_i \cap V = t_i \cap (X_i V X_i^{-1})$. Repeated application gives that

$$\begin{aligned} \mathfrak{t}_1 \cap \cdots \cap \mathfrak{t}_\ell &= \mathfrak{t}_1 \cap \cdots \cap (X_{\ell-1} \mathfrak{t}_\ell X_{\ell-1}^{-1}) \\ &= \mathfrak{t}_1 \cap \cdots \cap (X_{\ell-2} \mathfrak{t}_{\ell-1} X_{\ell-2}^{-1}) \cap (X_{\ell-2} X_{\ell-1} \mathfrak{t}_\ell X_{\ell-1}^{-1} X_{\ell-2}^{-1}) \\ &\;\;\vdots \\ &= \mathfrak{t}_1 \cap (X_1 \mathfrak{t}_2 X_1^{-1}) \cap \cdots \cap (X_1 \cdots X_{\ell-1}) \mathfrak{t}_\ell (X_{\ell-1}^{-1} \cdots X_1^{-1}). \end{aligned}$$

The condition of irreducibility implies that $\mathfrak{t}_1 \cap \cdots \cap \mathfrak{t}_\ell = \{0\}$, and it then follows from the above that (11) can be solved for any $v \in \mathfrak{su}(N)$. This concludes the argument that df_X is a surjection whenever the ℓ-tuple $X = (X_1, \ldots, X_\ell)$ is irreducible. $\square$

Since both Δ_k and $\Gamma_{\varepsilon\sigma}$ preserve the product $X_1 \cdots X_k$ (see (9)), we can restrict from $Q_k \times Q_k$ to H_k.

Now we are in a position to define the invariant $h_{N,a}(\varepsilon\sigma)$. Set $\Gamma^*_{\varepsilon\sigma} = \Gamma_{\varepsilon\sigma} \cap H^*_k$ and $\Delta^*_k = \Delta_k \cap H^*_k$. Since S_k is closed, it follows that both Γ^* and Δ^*_k are open submanifolds of H^*_k of dimension $k(N^2 - N)$.

Both Δ_k and $\Gamma_{\varepsilon\sigma}$ are compact, and so is their intersection $\Delta_k \cap \Gamma_{\varepsilon\sigma}$. Consequently, as Proposition 2.2 implies that every point in this intersection is irreducible, we have the following result.

Corollary 3.3. *The intersection $\Delta^*_k \cap \Gamma^*_{\varepsilon\sigma} \subset H^*_k$ is compact.*

The group $\mathrm{PU}(N)$ acts freely by conjugation on each of H^*_k, Δ^*_k, and $\Gamma^*_{\varepsilon\sigma}$, and the quotients by this action are the manifolds we denote as

$$\hat{H}_k = H^*_k / \mathrm{PU}(N), \quad \hat{\Delta}_k = \Delta^*_k / \mathrm{PU}(N), \quad \text{and} \quad \hat{\Gamma}_{\varepsilon\sigma} = \Gamma^*_{\varepsilon\sigma} / \mathrm{PU}(N).$$

Here the dimension of $\hat{H}_k$ equals $2k(N^2 - N) - 2(N^2 - 1)$, and both $\hat{\Delta}_k$ and $\hat{\Gamma}_{\varepsilon\sigma}$ are half-dimensional submanifolds of $\hat{H}_k$. Since the intersection $\hat{\Delta}_k \cap \hat{\Gamma}_{\varepsilon\sigma}$ is compact, we can deform $\hat{\Gamma}_{\varepsilon\sigma}$ into a submanifold $\tilde{\Gamma}_{\varepsilon\sigma}$ using an isotopy with compact support so that the intersection $\hat{\Delta}_k \cap \tilde{\Gamma}_{\varepsilon\sigma}$ is transverse and consists of finitely many points. Define

$$h_{N,a}(\varepsilon\sigma) = \#_{\hat{H}_k}(\hat{\Delta}_k \cap \tilde{\Gamma}_{\varepsilon\sigma})$$

as the oriented intersection number. We will describe the orientations in the following subsection. The intersection number $h_{N,a}(\varepsilon\sigma)$ is independent of the choice of isotopy of $\hat{\Gamma}_{\varepsilon\sigma}$, and we denote

$$h_{N,a}(\varepsilon\sigma) = \langle \hat{\Delta}_k, \hat{\Gamma}_{\varepsilon\sigma} \rangle_{\hat{H}_k}.$$

3B. *Orientations.* The following argument is similar to the one found in [Harper and Saveliev 2010, Section 3.4], and we include it here for completeness.

First, observe that the conjugacy class $C_A \subset \mathrm{SU}(N)$ is orientable, which follows for instance by identifying it with a flag variety. So choose an orientation for C_A and give $Q_k = (C_A)^k$ and $Q_k \times Q_k$ the induced product orientations. The diagonal Δ_k and the graph $\Gamma_{\varepsilon\sigma}$ are naturally diffeomorphic to Q_k via projection and so an orientation for Q_k determines orientations for both Δ_k and $\Gamma_{\varepsilon\sigma}$.

Using the standard orientation of $\mathrm{SU}(N)$, we obtain an orientation on $H^*_k = f^{-1}(I)$ using the base-fiber rule. Since the adjoint action of $\mathrm{PU}(N)$ on C_A is orientation preserving, the quotients $\hat{H}_k$, $\hat{\Delta}_k$, and $\hat{\Gamma}_{\varepsilon\sigma}$ are all orientable, and we orient them using the base-fiber rule.

Reversing the orientation of C_A reverses the orientation of Q_k only when k is odd, and in this case it reverses the orientations of both $\hat{\Delta}_k$, and $\hat{\Gamma}_{\varepsilon\sigma}$ but it does not affect the oriented intersection number $\langle \hat{\Delta}_k, \hat{\Gamma}_{\varepsilon\sigma}\rangle_{\hat{H}_k}$. This shows that the intersection number is actually independent of the choice of orientation on the conjugacy class C_A.

3C. ***Independence of $\boldsymbol{\varepsilon}$.*** The next result shows that $h_{N,a}(\varepsilon\sigma)$ is independent of the choice of ε compatible with a.

Proposition 3.4. *Fix a link L with $n > 1$ components and an allowable n-tuple $a = (a_1, \dots, a_n)$ of labels. Fix also a braid $\sigma \in B_k$ with closure $\hat{\sigma} = L$. If $\varepsilon, \varepsilon' \in (\mathbb{Z}_N)^k$ are k-tuples compatible with a, i.e., satisfying* (8), *then $h_{N,a}(\varepsilon\sigma) = h_{N,a}(\varepsilon'\sigma)$.*

Proof. We will define an orientation preserving automorphism $\varphi : \hat{H}_k \to \hat{H}_k$ such that $\varphi(\hat{\Delta}_k) = \hat{\Delta}_k$ and $\varphi(\hat{\Gamma}_{\varepsilon\sigma}) = \hat{\Gamma}_{\varepsilon'\sigma}$. Write the permutation

$$\bar{\sigma} = (i_1, \dots, i_{k_1})(i_{k_1+1}, \dots, i_{k_2}) \cdots (i_{k_{n-1}+1}, \dots, i_{k_n}) \tag{12}$$

as a product of disjoint cycles as in (4) and define $\delta = (\delta_1, \dots, \delta_k) \in (\mathbb{Z}_N)^k$ recursively with initial values

$$\delta_{i_1} = 1 = \delta_{i_{k_1+1}} = \cdots = \delta_{i_{k_{n-1}+1}} \tag{13}$$

and by setting

$$\delta_{(j)^{\bar{\sigma}}} = \delta_j\, \varepsilon_j\, (\varepsilon'_j)^{-1}. \tag{14}$$

Writing $\bar{\sigma}$ as a product of disjoint cycles as in (12) and noting that ε and ε' both satisfy (8), repeated application of the recursion (14) shows that the definition of $\delta = (\delta_1, \dots, \delta_k)$ is compatible with the initial values taken in (13).

Define the diffeomorphism $\tau : Q_k \to Q_k$ by

$$\tau(X) = \delta X = (\delta_1 X_1, \dots, \delta_k X_k).$$

Note that τ may be orientation preserving or reversing. Furthermore, τ preserves irreducibility and commutes with conjugation.

Consider the product map $\tau \times \tau : Q_k \times Q_k \to Q_k \times Q_k$. Observe that $\tau \times \tau$ preserves the orientation of $Q_k \times Q_k$ and hence the induced map $\varphi : \hat{H}_k \to \hat{H}_k$ is orientation preserving.

Since τ may be orientation reversing, φ restricted to $\hat{\Delta}_k$ or $\hat{\Gamma}_{\varepsilon\sigma}$ may be orientation reversing. The key observation is that if φ is orientation reversing on one, then it must be orientation reversing on the other. Hence, φ preserves the intersection number $h_{N,a}(\varepsilon\sigma)$,

$$\langle \hat{\Delta}_k, \hat{\Gamma}_{\varepsilon\sigma}\rangle_{\hat{H}_k} = \langle \varphi(\hat{\Delta}_k), \varphi(\hat{\Gamma}_{\varepsilon\sigma})\rangle_{\hat{H}_k}.$$

Clearly, $\varphi(\hat{\Delta}_k) = \hat{\Delta}_k$, so to finish off the proof we check that $\varphi(\hat{\Gamma}_{\varepsilon\sigma}) = \hat{\Gamma}_{\varepsilon'\sigma}$, or that the pair $(\delta X, \delta\varepsilon\sigma(X)) \in \hat{\Gamma}_{\varepsilon'\sigma}$. By the calculation

$$(\delta X, \delta\varepsilon\sigma(X)) = (\delta X, \delta\varepsilon\sigma(\delta^{N-1}\delta X)) = (\delta X, \delta\varepsilon\bar{\sigma}(\delta^{N-1})\sigma(\delta X)),$$

this will follow once we verify that $\delta\varepsilon\bar{\sigma}(\delta^{N-1}) = \varepsilon'$. Since $\delta_j^{N-1} = \delta_j^{-1}$, this is equivalent to showing that $\delta_j\varepsilon_j = \bar{\sigma}(\delta_j)\varepsilon'_j$ for $j = 1, \ldots, k$, which follows directly from (14), and this completes the proof of the proposition. □

3D. ***Independence under Markov moves.*** Based on the previous result, we denote $h_{N,a}(\varepsilon\sigma)$ by $h_{N,a}(\sigma)$ assuming that a choice of compatible ε has been made. In this subsection, we show that $h_{N,a}$ defines an invariant of n-component links, and this is achieved by showing that $h_{N,a}(\sigma)$ is invariant under the Markov moves.

Recall that two braids $\sigma, \tau \in B_k$ have isotopic closures $\hat{\sigma} = \hat{\tau}$ if and only if σ can be obtained from τ by a finite sequence of Markov moves; see for example [Birman 1974]. The first Markov move replaces $\sigma \in B_k$ by $\xi^{-1}\sigma\xi \in B_k$ for $\xi \in B_k$, and the second Markov move exchanges $\sigma \in B_k$ with $\sigma\sigma_k^{\pm 1} \in B_{k+1}$.

The following propositions give the SU(N) analogues of the SU(2) results in [Harper and Saveliev 2010, Propositions 4.2 and 4.3]; see the proof of [Lin 1992, Theorem 1.8].

Proposition 3.5. *The quantity $h_{N,a}(\sigma)$ is invariant under type 1 Markov moves.*

Proof. Suppose $\sigma \in B_k$ is a braid with

$$\bar{\sigma} = (I_1) \cdots (I_n) = (i_1, \ldots, i_{k_1}) \cdots (i_{k_{n-1}+1}, \ldots, i_{k_n})$$

in multi-index notation. Given a braid $\xi \in B_k$, let $\sigma' = \xi^{-1}\sigma\xi$ and note that $\overline{\sigma'}$ has the same cycle structure as $\bar{\sigma}$, in fact it is given by

$$\overline{\sigma'} = (I_1^{\xi}) \cdots (I_n^{\xi}) = ((i_1)^{\bar{\xi}}, \ldots, (i_{k_1})^{\bar{\xi}}) \cdots ((i_{k_{n-1}+1})^{\bar{\xi}}, \ldots, (i_{k_n})^{\bar{\xi}}).$$

We choose $\varepsilon' \in (\mathbb{Z}_N)^k$ compatible with the given labels, which means that ε' satisfies (8) with respect to the braid σ', namely

$$\prod_{i \in I_j^{\xi}} \varepsilon'_i = \omega^{a_j}$$

holds for $j = 1, \ldots, k$. Notice that if we define the k-tuple ε by setting $\varepsilon_i = \varepsilon'_{(i)^{\bar{\xi}}}$, then one can show that ε satisfies (8) with respect to $\bar{\sigma} = (I_1) \cdots (I_n)$; hence ε is also compatible with the given labels.

The braid ξ determines a map $\xi : Q_k \to Q_k$, and since it acts by permutation and conjugation on each of the factors in $Q_k = C_A \times \cdots \times C_A$, the fact that C_A is even-dimensional implies that this map is orientation preserving. This induces the map $\xi \times \xi$ on $Q_k \times Q_k$ preserving irreducibility, commuting with the adjoint action of PU(n), and preserving (2), thus we obtain a well-defined orientation preserving map $\xi \times \xi : \hat{H}_k \to \hat{H}_k$.

Clearly, $(\xi \times \xi)(\hat{\Delta}_k) = \hat{\Delta}_k$, so the diagonal is preserved, and we consider the effect of $\xi \times \xi$ on the graph $\hat{\Gamma}_{\varepsilon'\sigma'}$. If $(X, \varepsilon'\sigma'(X)) \in \hat{\Gamma}_{\varepsilon'\sigma'}$, then

$$(\xi \times \xi)(X, \varepsilon'\sigma'(X)) = (\xi \times \xi)(X, \varepsilon'\xi^{-1}\sigma\xi(X)) = (\xi(X), \xi(\varepsilon')\sigma\xi(X)) \in \hat{\Gamma}_{\varepsilon\sigma},$$

since $\xi(\varepsilon')_i = \varepsilon'_{(i)\xi} = \varepsilon_i$. Thus $(\xi \times \xi)(\hat{\Gamma}_{\varepsilon'\sigma'}) = \hat{\Gamma}_{\varepsilon\sigma}$, and we see that

$$\begin{aligned} h_{N,a}(\sigma') = \langle \hat{\Delta}_k, \hat{\Gamma}_{\varepsilon'\sigma'} \rangle_{\hat{H}_k} &= \langle (\xi \times \xi)(\hat{\Delta}_k), (\xi \times \xi)(\hat{\Gamma}_{\varepsilon'\sigma'}) \rangle_{\xi \times \xi(\hat{H}_k)} \\ &= \langle \hat{\Delta}_k, \hat{\Gamma}_{\varepsilon\sigma} \rangle_{\hat{H}_k} = h_{N,a}(\sigma). \end{aligned}$$

□

The next result is established using the same argument that is used to prove [Harper and Saveliev 2010, Proposition 4.3; Lin 1992, Theorem 1.8. We leave the details of the proof to the reader.

Proposition 3.6. *The quantity $h_{N,a}(\sigma)$ is invariant under type 2 Markov moves.*

4. Computations

In this section, we perform computations of $h_{N,a}(L)$ for various links L and we prove a vanishing condition for $h_{N,a}(L)$ for split links.

4A. ***The Hopf link and chain links.*** The chain link L is obtained as the closure of the braid $\sigma = \sigma_1^2\sigma_2^2 \cdots \sigma_{n-1}^2 \in B_n$. In this subsection, we compute $h_{N,a}(L)$ for L the Hopf link and the chain link with $N = n$ components. In particular, if d is chosen relatively prime to N and $a = (d, \ldots, d)$, then we will show that $h_{N,a}(L) = 0$ for the chain link with $n > 2$ components. For $n = 2$, L is just the Hopf link, which we denote by $H \subset S^3$. Harper and Saveliev [2010] proved that $h_{2,a}(H) = \pm 1$ for $a = (1, 1)$. We generalize this by showing that $h_{N,a}(H) = \pm 1$ if $a = (d, N - d)$, where d satisfies $1 \le d < N$ and is relatively prime to N.

The next result will be used repeatedly in the computations that follow.

Theorem 4.1. *Suppose $N \ge 2$ and set $\omega = e^{2\pi i/N}$ and $\xi = e^{2\pi i/2N}$. Any pair of matrices $(X, Y) \in \mathrm{SU}(N) \times \mathrm{SU}(N)$ satisfying $[X, Y] = \omega I$ is, up to conjugation, given by*

$$X = \begin{cases} \operatorname{diag}(1, \omega, \ldots, \omega^{N-1}) & \text{if } N \text{ is odd}, \\ \operatorname{diag}(\xi, \xi\omega, \ldots, \xi\omega^{N-1}) & \text{if } N \text{ is even}, \end{cases}$$

and

$$Y = \begin{pmatrix} 0 & \cdots & & \pm 1 \\ 1 & \ddots & & \vdots \\ & \ddots & & \\ 0 & \cdots & 1 & 0 \end{pmatrix}.$$

The pair $(X, Y) \in \mathrm{SU}(N) \times \mathrm{SU}(N)$ *determines an irreducible projective* $\mathrm{SU}(N)$ *representation of the free abelian group* $\mathbb{Z} \oplus \mathbb{Z}$ *of rank* 2.

Proof. First notice that $XYX^{-1}Y^{-1} = \omega I$ if and only if $Y^{-1}XY = \omega X$. Every element of $\mathrm{SU}(N)$ is conjugate to a diagonal matrix, and so we can write

$$X = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix},$$

where $\lambda_i \in U(1)$ and $\lambda_1 \cdots \lambda_N = 1$. However, because X is conjugate to ωX, we must have

$$\{\lambda_1, \ldots, \lambda_N\} = \{\omega\lambda_1, \ldots, \omega\lambda_N\}.$$

Reordering the terms, we can arrange that $\lambda_i = \omega^{i-1}\lambda_1$ for $i = 1, \ldots, N$. Since $\det X = 1$, we have $\lambda_1^N = (-1)^{N-1}$, and so without loss of generality we can take

$$\lambda_1 = \begin{cases} 1 & \text{if } N \text{ is odd,} \\ \xi & \text{if } N \text{ is even.} \end{cases}$$

This shows that X is of the required form.

Next, observe that $XYX^{-1}Y^{-1} = \omega I$ if and only if $XY = \omega YX$. Writing $Y = (y_{ij})$ and comparing the (ij) entries on right and left, it follows that

$$\omega^{i-1} y_{ij} = y_{ij}\omega^j.$$

This implies that $y_{ij} = 0$ unless $i \equiv j + 1 \mod N$. Furthermore, since Y has only one nonzero entry in each row and column, each entry must lie in $U(1)$ and we find that

$$Y = \begin{pmatrix} 0 & \cdots & & \mu_1 \\ \mu_2 & \ddots & & \vdots \\ & \ddots & & \\ 0 & & \mu_N & 0 \end{pmatrix},$$

where $\mu_i \in U(1)$ satisfy $\mu_1 \cdots \mu_N = (-1)^{N-1}$ (since $\det Y = 1$). Because X is diagonal with N distinct eigenvalues, the stabilizer subgroup $\mathrm{Stab}(X)$ is a copy of the standard maximal torus, i.e.,

$$\mathrm{Stab}(X) = \{\mathrm{diag}(\theta_1, \ldots, \theta_N) \mid \theta_i \in U(1), \theta_1 \cdots \theta_N = 1\} \cong T^{N-1}.$$

A matrix $P = \mathrm{diag}(\theta_1, \ldots, \theta_N) \in \mathrm{Stab}(X)$ acts on Y by

$$PYP^{-1} = \begin{pmatrix} 0 & \cdots & & \theta_1\theta_N^{-1}\mu_1 \\ \theta_2\theta_1^{-1}\mu_2 & \ddots & & \vdots \\ & \ddots & & \\ 0 & & \theta_N\theta_{N-1}^{-1}\mu_N & 0 \end{pmatrix}.$$

Setting

$$\begin{aligned}
\theta_1 &= \mu_1^{-1}\\
\theta_2 &= \theta_1\mu_2^{-1} = \mu_1^{-1}\mu_2^{-1}\\
\theta_3 &= \theta_2\mu_3^{-1} = \mu_1^{-1}\mu_2^{-1}\mu_3^{-1}\\
&\vdots\\
\theta_N &= \theta_{N-1}\mu_N^{-1} = \mu_1^{-1}\cdots\mu_N^{-1} = (-1)^{N-1},
\end{aligned}$$

it follows that

$$PYP^{-1} = \begin{pmatrix} 0 & \cdots & & (-1)^{N-1} \\ 1 & \ddots & & \vdots \\ & \ddots & & \\ 0 & & 1 & 0 \end{pmatrix}.$$

Since $PXP^{-1} = X$, this shows that, up to conjugation, Y is of the required form. Irreducibility of the pair (X, Y) follows from the fact that $\mathrm{Stab}(X)\cap\mathrm{Stab}(Y) = \mathbb{Z}_N$. $\square$

Remark 4.2. If $XYX^{-1} = \omega Y$, then $X^dYX^{-d} = \omega^d Y$ by induction. This shows that if (X, Y) are as in Theorem 4.1, then (X^d, Y) satisfies $[X^d, Y] = \omega^d I$. Using this observation, one can show that solutions (X', Y') to $[X', Y'] = \omega^d I$ are irreducible and unique up to conjugation provided d is relatively prime to N. This fails if d is not relatively prime to N; when $N = 4$ and $d = 2$, one can construct nonconjugate families of pairs $(X, Y) \in \mathrm{SU}(4)\times\mathrm{SU}(4)$ satisfying $[X, Y] = -I$. All of these pairs are reducible.

We now use this to evaluate $h_{N,a}(H)$ for the Hopf link H.

Proposition 4.3. *Suppose H is the Hopf link and $1 \le d < N$ is relatively prime to N. Then $h_{N,a}(H) = \pm 1$ for $a = (d, N-d)$.*

Proof. We motivate the proof with the following argument. The Hopf link H has link group $G_H = \langle x, y \mid [x, y] = 1\rangle$, and Theorem 4.1 implies there is a unique irreducible projective representation $\varrho : G_H \to \mathrm{SU}(N)$ with $[\varrho(x), \varrho(y)] = \omega^d I$. Uniqueness of ϱ up to conjugacy implies that $h_{N,a}(H) = \pm 1$.

More precisely, notice that the Hopf link is the closure of the braid $\sigma_1^2 \in B_2$ and fix the labels $a = (d, N-d)$ for H, where $1 \le d < N$ is relatively prime to N. The braid $\sigma = \sigma_1^2$ acts on pairs $(X_1, X_2) \in \mathrm{R}_2 = \mathrm{SU}(N)\times\mathrm{SU}(N)$ in the usual way (see Example 1.1), and for $\varepsilon = (\varepsilon_1, \varepsilon_2)$ we have

$$\varepsilon\sigma(X_1, X_2) = (\varepsilon_1 X_2^{-1}X_1X_2,\ \varepsilon_2 X_2^{-1}X_1^{-1}X_2X_1X_2).$$

For $(\varepsilon_1, \varepsilon_2) = (\omega^d, \omega^{N-d})$, one can easily see that $(X_1, X_2) \in \mathrm{Fix}(\varepsilon\sigma)$ if and only if $[X_1, X_2] = \omega^d$. By Theorem 4.1 and the preceding remarks, this equation has one solution which is irreducible and unique up to conjugation. Lemma 4.4 below shows that the solution is nondegenerate, and this implies that $h_{N,a}(H) = \pm 1$ for the Hopf link. □

The next result establishes the nondegeneracy result required for the above computation of $h_{N,a}(H)$.

Lemma 4.4. *Let H be the Hopf link, G_H its link group, and $1 \le d < N$ relatively prime to N. Suppose $\varrho : G_H \to \mathrm{SU}(N)$ is the projective representation, unique up to conjugation, of the link group G_H with $a = (d, N-d)$. Then $H^1(G_H; \mathfrak{su}(N)_{\mathrm{Ad}\,\varrho}) = 0$.*

Proof. Let $Z = S^3 \smallsetminus \tau H$ be the link exterior, and recall that the exterior of every nonsplit link in S^3 is a $K(\pi, 1)$. Thus $H^i(Z; \mathfrak{su}(N)_{\mathrm{Ad}\,\varrho}) = H^i(G_H; \mathfrak{su}(N)_{\mathrm{Ad}\,\varrho})$, where $G_H = \pi_1(Z)$ is the link group.

For the Hopf link, the link group $G_H = \mathbb{Z} \times \mathbb{Z}$ is the free abelian group of rank two. Since ϱ is irreducible, it follows that $H^0(G_H; \mathfrak{su}(N)_{\mathrm{Ad}\,\varrho}) = 0$, and Poincaré duality implies that $H^2(G_H; \mathfrak{su}(N)_{\mathrm{Ad}\,\varrho}) = 0$. Using $\chi(Z) = 0$, this shows that $H^1(G_H; \mathfrak{su}(N)_{\mathrm{Ad}\,\varrho}) = 0$, which completes the proof of the lemma. □

Next, we consider a chain link L and we establish the following vanishing result for $h_{N,a}(L)$.

Proposition 4.5. *Suppose L is a chain link with $n > 2$ components and that $n = N$. Then $h_{N,a}(L) = 0$ for $a = (d, \ldots, d)$, where d is relatively prime to N.*

Proof. We start with the chain link L with $n = 3$ components. It has link group with presentation

$$G_L = \langle x, y, z \mid [x, y] = 1 = [y, z] \rangle.$$

We will parametrize all triples $(X, Y, Z) \in \mathrm{SU}(3) \times \mathrm{SU}(3) \times \mathrm{SU}(3)$ satisfying $[X, Y] = \omega I = [Y, Z]$, and we will use this to show that $h_{3,a}(L) = 0$ for $a = (1, 1, 1)$.

Applying Theorem 4.1, up to conjugacy, there is a unique irreducible pair $(X, Y) \in \mathrm{SU}(3) \times \mathrm{SU}(3)$ satisfying the equation $[X, Y] = \omega I$. This pair is given by

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

In a general group, the commutator satisfies the relations

$$[x, y]^{-1} = [y, x] = y[x, y^{-1}]y^{-1} = x[x^{-1}, y]x^{-1}.$$

Setting $Z = X^{-1}$, this shows that $[Y, Z] = \omega I$, and thus the triple (X, Y, Z) gives rise to a projective representation $\varrho : G_L \to \mathrm{SU}(3)$ satisfying

$$[\varrho(x), \varrho(y)] = \omega I = [\varrho(y), \varrho(z)].$$

If $P \in \mathrm{Stab}(Y)$, then

$$[Y, PZP^{-1}] = P[Y, Z]P^{-1} = \omega I,$$

so the action of $\mathrm{Stab}(Y)$ on triples, given by $(X, Y, Z) \mapsto (X, Y, PZP^{-1})$, preserves the relations and is nontrivial on conjugacy classes. It follows that the solution set is 2-dimensional and parametrized by $\mathrm{Stab}(Y)/\mathbb{Z}_3$, which has Euler characteristic zero since $\mathrm{Stab}(Y) \cong T^2$ is a copy of a maximal torus. A calculation similar to the one in the proof of Lemma 4.4 shows that the solution set is a nondegenerate critical submanifold, and a standard argument then shows that its contribution to the invariant is given by plus or minus its Euler characteristic; see the proof of Proposition 8 in [Boden and Herald 1999]. It follows that $h_{3,a}(L) = 0$ for $a = (1, 1, 1)$, and a similar argument shows that $h_{3,a}(L) = 0$ for $a = (2, 2, 2)$.

One can also prove this via a direct approach making use of the fact that L is the closure of the braid $\sigma = \sigma_1^2\sigma_2^2$ and parametrizing the fixed point set $\mathrm{Fix}(\varepsilon\sigma)$ as was done for the Hopf link. We leave the details to the reader.

Next, consider the chain link L with 4 components. It has link group with presentation

$$G_L = \langle x, y, z, w \mid [x, y] = 1 = [y, z] = [z, w]\rangle.$$

By Theorem 4.1, up to conjugacy, there is a unique irreducible pair $(X, Y) \in \mathrm{SU}(4) \times \mathrm{SU}(4)$ satisfying the equation $[X, Y] = \omega I$. This pair is given by

$$X = \begin{pmatrix} \xi & 0 & 0 & 0 \\ 0 & \xi^3 & 0 & 0 \\ 0 & 0 & \xi^5 & 0 \\ 0 & 0 & 0 & \xi^7 \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Taking $Z = X^{-1}$ and $W = Y^{-1}$, one can show that the 4-tuple (X, Y, Z, W) gives rise to a projective representation $\varrho : G_L \to \mathrm{SU}(3)$ with $[\varrho(x), \varrho(y)] = \omega I = [\varrho(y), \varrho(z)] = [\varrho(z), \varrho(w)]$. The two groups $\mathrm{Stab}(Y)$ and $\mathrm{Stab}(Z)$ act on 4-tuples by

$$\begin{cases} (X, Y, Z, W) \mapsto (X, Y, PZP^{-1}, PWP^{-1}) & \text{for } P \in \mathrm{Stab}(Y), \\ (X, Y, Z, W) \mapsto (X, Y, Z, QWQ^{-1}) & \text{for } Q \in \mathrm{Stab}(Z), \end{cases}$$

and these actions preserve the relations and are nontrivial on conjugacy classes. It follows that the solution set is 6-dimensional and parametrized by $\mathrm{Stab}(Y)/\mathbb{Z}_4 \times \mathrm{Stab}(Z)/\mathbb{Z}_4$, which has Euler characteristic zero. By similar considerations as in the previous case, it follows that $h_{4,a}(L) = 0$ for $a = (1, 1, 1, 1)$, and a similar argument shows that $h_{4,a}(L) = 0$ for $a = (3, 3, 3, 3)$.

As before, one can perform these computations directly by noting that L is the closure of the braid $\sigma = \sigma_1^2\sigma_2^2\sigma_3^2$ and parametrizing the fixed point set $\mathrm{Fix}(\varepsilon\sigma)$.

This argument generalizes to the n-component chain link in a straightforward manner, as we now explain. The chain link L with n components has link group with presentation

$$G_L = \langle x_1, \dots, x_n \mid [x_1, x_2] = \cdots = [x_{n-1}, x_n] = 1 \rangle.$$

By Theorem 4.1, up to conjugacy, there is a unique irreducible pair $(X_1, X_2) \in \mathrm{SU}(N) \times \mathrm{SU}(N)$ satisfying the equation $[X_1, X_2] = \omega I$. A solution is obtained by taking $X_1 = X$ and $X_2 = Y$ for X, Y as in the statement of the theorem, and setting $X_{i+2} = X_i^{-1}$ for $i = 1, \dots, n-2$, the n-tuple

$$(X_1, \dots, X_n) \in \mathrm{SU}(N) \times \cdots \times \mathrm{SU}(N)$$

is easily seen to satisfy the relations

$$[X_1, X_2] = \cdots = [X_{n-1}, X_n] = \omega I.$$

For $i = 3, \dots, n$, the group $\mathrm{Stab}(X_i)$ acts on n-tuples by

$$(X_1, \dots, X_n) \mapsto (X_1, \dots, X_i, P X_{i+1} P^{-1}, \dots, P X_n P^{-1})$$

for $P \in \mathrm{Stab}(X_i)$. These actions preserve the relations and are nontrivial on conjugacy classes. Since each $\mathrm{Stab}(X_i) \cong T^{N-1}$ is a maximal torus, it follows that the solution set has dimension $(N-1)(N-2)$ and is parametrized by $\mathrm{Stab}(X_3)/\mathbb{Z}_N \times \cdots \times \mathrm{Stab}(X_n)/\mathbb{Z}_N$, which has Euler characteristic zero. It follows that $h_{N,a}(L) = 0$ for $a = (1, \dots, 1)$, and a similar argument shows that $h_{N,a}(L) = 0$ for $a = (d, \dots, d)$ for any d relatively prime to N. $\square$

4B. ***Split Links.*** In this section, we consider links $L \subset S^3$ that are geometrically split and prove a vanishing result for $h_{N,a}(L)$ provided that the labels satisfy the following condition. Assume L is a link with n components, and suppose it is split. Denoting the components of L by $\ell_1 \cup \cdots \cup \ell_n$, this means that $L = L_1 \cup L_2$, where up to reordering $L_1 = \ell_1 \cup \cdots \cup \ell_{n_1}$ and $L_2 = \ell_{n_1+1} \cup \cdots \cup \ell_n$ are sublinks that are separated by a 2-sphere. We shall assume that the labels $(a_1, \dots, a_n)$ satisfy the additional condition:

$$a_1 + a_2 + \cdots + a_{n_1} \text{ is not a multiple of } N. \tag{15}$$

Using Markov moves we can always find a *split braid* representative $\beta \in B_k$ of L; see Figure 3. This means that $\beta = \beta_1 \beta_2$ where

$$\beta_1 \in \mathrm{Im}(B_{k_1} \overset{i_1}{\hookrightarrow} B_k) \quad \text{and} \quad \beta_2 \in \mathrm{Im}(B_{k_2} \overset{i_2}{\hookrightarrow} B_k)$$

for $k = k_1 + k_2$ and i_1, i_2 are injective maps obtained by stabilizing on the right and left, respectively. More precisely, i_1 takes a braid in B_{k_1} and adds k_2 trivial strands on the right, and i_2 takes a braid in B_{k_2} and adds k_1 trivial strands on the left. Any

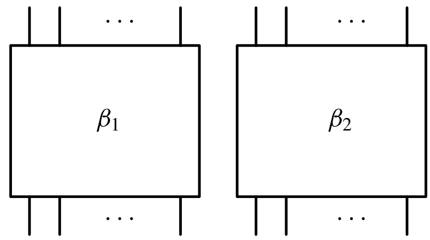


Figure 3. A split braid.

link L obtained as the closure $\hat{\beta}$ of a split braid is obviously a split link, and any split link L can be obtained as the closure of a split braid.

Proposition 4.6. *Suppose L is a split link and that β is a split braid with $\hat{\beta} = L$. Suppose further that $a = (a_1, \dots, a_n)$ is an n-tuple of labels satisfying* (15), *and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k)$ is a compatible k-tuple. Then the intersection $\Delta_k \cap \Gamma_{\varepsilon\beta} = \varnothing$, and consequently $h_{N,a}(L) = 0$.*

Proof. Let $X \in \Delta_k \cap \Gamma_{\varepsilon\beta}$, then by (8),

$$X_1 \cdots X_{k_1} = \omega^{dn_1} \beta(X)_1 \cdots \beta(X)_{k_1}.$$

Since $\beta = \beta_1\beta_2$ is a split braid with $\beta_1 \in B_{k_1}$, by (2) we have that

$$\beta(X)_1 \cdots \beta(X)_{k_1} = \beta_1(X)_1 \cdots \beta_1(X)_{k_1} = X_1 \cdots X_{k_1},$$

and this implies

$$X_1 \cdots X_{k_1} = \omega^{a_1+\cdots+a_{n_1}} X_1 \cdots X_{k_1}.$$

But $\omega^{a_1+\cdots+a_{n_1}} \neq 1$ by assumption (15), and this gives the desired contradiction. □

4C. ***Concluding remarks.*** One can give an alternative interpretation of the invariants $h_{N,a}(L)$ in terms of a signed count of conjugacy classes of representations $\varrho : G_L \to \mathrm{PU}(N)$ of the link group as follows. We begin by recalling the classification of principal $\mathrm{PU}(N)$ bundles from [Woodward 1982].

The classifying space $BPU(N)$ is simply connected and has $\pi_2(BPU(N)) = \mathbb{Z}_N$, and an application of the main theorem of [loc. cit.] implies that principal $\mathrm{PU}(N)$ bundles $P \to X$ over a 3-complex X are determined by the characteristic class $w(P) \in H^2(X; \mathbb{Z}_N)$. In case $N = 2$, $\mathrm{PU}(2) = SO(3)$ and $w(P)$ coincides with the second Stiefel-Whitney class.

Let $L \subset S^3$ be a link and $M_L = S^3 \smallsetminus \tau(L)$ its exterior. A projective $\mathrm{SU}(N)$ representation induces a representation $\varrho : G_L \to \mathrm{PU}(N)$, and we denote the associated cohomology class by $w(\varrho) \in H^2(G_L; \mathbb{Z}_N)$. The class $w(\varrho)$ vanishes if and only if ϱ lifts to an $\mathrm{SU}(N)$ representation. Further, there is a canonical injection $H^2(G_L; \mathbb{Z}_N) \to H^2(M_L; \mathbb{Z}_N)$ which is an isomorphism if and only if M_L is aspherical, i.e., if and only if L is nonsplit.

By reduction mod N, an allowable n-tuple $(a_1, \dots, a_n) \in \mathbb{Z}^n$ determines a unique cohomology class $\overline{w}(a_1, \dots, a_n) \in H^2(\partial M_L; \mathbb{Z}_N) \cong (\mathbb{Z}_N)^n$. The exact sequence in cohomology for the pair $(M_L, \partial M_L)$ gives

$$\to H^2(M_L; \mathbb{Z}_N) \xrightarrow{i^*} H^2(\partial M_L; \mathbb{Z}_N) \to H^3(M_L, \partial M_L; \mathbb{Z}_N) \to 0;$$

and condition (iii) from Section 2B guarantees that $\overline{w}(a_1, \dots, a_n)$ lies in the image of i^* and hence determines a class $w(a_1, \dots, a_n) \in H^2(M_L; \mathbb{Z}_N)$. Condition (ii) implies that the class $w(a_1, \dots, a_n)$ has order N.

From this point of view, the invariant $h_{N,a}(L)$ is closely related to the signed count of conjugacy classes of representations $\varrho : G_L \to \mathrm{PU}(N)$ such that $w(\varrho) = w(a_1, \dots, a_n)$. Proposition 4.6 is therefore a direct consequence of the fact that for split links L and allowable n-tuples $(a_1, \dots, a_n)$ satisfying condition (15), the associated cohomology class $w(a_1, \dots, a_n)$ does not lie in the image of the map $H^2(G_L; \mathbb{Z}_N) \to H^2(M_L; \mathbb{Z}_N)$.

As mentioned in the introduction, it would be interesting to investigate the relationship between the $\mathrm{SU}(N)$ Casson–Lin invariants studied here and the $\mathrm{SU}(N)$ instanton Floer groups constructed by Kronheimer and Mrowka [2011b; 2011a]. It would also be interesting to understand the relationship between the $\mathrm{SU}(N)$ Casson–Lin invariants and classical link invariants. For example, the main result of [Harper and Saveliev 2010] equates the SU(2) Casson–Lin invariant $h_2(L)$ of a two component link $L = \ell_1 \cup \ell_2$ with the linking number $\mathrm{lk}(\ell_1, \ell_2)$. The following conjecture, if true, would give a generalization to the higher rank setting.

Conjecture 4.7. *If $L = \ell_1 \cup \ell_2$ is a two component link in S^3, then the* $\mathrm{SU}(N)$ *Casson–Lin invariant satisfies*

$$h_{N,a}(L) = \mathrm{lk}(\ell_1, \ell_2)^{N-1}.$$

This conjecture is consistent with all known computations of the $\mathrm{SU}(N)$ Casson–Lin invariants, and it's possible that the invariants $h_{N,a}(L)$ are generally invariant under link homotopy. We hope to explore these topics in future work.

Acknowledgements

Both authors would like to thank Chris Herald, Andy Nicas, and Nikolai Saveliev for their valuable input. We would also like to thank an anonymous referee for many helpful suggestions. Boden expresses his gratitude to the Max Planck Institute for Mathematics in Bonn for its support.

References

[Alexander 1923] J. Alexander, "A lemma on systems of knotted curves", *Proc. Natl. Acad. Sci. USA* **9**:3 (1923), 93–95.

[Birman 1974] J. S. Birman, *Braids, links, and mapping class groups*, Princeton University Press, 1974. MR 0375281

[Boden and Herald 1999] H. U. Boden and C. M. Herald, "A connected sum formula for the SU(3) Casson invariant", *J. Differential Geom.* **53**:3 (1999), 443–464. MR 1806067

[Collin and Steer 1999] O. Collin and B. Steer, "Instanton Floer homology for knots via 3-orbifolds", *J. Differential Geom.* **51**:1 (1999), 149–202. MR 1703606

[Fenn et al. 1997] R. Fenn, R. Rimányi, and C. Rourke, "The braid-permutation group", *Topology* **36**:1 (1997), 123–135. MR 1410467

[Floer 1988] A. Floer, "An instanton-invariant for 3-manifolds", *Comm. Math. Phys.* **118**:2 (1988), 215–240. MR 956166

[Harper and Saveliev 2010] E. Harper and N. Saveliev, "A Casson–Lin type invariant for links", *Pacific J. Math.* **248**:1 (2010), 139–154. MR 2734168

[Harper and Saveliev 2012] E. Harper and N. Saveliev, "Instanton Floer homology for two-component links", *J. Knot Theory Ramifications* **21**:5 (2012), 1250054, 8. MR 2902278

[Herald 1997] C. M. Herald, "Flat connections, the Alexander invariant, and Casson's invariant", *Comm. Anal. Geom.* **5**:1 (1997), 93–120. MR 1456309

[Heusener and Kroll 1998] M. Heusener and J. Kroll, "Deforming abelian SU(2)-representations of knot groups", *Comment. Math. Helv.* **73**:3 (1998), 480–498. MR 1633375

[Kassel and Turaev 2008] C. Kassel and V. Turaev, *Braid groups*, Graduate Texts in Mathematics **247**, Springer, New York, 2008. MR 2435235

[Kronheimer and Mrowka 2011a] P. B. Kronheimer and T. S. Mrowka, "Khovanov homology is an unknot-detector", *Publ. Math. Inst. Hautes Études Sci.* **113** (2011), 97–208. MR 2805599

[Kronheimer and Mrowka 2011b] P. B. Kronheimer and T. S. Mrowka, "Knot homology groups from instantons", *J. Topol.* **4**:4 (2011), 835–918. MR 2860345

[Lin 1992] X.-S. Lin, "A knot invariant via representation spaces", *J. Differential Geom.* **35**:2 (1992), 337–357. MR 1158339

[Long 1989] D. D. Long, "On the linear representation of braid groups", *Trans. Amer. Math. Soc.* **311**:2 (1989), 535–560. MR 943606

[Ruberman and Saveliev 2004] D. Ruberman and N. Saveliev, "Rohlin's invariant and gauge theory, I: Homology 3-tori", *Comment. Math. Helv.* **79**:3 (2004), 618–646. MR 2081729

[Taubes 1990] C. H. Taubes, "Casson's invariant and gauge theory", *J. Differential Geom.* **31**:2 (1990), 547–599. MR 1037415

[Woodward 1982] L. M. Woodward, "The classification of principal PU_n-bundles over a 4-complex", *J. London Math. Soc.* (2) **25**:3 (1982), 513–524. MR 657508

Received June 11, 2015. Revised March 5, 2016.

HANS U. BODEN
DEPARTMENT OF MATHEMATICS AND STATISTICS
MCMASTER UNIVERSITY
1280 MAIN ST. W.
HAMILTON ON L8S 4K1
CANADA
boden@mcmaster.ca

ERIC HARPER
4704 CEDAR GLEN PLACE
CASTLE ROCK, CO 80109
UNITED STATES
ericharper228@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 285, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.285.283

THE SU(2) CASSON–LIN INVARIANT OF THE HOPF LINK

HANS U. BODEN AND CHRISTOPHER M. HERALD

We compute the SU(2) Casson–Lin invariant for the Hopf link and determine the sign in the formula of Harper and Saveliev relating this invariant to the linking number.

The Casson–Lin invariant $h(K)$ was defined for knots K by X.-S. Lin [1992] as a signed count of conjugacy classes of irreducible SU(2) representations of the knot group $G_K = \pi_1(S^3 \setminus K)$ with traceless meridional image, and Corollary 2.10 of the same paper shows that $h(K)$ is equal to $\frac{1}{2}\operatorname{sign}(K)$, one half the knot signature. E. Harper and N. Saveliev [2010] introduced the Casson–Lin invariant $h_2(L)$ of 2-component links, which they defined as a signed count of certain projective SU(2) representations of the link group $G_L = \pi_1(S^3 \setminus L)$. They showed that $h_2(L)$ equals the linking number of $L = \ell_1 \cup \ell_2$, up to an overall sign: $h_2(L) = \pm \operatorname{lk}(\ell_1, \ell_2)$. Harper and Saveliev [2012] also show that $h_2(L)$ can be regarded as an Euler characteristic associated to a certain SU(2) instanton Floer homology theory, defined by Kronheimer and Mrowka [2011].

The purpose of this note is to determine the sign in the formula of Harper and Saveliev, establishing the following.

Theorem 1. *If $L = \ell_1 \cup \ell_2$ is an oriented 2-component link in S^3, then its Casson–Lin invariant satisfies $h_2(L) = -\operatorname{lk}(\ell_1, \ell_2)$.*

We remark that the braid approach in [Harper and Saveliev 2010] is close in spirit to Lin's original definition, and it shows that $h_2(L)$ is an invariant of *oriented* links, because the Alexander and Markov theorems hold for oriented links; see Theorems 2.3 and 2.8 of [Kassel and Turaev 2008]. The sign of the invariant $h_2(L)$ depends not only on the choice of orientation on the braid, but also on the more subtle choice of identification of geometric braids with elements in the abstract braid group B_n, viewed as a subgroup of $\operatorname{Aut}(F_n)$. Here we follow Conventions 1.13 of [Kassel and Turaev 2008] in making this choice.

Note that extensions of the Casson–Lin invariants to SU(N) and to oriented links L in S^3 with at least two components are presented in [Boden and Harper

Boden was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.
MSC2010: primary 57M25; secondary 20C15.
Keywords: braids, links, representation spaces, Casson–Lin invariant.

2016], where, as before, they are defined by counting certain projective $\mathrm{SU}(N)$ representations of the link group G_L.

The rest of this paper is devoted to proving Theorem 1.

Proof. The proof of Proposition 5.7 in [Harper and Saveliev 2010] shows that the sign of $\mathrm{lk}(\ell_1, \ell_2)$ in our theorem is independent of L. (See also the proof of their Theorem 2 and their general discussion in Section 5.) Thus Theorem 1 will follow from a single computation.

To that end, we will determine the Casson–Lin invariant for the right-handed Hopf link. Since there is just one irreducible projective SU(2) representation of the link group, up to conjugation, it suffices to determine the sign associated to this one point.

We identify

$$\mathrm{SU}(2) = \left\{ x + yi + zj + wk \;\middle|\; |x|^2 + |y|^2 + |z|^2 + |w|^2 = 1 \right\}$$

with the group of unit quaternions and consider the conjugacy class

$$C_i = \left\{ yi + zj + wk \;\middle|\; |y|^2 + |z|^2 + |w|^2 = 1 \right\} \subset \mathrm{SU}(2)$$

of purely imaginary unit quaternions. Notice that C_i is diffeomorphic to S^2 and coincides with the set of SU(2) matrices of trace zero.

Let L be an oriented link in S^3, represented as the closure of an n-strand braid $\sigma \in B_n$. We follow Conventions 1.13 on page 17 of [Kassel and Turaev 2008] for writing geometric braids σ as words in the standard generators $\sigma_1, \dots, \sigma_{n-1}$. In particular, braids are oriented from top to bottom and σ_i denotes a right-handed crossing in which the $(i+1)$-st strand crosses over the i-th strand. The braid group B_n gives a faithful right action on the free group F_n on n generators, and here we follow the conventions in [Boden and Harper 2016] for associating an automorphism of F_n to a given braid $\sigma \in B_n$, which we write as $x_i \mapsto x_i^\sigma$ for $i = 1, \dots, n$. To be precise, to each braid group generator σ_i we associate the map $\sigma_i : F_n \to F_n$ given by

$$x_i \mapsto x_{i+1}, \quad x_{i+1} \mapsto (x_{i+1})^{-1} x_i x_{i+1}, \quad x_j \mapsto x_j \;\; (j \neq i,\; i+1),$$

and this is a right action, i.e., if $\sigma, \sigma' \in B_n$ are two braids, then $(x_i)^{\sigma\sigma'} = (x_i^\sigma)^{\sigma'}$ for all $1 \leq i \leq n$. Note that each braid $\sigma \in B_n$ fixes the product $x_1 \cdots x_n$.

A standard application of the Seifert–van Kampen theorem shows that the link complement $S^3 \setminus L$ has fundamental group

$$\pi_1(S^3 \setminus L) = \langle x_1, \dots, x_n \mid x_i^\sigma = x_i,\; i = 1, \dots, n \rangle.$$

We can therefore identify representations in $\mathrm{Hom}(\pi_1(S^3 \setminus L), \mathrm{SU}(2))$ with fixed points in $\mathrm{Hom}(F_n, \mathrm{SU}(2))$ under the induced action of the braid σ. We further identify $\mathrm{Hom}(F_n, \mathrm{SU}(2))$ with $\mathrm{SU}(2)^n$ by associating to a homomorphism ϱ the

n-tuple $(X_1, \ldots, X_n) = (\varrho(x_1), \ldots, \varrho(x_n))$. Note that $\sigma : \mathrm{SU}(2)^n \to \mathrm{SU}(2)^n$ is equivariant with respect to conjugation, so that fixed points come in whole orbits.

Every projective SU(2) representation can be identified with a fixed point in $\mathrm{Hom}(F_n, \mathrm{SU}(2))$ under the action of $\varepsilon\sigma$ for some n-tuple $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ with $\varepsilon_i = \pm 1$ such that $\varepsilon_1 \cdots \varepsilon_n = 1$. Notice that the action of $\varepsilon\sigma$ on $(X_1, \ldots, X_n) \in \mathrm{SU}(2)^n$ preserves the product $X_1 \cdots X_n$ and is equivariant with respect to conjugation. The Casson–Lin invariant $h_2(L)$ is then defined as a signed count of orbits of fixed points of $\varepsilon\sigma$ for a suitably chosen n-tuple $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$. The choice is made so that the resulting projective representations ϱ all have $w_2(\mathrm{Ad}\,\varrho) \neq 0$, meaning that the representations $\mathrm{Ad}\,\varrho$ do not lift to SU(2) representations. It has the consequence that for all fixed points ϱ of $\varepsilon\sigma$, each $\varrho(x_i)$ is a traceless SU(2) element.

We therefore restrict our attention to the subset of traceless representations, which are elements $\varrho \in \mathrm{Hom}(F_n, \mathrm{SU}(2))$ with $\varrho(x_j) \in C_i$ for $j = 1, \ldots, n$. Define $f : C_i^n \times C_i^n \to \mathrm{SU}(2)$ by setting

$$f(X_1, \ldots, X_n, Y_1, \ldots, Y_n) = (X_1 \cdots X_n)(Y_1 \cdots Y_n)^{-1}.$$

We obtain an orientation on $f^{-1}(1)$ by applying the base-fiber rule, using the product orientation on $C_i^n \times C_i^n$ and the standard orientation on the codomain of f. The quotient $f^{-1}(1)/\mathrm{conj}$ is then oriented by another application of the base-fiber rule, using the standard orientation on SU(2). This step uses the fact that, if $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ is chosen so that the associated SO(3) representation $\mathrm{Ad}\,\varrho$ has nontrivial second Stiefel–Whitney class $w_2 \neq 0$, then every fixed point of $\varepsilon\sigma$ in $\mathrm{Hom}(F_n, \mathrm{SU}(2))$ is necessarily irreducible.

We view conjugacy classes of fixed points of $\varepsilon\sigma$ as points in the intersection $\widehat{\Delta} \cap \widehat{\Gamma}_{\varepsilon\sigma}$, where $\widehat{\Delta} = \Delta/\mathrm{conj}$ is the quotient of the diagonal $\Delta \subset C_i^n \times C_i^n$, and where $\widehat{\Gamma}_{\varepsilon\sigma} = \Gamma_{\varepsilon\sigma}/\mathrm{conj}$ is the quotient of the graph $\Gamma_{\varepsilon\sigma}$ of $\varepsilon\sigma : C_i^n \to C_i^n$.

If the link L is the closure of a 2-strand braid, as it is for the Hopf link, then $\varepsilon = (-1, -1)$ is the only choice whose associated SO(3) bundle has $w_2 \neq 0$. Furthermore, in this case the intersection $\widehat{\Delta} \cap \widehat{\Gamma}_{\varepsilon\sigma}$ takes place in the pillowcase $f^{-1}(1)/\mathrm{conj}$, which is defined as the quotient

$$P = \{(a, b, c, d) \in C_i^4 \mid ab = cd\}/\mathrm{conj}. \tag{1}$$

It is well known that P is homeomorphic to S^2. To see this, first conjugate so that $a = i$, then conjugate by elements of the form $e^{i\theta}$ to arrange that b lies in the (i, j)-circle. A straightforward calculation using the equation $ab = cd$ shows that d must also lie on the (i, j)-circle. Clearly c is determined by a, b, d. We thus obtain an embedded 2-torus of elements of C_i^4 satisfying $ab = cd$, parametrized by

$$g(\theta_1, \theta_2) = (i, e^{k\theta_1} i, e^{k(\theta_2 - \theta_1)} i, e^{k\theta_2} i)$$

for $\theta_1, \theta_2 \in [0, 2\pi)$, which maps onto P. It is easy to verify that this is a two-to-one submersion, except when $\theta_1, \theta_2 \in \{0, \pi\}$. This realizes P as a quotient of the torus by the hyperelliptic involution. In particular, this involution is orientation-preserving, and away from the four singular points of P, we can lift all orientation questions up to the torus.

Let L be the right-handed Hopf link, which we view as the closure of the braid $\sigma = \sigma_1^2 \in B_2$, and suppose $\varepsilon = (-1, -1)$. The intersection $\widehat{\Delta} \cap \widehat{\Gamma}_{\varepsilon\sigma}$ consists of only one point, the conjugacy class of $g\left(\frac{\pi}{2}, \frac{\pi}{2}\right)$, that is, the point $[(i, j, i, j)] \in P$. (This is easily verified using the action of σ_1^2 on $F_2 = \langle x, y\rangle$; see Figure 1.) Thus, in order to pin down the sign of the Casson–Lin invariant $h_2(L)$, we must determine the orientations of $\widehat{\Delta}$, $\widehat{\Gamma}_{\varepsilon\sigma}$, and P at this point.

Notice that

$$\frac{\partial}{\partial\theta_1} g(\theta_1, \theta_2) = (0, e^{k\theta_1} j, -e^{k(\theta_2-\theta_1)} j, 0),$$
$$\frac{\partial}{\partial\theta_2} g(\theta_1, \theta_2) = (0, 0, e^{k(\theta_2-\theta_1)} j, e^{k\theta_2} j).$$

Evaluating at $\theta_1 = \theta_2 = \frac{\pi}{2}$ gives two tangent vectors $u_1 := (0, -i, -j, 0)$ and $u_2 := (0, 0, j, -i)$ to C_i^4 which span a complementary subspace in $\ker df$ to the orbit tangent space. Therefore, an ordering of these vectors determines an orientation on $P = f^{-1}(1)/\text{conj}$.

The orbit tangent space is spanned by the three tangent vectors

$$v_1 := \frac{\partial}{\partial t}\Big|_{t=0} e^{it}(i, j, i, j)e^{-it} = (0, 2k, 0, 2k),$$
$$v_2 := \frac{\partial}{\partial t}\Big|_{t=0} e^{jt}(i, j, i, j)e^{-jt} = (-2k, 0, -2k, 0),$$
$$v_3 := \frac{\partial}{\partial t}\Big|_{t=0} e^{kt}(i, j, i, j)e^{-kt} = (2j, -2i, 2j, -2i).$$

Then $\{u_1, u_2, v_1, v_2, v_3\}$ is a basis for $\ker(df|_{(i,j,i,j)}) = T_{(i,j,i,j)} f^{-1}(1)$. We choose vectors $w_1 = (k, 0, 0, 0)$, $w_2 = (0, k, 0, 0)$, $w_3 = (j, 0, 0, 0)$ to extend this to a basis for $T_{(i,j,i,j)} C_i^4$.

The orientation conventions in the definition of $h_2(L)$ (see Section 5d of [Harper and Saveliev 2010]) involve pulling back the orientation from $\mathfrak{su}(2) = T_1\,\mathrm{SU}(2)$ by df to obtain a coorientation for $\ker(df|_{(i,j,i,j)})$. With that in mind, we compute the action of df on $\{w_1, w_2, w_3\}$, namely, $df(w_1) = -j$, $df(w_2) = i$, $df(w_3) = k$.

Notice that the ordered triple $\{df(w_1), df(w_2), df(w_3)\} = \{-j, i, k\}$ gives the same orientation as the standard basis for $\mathfrak{su}(2)$. Thus, the base-fiber rule gives the coorientation $\{w_1, w_2, w_3\}$ on $\ker df$, so we choose the orientation $\mathcal{O}_{\ker df}$ on $\ker df$ such that $\mathcal{O}_{\{w_1,w_2,w_3\}} \oplus \mathcal{O}_{\ker df}$ agrees with the product orientation on $C_i^2 \times C_i^2$.

The orientation on the pillowcase P is then obtained by applying the base-fiber rule a second time to the quotient (1), using $\mathcal{O}_{\ker df}$ to orient $f^{-1}(1)$ and giving the

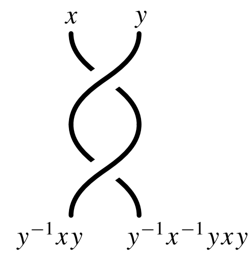


Figure 1. The action of $\sigma = \sigma_1^2$ on $F_2 = \langle x, y\rangle$.

orbit tangent space the orientation induced from that on SU(2) as well. We claim that the basis $\{u_1, u_2\}$ for the tangent space to the pillowcase has the opposite orientation. To see this, we note that $\{v_1, v_2, v_3\}$ is the fiber orientation for $\mathrm{SO}(3) \to f^{-1}(1) \to P$ and compare $S = \{w_1, w_2, w_3, u_1, u_2, v_1, v_2, v_3\}$ to the product orientation on $C_i^2 \times C_i^2$. Using the basis $\{(j, 0), (k, 0), (0, k), (0, i)\}$ for $T_{(i,j)}(C_i^2)$, we see that

$$\beta = \{(j, 0, 0, 0), (k, 0, 0, 0), (0, k, 0, 0), (0, i, 0, 0),\\ (0, 0, j, 0), (0, 0, k, 0), (0, 0, 0, k), (0, 0, 0, i)\}$$

is an oriented basis for $T_{(i,j,i,j)}C_i^4 = T_{(i,j)}C_i^2 \times T_{(i,j)}C_i^2$ with the product orientation.[1]

Let M be the matrix expressing the vectors in S in terms of the basis β. Since

$$M = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 2 \\ 1 & 0 & 0 & 0 & 0 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & -2 \end{bmatrix},$$

one easily computes that $\det M = -8$, confirming our claim that $\{u_2, u_1\}$ is a positively oriented basis for the pillowcase tangent space.

Recall that L is the right-handed Hopf link, which we represent as the closure of the braid $\sigma = \sigma_1^2 \in B_2$. For $\varepsilon = (-1, -1)$, as in Figure 1, one can verify that

$$\varepsilon\sigma(X, Y) = (-Y^{-1}XY, -Y^{-1}X^{-1}YXY).$$

Consider the curve $\alpha(\theta) = (i, e^{k\theta}i)$, passing through the point $(i, j) \in C_i^2$ when $\theta = \frac{\pi}{2}$, which is transverse to the orbit $[(i, j)]$. Then $(\alpha(\theta), \alpha(\theta))$ and $(\alpha(\theta), \varepsilon\sigma \circ \alpha(\theta))$ are curves in Δ and $\Gamma_{\varepsilon\sigma}$, respectively, and both are necessarily transverse to the orbit in

[1] As explained in Section 5d of [Harper and Saveliev 2010], the invariant $h_2(L)$ is independent of the choice of orientation on C_i. In fact, C_i^2 can be oriented arbitrarily provided one uses the *product* orientation on $C_i^2 \times C_i^2$.

C_i^4/conj. Thus, we can compare the orientations induced by the parametrizations $[(\alpha(\theta), \alpha(\theta))]$ and $[(\alpha(\theta), \varepsilon\sigma \circ \alpha(\theta))]$ of $\widehat{\Delta}$ and $\widehat{\Gamma}_{\varepsilon\sigma}$ to the pillowcase orientation determined above, namely $\{u_2, u_1\}$. The velocity vectors for the paths $(\alpha(\theta), \alpha(\theta)) = (i, e^{k\theta} i, i, e^{k\theta} i)$ and $(\alpha(\theta), \varepsilon\sigma \circ \alpha(\theta)) = (i, e^{k\theta} i, -e^{2k\theta} i, -e^{3k\theta} i)$ at $\theta = \frac{\pi}{2}$ are given by $(0, -i, 0, -i) = u_1 + u_2$ and $(0, -i, 2j, -3i) = u_1 + 3u_2$, respectively.

The Casson–Lin invariant is defined as the intersection number $h_2(L) = \langle \widehat{\Delta}, \widehat{\Gamma}_{\varepsilon\sigma} \rangle$, and in our case the sign of the unique intersection point in $\widehat{\Delta} \cap \widehat{\Gamma}_{\varepsilon\sigma}$ is determined by comparing the orientation of $\{u_1 + u_2, u_1 + 3u_2\}$ with $\{u_2, u_1\}$. Since the change of basis matrix $\left[\begin{smallmatrix} 1 & 3 \\ 1 & 1 \end{smallmatrix}\right]$ has negative determinant, it follows that $h_2(L) = -1$. $\square$

Acknowledgements

The authors thank Eric Harper and Nikolai Saveliev for many helpful discussions.

References

[Boden and Harper 2016] H. U. Boden and E. Harper, "The SU(N) Casson–Lin invariants for links", *Pacific J. Math.* **285**:2 (2016).

[Harper and Saveliev 2010] E. Harper and N. Saveliev, "A Casson–Lin type invariant for links", *Pacific J. Math.* **248**:1 (2010), 139–154. MR 2734168 Zbl 1206.57013

[Harper and Saveliev 2012] E. Harper and N. Saveliev, "Instanton Floer homology for two-component links", *J. Knot Theory Ramifications* **21**:5 (2012), 1250054, 8 pp. MR 2902278 Zbl 1237.57034

[Kassel and Turaev 2008] C. Kassel and V. Turaev, *Braid groups*, Graduate Texts in Mathematics **247**, Springer, 2008. MR 2435235 Zbl 1208.20041

[Kronheimer and Mrowka 2011] P. B. Kronheimer and T. S. Mrowka, "Knot homology groups from instantons", *J. Topol.* **4**:4 (2011), 835–918. MR 2860345 Zbl 1302.57064

[Lin 1992] X.-S. Lin, "A knot invariant via representation spaces", *J. Differential Geom.* **35**:2 (1992), 337–357. MR 1158339 Zbl 0774.57007

Received January 22, 2016. Revised April 20, 2016.

HANS U. BODEN
DEPARTMENT OF MATHEMATICS AND STATISTICS
MCMASTER UNIVERSITY
1280 MAIN ST. W.
ROOM 218
HAMILTON, ON L8S 4K1
CANADA
boden@mcmaster.ca

CHRISTOPHER M. HERALD
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NEVADA, RENO
RENO, NV 89557
UNITED STATES
herald@unr.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 285, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.285.289

COMMENSURATIONS AND METRIC PROPERTIES OF HOUGHTON'S GROUPS

JOSÉ BURILLO, SEAN CLEARY, ARMANDO MARTINO AND CLAAS E. RÖVER

We describe the automorphism groups and the abstract commensurators of Houghton's groups. Then we give sharp estimates for the word metric of these groups and deduce that the commensurators embed into the corresponding quasi-isometry groups. As a further consequence, we obtain that the Houghton group on two rays is at least quadratically distorted in those with three or more rays.

Introduction

The Houghton groups $\mathcal{H}_n$, introduced in [Houghton 1978], form an interesting family whose homological finiteness properties were described in [Brown 1987]. Röver [1999] showed that the $\mathcal{H}_n$ are all subgroups of Thompson's group V, and Lehnert [2009] described the metric for $\mathcal{H}_2$. Lee [2012] described isoperimetric bounds, and de Cornulier, Guyot and Pitsch [CGP 2007] showed that they are isolated points in the space of groups.

Here, we classify automorphisms and determine the abstract commensurator of $\mathcal{H}_n$. We also give sharp estimates for the word metric which are sufficient to show that the map from the abstract commensurator to the group of quasi-isometries of $\mathcal{H}_n$ is an injection.

1. Definitions and background

Let $\mathbb{N}$ be the set of natural numbers (positive integers) and $n \geq 1$ be an integer. We write $\mathbb{Z}_n$ for the integers modulo n with addition and put $R_n = \mathbb{Z}_n \times \mathbb{N}$. We interpret R_n as the graph of n pairwise disjoint rays; each vertex (i, k) is connected to $(i, k+1)$. We denote by Sym_n, FSym_n and FAlt_n, or simply Sym, FSym and FAlt if n is understood, the full symmetric group, the finitary symmetric group and the finitary alternating group on the set R_n, respectively.

The authors thank Charles Cox for his remarks on an earlier version of this paper and NUI Galway for its hospitality. This work was partially supported by MEC grant #MTM2011-25955 (Burillo), National Science Foundation grant #1417820 (Cleary) and Simons Foundation grant #234548 (Cleary).
MSC2010: 20F65.
Keywords: Houghton's groups, commensurations.

The *Houghton group* $\mathcal{H}_n$ is the subgroup of Sym consisting of those permutations that are eventually translations (of each of the rays). In other words, the permutation σ of the set R_n is in $\mathcal{H}_n$ if there exist integers $N \geq 0$ and $t_i = t_i(\sigma)$ for $i \in \mathbb{Z}_n$ such that for all $k \geq N$, we have $(i, k)\sigma = (i, k + t_i)$; throughout we will use right actions.

Note that necessarily the sum of the translations t_i must be zero because the permutation needs of course to be a bijection. This implies that $\mathcal{H}_1 \cong \mathsf{FSym}$.

For $i, j \in \mathbb{Z}_n$ with $i \neq j$ let $g_{ij} \in \mathcal{H}_n$ be the element which translates the line obtained by joining rays i and j, given by

$$\begin{aligned}
(i, n)g_{ij} &= (i, n-1) && \text{if } n > 1,\\
(i, 1)g_{ij} &= (j, 1),\\
(j, n)g_{ij} &= (j, n+1) && \text{if } n \geq 1, \text{ and}\\
(k, n)g_{ij} &= (k, n) && \text{if } k \notin \{i, j\}.
\end{aligned}$$

We also write g_i instead of $g_{i\,i+1}$. It is easy to see that $\{g_i \mid i \in \mathbb{Z}_n\}$, and $\{g_{ij} \mid i, j \in \mathbb{Z}_n,\ i \neq j\}$, are generating sets for $\mathcal{H}_n$ if $n \geq 3$ as we can simply check that the commutator $[g_0, g_1] = g_0^{-1} g_1^{-1} g_0 g_1$ transposes $(1, 1)$ and $(2, 1)$. In the special case of $\mathcal{H}_2$, the element g_1 is redundant as $g_1 = g_0^{-1}$. Further, an additional generator to g_0 is required to generate the group; we choose τ which fixes all points except for transposing $(0, 1)$ and $(1, 1)$.

It is now clear that the commutator subgroup of $\mathcal{H}_n$ is given by

$$\mathcal{H}_n' = \begin{cases} \mathsf{FAlt} & \text{if } n \leq 2,\\ \mathsf{FSym} & \text{if } n \geq 3. \end{cases}$$

For $n \geq 3$, we thus have a short exact sequence

$$1 \to \mathsf{FSym} \to \mathcal{H}_n \xrightarrow{\pi} \mathbb{Z}^{n-1} \to 1, \tag{1}$$

where $\pi(\sigma) = (t_0(\sigma), \ldots, t_{n-2}(\sigma))$ is the abelianization homomorphism. We note that as the sum of all the eventual translations must be zero, the last translation is determined by the preceding ones:

$$t_{n-1}(\sigma) = -\sum_{i=0}^{n-2} t_i(\sigma). \tag{2}$$

We will use the following facts freely throughout this paper; see [Dixon and Mortimer 1996] or [Scott 1964].

Lemma 1.1. *The group* FAlt *is simple and equal to the commutator subgroup of* FSym, *and* $\mathrm{Aut}(\mathsf{FAlt}) = \mathrm{Aut}(\mathsf{FSym}) = \mathsf{Sym}$.

2. Automorphisms of $\mathcal{H}_n$

Here we determine the automorphism group of $\mathcal{H}_n$. First we establish that we have to look no further than Sym. We let $N_G(H)$ denote the normalizer, in G, of the subgroup H of G.

Proposition 2.1. *Every automorphism of $\mathcal{H}_n$, $n \geq 1$, is given by conjugation by an element of* Sym; *that is to say*, $\mathrm{Aut}(\mathcal{H}_n) = N_{\mathsf{Sym}}(\mathcal{H}_n)$.

Proof. From the above, the finitary alternating group FAlt is the second derived subgroup of $\mathcal{H}_n$, and hence characteristic in $\mathcal{H}_n$. So every automorphism of $\mathcal{H}_n$ restricts to an automorphism of FAlt. Since Aut(FAlt) = Sym, this restriction yields a homomorphism $\mathrm{Aut}(\mathcal{H}_n) \to \mathsf{Sym}$ and we need to show that it is injective with image equal to $N_{\mathsf{Sym}}(\mathcal{H}_n)$.

In order to see this let $\psi \in \mathrm{Aut}(\mathcal{H}_n)$ be an automorphism. Compose this with an inner automorphism (of Sym) so that the result is an (injective) homomorphism $\alpha : \mathcal{H}_n \to \mathsf{Sym}$ whose restriction to FAlt is trivial. We let $k \in \mathbb{N}$ and consider the six consecutive points $a_\ell = (i, k + \ell)$ of R_n for $\ell \in \{0, 1, \ldots, 5\}$.

We denote by g_i^α the image of g_i under α, and by $(x\, y\, z)$ the 3-cycle of the points x, y and z. Using the identities

$$g_i^{-1}(a_1\, a_2\, a_3)g_i = (a_0\, a_1\, a_2) \quad \text{and} \quad g_i^{-1}(a_3\, a_4\, a_5)g_i = (a_2\, a_3\, a_4)$$

and applying α, which is trivial on FAlt, we get

$$(g_i^\alpha)^{-1}(a_1\, a_2\, a_3)g_i^\alpha = (a_0\, a_1\, a_2) \quad \text{and} \quad (g_i^\alpha)^{-1}(a_3\, a_4\, a_5)g_i^\alpha = (a_2\, a_3\, a_4).$$

Hence, g_i^α maps $\{a_1, a_2, a_3\}$ to $\{a_0, a_1, a_2\}$, and then also $\{a_3, a_4, a_5\}$ to $\{a_2, a_3, a_4\}$. The conclusion is that it maps a_3 to a_2. Applying a similar argument to all points in the branches i and $i + 1$, it follows that $g_i^\alpha = g_i$, and since i was arbitrary, this means that α is the identity map. □

With Lemma 1.1 in mind we now present the complete description of $\mathrm{Aut}(\mathcal{H}_n)$.

Theorem 2.2. *For $n \geq 2$, the automorphism group* $\mathrm{Aut}(\mathcal{H}_n)$ *of the Houghton group $\mathcal{H}_n$ is isomorphic to the semidirect product $\mathcal{H}_n \rtimes \mathcal{S}_n$, where $\mathcal{S}_n$ is the symmetric group that permutes the n rays.*

Proof. By the proposition, it suffices to prove that every $\alpha \in \mathsf{Sym}$ which normalizes $\mathcal{H}_n$ must map $(i, k + m)$ to $(j, l + m)$ for some $k, l \geq 1$ and all $m \geq 0$.

To this end, we pick $\alpha \in N_{\mathsf{Sym}}(\mathcal{H}_n)$ and $\sigma \in \mathcal{H}_n$. Since $\sigma^\alpha\ (= \alpha^{-1}\sigma\alpha)$ is in $\mathcal{H}_n$ and maps the point $x\alpha$ to $(x\sigma)\alpha$, these two points must lie on the same ray for all but finitely many $x \in R_n$. Similarly, x and $x\sigma$ lie on the same ray for all but finitely many $x \in R_n$, as $\sigma \in \mathcal{H}_n$. In fact, given a ray, we can choose σ so that whenever x lies on that ray, x and $x\sigma$ are successors on the same ray. Combining these facts, we see that α maps all but finitely many points of ray i to one and the same ray, say ray j.

This defines a homomorphism from $\mathrm{Aut}(\mathcal{H}_n)$ onto $\mathcal{S}_n$, which is obviously split, since given a permutation of the n rays, it clearly defines an automorphism of $\mathcal{H}_n$.

So assume that α is now in the kernel of that map, so it does not permute the rays, and take σ to be a g_{ji} generator of $\mathcal{H}_n$, i.e., an infinite cycle inside Sym. Then, since conjugating inside Sym preserves cycle types, the element $\sigma^\alpha \in \mathcal{H}_n$ is also a single infinite cycle. This means that σ^α has nonzero translations in only two rays, and these translations are 1 and -1. Any point in the support of σ is sent into the i-th ray by σ^k, for all sufficiently large k. Therefore, as α sends almost all points in the i-th ray into the i-th ray, the same is true for σ^α. Hence $t_i(\sigma^\alpha)$ is positive, so it must be $t_i(\sigma^\alpha) = 1$. It is quite clear now that α translates by an integer in the ray i, sufficiently far out. This finishes the proof since this could be done for any i, and hence $\alpha \in \mathcal{H}_n$. $\square$

3. Commensurations of $\mathcal{H}_n$

First, we recall that a commensuration of a group G is an isomorphism $A \xrightarrow{\phi} B$, where A and B are subgroups of finite index in G. Two commensurations ϕ and ψ of G are equivalent if there exists a subgroup A of finite index in G such that the restrictions of ϕ and ψ to A are equal. The set of all commensurations of G modulo this equivalence relation forms a group, known as the (abstract) *commensurator of* G, and is denoted $\mathrm{Com}(G)$. In this section we will determine the commensurator of $\mathcal{H}_n$.

For a moment, we let H be a subgroup of a group G. The *relative commensurator of* H *in* G is denoted $\mathrm{Com}_G(H)$ and consists of those $g \in G$ such that $H \cap H^g$ has finite index in both H and H^g. Similar to the homomorphism from $N_G(H)$ to $\mathrm{Aut}(H)$, there is a homomorphism from $\mathrm{Com}_G(H)$ to $\mathrm{Com}(H)$; its kernel consists of those elements of G that centralize a finite index subgroup of H.

To pin down $\mathrm{Com}(\mathcal{H}_n)$, we first establish that every commensuration of $\mathcal{H}_n$ can be seen as conjugation by an element of Sym, and then characterize $\mathrm{Com}_{\mathsf{Sym}}(\mathcal{H}_n)$.

Since a commensuration ϕ and its restriction to a subgroup of finite index in its domain are equivalent, we can restrict our attention to the following family of subgroups of finite index in $\mathcal{H}_n$, in order to exhibit $\mathrm{Com}(\mathcal{H}_n)$. For every integer $p \geq 1$, we define the subgroup U_p of $\mathcal{H}_n$ by

$$U_p = \langle \mathsf{FAlt}, g_i^p \mid i \in \mathbb{Z}_n \rangle.$$

We collect some useful properties of these subgroups first, where $A \subset_f B$ means that A is a subgroup of finite index in B.

Lemma 3.1. *Let* $n \geq 3$.

(i) *For every* p, *the group* U_p *coincides with* $\mathcal{H}_n^p$, *the subgroup generated by all* p*-th powers in* $\mathcal{H}_n$, *and hence is characteristic in* $\mathcal{H}_n$.

(ii) $U_p' = \begin{cases} \mathsf{FAlt}, & p \textit{ even}, \\ \mathsf{FSym}, & p \textit{ odd}; \end{cases}$ *and* $|\mathcal{H}_n : U_p| = \begin{cases} 2p^{n-1}, & p \textit{ even}, \\ p^{n-1}, & p \textit{ odd}. \end{cases}$

(iii) *For every finite index subgroup U of $\mathcal{H}_n$, there exists a $p \geq 1$ with* $\mathsf{FAlt} = U_p' \subset U_p \subset_f U \subset_f \mathcal{H}_n$.

The same is essentially true for the case $n = 2$, except that U_p' is always equal to FAlt in this case, with the appropriate change in the index.

Proof. First we establish (ii). We know $[g_i, g_j]$ is either trivial, when $j \notin \{i-1, i+1\}$, or an odd permutation. So the commutator identities $[ab, c] = [a, c]^b[b, c]$ and $[a, bc] = [a, c][a, b]^c$ imply the first part, and the second part follows immediately using the short exact sequence (1) from Section 1.

Part (i) is now an exercise, using that $\mathsf{FAlt}^p = \mathsf{FAlt}$.

To show (iii), let U be a subgroup of finite index in $\mathcal{H}_n$. The facts that FAlt is simple and U contains a normal finite index subgroup of $\mathcal{H}_n$ imply that $\mathsf{FAlt} \subset U$. Let p be the smallest even integer such that $(p\mathbb{Z})^{n-1}$ is contained in the image of U in the abelianization of $\mathcal{H}_n$. It is now clear that U_p is contained in U. □

Noting that $\mathrm{Com}(\mathcal{H}_1) = \mathrm{Aut}(\mathcal{H}_1) = \mathsf{Sym}$, we now characterize the commensurators of Houghton's groups.

Theorem 3.2. *Let $n \geq 2$. Every commensuration of $\mathcal{H}_n$ normalizes U_p for some even integer p. The group $N_p = N_{\mathsf{Sym}}(U_p)$ is isomorphic to $\mathcal{H}_{np} \rtimes (\mathcal{S}_p \wr \mathcal{S}_n)$, and* $\mathrm{Com}(\mathcal{H}_n)$ *is the direct limit of N_p with even p under the natural embeddings $N_p \to N_{pq}$ for $q \in \mathbb{N}$.*

Proof. Let $\phi \in \mathrm{Com}(\mathcal{H}_n)$. By Lemma 3.1, we can assume that U_p is contained in the domain of both ϕ and ϕ^{-1} for some even p. Let V be the image of U_p under ϕ. Then V has finite index in $\mathcal{H}_n$ and so contains FAlt, by Lemma 3.1. However, the set of elements of finite order in V equals $[V, V]$, whence $[V, V] = \mathsf{FAlt}$, as FAlt and FSym are not isomorphic. This means that the restriction of ϕ to FAlt is an automorphism of FAlt, and hence yields a homomorphism

$$\iota : \mathrm{Com}(\mathcal{H}_n) \to \mathsf{Sym}.$$

That ι is an injective homomorphism to $\mathrm{Com}_{\mathsf{Sym}}(\mathcal{H}_n)$ follows from an argument similar to the one in Proposition 2.1 applied to g_i^p and six points of the form $a_\ell = (i, k + p\ell)$ with $\ell \in \{0, 1, \ldots, 5\}$. Since the centralizer in Sym of FAlt, and hence of any finite index subgroup of $\mathcal{H}_n$, is trivial, the natural homomorphism from $\mathrm{Com}_{\mathsf{Sym}}(\mathcal{H}_n)$ to $\mathrm{Com}(\mathcal{H}_n)$ mentioned above is also injective, and we conclude that $\mathrm{Com}(\mathcal{H}_n)$ is isomorphic to $\mathrm{Com}_{\mathsf{Sym}}(\mathcal{H}_n)$, and that ι is the isomorphism:

$$\iota : \mathrm{Com}(\mathcal{H}_n) \to \mathrm{Com}_{\mathsf{Sym}}(\mathcal{H}_n).$$

From now on, we assume that $\phi \in \mathrm{Com}_{\mathsf{Sym}}(\mathcal{H}_n)$. In particular, the action of ϕ is given by conjugation, and our hypothesis is that $U_p^\phi \subset \mathcal{H}_n$. Now we can apply the argument from the proof of Theorem 2.2 to $\sigma \in U$ and σ^ϕ (instead of σ^α). Namely, consider the i-th ray, and choose a $\sigma = g_{ji}^p \in U_p$. So we have $t_i(\sigma) = p,\ t_j(\sigma) = -p$ and zero translation elsewhere. Now, except for finitely many points, σ^ϕ preserves the rays and sends $x\phi$ to $x\sigma\phi$. Thus there is an infinite subset of the i-th ray which is sent to the same ray by ϕ, say ray k. The infinite subset should be thought of as a union of congruence classes modulo p, except for finitely many points. We claim that no infinite subset of ray j can be mapped by ϕ to ray k. This is because if it were, the infinite subset would contain a congruence class modulo p (except for finitely many points) from which we would be able to produce two points x, y in the support of σ^ϕ such that all sufficiently large positive powers of σ^ϕ send x into ray k and all sufficiently large negative powers of σ^ϕ send y into ray k, and this is not possible for an element of $\mathcal{H}_n$. This means that ϕ maps an infinite subset of ray i onto almost all of ray k (observe that ray k is almost contained in the support of σ^ϕ so must be almost contained in the image of the union of rays i and j). Now applying similar arguments to ϕ^{-1} we get that ϕ is a bijection between rays i and k except for finitely many points. In fact, ϕ must induce bijections between the congruence classes modulo p (except for finitely many points) inside those two rays.

Thus ϕ induces a permutation of the ray system. Again, looking at large positive powers, we deduce that $t_k(\sigma^\phi) > 0$ and since the support of σ^ϕ is almost equal to the union of two rays, we must have $t_k(\sigma^\phi) = p$, as σ and hence σ^ϕ have exactly p orbits. In particular, we may deduce that ϕ normalizes U_p. So $U_p^\phi = U_p$.

In order to proceed, it will be useful to change the ray system. Specifically, each ray can be split into p rays, preserving the order. This realizes U_p as a (normal) subgroup of $\mathcal{H}_{np}$. We say two of these new rays are equivalent if they came from the same old ray. Thus there are n equivalence classes, each having p elements. The group U_p acts on this new ray system as the subgroup of $\mathcal{H}_{np}$ consisting of all $\sigma \in \mathcal{H}_{np}$ such that $t_i(\sigma) = t_j(\sigma)$ whenever the i-th and j-th rays are equivalent. Because we have split the rays, these translation amounts can be arbitrary in U_p (as a subgroup of $\mathcal{H}_{np}$) and not just multiples of p. In particular, we have $U_p = U_p^\phi \subset \mathcal{H}_{np}$ and the previous arguments imply that ϕ induces a bijection on the new ray system and sends equivalent rays to equivalent rays (since it is actually permuting the old ray system). Since ϕ permutes the rays, but must preserve equivalence classes, we get a homomorphism from $N_{\mathsf{Sym}}(U_p)$ to the subgroup of $\mathcal{S}_{np}$ which preserves the equivalence classes — this is easily seen to be $(\mathcal{S}_p \wr \mathcal{S}_n)$ and the above homomorphism is split.

As in Theorem 2.2 we now conclude that the kernel of this homomorphism is $\mathcal{H}_{np}$ and hence we get that $N_p = N_{\mathsf{Sym}}(U_p)$ is $\mathcal{H}_{np} \rtimes (\mathcal{S}_p \wr \mathcal{S}_n)$. □

We note that $\mathrm{Com}(\mathcal{H}_n)$ is not finitely generated, for if it were, it would lie in some maximal N_p.

4. Metric estimates for $\mathcal{H}_n$

In this section we will give sharp estimates for the word length of elements of Houghton's groups. This makes no sense for $\mathcal{H}_1$, which is not finitely generated. As mentioned in the Introduction, the metric in $\mathcal{H}_2$ was described by Lehnert [2009]. In order to deal with $\mathcal{H}_n$ for $n \geq 3$, we introduce the following measure of complexity of an element.

Given $\sigma \in \mathcal{H}_n$, we define $p_i(\sigma)$, for $i \in \mathbb{Z}_n$, to be the largest integer such that $(i, p_i(\sigma))\sigma \neq (i, p_i(\sigma) + t_i(\sigma))$. Note that if $t_i(\sigma) < 0$, then necessarily $p_i(\sigma) \geq |t_i(\sigma)|$, as the first element in each ray is numbered 1.

The *complexity* of $\sigma \in \mathcal{H}_n$ is the natural number $P(\sigma)$ defined by

$$P(\sigma) = \sum_{i \in \mathbb{Z}_n} p_i(\sigma).$$

And the *translation amount* of σ is

$$T(\sigma) = \frac{1}{2} \sum_{i \in \mathbb{Z}_n} |t_i(\sigma)|.$$

The above remark combined with (2) immediately implies $P(\sigma) \geq T(\sigma)$. It is easy to see that an element with complexity zero is trivial, and only the generators g_{ij} have complexity one.

Theorem 4.1. *Let $n \geq 3$ and $\sigma \in \mathcal{H}_n$, with complexity $P = P(\sigma) \geq 2$. Then the word length $|\sigma|$ of σ with respect to any finite generating set satisfies*

$$P/C \leq |\sigma| \leq KP \log P,$$

where the constants C and K only depend on the choice of generating set.

Proof. Since the word length with respect to two different finite generating sets differs only by a multiplicative constant, we can and will choose $\{g_{ij} \mid i, j \in \mathbb{Z}_n,\ i \neq j\}$ as the generating set to work with, and show that the statement holds with $C = 1$ and $K = 7$.

The lower bound is established by examining how multiplication by a generator can change the complexity. Suppose σ has complexity P and consider σg_{ij}. It is not difficult to see that

$$(3) \quad p_k(\sigma g_{ij}) = \begin{cases} p_k(\sigma) + 1 & \text{if } k = i \text{ and } (i, p_i(\sigma) + 1)\sigma = (i, 1), \\ p_k(\sigma) - 1 & \text{if } k = j,\ (j, p_j(\sigma) + 1)\sigma = (j, 1), \\ & \qquad\qquad \text{and } (j, p_j(\sigma))\sigma = (i, 1), \\ p_k(\sigma) & \text{otherwise,} \end{cases}$$

where the first two cases are mutually exclusive, as $i \neq j$. Thus $|P(\sigma g_{ij}) - P(\sigma)| \leq 1$, which establishes the lower bound.

The upper bound is obtained as follows. Suppose $\sigma \in \mathcal{H}_n$ has complexity P. First we show by induction on $T = T(\sigma)$ that there is a word ρ of length at most $T \le P$ such that the complexity of $\sigma\rho$ is $\overline{P}$ with $\overline{P} \le P$ and $T(\sigma\rho) = 0$. The case $T = 0$ is trivial. If $T > 0$, then there are $i, j \in \mathbb{Z}_n$ with $t_i(\sigma) > 0$ and $t_j(\sigma) < 0$. So $T(\sigma g_{ij}) = T - 1$. Moreover, $P(\sigma g_{ij}) \le P$, because the first case of (3) is excluded, as it implies that $t_i(\sigma) = -p_i(\sigma) \le 0$, contrary to our assumption. This completes the induction step.

We are now in the situation that $\sigma\rho \in \mathsf{FSym}$ and loosely speaking we proceed as follows.

1. We push all irregularities into ray 0, i.e., multiply by $\prod g_{i0}^{p_i(\sigma\rho)}$.
2. We push all points back into the ray to which they belong, except for those from ray 0 which we mix into ray 1, say.
3. We push out of ray 1 separating the points belonging to rays 0 and 1 into ray 0 and any other ray, say ray 2, respectively.
4. We push the points belonging to ray 1 back from ray 2 into it.

These four steps can be achieved by multiplying by an element μ of length at most $4\overline{P}$, such that $\sigma\rho\mu$ is an element which, for each i, permutes an initial segment I_i of ray i. Notice that $\sigma\rho\mu$ is now an element of $\mathcal{H}_n$ which maps each ray to itself, and hence $t_i(\sigma\rho\mu) = 0$ for all i.

It is clear that μ can be chosen so that the total length of the moved intervals $\sum |I_i|$ is at most $\overline{P}$. Finally, we sort each of these intervals using a recursive procedure, modeled on standard merge sort.

To sort the interval $I = I_2$ say, we push each of its points out of ray 2 and into either ray 0 if it belongs to the lower half, or ray 1 if it belongs to the upper half of I. If each of the two halves occurs in the correct order, then we only have to push them back into ray 2 and are done, having used $2|I|$ generators. If the two halves are not yet sorted, then we use the same "separate the upper and lower halves" approach on each of them recursively in order to sort them. In total this takes at most $2|I|\log_2|I|$ steps.

Altogether we have used at most

$$P + 4\overline{P} + 2\sum_{i\in\mathbb{Z}_n} |I_i|\log_2|I_i| \le 7P\log_2 P$$

generators to represent the inverse of σ; we used the hypothesis $P \ge 2$ in the last inequality. □

We note that because there are many permutations, the fraction of elements which are close to the lower bound goes to zero in much the same way as shown for Thompson's group V by Birget [2004] and its generalization nV by Burillo and Cleary [2010].

Lemma 4.2. *Let $n \geq 3$. For $\mathcal{H}_n$ take the generating set $g_1, \ldots, g_{n-1}$ with $n-1$ elements. Consider the following sets:*

- *B_k is the ball of radius k,*
- *C_k is the set of elements in $\mathcal{H}_n$ which have complexity $P = k$,*
- *$D_k \subset C_k$ is the set of elements of C_k which have word length at most $k \log_{2n-2} k$.*

Then

$$\lim_{k\to\infty} \frac{|D_k|}{|C_k|} = 0.$$

An element of complexity P, according to the metric estimates proved above, has word length between P and $P \log P$. What this lemma means is that most elements with complexity P will have word length closer to $P \log P$ than to P.

Proof. Observe that

$$\frac{|D_k|}{|B_{k \log_{2n-2} k}|} \leq 1$$

because it is a subset. Now, introduce the C_k as

$$\frac{|D_k|}{|C_k|} \frac{|C_k|}{|B_{k \log_{2n-2} k}|}$$

and the proof will be complete if we show that

$$\lim_{k\to\infty} \frac{|C_k|}{|B_{k \log_{2n-2} k}|} = \infty.$$

In C_k there are at least $(nk-2)!$ elements. This is because we can take a transposition of a point at distance k down one of the rays with another point. Since this already ensures $P = k$, we are free to choose any permutation of the other $nk-2$ points that are at one of the first k positions in each ray. And inside $|B_{k \log_{2n-2} k}|$, counting grossly according to the number of generators, there are at most $(2n-2)^{k \log_{2n-2} k} = k^k$ elements. Now the limit becomes

$$\lim_{k\to\infty} \frac{(nk-2)!}{k^k},$$

which is easily seen to approach infinity using Stirling's formula and the fact that $n \geq 3$. □

Consequentially, these estimates give an easy way to see that the group has exponential growth. We note that exponential growth also follows easily from the fact that g_{01} and g_{02} generate a free subsemigroup.

Proposition 4.3. *Let $n \geq 3$. Then $\mathcal{H}_n$ has exponential growth.*

Proof. Consider a finitary permutation of complexity P, and observe that there are at least $P!$ of those. By the metric estimate, its word length is at most $KP\log P$. Using again as in the previous lemma the notation B_k for a ball, the group will have exponential growth if

$$\lim_{k\to\infty}\frac{\log|B_k|}{k}>0.$$

In our case, this amounts to

$$\lim_{P\to\infty}\frac{\log|B_{KP\log P}|}{KP\log P}\geq\lim_{P\to\infty}\frac{\log(P!)}{KP\log P}=\frac{1}{K}.$$ □

5. Subgroup embeddings

We note that each $\mathcal{H}_n$ is a subgroup of $\mathcal{H}_m$ for $n<m$ and that our estimates together with work of Lehnert are enough to give at least quadratic distortion for some of these embeddings.

Theorem 5.1. *The group $\mathcal{H}_2$ is at least quadratically distorted in $\mathcal{H}_m$ for $m\geq 3$.*

Proof. We consider the element σ_n of $\mathcal{H}_2$ which has $T(\sigma_n)=0$ and transposes $(0,k)$ and $(1,k)$ for all $k\leq n$. Then σ_n corresponds to the word g_n defined in Theorem 8 of [Lehnert 2009], where it is shown to have length of the order of n^2 with respect to the generators of $\mathcal{H}_2$ in Lemma 10 there, which are exactly the generators for $\mathcal{H}_2$ given in Section 1. One can easily check that $\sigma_n=g_{02}^n g_{12}^n g_{02}^{-n} g_{12}^{-n}$ in $\mathcal{H}_3$. Thus a family of words of quadratically growing length in $\mathcal{H}_2$ has linearly growing length in $\mathcal{H}_3$, which proves the theorem. □

A natural, but seemingly difficult, question is whether $\mathcal{H}_n$ is distorted in $\mathcal{H}_m$ for $3\leq n<m$. Another question, which also seems difficult, is whether $\mathcal{H}_n$ is distorted in Thompson's group V, under the embeddings mentioned in the Introduction [Röver 1999].

6. Some quasi-isometries of $\mathcal{H}_n$

Commensurations give rise to quasi-isometries and are often a rich source of examples of quasi-isometries. Here we show that the natural map from the commensurator of $\mathcal{H}_n$ to the quasi-isometry group of $\mathcal{H}_n$, which we denote by $\mathrm{QI}(\mathcal{H}_n)$, is an injection. That is, we show that each commensuration is not within a bounded distance of the identity. That this is an injection also follows from the more general argument of Whyte which appears as Proposition 7.5 in [Farb and Mosher 2002].

Theorem 6.1. *The natural homomorphism from* $\mathrm{Com}(\mathcal{H}_n)$ *to* $\mathrm{QI}(\mathcal{H}_n)$ *is an embedding for $n\geq 2$.*

Proof. We will show that for each nontrivial $\phi \in \mathrm{Com}(\mathcal{H}_n)$ and every $N \in \mathbb{N}$ we can find a $\sigma \in \mathcal{H}_n$ such that $d(\sigma, \sigma^\phi) \geq N$, so none of the nontrivial images are within a bounded distance of the identity. By Theorem 3.2, we can and will view ϕ as a nontrivial element of $N_p \subset \mathsf{Sym}$ for some even p.

If ϕ eventually translates a ray i nontrivially to a possibly different ray j, then we let $\sigma = ((i, N)\,(i, N+1))$, a transposition in the translated ray. The image of σ under conjugation by ϕ is the transposition $((j, N+t)\,(j, N+t'+1))$, and the distance $d(\sigma, \sigma^\phi)$ is the length of $\sigma^{-1}\sigma^\phi$, which is at least N since it moves at least one point at distance N down one of the rays.

If ϕ does not eventually translate a ray but maps almost all of ray j to ray i with $i \neq j$, then we can show boundedness away from the identity by taking $\sigma = ((j, N)\,(j, N+1))$. The point (i, N) is fixed by σ but is moved to $(i, N+1)$ under σ^ϕ ensuring that the length of $\sigma^{-1}\sigma^\phi$ is at least N.

Finally, if ϕ does not have the preceding two properties, then ϕ is a nontrivial finitary permutation. Since Houghton's group is k-transitive, for every k, we can find a $\sigma \in \mathcal{H}_n$ such that ϕ^σ has support disjoint from that of ϕ, and at distance at least N down one of the rays. Hence $\sigma^{-1}\phi^{-1}\sigma\phi = \sigma^{-1}\sigma^\phi$ has length at least N. $\square$

7. Co-Hopficity

Houghton's groups have been long known to be Hopfian although they are not residually finite; see [CGP 2007]. In this section we will prove that $\mathcal{H}_n$ is not co-Hopfian, by exhibiting a map which is injective but not surjective:

$$f : \mathcal{H}_n \to \mathcal{H}_n,$$
$$s \mapsto f(s),$$

defined by: if $(i, n)s = (j, m)$, then

$$(i, 2n-1)f(s) = (j, 2m-1) \quad \text{and} \quad (i, 2n)f(s) = (j, 2m).$$

It is straightforward to show that f is a homomorphism. It is injective, because if s is not the identity with $(i, n)s \neq (i, n)$, then $(i, 2n)f(s) \neq (i, 2n)$. And clearly the map is not surjective, because the permutation always sends adjacent points $(i, 2n-1)$, $(i, 2n)$ to adjacent points, and a permutation which does not do this cannot be in the image.

In fact, $\mathcal{H}_n$ has many proper subgroups isomorphic to the whole group. The following argument was pointed out to us by Peter Kropholler.

One can well-order the ray system by taking a lexicographic order. The group $\mathcal{H}_n$ is then the group of all almost order-preserving bijections of the well-ordered ray system. It is then clear that the ray system minus a point is order isomorphic to the original ray system, which demonstrates that a point stabilizer is a subgroup isomorphic to $\mathcal{H}_n$.

Theorem 7.1. *Houghton's groups $\mathcal{H}_n$ are not co-Hopfian.*

References

[Birget 2004] J.-C. Birget, "The groups of Richard Thompson and complexity", *Internat. J. Algebra Comput.* **14**:5–6 (2004), 569–626. MR 2104771 Zbl 1088.20016

[Brown 1987] K. S. Brown, "Finiteness properties of groups", *J. Pure Appl. Algebra* **44**:1–3 (1987), 45–75. MR 885095 Zbl 0613.20033

[Burillo and Cleary 2010] J. Burillo and S. Cleary, "Metric properties of higher-dimensional Thompson's groups", *Pacific J. Math.* **248**:1 (2010), 49–62. MR 2734164 Zbl 1236.20045

[CGP 2007] Y. de Cornulier, L. Guyot, and W. Pitsch, "On the isolated points in the space of groups", *J. Algebra* **307**:1 (2007), 254–277. MR 2278053 Zbl 1132.20018

[Dixon and Mortimer 1996] J. D. Dixon and B. Mortimer, *Permutation groups*, Graduate Texts in Mathematics **163**, Springer, New York, 1996. MR 1409812 Zbl 0951.20001

[Farb and Mosher 2002] B. Farb and L. Mosher, "The geometry of surface-by-free groups", *Geom. Funct. Anal.* **12**:5 (2002), 915–963. MR 1937831 Zbl 1048.20026

[Houghton 1978] C. H. Houghton, "The first cohomology of a group with permutation module coefficients", *Arch. Math.* (*Basel*) **31**:3 (1978), 254–258. MR 521478 Zbl 0377.20044

[Lee 2012] S. R. Lee, "Geometry of Houghton's groups", preprint, 2012. arXiv 1212.0257

[Lehnert 2009] J. Lehnert, "Some remarks on depth of dead ends in groups", *Internat. J. Algebra Comput.* **19**:4 (2009), 585–594. MR 2536193 Zbl 1191.20027

[Röver 1999] C. E. Röver, *Subgroups of finitely presented simple groups*, Ph.D. thesis, University of Oxford, 1999, available at http://www.maths.nuigalway.ie/~chew/papers/thesis.shtml.

[Scott 1964] W. R. Scott, *Group theory*, Prentice-Hall, Englewood Cliffs, NJ, 1964. MR 0167513 Zbl 0126.04504

Received February 22, 2016.

JOSÉ BURILLO
DEPARTAMENT DE MATEMÀTICA APLICADA IV
EETAC-UPC
ESTEVE TERRADES 5
08860 CASTELLDEFELS
SPAIN
burillo@ma4.upc.edu

SEAN CLEARY
DEPARTMENT OF MATHEMATICS R8133
THE CITY COLLEGE OF NEW YORK
160 CONVENT AVENUE
NEW YORK, NY 10031
UNITED STATES
scleary@ccny.cuny.edu

ARMANDO MARTINO
MATHEMATICAL SCIENCES
UNIVERSITY OF SOUTHAMPTON
UNIVERSITY ROAD
SOUTHAMPTON SO17 1BJ
UNITED KINGDOM
a.martino@soton.ac.uk

CLAAS E. RÖVER
SCHOOL OF MATHEMATICS, STATISTICS AND APPLIED MATHEMATICS
NUI GALWAY
UNIVERSITY ROAD
GALWAY
IRELAND
claas.roever@nuigalway.ie

PACIFIC JOURNAL OF MATHEMATICS
Vol. 285, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.285.303

CONFORMAL HOLONOMY EQUALS AMBIENT HOLONOMY

ANDREAS ČAP, A. ROD GOVER,
C. ROBIN GRAHAM AND MATTHIAS HAMMERL

We study the relation between two notions of holonomy on a conformal manifold. The first is the conformal holonomy, defined to be the holonomy of the normal tractor connection. The second is the holonomy of the Fefferman–Graham ambient metric of the conformal manifold. It is shown that the infinitesimal conformal holonomy and the infinitesimal ambient holonomy always agree up to the order that the ambient metric is defined.

1. Introduction

The tractor bundle $\mathcal{T}$ of a smooth conformal manifold (M, c) of dimension $n \geq 3$ and signature (p, q), $p+q=n$, is a rank-$(n+2)$ vector bundle naturally associated to the conformal structure, which carries a canonical connection ∇; see [Bailey et al. 1994]. This connection is characterized by a normalization condition on its curvature, whence it is called the normal tractor connection [Čap and Gover 2003]. It can be viewed as a conformally invariant analog of the Levi-Civita connection of a Riemannian manifold and has played an essential role in many recent developments in conformal geometry. The holonomy of $(\mathcal{T}, \nabla)$ is called the conformal holonomy of (M, c). Following early work [Armstrong 2007; Leistner 2006; Leitner 2005], its study has been the focus of active recent research; see, e.g., [Alt 2012; Armstrong and Leitner 2012; Lischewski 2015].

Another invariant object associated to a conformal manifold is the ambient metric of [Fefferman and Graham 1985; 2012]. This is a smooth pseudo-Riemannian metric of signature $(p+1, q+1)$ on a space of dimension $n+2$, determined up to diffeomorphism along a canonical hypersurface, to infinite order if n is odd, and to order $\frac{n}{2}-1$ if n is even. Its Levi-Civita connection is another connection associated to the conformal manifold and one can also consider its holonomy. Because the

Čap and Gover gratefully acknowledge support from the Royal Society of New Zealand via Marsden Grant 13-UOA-018. Further, support by projects P23244-N13 (Čap and Hammerl) and P27072-N25 (Čap) of the Austrian Science Fund (FWF) is gratefully acknowledged. Research of Graham is partially supported by NSF grant # DMS 1308266.
MSC2010: primary 53A30; secondary 53C29.
Keywords: holonomy, ambient metric, tractor, conformal geometry.

holonomy group of a connection is a global invariant and the ambient metric is only invariantly defined as a jet along a hypersurface, its holonomy group is not the appropriate object to study. Instead we consider the infinitesimal holonomy, which depends only on the jet at a point. The main result of this paper asserts that, suitably interpreted, the infinitesimal holonomies of the tractor connection and the Levi-Civita connection of the ambient metric agree at each point.

In order to formulate the result precisely, we describe a realization of the tractor bundle in ambient terms which was derived in [Čap and Gover 2003]. Details will be provided in Section 2. If (M, c) is a conformal manifold, its metric bundle is the ray bundle $\mathcal{G} \subset S^2T^*M$ whose sections are the metrics $g \in c$. The ambient space is $\mathcal{G} \times \mathbb{R}$, in which $\mathcal{G}$ is embedded as the hypersurface $\mathcal{G} \times \{0\}$. There are dilations $\delta_s : \mathcal{G} \to \mathcal{G}$ given by $\delta_s(x, g_x) = (x, s^2 g_x)$, $s > 0$, which extend to $\mathcal{G} \times \mathbb{R}$ acting in the first factor. For $x \in M$, we denote by $\mathcal{G}_x$ the fiber of $\mathcal{G}$ over x, and we view $\mathcal{G}_x$ as a 1-dimensional submanifold of $\mathcal{G} \times \mathbb{R}$ via $\mathcal{G}_x \subset \mathcal{G} = \mathcal{G} \times \{0\} \subset \mathcal{G} \times \mathbb{R}$. Then $T(\mathcal{G} \times \mathbb{R})|_{\mathcal{G}_x}$ denotes the tangent bundle to $\mathcal{G} \times \mathbb{R}$ restricted to the submanifold $\mathcal{G}_x$, a rank-$(n+2)$ vector bundle over $\mathcal{G}_x$. The standard tractor bundle of (M, c) can be realized as the rank-$(n+2)$ vector bundle $\mathcal{T} \to M$ with fiber

$$\mathcal{T}_x = \big\{U \in \Gamma(T(\mathcal{G} \times \mathbb{R})|_{\mathcal{G}_x}) : (\delta_s)^* U = s^{-1} U,\ s > 0\big\}. \tag{1-1}$$

The right-hand side of (1-1) is clearly a vector space of dimension $n+2$ varying smoothly with x. A section of $\mathcal{T}$ on M is thus a vector field in $\mathcal{G} \times \mathbb{R}$ defined on $\mathcal{G}$ which is homogeneous of degree -1 with respect to the δ_s.

As we will also review in Section 2, an ambient metric for (M, c) is a pseudo-Riemannian metric $\tilde{g}$ which is defined in a dilation-invariant neighborhood $\widetilde{\mathcal{G}}$ of $\mathcal{G}$ in $\mathcal{G} \times \mathbb{R}$ by certain conditions. As indicated above, it is uniquely determined by (M, c) up to diffeomorphism to infinite order if n is odd and to order $\frac{n}{2} - 1$ if n is even.

It seems that the notion of infinitesimal holonomy was first introduced and studied systematically by Nijenhuis [1953a; 1953b; 1954]. A standard reference is [Kobayashi and Nomizu 1963]. If $(\mathcal{V}, \nabla)$ is a smooth vector bundle with connection on a manifold M and $x \in M$, the infinitesimal holonomy algebra $\mathfrak{hol}_x$ of $(\mathcal{V}, \nabla)$ at x is the subspace of $\operatorname{End} \mathcal{V}_x$ defined by

$$\mathfrak{hol}_x = \operatorname{span}_{\mathbb{R}}\big\{\nabla_{\eta_k} \nabla_{\eta_{k-1}} \cdots \nabla_{\eta_3}(R(\eta_1, \eta_2))(x) : k \geq 2,\ \eta_1, \ldots, \eta_k \in \mathfrak{X}(M)\big\}. \tag{1-2}$$

Here $\mathfrak{X}(M)$ denotes the space of smooth vector fields on M and $R : \Lambda^2 TM \to \operatorname{End} \mathcal{V}$ the curvature of ∇. It is a standard fact that $\mathfrak{hol}_x$ is a subalgebra of $\operatorname{End} \mathcal{V}_x$ for its natural Lie algebra structure with bracket the commutator of endomorphisms. Clearly $\mathfrak{hol}_x$ depends only on the infinite order jet of ∇ at x, and so in particular there is generally no relation between $\mathfrak{hol}_x$ and $\mathfrak{hol}_y$ for $x \neq y$. However, if M and $(\mathcal{V}, \nabla)$ are real-analytic, then $\mathfrak{hol}_x$ is the Lie algebra of Hol_x, where $\operatorname{Hol}_x \subset \operatorname{Aut} \mathcal{V}_x$ is the usual holonomy group of $(\mathcal{V}, \nabla)$ defined by parallel translation around loops based at x. Of course, Hol_x is always isomorphic to Hol_y for M smooth and connected.

For a conformal manifold (M, c), we denote by $\mathfrak{hol}_x$ the infinitesimal holonomy at x of $(\mathcal{T}, \nabla)$, where ∇ is the normal tractor connection. Thus $\mathfrak{hol}_x$ is a subalgebra of $\operatorname{End}\mathcal{T}_x$. The realization (1-1) of $\mathcal{T}_x$ induces the realization

$$\operatorname{End}\mathcal{T}_x = \left\{E \in \Gamma(\operatorname{End}T(\mathcal{G}\times\mathbb{R})|_{\mathcal{G}_x}) : (\delta_s)^* E = E,\ s > 0\right\} \tag{1-3}$$

of $\operatorname{End}\mathcal{T}_x$. Thus an element of $\mathfrak{hol}_x$ is realized as a section of the vector bundle $\operatorname{End}T(\mathcal{G}\times\mathbb{R})|_{\mathcal{G}_x}$ over $\mathcal{G}_x$ which is homogeneous of degree 0 with respect to the δ_s. For any $z \in \mathcal{G}_x$, evaluation at z is an isomorphism

$$\mathrm{ev}_z : \operatorname{End}\mathcal{T}_x \to \operatorname{End}T_z(\mathcal{G}\times\mathbb{R}).$$

So $\mathrm{ev}_z(\mathfrak{hol}_x)$ is an isomorphic copy of $\mathfrak{hol}_x$ in $\operatorname{End}T_z(\mathcal{G}\times\mathbb{R})$.

If $\tilde{g}$ is an ambient metric for (M, c) and $x \in M$, the infinitesimal holonomy at $z \in \mathcal{G}_x$ of the Levi-Civita connection $\widetilde{\nabla}$ of $\tilde{g}$ is a subalgebra of $\operatorname{End}T_z\widetilde{\mathcal{G}} = \operatorname{End}T_z(\mathcal{G}\times\mathbb{R})$. If n is odd, we denote this subalgebra $\widetilde{\mathfrak{hol}}_z$. This is clearly independent of the infinite-order ambiguity in $\tilde{g}$. However, when n is even, the ambient metric is determined by (M, c) only to order $\frac{n}{2}-1$ along $\mathcal{G}$. So we need to restrict the number of differentiations transverse to $\mathcal{G}$ to avoid this ambiguity. Therefore, when $n \geq 4$ is even, we define

$$\widetilde{\mathfrak{hol}}_z = \operatorname{span}_{\mathbb{R}}\left\{\widetilde{\nabla}_{\tilde{\xi}_k}\widetilde{\nabla}_{\tilde{\xi}_{k-1}}\cdots\widetilde{\nabla}_{\tilde{\xi}_3}(\widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2))(z) : k \geq 2,\ \tilde{\xi}_1, \dots, \tilde{\xi}_k \in \mathfrak{X}(\widetilde{\mathcal{G}})\right\}, \tag{1-4}$$

where $\widetilde{R}$ is the curvature of $\widetilde{\nabla}$, but we impose the requirement that no more than $\frac{n}{2}-2$ of the vector fields $\tilde{\xi}_1, \dots, \tilde{\xi}_k$ are somewhere transverse to $\mathcal{G}$. Then $\widetilde{\nabla}_{\tilde{\xi}_k}\widetilde{\nabla}_{\tilde{\xi}_{k-1}}\cdots\widetilde{\nabla}_{\tilde{\xi}_3}(\widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2))$ depends on at most $\frac{n}{2}-1$ transverse derivatives of $\tilde{g}$, so its value at z is independent of the ambiguity at order $\frac{n}{2}$. A priori, (1-4) is only defined as a vector space, but it is a consequence of Theorem 1.1 that it is a Lie subalgebra of $\operatorname{End}T_z\widetilde{\mathcal{G}}$.

Our main result is the following.

Theorem 1.1. *Let (M, c) be a conformal manifold of dimension $n \geq 3$ and $\tilde{g}$ an ambient metric for (M, c). If $x \in M$ and $z \in \mathcal{G}_x$, then*

$$\mathrm{ev}_z(\mathfrak{hol}_x) = \widetilde{\mathfrak{hol}}_z.$$

An immediate corollary is the equality of restricted tractor and ambient holonomy groups in the odd-dimensional real-analytic case. Recall that if $(\mathcal{V}, \nabla)$ is a vector bundle with connection on a smooth manifold M and $x \in M$, then the restricted holonomy group is

$$\mathrm{Hol}^0_x(\mathcal{V}, \nabla) = \{L_\gamma\} \subset \operatorname{Aut}\mathcal{V}_x,$$

where γ is a smooth contractible loop based at x and L_γ is the linear transformation of $\mathcal{V}_x$ obtained by parallel translation around γ. Just as with infinitesimal holonomy, for the tractor connection of a conformal manifold we have that if $z \in \mathcal{G}_x$, then $\mathrm{ev}_z(\mathrm{Hol}^0_x(\mathcal{T}, \nabla))$ is an isomorphic copy of $\mathrm{Hol}^0_x(\mathcal{T}, \nabla)$ in $\mathrm{Aut}\, T_z(\mathcal{G} \times \mathbb{R})$.

Corollary 1.2. *Let (M, c) be an odd-dimensional real-analytic conformal manifold and $\tilde{g}$ a real-analytic ambient metric for (M, c). If $x \in M$ and $z \in \mathcal{G}_x$, then*

$$\mathrm{ev}_z(\mathrm{Hol}^0_x(\mathcal{T}, \nabla)) = \mathrm{Hol}^0_z(T\widetilde{\mathcal{G}}, \widetilde{\nabla}).$$

Corollary 1.2 follows from Theorem 1.1 since $\mathrm{ev}_z(\mathrm{Hol}^0_x(\mathcal{T}, \nabla))$ and $\mathrm{Hol}^0_z(T\widetilde{\mathcal{G}}, \widetilde{\nabla})$ are connected Lie subgroups of $\mathrm{Aut}\, T_z\widetilde{\mathcal{G}}$ with the same Lie algebra $\mathrm{ev}_z(\mathfrak{hol}_x) = \widetilde{\mathfrak{hol}}_z$.

The tractor bundle $\mathcal{T}$ carries a tractor metric h of signature $(p+1, q+1)$ which is parallel with respect to ∇. So by choosing a frame for $\mathcal{T}_x$, one can identify $\mathrm{Hol}^0(\mathcal{T}, \nabla)$ with a subgroup of $\mathrm{SO}_e(p+1, q+1)$ which is well-defined up to conjugacy independently of x and the choice of frame (assuming M is connected). Corollary 1.2 immediately implies:

Corollary 1.3. *Let (M, c) be an odd-dimensional connected real-analytic conformal manifold. Then its restricted conformal holonomy group $\mathrm{Hol}^0(\mathcal{T}, \nabla) \subset \mathrm{SO}_e(p+1, q+1)$ is realizable as the restricted holonomy group of a real-analytic pseudo-Riemannian manifold of signature $(p+1, q+1)$.*

Corollary 1.3 is interesting because of the wealth of known information concerning pseudo-Riemannian holonomy (in particular, Berger's list) and the restriction it places on conformal holonomy groups.

If a pseudo-Riemannian manifold admits a nonzero parallel tensor field, then its holonomy group is constrained to lie in the isotropy group consisting of the linear transformations preserving the tensor at a point. Of course, many interesting pseudo-Riemannian holonomy groups arise in this fashion. Likewise, interesting classes of conformal manifolds are characterized by admitting a parallel tractor-tensor field (i.e., a section of $\otimes^r \mathcal{T}^*$ for some $r \geq 1$) of a particular algebraic type. A precursor to Theorem 1.1 is the result of [Graham and Willse 2012] asserting that a parallel tractor-tensor field on a conformal manifold admits an extension to the ambient space which is parallel with respect to the ambient metric (to infinite order for n odd, to order $\frac{n}{2} - 1$ for n even). This result was one motivation for our consideration of the question of equality of infinitesimal holonomy in general.

In order to prove Theorem 1.1, one must express the ambient connection and its curvature in tractor terms. Čap and Gover [2003] showed how the tractor bundle and connection could be written in ambient terms. This gives the inclusion $\mathrm{ev}_z(\mathfrak{hol}_x) \subset \widetilde{\mathfrak{hol}}_z$ in Theorem 1.1. Gover and Peterson [2003] reversed the direction and showed how to express the full ambient curvature and its covariant derivatives in terms of tractor calculus. Our proof of the reverse inclusion in Theorem 1.1, i.e., of ambient holonomy in tractor holonomy, is based on these relations.

In Section 2 we review the ambient metric construction and the realization of the tractor bundle and connection in ambient terms. In Section 3 we discuss infinitesimal holonomy and prove Theorem 1.1, in the process recalling the tractor expressions for the ambient curvature and connection.

2. Ambient metrics and tractors

We begin by reviewing background material concerning ambient metrics and tractors. The main reference for the material on ambient metrics is [Fefferman and Graham 2012]. References for the ambient formulation of tractors are [Čap and Gover 2003] and [Gover and Peterson 2003].

Let (M, c) be a conformal manifold of dimension $n \geq 3$ and signature (p, q), $p + q = n$. Metrics in the conformal class c are sections of the metric bundle $\mathcal{G} := \{(x, g_x) : x \in M,\ g \in c\} \subset S^2T^*M$. Let $\pi : \mathcal{G} \to M$ denote the projection and $\delta_s : \mathcal{G} \to \mathcal{G}$ the dilations defined by $\delta_s(x, g_x) = (x, s^2 g_x),\ s > 0$. Let $T = \frac{d}{ds}\delta_s|_{s=1}$ be the infinitesimal generator of the dilations. There is a tautological symmetric 2-tensor $\boldsymbol{g}$ on $\mathcal{G}$ defined for $X, Y \in T_{(x,g_x)}\mathcal{G}$ by $\boldsymbol{g}(X, Y) = g_x(\pi_* X, \pi_* Y)$.

Regard $\mathcal{G}$ as a hypersurface in $\mathcal{G} \times \mathbb{R}$ via $\iota(z) = (z, 0)$, $z \in \mathcal{G}$. The variable in the $\mathbb{R}$ factor is denoted ρ. A straight preambient metric for (M, c) is a smooth metric $\tilde{g}$ of signature $(p + 1, q + 1)$ on a dilation-invariant neighborhood $\widetilde{\mathcal{G}}$ of $\mathcal{G}$ satisfying

(1) $\delta_s^* \tilde{g} = s^2 \tilde{g}$ for $s > 0$;

(2) $\iota^* \tilde{g} = \boldsymbol{g}$;

(3) $\widetilde{\nabla} T = \mathrm{Id}$, where Id denotes the identity endomorphism and $\widetilde{\nabla}$ the Levi-Civita connection of $\tilde{g}$.

If n is odd, an ambient metric for (M, c) is a straight preambient metric for (M, c) such that $\mathrm{Ric}(\tilde{g})$ vanishes to infinite order on $\mathcal{G}$. (To infinite order, the straightness condition (3) is a consequence of the infinite order vanishing of $\mathrm{Ric}(\tilde{g})$. But this is a nontrivial result (see [Fefferman and Graham 2012]), and it is convenient to have (3) holding in a full neighborhood of $\mathcal{G}$. So (3) is included in the definition.) There exists an ambient metric for (M, c) and it is unique to infinite order up to pullback by a diffeomorphism defined on a dilation-invariant neighborhood of $\mathcal{G} \times \mathbb{R}$ which commutes with dilations and which restricts to the identity on $\mathcal{G}$. If M is a real-analytic manifold and there is a real-analytic metric in the conformal class, then there exists a real-analytic ambient metric for (M, c) satisfying $\mathrm{Ric}(\tilde{g}) = 0$ on some dilation-invariant $\widetilde{\mathcal{G}}$ as above.

In order to formulate the definition of ambient metrics for n even, let S_{IJ} be a symmetric 2-tensor field on an open neighborhood of $\mathcal{G}$ in $\mathcal{G} \times \mathbb{R}$ and $m \geq 0$. We write

$S_{IJ} = O^+_{IJ}(\rho^m)$ if $S_{IJ} = O(\rho^m)$ and, for each point $z \in \mathcal{G}$, the symmetric 2-tensor $(\iota^*(\rho^{-m}S))(z)$ is of the form $\pi^* s$ for some symmetric 2-tensor s at $x = \pi(z) \in M$ satisfying $\operatorname{tr}_{g_x} s = 0$. If n is even, an ambient metric for (M, c) is a straight preambient metric such that $\operatorname{Ric}(\tilde{g}) = O^+_{IJ}(\rho^{n/2-1})$. There exists an ambient metric for (M, c) and it is unique up to addition of a term which is $O^+_{IJ}(\rho^{n/2})$ and up to pullback by a diffeomorphism defined on a dilation-invariant neighborhood of $\mathcal{G}$ which commutes with dilations and which restricts to the identity on $\mathcal{G}$. For n even, a conformally invariant tensor, the ambient obstruction tensor, obstructs the existence of smooth solutions to $\operatorname{Ric}(\tilde{g}) = O(\rho^{n/2})$.

Let (M, c) be a conformal manifold with metric bundle $\mathcal{G} \xrightarrow{\pi} M$. For $x \in M$, write $\mathcal{G}_x = \pi^{-1}(\{x\})$ for the fiber of $\mathcal{G}$ over x. Recall that the bundle $\mathcal{D}(w)$ of conformal densities of weight $w \in \mathbb{C}$ has fiber $\mathcal{D}_x(w) = \{f : \mathcal{G}_x \to \mathbb{C} : (\delta_s)^* f = s^w f,\ s > 0\}$. Thus sections of $\mathcal{D}(w)$ on M are functions on $\mathcal{G}$ homogeneous of degree w.

The standard tractor bundle and its normal connection can be similarly realized in terms of homogeneous vector fields on $\mathcal{G}_x$. As described in the introduction, the standard tractor bundle can be realized as the rank-$(n+2)$ vector bundle $\mathcal{T} \to M$ with fiber over x given by (1-1). It can equivalently be described as an $\mathbb{R}_+$-quotient of $T\widetilde{\mathcal{G}}|_{\mathcal{G}}$; see [Čap and Gover 2003]. If $\tilde{g}$ is an ambient metric for (M, c) and if $U, W \in \mathcal{T}_x$, then $\tilde{g}(U, W)$ is homogeneous of degree 0 on $\mathcal{G}_x$, i.e., $\tilde{g}(U, W) \in \mathbb{R}$. Therefore $h(U, W) = \tilde{g}(U, W)$ defines a metric h of signature $(p+1, q+1)$ on $\mathcal{T}$, the tractor metric. Since T is homogeneous of degree 0 with respect to the δ_s, it defines a section of $\mathcal{T}(1)$, where in general we denote the effect of tensoring a bundle with $\mathcal{D}(w)$ by appending (w). The set of U in (1-1) which at each point of $\mathcal{G}_x$ is a multiple of T determines a subbundle of $\mathcal{T}$ which we denote $\operatorname{span}\{T\}$. Its orthogonal complement $\operatorname{span}\{T\}^\perp$ is the set of U which at each point of $\mathcal{G}_x$ is tangent to $\mathcal{G}$. This gives the filtration

$$0 \subset \operatorname{span}\{T\} \subset \operatorname{span}\{T\}^\perp \subset \mathcal{T}. \tag{2-1}$$

In order to realize the tractor connection, observe that $\pi_* : T\mathcal{G} \to TM$ induces a realization of the tangent bundle TM as

$$T_x M = \bigl\{\bar{\eta} \in \Gamma(T\mathcal{G}|_{\mathcal{G}_x}) : (\delta_s)^* \bar{\eta} = \bar{\eta},\ s > 0\bigr\}/\operatorname{span}\{T\}, \tag{2-2}$$

where here $\operatorname{span}\{T\}$ really means the constant multiples of T. If $\eta \in T_x M$, choose $\bar{\eta} \in \Gamma(T\mathcal{G}|_{\mathcal{G}_x})$ representing η. We will call such an $\bar{\eta}$ an invariant lift of η. Let $\tilde{g}$ be an ambient metric for (M, c) and $\widetilde{\nabla}$ its Levi-Civita connection. If U is a section of $\mathcal{T}$ near x, then $\widetilde{\nabla}_{\bar{\eta}} U \in \Gamma(T\widetilde{\mathcal{G}}|_{\mathcal{G}_x})$ makes sense since U is defined on $\mathcal{G}$ and $\bar{\eta}$ is tangent to $\mathcal{G}$. The straightness of $\tilde{g}$ and the homogeneity of U imply that $\widetilde{\nabla}_T U = 0$. Therefore $\widetilde{\nabla}_{\bar{\eta}} U$ is independent of the choice of invariant lift $\bar{\eta}$. Also $\widetilde{\nabla}_{\bar{\eta}} U$ has the same homogeneity as U, so $\widetilde{\nabla}_{\bar{\eta}} U$ defines an element of $\mathcal{T}_x$. This realizes the tractor connection ∇ on $\mathcal{T}$:

$$\nabla_\eta U = \widetilde{\nabla}_{\bar{\eta}} U. \tag{2-3}$$

The tractor metric h is parallel with respect to ∇ since $\widetilde{\nabla}\tilde{g} = 0$. These realizations of the tractor metric and connection depend on the choice of ambient metric $\tilde{g}$. But the realizations obtained by changing $\tilde{g}$ by a diffeomorphism are equivalent.

The realization (1-1) of the tractor bundle induces the realizations

$$(\otimes^r \mathcal{T}^*)_x = \big\{\chi \in \Gamma(\otimes^r T^*\widetilde{\mathcal{G}}|_{\mathcal{G}_x}) : (\delta_s)^*\chi = s^r\chi,\ s > 0\big\}, \quad r \in \mathbb{N}, \tag{2-4}$$

of the bundles of cotractor-tensors, as well as the realization (1-3) of the bundle of tractor endomorphisms. The induced tractor connections on these bundles are also given in terms of the ambient connection and an invariant lift $\bar{\eta}$ as in (2-3). Throughout this paper we will identify weighted tractor-tensors with homogeneous sections of bundles on $\mathcal{G}$ as in (1-1), (1-3), (2-4).

The curvature R of the tractor connection can be expressed in terms of the curvature $\widetilde{R}$ of an ambient metric. We have $R : \Lambda^2 TM \to \operatorname{End}\mathcal{T}$ and $\widetilde{R} : \Lambda^2 T\widetilde{\mathcal{G}} \to \operatorname{End} T\widetilde{\mathcal{G}}$. The straightness of the ambient metric implies that $T \lrcorner \widetilde{R} = 0$ on $\mathcal{G}$. So if $\eta_1, \eta_2 \in T_x M$ and $\bar{\eta}_1, \bar{\eta}_2 \in \Gamma(T\mathcal{G}|_{\mathcal{G}_x})$ are invariant lifts, then $\widetilde{R}(\bar{\eta}_1, \bar{\eta}_2) \in \Gamma(\operatorname{End} T\widetilde{\mathcal{G}}|_{\mathcal{G}_x})$ is independent of the choices of $\bar{\eta}_1, \bar{\eta}_2$. Moreover, $\widetilde{R}(\bar{\eta}_1, \bar{\eta}_2)$ is homogeneous of degree 0 with respect to the δ_s, so it realizes an element of $\operatorname{End}\mathcal{T}_x$, and one has

$$R(\eta_1, \eta_2) = \widetilde{R}(\bar{\eta}_1, \bar{\eta}_2). \tag{2-5}$$

We follow usual notational conventions. We label tensors on the ambient space and therefore also tractors with capital Latin indices and vectors on M with lower case Latin indices. We use $\mathcal{E}$ to denote the space of smooth sections of a bundle on M, the bundle specified by the accompanying indices. Just as with the bundles themselves, we denote the spaces of sections of the corresponding weighted bundles by appending (w). The notation $\mathcal{E}^\Phi(w)$ signifies the space of sections of a generic weighted tractor bundle, where Φ denotes an arbitrary collection of upper and lower capital indices. If Φ consists of r upper indices and s lower indices, we denote by $\widetilde{\mathcal{E}}^\Phi(w)$ the space of sections of $(\otimes^r T\widetilde{\mathcal{G}}) \otimes (\otimes^s T^*\widetilde{\mathcal{G}})$ on $\widetilde{\mathcal{G}}$ of the same homogeneity degree as sections of $\mathcal{E}^\Phi(w)$, i.e., of homogeneity degree $w - r + s$. Ambient/tractor indices are raised and lowered using the ambient/tractor metric $\tilde{g}_{AB}/h_{AB}$ and lower case indices using the conformal metric $\boldsymbol{g}_{ij} \in \mathcal{E}_{ij}(2)$.

A choice of metric g in the conformal class induces a splitting of the cotractor bundle

$$\mathcal{T}^* = \mathcal{D}(-1) \oplus T^*M(1) \oplus \mathcal{D}(1). \tag{2-6}$$

This is the formulation in the original definition of the tractor bundle in [Bailey et al. 1994]. It can also be viewed in terms of the ambient realization by putting $\tilde{g}$

in normal form relative to g (see [Gover and Peterson 2003] or [Graham and Willse 2012]). The three inclusions determined by this splitting determine sections

$$X_A \in \mathcal{E}_A(1), \quad Z_A{}^i \in \mathcal{E}_A{}^i(-1), \quad Y_A \in \mathcal{E}_A(-1)$$

so that

$$V_A = \varphi X_A + \psi_i Z_A{}^i + \rho Y_A \tag{2-7}$$

corresponds to $V_A = (\varphi, \psi_i, \rho) \in \mathcal{E}(-1) \oplus \mathcal{E}_i(1) \oplus \mathcal{E}(1)$. The sections Y_A and $Z_A{}^i$ are scale-dependent, i.e., they depend on the choice of g, while X_A is scale-independent: $X^A \in \mathcal{E}^A(1)$ is another notation for the weighted tractor defined by the vector field $T|_{\mathcal{G}}$.

3. Holonomy

Recall from the introduction that the infinitesimal holonomy $\mathfrak{hol}_x$ of a vector bundle with connection $(\mathcal{V}, \nabla)$ on a manifold M is defined pointwise by (1-2), and Theorem 1.1 is stated in terms of pointwise infinitesimal holonomy. The proof of Theorem 1.1 goes by induction on the order of differentiation. Thus it is natural to formulate an induction hypothesis involving objects which can be differentiated. So we introduce spaces consisting of global sections which restrict at each point to the infinitesimal holonomy. For $k \geq 2$, we define

$$\mathfrak{hol}_M^k = \operatorname{span}_{C^\infty(M)}\big\{\nabla_{\eta_l}\nabla_{\eta_{l-1}}\cdots\nabla_{\eta_3}(R(\eta_1,\eta_2)) : 2 \leq l \leq k,\ \eta_1,\dots,\eta_l \in \mathfrak{X}(M)\big\} \tag{3-1}$$

and

$$\mathfrak{hol}_M = \bigcup_{k \geq 2} \mathfrak{hol}_M^k$$

so that $\mathfrak{hol}_M^k, \mathfrak{hol}_M \subset \Gamma(\operatorname{End}\mathcal{V})$. Clearly $\mathfrak{hol}_x = \{E(x) : E \in \mathfrak{hol}_M\}$. One has

$$[\mathfrak{hol}_M^k, \mathfrak{hol}_M^l] \subset \mathfrak{hol}_M^{k+l}. \tag{3-2}$$

In fact, the proof in [Kobayashi and Nomizu 1963] that $\mathfrak{hol}_x$ is a subalgebra of $\operatorname{End}\mathcal{V}_x$ establishes the analog of (3-2) in the principal bundle setting.

There is an alternate characterization of these spaces in terms of iterated covariant derivatives with respect to a coupled connection. If we choose arbitrarily a connection on TM and denote also by ∇ the coupled connection on $\mathcal{V} \otimes TM$, then the Leibniz formula and induction show that

$$\mathfrak{hol}_M^k = \operatorname{span}_{C^\infty(M)}\big\{(\nabla^{l-2}R)(\eta_1,\eta_2,\dots,\eta_l) : 2 \leq l \leq k,\ \eta_1,\dots,\eta_l \in \mathfrak{X}(M)\big\}. \tag{3-3}$$

R again denotes the curvature of the connection on $\mathcal{V}$. Here it is viewed as a section of $\Lambda^2 T^*M \otimes \operatorname{End}\mathcal{V}$ and $\nabla^{l-2}R$ denotes its iterated covariant derivative with respect to the coupled connection.

If (M, c) is a conformal manifold, we take $\mathcal{V} = \mathcal{T}$ to be the tractor bundle with its normal connection and we denote the corresponding spaces by $\mathfrak{hol}^k_M$, $\mathfrak{hol}_M$. As usual, via our realization (1-3) we identify elements of $\mathfrak{hol}_M$ as global sections of $\operatorname{End} T(\mathcal{G} \times \mathbb{R})|_{\mathcal{G}}$ which are homogeneous of degree 0 with respect to the δ_s.

For the ambient metric we modify the definition slightly to respect homogeneity. If n is odd and $\tilde{g}$ is an ambient metric for (M, c), we define for $k \geq 2$

$$\widetilde{\mathfrak{hol}}^k_M = \operatorname{span}_{C^\infty(M)}\big\{\widetilde{\nabla}_{\tilde{\xi}_l}\widetilde{\nabla}_{\tilde{\xi}_{l-1}}\cdots\widetilde{\nabla}_{\tilde{\xi}_3}(\widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2))|_{\mathcal{G}} : 2 \leq l \leq k,\ \tilde{\xi}_1, \dots, \tilde{\xi}_l \in \mathfrak{X}_0(\widetilde{\mathcal{G}})\big\}, \tag{3-4}$$

where $\mathfrak{X}_0(\widetilde{\mathcal{G}})$ denotes the space of smooth vector fields on $\widetilde{\mathcal{G}}$ which are homogeneous of degree 0 with respect to the δ_s and $C^\infty(M)$ is viewed as the subspace of $C^\infty(\mathcal{G})$ of functions homogeneous of degree 0. Observe that by definition, $\widetilde{\mathfrak{hol}}^k_M \subset \Gamma(\operatorname{End} T\widetilde{\mathcal{G}}|_{\mathcal{G}})$ consists of sections which are homogeneous of degree 0. If n is even, we again define $\widetilde{\mathfrak{hol}}^k_M$ for $k \geq 2$ by (3-4), except that we require that at most $\frac{n}{2} - 2$ of the $\tilde{\xi}_i$ are somewhere transverse to $\mathcal{G}$. For general n, we then set

$$\widetilde{\mathfrak{hol}}_M = \bigcup_{k \geq 2} \widetilde{\mathfrak{hol}}^k_M.$$

As above, $\widetilde{\mathfrak{hol}}_M$ also has a description in terms of iterated derivatives of curvature:

$$\widetilde{\mathfrak{hol}}^k_M = \operatorname{span}_{C^\infty(M)}\big\{(\widetilde{\nabla}^{l-2}\widetilde{R})(\tilde{\xi}_1, \tilde{\xi}_2, \dots, \tilde{\xi}_l)|_{\mathcal{G}} : 2 \leq l \leq k,\ \tilde{\xi}_1, \dots, \tilde{\xi}_l \in \mathfrak{X}_0(\widetilde{\mathcal{G}})\big\}. \tag{3-5}$$

Here we take the coupling connection on $T\widetilde{\mathcal{G}}$ also to be the Levi-Civita connection $\widetilde{\nabla}$. As usual, for n even we require that at most $\frac{n}{2} - 2$ of the $\tilde{\xi}_i$ are somewhere transverse to $\mathcal{G}$. In this case, the equivalence of the descriptions (3-4) and (3-5) only holds for $k \leq \frac{n}{2} - 1$, since $\widetilde{\nabla}_{\tilde{\xi}}\tilde{\eta}$ can be transverse to $\mathcal{G}$ when both $\tilde{\xi}|_{\mathcal{G}}$ and $\tilde{\eta}|_{\mathcal{G}}$ are tangent to $\mathcal{G}$.

We claim $\widetilde{\mathfrak{hol}}_z = \{E(z) : E \in \widetilde{\mathfrak{hol}}_M\}$. To see this, choose a frame $\tilde{\zeta}_0, \tilde{\zeta}_1, \dots, \tilde{\zeta}_{n+1}$ for $T\widetilde{\mathcal{G}}$ near z such that $\tilde{\zeta}_A|_{\mathcal{G}}$ is tangent to $\mathcal{G}$ for $1 \leq A \leq n+1$, and such that each $\tilde{\zeta}_A$ is homogeneous of degree 0 with respect to the δ_s. By writing each $\tilde{\xi}_i$ in (1-4) as a linear combination of the $\tilde{\zeta}_A$, it is not hard to see that

$$\widetilde{\mathfrak{hol}}_z = \operatorname{span}_{\mathbb{R}}\big\{\widetilde{\nabla}_{\tilde{\zeta}_{A_k}}\widetilde{\nabla}_{\tilde{\zeta}_{A_{k-1}}}\cdots\widetilde{\nabla}_{\tilde{\zeta}_{A_3}}(\widetilde{R}(\tilde{\zeta}_{A_1}, \tilde{\zeta}_{A_2}))(z) : k \geq 2\big\},$$

where for n even at most $\frac{n}{2} - 2$ of the indices $A_1, \dots, A_k$ are equal to 0. It follows immediately that $\widetilde{\mathfrak{hol}}_z = \{E(z) : E \in \widetilde{\mathfrak{hol}}_M\}$.

In light of these observations, it is clear that Theorem 1.1 is a consequence of the following theorem.

Theorem 3.1. *Let (M, c) be a conformal manifold of dimension $n \geq 3$ and $\tilde{g}$ an ambient metric for (M, c). Then*

$$\mathfrak{hol}_M = \widetilde{\mathfrak{hol}}_M.$$

The inclusion $\mathfrak{hol}_M \subset \widetilde{\mathfrak{hol}}_M$ follows immediately from the ambient realizations of the tractor connection and curvature. If $\eta_1, \ldots, \eta_k \in \mathfrak{X}(M)$ and $\bar{\eta}_1, \ldots, \bar{\eta}_k$ are invariant lifts, then (2-3), (2-5) give

$$\nabla_{\eta_k} \nabla_{\eta_{k-1}} \cdots \nabla_{\eta_3} (R(\eta_1, \eta_2)) = \widetilde{\nabla}_{\bar{\eta}_k} \widetilde{\nabla}_{\bar{\eta}_{k-1}} \cdots \widetilde{\nabla}_{\bar{\eta}_3} (\widetilde{R}(\bar{\eta}_1, \bar{\eta}_2)), \tag{3-6}$$

so $\mathfrak{hol}_M \subset \widetilde{\mathfrak{hol}}_M$. The right-hand side is in $\widetilde{\mathfrak{hol}}_M$ also for n even since none of the $\bar{\eta}_i$ are transverse to $\mathcal{G}$.

We remark that (3-6) is already sufficient to prove Theorem 3.1, and therefore also Theorem 1.1, when $n = 4$. In fact, when $n = 4$, each $\tilde{\xi}_i|_{\mathcal{G}}$ in (3-4) is required to be everywhere tangent to $\mathcal{G}$, so is an invariant lift of some $\eta_i \in \mathfrak{X}(M)$.

To prove the opposite inclusion $\widetilde{\mathfrak{hol}}_M \subset \mathfrak{hol}_M$, we must rewrite expressions of the form $\widetilde{\nabla}_{\tilde{\xi}_l} \widetilde{\nabla}_{\tilde{\xi}_{l-1}} \cdots \widetilde{\nabla}_{\tilde{\xi}_3} (\widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2))|_{\mathcal{G}}$ purely in tractor terms when the $\tilde{\xi}_i$ are allowed to be transverse to $\mathcal{G}$. We do this using tractor representations of the curvature and connection of the ambient metric derived in [Gover and Peterson 2003]. These representations are expressed in terms of the splitting (2-6), (2-7) of the cotractor bundle determined by a choice of metric $g \in c$. Consider first the case n odd.

Proof of Theorem 3.1 for n odd. We show by induction on $k \geq 2$ that $\widetilde{\mathfrak{hol}}^k_M \subset \mathfrak{hol}_M$. For $k = 2$, we use the tractor expression for ambient curvature

$$\widetilde{R}_{AB}{}^P{}_Q|_{\mathcal{G}} = Z_A{}^a Z_B{}^b R_{ab}{}^P{}_Q - \frac{2}{n-4} X_{[A} Z_{B]}{}^b \nabla^c R_{cb}{}^P{}_Q.$$

This is (13), (35) of [Gover and Peterson 2003]. The ∇^c on the right-hand side refers to the connection obtained by coupling the tractor connection with the Levi-Civita connection of the chosen representative metric g. Now $\widetilde{\mathfrak{hol}}^2_M$ is spanned by contractions of the left-hand side against $\tilde{\xi}_1^A \tilde{\xi}_2^B$, where $\tilde{\xi}_1, \tilde{\xi}_2 \in \mathfrak{X}_0(\widetilde{\mathcal{G}})$. It is evident that after such a contraction, the first term on the right-hand side is in $\mathfrak{hol}^2_M$. For the second term, write $\nabla^c R_{cb}{}^P{}_Q = \boldsymbol{g}^{cd} \nabla_c R_{db}{}^P{}_Q$ and introduce a partition of unity subordinate to a covering of M in each open set of which $\boldsymbol{g}^{cd}$ can be expressed as a smooth linear combination of tensor products of vector fields. It follows that after contraction with $\tilde{\xi}_1^A \tilde{\xi}_2^B$, the second term is in $\mathfrak{hol}^3_M$. Thus the initial $k = 2$ step of the induction is established.

The induction step for higher k will be carried out using the tractor-D operator. If Φ denotes an arbitrary collection of upper and/or lower tractor indices, then

$$D_A : \mathcal{E}^\Phi(w) \to \mathcal{E}^\Phi{}_A(w-1)$$

is defined in terms of the splitting determined by a representative metric g by

$$D_A V = w(n+2w-2) Y_A V + (n+2w-2) Z_A{}^a \nabla_a V - X_A \Box V, \tag{3-7}$$

where $\Box V = \nabla^i \nabla_i V + wJV$ and $J = R/(2(n-1))$. D_A can also be expressed in ambient terms:

$$D_A V = (n+2w-2)\widetilde{\nabla}_A \widetilde{V}|_{\mathcal{G}} - X_A(\widetilde{\Delta}\widetilde{V})|_{\mathcal{G}}. \tag{3-8}$$

These are (8), (31) of [Gover and Peterson 2003]. On the right-hand side, $\widetilde{V} \in \widetilde{\mathcal{E}}^{\Phi}(w)$ is an arbitrary homogeneous extension of $V \in \mathcal{E}^{\Phi}(w)$ and $\widetilde{\Delta}$ denotes the ambient Laplacian acting on the corresponding space of tensors: $\widetilde{\Delta} = \widetilde{\nabla}^I \widetilde{\nabla}_I$. The expression on the right-hand side turns out to be independent of the choice of $\widetilde{V}$.

Assume now that $k \geq 2$ and $\widetilde{\mathfrak{hol}}^k_M \subset \mathfrak{hol}_M$. According to (3-5), in order to prove that $\widetilde{\mathfrak{hol}}^{k+1}_M \subset \mathfrak{hol}_M$, it suffices to show that $\bigl(\tilde{\xi}_1^A \tilde{\xi}_2^B \cdots \tilde{\xi}_{k+1}^E \widetilde{\nabla}^{k-1}_{A\cdots C} \widetilde{R}_{DE}{}^P{}_Q\bigr)\big|_{\mathcal{G}} \in \mathfrak{hol}_M$ for $\tilde{\xi}_1, \ldots, \tilde{\xi}_{k+1} \in \mathfrak{X}_0(\widetilde{\mathcal{G}})$. Set $\xi_s^A = \tilde{\xi}_s^A|_{\mathcal{G}} \in \mathcal{E}^A(1)$, $1 \leq s \leq k+1$.

Define

$$\widetilde{V} = \widetilde{\nabla}^{k-2}_{B\cdots C} \widetilde{R}_{DE}{}^P{}_Q \in \widetilde{\mathcal{E}}_{B\cdots E}{}^P{}_Q(-k)$$

and rewrite (3-8) as

$$(n-2k-2)\widetilde{\nabla}_A \widetilde{V}|_{\mathcal{G}} = D_A V + X_A(\widetilde{\Delta}\widetilde{V})|_{\mathcal{G}},$$

where $V := \widetilde{V}|_{\mathcal{G}} \in \mathcal{E}_{B\cdots E}{}^P{}_Q(-k)$. Since the coefficient $(n-2k-2)$ is nonzero for n odd, it suffices to show that

$$\xi_1^A \cdots \xi_{k+1}^E D_A V_{B\cdots E}{}^P{}_Q \in \mathfrak{hol}_M \tag{3-9}$$

and

$$\tilde{\xi}_2^B \cdots \tilde{\xi}_{k+1}^E \widetilde{\Delta} \widetilde{V}_{B\cdots E}{}^P{}_Q|_{\mathcal{G}} \in \mathfrak{hol}_M. \tag{3-10}$$

For (3-9), contract (3-7) against $\xi_1^A \cdots \xi_{k+1}^E$. The first term on the right-hand side gives a multiple of

$$(\xi_1^A Y_A)\xi_2^B \cdots \xi_{k+1}^E V_{B\cdots E}{}^P{}_Q,$$

which is in $\mathfrak{hol}_M$ by the induction hypothesis. The second term on the right-hand side gives a multiple of

$$(\xi_1^A Z_A{}^a)\xi_2^B \cdots \xi_{k+1}^E \nabla_a V_{B\cdots E}{}^P{}_Q.$$

If we set $\eta^a = \xi_1^A Z_A{}^a$, then this can be rewritten as

$$\eta^a \xi_2^B \cdots \xi_{k+1}^E \nabla_a V_{B\cdots E}{}^P{}_Q = \nabla_\eta(\xi_2^B \cdots \xi_{k+1}^E V_{B\cdots E}{}^P{}_Q) - \sum_{s=2}^{k+1} \xi_2^B \cdots (\nabla_\eta \xi_s^R) \cdots \xi_{k+1}^E V_{B\cdots R\cdots E}{}^P{}_Q.$$

The induction hypothesis shows that $\xi_2^B \cdots \xi_{k+1}^E V_{B\cdots E}{}^P{}_Q \in \mathfrak{hol}_M$, so we conclude that $\nabla_\eta(\xi_2^B \cdots \xi_{k+1}^E V_{B\cdots E}{}^P{}_Q) \in \mathfrak{hol}_M$. Each term in the sum on the right-hand side is clearly in $\mathfrak{hol}_M$ by the induction hypothesis. Thus the contraction of the second term of the right-hand side of (3-7) is in $\mathfrak{hol}_M$. The third term of (3-7) is handled similarly, namely by expanding the difference

$$\xi_2^B \cdots \xi_{k+1}^E \nabla^c \nabla_c V_{B\cdots E}{}^P{}_Q - \nabla^c \nabla_c (\xi_2^B \cdots \xi_{k+1}^E V_{B\cdots E}{}^P{}_Q)$$

using the Leibniz rule and introducing a partition of unity to rewrite sections of tensor product bundles as sums of tensor products of sections of the factors as in the proof in the case $k = 2$. This concludes the proof of (3-9).

It remains to prove (3-10). Now $\widetilde{\Delta}\widetilde{V} = \widetilde{\Delta}\widetilde{\nabla}^{k-2}\widetilde{R}$. It is well-known that the Laplacian of an iterated covariant derivative of the curvature tensor of a Ricci-flat metric can be reexpressed as a linear combination of quadratic terms in curvature by commuting both derivatives in $\widetilde{\Delta}$ all the way to the right and applying the second Bianchi identity. We will argue using the induction hypothesis that each commutator term is already in $\mathfrak{hol}_M$.

Write

$$\widetilde{\Delta}\widetilde{\nabla}^{k-2}\widetilde{R}_{DE}{}^P{}_Q = \tilde{g}^{IJ}\widetilde{\nabla}_I\widetilde{\nabla}_J\widetilde{\nabla}^{k-2}\widetilde{R}_{DE}{}^P{}_Q.$$

First commute $\widetilde{\nabla}_J$ to the right of all derivatives in $\widetilde{\nabla}^{k-2}$. Modulo commutator terms, one obtains

$$\tilde{g}^{IJ}\widetilde{\nabla}_I\widetilde{\nabla}^{k-2}\widetilde{\nabla}_J\widetilde{R}_{DE}{}^P{}_Q = \tilde{g}^{IJ}\widetilde{\nabla}_I\widetilde{\nabla}^{k-2}\widetilde{\nabla}_D\widetilde{R}_{JE}{}^P{}_Q + \tilde{g}^{IJ}\widetilde{\nabla}_I\widetilde{\nabla}^{k-2}\widetilde{\nabla}_E\widetilde{R}_{DJ}{}^P{}_Q.$$

Now commuting $\widetilde{\nabla}_I$ all the way to the right shows that modulo commutators the above is equal to

$$\tilde{g}^{IJ}\widetilde{\nabla}^{k-2}\widetilde{\nabla}_D\widetilde{\nabla}_I\widetilde{R}_{JE}{}^P{}_Q + \tilde{g}^{IJ}\widetilde{\nabla}^{k-2}\widetilde{\nabla}_E\widetilde{\nabla}_I\widetilde{R}_{DJ}{}^P{}_Q.$$

This vanishes on $\mathcal{G}$ by the second Bianchi identity and the infinite-order vanishing of $\operatorname{Ric}\tilde{g}$.

To analyze the commutator terms, it is convenient to suppress writing the $\operatorname{End} T\widetilde{\mathcal{G}}$ indices ${}^P{}_Q$. We will denote by $\widetilde{\mathcal{R}}_{BC}$ the curvature tensor of $\tilde{g}$ viewed as an $\operatorname{End} T\widetilde{\mathcal{G}}$-valued section of $\Lambda^2 T^*\widetilde{\mathcal{G}}$. If U is an $\operatorname{End} T\widetilde{\mathcal{G}}$-valued section of $\otimes^r T^*\widetilde{\mathcal{G}}$ and V is an $\operatorname{End} T\widetilde{\mathcal{G}}$-valued section of $\otimes^s T^*\widetilde{\mathcal{G}}$, we will denote by $[U, V]$ the $\operatorname{End} T\widetilde{\mathcal{G}}$-valued section of $\otimes^{r+s} T^*\widetilde{\mathcal{G}}$ which is the commutator in the $\operatorname{End} T\widetilde{\mathcal{G}}$ indices and the tensor product in the $T^*\widetilde{\mathcal{G}}$ indices. The Leibniz formula gives

$$\widetilde{\nabla}[U, V] = [\widetilde{\nabla}U, V] + [U, \widetilde{\nabla}V]. \tag{3-11}$$

The Ricci identity for commuting covariant derivatives can be written

$$[\widetilde{\nabla}_B, \widetilde{\nabla}_C]U = \widetilde{\mathcal{R}}_{BC}.U + [\widetilde{\mathcal{R}}_{BC}, U], \tag{3-12}$$

where $\widetilde{\mathcal{R}}_{BC}.U$ denotes the action of the endomorphism $\widetilde{\mathcal{R}}_{BC}$ on the $\otimes^r T^*\widetilde{\mathcal{G}}$ indices of U.

Every commutator which arose in the above argument was of the form

$$\widetilde{\nabla}^i[\widetilde{\nabla}_B, \widetilde{\nabla}_C]\widetilde{\nabla}^j\widetilde{\mathcal{R}}$$

for some choice of indices B, C, where $i \geq 0$, $j \geq 0$, and $i + j = k - 2$. Express the commutator $[\widetilde{\nabla}_B, \widetilde{\nabla}_C]\widetilde{\nabla}^j\widetilde{\mathcal{R}}$ using (3-12) with $U = \widetilde{\nabla}^j\widetilde{\mathcal{R}}$. The first term on the right-hand side of (3-12) gives terms of the form $\widetilde{\nabla}^i(\widetilde{\mathcal{R}}.\widetilde{\nabla}^j\widetilde{\mathcal{R}})$. Expanding the $\widetilde{\nabla}^i$ with the Leibniz rule, it is clear that one obtains a sum of terms, each of which has the form

$$\operatorname{contr}(\widetilde{\nabla}^p\widetilde{\mathcal{R}} \otimes \widetilde{\nabla}^q\widetilde{\mathcal{R}}), \tag{3-13}$$

with $p \geq 0$, $q \geq 0$, and $p + q = k - 2$. Here contr indicates a single contraction of the upper $\operatorname{End} T^*\widetilde{\mathcal{G}}$ index of $\widetilde{\nabla}^p\widetilde{R}$ against one of the $\otimes^{q+2}T^*\widetilde{\mathcal{G}}$ indices of $\widetilde{\nabla}^q\widetilde{\mathcal{R}}$. In particular, the suppressed $\operatorname{End} T^*\widetilde{\mathcal{G}}$ indices are those on $\widetilde{\nabla}^q\widetilde{\mathcal{R}}$. The second term on the right-hand side of (3-12) gives terms of the form $\widetilde{\nabla}^i[\widetilde{\mathcal{R}}, \widetilde{\nabla}^j\widetilde{\mathcal{R}}]$. Expanding the $\widetilde{\nabla}^i$ using (3-11), one obtains a sum of terms of the form

$$[\widetilde{\nabla}^p\widetilde{\mathcal{R}}, \widetilde{\nabla}^q\widetilde{\mathcal{R}}], \tag{3-14}$$

again with $p \geq 0$, $q \geq 0$, and $p + q = k - 2$.

We need to show (3-10). Suppressing the $\operatorname{End} T\widetilde{\mathcal{G}}$ indices, we have

$$\tilde{\xi}_2^B \cdots \tilde{\xi}_{k+1}^E \widetilde{\Delta}\widetilde{V}_{B\cdots E}|_{\mathcal{G}} = \tilde{\xi}_2^B \cdots \tilde{\xi}_{k+1}^E \tilde{g}^{IJ}\widetilde{\nabla}_I\widetilde{\nabla}_J\widetilde{\nabla}^{k-2}_{B\cdots C}\widetilde{\mathcal{R}}_{DE}|_{\mathcal{G}}.$$

Upon commuting $\widetilde{\nabla}_I$ and $\widetilde{\nabla}_J$ to the right as described above, it follows that this may be written as a sum of contractions of terms of the form (3-13), (3-14) against $\tilde{\xi}_i$ and $\tilde{g}^{IJ}$ with all indices contracted except for the suppressed $\operatorname{End} T\widetilde{\mathcal{G}}$ indices. In a term (3-13), the free $\operatorname{End} T\widetilde{\mathcal{G}}$ indices are those on the second factor $\widetilde{\nabla}^q\widetilde{\mathcal{R}}$. Consequently, we can introduce a partition of unity and express locally the tensor arising from $\tilde{g}^{IJ}$, $\widetilde{\nabla}^p\widetilde{\mathcal{R}}$, and the $\tilde{\xi}_i$ which contracts against the other $q + 2$ indices of $\widetilde{\nabla}^q\widetilde{\mathcal{R}}$ as a sum of tensor products of vector fields. Since $q \leq k - 2$, it follows by the induction hypothesis that all these terms are in $\mathfrak{hol}_M$ when restricted to $\mathcal{G}$. In a term (3-14), all the indices except the endomorphism indices are contracted against $\tilde{g}^{IJ}$ and the $\tilde{\xi}_i$. Again use a partition of unity and express locally $\tilde{g}^{IJ}$ as a sum of tensor products of vector fields. Then the induction hypothesis implies that the restriction to $\mathcal{G}$ of the contractions against $\widetilde{\nabla}^p\widetilde{\mathcal{R}}$ and $\widetilde{\nabla}^q\widetilde{\mathcal{R}}$ are separately in $\mathfrak{hol}_M$. It follows from (3-2) that the commutator is also in $\mathfrak{hol}_M$. □

Proof of Theorem 3.1 for n even. We have already observed that (3-6) is sufficient to prove the case $n = 4$. So we assume that $n \geq 6$. We next observe that the same argument used for n odd applies also when n is even to show $\widetilde{\mathfrak{hol}}_M^{n/2-1} \subset \mathfrak{hol}_M$. In fact, up to this order the relevant constant $n + 2w - 2$ in (3-8) is nonzero and the argument only uses $\operatorname{Ric}(\tilde{g}) = O(\rho^{n/2-1})$.

For $n \geq 6$ even, we prove $\widetilde{\mathfrak{hol}}_M^k \subset \mathfrak{hol}_M$ by induction on k, beginning with the case $k = \frac{n}{2} - 1$. So assume for some $k \geq \frac{n}{2} - 1$ that $\widetilde{\mathfrak{hol}}_M^k \subset \mathfrak{hol}_M$ and we will show $\widetilde{\mathfrak{hol}}_M^{k+1} \subset \mathfrak{hol}_M$. According to (3-4), we have to show that

$$\widetilde{\nabla}_{\tilde{\xi}_{k+1}} \widetilde{\nabla}_{\tilde{\xi}_k} \cdots \widetilde{\nabla}_{\tilde{\xi}_3} (\widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2))|_{\mathcal{G}} \in \mathfrak{hol}_M$$

whenever $\tilde{\xi}_1, \ldots, \tilde{\xi}_{k+1} \in \mathfrak{X}_0(\widetilde{\mathcal{G}})$ and at most $\frac{n}{2} - 2$ of the $\tilde{\xi}_i$ are somewhere transverse to $\mathcal{G}$. Since $k + 1 \geq \frac{n}{2}$, at least two of the $\tilde{\xi}_i$ are everywhere tangent to $\mathcal{G}$. If $\tilde{\xi}_{k+1}$ is everywhere tangent to $\mathcal{G}$, then its restriction to $\mathcal{G}$ is the invariant lift of some $\eta \in \mathfrak{X}(M)$. In this case (2-3) gives

$$\widetilde{\nabla}_{\tilde{\xi}_{k+1}} \widetilde{\nabla}_{\tilde{\xi}_k} \cdots \widetilde{\nabla}_{\tilde{\xi}_3} (\widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2))|_{\mathcal{G}} = \nabla_\eta \big(\widetilde{\nabla}_{\tilde{\xi}_k} \cdots \widetilde{\nabla}_{\tilde{\xi}_3} (\widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2))|_{\mathcal{G}} \big).$$

The induction hypothesis shows that $\widetilde{\nabla}_{\tilde{\xi}_k} \cdots \widetilde{\nabla}_{\tilde{\xi}_3} (\widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2))|_{\mathcal{G}} \in \mathfrak{hol}_M$, from which it follows that $\nabla_\eta \big(\widetilde{\nabla}_{\tilde{\xi}_k} \cdots \widetilde{\nabla}_{\tilde{\xi}_3} (\widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2))|_{\mathcal{G}} \big) \in \mathfrak{hol}_M$, as desired.

If $\tilde{\xi}_i$ is everywhere tangent to $\mathcal{G}$ for some i, $3 \leq i \leq k$, then we can commute $\widetilde{\nabla}_{\tilde{\xi}_i}$ all the way to the left and reduce to the previous case. Modulo relabeling the indices, each commutator is of the form

$$\begin{aligned} \widetilde{\nabla}_{\tilde{\xi}_{k+1}} \cdots \widetilde{\nabla}_{\tilde{\xi}_{j+1}} \big[\widetilde{\nabla}_{\tilde{\xi}_j}, \widetilde{\nabla}_{\tilde{\xi}_{j-1}} \big] \widetilde{\nabla}_{\tilde{\xi}_{j-2}} \cdots \widetilde{\nabla}_{\tilde{\xi}_3} \widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2) \\ = \widetilde{\nabla}_{\tilde{\xi}_{k+1}} \cdots \widetilde{\nabla}_{\tilde{\xi}_{j+1}} \widetilde{\nabla}_{[\tilde{\xi}_j, \tilde{\xi}_{j-1}]} \widetilde{\nabla}_{\tilde{\xi}_{j-2}} \cdots \widetilde{\nabla}_{\tilde{\xi}_3} \widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2) \\ + \widetilde{\nabla}_{\tilde{\xi}_{k+1}} \cdots \widetilde{\nabla}_{\tilde{\xi}_{j+1}} \big[\widetilde{R}(\tilde{\xi}_j, \tilde{\xi}_{j-1}), \widetilde{\nabla}_{\tilde{\xi}_{j-2}} \cdots \widetilde{\nabla}_{\tilde{\xi}_3} \widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2) \big]. \end{aligned}$$

In the first term on the right-hand side, the number of differentiations has decreased by 1 without increasing the number of vector fields somewhere transverse to $\mathcal{G}$, since the commutator of two vector fields tangent to $\mathcal{G}$ is also tangent to $\mathcal{G}$. So the restriction to $\mathcal{G}$ of this term is in $\mathfrak{hol}_M$ by the induction hypothesis. In the second term on the right-hand side, expand the derivatives outside the commutator using the Leibniz rule. One obtains a linear combination of commutators of covariant derivatives of curvature endomorphisms. The restriction to $\mathcal{G}$ of each such covariant derivative itself is in $\mathfrak{hol}_M$ by the induction hypothesis. Equation (3-2) then shows that the commutator is in $\mathfrak{hol}_M$.

Finally we must consider the possibility that none of $\tilde{\xi}_3, \ldots, \tilde{\xi}_{k+1}$ is everywhere tangent to $\mathcal{G}$. (This can only happen in the beginning case $k = \frac{n}{2} - 1$, but we will not use this.) It must be that $\tilde{\xi}_1$ and $\tilde{\xi}_2$ are everywhere tangent to $\mathcal{G}$. In this case, we apply the second Bianchi identity to write

$$\begin{aligned} \widetilde{\nabla}_{\tilde{\xi}_3} \widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_2) = \widetilde{\nabla}_{\tilde{\xi}_1} \widetilde{R}(\tilde{\xi}_3, \tilde{\xi}_2) + \widetilde{\nabla}_{\tilde{\xi}_2} \widetilde{R}(\tilde{\xi}_1, \tilde{\xi}_3) + \widetilde{R}(\widetilde{\nabla}_{\tilde{\xi}_3} \tilde{\xi}_1, \tilde{\xi}_2) + \widetilde{R}(\tilde{\xi}_1, \widetilde{\nabla}_{\tilde{\xi}_3} \tilde{\xi}_2) \\ - \widetilde{R}(\widetilde{\nabla}_{\tilde{\xi}_1} \tilde{\xi}_3, \tilde{\xi}_2) - \widetilde{R}(\tilde{\xi}_1, \widetilde{\nabla}_{\tilde{\xi}_2} \tilde{\xi}_3) + \widetilde{R}(\widetilde{\nabla}_{\tilde{\xi}_1} \tilde{\xi}_2 - \widetilde{\nabla}_{\tilde{\xi}_2} \tilde{\xi}_1, \tilde{\xi}_3). \end{aligned}$$

The first two terms of the right-hand side reduce to the previous case. The next four terms reduce to the induction hypothesis since $\tilde{\xi}_1$ and $\tilde{\xi}_2$ are tangential and at least one occurs as an argument in each term, so the number of transversal vector fields does not increase. The last term also reduces to the induction hypothesis since $\widetilde{\nabla}_{\tilde{\xi}_1}\tilde{\xi}_2 - \widetilde{\nabla}_{\tilde{\xi}_2}\tilde{\xi}_1 = [\tilde{\xi}_1, \tilde{\xi}_2]$ is tangential. □

References

[Alt 2012] J. Alt, "Transitive conformal holonomy groups", *Cent. Eur. J. Math.* **10**:5 (2012), 1710–1720. MR 2949648 Zbl 1278.53045

[Armstrong 2007] S. Armstrong, "Definite signature conformal holonomy: a complete classification", *J. Geom. Phys.* **57**:10 (2007), 2024–2048. MR 2348277 Zbl 05201910

[Armstrong and Leitner 2012] S. Armstrong and F. Leitner, "Decomposable conformal holonomy in Riemannian signature", *Math. Nachr.* **285**:2-3 (2012), 150–163. MR 2881272 Zbl 1246.53064

[Bailey et al. 1994] T. N. Bailey, M. G. Eastwood, and A. R. Gover, "Thomas's structure bundle for conformal, projective and related structures", *Rocky Mountain J. Math.* **24**:4 (1994), 1191–1217. MR 1322223 Zbl 0828.53012

[Čap and Gover 2003] A. Čap and A. R. Gover, "Standard tractors and the conformal ambient metric construction", *Ann. Global Anal. Geom.* **24**:3 (2003), 231–259. MR 1996768 Zbl 1039.53021

[Fefferman and Graham 1985] C. Fefferman and C. R. Graham, "Conformal invariants", pp. 95–116 in *The mathematical heritage of Élie Cartan* (Lyon, 1984), Astérisque (numéro hors série), Société Mathématique de France, Paris, 1985. MR 87g:53060 Zbl 0602.53007

[Fefferman and Graham 2012] C. Fefferman and C. R. Graham, *The ambient metric*, Annals of Mathematics Studies **178**, Princeton University Press, 2012. MR 2858236 Zbl 1243.53004

[Gover and Peterson 2003] A. R. Gover and L. J. Peterson, "Conformally invariant powers of the Laplacian, Q-curvature, and tractor calculus", *Comm. Math. Phys.* **235**:2 (2003), 339–378. MR 1969732 Zbl 1022.58014

[Graham and Willse 2012] C. R. Graham and T. Willse, "Parallel tractor extension and ambient metrics of holonomy split G_2", *J. Differential Geom.* **92**:3 (2012), 463–505. MR 3005060 Zbl 1268.53075

[Kobayashi and Nomizu 1963] S. Kobayashi and K. Nomizu, *Foundations of differential geometry, I*, Interscience Publishers, New York, 1963. Zbl 0119.37502

[Leistner 2006] T. Leistner, "Conformal holonomy of C-spaces, Ricci-flat, and Lorentzian manifolds", *Differential Geom. Appl.* **24**:5 (2006), 458–478. MR 2254049 Zbl 1109.53052

[Leitner 2005] F. Leitner, "Conformal Killing forms with normalisation condition", pp. 279–292 in *Proceedings of the 24th winter school "geometry and physics"* (Srní, Czech Republic, 2005), II **75**, 2005. MR 2152367 Zbl 1101.53040

[Lischewski 2015] A. Lischewski, "Reducible conformal holonomy in any metric signature and application to twistor spinors in low dimension", *Differential Geom. Appl.* **40** (2015), 252–268. MR 3333107 Zbl 1318.53047

[Nijenhuis 1953a] A. Nijenhuis, "On the holonomy groups of linear connections, IA: General properties of affine connections", *Nederl. Akad. Wetensch. Proc. Ser. A.* **15** (1953), 233–240. MR 0063737 Zbl 0051.13203

[Nijenhuis 1953b] A. Nijenhuis, "On the holonomy groups of linear connections, IB: General properties of affine connections", *Nederl. Akad. Wetensch. Proc. Ser. A* **15** (1953), 241–249. MR 0063737 Zbl 0051.13203

[Nijenhuis 1954] A. Nijenhuis, "On the holonomy groups of linear connections, II: Properties of general linear connections", *Nederl. Akad. Wetensch. Proc. Ser. A.* **16** (1954), 17–25. MR 0063738 Zbl 0055.40703

Received April 3, 2015. Revised March 9, 2016.

ANDREAS ČAP
FACULTY OF MATHEMATICS
UNIVERSITY OF VIENNA
OSKAR-MORGENSTERN-PLATZ 1
1090 WIEN
AUSTRIA
andreas.cap@univie.ac.at

A. ROD GOVER
DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF AUCKLAND
PRIVATE BAG 92019
AUCKLAND 1142
NEW ZEALAND
r.gover@auckland.ac.nz

C. ROBIN GRAHAM
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
BOX 354350
SEATTLE, WA 98195-4350
UNITED STATES
robin@math.washington.edu

MATTHIAS HAMMERL
DEPARTMENT OF MATHEMATICS AND INFORMATICS
UNIVERSITY OF GREIFSWALD
WALTHER-RATHENAU-STR. 47
D-17489 GREIFSWALD
GERMANY
matthias.hammerl@uni-greifswald.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 285, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.285.319

NONORIENTABLE LAGRANGIAN COBORDISMS BETWEEN LEGENDRIAN KNOTS

ORSOLA CAPOVILLA-SEARLE AND LISA TRAYNOR

In the symplectization of standard contact 3-space, $\mathbb{R} \times \mathbb{R}^3$, it is known that an orientable Lagrangian cobordism between a Legendrian knot and itself, also known as an orientable Lagrangian endocobordism for the Legendrian knot, must have genus 0. We show that any Legendrian knot has a nonorientable Lagrangian endocobordism, and that the cross-cap genus of such a nonorientable Lagrangian endocobordism must be a positive multiple of 4. The more restrictive exact, nonorientable Lagrangian endocobordisms do not exist for any exactly fillable Legendrian knot but do exist for any stabilized Legendrian knot. Moreover, the relation defined by exact, nonorientable Lagrangian cobordism on the set of stabilized Legendrian knots is symmetric and defines an equivalence relation, a contrast to the nonsymmetric relation defined by orientable Lagrangian cobordisms.

1. Introduction

Smooth cobordisms are a common object of study in topology. Motivated by ideas in symplectic field theory [Eliashberg et al. 2000], Lagrangian cobordisms that are cylindrical over Legendrian submanifolds outside a compact set have been an active area of research interest. Throughout this paper, we will study Lagrangian cobordisms in the symplectization of the standard contact $\mathbb{R}^3$, namely the symplectic manifold $(\mathbb{R} \times \mathbb{R}^3, d(e^t\alpha))$, where $\alpha = dz - y\,dx$, that coincide with the cylinders $\mathbb{R} \times \Lambda_+$ (resp., $\mathbb{R} \times \Lambda_-$) when the $\mathbb{R}$-coordinate is sufficiently positive (resp., negative). Our focus will be on nonorientable Lagrangian cobordisms between Legendrian knots Λ_+ and Λ_- and nonorientable Lagrangian endocobordisms, which are nonorientable Lagrangian cobordisms with $\Lambda_+ = \Lambda_-$.

Smooth endocobordisms in $\mathbb{R}\times\mathbb{R}^3$ without the Lagrangian condition are abundant: for any smooth knot $K \subset \mathbb{R}^3$ and an arbitrary $j \geq 0$, there is a smooth 2-dimensional orientable submanifold M of genus j such that M agrees with the cylinder $\mathbb{R} \times K$ when the $\mathbb{R}$-coordinate lies outside an interval $[T_-, T_+]$; the analogous statement

MSC2010: primary 57R17, 53D42; secondary 57M25.
Keywords: Legendrian knot, Lagrangian cobordism, Lagrangian endocobordism, exact Lagrangian, fillable Legendrian.

holds for nonorientable M and cross-cap genus[1] when $j > 0$. For any Legendrian knot Λ, it is easy to construct an orientable Lagrangian endocobordism of genus 0, namely the trivial Lagrangian cylinder $\mathbb{R} \times \Lambda$. In fact, with the added Lagrangian condition, *orientable* Lagrangian endocobordisms must be concordances:

Theorem [Chantraine 2010]. *Any orientable Lagrangian endocobordism of any Legendrian knot has genus* 0.

Nonorientable Lagrangian endocobordisms also exist and have topological restrictions:

Theorem 1.1. *For an arbitrary Legendrian knot* Λ, *there exists a nonorientable Lagrangian endocobordism for* Λ *of cross-cap genus* g *if and only if* $g \in 4\mathbb{Z}^+$.

Theorem 1.1 is proved in Theorems 3.2 and 3.3. The fact that the cross-cap genus of a nonorientable Lagrangian endocobordism must be a positive multiple of 4 follows from a result of Audin [1988] about the obstruction to the Euler characteristic for closed, Lagrangian submanifolds in $\mathbb{R}^4$. It is easy to construct *immersed* Lagrangian endocobordisms; the existence of the desired embedded endocobordisms follows from Lagrangian surgery, as developed, for example, by Polterovich [1991].

Of special interest are Lagrangian cobordisms that satisfy an additional "exactness" condition. Exactness is known to be quite restrictive: by a foundational result of Gromov [1985], there are no closed, exact Lagrangian submanifolds in $\mathbb{R}^{2n}$ with its standard symplectic structure. The nonclosed trivial Lagrangian cylinder $\mathbb{R} \times \Lambda$ is exact, and Section 2 describes some general methods to construct exact Lagrangian cobordisms. In contrast to Theorem 1.1, there are some Legendrians that do not admit *exact*, nonorientable Lagrangian endocobordisms:

Theorem 1.2. *There does not exist an exact*, *nonorientable Lagrangian endocobordism for any Legendrian knot* Λ *that is exactly orientably or nonorientably fillable.*

A Legendrian knot Λ is exactly fillable if there exists an exact Lagrangian cobordism that is cylindrical over Λ at the positive end and does not intersect $\{T_-\} \times \mathbb{R}^3$ for $T_- \ll 0$; precise definitions can be found in Section 2. Theorem 1.2 is proved in Section 4; it follows from the Seidel isomorphism (Theorem 4.1), which relates the topology of a filling to the linearized contact cohomology of the Legendrian at the positive end. Theorem 1.2 implies that on the set of Legendrian knots in $\mathbb{R}^3$ that are exactly fillable, orientably or not, the relation defined by exact, nonorientable Lagrangian cobordism is *antireflexive* and *antisymmetric*; see Corollary 4.2. Figure 6 gives some particular examples of Legendrians that are exactly fillable and thus do not admit exact, nonorientable Lagrangian endocobordisms. Many of these examples are maximal tb Legendrian representatives of twist or torus knots. In fact,

[1] the number of real projective planes in a connected sum decomposition

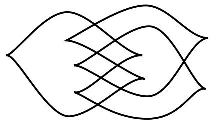


Figure 1. The max tb Legendrian representative of $m(8_{19})$.

using the classification results of Etnyre and Honda [2001] and of Etnyre, Ng, and Vértesi [Etnyre et al. 2013], we show:

Corollary 1.3. *Let K be the smooth knot type of either a twist knot or a positive torus knot or a negative torus knot of the form $T(-p, 2k)$ for p odd and $p > 2k > 0$. Then any maximal* tb *Legendrian representative of K does not have an exact, nonorientable Lagrangian endocobordism.*

However, stabilized Legendrian knots do admit exact, nonorientable Lagrangian endocobordisms: a Legendrian knot is said to be stabilized if, after Legendrian isotopy, a strand contains a zig-zag as shown in Figure 4.

Theorem 1.4. *For any stabilized Legendrian knot Λ and any $k \in \mathbb{Z}^+$, there exists an exact, nonorientable Lagrangian endocobordism for Λ of cross-cap genus $4k$.*

Some Legendrian knots are neither exactly fillable nor stabilized. Thus, a natural question is:

Question 1.5. If a Legendrian knot is not exactly fillable and is not stabilized, does it have an exact, nonorientable Lagrangian endocobordism? In particular, does the unique Legendrian representative of $m(8_{19}) = T(-4, 3)$ with maximal tb whose front projection is shown in Figure 1 have an exact, nonorientable Lagrangian endocobordism?

A description of how the Legendrian knot can be recovered from the front projection is given on page 322. The max tb version of $m(8_{19})$ is not exactly fillable since the upper bound on the tb invariant for all Legendrian representatives of $m(8_{19})$ given by the Kauffman polynomial is not sharp; see Section 6 for more details. In response to Question 1.5, Chantraine, Dimitroglou Rizell, Ghiggini, and Golovko [Chantraine et al. 2015, Corollary 12.3] proved an extension of Theorem 1.2 that shows an exact, nonorientable Lagrangian endocobordism does not exist for an orientable Legendrian that admits an augmentation or, more generally, for an orientable Legendrian whose characteristic algebra admits a finite-dimensional representation. The max tb Legendrian representative of $m(8_{19}) = T(-4, 3)$ does not have an augmentation, but by results of Sivek [2013, Corollary 3.5], the characteristic algebra of this Legendrian does have a 2-dimensional representation. Thus the answer to Question 1.5 is no; see Section 6 for additional questions.

Given the existence of exact, nonorientable Lagrangian endocobordisms for a stabilized Legendrian, it is natural to ask: What Legendrian knots can appear as a "slice" of such an endocobordism? The parallel question for orientable Lagrangian endocobordisms has been studied in [Chantraine 2015; Baldwin and Sivek 2014; Cornwell et al. 2016]. The nonorientable version of this question is closely tied to the question of whether or not nonorientable Lagrangian cobordisms define an equivalence relation on the set of Legendrian knots. By a result of Chantraine [2010], it is known that the relation defined on the set of Legendrian knots by *orientable* Lagrangian cobordism is *not* an equivalence relation since symmetry fails. In fact, the relation defined on the set of *stabilized* Legendrian knots by exact, nonorientable Lagrangian cobordism is symmetric: see Theorem 5.2. In addition, this relation is transitive by "stacking" (Lemma 2.2) and reflexive by Theorem 1.4. Thus we get:

Theorem 1.6. *On the set of stabilized Legendrian knots, the relation defined by exact, nonorientable Lagrangian cobordism is an equivalence relation. Moreover, all stabilized Legendrian knots are equivalent with respect to this relation.*

2. Background

In this section we review Legendrian and Lagrangian submanifolds.

Contact manifolds and Legendrian submanifolds. Below is some basic background on contact manifolds and Legendrian knots. More information can be found, for example, in [Etnyre 2003; 2005].

A *contact manifold* (Y, ξ) is an odd-dimensional manifold together with a contact structure, which consists of a maximally nonintegrable field of tangent hyperplanes. The *standard contact structure* on $\mathbb{R}^3$ is the field $\xi_p = \ker \alpha_0(p)$ for $\alpha_0(x, y, z) = dz - y\,dx$. A *Legendrian link* Λ is a submanifold of $\mathbb{R}^3$ diffeomorphic to a disjoint union of circles such that for all $p \in \Lambda$, we have $T_p\Lambda \subset \xi_p$; if, in addition, Λ is connected, Λ is a *Legendrian knot*. It is common to examine Legendrian links from their xz-projections, known as their *front projections*. A Legendrian link will generically have an immersed front projection with semicubical cusps, no vertical tangents, and no self-tangencies; any such projection can be uniquely lifted to a Legendrian link using $y = dz/dx$.

Two Legendrian links Λ_0 and Λ_1 are *equivalent Legendrian links* if there exists a 1-parameter family of Legendrian links Λ_t, $t \in [0, 1]$, joining Λ_0 and Λ_1. In fact, Legendrian links Λ_0, Λ_1 are equivalent if and only if their front projections are equivalent by planar isotopies that do not introduce vertical tangents and the *Legendrian Reidemeister moves* as shown in Figure 2.

Every knot has a Legendrian representative. In fact, every knot has an infinite number of different Legendrian representatives. For example, Figure 3 shows three different oriented Legendrians that are all topologically the unknot. These unknots can be distinguished by classical Legendrian invariants: the Thurston–Bennequin

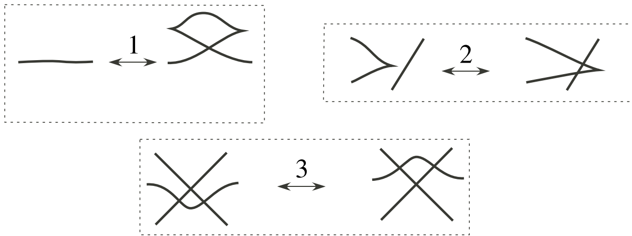


Figure 2. The three Legendrian Reidemeister moves. There is another type-1 move obtained by flipping the planar figure about a horizontal line, and there are three additional type-2 moves obtained by flipping the planar figure about a vertical, a horizontal, and both a vertical and a horizontal line.

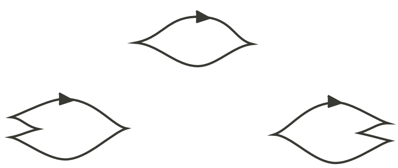


Figure 3. Three different Legendrian unknots: the one with maximal tb invariant of -1 and two others obtained by $\pm$-stabilizations.

number, tb, and the rotation number, r. These invariants can easily be computed from a front projection; see, for example, [Boranda et al. 2013].

The two unknots in the second line of Figure 3 are obtained from the one at the top by stabilization. In general, from an oriented Legendrian Λ, one can obtain oriented Legendrians $S_{\pm}(\Lambda)$: the *positive* (*negative*) *stabilization*, S_+ (S_-), is obtained by replacing a portion of a strand with a strand that contains a down (up) *zig-zag*, as shown in Figure 4. This stabilization procedure will not change the underlying smooth knot type but will decrease the Thurston–Bennequin number by 1; adding an up (down) zig-zag will decrease (increase) the rotation number by 1. It is possible to move a zig-zag to any strand of a Legendrian knot [Fuchs and Tabachnikov 1997]. Bennequin and slice-Bennequin inequalities (see, for example, [Etnyre 2005]) show that for any Legendrian representative Λ of a fixed smooth knot type K, $\mathrm{tb}(\Lambda) + |r(\Lambda)|$ is bounded above. Because of such bounds, the set of oriented Legendrian representatives of a fixed smooth knot type can be visualized by a "mountain range" in the plane where each Legendrian representative Λ is recorded by a vertex at coordinates $(r(\Lambda), \mathrm{tb}(\Lambda))$; two vertices are connected by an edge if the corresponding Legendrians are related by stabilization. Many examples of known and conjectured mountain ranges can be found in the Legendrian knot atlas of Chongchitmate and Ng [2013].

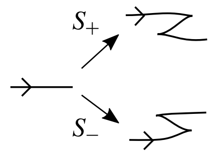


Figure 4. The positive (negative) stabilization of an oriented knot is obtained by introducing a down (an up) zig-zag.

Symplectic manifolds, Lagrangian submanifolds, and Lagrangian cobordisms. We will now discuss some basic concepts in symplectic geometry. Additional background can be found, for example, in [McDuff and Salamon 1998].

A *symplectic manifold* (M, ω) is an even-dimensional manifold together with a 2-form ω that is closed and nondegenerate; when ω is an exact 2-form, $(M, \omega = d\beta)$ is said to be an *exact symplectic manifold.* A basic example of an exact symplectic manifold is $(\mathbb{R}^4, \omega_0 = dx_1 \wedge dy_1 + dx_2 \wedge dy_2)$. The cobordisms constructed in this paper live inside the symplectic manifold that is constructed as the symplectization of $(\mathbb{R}^3, \xi_0 = \ker \alpha_0)$, namely, $\mathbb{R} \times \mathbb{R}^3$ with symplectic form given by $\omega = d(e^t \alpha_0)$. In fact, there is an exact symplectic diffeomorphism between the symplectization $(\mathbb{R} \times \mathbb{R}^3, \omega)$ and the standard $(\mathbb{R}^4, \omega_0)$; see, for example, [Bourgeois et al. 2015].

A *Lagrangian submanifold* L of a 4-dimensional symplectic manifold (M, ω) is a 2-dimensional submanifold such that $\omega|_L = 0$. When M is an exact symplectic manifold, $\omega = d\beta$, $\beta|_L$ is necessarily a closed 1-form; when, in addition, $\beta|_L$ is an exact 1-form, $\beta|_L = df$, then L is said to be an *exact Lagrangian submanifold.* It is easy to verify that the exactness of the Lagrangian does not depend on the choice of β.

Remark. There is a (nonexact) Lagrangian torus in the standard symplectic $\mathbb{R}^4$: this can be seen as the product of two embedded circles in each of the (x_1, y_1) and (x_2, y_2) planes. By classical algebraic topology, it follows that the torus is the only compact, orientable surface that admits a Lagrangian embedding into $\mathbb{R}^4$: a result of Whitney equates a signed count of double points of an immersion to the Euler number of the normal bundle, but for a Lagrangian submanifold, the normal and tangent bundles are isomorphic [Audin et al. 1994].

We now turn our focus to noncompact Lagrangians that are cylindrical over Legendrians.

Definition 2.1. Let Λ_-, Λ_+ be Legendrian links in $\mathbb{R}^3$.

(1) A Lagrangian submanifold without boundary $L \subset \mathbb{R} \times \mathbb{R}^3$ is a *Lagrangian cobordism from* Λ_+ *to* Λ_- if it is of the form

$$L = \big((-\infty, T_-] \times \Lambda_-\big) \cup \bar{L} \cup \big([T_+, +\infty) \times \Lambda_+\big)$$

for some $T_- < T_+$, where $\bar{L} \subset [T_-, T_+] \times \mathbb{R}^3$ is compact with boundary $\partial \bar{L} = (\{T_-\} \times \Lambda_-) \cup (\{T_+\} \times \Lambda_+)$.

(2) A Lagrangian cobordism from Λ_+ to Λ_- is *orientable* (resp., *nonorientable*) if L is orientable (resp., nonorientable).

(3) A Lagrangian cobordism from Λ_+ to Λ_- is *exact* if L is exact, namely $e^t \alpha_0|_L = df|_L$, and the primitive, f, is constant on the cylindrical ends: there exist constants $C_\pm$ such that

$$f|_{L\cap((-\infty,T_-)\times\mathbb{R}^3)} = C_-, \quad f|_{L\cap((T_+,+\infty)\times\mathbb{R}^3)} = C_+.$$

A Legendrian knot Λ is (*exactly*) *fillable* if there exists an (exact) Lagrangian cobordism from $\Lambda_+ = \Lambda$ to $\Lambda_- = \varnothing$.

An important property of Lagrangian cobordisms is that they can be stacked (or composed):

Lemma 2.2 (stacking cobordisms [Ekholm et al. 2012]). *If L_{12} is a Lagrangian cobordism from $\Lambda_+ = \Lambda_1$ to $\Lambda_- = \Lambda_2$, and L_{23} is a Lagrangian cobordism from $\Lambda_+ = \Lambda_2$ to $\Lambda_- = \Lambda_3$, then there exists a Lagrangian cobordism L_{13} from $\Lambda_+ = \Lambda_1$ to $\Lambda_- = \Lambda_3$. Furthermore, if L_{12} and L_{23} are exact, then there exists an exact L_{13}.*

Constructions of exact Lagrangian cobordisms are an active area of research. In this paper, we will use the fact that there exist exact Lagrangian cobordisms between Legendrians related by isotopy and certain surgeries. The existence of exact Lagrangian cobordisms from isotopy is well known; see, for example, [Eliashberg and Gromov 1998; Chantraine 2010; Ekholm et al. 2012; Bourgeois et al. 2015].

Lemma 2.3 (exact cobordisms from isotopy). *Suppose that Λ and Λ' are isotopic Legendrian links. Then there exists an exact, orientable Lagrangian cobordism, in fact concordance, from $\Lambda_+ = \Lambda$ to $\Lambda_- = \Lambda'$.*

Remark. In general, the trace of a Legendrian isotopy is not a Lagrangian cobordism. However it is possible to add a "correction term" so that it will be Lagrangian. More precisely, let $\lambda_t(u) = (x(t,u), y(t,u), z(t,u))$, $t \in \mathbb{R}$, be a Legendrian isotopy such that $\frac{\partial\lambda}{\partial t}(t,u)$ has compact support with $\operatorname{Im}\lambda_t(u) = \Lambda_-$ for $t \leq -T$ and $\operatorname{Im}\lambda_t(u) = \Lambda_+$ for $t \geq T$, and let

$$\eta(t,u) = \alpha_0\Big(\frac{\partial\lambda}{\partial t}(t,u)\Big).$$

Then $\Gamma(t,u) = (t, x(t,u), y(t,u), z(t,u) + \eta(t,u))$ is an exact Lagrangian immersion. If $\eta(t,u)$ is sufficiently small, which can be guaranteed by "slowing down" the isotopy via a t-reparametrization, then $\Gamma(t,u)$ is an exact Lagrangian embedding.

In addition, Legendrians Λ and Λ' that differ by "surgery" can be connected by an exact Lagrangian cobordism. In one of these surgery operations, a Legendrian

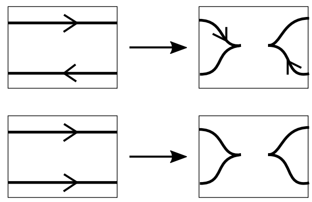


Figure 5. Orientable and nonorientable Legendrian surgeries.

0-tangle, consisting of two strands with no crossings and no cusp points, is replaced with a Legendrian ∞-tangle, consisting of two strands that each have one cusp and no crossings; see Figure 5. When the strands of the 0-tangle are oppositely oriented, this is an *orientable surgery*; otherwise this is a *nonorientable surgery*. There is another surgery operation that shows that the maximal tb Legendrian representative of the unknot, shown at the top of Figure 3, can be filled.

Lemma 2.4 (exact cobordisms from surgery [Ekholm et al. 2012; Dimitroglou Rizell 2014; Bourgeois et al. 2015]).

(1) *Suppose Λ_+ and Λ_- are Legendrian knots, where Λ_- is obtained from Λ_+ by orientable (nonorientable) surgery, as shown in Figure 5. Then there exists an exact, orientable (nonorientable) Lagrangian cobordism from Λ_+ to Λ_-.*

(2) *Suppose Λ_+ is the Legendrian unknot with* tb *equal to the maximum value of -1. Then there exists an exact, orientable Lagrangian filling of Λ_+ by a disk.*

Remark. By Lemmas 2.2, 2.3, and 2.4, to show there exists an exact Lagrangian cobordism from Λ_+ to Λ_-, it suffices to show that there is a string of Legendrian links $(\Lambda_+ = \Lambda_0, \Lambda_1, \dots, \Lambda_n = \Lambda_-)$, where each Λ_{i+1} is obtained from Λ_i by a single surgery, as shown in Figure 5, and Legendrian isotopy. In the case where each surgery is orientable, the exact Lagrangian cobordism will be orientable. If all surgeries are orientable and $\Lambda_\pm$ are both knots, then the length, n, of this string must be even and will agree with twice the genus of the Lagrangian cobordism; for more details, see [Boranda et al. 2013]. If there is at least one nonorientable surgery, the exact Lagrangian cobordism will be nonorientable and the length of the string agrees with the cross-cap genus of the Lagrangian cobordism. To construct an exact Lagrangian filling of Λ_+, it suffices to construct such a string to $\Lambda_- = U$, where U is a trivial link of maximal tb Legendrian unknots.

3. Constructions of nonorientable Lagrangian endocobordisms

We show that *any* Legendrian knot has a nonorientable Lagrangian endocobordism with cross-cap genus an arbitrary multiple of 4. We then show that it is not possible to get any other cross-cap genera.

The strategy to show existence is to first construct an immersed orientable Lagrangian cobordism, and then apply "Lagrangian surgery" to modify it so that it is embedded. The following description of Lagrangian surgery follows Polterovich's construction [1991]; see also [Lalonde and Sikorav 1991].

To describe Lagrangian surgery precisely, we first need to explain the "sign" of a double point. Suppose that x is a point of self-intersection of a generic, immersed, oriented 2-dimensional submanifold L of $\mathbb{R}^4$. Then $\operatorname{sgn}(x) \in \{\pm 1\}$ will denote the *sign of self-intersection of L* at x: let (v_1, v_2) and (w_1, w_2) be positively oriented bases of the transverse tangent spaces at x; then

$$\operatorname{sgn}(x) = +1 \iff (v_1, v_2, w_1, w_2) \text{ is a positively oriented basis of } \mathbb{R}^4,$$

and otherwise $\operatorname{sgn}(x) = -1$.

By constructing a Lagrangian handle in a Darboux chart, it is possible to remove double points of a Lagrangian immersion:

Lemma 3.1 (Lagrangian surgery [Polterovich 1991]). *Let Σ be a 2-dimensional manifold. Suppose $\phi : \Sigma \to \mathbb{R}^4$ is a Lagrangian immersion and $U \subset \mathbb{R}^4$ contains a single transversal double point x of ϕ. Then there exists a 2-dimensional manifold Σ' and a Lagrangian immersion $\phi' : \Sigma' \to \mathbb{R}^4$ such that*

(1) $\operatorname{Im} \phi = \operatorname{Im} \phi'$ *on* $\mathbb{R}^4 - U$,

(2) *ϕ' has no double points in U.*

Furthermore, let $\phi^{-1}(\{x\}) = \{p_1, p_2\} \subset \Sigma$.

(1) *If p_1, p_2 are in disjoint components of Σ, then Σ' is obtained from Σ by a connect sum operation.*

(2) *If p_1, p_2 are in the same component of Σ, then*
   (a) *if Σ is not orientable*, $\Sigma' = \Sigma \# K \; (= \Sigma \# T)$,
   (b) *if Σ is orientable*, $\Sigma' = \Sigma \# T$ *when* $\operatorname{sgn}(x) = +1$, *and* $\Sigma' = \Sigma \# K$ *when* $\operatorname{sgn}(x) = -1$,

   where K denotes the Klein bottle and T denotes the torus.

We now have the necessary background to show the existence of a nonorientable Lagrangian endocobordism for any Legendrian knot:

Theorem 3.2. *For any Legendrian knot Λ and any $k \in \mathbb{Z}^+$, there exists a nonorientable Lagrangian endocobordism for Λ of cross-cap genus $4k$.*

Proof. For an arbitrary Legendrian knot Λ, begin with a cylindrical Lagrangian cobordism, $L = \mathbb{R} \times \Lambda$ in $\mathbb{R} \times \mathbb{R}^3$, which is a space that is symplectically equivalent to the standard $\mathbb{R}^4$. As explained in the remark on page 324, there exists an embedded Lagrangian torus, T, such that $T \cap L = \varnothing$. After a suitable shift and perturbation, we can assume that L and T intersect at exactly two points, x_1 and x_2, where

$\operatorname{sgn}(x_1) = +1$ and $\operatorname{sgn}(x_2) = -1$. By Lemma 2.4, Lagrangian surgery at x_1 results in the connected, oriented, immersed Lagrangian diffeomorphic to $(\mathbb{R} \times S^1) \# T$ with a double point at x_2 of index -1; a second Lagrangian surgery at x_2 results in an embedded, nonorientable Lagrangian cobordism diffeomorphic to $(\mathbb{R} \times S^1) \# T \# K$, and thus of cross-cap genus 4. Stacking these endocobordisms, using Lemma 2.2, produces an embedded, nonorientable Lagrangian cobordism of cross-cap genus $4k$ for any $k \in \mathbb{Z}^+$. □

In fact, the possible cross-cap genera that appear in Theorem 3.2 are all that can exist:

Theorem 3.3. *Any nonorientable Lagrangian endocobordism in $\mathbb{R} \times \mathbb{R}^3$ must have cross-cap genus $4k$ for some $k \in \mathbb{Z}^+$.*

This cross-cap genus restriction is closely tied to Euler characteristic obstructions for *compact*, nonorientable submanifolds that admit Lagrangian embeddings in $(\mathbb{R}^4, \omega_0)$, or equivalently in $(\mathbb{R} \times \mathbb{R}^3, d(e^t\alpha))$:

Lemma 3.4 [Audin 1988]. *Any compact, nonorientable Lagrangian submanifold of $\mathbb{R} \times \mathbb{R}^3$ has an Euler characteristic divisible by* 4.

This result can be seen as an extension, to the nonorientable setting, of a formula of Whitney that relates the number of double points of a smooth immersion to the Euler number of the normal bundle of the immersion; see [Audin 1988; Audin et al. 1994].

Remark. Lemma 3.4 implies that any compact, nonorientable Lagrangian submanifold L in $\mathbb{R} \times \mathbb{R}^3$ has cross-cap genus $2 + 4j$ for some $j \geq 0$. There are explicit constructions of compact, nonorientable Lagrangian submanifolds of cross-cap genus $2 + 4j$ for all $j > 0$ [Givental 1986; Audin 1990]. It has been shown that there is no embedded, Lagrangian Klein bottle ($j = 0$) [Nemirovskiĭ 2009; Shevchishin 2009].

To utilize the cross-cap genus restrictions for compact Lagrangians, we will employ the following lemma, which shows that for any Lagrangian endocobordism, it is possible to construct a compact, nonorientable Lagrangian submanifold into which we can glue the compact portion of a Lagrangian endocobordism.

Lemma 3.5. *For any Legendrian knot $\Lambda \subset \mathbb{R}^3$, any open set $D \subset \mathbb{R}^3$ containing Λ, and any $T \in \mathbb{R}^+$, there exists a compact, nonorientable Lagrangian submanifold L in $\mathbb{R} \times \mathbb{R}^3$ such that*

$$L \cap ([-T, T] \times D) = [-T, T] \times \Lambda.$$

Proof. The strategy will be to construct a Lagrangian torus with double points, thought of as two finite cylinders with top and bottom circles identified, and then apply Lagrangian surgery to remove the double points. As a first step, we construct (nondisjoint) Lagrangian embeddings of two cylinders via Legendrian isotopies (see Lemma 2.3). Namely, start with two disjoint copies of Λ: Λ in D and a translated version $\Lambda' \subset \mathbb{R}^3 - D$. Now, for $t \in [0, U]$, consider Legendrian isotopies Λ_t of Λ and Λ'_t of Λ' that satisfy the following conditions: $\Lambda_t = \Lambda$ for $t \in [0, U]$; $\Lambda'_t = \Lambda'$ for $t \in [0, T]$; and for $t \in [T, U]$, Λ'_t is a Legendrian isotopy of Λ' such that $\Lambda'_t = \Lambda$ for t near U. By repeating an analogous procedure for $t \in [-U, 0]$, we can obtain a smooth immersion of the torus into $[-U, U] \times \mathbb{R}^3$. The arguments used to prove Lemma 2.3 (see the remark on page 325) show that for $U - T$ sufficiently large, the trace of these isotopies can be perturbed to two nondisjoint embedded Lagrangian cylinders that do not have any intersection points in $[-T, T] \times \mathbb{R}^3$, and a direct calculation shows that each double point with $t \in [T, U]$ can be paired up with a double point with $t \in [-U, -T]$ of opposite sign. Then by applying Lagrangian surgery (see Lemma 3.1) at each double point we get a compact, nonorientable Lagrangian submanifold L in $\mathbb{R} \times \mathbb{R}^3$ with the desired properties. □

We are now ready to prove the cross-cap genus restriction for arbitrary nonorientable Lagrangian endocobordisms:

Proof of Theorem 3.3. Let C be a nonorientable Lagrangian endocobordism. Suppose $C \subset \mathbb{R} \times D$ and C agrees with a standard cylinder outside $[-T, T] \times \mathbb{R}^3$. By Lemma 3.5, there is a compact, nonorientable Lagrangian submanifold L in $\mathbb{R} \times \mathbb{R}^3$ such that

$$L \cap ([-T, T] \times D) = [-T, T] \times \Lambda.$$

Let L' be the Lagrangian submanifold obtained by removing the standard cylindrical portion of L in $[-T, T] \times D$ and replacing it with $C \cap ([-T, T] \times \mathbb{R}^3)$. Then L' will be a compact, nonorientable Lagrangian submanifold whose cross-cap genus, $k(L')$, differs from the cross-cap genus of L, $k(L)$, by the cross-cap genus of C, $k(C)$: $k(L') = k(L) + k(C)$. By Lemma 3.4, there exist $j, j' \in \mathbb{Z}^+$ such that $k(L) = 2 + 4j$ and $k(L') = 2 + 4j'$. Thus we find that $k(C)$ must be divisible by 4. □

Remark. For exact Lagrangian cobordisms that are constructed from isotopy and surgery (see Lemmas 2.3 and 2.4) it is possible to show that the cross-cap genus must be a multiple of 4 by an alternate argument that relies on a careful analysis of the possible changes to $\mathrm{tb}(\Lambda)$ under surgery [Capovilla-Searle 2015].

4. Obstructions to exact, nonorientable Lagrangian endocobordisms

We will now focus on *exact*, nonorientable Lagrangian cobordisms. We will prove Theorem 1.2, which states that any Legendrian knot that is exactly fillable does not have an exact, nonorientable Lagrangian endocobordism. The proof of this theorem will involve applying the Seidel isomorphism, which relates the topology of a filling to the linearized Legendrian contact cohomology of the Legendrian at the positive end. We will then apply Theorem 1.2 and give examples of maximal tb Legendrian knots that do not have exact, nonorientable Lagrangian endocobordisms.

We begin with a brief description of Legendrian contact homology; additional background information can be found, for example, in [Etnyre 2005]. Legendrian contact homology is a Floer-type invariant of a Legendrian submanifold that lies within Eliashberg, Givental, and Hofer's symplectic field theory framework [Eliashberg 1998; Eliashberg et al. 2000; Chekanov 2002]. It is possible to associate to a Legendrian submanifold $\Lambda \subset \mathbb{R}^3$ the stable, tame isomorphism class of an associative differential graded algebra (DGA), $(\mathcal{A}(\Lambda), \partial)$. The algebra is freely generated by the so-called Reeb chords of Λ, and is graded using a Maslov index. The differential comes from counting pseudoholomorphic curves in the symplectization of $\mathbb{R}^3$; for our interests, we will always use $\mathbb{Z}/2$ coefficients. *Legendrian contact homology*, namely the homology of $(\mathcal{A}(\Lambda), \partial)$, is a Legendrian isotopy invariant. Legendrian contact homology has been defined for Legendrians in contact manifolds other than $\mathbb{R}^3$; see, for example, [Ekholm et al. 2007; Sabloff 2003].

In general, it is difficult to extract information directly from the Legendrian contact homology. An important computational technique arises from the existence of augmentations of the DGA. An *augmentation* ε of $\mathcal{A}(\Lambda)$ is a differential algebra homomorphism $\varepsilon : (\mathcal{A}(\Lambda), \partial) \to (\mathbb{Z}_2, 0)$; a *graded augmentation* is an augmentation such that ε is supported on elements of degree 0. Not all Legendrians have an augmentation; for any Legendrian Λ, there are only a finite number of augmentations. Given a graded augmentation ε, one can linearize $(\mathcal{A}(\Lambda), \partial)$ to a finite-dimensional differential graded complex $(A(\Lambda), \partial^{\varepsilon})$ and obtain *linearized contact homology*, denoted $\mathrm{LCH}_*(\Lambda, \varepsilon; \mathbb{Z}/2)$, and its *dual linearized contact cohomology*, $\mathrm{LCH}^*(\Lambda, \varepsilon; \mathbb{Z}/2)$. The set of all linearized (co)homology groups with respect to all possible graded augmentations is an invariant of Λ. If the augmentation is ungraded, one can still examine the rank of the nongraded linearized (co)homology, $\dim \mathrm{LCH}(\Lambda, \varepsilon; \mathbb{Z}/2)$, and obtain as an invariant of Λ the set of ranks of this total linearized (co)homology for all possible augmentations. Ungraded linearized (co)homology is not an effective invariant: of the many examples of Legendrian knots in the Legendrian knot atlas of Chongchitmate and Ng [2013] that have the same classical invariants yet can be distinguished through graded linearized homology, none can be distinguished by examining ungraded homology. However, ungraded (co)homology will be useful in arguments below.

Ekholm [2008] has shown that an exact Lagrangian filling, F, of a Legendrian submanifold $\Lambda \subset \mathbb{R}^3$ induces an augmentation ε_F of $(\mathcal{A}(\Lambda), \partial)$. When this filling has Maslov class 0, the augmentation will be graded. Informally, Maslov 0 means that along each loop in the filling, the corresponding loop of Lagrangian tangent planes is trivial in the Lagrangian Grassmannian.

The following result of Seidel will play a central role in showing obstructions to exact, nonorientable Lagrangian endocobordisms. A proof of this result was sketched in [Ekholm 2012] and given in detail in [Dimitroglou Rizell 2016]; a parallel result using generating family homology was given in [Sabloff and Traynor 2013].

Theorem 4.1 (Seidel isomorphism [Ekholm 2012; Dimitroglou Rizell 2016; Ekholm et al. 2012]). *Let $\Lambda \subset \mathbb{R}^{2n+1}$ be an n-dimensional Legendrian submanifold with an exact Lagrangian filling F; let ε_F denote the augmentation induced by the filling. Then*

$$\dim H(F; \mathbb{Z}/2) = \dim \mathrm{LCH}(\Lambda, \varepsilon_F; \mathbb{Z}/2).$$

If the filling F of the n-dimensional Legendrian has Maslov class 0, *then a graded version of the above equality holds*:

$$\dim H_{n-*}(F; \mathbb{Z}/2) = \dim \mathrm{LCH}^*(\Lambda, \varepsilon_F; \mathbb{Z}/2).$$

The ungraded version of the Seidel isomorphism will be used to prove that any Legendrian Λ that is exactly fillable does not have an exact, nonorientable Lagrangian endocobordism:

Proof of Theorem 1.2. For a contradiction, suppose that there is a Legendrian knot Λ that has an exact Lagrangian filling and an exact, nonorientable Lagrangian endocobordism. Then by stacking the endocobordisms (see Lemma 2.2) it follows that Λ has an infinite number of topologically distinct exact, nonorientable Lagrangian fillings. Each of these exact Lagrangian fillings induces an augmentation. Since there are only a finite number of possible augmentations, there must exist two topologically distinct fillings that induce the same augmentation. However, this gives a contradiction to the Seidel isomorphism, Theorem 4.1. □

Theorem 1.2 implies that on the set of Legendrian knots in $\mathbb{R}^3$ that are exactly fillable, orientably or not, the relation defined by exact, nonorientable Lagrangian cobordism is antireflexive. Thus, by stacking (Lemma 2.2) we immediately also see:

Corollary 4.2. *On the set of Legendrian knots in $\mathbb{R}^3$ that are exactly fillable*, *orientably or not*, *the relation $\sim$ defined by exact*, *nonorientable Lagrangian cobordism is antisymmetric*: $\Lambda_1 \sim \Lambda_2 \implies \Lambda_2 \nsim \Lambda_1$.

We now apply Theorem 1.2 to give examples of Legendrians that do not have exact, nonorientable Lagrangian endocobordisms. Hayden and Sabloff [2015] showed that every positive knot type has a Legendrian representative that has an exact, orientable Lagrangian filling. Combining this with Theorem 1.2 immediately gives the next result.

Corollary 4.3 [Hayden and Sabloff 2015]. *Each positive knot has a Legendrian representative that does not have an exact*, *nonorientable Lagrangian endocobordism.*

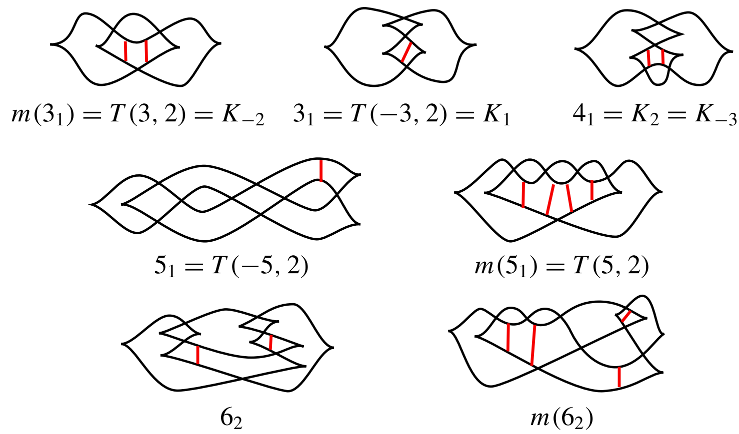


Figure 6. Examples of max tb Legendrians that do not have exact, nonorientable Lagrangian endocobordisms. The red lines indicate points for surgeries.

There is work in progress to show that every $+$-adequate knot has a Legendrian representative with an exact filling (J. M. Sabloff, private communication).

Many maximal tb representatives of low crossing knots have fillings, orientable or not. Figure 6 illustrates some Legendrians that can be verified to have exact, Lagrangian fillings; see the remark on page 326. Many of the examples in Figure 6 are Legendrian representatives of twist knots, K_m, or torus knots, $T(p, q)$. Using Theorem 1.2 together with classification results of Etnyre and Honda [2001] and of Etnyre, Ng, and Vértesi [Etnyre et al. 2013], we show that *all* maximal tb representatives of twist knots, positive torus knots, and negative torus knots of the form $T(-p, 2k)$, $p > 2k > 0$, do not have exact, nonorientable Lagrangian endocobordisms:

Proof of Corollary 1.3. By Theorem 1.2, to show the nonexistence of an exact, nonorientable Lagrangian endocobordism, it suffices to show the existence of an exact Lagrangian filling.

First consider the case where Λ is a maximal tb representative of a twist knot, whose form is shown in Figure 7. Etnyre, Ng, and Vértesi [Etnyre et al. 2013] have

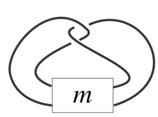


Figure 7. The smooth twist knot K_m; the box contains m right-handed half-twists if $m \geq 0$, and $|m|$ left-handed twists if $m < 0$. Notice that K_0 and K_{-1} are unknots.

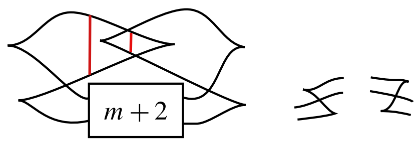


Figure 8. Any maximal tb Legendrian representative of a negative twist knot K_m, with $m \leq -2$, is Legendrian isotopic to a Legendrian of the form on the left with the box containing $|m+2|$ half-twists, where each half-twist has form S (middle) or form Z (right). Two surgeries produces a max tb Legendrian unknot.

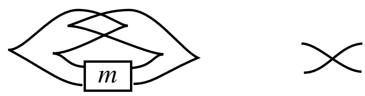


Figure 9. Any maximal tb Legendrian representative of a positive twist knot K_m, with $m \geq 1$, is Legendrian isotopic to a Legendrian in the form at left, where the box contains m half-twists of form X, right.

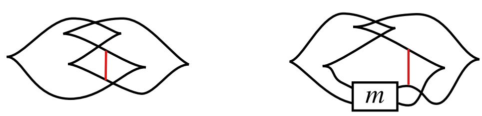


Figure 10. An inductive argument shows that every max tb representative of a positive twist knot has an exact Lagrangian filling.

classified all Legendrian twist knots: every maximal tb Legendrian representative of K_m, for $m \leq -2$, is Legendrian isotopic to one of the form in Figure 8, and every maximal tb Legendrian representative of K_m, for $m \geq 1$, is Legendrian isotopic to one of the form in Figure 9. For a max tb representative of a negative twist knot, Figure 8 illustrates the two surgeries that show the existence of an exact Lagrangian filling. For a max tb Legendrian representative of a positive twist knot, the existence of an exact filling can be shown by an induction argument: Figure 10 (left) indicates the surgery point when $m = 1$; for all $m \geq 1$, a maximal tb representative of K_{m+1} can be reduced to a maximal tb representative of K_m by one surgery, as indicated in Figure 10 (right).

Next consider maximal tb Legendrian representatives of a torus knot, a knot that can be smoothly isotoped so that it lies on the surface of an unknotted torus in $\mathbb{R}^3$. Every torus knot can be specified by a pair (p, q) of coprime integers: we will use the convention that the (p, q)-torus knot, $T(p, q)$, winds p times around a meridional curve of the torus and q times in the longitudinal direction. In fact,

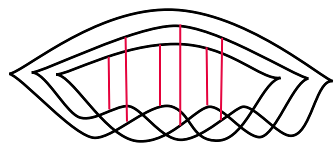


Figure 11. Surgeries that result in an exact filling of the maximal tb representative of the positive torus knot $T(5, 3)$.

$T(p, q)$ is equivalent to $T(q, p)$ and to $T(-p, -q)$. We will always assume that $|p| > q \geq 2$, since we are interested in nontrivial torus knots.

Etnyre and Honda [2001] showed there is a unique maximal tb representative of a positive torus knot $T(p, q)$ with $p > 0$. The surgeries used in [Boranda et al. 2013, Theorem 4.2] show that this maximal representative is exactly fillable. Figure 11 illustrates the orientable surgeries for the (5, 3)-torus knot; in this sequence of surgeries, one begins with surgeries on the innermost strands, and then performs a Legendrian isotopy so that it is possible to do a surgery on the next set of innermost strands.

Lastly consider the case where Λ is topologically a negative torus knot, $T(-p, 2k)$ with $p > 2k > 0$. In this case, Etnyre and Honda have shown that the number of different maximal tb Legendrian representations depends on the divisibility of p by $2k$: if $|p| = m2k + e$, $0 < e < 2k$, there are m nonoriented Legendrian representatives of $T(-p, 2k)$ with maximal tb. These different representatives with maximal tb are obtained by writing $m = 1 + n_1 + n_2$, where $n_1, n_2 \geq 0$, and then $\Lambda_{(n_1,n_2)}$ is constructed using the form shown in Figure 12 with n_1 and n_2 copies of the tangle B inserted as indicated; this figure also shows k surgeries that guarantee the existence of an exact Lagrangian filling. $\square$

Some comments on obstructions to exact fillings are discussed in Section 6.

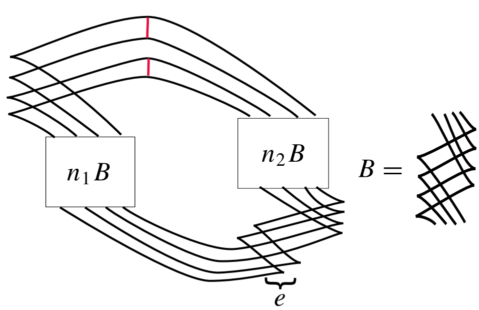


Figure 12. The general form of a maximal tb representative of a negative torus knot $T(-p, 2k)$, with $p > 2k > 0$, with $k = 2$ and $|p| = (1 + n_1 + n_2)2k + e$; the indicated k surgeries produce a Legendrian trivial link of maximal tb unknots.

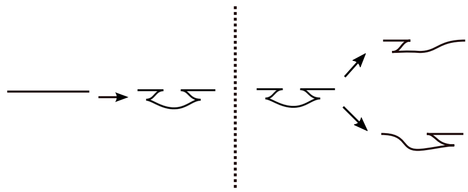


Figure 13. It is possible to construct exact, nonorientable Lagrangian cobordisms between $\Lambda_+ = \Lambda$ and $S_-S_+(\Lambda)$, left, and between $\Lambda_+ = S_-S_+(\Lambda)$ and $\Lambda_- = S_+(\Lambda)$ or $\Lambda_- = S_-(\Lambda)$, right.

5. Constructions of exact, nonorientable Lagrangian cobordisms

We will construct an exact, nonorientable Lagrangian endocobordism of cross-cap genus 4 for any stabilized Legendrian knot, and a nonorientable Lagrangian cobordism between any two stabilized Legendrian knots. All these exact Lagrangian cobordisms are constructed through isotopy and surgery; see the remark on page 326.

Central to these constructions will be the following lemma, which says that one can always introduce a pair of "oppositely oriented" zig-zags to Λ_+, and if one has a pair of oppositely oriented zig-zags in Λ_+, then one can remove either element of this pair; see Figure 13. One needs to be careful when discussing orientations for the ends of a nonorientable Lagrangian cobordism: given an orientation on Λ_+, there is no canonical orientation for Λ_-. However, although an orientation is needed on Λ to distinguish between $S_+(\Lambda)$ and $S_-(\Lambda)$, $S_-S_+(\Lambda)$ is a well-defined operation on unoriented knots.

Lemma 5.1. *Let Λ be a Legendrian knot. Then there exists an exact, nonorientable Lagrangian cobordism*

(1) *of cross-cap genus* 2 *between* $\Lambda_+ = \Lambda$ *and* $\Lambda_- = S_-S_+(\Lambda)$,

(2) *of cross-cap genus* 1 *between* $\Lambda_+ = S_-S_+(\Lambda)$ *and* $\Lambda_- = S_+(\Lambda)$ *or* $\Lambda_- = S_-(\Lambda)$.

Proof. The strategy will be to construct the desired exact, nonorientable Lagrangian cobordism via Legendrian isotopy and surgeries that are performed locally, near the site of the stabilizations. Figure 14 illustrates the isotopy and surgeries, the second of which is nonorientable, that imply the existence of a cross-cap genus 2 Lagrangian cobordism between $\Lambda_+ = \Lambda$ and $\Lambda_- = S_-S_+(\Lambda)$. Figure 15 illustrates the isotopy and surgery that imply the existence of a cross-cap genus 1 Lagrangian cobordism between $\Lambda_+ = S_-S_+(\Lambda)$ and $\Lambda_- = S_+(\Lambda)$, when the original strand is oriented from right to left, or $\Lambda_- = S_-(\Lambda)$, when the original strand is oriented from left to right. ☐

Exact, nonorientable Lagrangian endocobordisms. In Theorem 1.2, it was shown that Legendrians that are exactly fillable do not have exact, nonorientable Lagrangian

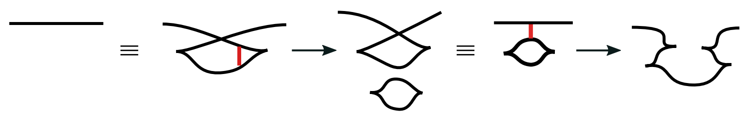


Figure 14. By applying an orientable and a nonorientable surgery, any strand can have a pair of oppositely oriented zig-zags introduced.

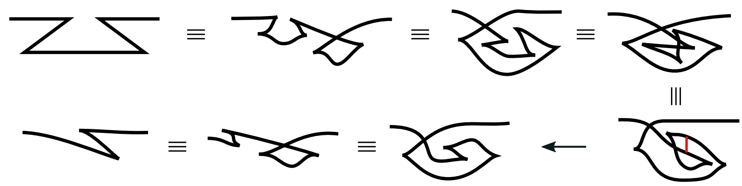


Figure 15. In the presence of oppositely oriented zig-zags, via one nonorientable surgery, one of the zig-zags can be removed.

endocobordisms. However exact, nonorientable Lagrangian endocobordisms do exist for stabilized knots:

Proof of Theorem 1.4. First consider the case where Λ is the negative stabilization of a Legendrian: $\Lambda = S_-(\hat{\Lambda})$. Then by Lemma 5.1, there exists an exact, nonorientable Lagrangian cobordism

(1) of cross-cap genus 2 between Λ and $S_-S_+(\Lambda)$,

(2) of cross-cap genus 1 between $S_-S_+(\Lambda)$ and $S_+(\Lambda)$,

(3) of cross-cap genus 1 between $S_+(\Lambda) = S_+(S_-(\hat{\Lambda}))$ and $S_-(\hat{\Lambda}) = \Lambda$.

Stacking these cobordisms results in an exact, nonorientable Lagrangian endocobordism of cross-cap genus 4. Additional stacking results in arbitrary multiples of cross-cap genus 4.

An analogous argument proves the case where Λ is the positive stabilization of a Legendrian: $\Lambda = S_+(\hat{\Lambda})$. □

Exact, nonorientable Lagrangian cobordisms between stabilized Legendrians. Given that every stabilized Legendrian knot has a nonorientable Lagrangian endocobordism, a natural question is: What Legendrian knots can appear as a "slice" of such an endocobordism? We show that *any* stabilized Legendrian knot can appear as such a slice.

Theorem 5.2. *For smooth knot types K, K', let Λ be any Legendrian representative of K and let Λ' be a stabilized Legendrian representative of K'. Then there exists an exact, nonorientable Lagrangian cobordism between $\Lambda_+ = \Lambda$ and $\Lambda_- = \Lambda'$.*

Before moving to the proof of Theorem 5.2, we show that nonorientable Lagrangian cobordisms define an equivalence relation on the set of stabilized Legendrian knots:

Proof of Theorem 1.6. Let $\mathcal{L}^s$ denote the set of all stabilized Legendrian knots of any smooth knot type. Define the relation $\sim$ on $\mathcal{L}^s$ by $\Lambda_1 \sim \Lambda_2$ if there exists an exact, nonorientable Lagrangian cobordism from $\Lambda_+ = \Lambda_1$ to $\Lambda_- = \Lambda_2$. Reflexivity of $\sim$ follows from Theorem 1.4, symmetry follows from Theorem 5.2, and transitivity follows from Lemma 2.2. Thus $\sim$ defines an equivalence relation. Moreover, by Theorem 5.2, we see that with respect to this equivalence relation, there is only one equivalence class. □

To prove Theorem 5.2, it will be useful to first show that there is an exact, nonorientable Lagrangian cobordism between any two stabilized Legendrians of a fixed knot type:

Proposition 5.3. *Let K be any smooth knot type, and let Λ, Λ' be Legendrian representatives of K, where Λ' is stabilized. Then there exists an exact, nonorientable Lagrangian cobordism between $\Lambda_+ = \Lambda$ and $\Lambda_- = \Lambda'$.*

Proof. Fix a smooth knot type K, and let Λ_1, Λ_2 be Legendrian representatives, where Λ_2 is stabilized. By results of Fuchs and Tabachnikov [1997], we know that there exist r_1, ℓ_1, r_2, ℓ_2 such that

$$S_-^{\ell_1} S_+^{r_1}(\Lambda_1) = S_-^{\ell_2} S_+^{r_2}(\Lambda_2).$$

By applying additional positive stabilizations, if needed, we can assume $r_1 > \ell_1$.

Consider the case where Λ_2 is the negative stabilization of some Legendrian: $\Lambda_2 = S_-(\hat{\Lambda}_2)$. By applications of Lemma 5.1, there exists an exact, nonorientable Lagrangian cobordism between

(1) Λ_1 and $S_-^{r_1} S_+^{r_1}(\Lambda_1)$,

(2) $S_-^{r_1} S_+^{r_1}(\Lambda_1)$ and $S_-^{\ell_1} S_+^{r_1}(\Lambda_1)$, and thus between $S_-^{r_1} S_+^{r_1}(\Lambda_1)$ and $S_-^{\ell_2} S_+^{r_2}(\Lambda_2)$,

(3) $S_-^{\ell_2} S_+^{r_2}(\Lambda_2)$ and $S_+^{r_2}(\Lambda_2)$,

(4) $S_+^{r_2}(\Lambda_2) = S_+^{r_2}(S_-(\hat{\Lambda}_2))$ and $S_-(\hat{\Lambda}_2) = \Lambda_2$.

By stacking these cobordisms (Lemma 2.2), we have our desired exact, nonorientable Lagrangian cobordism between Λ_1 and Λ_2. An analogous argument proves the case where Λ_2 is the positive stabilization of some Legendrian. □

Proof of Theorem 5.2. The strategy here is to first show that one can construct an exact, nonorientable Lagrangian cobordism between Λ and a stabilized Legendrian unknot Λ_0. Similarly, it is possible to construct an exact, nonorientable Lagrangian cobordism between Λ' and a stabilized Legendrian unknot Λ_0'; we will show it

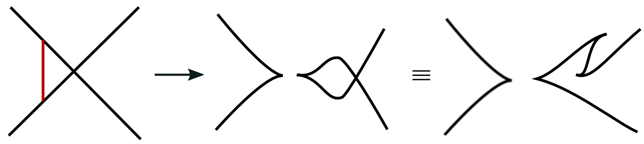


Figure 16. For any Legendrian knot Λ, perform a surgery near each crossing to get a set of disjoint Legendrian unknots.

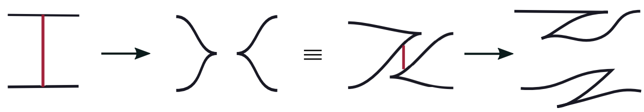


Figure 17. Surgeries used to convert to a link of Legendrian unknots can be "undone", at the cost of additional stabilizations.

is possible to "reverse" this sequence of surgeries and construct an exact, nonorientable Lagrangian cobordism between Λ_0' and $\widetilde{\Lambda}'$, which is a stabilization of Λ'. By Proposition 5.3, there exists an exact, nonorientable Lagrangian cobordism between Λ_0 and Λ_0' and between $\widetilde{\Lambda}'$ and Λ'. Thus we will have the desired exact, nonorientable Lagrangian cobordism between Λ and Λ' by stacking the cobordisms between Λ and Λ_0, between Λ_0 and Λ_0', between Λ_0' and $\widetilde{\Lambda}'$, and between $\widetilde{\Lambda}'$ and Λ'.

We first show how it is possible to construct an exact, nonorientable Lagrangian cobordism from Λ to a Legendrian unknot; cf. [Boranda et al. 2013]. Let Λ be an arbitrary Legendrian knot. We can assume that Λ has at least one positive crossing by, if necessary, applying a Legendrian Reidemeister 1 move. As shown in Figure 16, performing an orientable or nonorientable surgery near a crossing produces a crossing that can be removed through Legendrian Reidemeister moves. Perform such a surgery on every crossing in Λ until you have obtained k disjoint stabilized Legendrian unknots; since Λ has at least one positive crossing, we have performed at least one nonorientable surgery. Align the k Legendrian unknots vertically and perform surgeries so that we obtain a single stabilized Legendrian unknot Λ_0. In this way, we have constructed an exact, nonorientable Lagrangian cobordism between Λ and Λ_0.

A similar procedure can be used to construct a sequence of surgeries from Λ' to another Legendrian unknot Λ_0'; now we show it is possible to "reverse" this procedure and construct a sequence of surgeries from Λ_0' to $\widetilde{\Lambda}'$, a Legendrian obtained by applying stabilizations to Λ'. Figure 17 illustrates how every surgery that was used to get to a Legendrian unknot can be undone at the cost of adding additional zig-zags into the original strands. Figure 18 illustrates this procedure with an example.

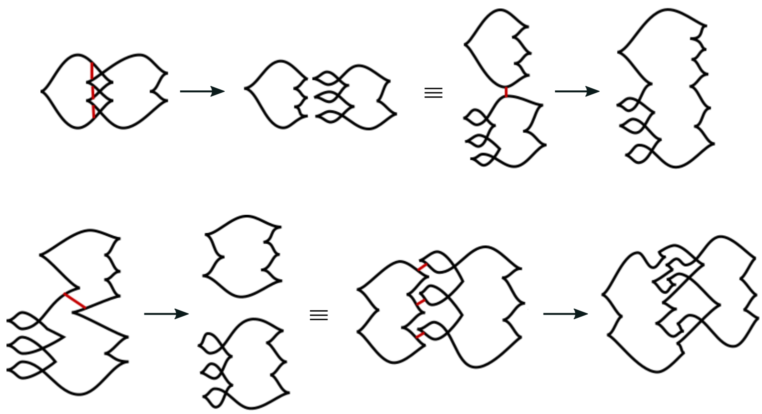


Figure 18. Top: surgeries that give rise to an exact, nonorientable Lagrangian cobordism from the max tb version of 3_1 to a stabilized unknot. Bottom: surgeries that give rise to an exact, nonorientable Lagrangian cobordism from the stabilized unknot to a stabilized representative of 3_1.

As outlined at the beginning of the proof, these constructions prove the existence of an exact Lagrangian cobordism from $\Lambda_+ = \Lambda$ to $\Lambda_- = \Lambda'$. $\square$

6. Additional questions

We end with a brief discussion of some additional questions.

From results above, we know that exactly fillable Legendrian knots do not admit exact, nonorientable Lagrangian endocobordisms, while stabilized Legendrian knots do. There are examples of Legendrian knots that are neither exactly fillable nor stabilized. As mentioned above, Ekholm [2008] has shown that if Λ is exactly fillable, then there exists an ungraded augmentation of $\mathcal{A}(\Lambda)$. By work of Sabloff [2005] and, independently, Fuchs and Ishkhanov [2004], we then know that there exists an ungraded ruling of Λ. (Definitions of graded and ungraded rulings can be found, for example, in [Kálmán 2008].) Then it follows by [Rutherford 2006] that the Kauffman bound on the maximal tb value for all Legendrian representatives of the smooth knot type of Λ is sharp. Thus, if the Kauffman bound is not sharp for the smooth knot type K, then no Legendrian representative of K is exactly fillable. So a natural question is:

Question 6.1. If Λ is a maximal tb representative of a knot type K for which the upper bound on tb for all Legendrian representatives given by the Kauffman polynomial is not sharp, does Λ have an exact, nonorientable Lagrangian endocobordism?

The Legendrian representative of $m(8_{19})$ shown in Figure 1 satisfies the hypotheses in Question 6.1; the Kauffman bound is known to be sharp for all knots with 10 or fewer crossings except $m(8_{19})$, $m(9_{42})$, $m(10_{124})$, $m(10_{128})$, $m(10_{132})$, and $m(10_{136})$ [Ng 2001; 2005; Rutherford 2006]. As mentioned in the Introduction, [Chantraine et al. 2015] contains results which imply that the answer to Question 6.1 is no. However, this now spawns new questions. For example, consider the max tb Legendrian representative of $m(10_{132})$ given as K_2 in [Sivek 2013, Figure 2]. The Legendrian K_2 is not stabilized, does not have an augmentation (and thus is not exactly fillable), and does not have a finite-dimensional representation. Does K_2 have an exact, nonorientable Lagrangian endocobordism?

There are also examples of Legendrians with nonmaximal tb that are not stabilized. For example, $m(10_{161})$ is a knot type where the unique maximal tb representative has a filling. However, there is a Legendrian representative with nonmaximal tb that does not arise as a stabilization. As shown in [Shonkwiler and Vela-Vick 2011, Figure 1], this nonmaximal tb, nonstabilized Legendrian does have an ungraded ruling, and the characteristic algebra of K_2 does not have a finite-dimensional representation [Sivek 2013].

Question 6.2. Does the nonstabilized, nonmaximal tb Legendrian representative of $m(10_{161})$ have an exact, nonorientable Lagrangian endocobordism?

Additional examples of nonstabilized and nonmaximal tb representatives can be found in the Legendrian knot atlas of Chongchitmate and Ng [2013].

There are additional questions that arise from the constructions of fillings. For example, it is known by results of Chantraine [2010] that orientable fillings realize the smooth 4-ball genus. In Figure 6, examples are given of nonorientable Lagrangian fillings of maximal tb representatives of 6_2 and $m(6_2)$ of cross-cap genus 2 and 4, respectively: the smooth 4-dimensional cross-cap number of both 6_2 and $m(6_2)$ is 1.

Question 6.3. Does there exist a nonorientable Lagrangian filling of these Legendrian representatives of 6_2 and $m(6_2)$ of cross-cap genus 1?

Acknowledgements

We thank Baptiste Chantraine, Richard Hind, and Josh Sabloff for stimulating discussions. We also thank Georgios Dimitroglou Rizell and Tobias Ekholm for helpful comments. We are also grateful to the referee for numerous comments and suggestions that made the exposition better. Both authors thank the Andrew W. Mellon Foundation for supporting the first author with a Mellon Mays Undergraduate Fellowship; this paper grew out of her thesis project [Capovilla-Searle 2015]. Traynor gratefully acknowledges the hospitality of the Institute for Advanced Study in Princeton and support at IAS from The Fund for Mathematics during a portion of this work.

References

[Audin 1988] M. Audin, "Fibrés normaux d'immersions en dimension double, points doubles d'immersions lagragiennes et plongements totalement réels", *Comment. Math. Helv.* **63**:4 (1988), 593–623. MR 966952 Zbl 0666.57024

[Audin 1990] M. Audin, "Quelques remarques sur les surfaces lagrangiennes de Givental", *J. Geom. Phys.* **7**:4 (1990), 583–598. MR 1131914 Zbl 0728.57020

[Audin et al. 1994] M. Audin, F. Lalonde, and L. Polterovich, "Symplectic rigidity: Lagrangian submanifolds", pp. 271–321 in *Holomorphic curves in symplectic geometry*, edited by M. Audin and J. Lafontaine, Progr. Math. **117**, Birkhäuser, Basel, 1994. MR 1274934 Zbl 0802.53001

[Baldwin and Sivek 2014] J. A. Baldwin and S. Sivek, "Invariants of Legendrian and transverse knots in monopole knot homology", preprint, 2014. arXiv 1405.3275

[Boranda et al. 2013] B. Boranda, L. Traynor, and S. Yan, "The surgery unknotting number of Legendrian links", *Involve* **6**:3 (2013), 273–299. MR 3101761 Zbl 06227497

[Bourgeois et al. 2015] F. Bourgeois, J. M. Sabloff, and L. Traynor, "Lagrangian cobordisms via generating families: construction and geography", *Algebr. Geom. Topol.* **15**:4 (2015), 2439–2477. MR 3402346 Zbl 1330.57037

[Capovilla-Searle 2015] O. Capovilla-Searle, *Surgery strings and loops of Legendrian knots*, bachelor's thesis, Bryn Mawr College, 2015.

[Chantraine 2010] B. Chantraine, "Lagrangian concordance of Legendrian knots", *Algebr. Geom. Topol.* **10**:1 (2010), 63–85. MR 2580429 Zbl 1203.57010

[Chantraine 2015] B. Chantraine, "Lagrangian concordance is not a symmetric relation", *Quantum Topol.* **6**:3 (2015), 451–474. MR 3392961 Zbl 1337.57056

[Chantraine et al. 2015] B. Chantraine, G. Dimitroglou Rizell, P. Ghiggini, and R. Golovko, "Floer theory for Lagrangian cobordisms", preprint, 2015. arXiv 1511.09471v1

[Chekanov 2002] Y. Chekanov, "Differential algebra of Legendrian links", *Invent. Math.* **150**:3 (2002), 441–483. MR 1946550 Zbl 1029.57011

[Chongchitmate and Ng 2013] W. Chongchitmate and L. Ng, "An atlas of Legendrian knots", *Exp. Math.* **22**:1 (2013), 26–37. MR 3038780 Zbl 1267.57004

[Cornwell et al. 2016] C. Cornwell, L. Ng, and S. Sivek, "Obstructions to Lagrangian concordance", *Algebr. Geom. Topol.* **16**:2 (2016), 797–824. MR 3493408 Zbl 06577067

[Dimitroglou Rizell 2014] G. Dimitroglou Rizell, "Legendrian ambient surgery and Legendrian contact homology", preprint, 2014. arXiv 1205.5544v5

[Dimitroglou Rizell 2016] G. Dimitroglou Rizell, "Lifting pseudo-holomorphic polygons to the symplectisation of $P \times \mathbb{R}$ and applications", *Quantum Topol.* **7**:1 (2016), 29–105. MR 3459958 Zbl 06558091

[Ekholm 2008] T. Ekholm, "Rational symplectic field theory over $\mathbb{Z}_2$ for exact Lagrangian cobordisms", *J. Eur. Math. Soc.* (*JEMS*) **10**:3 (2008), 641–704. MR 2421157 Zbl 1154.57020

[Ekholm 2012] T. Ekholm, "Rational SFT, linearized Legendrian contact homology, and Lagrangian Floer cohomology", pp. 109–145 in *Perspectives in analysis, geometry, and topology*, edited by I. Itenberg et al., Progr. Math. **296**, Birkhäuser/Springer, New York, 2012. MR 2884034 Zbl 1254.57024

[Ekholm et al. 2007] T. Ekholm, J. Etnyre, and M. Sullivan, "Legendrian contact homology in $P \times \mathbb{R}$", *Trans. Amer. Math. Soc.* **359**:7 (2007), 3301–3335. MR 2299457 Zbl 1119.53051

[Ekholm et al. 2012] T. Ekholm, K. Honda, and T. Kálmán, "Legendrian knots and exact Lagrangian cobordisms", preprint, 2012. arXiv 1212.1519

[Eliashberg 1998] Y. Eliashberg, "Invariants in contact topology", *Doc. Math.* Extra Vol. II (1998), 327–338. MR 1648083 Zbl 0913.53010

[Eliashberg and Gromov 1998] Y. Eliashberg and M. Gromov, "Lagrangian intersection theory: finite-dimensional approach", pp. 27–118 in *Geometry of differential equations*, edited by A. Khovanskiĭ et al., Amer. Math. Soc. Transl. Ser. 2 **186**, Amer. Math. Soc., Providence, RI, 1998. MR 1732407 Zbl 0919.58015

[Eliashberg et al. 2000] Y. Eliashberg, A. Givental, and H. Hofer, "Introduction to symplectic field theory", *Geom. Funct. Anal.* Special Vol., Part II (2000), 560–673. MR 1826267 Zbl 0989.81114

[Etnyre 2003] J. B. Etnyre, "Introductory lectures on contact geometry", pp. 81–107 in *Topology and geometry of manifolds* (Athens, GA, 2001), edited by G. Matić and C. McCrory, Proc. Sympos. Pure Math. **71**, Amer. Math. Soc., Providence, RI, 2003. MR 2024631 Zbl 1045.57012

[Etnyre 2005] J. B. Etnyre, "Legendrian and transversal knots", pp. 105–185 in *Handbook of knot theory*, edited by W. Menasco and M. Thistlethwaite, Elsevier, Amsterdam, 2005. MR 2179261 Zbl 1095.57006

[Etnyre and Honda 2001] J. B. Etnyre and K. Honda, "Knots and contact geometry, I: Torus knots and the figure eight knot", *J. Symplectic Geom.* **1**:1 (2001), 63–120. MR 1959579 Zbl 1037.57021

[Etnyre et al. 2013] J. B. Etnyre, L. L. Ng, and V. Vértesi, "Legendrian and transverse twist knots", *J. Eur. Math. Soc.* (*JEMS*) **15**:3 (2013), 969–995. MR 3085098 Zbl 1271.57060

[Fuchs and Ishkhanov 2004] D. Fuchs and T. Ishkhanov, "Invariants of Legendrian knots and decompositions of front diagrams", *Mosc. Math. J.* **4**:3 (2004), 707–717, 783. MR 2119145 Zbl 1073.53106

[Fuchs and Tabachnikov 1997] D. Fuchs and S. Tabachnikov, "Invariants of Legendrian and transverse knots in the standard contact space", *Topology* **36**:5 (1997), 1025–1053. MR 1445553 Zbl 0904.57006

[Givental 1986] A. B. Givental', "Lagrangian imbeddings of surfaces and the open Whitney umbrella", *Funktsional. Anal. i Prilozhen.* **20**:3 (1986), 35–41, 96. In Russian; translated in *Funct. Anal. Appl.* **20**:3 (1986), 197–203. MR 868559 Zbl 0621.58025

[Gromov 1985] M. Gromov, "Pseudoholomorphic curves in symplectic manifolds", *Invent. Math.* **82**:2 (1985), 307–347. MR 809718 Zbl 0592.53025

[Hayden and Sabloff 2015] K. Hayden and J. M. Sabloff, "Positive knots and Lagrangian fillability", *Proc. Amer. Math. Soc.* **143**:4 (2015), 1813–1821. MR 3314092 Zbl 1311.57033

[Kálmán 2008] T. Kálmán, "Rulings of Legendrian knots as spanning surfaces", *Pacific J. Math.* **237**:2 (2008), 287–297. MR 2421123 Zbl 1153.53058

[Lalonde and Sikorav 1991] F. Lalonde and J.-C. Sikorav, "Sous-variétés lagrangiennes et lagrangiennes exactes des fibrés cotangents", *Comment. Math. Helv.* **66**:1 (1991), 18–33. MR 1090163 Zbl 0759.53022

[McDuff and Salamon 1998] D. McDuff and D. Salamon, *Introduction to symplectic topology*, 2nd ed., Clarendon/Oxford University Press, New York, 1998. MR 1698616 Zbl 1066.53137

[Nemirovskiĭ 2009] S. Y. Nemirovskiĭ, "The homology class of a Lagrangian Klein bottle", *Izv. Ross. Akad. Nauk Ser. Mat.* **73**:4 (2009), 37–48. In Russian; translated in *Izv. Math.* **73**:4 (2009), 689–698. MR 2583965 Zbl 1191.57019

[Ng 2001] L. L. Ng, “Maximal Thurston–Bennequin number of two-bridge links”, *Algebr. Geom. Topol.* **1** (2001), 427–434. MR 1852765 Zbl 1056.57010

[Ng 2005] L. Ng, “A Legendrian Thurston–Bennequin bound from Khovanov homology”, *Algebr. Geom. Topol.* **5** (2005), 1637–1653. MR 2186113 Zbl 1091.57009

[Polterovich 1991] L. Polterovich, “The surgery of Lagrange submanifolds”, *Geom. Funct. Anal.* **1**:2 (1991), 198–210. MR 1097259 Zbl 0754.57027

[Rutherford 2006] D. Rutherford, “Thurston–Bennequin number, Kauffman polynomial, and ruling invariants of a Legendrian link: the Fuchs conjecture and beyond”, *Int. Math. Res. Not.* **2006** (2006), Art. ID 78591. MR 2219227 Zbl 1106.57012

[Sabloff 2003] J. M. Sabloff, “Invariants of Legendrian knots in circle bundles”, *Commun. Contemp. Math.* **5**:4 (2003), 569–627. MR 2003211 Zbl 1046.57013

[Sabloff 2005] J. M. Sabloff, “Augmentations and rulings of Legendrian knots”, *Int. Math. Res. Not.* **2005**:19 (2005), 1157–1180. MR 2147057 Zbl 1082.57020

[Sabloff and Traynor 2013] J. M. Sabloff and L. Traynor, “Obstructions to Lagrangian cobordisms between Legendrians via generating families”, *Algebr. Geom. Topol.* **13**:5 (2013), 2733–2797. MR 3116302 Zbl 1270.53096

[Shevchishin 2009] V. V. Shevchishin, “Lagrangian embeddings of the Klein bottle and the combinatorial properties of mapping class groups”, *Izv. Ross. Akad. Nauk Ser. Mat.* **73**:4 (2009), 153–224. In Russian; translated in *Izv. Math.* **73**:4 (2009), 797–859. MR 2583968 Zbl 1196.57021

[Shonkwiler and Vela-Vick 2011] C. Shonkwiler and D. S. Vela-Vick, “Legendrian contact homology and nondestabilizability”, *J. Symplectic Geom.* **9**:1 (2011), 33–44. MR 2787360 Zbl 1226.57013

[Sivek 2013] S. Sivek, “The contact homology of Legendrian knots with maximal Thurston–Bennequin invariant”, *J. Symplectic Geom.* **11**:2 (2013), 167–178. MR 3046489 Zbl 1277.53082

Received September 27, 2015. Revised April 19, 2016.

ORSOLA CAPOVILLA-SEARLE
DEPARTMENT OF MATHEMATICS
DUKE UNIVERSITY
DURHAM, NC 27708
UNITED STATES
orsola.capovilla.searle@duke.edu

LISA TRAYNOR
DEPARTMENT OF MATHEMATICS
BRYN MAWR COLLEGE
BRYN MAWR, PA 19010
UNITED STATES
ltraynor@brynmawr.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 285, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.285.345

A STRONG MULTIPLICITY ONE THEOREM FOR SL_2

JINGSONG CHAI AND QING ZHANG

It is known that the multiplicity one property holds for SL_2 while the strong multiplicity one property fails. However, in this paper we show that if we require further that a pair of cuspidal representations π and π' of SL_2 have the same local components at the archimedean places and the places above 2, and they are generic with respect to the same additive character, then they also satisfy the strong multiplicity one property. The proof is based on a local converse theorem for SL_2.

1. Introduction

Let F be a number field and $\mathbb{A} = \mathbb{A}_F$ be its ring of adeles. Let G be a linear reductive algebraic group defined over F. The study of the space of automorphic forms $L^2(G(F)\backslash G(\mathbb{A}))$ has been a central topic in the Langlands program and representation theory. Let $L^2_0(G(F)\backslash G(\mathbb{A}))$ be the subspace of cuspidal representations. Suppose π is an irreducible automorphic representation of $G(\mathbb{A})$. It is known that π occurs discretely with finite multiplicity m_π in $L^2_0(G(F)\backslash G(\mathbb{A}))$.

The multiplicities m_π are important in the study of automorphic forms and number theory. By [Jacquet and Shalika 1981; Badulescu 2008] and the work of Piatetski-Shapiro, the group $G = \mathrm{GL}_n$ and its inner forms have the property of multiplicity one, that is, $m_\pi \le 1$ for any π. This is also true for SL_2 by the famous work of D. Ramakrishnan [2000]. But in general the multiplicity one property fails, for example [Blasius 1994; Gan et al. 2002; Li 1997; Labesse and Langlands 1979] to list a few.

In the case of GL_n a stronger theorem, called the strong multiplicity one, holds. It says that for two cuspidal representations π_1 and π_2, if they have isomorphic local components almost everywhere, then they coincide in the space of cusp forms (not only isomorphic). It follows from the results in [Labesse and Langlands 1979] that SL_2 does not have this strong multiplicity one property. The multiplicity one property is already rare and the strong multiplicity one is even rarer. To the authors' knowledge the examples other than GL_n in this direction are the strong multiplicity one theorems for $\mathrm{U}(2, 1)$ [Gelbart and Piatetski-Shapiro 1984; Baruch 1997] and

MSC2010: 11F70, 22E50.
Keywords: strong multiplicity one theorem, local converse theorem, Howe vectors.

GSp_4 [Soudry 1987] and the rigidity theorem for $\mathrm{SO}(2n+1)$ [Jiang and Soudry 2003, Theorem 5.3].

The main purpose of this paper is to prove a weaker version of the strong multiplicity one result for $\mathrm{Sp}_2 = \mathrm{SL}_2$. Although we know strong multiplicity one does not hold in general for a pair of cuspidal representations π_1 and π_2 of $\mathrm{SL}_2(\mathbb{A})$, if we require that both π_1 and π_2 are generic with respect to the same additive character ψ of $\mathbb{A}$, then we can show that they also satisfy the strong multiplicity one property.

The reason for the failure of the strong multiplicity one for SL_2 is the existence of L-packets. According to the local conjecture of Gan–Gross–Prasad [2012, Conjecture 17.3] there is at most one ψ-generic representation in each L-packet. For SL_2, the result is known by the local discussion in [Labesse and Langlands 1979]. In this paper, we prove a local converse theorem for $\mathrm{SL}_2(F)$ when F is a p-adic field such that its residue characteristic is not 2, which will reprove the result of Labesse and Langlands [1979] and confirm a local converse conjecture of Jiang, see [Jiang 2006, Conjecture 3.7] and [Jiang and Nien 2013, Conjecture 6.3]. This also implies our version of strong multiplicity one easily.

We now give some details of our results. Gelbart and Piatetski-Shapiro [1987] constructed some Rankin–Selberg integrals to study L-functions on the group $G_n \times \mathrm{GL}(n)$, for $G_n = \mathrm{Sp}(n)$ and $\mathrm{U}(n,n)$. In particular, in Method C in that paper, if π is a globally generic cuspidal representation of $\mathrm{Sp}_{2n}(\mathbb{A})$, then τ is a cuspidal representation of $\mathrm{GL}_n(\mathbb{A})$. Consider the global Shimura type zeta-integral

$$I(s,\phi,E) = \int_{\mathrm{Sp}_{2n}(F)\backslash \mathrm{Sp}_{2n}(\mathbb{A})} \phi(g)\theta(g)E(g,s)dg,$$

where ϕ belongs to the space of π, $E(g,s)$ is a genuine Eisenstein series on $\widetilde{\mathrm{Sp}}_{2n}(\mathbb{A})$ built from the representation induced from the representation τ of $\mathrm{GL}_n(\mathbb{A})$ twisted by $|\det|^s$ and $\theta(g)$ is some theta series on $\widetilde{\mathrm{Sp}}_{2n}(\mathbb{A})$. Note that the product $\theta(g)E(g,s)$ is well-defined on Sp_{2n}. The global integral is shown to be Eulerian. The local functional equations and unramified calculations were also carried out by Gelbart and Piatetski-Shapiro [1987]. Although we will only consider the easiest case when $n=1$ of Gelbart and Piatetski-Shapiro's construction, we remark here that Ginzburg, Rallis and Soudry [1997; 1998] generalized the above construction to $\mathrm{Sp}_{2n} \times \mathrm{GL}_k$, for any k.

We study more details of Gelbart and Piatetski-Shapiro's local integral

$$\Psi(W_v,\phi_v,f_{s,v}) = \int_{N(F_v)\backslash \mathrm{SL}_2(F_v)} W_v(h)(\omega_{\psi_v^{-1}}(h)\phi_v)(1)f_{s,v}(h)dh$$

(for the unexplained notations, see sections below) when v is finite. These local zeta-integrals satisfy certain functional equations, which come from the intertwining operators on induced representations and certain uniqueness statements. These functional equations can then be used to define local gamma factors $\gamma(s, \pi_v, \eta_v, \psi_v)$, where π_v is a generic representation of $\mathrm{SL}_2(F_v)$, η_v is a character of $F_v^\times$ and ψ_v is a nontrivial additive character. The main local result of this paper can be formulated as follows.

Theorem 3.10 (Local converse theorem and stability of γ). *Suppose that the residue characteristic of the p-adic field F is not* 2 *and ψ is a nontrivial additive character of F. Let (π, V_π) and $(\pi', V_{\pi'})$ be two ψ-generic representations of* $\mathrm{SL}_2(F)$ *with the same central character.*

(1) *If $\gamma(s, \pi, \eta, \psi) = \gamma(s, \pi', \eta, \psi)$ for all quasicharacters η of $F^\times$, then $\pi \cong \pi'$.*

(2) *There is an integer $l = l(\pi, \pi')$ such that if η is a quasicharacter of $F^\times$ with conductor* $\mathrm{cond}(\eta) > l$, *then*

$$\gamma(s, \pi, \eta, \psi) = \gamma(s, \pi', \eta, \psi).$$

The proof of this result follows closely [Baruch 1995; 1997; Zhang 2015] and Howe vectors play an important role. With the help of this result, combined with a nonvanishing result on archimedean local integrals proved in Lemma 4.9, we follow the argument in [Baruch 1997, Theorem 7.2.13], or in [Casselman 1973, Theorem 2], to prove the main global result of this paper.

Theorem 4.8 (Strong multiplicity one for SL_2). *Let ψ be a nontrivial additive character of $F \setminus \mathbb{A}$ and let $\pi = \otimes \pi_v$ and $\pi' = \otimes \pi'_v$ be two irreducible cuspidal automorphic representations of* $\mathrm{SL}_2(\mathbb{A})$ *with the same central character. Suppose that π and π' are both ψ-generic. Let S be a finite set of* ***finite*** *places such that no place in S is above* 2. *If $\pi_v \cong \pi'_v$ for all $v \notin S$, then $\pi = \pi'$.*

Remark. The restriction on residue characteristic comes from Lemma 3.3. It is expected that this restriction can be removed.

Besides the above, we also in this paper include a discussion of relations between global genericity and local genericity. An irreducible cuspidal automorphic representation (π, V_π) is called globally ψ-generic if for some $\phi \in V_\pi$, the integral

$$\int_{N(F)\setminus N(\mathbb{A})} \phi(ug)\psi^{-1}(u)du \neq 0$$

for some $g \in \mathrm{SL}_2(\mathbb{A})$. The representation π is called locally ψ-generic if each of its local component is generic for the corresponding local components of ψ. It is easy to see that if π is globally ψ-generic, then π is also locally ψ-generic. It is a conjecture that on a reductive algebraic group G, the converse is also true. This conjecture is closely related to the Ramanujan conjecture. See [Shahidi 2011] for more detailed discussions. We confirm this conjecture for SL_2.

Theorem 4.3. *Let $\pi = \otimes_v \pi_v$ be an irreducible cuspidal automorphic representation of $\mathrm{SL}_2(\mathbb{A})$ and $\psi = \otimes \psi_v$ be a nontrivial additive character of $F \setminus \mathbb{A}$. Then π is ψ-generic if and only if each π_v is ψ_v-generic.*

Gelbart, Rogawski and Soudry [1997, Proposition 2.5] proved similar results for $\mathrm{U}(1, 1)$ and for endoscopic cuspidal automorphic representations of $\mathrm{U}(2, 1)$. From the discussions given in [Gelbart et al. 1997] Theorem 4.3 follows directly from the results of Labesse and Langlands [1979]. Here, we include this result because we adopt a local argument (see Proposition 2.1) which is different from that given in [Labesse and Langlands 1979]. Hopefully, this local argument can be extended to more general groups.

As explained above, there is essentially nothing new in this paper. All the results and proofs should be known to the experts. Our task here is simply to try to write down the details and to check everything works out as expected.

This paper is organized as follows. In Section 2 we collect basic results about the local zeta-integrals which will be needed. In Section 3 we study the Howe vectors and use them to prove the local converse theorem and stability of local gamma factors. In Section 4 we prove the main global results.

1A. ***Notations.*** Let F be a field. In $\mathrm{SL}_2(F)$, we consider the following subgroups. Let B be the upper triangular subgroup. Let $B = TN$ be the Levi-decomposition, where T is the diagonal torus and N is the upper triangular unipotent. Denote

$$t(a) = \begin{pmatrix} a & \\ & a^{-1} \end{pmatrix} \in T, \text{ for } a \in F^\times, \quad \text{and} \quad n(b) = \begin{pmatrix} 1 & b \\ & 1 \end{pmatrix} \in N, \text{ for } b \in F.$$

Let $\overline{N}$ be the lower triangular unipotent and denote

$$\bar{n}(x) = \begin{pmatrix} 1 & \\ x & 1 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} & 1 \\ -1 & \end{pmatrix}.$$

Denote by St the natural inclusion $\mathrm{SO}_3(\mathbb{C}) \to \mathrm{GL}_3(\mathbb{C})$ and view it as the "standard" representation of ${}^L\mathrm{SL}_2 = \mathrm{SO}_3(\mathbb{C})$.

2. The local zeta-integral

2A. ***The genericity of representations of $\mathbf{SL_2}(F)$.*** In this section let F be a local field and ψ be a nontrivial additive character of F, which is also viewed as a character of $N(F)$. For $\kappa \in F^\times$ and $g \in \mathrm{SL}_2(F)$ we define

$$g^\kappa = \begin{pmatrix} \kappa & \\ & 1 \end{pmatrix} g \begin{pmatrix} \kappa^{-1} & \\ & 1 \end{pmatrix}.$$

Explicitly

$$\begin{pmatrix} x & y \\ z & w \end{pmatrix}^{\kappa} = \begin{pmatrix} x & \kappa y \\ \kappa^{-1} z & w \end{pmatrix}.$$

Note that if $\kappa \in F^{\times,2}$, say $\kappa = a^2$, then $g^\kappa = t(a) g t(a)^{-1}$, i.e., $g \mapsto g^\kappa$ is an inner automorphism on $\mathrm{SL}_2(F)$. Let (π, V_π) be an infinite dimensional irreducible smooth representation of $\mathrm{SL}_2(F)$. We consider the representation $(\pi^\kappa, V_{\pi^\kappa})$ defined by

$$V_{\pi^\kappa} = V_\pi \quad \text{and} \quad \pi^\kappa(g) = \pi(g^\kappa).$$

Let ψ_κ be the character of F defined by $\psi_\kappa(b) = \psi(\kappa b)$. If (π, V_π) is ψ-generic with a nonzero ψ Whittaker functional $\Lambda : V_\pi \to \mathbb{C}$, one verifies that

$$\Lambda(\pi^\kappa(n)v) = \Lambda(\pi(n^\kappa)v) = \psi_\kappa(n)\Lambda(v)$$

for all $n \in N(F)$ and all $v \in V_{\pi^\kappa} = V_\pi$. Hence $(\pi^\kappa, V_{\pi^\kappa})$ is ψ_κ-generic.

Proposition 2.1. *If π is both ψ- and ψ_κ-generic, then $\pi \cong \pi^\kappa$.*

Proof. If F is nonarchimedean, the proof is similar to the $U(1,1)$ case as in [Zhang 2015].

If F is archimedean the case $F = \mathbb{C}$ is easy, as every κ has a square root in $\mathbb{C}$. Now consider $F = \mathbb{R}$. We will work with the category of smooth representations of moderate growth of finite length. The Whittaker functional is an exact functor from this category to the category of vector spaces by [Casselman et al. 2000, Theorem 8.2].

We first consider the case when $I(\chi) = \mathrm{Ind}_B^G(\chi)$ for some quasicharacter χ of $F^\times$. For $f \in I(\chi)$, consider the function f^κ on $\mathrm{SL}_2(F)$ defined by $f^\kappa(g) = f(g^{\kappa^{-1}})$. It is clear that $f^\kappa \in I(\chi)^\kappa$ and the map $f \mapsto f^\kappa$ defines an isomorphism $I(\chi) \to I(\chi)^\kappa$.

By results in [Vogan 1981, Chapter 2], if π is not a fully induced representation then it can be embedded into a principal series $I(\chi)$. This $I(\chi)$ has two irreducible infinite dimensional subrepresentations, use π' to denote the other one. The quotient of $I(\chi)$ by the sum of π and π', denoted by π'', is finite dimensional, i.e., we have a short exact sequence

$$0 \to \pi \oplus \pi' \to I(\chi) \to \pi'' \to 0.$$

First, by [Casselman et al. 2000, Theorem 6.1], we know that the Whittaker functionals on $I(\chi)$ are one dimensional for either ψ or ψ_κ. Note that π'' cannot be generic as it is finite dimensional. Since the Whittaker functor is exact, it follows that the dimension of Whittaker functionals on $\pi \oplus \pi'$ is also one for either ψ or ψ_κ. By the assumption π is both ψ- and ψ_κ-generic, thus π' is neither ψ- nor ψ_κ-generic.

Now since π is ψ-generic, π^κ is ψ_κ-generic. Hence the image of π under the isomorphism $I(\chi) \to I(\chi)^\kappa$ given by $f \mapsto f^\kappa$ is again ψ_κ-generic and hence it has to be ψ-generic and isomorphic to π, which finishes the proof. $\square$

2B. *Weil representations of* $\widetilde{\mathrm{SL}}_2$. Let $\widetilde{\mathrm{SL}}_2$ be the metaplectic double cover of SL_2. Then we have an exact sequence

$$0 \to \mu_2 \to \widetilde{\mathrm{SL}}_2 \to \mathrm{SL}_2 \to 0,$$

where $\mu_2 = \{\pm 1\}$.

The product on $\widetilde{\mathrm{SL}}_2(F)$ is given by

$$(g_1, \zeta_1)(g_2, \zeta_2) = (g_1 g_2, \zeta_1 \zeta_2 c(g_1, g_2)),$$

where $c : \mathrm{SL}_2(F) \times \mathrm{SL}_2(F) \to \{\pm 1\}$ is defined by Hilbert symbols as

$$c(g_1, g_2) = (\mathbf{x}(g_1), \mathbf{x}(g_2))_F (-\mathbf{x}(g_1)\mathbf{x}(g_2), \mathbf{x}(g_1 g_2))_F,$$

where

$$\mathbf{x}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{cases} c & c \neq 0, \\ d & c = 0, \end{cases}$$

and $(\ ,\ )_F$ is the Hilbert symbol. For these formulas for the Kubota cocycle see [Kubota 1969, Section 3].

For a subgroup A of $\mathrm{SL}_2(F)$, we denote by $\widetilde{A}$ the preimage of A in $\widetilde{\mathrm{SL}}_2(F)$, which is a subgroup of $\widetilde{\mathrm{SL}}_2(F)$. For an element $g \in \mathrm{SL}_2(F)$, we sometimes abuse notation by writing $(g, 1) \in \widetilde{\mathrm{SL}}_2(F)$ as g.

A representation π of $\widetilde{\mathrm{SL}}_2(F)$ is called genuine if $\pi(\zeta g) = \zeta \pi(g)$ for all $g \in \widetilde{\mathrm{SL}}_2(F)$ and $\zeta \in \mu_2$. Let ψ be an additive character of F. Then there is a Weil representation ω_ψ of $\widetilde{\mathrm{SL}}_2(F)$ on the space $\mathcal{S}(F)$ of Schwartz–Bruhat functions on F. For $f \in \mathcal{S}(F)$, we have the well-known formulas:

$$\left(\omega_\psi \begin{pmatrix} & 1 \\ -1 & \end{pmatrix}\right) f(x) = \gamma(\psi) \hat{f}(x),$$

$$\left(\omega_\psi \begin{pmatrix} 1 & b \\ & 1 \end{pmatrix}\right) f(x) = \psi(bx^2) f(x), \quad b \in F$$

$$\left(\omega_\psi \begin{pmatrix} a & \\ & a^{-1} \end{pmatrix}\right) f(x) = |a|^{1/2} \frac{\gamma(\psi)}{\gamma(\psi_a)} f(ax), \quad a \in F^\times.$$

$$\omega_\psi(\zeta) f(x) = \zeta f(x), \quad \zeta \in \mu_2.$$

Here $\hat{f}(x) = \int_F f(y)\psi(2xy)dy$, where dy is normalized so that $\hat{\hat{f}}(x) = f(-x)$, $\gamma(\psi)$ is the Weil index and $\psi_a(x) = \psi(ax)$.

Let $\widetilde{T}$ be the inverse image of $T = \{t(a) := \left(\begin{smallmatrix} a & \\ & a^{-1}\end{smallmatrix}\right), a \in F^\times\} \subset \mathrm{SL}_2(F)$ in $\widetilde{\mathrm{SL}}_2(F)$. The product in $\widetilde{T}$ is given by the Hilbert symbol, i.e.,

$$(t(a), \zeta_1)(t(b), \zeta_2) = (t(ab), \zeta_1\zeta_2(a, b)_F).$$

The function

$$\mu_\psi(a) = \frac{\gamma(\psi)}{\gamma(\psi_a)}$$

satisfies

$$\mu_\psi(a)\mu_\psi(b) = \mu_\psi(ab)(a, b)_F,$$

and thus extends to a genuine character of $\widetilde{T}$.

The representation ω_ψ is not irreducible and we have $\omega_\psi = \omega_\psi^+ \oplus \omega_\psi^-$, where ω_ψ^+ and ω_ψ^- are the subrepresentations on even and odd functions in $\mathcal{S}(F)$, respectively. All the above facts can be found in [Gelbart and Piatetski-Shapiro 1980, Section 1].

2C. ***The local zeta-integral.*** Let $\mu_\psi(a) = \gamma(\psi)/\gamma(\psi_a)$ be as above, which is viewed as a character of $\widetilde{T}$. Let η be a quasicharacter of $F^\times$ and η_s denote the character $\eta|\cdot|^s$ of $F^\times$. Consider the induced representation $I(s, \eta, \psi) = \mathrm{Ind}_{\widetilde{B}(F)}^{\widetilde{\mathrm{SL}}_2(F)}(\eta_{s-1/2}\mu_\psi)$.

Let (π, V) be a ψ-generic representation of $\mathrm{SL}_2(F)$ with its Whittaker model $\mathcal{W}(\pi, \psi)$. Choose $W \in \mathcal{W}(\pi, \psi)$, $\phi \in \mathcal{S}(F)$ and $f_s \in I(s, \eta, \psi^{-1})$. Note that $(\omega_{\psi^{-1}}(h)\phi)(1) f_s(h)$ is well-defined as a function on $\mathrm{SL}_2(F)$ and consider the integral

$$\Psi(W, \phi, f_s) = \int_{N(F)\backslash \mathrm{SL}_2(F)} W(h)(\omega_{\psi^{-1}}(h)\phi)(1) f_s(h)\,dh.$$

By results in [Gelbart et al. 1987, Sections 5 and 12], the above integral is absolutely convergent when $\mathrm{Re}(s)$ is large enough and has a meromorphic continuation to the whole plane.

Remark. Gelbart and Piatetski-Shapiro [1987, Method C] constructed a global zeta-integral for $\mathrm{Sp}_{2n} \times \mathrm{GL}_n$ which showed that it is Eulerian. They also sketched a proof of the local functional equation. The above integral is the simplest case of the Gelbart and Piatetski-Shapiro integral, namely when $n = 1$.

2D. ***Local functional equation.*** The trilinear form $(W, \phi, f_s) \mapsto \Psi(W, \phi, f_s)$ defines an element in

$$\mathrm{Hom}_{\mathrm{SL}_2}(\pi \otimes \omega_{\psi^{-1}} \otimes I(s, \eta, \psi^{-1}), \mathbb{C}),$$

which has dimension at most one. The proof of this fact is given in [Gelbart et al. 1987, §11] and also can be deduced by the uniqueness of the Fourier–Jacobi model for SL_2, see [Sun 2012]. Let

$$M_s : I(s, \eta, \psi^{-1}) \to I(1-s, \eta^{-1}, \psi^{-1})$$

be the standard intertwining operator, i.e.,

$$M_s(f_s)(g) = \int_N f_s(wng)dn.$$

By the one dimensionality of the above Hom space we get the following:

Proposition 2.2. *There is a meromorphic function $\gamma(s, \pi, \eta, \psi)$ such that*

$$\Psi(W, \phi, M_s(f_s)) = \gamma(s, \pi, \eta, \psi)\Psi(W, \phi, f_s),$$

for all $W \in \mathcal{W}(\pi, \psi)$, $\phi \in \mathcal{S}(F)$ and $f_s \in I(s, \eta, \psi^{-1})$.

2E. *Unramified calculation.* The unramified calculation of Method C is in fact not included in [Gelbart et al. 1987], but it can be simply done in the SL_2-case.

Let F be a nonarchimedian local field with odd residue characteristic. Suppose everything is unramified. Then the character μ_ψ is unramified, [Szpruch 2009, p. 2188]. Suppose the representation (π, V) has Satake parameter a, which means that π is the unramified component $\mathrm{Ind}_{B(F)}^{\mathrm{SL}_2(F)}(\nu)$ for an unramified character ν and $a = \nu(p_F)$, where p_F is some prime element of F. Let

$$b_k = t(p_F^k) = \mathrm{diag}(p_F^k, p_F^{-k}),$$

and W be the spherical Whittaker functional normalized by $W(e) = 1$. Then $W(b_k) = 0$ for $k < 0$ and

$$W(b_k) = \frac{q^{-k}}{a-1}(a^{k+1} - a^{-k}),$$

by the general Casselman–Shalika formula [1980, Theorem 5.4]. For $k \geq 0$ we have

$$(\omega_{\psi^{-1}}(b_k)\phi)(1) = \mu_{\psi^{-1}}(p_F^k)|p_F^k|^{1/2},$$

where ϕ is the characteristic function of the ring of integers $\mathcal{O}_F$. On the other hand, let f_s be the standard spherical section of $I(s, \eta, \psi^{-1})$ normalized by $f_s(1) = 1$. Then we have

$$f_s(b_k) = \eta(p_F^k)|p_F^k|^{s+1/2}\mu_{\psi^{-1}}(p_F^k).$$

Since $\mu_{\psi^{-1}}(p_F^k)\mu_{\psi^{-1}}(p_F^k) = (p_F^k, p_F^k)_F = (p_F, -1)_F^k$, we have

$$\begin{aligned}
\Psi(W, \phi, f_s) &= \int_{F^\times}\int_K W(t(a)k)\omega_{\psi^{-1}}(t(ak)\phi)(1) f_s(t(a)k)|a|^{-2}dkda \\
&= \int_{F^\times} W(t(a))\omega_{\psi^{-1}}(t(a))\phi(1) f_s(t(a))|a|^{-2}da \\
&= \sum_{k\geq 0} W(b_k)(\omega_{\psi^{-1}}(b_k)\phi)(1) f_s(b_k)|p_F^k|^{-2} \\
&= \frac{1}{a-1}\sum_{k\geq 0}(a^{k+1}-a^{-k})(p_F, -1)^k\eta(p_F)^k q_F^{-ks} \\
&= \frac{1+c}{(1-ac)(1-a^{-1}c)} = \frac{1-c^2}{((1-ac)(1-c)(1-a^{-1}c)} \\
&= \frac{L(s, \pi, St\otimes\eta\chi)}{L(2s, \eta^2)},
\end{aligned}$$

where $c = (p_F, -1)\eta(p_F)q_F^{-s}$, and $\chi(a) = (a, -1)_F$. Recall that St is the standard representation of ${}^L\,\mathrm{SL}_2 = \mathrm{SO}_3(\mathbb{C})$.

Remark. From the calculation of the μ_ψ given in [Szpruch 2009, Lemmas 1.5 and 1.10], one can check that

$$M_s(f_s) = \frac{L(2s-1, \eta^2)}{L(2s, \eta^2)} f_{1-s},$$

where f_s and f_{1-s} are the standard spherical sections in, respectively, $I(s, \eta, \psi^{-1})$ and $I(1-s, \eta^{-1}, \psi^{-1})$. Thus the factor $L(2s, \eta^2)$ appearing in the above unramified calculation will play the role of the normalizing factor of a global intertwining operator or Eisenstein series.

3. Howe vectors and the local converse theorem

In this section, we assume F is a p-adic field with odd residue characteristic. We will follow Baruch's method [1995; 1997] to give a proof of the local converse theorem for generic representations of $\mathrm{SL}_2(F)$.

3A. ***Howe vectors.*** Let ψ be an unramified character. For a positive integer m, let $K_m = (1 + M_{2\times 2}(\mathcal{P}_F^m)) \cap \mathrm{SL}_2(F)$ where $\mathcal{P}_F = (p_F)$ denotes the maximal ideal in $\mathcal{O}_F$. Define a character τ_m of K_m by

$$\tau_m(k) = \psi(p_F^{-2m}k_{12})$$

for $k = (k_{ij}) \in K_m$. It is easy to see that τ_m is indeed a character on K_m.

Let $d_m = t(p_F^{-m})$. Consider the subgroup $J_m = d_m K_m d_m^{-1}$. Then

$$J_m = \begin{pmatrix} 1+\mathcal{P}_F^m & \mathcal{P}_F^{-m} \\ \mathcal{P}_F^{3m} & 1+\mathcal{P}_F^m \end{pmatrix} \cap \mathrm{SL}_2(F).$$

Define $\psi_m(j) = \tau_m(d_m^{-1} j d_m)$ for $j \in J_m$. For a subgroup $H \subset \mathrm{SL}_2(F)$, denote $H_m = H \cap J_m$. It is easy to check that $\psi_m|_{N_m} = \psi|_{N_m}$.

Let π be an irreducible smooth ψ-generic representation of $\mathrm{SL}_2(F)$ and let $v \in V_\pi$ be a vector such that $W_v(1) = 1$. For $m \geq 1$, as in [Baruch 1995; 1997] we consider

$$v_m = \frac{1}{\mathrm{Vol}(N_m)} \int_{N_m} \psi(n)^{-1} \pi(n) v dn. \tag{3-1}$$

Let $L \geq 1$ be an integer such that v is fixed by K_L. Following E. M. Baruch, we call $v_m, m \geq L$ **Howe vectors**.

Lemma 3.1. *We have*:

(1) $W_{v_m}(1) = 1$.

(2) *If* $m \geq L$ *then* $\pi(j) v_m = \psi_m(j) v_m$ *for all* $j \in J_m$.

(3) *If* $k \leq m$ *then*

$$v_m = \frac{1}{\mathrm{Vol}(N_m)} \int_{N_m} \psi(u)^{-1} \pi(u) v_k du.$$

The proof of this lemma is the same as the proof in the U(2, 1) case, which is given in [Baruch 1997, Lemma 5.2].

Lemma 3.2. *Let* $m \geq L$ *and* $t = t(a)$ *for* $a \in F^\times$:

(1) *If* $W_{v_m}(t) \neq 0$, *we have*
$$a^2 \in 1 + \mathcal{P}_F^m.$$

(2) *If* $W_{v_m}(tw) \neq 0$, *we have*
$$a^2 \in \mathcal{P}^{-3m}.$$

Proof.

(1) Take $x \in \mathcal{P}^{-m}$. We then have $n(x) \in N_m \subset J_m$. From the relation

$$tn(x) = n(a^2 x)t$$

and (2) of Lemma 3.1 we have

$$\psi(x) W_{v_m}(t) = \psi(a^2 x) W_{v_m}(t).$$

If $W_{v_m}(t) \neq 0$ we get $\psi(x) = \psi(a^2 x)$ for all $x \in \mathcal{P}^{-m}$. Since ψ is unramified we get $a^2 \in 1 + \mathcal{P}^m$.

(2) For $x \in \mathcal{P}^{3m}$ we have $\bar{n}(x) \in \bar{N}_m$. From the relation $tw\bar{n}(x) = n(-a^2x)tw$ and Lemma 3.1 (2) we get

$$W_{v_m}(tw) = \psi(-a^2x)W_{v_m}(tw).$$

Thus if $W_{v_m}(tw) \neq 0$ we get $\psi(-a^2x) = 1$ for all $x \in \mathcal{P}^{3m}$. Thus $a^2 \in \mathcal{P}^{-3m}$. □

Lemma 3.3. *For $m \geq 1$ the squaring map from $1 + \mathcal{P}^m \to 1 + \mathcal{P}^m$, sending $a \mapsto a^2$, is well-defined and surjective.*

This lemma requires that the residue field of F is not of characteristic 2 which we assume throughout this section.

Proof. For $x \in \mathcal{P}^m$, it is clear that $(1+x)^2 = 1 + 2x + x^2 \in 1 + \mathcal{P}^m$. Thus the square map is well-defined. On the other hand, we take $u \in 1 + \mathcal{P}^m$ and consider the equation $f(X) := X^2 - u = 0$. We have $f'(X) = 2X$. Since $q^{-m} = |1-u| = |f(1)| < |f'(1)|^2 = |2|^2 = 1$ by Newton's Lemma, see for example [Lang 1994, Proposition 2, Chapter II], there is a root $a \in \mathcal{O}_F$ of $f(X)$ such that

$$|a-1| \leq \frac{|f(1)|}{|f'(1)|^2} = |1-u| = q^{-m}.$$

Thus we get a root $a \in 1 + \mathcal{P}^m$ of $f(X)$. This completes the proof. □

Let $Z = \{\pm 1\}$ and identify Z with the center of $\mathrm{SL}_2(F)$. Denote by ω_π the central character of π.

Corollary 3.4. *Let $m \geq L$. Then we have*

$$W_{v_m}(t(a)) = \begin{cases} \omega_\pi(z) & \textit{if } a = za' \textit{ for some } z \in Z \textit{ and } a' \in 1 + \mathcal{P}^m, \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. Suppose that $W_{v_m}(t(a)) \neq 0$. Then by Lemma 3.2 we have $a^2 \in 1 + \mathcal{P}^m$. By Lemma 3.3 there exists an $a' \in 1 + \mathcal{P}^m$ such that $a^2 = (a')^2$. Thus $a = za'$ for some $z \in Z$. Since $a' \in 1 + \mathcal{P}^m$ we get $t(a') \in J_m$. The assertion follows from Lemma 3.1. □

From now on, we fix two ψ-generic representations (π, V_π) and $(\pi', V_{\pi'})$ with the same central characters. Fix v and v' such that $W_v(1) = 1 = W_{v'}(1)$. Let L be an integer such that both v and v' are fixed by K_L. For $m \geq 1$ consider the Howe vectors v_m and v'_m.

By Corollary 3.4 and the fact that $\omega_\pi = \omega_{\pi'}$ we get the following:

Corollary 3.5. *For $m \geq L$ we have $W_{v_m}(g) = W_{v'_m}(g)$ for all $g \in B$.*

Lemma 3.6 (Baruch). *If $m \geq 4L$ and $n \in N - N_m$ we have*

$$W_{v_m}(twn) = W_{v'_m}(twn),$$

for all $t \in T$.

Proof. This is a special case of [Baruch 1995, Lemma 6.2.2]. A similar result for U(2, 1) is given in [Baruch 1997, Proposition 5.7]. We just remark that the proof of this lemma depends on Corollary 3.5, and hence requires that the residue characteristic of F is not 2. □

3B. *Induced representations.* Note that $\overline{N}(F)$ and $N(F)$ split in $\widetilde{\mathrm{SL}}_2(F)$. Moreover, for $g_1 \in N$ and $g \in \overline{N}$ we have $c(g_1, g_2) = 1$. In fact if $g_1 = n(y)$ and $g_2 = \bar{n}(x)$ with $x \neq 0$ we have $\mathbf{x}(g_1) = 1$ and $\mathbf{x}(g_2) = x$. Thus

$$c(g_1, g_2) = (1, x)_F(-x, x)_F = 1.$$

This shows that $N(F) \cdot \overline{N}(F) \subset \mathrm{SL}_2(F)$, where $\mathrm{SL}_2(F)$ denotes the subset of $\widetilde{\mathrm{SL}}_2(F)$ which consists of elements of the form $(g, 1)$ for $g \in \mathrm{SL}_2(F)$.

Let X be an open compact subgroup of $N(F)$. For $x \in X$ and $i > 0$ consider the set $A(x, i) = \{\bar{n} \in \overline{N}(F) : \bar{n}x \in B \cdot \overline{N}_i\}$.

Lemma 3.7. (1) *For any positive integer c there exists an integer $i_1 = i_1(X, c)$ such that for all $i \geq i_1$, $x \in X$ and $\bar{n} \in A(x, i)$ we have*

$$\bar{n}x = nt(a)\bar{n}_0,$$

with $n \in N$, $\bar{n}_0 \in \overline{N}_i$ and $a \in 1 + \mathcal{P}^c$.

(2) *There exists an integer $i_0 = i_0(X)$ such that for all $i \geq i_0$ we have $A(x, i) = \overline{N}_i$.*

Proof. By abuse of notation, for $x \in X$ we write $x = n(x)$. Since X is compact there is a constant C such that $|x| < C$ for all $n(x) \in X \subset N$.

For $n(x) \in X$ and $\bar{n}(y) \in A(x, i)$ we have $\bar{n}(y)n(x) \in B \cdot \overline{N}_i$. Thus we can assume that

$$\bar{n}(y)n(x) = \begin{pmatrix} a & b \\ & a^{-1} \end{pmatrix} \bar{n}(\bar{y})$$

for $a \in F^\times$, $b \in F$ and $\bar{y} \in \mathcal{P}^{3i}$. Rewrite the above expression as

$$\bar{n}(-y) \begin{pmatrix} a & b \\ & a^{-1} \end{pmatrix} = n(x)\bar{n}(-\bar{y}),$$

or

$$\begin{pmatrix} a & b \\ -ay & a^{-1} - by \end{pmatrix} = \begin{pmatrix} 1 - x\bar{y} & x \\ -\bar{y} & 1 \end{pmatrix}.$$

Thus we get

$$a = 1 - x\bar{y} \text{ and } ay = \bar{y}.$$

Since $|x| < C$ and $\bar{y} \in \mathcal{P}^{3i}$ it is clear that for any positive integer c we can choose $i_1(X, c)$ such that $a = 1 - x\bar{y} \in 1 + \mathcal{P}^c$ for all $n(x) \in X$ and $\bar{n}(y) \in A(x, i)$. This proves (1).

If we take $i_0(X) = i_1(X, 1)$ we get $a \in 1+\mathcal{P} \subset \mathcal{O}^\times$ for $i \geq i_0$. From $ay = \bar{y}$ we get $y \in \mathcal{P}^{3i}$. Thus for $i \geq i_0(X)$ we have that $\bar{n}(y) \in \overline{N}_i$, i.e., $A(x, i) \subset \overline{N}_i$.

The other direction can be checked similarly if i is large. We omit the details. □

Given a positive integer i and a complex number $s \in \mathbb{C}$ we consider the following function f_s^i on $\widetilde{\mathrm{SL}}_2(F)$:

$$f_s^i(\tilde{g}) = \begin{cases} \zeta \mu_{\psi^{-1}}(a)\eta_{s+1/2}(a) & \text{if } \tilde{g} = \left(\begin{pmatrix} a & b \\ & a^{-1}\end{pmatrix}, \zeta\right)\bar{n}(x), \\ & \quad \text{with } a \in F^\times, b \in F, \zeta \in \mu_2, x \in \mathcal{P}^{3i}, \\ 0 & \text{otherwise.}\end{cases}$$

Lemma 3.8. (1) *There exists an integer $i_2(\eta)$ such that for all $i \geq i_2(\eta)$, f_s^i defines a section in $I(s, \eta, \psi^{-1})$.*

(2) *Let X be an open compact subset of N. There exists an integer $I(X, \eta) \geq i_2(\eta)$ such that for all $i \geq I(X, \eta)$ we have*

$$\tilde{f}_s^i(wx) = \mathrm{vol}(\overline{N}_i) = q_F^{-3i}$$

for all $x \in X$, where $\tilde{f}_s^i = M_s(f_s^i)$ and $w = \begin{pmatrix} & 1 \\ -1 & \end{pmatrix}$.

Proof. (1) From the definition it is clear that

$$f_s^i\left(\left(\begin{pmatrix} a & b \\ & a^{-1}\end{pmatrix}, \zeta\right)\tilde{g}\right) = \zeta \mu_{\psi^{-1}}(a)\eta_{s+1/2}(a) f_s^i(\tilde{g}),$$

for $a \in F^\times$, $b \in F$, $\zeta \in \mu_2$ and $\tilde{g} \in \widetilde{\mathrm{SL}}_2(F)$. It suffices to show that for i large there is an open compact subgroup $\widetilde{H}_i \subset \widetilde{\mathrm{SL}}_2(F)$ such that $f_s^i(\tilde{g}\tilde{h}) = f_s^i(\tilde{g})$ for all $\tilde{g} \in \widetilde{\mathrm{SL}}_2(F)$ and $\tilde{h} \in \widetilde{H}_i$.

If ψ is unramified and the residue characteristic is not 2 as we assumed then the character $\mu_{\psi^{-1}}$ is trivial on $\mathcal{O}_F^\times$, see for example [Szpruch 2009, p. 2188].

Let c be a positive integer such that η is trivial on $1 + \mathcal{P}^c$. Let $i_2(\eta) = \max\{c, i_0(N \cap K_c), i_1(N \cap K_c, c)\}$. For $i \geq i_2(\eta)$ we take $\widetilde{H}_i = K_{4i} = 1 + M_2(\mathcal{P}^{4i})$. Note that K_{4i} splits and thus can be viewed as a subgroup of $\widetilde{\mathrm{SL}}_2$. We now check that for $i \geq i_2(\eta)$ we have $f_s^i(\tilde{g}h) = f_s(\tilde{g})$ for all $\tilde{g} \in \widetilde{\mathrm{SL}}_2$ and $h \in K_{4i}$. We have the decomposition $K_{4i} = (N \cap K_{4i})(T \cap K_{4i})(\overline{N} \cap K_{4i})$. For $h \in \overline{N} \cap K_{4i} \subset \overline{N}_i$ we have $f_s^i(\tilde{g}h) = f_s^i(\tilde{g})$ by the definition of f_s^i. Now we take $h \in T \cap K_{4i}$. Write $h = t(a_0)$ with $a_0 \in 1+\mathcal{P}^{4i}$. We have $\bar{n}(x)h = h\bar{n}(a_0^{-2}x)$. It is clear that $x \in \mathcal{P}^{3i}$ if and only if $a_0^{-2}x \in \mathcal{P}^{3i}$. On the other hand, for any $a \in F^\times$ and $b \in F$ we have

$$c\left(\begin{pmatrix} a & b \\ & a^{-1}\end{pmatrix}, t(a_0)\right) = (a, a_0) = 1,$$

since $a_0 \in 1+\mathcal{P}_F^{4i} \subset F^{\times,2}$ by Lemma 3.3. Thus we get

$$\left(\begin{pmatrix} a & b \\ & a^{-1} \end{pmatrix}, \zeta\right)\bar{n}(x)h = \left(\begin{pmatrix} aa_0 & ba_0^{-1} \\ & a^{-1}a_0^{-1} \end{pmatrix}, \zeta\right)\bar{n}(a_0^{-2}x).$$

By the definition of f_s^i, if $x \in \mathcal{P}^{3i}$ for $g = \left(\begin{pmatrix} a & b \\ & a^{-1} \end{pmatrix}, \zeta\right)\bar{n}(x)$ then we get

$$f_s^i(gh) = \mu_{\psi^{-1}}(a_0 a)\eta_{s+1/2}(aa_0) = \mu_{\psi^{-1}}(a)\eta_{s+1/2}(a) = f_s^i(g)$$

by the assumption on i.

Finally, we consider $h \in N \cap K_{4i} \subset N \cap K_c$. By the assumption on i we get

$$A(h,i) = A(h^{-1},i) = \overline{N}_i.$$

In particular, for $\bar{n} \in \overline{N}_i$ we have $\bar{n}h \in B\cdot\overline{N}_i$ and $\bar{n}h^{-1} \in B\cdot\overline{N}_i$. Now it is clear that $\tilde{g} \in \widetilde{B}\cdot\overline{N}_i$ if and only if $\tilde{g}h \in \widetilde{B}\cdot\overline{N}_i$. Thus $f_s^i(\tilde{g}) = 0$ if and only if $f_s^i(\tilde{g}h) = 0$. Moreover, for $\bar{n} \in \overline{N}_i$, we have

$$\bar{n}h = \begin{pmatrix} a_0 & b_0 \\ & a_0^{-1} \end{pmatrix}\bar{n}_0$$

for $a_0 \in 1+\mathcal{P}^c$, $b_0 \in F$ and $\bar{n}_0 \in \overline{N}_i$. Thus for $\tilde{g} = \left(\begin{pmatrix} a & b \\ & a^{-1} \end{pmatrix}, \zeta\right)\bar{n}$ with $\bar{n} \in \overline{N}_i$ we get

$$\tilde{g}h = \left(\begin{pmatrix} aa_0 & ab_0 + a_0^{-1}b \\ & a_0^{-1}a^{-1} \end{pmatrix}, \zeta\right)\bar{n}_0.$$

Here we used the fact that $a_0 \in 1+\mathcal{P}^c$ is a square and thus

$$c\left(\begin{pmatrix} a & b \\ & a^{-1} \end{pmatrix}, \begin{pmatrix} a_0 & b_0 \\ & a_0^{-1} \end{pmatrix}\right) = 1.$$

Since $\mu_{\psi^{-1}}(a_0) = 1$, $(a, a_0) = 1$ and $\eta_{s+1/2}(a_0) = 1$ we get

$$f_s^i(\tilde{g}h) = f_s^i(g).$$

This finishes the proof of (1).

(2) As in the proof of (1) let c be a positive integer such that η is trivial on $1+\mathcal{P}^c$. Take $I(X,\eta) = \max\{i_1(X,c), i_0(X)\}$. We have

$$\tilde{f}_s^i(wx) = \int_N f_s^i(w^{-1}nwx)\,dn.$$

By the definition of f_s^i, $f_s^i(w^{-1}nwx) \neq 0$ if and only if $w^{-1}nwx \in B\overline{N}_i$ if and only if $w^{-1}nw \in A(x,i) = \overline{N}_i$ for all $i \geq I(X)$ and $x \in X$. On the other hand, if

$w^{-1}nw \in A(x,i)$, we have

$$w^{-1}nwx = \begin{pmatrix} a & b \\ & a^{-1} \end{pmatrix} \bar{n}_0$$

with $a \in 1 + \mathcal{P}_F^c$. Thus

$$f_s^i(w^{-1}nwx) = \eta_{s+1/2}(a)\mu_{\psi^{-1}}(a) = 1.$$

Now the assertion is clear. □

3C. *The local converse theorem.*

Lemma 3.9. *Let ϕ^m be the characteristic function of $1 + \mathcal{P}^m$. Then*

(1) *for $n \in N_m$ we have $\omega_{\psi^{-1}}(n)\phi^m = \psi^{-1}(n)\phi^m$, and*

(2) *for $\bar{n} \in \overline{N}_m$ we have $\omega_{\psi^{-1}}(\bar{n})\phi^m = \phi^m$.*

Proof.

(1) For $n = n(b) \in N_m$ we have $b \in \mathcal{P}^{-m}$. For $x \in 1 + \mathcal{P}^m$ we have $bx^2 - b \in \mathcal{O}_F$. Thus

$$\omega_{\psi^{-1}}(n)\phi^m(x) = \psi^{-1}(bx^2)\phi^m(x) = \psi^{-1}(b)\phi^m(x).$$

For $x \notin 1 + \mathcal{P}^m$ we have $\omega_{\psi^{-1}}(n)\phi^m(x) = \psi^{-1}(bx^2)\phi^m(x) = 0$. The first assertion follows.

(2) For $\bar{n} \in \overline{N}_m$ we can write $\bar{n} = w^{-1}n(b)w$ with $b \in \mathcal{P}^{3m}$. Let $\phi' = \omega_{\psi^{-1}}(w)\phi^m$. We have

$$\begin{aligned}\phi'(x) &= \gamma(\psi^{-1}) \int_F \phi^m(y)\psi^{-1}(2xy)dy \\ &= \gamma(\psi^{-1})\psi^{-1}(2x) \int_{\mathcal{P}^m} \psi^{-1}(2xz)dz \\ &= \gamma(\psi^{-1})\psi^{-1}(2x)\,\mathrm{vol}(\mathcal{P}^m)\,\mathrm{Char}(\mathcal{P}^{-m})(x),\end{aligned}$$

where $\mathrm{Char}(\mathcal{P}^{-m})$ denotes the characteristic function of the set $\mathcal{P}^{-m}$. It is clear that $\omega_{\psi^{-1}}(n(b))\phi' = \phi'$. Thus we have

$$\omega_{\psi^{-1}}(\bar{n})\phi^m = \omega_{\psi^{-1}}(w^{-1}n(b))\phi' = \omega_{\psi^{-1}}(w^{-1})\phi' = \omega_{\psi^{-1}}(w^{-1})\omega_{\psi^{-1}}(w)\phi^m = \phi^m.$$

This completes the proof. □

Given a quasicharacter η of $F^\times$ recall that we have defined a local gamma factor $\gamma(s, \pi, \eta, \psi)$ in Proposition 2.2.

Theorem 3.10. *Suppose that the residue characteristic of F is not 2 and ψ is a nontrivial additive character of F. Let (π, V_π) and $(\pi', V_{\pi'})$ be two ψ-generic representations of $\mathrm{SL}_2(F)$ with the same central character.*

(1) *If* $\gamma(s, \pi, \eta, \psi) = \gamma(s, \pi', \eta, \psi)$ *for all quasicharacters* η *of* $F^\times$, *then* $\pi \cong \pi'$.

(2) *There is an integer* $l = l(\pi, \pi')$ *such that if* η *is quasicharacter of* $F^\times$ *with conductor* $\mathrm{cond}(\eta) > l$, *then*

$$\gamma(s, \pi, \eta, \psi) = \gamma(s, \pi', \eta, \psi).$$

Remark. Theorem 3.10 can be viewed as one example of a general local converse conjecture for classical groups, see [Jiang 2006, Conjecture 3.7] or [Jiang and Nien 2013, Conjecture 6.3].

Proof. We will first treat the case where ψ is unramified and prove the general case at the end. We fix the notations $v \in V_\pi$, $v' \in V_{\pi'}$ and L as before.

Let η be a quasicharacter of $F^\times$. We take an integer $m \geq \max\{6L, \mathrm{cond}(\eta)\}$ and consider the Howe vectors v_m and v'_m. Additionally, we take an integer $i \geq \max\{i_2(\eta), I(N_m, \eta), m\}$. In particular we have a section $f_s^i \in I(s, \eta, \psi)$ as in Section 3C. Let $W_m = W_{v_m}$ or $W_{v'_m}$. We compute the integral of $\Psi(W_m, \phi^m, f_s^i)$ on the open dense subset $T\overline{N}(F) = N(F) \setminus N(F)T\overline{N}(F)$ of $N(F) \setminus \mathrm{SL}_2(F)$. For $g = nt(a)\bar{n}$ we can take the quotient measure as $dg = |a|^{-2} d\bar{n} da$. By the definition of f_s^i we get

$$\begin{aligned}
\Psi(W_m, \phi^m, f_s^i) &= \int_{T \times \overline{N}(F)} W_m(t(a)\bar{n})(\omega_{\psi^{-1}}(t(a)\bar{n})\phi^m)(1) f_s^i(t(a)\bar{n})|a|^{-2} d\bar{n} da \\
&= \int_{T \times \overline{N}_i} W_m(t(a)\bar{n})\mu_{\psi^{-1}}(a)|a|^{1/2}\omega_{\psi^{-1}}(\bar{n}) \\
&\qquad\qquad \cdot \phi^m(a)\mu_{\psi^{-1}}(a)\eta_{s+1/2}(a)|a|^{-2} d\bar{n} da \\
&= \int_{T \times \overline{N}_i} W_m(t(a)\bar{n})\omega_{\psi^{-1}}(\bar{n})\phi^m(a)\chi(a)\eta_{s-1}(a) d\bar{n} da,
\end{aligned}$$

where $\chi(a) = \mu_{\psi^{-1}}(a)\mu_{\psi^{-1}}(a) = (a, -1)_F$. Since $i \geq m$ we get $\overline{N}_i \subset \overline{N}_m$. By Lemmas 3.1 and 3.9 we get $W_m(t(a)\bar{n}) = W_m(t(a))$ and $\omega_{\psi^{-1}}(\bar{n})\phi^m = \phi^m$. Thus we get

$$\Psi(W_m, \phi^m, f_s^i) = q^{-3i} \int_{F^\times} W_m(t(a))\phi^m(a)\chi(a)\eta_{s-1}(a) da.$$

Since $\phi^m = \mathrm{Char}(1 + \mathcal{P}^m)$ and, for $a \in 1 + \mathcal{P}^m$, we have $W_m(t(a)) = 1$. By Lemma 3.1 we get

$$\Psi(W_m, \phi^m, f_s^i) = q^{-3i} \int_{1+\mathcal{P}^m} \chi(a)\eta(a) da.$$

Since $\chi(a) = 1$ for $a \in 1 + \mathcal{P}^m$ and $m \geq \mathrm{cond}(\eta)$ by assumption we get

$$\Psi(W_m, \phi^m, f_s^i) = q^{-3i-m}.$$

The above calculation works for both W_{v_m} and $W_{v'_m}$. Thus we have

$$\Psi(W_{v_m}, \phi^m, f_s^i) = \Psi(W_{v'_m}, \phi^m, f_s^i) = q^{-3i-m}. \tag{3-2}$$

Next we compute the other side of the local functional equation, $\Psi(W_m, \phi^m, \tilde{f}_s^i)$, on the open dense subset $N(F)\backslash N(F)TwN(F) \subset N(F)\backslash \mathrm{SL}_2(F)$, where $\tilde{f}_s^i = M_s(f_s^i)$. We have

$$\begin{aligned}
&\Psi(W_m, \phi^m, \tilde{f}_s^i)\\
&\quad = \int_{T\times N(F)} W_m(t(a)wn)(\omega_{\psi^{-1}}(t(a)wn)\phi^m)(1)\tilde{f}_s^i(t(a)wn)|a|^{-2}dnda\\
&\quad = \int_{T\times N_m} W_m(t(a)wn)(\omega_{\psi^{-1}}(t(a)wn)\phi^m)(1)\tilde{f}_s^i(t(a)wn)|a|^{-2}dnda\\
&\qquad + \int_{T\times (N(F)-N_m)} W_m(t(a)wn)(\omega_{\psi^{-1}}(t(a)wn)\phi^m)(1)\tilde{f}_s^i(t(a)wn)|a|^{-2}dnda.
\end{aligned}$$

By Lemma 3.6 we get $W_{v_m}(t(a)wn) = W_{v'_m}(t(a)wn)$ for all $n \in N(F) - N_m$. Thus

$$\begin{aligned}
&\Psi(W_{v_m}, \phi^m, \tilde{f}_s^i) - \Psi(W_{v'_m}, \phi^m, \tilde{f}_s^i)\\
&\qquad = \int_{T\times N_m} (W_{v_m}(t(a)wn) - W_{v'_m}(t(a)wn))(\omega_{\psi^{-1}}(t(a)wn)\phi^m)(1)\\
&\qquad\qquad\qquad \cdot \tilde{f}_s^i(t(a)wn)|a|^{-2}dnda.
\end{aligned}$$

Since $i \geq I(N_m, \eta)$ we get

$$\tilde{f}_s^i(t(a)wn) = \mu_{\psi^{-1}}(a)\eta_{3/2-s}^{-1}(a)q_F^{-3i}$$

by Lemma 3.8. On the other hand, by Lemma 3.1 and Lemma 3.9, for $n \in N_m$ we get

$$\begin{gathered}
W_m(t(a)wn) = \psi(n)W_m(t(a)w),\\
(\omega_{\psi^{-1}}(t(a)wn)\phi^m)(1) = \psi^{-1}(n)(\omega_{\psi^{-1}}(t(a)w)\phi^m)(1).
\end{gathered}$$

Thus

$$\begin{aligned}
&\Psi(W_{v_m}, \phi^m, \tilde{f}_s^i) - \Psi(W_{v'_m}, \phi^m, \tilde{f}_s^i)\\
&\qquad = q_F^{-3i+m}\int_T (W_{v_m}(t(a)w) - W_{v'_m}(t(a)w))(\omega_{\psi^{-1}}(w)\phi^m)(a)\\
&\qquad\qquad \cdot \chi(a)\eta^{-1}(a)|a|^{-s}da.
\end{aligned} \tag{3-3}$$

By (3-2), (3-3) and the local functional equation we get

$$\begin{aligned}
&q^{-2m}(\gamma(s, \pi, \eta, \psi) - \gamma(s, \pi', \eta, \psi))\\
&\qquad = \int_{F^\times} (W_{v_m}(t(a)w) - W_{v'_m}(t(a)w))(\omega_{\psi^{-1}}(w)\phi^m)(a)\chi(a)\eta^{-1}(a)|a|^{-s}da.
\end{aligned} \tag{3-4}$$

Let $k = 4L$. Since $m \geq 6L > k$, by Lemmas 3.1 and 3.6, we get

$$\begin{aligned} W_{v_m}(t(a)w) - W_{v'_m}(t(a)w) &= \frac{1}{\mathrm{vol}(N_m)} \int_{N_m} (W_{v_k}(t(a)wn) - W_{v'_k}(t(a)wn))\psi^{-1}(n)dn \\ &= \frac{1}{\mathrm{vol}(N_m)} \int_{N_k} (W_{v_k}(t(a)wn) - W_{v'_k}(t(a)wn))\psi^{-1}(n)dn \\ &= \frac{\mathrm{vol}(N_k)}{\mathrm{vol}(N_m)} (W_{v_k}(t(a)w) - W_{v'_k}(t(a)w)) \\ &= q^{k-m}(W_{v_k}(t(a)w) - W_{v'_k}(t(a)w)). \end{aligned}$$

Now we can rewrite (3-4) as

$$\begin{aligned} \text{(3-5)} \quad & q^{-m-k}(\gamma(s, \pi, \eta, \psi) - \gamma(s, \pi', \eta, \psi)) \\ &= \int_{F^\times} (W_{v_k}(t(a)w) - W_{v'_k}(t(a)w))(\omega_{\psi^{-1}}(w)\phi^m)(a)\chi(a)\eta^{-1}(a)|a|^{-s}da. \end{aligned}$$

By Lemma 3.2, if $a \notin \mathcal{P}^{-6L}$, i.e., $a^2 \notin \mathcal{P}^{-3k}$, we get $W_{v_k}(t(a)w) = 0 = W_{v'_k}(t(a)w)$. Thus the integral on the right side of formula (3-5) can be taken over $\mathcal{P}^{-6L}$. For $a \in \mathcal{P}^{-6L}$ and $m \geq 6L$ (as we assumed), by the calculation given in the proof of Lemma 3.9, we have

$$\begin{aligned} (\omega_{\psi^{-1}}(w)\phi^m)(a) &= \gamma(\psi^{-1})\psi^{-1}(2a)\,\mathrm{vol}(\mathcal{P}^m)\,\mathrm{Char}(\mathcal{P}^{-m})(a) \\ &= \gamma(\psi^{-1})\psi^{-1}(2a)q^{-m}. \end{aligned}$$

Plugging this into (3-5) we get

$$\begin{aligned} \text{(3-6)} \quad & q^{-k}\gamma(\psi^{-1})^{-1}(\gamma(s, \pi, \eta, \psi) - \gamma(s, \pi', \eta, \psi)) \\ &= \int_{F^\times} (W_{v_k}(t(a)w) - W_{v'_k}(t(a)w))\psi^{-1}(2a)\chi(a)\eta^{-1}(a)|a|^{-s}da. \end{aligned}$$

Now we can prove our theorem. We consider (1) first. Suppose $\gamma(s, \pi, \eta, \psi) = \gamma(s, \pi', \eta, \psi)$ for all quasicharacters η of $F^\times$. Then we get

$$\int_{F^\times} (W_{v_k}(t(a)w) - W_{v'_k}(t(a)w))\psi^{-1}(2a)\chi(a)\eta^{-1}(a)|a|^{-s}da = 0$$

for all quasicharacters η.

We rewrite the equality as

$$\begin{aligned} 0 &= \int_{F^\times} (W_{v_k}(t(a)w) - W_{v'_k}(t(a)w))\psi^{-1}(2a)\chi(a)\eta^{-1}(a)|a|^{-s}da \\ &= \sum_{m=-\infty}^{\infty} \int_{|a|=q^m} (W_{v_k}(t(a)w) - W_{v'_k}(t(a)w))\psi^{-1}(2a)\chi(a)\eta^{-1}(a)da\,q^{-ms}. \end{aligned}$$

It follows that all the coefficients in the above Laurent series in q^s have to be zero. So

$$\int_{|a|=q^m} (W_{v_k}(t(a)w) - W_{v'_k}(t(a)w))\psi^{-1}(2a)\chi(a)\eta^{-1}(a)da = 0 \tag{3-7}$$

for all quasicharacters η.

Since the set $\{a \in F^\times : |a| = q^m\}$ is compact open in $F^\times$, the left side of equation (3-7) can be viewed as Mellin transform of a compactly supported function on $F^\times$. By the inverse Mellin transform we get

$$(W_{v_k}(t(a)w) - W_{v'_k}(t(a)w))\psi^{-1}(2a) = 0,$$

or

$$W_{v_k}(t(a)w) = W_{v'_k}(t(a)w).$$

By Lemmas 3.1 and 3.6, Corollary 3.5 and the Bruhat decomposition $SL_2(F) = B \cup BwB$ we get

$$W_{v_k}(g) = W_{v'_k}(g)$$

for all $g \in SL_2(F)$. By the uniqueness of Whittaker model we get $\pi \cong \pi'$. This proves (1).

Next we consider (2). Let $l = l(\pi, \pi')$ be an integer such that $l \geq 6L$, then

$$W_{v_k}(t(a_0a)w) = W_{v_k}(t(a)w) \quad \text{and} \quad W_{v'_k}(t(a_0a)w) = W_{v'_k}(t(a)w)$$

for all $a_0 \in 1 + \mathcal{P}^l$ and all $a \in \mathcal{P}^{-6L}$. Such an l exists because the functions $a \mapsto W_{v_k}(t(a)w)$ and $a \mapsto W_{v'_k}(t(a)w)$ on $\mathcal{P}^{-6L} \subset F^\times$ are continuous. Note that $k = 4L$ and L only depends on the choices of v and v'. On the other hand, for $a \in \mathcal{P}^{-6L}$, it is easy to see that

$$\psi^{-1}(2a_0a) = \psi^{-1}(2a) \quad \text{for all } a_0 \in 1 + \mathcal{P}^l,$$

since $l \geq 6L$. It is also clear that $\chi(a_0a) = \chi(a)$ for all $a_0 \in 1 + \mathcal{P}^l$, since the character χ is unramified. As we noted before, the integrand of the right side integral of (3-6) has support in $\mathcal{P}^{-6L}$. Let η be a quasicharacter of $F^\times$ with $\mathrm{cond}(\eta) > l$. Then it is clear that the integral of the right side of (3-6) vanishes. Thus we get

$$\gamma(s, \pi, \eta, \psi) = \gamma(s, \pi', \eta, \psi).$$

This finishes the proof when ψ is unramified.

Now let us consider the general case when ψ is ramified. The proof is essentially the same as the unramified case. We will indicate the necessary changes in the above proof. If ψ has conductor c, i.e., $\psi(\mathcal{P}_F^c) = 1$ but $\psi(\mathcal{P}_F^{c-1}) \neq 1$, we define

$d_m = \mathrm{diag}(p_F^{-2m+c}, 1) \in \mathrm{GL}_2(F)$ and $J_m = d_m K_m d_m^{-1}$. Then

$$J_m = \begin{pmatrix} 1+\mathcal{P}_F^m & \mathcal{P}_F^{-m+c} \\ \mathcal{P}_F^{3m-c} & 1+\mathcal{P}_F^m \end{pmatrix} \cap \mathrm{SL}_2(F).$$

For $j = (j_{il})_{1 \le i,l \le 2} \in J_m$ we define $\psi_m(j) = \psi(j_{12})$. It is clear that ψ_m defines a character of J_m. Given a ψ-generic representation (π, V) of $\mathrm{SL}_2(F)$ and a vector $v \in V$ we define v_m in the same way as before, i.e., by (3-1). In this case, we fix an integer L such that $L \ge c$ and v is fixed by K_L. We call $\{v_m\}_{m \ge L}$ the Howe vectors. We note that in the proof of Lemma 3.8, we used that ψ is unramified to make sure $\mu_{\psi^{-1}}$ is trivial on $\mathcal{O}_F^\times$. If ψ is ramified, by continuity, $\mu_{\psi^{-1}}$ is trivial on $1+\mathcal{P}_F^i$ for i large. This is all what we need in the proof of Lemma 3.8 to extend it to the ramified case. Now one can check easily that all of the above proofs go through and we get the theorem in general. □

4. A strong multiplicity one theorem

Let F be a number field and $\mathbb{A}$ be its adele ring.

4A. ***Global genericity.*** In this subsection we discuss the relation between global genericity and local genericity. Let φ be a cusp form on $\mathrm{SL}_2(F) \setminus \mathrm{SL}_2(\mathbb{A})$. Since the group $N(F) \setminus N(\mathbb{A})$ is compact and abelian we have the Fourier expansion

$$\varphi(g) = \sum_{\psi \in \widehat{N(F)\setminus N(\mathbb{A})}} W_\varphi^\psi(g),$$

where

$$W_\varphi^\psi(g) = \int_{N(F)\setminus N(\mathbb{A})} \varphi(ng)\psi^{-1}(n)dg.$$

Since φ is a cusp form we get $W_\varphi^{\psi_0} \equiv 0$, where ψ_0 is the trivial character of $F \setminus \mathbb{A}$. Thus we get

$$\varphi(g) = \sum_{\substack{\psi \in \widehat{N(F)\setminus N(\mathbb{A})} \\ \psi \ne \psi_0}} W_\varphi^\psi(g).$$

Fix a nontrivial additive character ψ of $N(F) \setminus N(\mathbb{A})$. Then

$$(\widehat{N(F) \setminus N(\mathbb{A})}) \setminus \{\psi_0\} = \{\psi_\kappa : \kappa \in F^\times\},$$

where $\psi_\kappa(a) = \psi(\kappa a)$ and $a \in \mathbb{A}$. If $\kappa \in F^{\times,2}$, say $\kappa = a^2$, we have

$$W_\varphi^{\psi_\kappa}(g) = W_\varphi^\psi(t(a)g).$$

Thus we get

$$\varphi(g) = \sum_{\kappa \in F^\times/F^{\times,2}} \sum_{a \in F^\times} W_\varphi^{\psi_\kappa}(t(a)g).$$

Corollary 4.1. *If φ is a nonzero cusp form, there exists $\kappa \in F^\times$ such that*

$$W_\varphi^{\psi_\kappa} \neq 0.$$

Let (π, V_π) be a cuspidal automorphic representation of $\mathrm{SL}_2(F) \backslash \mathrm{SL}_2(\mathbb{A})$. We say π is ψ_κ-generic if there exists $\varphi \in V_\pi$ such that

$$W_\varphi^{\psi_\kappa} \not\equiv 0.$$

Corollary 4.2. *Let π be a cuspidal automorphic representation of $\mathrm{SL}_2(F)\backslash\mathrm{SL}_2(\mathbb{A})$ and ψ be a nontrivial additive character of $F \backslash \mathbb{A}$. Then there exists $\kappa \in F^\times$ such that π is ψ_κ-generic.*

Theorem 4.3. *Let $\pi = \otimes_v \pi_v$ be an irreducible cuspidal automorphic representation of $\mathrm{SL}_2(\mathbb{A})$ and $\psi = \otimes \psi_v$ be a nontrivial additive character of $F \backslash \mathbb{A}$. Then π is ψ-generic if and only if each π_v is ψ_v-generic.*

Proof. A similar result is proved for $\mathrm{U}(1, 1)$ by Gelbart, Rogawski and Soudry [1997, Proposition 2.5].

It is clear that global genericity implies local genericity. Now we consider the other direction. We assume each π_v is ψ_v-generic.

We assume π is ψ_κ-generic for some $\kappa \in F^\times$, i.e., there exists $\varphi \in V_\pi$ such that

$$W_\varphi^{\psi_\kappa}(g) = \int_{N(F)\backslash N(\mathbb{A})} \varphi(ng)\psi_\kappa^{-1}(n)dn \neq 0.$$

Then π_v is also $\psi_{\kappa,v}$-generic, where $\psi_{\kappa,v}(a) = \psi_v(\kappa a)$. By Proposition 2.1 we get $\pi_v \cong \pi_v^\kappa$.

For $\varphi \in V_\pi$ consider the function $\varphi^\kappa(g) = \varphi(g^\kappa)$, where g^k is defined by

$$g^\kappa = \mathrm{diag}(\kappa, 1)g\, \mathrm{diag}(\kappa^{-1}, 1).$$

Then

$$\begin{aligned}\int_{N(F)\backslash N(\mathbb{A})} \varphi^\kappa(ng)dn &= \int_{N(F)\backslash N(\mathbb{A})} \varphi((ng)^\kappa)dn = \int_{N(F)\backslash N(\mathbb{A})} \varphi(n^\kappa g^\kappa)dn \\ &= \int_{N(F)\backslash N(\mathbb{A})} \varphi(ng^\kappa)dn = 0,\end{aligned}$$

hence φ^κ is also a cusp form. Let V_π^κ be the space which consists of functions of the form φ^κ for all $\varphi \in V_\pi$. Let π^κ denote the cuspidal automorphic representation of $\mathrm{SL}_2(\mathbb{A})$ on V_π^κ.

Claim. $(\pi^\kappa)_v = \pi_v^\kappa$.

Proof. Let $\Lambda : V_\pi \to \mathbb{C}$ be a nonzero ψ_κ-Whittaker functional for π and let Λ_v be a nonzero $(\psi_\kappa)_v$-Whittaker functional on V_{π_v} satisfying that if $\varphi = \otimes_v \varphi_v$ is a pure tensor, then

$$\Lambda(\pi(g)\varphi) = \prod_v \Lambda_v(\pi_v(g_v)\varphi_v).$$

Note that Λ is in fact given by

$$\Lambda(\varphi) = \int_{N(F)\backslash N(\mathbb{A})} \varphi(n)\psi_\kappa^{-1}(n)dn.$$

Then the ψ_{κ^2}-Whittaker functional of π^κ is given by

$$\int_{N(F)\backslash N(\mathbb{A})} \varphi^\kappa(n)\psi_{\kappa^2}^{-1}(n)dn.$$

This means that if $W_\varphi(g)$ is a ψ_κ-Whittaker function of π, then $W_{\varphi^\kappa}(g) = W_\varphi(g^\kappa)$ is a ψ_{κ^2}-Whittaker function of π^κ.

Hence, with $\varphi = \otimes_v\varphi_v$ a pure tensor, we have $W_\varphi(g) = \prod_v W_{\varphi_v}(g_v)$ and $\{W_{\varphi_v}(g_v)\}$ is the Whittaker model of π_v, while $W_{\varphi^\kappa}(g) = W_\varphi(g^\kappa) = \prod_v W_{\varphi_v}(g_v^\kappa)$ and $\{W_{\varphi_v}(g_v^\kappa)\}$ is the Whittaker model of $(\pi^\kappa)_v$. Now $W_v(g_v) \to W_v(g_v^\kappa)$ gives an isomorphism between π_v^κ and $(\pi^\kappa)_v$, which proves the claim. □

Now let us continue the proof of the theorem. By the claim we have $\pi_v \cong (\pi^\kappa)_v$ or $\pi \cong \pi^\kappa$. By the multiplicity one theorem for SL_2 of Ramakrishnan [2000] we get $\pi = \pi^\kappa$. Since π is ψ_κ-generic we get that π^κ is ψ_{κ^2}-generic and hence ψ-generic. Since $\pi = \pi^\kappa$ the theorem follows. □

4B. ***Eisenstein series on*** $\widetilde{\mathbf{SL}}_2(\mathbb{A})$**.** Let $\widetilde{\mathrm{SL}}_2(\mathbb{A})$ be the double cover of $\mathrm{SL}_2(\mathbb{A})$. It is well-known that $\mathrm{SL}_2(F)$ splits over the projection $\widetilde{\mathrm{SL}}_2(\mathbb{A}) \to \mathrm{SL}_2(\mathbb{A})$. Let μ_ψ be the genuine character of $T(F) \setminus \widetilde{T}(\mathbb{A})$ whose local components are μ_{ψ_v} as given in §2.

Let η be a quasicharacter of $F^\times \setminus \mathbb{A}^\times$ and $s \in \mathbb{C}$. We consider the induced representation

$$I(s, \chi, \psi) = \mathrm{Ind}_{\widetilde{B}(\mathbb{A})}^{\widetilde{\mathrm{SL}}_2(\mathbb{A})}(\mu_\psi\eta_{s-1/2}).$$

For $f_s \in I(s, \eta, \psi)$ we consider the Eisenstein series $E(g, f_s)$ on $\widetilde{\mathrm{SL}}_2(\mathbb{A})$:

$$E(g, f_s) = \sum_{B(F)\backslash \mathrm{SL}_2(F)} f_s(\gamma g),\ g \in \widetilde{\mathrm{SL}}_2(\mathbb{A}).$$

The above sum is absolutely convergent when $\mathrm{Re}(s) \gg 0$ and can be meromorphically continued to the whole s-plane.

There is an intertwining operator $M_s = M_s(\eta) : I(s, \eta, \psi) \to I(1-s, \eta^{-1}, \psi)$ with

$$M_s(f_s)(g) = \int_{N(F)\backslash N(\mathbb{A})} f_s(wng)dn.$$

The above integral is absolutely convergent for $\mathrm{Re}(s) \gg 0$ and defines a meromorphic function of $s \in \mathbb{C}$.

Proposition 4.4. (1) *If $\eta^2 \neq 1$, then the Eisenstein series $E(g, f_s)$ is holomorphic for all s. If $\eta^2 = 1$, the only possible poles of $E(g, f_s)$ are at $s = 0$ and $s = 1$. Moreover, the order of the poles are at most* 1.

(2) *We have the functional equation*

$$E(g, f_s) = E(g, M_s(f_s)) \quad \textit{and} \quad M_s(\eta) \circ M_{1-s}(\eta^{-1}) = 1.$$

See [Gan et al. 2014, Proposition 6.1] for example.

4C. *The global zeta-integral.* Let ψ be a nontrivial additive character of $F \setminus \mathbb{A}$. Then there is a global Weil representation ω_ψ of $\widetilde{\mathrm{SL}}_2(\mathbb{A})$ on $\mathcal{S}(\mathbb{A})$. For $\phi \in \mathcal{S}(\mathbb{A})$ we consider the theta series

$$\theta_\psi(\phi)(g) = \sum_{x \in F} (\omega_\psi(g)\phi)(x).$$

It is well-known that θ_ψ defines an automorphic form on $\widetilde{\mathrm{SL}}_2(\mathbb{A})$.

Let (π, V_π) be a ψ-generic cuspidal automorphic representation of $\mathrm{SL}_2(\mathbb{A})$. For $\varphi \in V_\pi$, $\phi \in \mathcal{S}(\mathbb{A})$ and $f_s \in I(s, \eta, \psi^{-1})$ consider the integral

$$Z(\varphi, \theta_{\psi^{-1}}(\phi), E(\,\cdot\,, f_s)) = \int_{\mathrm{SL}_2(F)\backslash \mathrm{SL}_2(\mathbb{A})} \varphi(g)\theta_{\psi^{-1}}(\phi)(g)E(g, f_s)dg. \tag{4-1}$$

Proposition 4.5 [Gelbart et al. 1987, Theorem 4.C]. *For* $\mathrm{Re}(s) \gg 0$, *the integral $Z(\varphi, \theta_{\psi^{-1}}(\phi), E(\,\cdot\,, f_s))$ is absolutely convergent and*

$$Z(\varphi, \theta_{\psi^{-1}}(\phi), E(\,\cdot\,, f_s)) = \int_{N(\mathbb{A})\backslash \mathrm{SL}_2(\mathbb{A})} W_\varphi^\psi(g)(\omega_{\psi^{-1}}(g))\phi(1) f_s(g)dg,$$

where $W_\varphi^\psi(g) = \int_{N(F)\backslash N(\mathbb{A})} \varphi(ng)\psi^{-1}(n)dn$ is the ψ-th Whittaker coefficient of φ.

Corollary 4.6. *We take $\varphi = \otimes\varphi_v$, $\phi = \otimes_v \phi_v$ and $f_s = \otimes f_{s,v}$ to be pure tensors. Let S be a finite set of places such that for all $v \notin S$, v is finite and π_v, ψ_v, $f_{s,v}$ are unramified. Then for* $\mathrm{Re}(s) \gg 0$ *we have*

$$Z(\varphi, \theta_{\psi^{-1}}(\phi), E(\,\cdot\,, f_s)) = \prod_{v \in S} \Psi(W_{\varphi_v}, \phi_v, f_{s,v}) \frac{L^S(s, \pi, St \otimes (\chi\eta))}{L^S(2s, \eta^2)},$$

where χ is the character of $F^\times \setminus \mathbb{A}^\times$ defined by

$$\chi((a_v)) = \prod_v (a_v, -1)_{F_v}, \qquad (a_v)_v \in \mathbb{A}^\times.$$

Moreover, we have the following functional equation

$$Z(\varphi, \theta_{\psi^{-1}}(\phi), E(\,\cdot\,, f_s)) = Z(\varphi, \theta_{\psi^{-1}}(\phi), E(\,\cdot\,, M_s(f_s))).$$

This follows directly from Proposition 2.2, the unramified calculation, and the functional equation of Eisenstein series in Proposition 4.4.

Corollary 4.7. (1) *The partial L-function $L^S(s, \pi, St \otimes \chi\eta)$ can be extended to a meromorphic function of s.*

(2) *If $\eta^2 \neq 1$, then $L^S(s, \pi, St \otimes \chi\eta)$ is holomorphic for* $\mathrm{Re}(s) > 1/2$.

(3) *If $\eta^2 = 1$, then, on the region* $\mathrm{Re}(s) > 1/2$, *the only possible pole of the function $L^S(s, \pi, St \otimes \chi\eta)$ is at $s = 1$. Moreover, the order of the pole of $L^S(s, \pi, St \otimes (\chi\eta))$ at $s = 1$ is at most* 1.

(4) *Let S_∞ be the set of infinity places of F, then we can find data $\varphi_v \in V_{\pi_v}$, $\phi_v \in \mathcal{S}(F_v)$ and $f_{s,v} \in I(s, \eta_v, \psi_v)$ for $v \in S_\infty$ such that*

$$\frac{L^S(s, \pi, St \otimes (\chi\eta))}{L^S(1-s, \pi, St \otimes (\chi\eta^{-1}))} = \prod_{v \in S_\infty} \frac{\Psi(W_{\varphi_v}, \phi_v, M_s(f_{s,v}))}{\Psi(W_{\varphi_v}, \phi_v, f_{s,v})} \cdot \prod_{v \in S - S_\infty} \gamma(s, \pi_v, \eta_v, \psi_v) \cdot \frac{L^S(2s-1, \eta^2)}{L^S(2-2s, \eta^{-2})},$$

where S is a large enough finite set of places which contains S_∞, all finite places v such that $v|2$ and all finite places such that our data is ramified. Here $\gamma(s, \pi_v, \eta_v, \psi_v)$ is the local gamma factors defined in Proposition 2.2.

Proof. By Proposition 4.4 and Corollary 4.6 to prove (1)-(3) it suffices to show that, for each place v and for any fixed point $s \in \mathbb{C}$, we can choose the data $(W_v, \phi_v, f_{s,v})$ such that $\Psi(W_{\varphi_v}, \phi_v, f_{s,v}) \neq 0$. If v is nonarchimedean this is shown in the proof of Theorem 3.10, see equation (3-2). We will prove the general case later, see Lemma 4.9. We now consider (4). For $s \in \mathbb{C}$ with $\mathrm{Re}(s) \gg 0$ we choose data $\varphi = \otimes\varphi_v$, $\phi = \otimes\phi_v$ and $f_s = \otimes f_{s,v}$ such that $\Psi(W_{\varphi_v}, \phi_v, f_{s,v}) \neq 0$ for each $v \in S$ and φ_v, ϕ_v, $f_{s,v}$ and ψ_v are unramified for $v \notin S$. By the Remark at the end of §2, for $v \notin S$, we have

$$M_s(f_{s,v}) = \frac{L(2s-1, \eta_v^2)}{L(2s, \eta_v^2)} f_{1-s,v}.$$

Thus, by Corollary 4.6, for $\mathrm{Re}(s) \ll 0$ we have

$$Z(\varphi, \theta_{\psi^{-1}}(\phi), E(\,\cdot\,, M_s(f_s))) = \prod_{v \in S} \Psi(W_{\varphi_v}, \phi_v, M_s(f_{s,v})) \frac{L^S(1-s, \pi, St \otimes (\chi\eta^{-1}))}{L^S(2-2s, \eta^{-2})} \cdot \frac{L^S(2s-1, \eta^2)}{L^S(2s, \eta^2)}.$$

Note that the above equation also holds after meromorphic continuation. Now (4) follows from Corollary 4.6 and Proposition 2.2 directly. □

4D. ***A strong multiplicity one theorem.*** With the above preparation, we are now ready to prove the main global result of this paper.

Theorem 4.8. *Let $\pi = \otimes\pi_v$ and $\pi' = \otimes\pi'_v$ be two irreducible cuspidal automorphic representations of* $\mathrm{SL}_2(\mathbb{A})$ *with the same central character. Suppose that π and π' are both ψ-generic. Let S be a finite set of* ***finite*** *places such that no place in S is above* 2. *If $\pi_v \cong \pi'_v$ for all $v \notin S$, then $\pi = \pi'$.*

Proof. The following argument follows from the proof of [Casselman 1973, Theorem 2, p. 307].

Let S_1 be a large finite set of places which contains $S_\infty \cup S$. Since $\pi_v \cong \pi'_v$ for all $v \notin S$, we have $L^{S_1}(s, \pi, St \otimes (\chi\eta)) = L^{S_1}(s, \pi', St \otimes (\chi\eta))$ and $L^{S_1}(1 - s, \pi, St \otimes (\chi\eta^{-1})) = L^{S_1}(1 - s, \pi, St \otimes (\chi\eta^{-1}))$. Thus, by Corollary 4.7 (4), for each quasicharacter η, we can find data $\varphi_v \in V_{\pi_v}$, $\phi_v \in \mathcal{S}(F_v)$ and $f_{s,v}$ for $v \in S_\infty$ such that

$$\prod_{v\in S_\infty} \frac{\Psi(W_{\varphi_v}, \phi_v, M_s(f_{s,v}))}{\Psi(W_{\varphi_v}, \phi_v, f_{s,v})} \cdot \prod_{v\in S_1 - S_\infty} \gamma(s, \pi_v, \eta_v, \psi_v) = \prod_{v\in S_\infty} \frac{\Psi(W_{\varphi'_v}, \phi_v, M_s(f_{s,v}))}{\Psi(W_{\varphi'_v}, \phi_v, f_{s,v})} \cdot \prod_{v\in S_1 - S_\infty} \gamma(s, \pi'_v, \eta_v, \psi_v),$$

where φ'_v is the image of φ_v under a fixed isomorphism $\pi_v \cong \pi'_v$ for $v \in S_\infty$. Since $\pi_v \cong \pi'_v$ for $v \in S_1 - S$, we get

$$\prod_{v\in S} \gamma(s, \pi_v, \eta_v, \psi_v) = \prod_{v\in S} \gamma(s, \pi'_v, \eta_v, \psi_v).$$

Fix $v_0 \in S$. By [Jacquet and Langlands 1970, Lemma 12.5], given an arbitrary character η_{v_0}, we can find a character η of $\mathbb{A}^\times$ which restricted to v_0 is η_{v_0} and has arbitrarily high conductor at the other places of S. By Theorem 3.10 (2) we conclude that

$$\gamma(s, \pi_{v_0}, \eta_{v_0}, \psi_{v_0}) = \gamma(s, \pi'_{v_0}, \eta_{v_0}, \psi_{v_0})$$

for all characters η_{v_0}. Thus, by Theorem 3.10 (1), we conclude that $\pi_{v_0} \cong \pi'_{v_0}$. This applies also to the other places of S. Thus we proved that $\pi_v \cong \pi'_v$ for all places v. Now the theorem follows from the multiplicity one theorem for SL_2 of [Ramakrishnan 2000]. □

Remark. We expect that the restriction about residue characteristics on the finite set S in Theorem 4.8 can be removed.

Finally, we prove a nonvanishing result about the archimedean local zeta-integrals which is used in the above proof. We formulate and prove the result both for the p-adic and the archimedean cases simultaneously.

Lemma 4.9. *Let F be a local field, ψ be a nontrivial additive character of F, η be a quasicharacter of $F^\times$ and π be a ψ-generic representation of* $\mathrm{SL}_2(F)$. *Then there exists $W \in \mathcal{W}(\pi, \psi)$, $\phi \in \mathcal{S}(F)$ and $f_s \in \mathrm{Ind}_{\widetilde{B}}^{\widetilde{\mathrm{SL}}_2(F)}(\eta_{s-1/2}\mu_\psi)$ such that*

$$\Psi(W, \phi, f_s) = \int_{N(F)\backslash \mathrm{SL}_2(F)} W(h)(\omega_{\psi^{-1}}\phi)(h) f_s(h) \neq 0.$$

Proof. We note that the Bruhat cell $\Omega = N(F)TwN(F)$ is open and dense in $\mathrm{SL}_2(F)$. Thus the above integral is reduced to

$$\Psi(W, \phi, f_s) = \int_{TN(F)} W(wt(a)n(u))(\omega_{\psi^{-1}}(wt(a)n(u))\phi)(1) f_s(wt(a)n(u))\Delta(a)\,da\,du,$$

where $\Delta(a) = |a|^{-2}$.

Using the formulas for the Weil representation $\omega_{\psi^{-1}}$ we find

$$(\omega_{\psi^{-1}}(wt(a)n(u))\phi)(x) = |a|^{1/2}\frac{\gamma(\psi^{-1})}{\gamma(\psi_a^{-1})}\int_F \psi(ua^2y^2)\phi(ay)\psi(2xy)\,dy = |a|^{1/2}\frac{\gamma(\psi^{-1})}{\gamma(\psi_a^{-1})}\hat{\Phi}_{a,u}(x),$$

where $\Phi_{a,u}(x) = \psi(ua^2x^2)\phi(ax)$ which is again a Schwartz function on F and depends continuously on a and u.

We next explain that the set $\{(g, 1) : g \in N(F)TwN(F)\}$, still denoted as Ω, is open in $\widetilde{\mathrm{SL}}_2(F)$. Note that there is a double covering map $p : \widetilde{\mathrm{SL}}_2(F) \to \mathrm{SL}_2(F)$. For any $(g, 1) \in \Omega$ its projection under p is g. As p is a covering map there exists an open neighborhood U_g of g contained in $N(F)TwN(F)$ such that $p^{-1}(U_g)$ is a disjoint union of two open subsets of $\widetilde{\mathrm{SL}}_2(F)$, each is homeomorphic to U_g by p. Then one component of $p^{-1}(U_g)$ is an open neighborhood of $(g, 1)$ in Ω, which shows that Ω is open in $\widetilde{\mathrm{SL}}_2(F)$.

Now define $f_s \in I(s, \eta, \psi^{-1})$ on the set $\{(g, 1) : g \in \mathrm{SL}_2(F)\}$ by

$$f_s(g) = \begin{cases} \delta(b)^{1/2}(\eta_{s-1/2}\mu_{\psi^{-1}})(b) f_2(u) & \text{if } g = bwn(u) \in \Omega, \\ 0 & \text{otherwise,} \end{cases}$$

where $b \in B(F) = TN(F)$, $u \in F$ and f_2 is a compactly supported function to be determined later. Then we extend the definition of f_2 to the set $\{(g, -1) : g \in \mathrm{SL}_2(F)\}$ to make it genuine, i.e., $f_s(g, -1) = -1 f_s(g, 1)$.

Then the integral Ψ can be reduced further to

$$\Psi(W, \phi, f_s) = \int_{TN(F)} W(wau)|a|^{1/2}\frac{\gamma(\psi^{-1})}{\gamma(\psi_a^{-1})}\hat{\Phi}_{a,u}(1)\delta(a)^{1/2}(\eta_{s-1/2}\mu_{\psi^{-1}})(a) \cdot f_2(u)\Delta(a)\,da\,du. \tag{4-2}$$

Case 1 (F is p-adic). Consider the Howe vector W_{v_m}. By Corollary 3.4, taking m large enough, W_{v_m} can have arbitrarily small compact open support around 1 when restricted to T. Then $W_{w.v_m}(t(a^{-1})w)$ has small compact open support around $a=1$.

First choose ϕ so that $\hat{\Phi}_{a,u}(1) \neq 0$ when $a = 1, u = 0$. Then choose m so that $W_{w.v_m}(wt(a)) = W_{w.v_m}(t(a^{-1})w)$ has small compact support around 1 and all the other data involving a in the integral (*) are nonzero constants. For this $W_{w.v_m}$, consider $W_{w.v_m}(wt(a)u)$ with $u \in N$. When u is close to 1 enough, we have $W_{w.v_m}(wt(a)u) = W_{w.v_m}(wt(a))$ for all a in that small compact support around 1. Then take f_2 with support u close to 1 satisfying the above. With these choices of $W_{w.v_m}(g)$, f_2, ϕ, the integral (4-2) is nonzero.

Case 2 (F is archimedean). We will concentrate on the case $F = \mathbb{R}$. The case $F = \mathbb{C}$ is similar as we have the same formulas for the Weil representation by [Jacquet and Langlands 1970, Proposition 1.3]. We begin with the formulas

$$\begin{aligned}(4\text{-}3)\quad \Psi(W, \phi, f_s) &= \int_{TN(F)} W(wau)|a|^{1/2}\frac{\gamma(\psi^{-1})}{\gamma(\psi_a^{-1})}\hat{\Phi}_{a,u}(1)\delta(a)^{1/2}(\eta_{s-1/2}\mu_{\psi^{-1}})(a) \\ &\qquad\qquad \cdot f_2(u)\Delta(a)\,da\,du,\end{aligned}$$

where $\Phi_{a,u}(x) = \psi(ua^2x^2)\phi(ax)$ is again a Schwartz function, as is ϕ, and it depends on a and u continuously. Since the Fourier transform is an isometry of the Schwartz space we can choose ϕ so that the Fourier transform $\hat{\Phi}_{a,u}(1) > 0$ when $a = 1$ and $u = 0$, and it depends on a and u continuously.

Now let (π, V) be an irreducible generic smooth representation of $\mathrm{SL}_2(\mathbb{R})$ of moderate growth. Realize π as a quotient of a smooth principal series $I(\chi, s)$, i.e.,

$$0 \to V' \to I(\chi, s) \to V \to 0.$$

Let $\lambda : V \to \mathbb{C}$ be the unique nonzero continuous Whittaker functional on V. Then the composition

$$\Lambda : I(\chi, s) \longrightarrow V \xrightarrow{\lambda} \mathbb{C}$$

gives the unique nonzero continuous Whittaker functional on $I(\chi, s)$ up to a scalar. It follows that the two spaces $\{\lambda(\pi(g)v) : g \in \mathrm{SL}_2(F), v \in V\}$ and $\{\Lambda(R(g).f) : g \in \mathrm{SL}_2(F), f \in I(\chi, s)\}$ are the same, although the first is the Whittaker model of π while the later may not be a Whittaker model of $I(\chi, s)$.

The Whittaker functional on $I(\chi, s)$ is given by the following

$$\Lambda(f) = \int_{N(F)} f(wu)\psi^{-1}(u)\,du,$$

when s is in some right half plane and its continuation gives Whittaker functionals for all $I(\chi, s)$. Also when f has support inside $\Omega = N(F)TwN(F)$ the above integral always converges for any s and gives the Whittaker functional.

Now for such f one computes that, for $a = t(a) \in T$,

$$\Lambda(I(a).f) = \int_{N(F)} f(wua)\psi^{-1}(u)du = \chi'(a) \int_F f(wu)\psi^{-1}(a^2u)du$$
$$= \chi'(a) \int_F f_1(u)\psi^{-1}(a^2u)du = \chi'(a)\hat{f}_1(a^2),$$

where f_1 is the restriction of f to wN which can be chosen to be a Schwartz function, $\hat{f}_1$ is its Fourier transform and χ' is a certain character. Again, as the Fourier transform gives an isometry of Schwartz functions, we can always choose f so that its Whittaker function $W_f(a)$ has arbitrarily small compact support around 1. By a right translation by w we show that one can always choose f so that $W_{w.f}(aw)$ has small compact support around 1.

In order to prove the proposition note that we have chosen Φ. Let

$$R(a,u) = |a|^{1/2}\frac{\gamma(\psi^{-1})}{\gamma(\psi_a^{-1})}\hat{\Phi}_{a,u}(1)\delta(a)^{1/2}(\eta_{s-1/2}\mu_{\psi^{-1}})(a)\Delta(a).$$

Then $R(a,u)$ is a continuous function of a and u, and $R(1,0) \neq 0$. This means that there exist neighborhoods U_1 of $a = 1$ and U_2 of $u = 0$, such that $R(a,u) > R(1,0)/2 > 0$ for all $a \in U_1$ and $u \in U_2$.

Now choose f so that $W_{w.f}(aw)$ has small compact support in a neighborhood V_1 of 1 with $V_1 \subset U_1$ and $W_{w.f}(w) > 0$. For this Whittaker function, since $W_{w.f}(awu)$ is continuous on u, we can choose f_2 so that it is positively supported in a neighborhood V_2 of 0 such that:

(1) $V_2 \subset U_2$.

(2) $W_{w.f}(awu) > W_{w.f}(w)/2 > 0$ for all $u \in V_2$.

Then (4-3) becomes

$$\int W_{w.f}(awu)R(a,u)f_2(u)dadu > \frac{W_{w.f}(w)}{2}\frac{R(1,0)}{2}\int_{V_1}\int_{V_2} f_2(u)dadu > 0,$$

which proves the nonvanishing. □

Acknowledgements

We would like to thank our advisor Professor Jim Cogdell for his guidance, encouragement and support over the years. We would also like to thank Professor Tonghai Yang for helpful discussions. We are grateful to the referee for the detailed comments which improved the exposition of the paper. The first author is supported by the National Natural Science Foundation of China grant 11401193 and the Young Teachers Program of Hunan University, NO. 531107040661 (Fundamental Research Funds for the Central Universities of China).

References

[Badulescu 2008] A. I. Badulescu, "Global Jacquet–Langlands correspondence, multiplicity one and classification of automorphic representations", *Invent. Math.* **172**:2 (2008), 383–438. MR 2390289

[Baruch 1995] E. M. Baruch, *Local factors attached to representations ofp-adic groups and strong multiplicity one*, Ph.D. thesis, Yale University, Ann Arbor, MI, 1995, http://search.proquest.com/docview/304230667. MR 2692992

[Baruch 1997] E. M. Baruch, "On the gamma factors attached to representations of U(2, 1) over a p-adic field", *Israel J. Math.* **102** (1997), 317–345. MR 1489111

[Blasius 1994] D. Blasius, "On multiplicities for SL(n)", *Israel J. Math.* **88**:1-3 (1994), 237–251. MR 1303497

[Casselman 1973] W. Casselman, "On some results of Atkin and Lehner", *Math. Ann.* **201** (1973), 301–314. MR 0337789

[Casselman and Shalika 1980] W. Casselman and J. Shalika, "The unramified principal series of p-adic groups. II. The Whittaker function", *Compositio Math.* **41**:2 (1980), 207–231. MR 581582

[Casselman et al. 2000] W. Casselman, H. Hecht, and D. Miličić, "Bruhat filtrations and Whittaker vectors for real groups", pp. 151–190 in *The mathematical legacy of Harish-Chandra* (Baltimore, MD), edited by R. S. Doran and V. S. Varadarajan, Proc. Sympos. Pure Math. **68**, Amer. Math. Soc., Providence, RI, 2000. MR 1767896

[Gan et al. 2002] W. T. Gan, N. Gurevich, and D. Jiang, "Cubic unipotent Arthur parameters and multiplicities of square integrable automorphic forms", *Invent. Math.* **149**:2 (2002), 225–265. MR 1918673

[Gan et al. 2012] W. T. Gan, B. H. Gross, and D. Prasad, "Symplectic local root numbers, central critical L values, and restriction problems in the representation theory of classical groups", 346 (2012), 1–109. MR 3202556

[Gan et al. 2014] W. T. Gan, Y. Qiu, and S. Takeda, "The regularized Siegel–Weil formula (the second term identity) and the Rallis inner product formula", *Invent. Math.* **198**:3 (2014), 739–831. MR 3279536

[Gelbart and Piatetski-Shapiro 1980] S. Gelbart and I. I. Piatetski-Shapiro, "Distinguished representations and modular forms of half-integral weight", *Invent. Math.* **59**:2 (1980), 145–188. MR 577359

[Gelbart and Piatetski-Shapiro 1984] S. Gelbart and I. Piatetski-Shapiro, "Automorphic forms and L-functions for the unitary group", pp. 141–184 in *Lie group representations, II.* (College Park, Md.), edited by R. Herb et al., Lecture Notes in Math. **1041**, Springer, Berlin, 1984. MR 748507

[Gelbart et al. 1987] S. Gelbart, I. Piatetski-Shapiro, and S. Rallis, *Explicit constructions of automorphic L-functions*, Lecture Notes in Mathematics **1254**, Springer, Berlin, 1987. MR 892097

[Gelbart et al. 1997] S. Gelbart, J. Rogawski, and D. Soudry, "Endoscopy, theta-liftings, and period integrals for the unitary group in three variables", *Ann. of Math.* (2) **145**:3 (1997), 419–476. MR 1454699

[Ginzburg et al. 1997] D. Ginzburg, S. Rallis, and D. Soudry, "Periods, poles of L-functions and symplectic-orthogonal theta lifts", *J. Reine Angew. Math.* **487** (1997), 85–114. MR 1454260

[Ginzburg et al. 1998] D. Ginzburg, S. Rallis, and D. Soudry, "L-functions for symplectic groups", *Bull. Soc. Math. France* **126**:2 (1998), 181–244. MR 1675971

[Jacquet and Langlands 1970] H. Jacquet and R. P. Langlands, *Automorphic forms on* GL(2), Springer, Berlin-New York, 1970. MR 0401654 Zbl 0236.12010

[Jacquet and Shalika 1981] H. Jacquet and J. A. Shalika, "On Euler products and the classification of automorphic forms. II.", *Amer. J. Math.* **103**:4 (1981), 777–815. MR 623137

[Jiang 2006] D. Jiang, "On local γ-factors", pp. 1–28 in *Arithmetic geometry and number theory*, edited by L. Weng and I. Nakamura, Ser. Number Theory Appl. **1**, World Sci. Publ., Hackensack, NJ, 2006. MR 2258071

[Jiang and Nien 2013] D. Jiang and C. Nien, "On the local Langlands conjecture and related problems over p-adic local fields", preprint, 2013, http://www.math.umn.edu/~jiang034/Papers/iccm13.pdf.

[Jiang and Soudry 2003] D. Jiang and D. Soudry, "The local converse theorem for $\mathrm{SO}(2n+1)$ and applications", *Ann. of Math.* (2) **157**:3 (2003), 743–806. MR 1983781

[Kubota 1969] T. Kubota, *On automorphic functions and the reciprocity law in a number field*, Kinokuniya Book-Store Co., Ltd., Tokyo, 1969. MR 0255490

[Labesse and Langlands 1979] J.-P. Labesse and R. P. Langlands, "L-indistinguishability for SL(2)", *Canad. J. Math.* **31**:4 (1979), 726–785. MR 540902

[Lang 1994] S. Lang, *Algebraic number theory*, 2nd ed., Graduate Texts in Mathematics **110**, Springer, New York, 1994. MR 1282723

[Li 1997] J.-S. Li, "Automorphic forms with degenerate Fourier coefficients", *Amer. J. Math.* **119**:3 (1997), 523–578. MR 1448215

[Ramakrishnan 2000] D. Ramakrishnan, "Modularity of the Rankin–Selberg L-series, and multiplicity one for SL(2)", *Ann. of Math.* (2) **152**:1 (2000), 45–111. MR 1792292

[Shahidi 2011] F. Shahidi, "Arthur packets and the Ramanujan conjecture", *Kyoto J. Math.* **51**:1 (2011), 1–23. MR 2784745

[Soudry 1987] D. Soudry, "A uniqueness theorem for representations of GSO(6) and the strong multiplicity one theorem for generic representations of GSp(4)", *Israel J. Math.* **58**:3 (1987), 257–287. MR 917359

[Sun 2012] B. Sun, "Multiplicity one theorems for Fourier–Jacobi models", *Amer. J. Math.* **134**:6 (2012), 1655–1678. MR 2999291

[Szpruch 2009] D. Szpruch, "Computation of the local coefficients for principal series representations of the metaplectic double cover of $\mathrm{SL}_2(\mathbb{F})$", *J. Number Theory* **129**:9 (2009), 2180–2213. MR 2528059

[Vogan 1981] D. A. Vogan, Jr., *Representations of real reductive Lie groups*, Progress in Mathematics **15**, Birkhäuser, Boston, 1981. MR 632407

[Zhang 2015] Q. Zhang, "A local converse theorem for U(1,1)", preprint, 2015. arXiv 1508.07062

Received November 19, 2015. Revised May 9, 2016.

JINGSONG CHAI
COLLEGE OF MATHEMATICS AND ECONOMETRICS
HUNAN UNIVERSITY
CHANGSHA, 410082
CHINA
jingsongchai@hotmail.com

QING ZHANG
DEPARTMENT OF MATHEMATICS
THE OHIO STATE UNIVERSITY
231 WEST 18TH AVENUE
100 MATH TOWER
COLUMBUS, OH 43210-1174
UNITED STATES
zhang.1649@osu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 285, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.285.375

THE YAMABE PROBLEM ON NONCOMPACT CR MANIFOLDS

PAK TUNG HO AND SEONGTAG KIM

Let (M,θ) be a noncompact complete strictly pseudoconvex CR manifold of real dimension $2n+1\geq 3$ with positive Webster scalar curvature. We show that there exists a conformal contact form $\tilde{\theta}=u^{2/n}\theta$ with positive constant Webster scalar curvature if the CR-Yamabe invariant $Y(M,\theta)$ of (M,θ) is positive and strictly less than the CR-Yamabe invariant at infinity $\overline{Y(M,\theta)}$.

1. Introduction

Suppose that (M,g) is a compact Riemannian manifold of dimension $n\geq 3$. As a generalization of the uniformization theorem, the Yamabe problem is to find a metric conformal to g such that its scalar curvature is constant. This was solved by Trudinger [1968], Aubin [1976] and Schoen [1984]. The uniqueness of the solution of the Yamabe problem was studied in [Kazdan and Warner 1975; Lou 1998]. See the survey article [Lee and Parker 1987] for more about the Yamabe problem. See also [Brendle 2005; 2007; Chow 1992; Schwetlick and Struwe 2003; Ye 1994] for results related to the Yamabe flow, which is the geometric flow introduced to study the Yamabe problem.

The Yamabe problem was also studied on complete noncompact Riemannian manifolds. In this case, there is a simple counterexample such that the Yamabe problem does not have a solution (see [Jin 1988]). See also [Aviles and McOwen 1988; Bland and Kalka 1989; Große and Nardmann 2014; Kim 1997; 2000; Zhang 2003] and references therein for results related to the Yamabe problem on noncompact Riemannian manifolds. In particular, we mention the following result which is related to our main theorem. If (M,g) is a noncompact Riemannian manifold with positive scalar curvature R_g, we define

$$Y(M,g)=\inf_{u\in C_0^\infty(M)}\frac{\displaystyle\int_M|\nabla_g u|^2+\frac{n-2}{4(n-1)}R_g u^2\,dV_g}{\left(\displaystyle\int_M u^{\frac{2n}{n-2}}\,dV_\theta\right)^{\frac{n-2}{n}}}$$

MSC2010: primary 32V20, 53C21; secondary 32V05, 58J05.
Keywords: CR manifolds, CR Yamabe problem, Webster scalar curvature, conformal changes, noncompact complete manifolds.

and

$$\overline{Y(M,g)} = \lim_{r\to\infty} \inf_{u\in C_0^\infty(M-B_r)} \frac{\displaystyle\int_{M-B_r} |\nabla_g u|^2 + \frac{n-2}{4(n-1)} R_g u^2 \, dV_g}{\left(\displaystyle\int_{M-B_r} u^{\frac{2n}{n-2}} \, dV_g\right)^{\frac{n-2}{n}}},$$

where r is the distance induced by the Riemannian metric g from x to a fixed point x_0 in M, and B_r is the ball of radius r centered at x_0. The second author [Kim 1996] proved the following:

Theorem 1.1. *Suppose (M,g) is a noncompact Riemannian manifold with positive scalar curvature with $0 < Y(M,g) < \overline{Y(M,g)}$. Then there exists a Riemannian metric conformal to g which has positive constant scalar curvature.*

The Yamabe problem can also be formulated in the context of CR manifolds. Suppose that (M,θ) is a compact strictly pseudoconvex CR manifold of real dimension $2n+1$ with a given contact form θ. The CR Yamabe problem is to find a contact form conformal to θ such that its Webster scalar curvature is constant. This was introduced by Jerison and Lee [1987], and was solved by them for the case when $n \geq 2$ and M is not locally CR equivalent to the sphere S^{2n+1} in [Jerison and Lee 1987; 1988; 1989]. The remaining cases, namely when $n = 1$ or when M is locally CR equivalent to the sphere, were studied respectively in [Gamara and Yacoub 2001] and in [Gamara 2001]. See also the recent work of Cheng, Chiu and Yang [Cheng et al. 2014] and Cheng, Malchiodi and Yang [Cheng et al. 2013] for the study of these two cases. The uniqueness of the solution of the CR Yamabe problem was studied in [Ho 2013; Jerison and Lee 1987]. On the other hand, the CR Yamabe flow, the geometric flow introduced to study the CR Yamabe problem, was studied in [Chang and Cheng 2002; Chang et al. 2010; Ho 2012; 2015].

In this paper, we study the CR Yamabe problem on noncompact manifolds. We suppose that (M,θ) is a noncompact strictly pseudoconvex CR manifold of real dimension $2n+1$ such that its Webster scalar curvature R_θ is positive. We would like to find another contact form conformal to θ such that its Webster scalar curvature is constant. This is equivalent to finding a positive solution to the equation

$$-\Delta_\theta u + \frac{n}{2n+2} R_\theta u = q u^{1+\frac{2}{n}}, \tag{1-1}$$

where q is a positive constant. We define

$$Y(M,\theta) = \inf_{u\in C_0^\infty(M)} \frac{\displaystyle\int_M |\nabla_\theta u|^2 + \frac{n}{2n+2} R_\theta u^2 \, dV_\theta}{\left(\displaystyle\int_M u^{2+\frac{2}{n}} \, dV_\theta\right)^{\frac{n}{n+1}}}$$

and

$$\overline{Y(M,\theta)} = \lim_{r\to\infty} \inf_{u\in C_0^\infty(M-B_r)} \frac{\displaystyle\int_{M-B_r} |\nabla_\theta u|^2 + \frac{n}{2n+2} R_\theta u^2 \, dV_\theta}{\left(\displaystyle\int_{M-B_r} u^{2+\frac{2}{n}} \, dV_\theta\right)^{\frac{n}{n+1}}},$$

where ∇_θ is the subgradient with respect to θ, $dV_\theta = \theta\wedge(d\theta)^n$ is the volume form of θ, r is the Carnot–Carathéodory distance from x to a fixed point $x_0 \in M$ with respect to the contact form θ, and B_r is the ball of radius r centered at x_0. We refer readers to the book [Dragomir and Tomassini 2006] or the paper [Jerison and Lee 1987] for more about the definitions and concepts related to CR manifolds.

Note that $\overline{Y(M,\theta)}$ is well defined. Indeed, if we let

$$f(r) = \inf_{u\in C_0^\infty(M-B_r)} \frac{\displaystyle\int_{M-B_r} |\nabla_\theta u|^2 + \frac{n}{2n+2} R_\theta u^2 \, dV_\theta}{\left(\displaystyle\int_{M-B_r} u^{2+\frac{2}{n}} \, dV_\theta\right)^{\frac{n}{n+1}}},$$

then it follows from the definition that $f(r)$ is nondecreasing as a function of r. Since $f(r)$ is bounded above by $Y(S^{2n+1}, \theta_{S^{2n+1}})$, $\lim_{r\to\infty} f(r)$ exists.

The following is our main theorem, which is the CR version of Theorem 1.1.

Theorem 1.2. *Let (M,θ) be a noncompact strictly pseudoconvex CR manifold of real dimension $2n+1\ge 3$ with positive Webster scalar curvature. Assume that*

$$0 < Y(M,\theta) < \overline{Y(M,\theta)}.$$

Then there exists a positive solution u of (1-1). *That is, the contact form $u^{2/n}\theta$ conformal to θ has positive constant Webster scalar curvature.*

2. Proof

Since we have assumed that

$$0 < Y(M,\theta) < \overline{Y(M,\theta)} \ (\le Y(S^{2n+1}, \theta_{S^{2n+1}})),$$

there exists a sequence of smooth compact domains K_i such that $Y(K_i,\theta) < \overline{Y(M,\theta)}$ with $K_i \subset K_{i+1}$ satisfying $\bigcup K_i = M$. Using the work on the CR Yamabe problem in the compact case (see [Jerison and Lee 1987]) for $2n+1\ge 3$, we have a positive smooth function u_i on each K_i with

$$-\Delta_\theta u_i + \frac{n}{2n+2} R_\theta u_i = q_i u_i^{1+\frac{2}{n}} \quad \text{on } K_i, \tag{2-1}$$

$u_i = 0$ on ∂K_i and

(2-2)
$$\int_{K_i} u_i^{2+\frac{2}{n}}\, dV_\theta = 1,$$

where

$$q_i = Y(K_i, \theta) \to Y(M, \theta) \quad \text{as } i \to \infty.$$

We extend the domain of u_i by defining $u_i = 0$ outside K_i, and we still denote its extension by u_i. Then the extension of u_i is in $S_1^2(M, \theta)$, the completion of $C_0^\infty(M)$ with the norm

$$\|u\|^2_{S_1^2(M,\theta)} = \int_M |\nabla_\theta u|^2 + \frac{n}{2n+2} R_\theta u^2\, dV_\theta.$$

For sufficiently large i, let $\widetilde{K}$ and K' be fixed smooth compact subsets of M with $\widetilde{K} \subset K' \subset K_i$. We shall show that $\int_{\widetilde{K}} u_i^{(1+b)(2+2/n)}\, dV_\theta$ is uniformly bounded for some positive b. The constant $c(\epsilon)$ in the Sobolev embedding for u_i on a noncompact complete Riemannian manifold depends on the domain and does not have to be uniformly bounded (see (2-7)); therefore the Sobolev embedding is not directly applicable in (6) of [Kim 1996]. However, the Sobolev embedding holds for $u_i\varphi$ on a fixed domain K', where φ is a cutoff function supported in K'. The uniform bound of u_i in $L_{(1+b)(2+2/n)}(\widetilde{K})$ can be obtained on each compact subset $\widetilde{K}$, by applying the same method of [Kim 1996] to $u_i\varphi$. The detailed proof for the CR case is provided in the following steps.

Take

$$\Omega = \{x \in K' \mid u_i(x) \geq 1\}$$

with

$$Y(K', \theta) < Y(S^{2n+1}, \theta_{S^{2n+1}}).$$

Then

(2-3)
$$|\Omega| = \int_\Omega dV_\theta < 1$$

by (2-2). Now let $u_i = 1 + w_i$. Then

(2-4)
$$\Omega = \{x \in K' \mid w_i(x) \geq 0\}$$

by definition, for sufficiently large i, and (2-1) is equivalent to

(2-5)
$$-\Delta_\theta w_i + \frac{n}{2n+2} R_\theta (1 + w_i) = q_i (1 + w_i)^{1+\frac{2}{n}} \quad \text{on } K_i.$$

Take a smooth cutoff function $\varphi \in C_0^\infty(K')$ with $\varphi \equiv 1$ on $\widetilde{K}$ and $|\varphi| \le 1$ on K'. Multiplying (2-5) by $\varphi^{2+2b} w_i^{1+2b}$, where $b > 0$, and integrating it over Ω, we get

(2-6)
$$\begin{aligned}
q_i \int_\Omega & \varphi^{2+2b} w_i^{1+2b} (1+w_i)^{1+\frac{2}{n}} \, dV_\theta \\
&= -\int_\Omega \varphi^{2+2b} w_i^{1+2b} \Delta_\theta w_i \, dV_\theta + \frac{n}{2n+2} \int_\Omega R_\theta (1+w_i) \varphi^{2+2b} w_i^{1+2b} \, dV_\theta \\
&= \int_\Omega \frac{1+2b}{(1+b)^2} \varphi^{2+2b} |\nabla_\theta (w_i^{1+b})|^2 + \frac{2}{1+b} \varphi^{1+b} w_i^{1+b} \nabla_\theta w_i^{1+b} \cdot \nabla_\theta \varphi^{1+b} \, dV_\theta \\
&\qquad + \frac{n}{2n+2} \int_\Omega R_\theta (1+w_i) w_i^{1+2b} \varphi^{2+2b} \, dV_\theta \\
&\ge \int_\Omega \frac{1+2b-\epsilon_1}{(1+b)^2} \varphi^{2+2b} |\nabla_\theta (w_i^{1+b})|^2 - \frac{1}{\epsilon_1} |\nabla_\theta \varphi^{1+b}|^2 w_i^{2+2b} \, dV_\theta \\
&\qquad + \frac{n}{2n+2} \int_\Omega R_\theta (1+w_i) w_i^{1+2b} \varphi^{2+2b} \, dV_\theta,
\end{aligned}$$

where we used integration by parts, Hölder's inequality and (2-4).

We are going to estimate the terms on the right-hand side of (2-6). Applying (A-1) in the Appendix for $\varphi w_i \in C_0^\infty(\Omega)$, where $\Omega \subset K_i \subset M$, we obtain that for any given $\epsilon > 0$, there exists $C(\epsilon)$, which depends on the given domain Ω, such that

(2-7)
$$\begin{aligned}
&\left(\int_\Omega (\varphi^{1+b} w_i^{1+b})^{2+\frac{2}{n}} \, dV_\theta \right)^{\frac{n}{n+1}} \\
&\le \frac{1+\epsilon}{Y(S^{2n+1}, \theta_{S^{2n+1}})} \int_\Omega |\nabla_\theta (\varphi^{1+b} w_i^{1+b})|^2 \, dV_\theta + C(\epsilon) \int_\Omega (\varphi^{1+b} w_i^{1+b})^2 \, dV_\theta \\
&\le \frac{1+\epsilon}{Y(S^{2n+1}, \theta_{S^{2n+1}})} \int_\Omega (1+\epsilon_2) \varphi^{2+2b} |\nabla_\theta w_i^{1+b}|^2 \\
&\qquad + \frac{1}{\epsilon_2} w_i^{2+2b} |\nabla_\theta \varphi^{1+b}|^2 \, dV_\theta + C(\epsilon) \int_\Omega (\varphi^{1+b} w_i^{1+b})^2 \, dV_\theta \\
&\le \frac{1+\epsilon}{Y(S^{2n+1}, \theta_{S^{2n+1}})} \Bigg(\int_\Omega \frac{(1+b)^2 (1+\epsilon_2)}{1+2b-\epsilon_1} \\
&\qquad\qquad \times \Bigg(q_i \varphi^{2+2b} w_i^{1+2b} (1+w_i)^{1+\frac{2}{n}} + \frac{1}{\epsilon_1} |\nabla_\theta \varphi^{1+b}|^2 w_i^{2+2b} \\
&\qquad\qquad\qquad - \frac{n}{2n+2} R_\theta (1+w_i) w_i^{1+2b} \varphi^{2+2b} \Bigg) \\
&\qquad\qquad\qquad + \frac{1}{\epsilon_2} w_i^{2+2b} |\nabla_\theta \varphi^{1+b}|^2 \, dV_\theta \Bigg) \\
&\qquad + C(\epsilon) \int_\Omega (\varphi^{1+b} w_i^{1+b})^2 \, dV_\theta,
\end{aligned}$$

where we have used (2-6) in the last inequality. Now let $\Omega_1 = \{x \in \Omega \mid w_i \geq 2\}$ and $\Omega_2 = \Omega - \Omega_1$. Let $a = 1 + 2/n$ and $x = 1/w_i$. Note that if $w_i \in \Omega_1$, i.e., $w_i \geq 2$, then $|x| \leq \frac{1}{2}$ and

$$
\begin{aligned}
(1 + w_i)^a - w_i^a &= w_i^a\Big(1 + \frac{1}{w_i}\Big)^a - w_i^a \\
&= w_i^a(1 + x)^a - w_i^a \\
&= w_i^a\big(1 + ax + \tfrac{1}{2}a(a-1)x^2 + \cdots - 1\big) \\
&\leq c_1 w_i^{2/n}
\end{aligned}
\tag{2-8}
$$

for some constant c_1. Using (2-8), the integral in (2-7) can be estimated as follows:

$$
\begin{aligned}
&\Big(\int_\Omega (\varphi^{1+b} w_i^{1+b})^{2+\frac{2}{n}}\, dV_\theta\Big)^{\frac{n}{n+1}} \\
&\quad\leq \frac{1+\epsilon}{Y(S^{2n+1}, \theta_{S^{2n+1}})}\Bigg(\frac{(1+b)^2(1+\epsilon_2)}{1+2b-\epsilon_1} \\
&\qquad\qquad \times \Big(\int_{\Omega_1} q_i \varphi^{2+2b} w_i^{2+2b+\frac{2}{n}} + C\varphi^{2+2b} w_i^{1+2b+\frac{2}{n}}\, dV_\theta \\
&\qquad\qquad\quad + \int_{\Omega_2} q_i \varphi^{2+2b} w_i^{1+2b}(1+w_i)^{1+\frac{2}{n}}\, dV_\theta \\
&\qquad\qquad\quad + \int_\Omega \frac{1}{\epsilon_1}|\nabla_\theta \varphi^{1+b}|^2 w_i^{2+2b} \\
&\qquad\qquad\quad - \frac{n}{2n+2} R_\theta (1+w_i) w_i^{1+2b} \varphi^{2+2b}\, dV_\theta\Big) \\
&\qquad\qquad\quad + \int_\Omega \frac{1}{\epsilon_2} w_i^{2+2b} |\nabla_\theta \varphi^{1+b}|^2\, dV_\theta\Bigg) \\
&\qquad + C(\epsilon)\int_\Omega (\varphi^{1+b} w_i^{1+b})^2\, dV_\theta.
\end{aligned}
\tag{2-9}
$$

By Hölder's inequality, we have

$$
\begin{aligned}
&\int_\Omega \varphi^{2+2b} w_i^{2+2b+\frac{2}{n}} dV_\theta \\
&\quad\leq \Big(\int_\Omega (\varphi^{1+b} w_i^{1+b})^{2+\frac{2}{n}} dV_\theta\Big)^{\frac{n}{n+1}} \Big(\int_\Omega w_i^{\frac{2}{n}(n+1)} dV_\theta\Big)^{\frac{1}{n+1}} \\
&\quad\leq \Big(\int_\Omega (\varphi^{1+b} w_i^{1+b})^{2+\frac{2}{n}}\, dV_\theta\Big)^{\frac{n}{n+1}},
\end{aligned}
\tag{2-10}
$$

where the last inequality follows from

$$\int_\Omega w_i^{2+\frac{2}{n}}\,dV_\theta \le \int_\Omega u_i^{2+\frac{2}{n}}\,dV_\theta \le 1$$

by (2-2) and the definition of w_i and Ω. Since $\Omega_1 \subset \Omega$, we can combine (2-9) and (2-10) to get

(2-11)
$$\begin{aligned}
&\Big(\int_\Omega (\varphi^{1+b} w_i^{1+b})^{2+\frac{2}{n}}\,dV_\theta\Big)^{\frac{n}{n+1}} \\
&\quad\le \frac{1+\epsilon}{Y(S^{2n+1},\theta_{S^{2n+1}})} \\
&\qquad\times\Bigg(\frac{(1+b)^2(1+\epsilon_2)}{1+2b-\epsilon_1} \\
&\qquad\quad\times\bigg(q_i\Big(\int_\Omega (\varphi^{1+b} w_i^{1+b})^{2+\frac{2}{n}}\,dV_\theta\Big)^{\frac{n}{n+1}} + C\int_{\Omega_1}\varphi^{2+2b} w_i^{1+2b+\frac{2}{n}}\,dV_\theta \\
&\qquad\qquad + \int_{\Omega_2} q_i\varphi^{2+2b} w_i^{1+2b}(1+w_i)^{1+\frac{2}{n}}\,dV_\theta + \int_\Omega \frac{1}{\epsilon_1}|\nabla_\theta\varphi^{1+b}|^2 w_i^{2+2b} \\
&\qquad\qquad - \frac{n}{2n+2}R_\theta(1+w_i)w_i^{1+2b}\varphi^{2+2b}\,dV_\theta\bigg) \\
&\qquad\qquad + \int_\Omega \frac{1}{\epsilon_2} w_i^{2+2b}|\nabla_\theta\varphi^{1+b}|^2\,dV_\theta\Bigg) \\
&\quad + C(\epsilon)\int_\Omega (\varphi^{1+b} w_i^{1+b})^2\,dV_\theta.
\end{aligned}$$

Since $q_i < Y(K',\theta) < \overline{Y(M,\theta)} \le Y(S^{2n+1},\theta_{S^{2n+1}})$, we can take $\epsilon, \epsilon_1, \epsilon_2$ and $0 < b < 1/n$ sufficiently small such that

$$\frac{1+\epsilon}{Y(S^{2n+1},\theta_{S^{2n+1}})}\,\frac{(1+b)^2(1+\epsilon_2)}{1+2b-\epsilon_1}q_i \le c_0 < 1.$$

Combining this with (2-11), we obtain

(2-12)
$$\begin{aligned}
&(1-c_0)\Big(\int_\Omega (\varphi w_i)^{(1+b)(2+2/n)}\,dV_\theta\Big)^{\frac{n}{n+1}} + C\int_\Omega R_\theta(1+w_i)w_i^{1+2b}\varphi^{2+2b}\,dV_\theta \\
&\qquad \le C\int_{\Omega_1} w_i^{1+\frac{2}{n}+2b}\varphi^{2+2b}\,dV_\theta + C\int_{\Omega_2} w_i^{1+2b}(1+w_i)^{1+\frac{2}{n}}\varphi^{2+2b}\,dV_\theta \\
&\qquad\quad + C\int_\Omega (w_i^{1+b})^2\varphi^{2+2b}\,dV_\theta + C\int_\Omega |\nabla_\theta\varphi^{1+b}|^2 w_i^{2+2b}\,dV_\theta
\end{aligned}$$

Here C is a constant independent of i. We are going to estimate the terms on the right-hand side of (2-12). Since $\Omega_2 = \Omega - \Omega_1$, we have

$$|\Omega_2| < 1 \qquad \text{and} \qquad 0 \le w_i \le 2 \quad \text{on } \Omega_2.$$

This implies that

$$\int_{\Omega_2} w_i^{1+2b}(1+w_i)^{1+\frac{2}{n}}\varphi^{2+2b}\,dV_\theta \le C$$

for some constant C independent of i. Also, since $b < 1/n$, we have

$$t_1 := \frac{1+b}{1+\frac{1}{n}} < 1.$$

Then it follows from Hölder's inequality that

$$\int_\Omega (w_i^{1+b})^2\varphi^{2+2b}\,dV_\theta \le \left(\int_\Omega w_i^{2+\frac{2}{n}}\,dV_\theta\right)^{t_1}|\Omega|^{1-t_1} \le \left(\int_\Omega u_i^{2+\frac{2}{n}}\,dV_\theta\right)^{t_1} \le 1,$$

where we have used (2-2), (2-3) and (2-4). On the other hand, since $w_i \ge 2$ in Ω_1 and $b < 1/n$, we have

$$\int_{\Omega_1} w_i^{1+\frac{2}{n}+2b}\varphi^{2+2b}\,dV_\theta \le \int_{\Omega_1} w_i^{2+\frac{2}{n}}\,dV_\theta \le \int_\Omega u_i^{2+\frac{2}{n}}\,dV_\theta \le 1,$$

where we have used (2-2). Since φ is a smooth fixed cutoff function, the last term of (2-12) is also bounded. Combining all these, we can conclude that the right-hand side of (2-12) is uniformly bounded. Thus, the left-hand side of (2-12) is uniformly bounded; i.e.,

$$(1-c_0)\left(\int_\Omega (\varphi w_i)^{(1+b)(2+2/n)}\,dV_\theta\right)^{\frac{n}{n+1}} + C\int_\Omega R_\theta(1+w_i)w_i^{1+2b}\varphi^{2+2b}\,dV_\theta \le C.$$

In particular, this implies that

$$\left(\int_\Omega (\varphi w_i)^{(1+b)(2+2/n)}\,dV_\theta\right)^{\frac{n}{n+1}} \le C_0 \tag{2-13}$$

and

$$\int_{\widetilde{K}} w_i^{(1+b)(2+2/n)}\,dV_\theta \le C_0' \tag{2-14}$$

for some constants C_0 and C_0' independent of i. Therefore, u_i is uniformly bounded in $L_{(1+b)(2+2/n)}(\widetilde{K})$ for each compact subset $\widetilde{K}$ of M and some positive b.

We can now show that w_i is $C^{2,\alpha}$ bounded on each compact subset of M in the following way: Consider sufficiently large compact subsets $K \subset K_0 \subset K_1 \subset K_2$

with smooth boundary satisfying $Y(K,\theta) < Y(S^{2n+1}, \theta_{S^{2n+1}})$. It follows from (2-14) that

$$\int_{K_2} w_i^{2+\frac{2}{n}+2\bar{b}}\, dV_\theta \le C_0,$$

where $\bar{b} = b(1+1/n)$ and C_0 is a constant independent of i. Also, we have

$$|\Delta_\theta w_i| = \left| \frac{n}{2n+2} R_\theta(1+w_i) - q_i(1+w_i)^{1+\frac{2}{n}} \right| \le C(1+w_i)^{1+\frac{2}{n}} \quad \text{on } K_2,$$

where C is a constant that depends only on K_2 and $\max_{K_2} R_\theta$. Hence, $\Delta_\theta w_i \in L^q(K_2)$, where $q = (2n+2+2n\bar{b})/(n+2)$. By the regularity theory (see [Jerison and Lee 1987, Proposition 5.7(c)]), we have $w_i \in S_2^q(K_1)$. From the Folland-Stein embedding theorem (see [Jerison and Lee 1987, Proposition 5.5]), we have $w_i \in L^s(K_1)$, where

$$s = \left(2 + \frac{2}{n} + 2\bar{b}\right) \frac{n+1}{n+1-2\bar{b}} > 2 + \frac{2}{n} + 2\bar{b}.$$

Continuing this procedure, we get $w_i \in S_2^t(K_0)$ for all $t > 1$. Again by the Folland–Stein embedding theorem (see [Jerison and Lee 1987, Proposition 5.7(a–b)]), we have $w_i^{2+2/n} \in C^\alpha(K_0)$ for some $\alpha > 0$. By the regularity theory again (see [Jerison and Lee 1987, Proposition 5.9(b)]), we can conclude that $w_i \in C^{2,\alpha}(K)$, as required.

By the definition of Ω and since $u_i = 1 + w_i$, we have a uniform $C^{2,\alpha}$ bound for u_i on each compact subset of M. Therefore, we can find a subsequence, which we still denote by $\{u_i\}$, that converges to some u uniformly on each compact subset by the Arzelà–Ascoli theorem.

To sum up, we have proved the following:

Lemma 2.1. *If $Y(M,\theta) < \overline{Y(M,\theta)}$, then there exists a subsequence $\{u_i\}$ which converges to a solution u of* (1-1) *uniformly on each compact subset of M.*

We remark that we do not know whether u is strictly positive. Note that if $u = 0$ at some point of M, then by applying Proposition 2.2 (stated below) to (1-1), we can conclude that u is identically equal to zero.

Proposition 2.2. *Suppose that u is a nonnegative function on M satisfying*

$$-\Delta_\theta u + P(x)u \ge 0,$$

where $P(x)$ is a smooth function on M. Then for any compact set K in M, there exists a constant C such that

$$\int_K u^{2+\frac{2}{n}}\, dV_\theta \le C\left(\min_K u\right)\left(\max_K u\right)^{\frac{n+2}{n}}.$$

We skip the proof of Proposition 2.2, because it is essentially the same as the proof of Proposition A.1 in [Ho 2012].

We are going to show that it is impossible for u to be identically equal to zero. First, we have the following:

Lemma 2.3. *As $i \to \infty$,*

$$\int_M |u_i|^{2+\frac{2}{n}}\,dV_\theta - \int_M |u-u_i|^{2+\frac{2}{n}}\,dV_\theta \to \int_M |u|^{2+\frac{2}{n}}\,dV_\theta.$$

Proof. Note that

$$\begin{aligned}
&\int_M |u_i|^{2+\frac{2}{n}}\,dV_\theta - \int_M |u-u_i|^{2+\frac{2}{n}}\,dV_\theta \\
&\qquad = -\int_M \int_0^1 \frac{\partial}{\partial t}|u_i - tu|^{2+\frac{2}{n}}\,dt\,dV_\theta \\
&\qquad = \left(2+\frac{2}{n}\right)\int_M \int_0^1 u(u_i - tu)|u_i - tu|^{\frac{2}{n}}\,dt\,dV_\theta \\
&\qquad \to \left(2+\frac{2}{n}\right)\int_M \int_0^1 u(u - tu)|u - tu|^{\frac{2}{n}}\,dt\,dV_\theta \\
&\qquad = \int_M |u|^{2+\frac{2}{n}}\,dV_\theta
\end{aligned}$$

as $i \to \infty$. □

For abbreviation, we let

$$v_i = u_i - u \quad \text{and} \quad E(v) = \int_M \left(|\nabla_\theta v|^2 + \frac{n}{2n+2}R_\theta v^2\right)dV_\theta.$$

Lemma 2.4. *As $i \to \infty$,*

$$E(u_i) - E(v_i) \to E(u).$$

Proof. We compute

$$\begin{aligned}
E(u_i) - E(v_i) &= E(u+v_i) - E(v_i) \\
&= E(u) + 2\int_M \left(-\Delta_\theta u + \frac{n}{2n+2}R_\theta u\right)v_i\,dV_\theta \\
&\to E(u)
\end{aligned}$$

as $i \to \infty$, since v_i tends to 0 weakly in $S_1^2(M)$. This proves the assertion. □

Lemma 2.5. *For any fixed B_r, we have*

$$E(v_i) \geq Y(M - B_r, \theta)\left(\int_{M-B_r} |v_i|^{2+\frac{2}{n}}\,dV_\theta\right)^{\frac{n}{n+1}} + o(1) \quad \textit{as } i \to \infty.$$

Proof. Note that

$$\begin{aligned} E(v_i) &= \int_M \left(|\nabla_\theta v_i|^2 + \frac{n}{2n+2} R_\theta v_i^2 \right) dV_\theta \\ &= \int_{M-B_r} \left(|\nabla_\theta v_i|^2 + \frac{n}{2n+2} R_\theta v_i^2 \right) dV_\theta + \int_{B_r} \left(|\nabla_\theta v_i|^2 + \frac{n}{2n+2} R_\theta v_i^2 \right) dV_\theta \\ &\geq \int_{M-B_r} \left(|\nabla_\theta v_i|^2 + \frac{n}{2n+2} R_\theta v_i^2 \right) dV_\theta + o(1) \\ &\geq Y(M-B_r,\theta) \left(\int_{M-B_r} |v_i|^{2+\frac{2}{n}} \, dV_\theta \right)^{\frac{n}{n+1}} + o(1), \end{aligned}$$

where the first inequality follows from the fact that $v_i \to 0$ uniformly on B_r by Lemma 2.1. This proves the assertion. □

Note that $u_i \rightharpoonup u$ weakly in $S_1^2(M,\theta)$. Assume that

$$\int_M |u|^{2+2/n} \, dV_\theta = \lambda.$$

Note that if $\lambda > 0$, then

$$E(u) = \lambda^{\frac{n}{n+1}} E\big(\lambda^{-\frac{n}{2n+2}} u\big) \geq \lambda^{\frac{n}{n+1}} Y(M,\theta). \tag{2-15}$$

Furthermore, if $\lambda < 1$, then

$$E(v_i) = (1-\lambda)^{\frac{n}{n+1}} E\big((1-\lambda)^{-\frac{n}{2n+2}} v_i\big) \geq (1-\lambda)^{\frac{n}{n+1}} \overline{Y(M,\theta)} + O(1) \tag{2-16}$$

by the definition of $\overline{Y(M,\theta)}$.

We have the following three cases:

Case 1. If $0 < \lambda < 1$, then

$$\begin{aligned} Y(M,\theta) &= E(u_i) + o(1) \\ &= E(u) + E(v_i) + o(1) \\ &\geq \lambda^{\frac{n}{n+1}} Y(M,\theta) + (1-\lambda)^{\frac{n}{n+1}} \overline{Y(M,\theta)} + o(1) \\ &\geq (\lambda^{\frac{n}{n+1}} + (1-\lambda)^{\frac{n}{n+1}}) Y(M,\theta) + o(1), \end{aligned}$$

where the second equality follows from Lemma 2.4, the first inequality follows from (2-15) and (2-16), and the last inequality follows from the assumption that $Y(M,\theta) < \overline{Y(M,\theta)}$. But this is a contradiction, since

$$\lambda^t + (1-\lambda)^t > (1-\lambda+\lambda)^t = 1 \qquad \text{for } 0 < \lambda < 1 \text{ and } 0 < t < 1.$$

Case 2. If $\lambda = 0$, then

$$\begin{aligned} Y(M,\theta) &= E(u_i) + o(1) \\ &= E(u) + E(v_i) + o(1) \\ &\geq E(v_i) + o(1) \\ &\geq \overline{Y(M,\theta)} + o(1), \end{aligned}$$

where the second equality follows from Lemma 2.4, and the last inequality follows from (2-16) with $\lambda = 0$. But this contradicts the assumption that $Y(M,\theta) < \overline{Y(M,\theta)}$.

Case 3. Therefore, we must have $\lambda = 1$; i.e.,

$$\int_M |u|^{2+\frac{2}{n}}\, dV_\theta = 1.$$

This implies that u is not identically equal to zero. As pointed out in the remark after Lemma 2.1, u is strictly positive. Therefore, we have a positive solution u in $S_1^2(M,\theta)$ for (1-1).

Now it follows from Theorem 5.15 in [Jerison and Lee 1987] that u is smooth. This proves Theorem 1.2.

Appendix

We prove the following inequality related to the Folland–Stein embedding:

Theorem A.1. *Suppose K is a smooth compact subset in M. For any $\epsilon > 0$, there exists a constant $C(\epsilon, K)$ such that*

$$\text{(A-1)}\quad Y(S^{2n+1}, \theta_{S^{2n+1}})\left(\int_K |\varphi|^{2+\frac{2}{n}}\, dV_\theta\right)^{\frac{n}{n+1}} \leq (1+\epsilon)\int_K |\nabla_\theta \varphi|^2\, dV_\theta + C(\epsilon, K)\int_K |\varphi|^2\, dV_\theta$$

for all $\varphi \in S_1^2(M,\theta)$ with its compact support lying in K.

We remark that Theorem A.1 is probably well known. But we cannot find it in the literature. Therefore we provide the proof here. In particular, the Riemannian version of Theorem A.1 can be found in Theorem 2.21 of [Aubin 1998].

Proof of Theorem A.1. Given any $\delta > 0$, for any point $p \in M$, there exists a neighborhood U_p of p and a diffeomorphism f_p from U_p to a neighborhood of the origin of $\mathbb{H}^n$ such that (see [Jerison and Lee 1987, Theorem 4.3])

$$\text{(A-2)}\quad \begin{aligned} (f_p)_*(dV_\theta) &= (1 + O(\delta))dV_{\theta_{\mathbb{H}^n}}, \\ (f_p)_*(|\nabla_\theta \varphi|^2) &= (1 + O(\delta))|\nabla_{\theta_{\mathbb{H}^n}}(\varphi \circ f)|^2 \end{aligned}$$

for any function φ in M. It follows from [Jerison and Lee 1988, Corollary C] that

$$\left(\int_{\mathbb{H}^n}|\varphi|^{2+\frac{2}{n}}\,dV_{\theta_{\mathbb{H}^n}}\right)^{\frac{n}{n+1}} \le K(n,2)\int_{\mathbb{H}^n}|\nabla_{\theta_{\mathbb{H}^n}}\varphi|^2\,dV_{\theta_{\mathbb{H}^n}} \tag{A-3}$$

for any smooth function φ which has compact support in $\mathbb{H}^n$, where

$$\begin{aligned} K(n,2) &= \frac{1}{2\pi n(n+1)} \\ &= \frac{1}{Y(S^{2n+1},\theta_{S^{2n+1}})}. \end{aligned}$$

This implies that (A-3) is also true for $\varphi \in S_1^2(\mathbb{H}^n,\theta_{\mathbb{H}^n})$ which is compactly supported. Combining (A-2) and (A-3), we get

$$\begin{aligned} \left(\int_{U_p}|\varphi|^{2+\frac{2}{n}}\,dV_\theta\right)^{\frac{n}{n+1}} &= \left(\int_{f_p(U_p)}|\varphi\circ f_p|^{2+\frac{2}{n}}\,(f_p)_*(dV_\theta)\right)^{\frac{n}{n+1}} \\ &\le (1+O(\delta))\left(\int_{f_p(U_p)}|\varphi\circ f_p|^{2+\frac{2}{n}}\,dV_{\theta_{\mathbb{H}^n}}\right)^{\frac{n}{n+1}} \\ &\le (1+O(\delta))K(n,2)\int_{f_p(U_p)}|\nabla_{\theta_{\mathbb{H}^n}}(\varphi\circ f_p)|^2\,dV_{\theta_{\mathbb{H}^n}} \\ &\le (1+O(\delta))K(n,2)\int_{U_p}|\nabla_\theta\varphi|^2\,dV_\theta \end{aligned} \tag{A-4}$$

for any function φ which has compact support in U_p.

Since K is compact, there exists a finite subcovering $\{U_{p_i}\}_{i=1}^k$; i.e.,

$$K = \bigcup_{i=1}^k U_{p_i}.$$

Suppose $\{h_i\}_{i=1}^k$ is a partition of unity subordinate to $\{U_{p_i}\}_{i=1}^k$; i.e., the support of h_i lies in U_{p_i},

$$\sum_{i=1}^k h_i = 1 \quad \text{and} \quad |\nabla_\theta(h_i^{1/2})| \le H. \tag{A-5}$$

For abbreviation, we write

$$\|\varphi\|_p = \left(\int_M |\varphi|^p\,dV_\theta\right)^{\frac{1}{p}}.$$

Therefore, for any function φ compactly supported in K, we have

$$
\begin{aligned}
\text{(A-6)}\quad \sum_{i=1}^{k}\|\varphi^2 h_i\|_{\frac{n+1}{n}}
&= \sum_{i=1}^{k}\|\varphi h_i^{1/2}\|_{2+\frac{2}{n}}^2\\
&\le (1+O(\delta))\,K(n,2)\sum_{i=1}^{k}\|\nabla_\theta(\varphi h_i^{1/2})\|_2^2\\
&\le (1+O(\delta))\,K(n,2)\sum_{i=1}^{k}\int(|\nabla_\theta\varphi|h_i^{1/2}+\varphi|\nabla_\theta(h_i^{1/2})|)^2\,dV_\theta\\
&\le (1+O(\delta))\,K(n,2)\\
&\quad\times\int\sum_{i=1}^{k}\big(|\nabla_\theta\varphi|^2h_i+2|\nabla_\theta\varphi|h_i^{1/2}|\varphi|\,|\nabla_\theta(h_i^{1/2})|+|\varphi|^2|\nabla_\theta(h_i^{1/2})|^2\big)dV_\theta\\
&\le (1+O(\delta))\,K(n,2)\big(\|\nabla_\theta\varphi\|_2^2+2kH\|\nabla_\theta\varphi\|_2\|\varphi\|_2+kH\|\varphi\|_2^2\big),
\end{aligned}
$$

where the first inequality follows from (A-4), the last inequality follows from (A-5) and

$$\Big(\sum_{i=1}^{k}h_i^{1/2}\Big)^2\le k\sum_{i=1}^{k}h_i=k$$

by Hölder's inequality.

For any $\epsilon>0$, we can choose δ small enough such that

$$(1+O(\delta))K(n,2)\le K(n,2)+\frac{\epsilon}{2}. \tag{A-7}$$

Since the last expression of (A-6) is independent of i, we establish the inequality

$$
\begin{aligned}
\|\varphi\|_{2+\frac{2}{n}}^2=\|\varphi^2\|_{\frac{n+1}{n}}=\Big\|\varphi^2\sum_{i=1}^{k}h_i\Big\|_{\frac{n+1}{n}}
&\le\sum_{i=1}^{k}\|\varphi^2h_i\|_{\frac{n+1}{n}}\\
&\le(1+O(\delta))K(n,2)\big(\|\nabla_\theta\varphi\|_2^2+2kH\|\nabla_\theta\varphi\|_2\|\varphi\|_2+kH\|\varphi\|_2^2\big)\\
&\le\Big(K(n,2)+\frac{\epsilon}{2}\Big)\big(\|\nabla_\theta\varphi\|_2^2+2kH\|\nabla_\theta\varphi\|_2\|\varphi\|_2+kH\|\varphi\|_2^2\big)\\
&\le\Big(K(n,2)+\frac{\epsilon}{2}\Big)\big((1+\epsilon)\|\nabla_\theta\varphi\|_2^2+C(\epsilon,k,H)\|\varphi\|_2^2\big),
\end{aligned}
$$

where we have used (A-7) and Young's inequality. Here $C(\epsilon, k, H)$ is a constant depending only on ϵ, k and H. This proves the assertion. □

Acknowledgements

The authors are grateful to the referee for valuable comments which improved the manuscript. Ho was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (No. 201531021.01) and Kim was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2011-0025674).

References

[Aubin 1976] T. Aubin, "Équations différentielles non linéaires et problème de Yamabe concernant la courbure scalaire", *J. Math. Pures Appl.* (9) **55**:3 (1976), 269–296. MR 0431287 Zbl 0336.53033

[Aubin 1998] T. Aubin, *Some nonlinear problems in Riemannian geometry*, Springer, Berlin, 1998. MR 1636569 Zbl 0896.53003

[Aviles and McOwen 1988] P. Aviles and R. C. McOwen, "Conformal deformation to constant negative scalar curvature on noncompact Riemannian manifolds", *J. Differential Geom.* **27**:2 (1988), 225–239. MR 925121 Zbl 0648.53021

[Bland and Kalka 1989] J. Bland and M. Kalka, "Negative scalar curvature metrics on noncompact manifolds", *Trans. Amer. Math. Soc.* **316**:2 (1989), 433–446. MR 987159 Zbl 0694.53041

[Brendle 2005] S. Brendle, "Convergence of the Yamabe flow for arbitrary initial energy", *J. Differential Geom.* **69**:2 (2005), 217–278. MR 2168505 Zbl 1085.53028

[Brendle 2007] S. Brendle, "Convergence of the Yamabe flow in dimension 6 and higher", *Invent. Math.* **170**:3 (2007), 541–576. MR 2357502 Zbl 1130.53044

[Chang and Cheng 2002] S.-C. Chang and J.-H. Cheng, "The Harnack estimate for the Yamabe flow on CR manifolds of dimension 3", *Ann. Global Anal. Geom.* **21**:2 (2002), 111–121. MR 1894940 Zbl 1007.53034

[Chang et al. 2010] S.-C. Chang, H.-L. Chiu, and C.-T. Wu, "The Li–Yau–Hamilton inequality for Yamabe flow on a closed CR 3-manifold", *Trans. Amer. Math. Soc.* **362**:4 (2010), 1681–1698. MR 2574873 Zbl 1192.32020

[Cheng et al. 2013] J.-H. Cheng, A. Malchiodi, and P. Yang, "A positive mass theorem in three dimensional Cauchy–Riemann geometry", preprint, 2013. arXiv 1312.7764

[Cheng et al. 2014] J.-H. Cheng, H.-L. Chiu, and P. Yang, "Uniformization of spherical *CR* manifolds", *Adv. Math.* **255** (2014), 182–216. MR 3167481 Zbl 1288.32051

[Chow 1992] B. Chow, "The Yamabe flow on locally conformally flat manifolds with positive Ricci curvature", *Comm. Pure Appl. Math.* **45**:8 (1992), 1003–1014. MR 1168117 Zbl 0785.53027

[Dragomir and Tomassini 2006] S. Dragomir and G. Tomassini, *Differential geometry and analysis on CR manifolds*, Progress in Mathematics **246**, Birkhäuser, Boston, 2006. MR 2214654 Zbl 1099.32008

[Gamara 2001] N. Gamara, "The CR Yamabe conjecture — the case $n = 1$", *J. Eur. Math. Soc.* (*JEMS*) **3**:2 (2001), 105–137. MR 1831872 Zbl 0988.53013

[Gamara and Yacoub 2001] N. Gamara and R. Yacoub, “CR Yamabe conjecture — the conformally flat case”, *Pacific J. Math.* **201**:1 (2001), 121–175. MR 1867895 Zbl 1054.32020

[Große and Nardmann 2014] N. Große and M. Nardmann, “The Yamabe constant on noncompact manifolds”, *J. Geom. Anal.* **24**:2 (2014), 1092–1125. MR 3192307 Zbl 1315.53028

[Ho 2012] P. T. Ho, “The long-time existence and convergence of the CR Yamabe flow”, *Commun. Contemp. Math.* **14**:2 (2012), 1250014, 50. MR 2901057 Zbl 1246.53087

[Ho 2013] P. T. Ho, “Results related to prescribing pseudo-Hermitian scalar curvature”, *Internat. J. Math.* **24**:3 (2013), 1350020, 29. MR 3048007 Zbl 1267.32034

[Ho 2015] P. T. Ho, “The Webster scalar curvature flow on CR sphere, I”, *Adv. Math.* **268** (2015), 758–835. MR 3276608 Zbl 1301.32028

[Jerison and Lee 1987] D. Jerison and J. M. Lee, “The Yamabe problem on CR manifolds”, *J. Differential Geom.* **25**:2 (1987), 167–197. MR 880182 Zbl 0661.32026

[Jerison and Lee 1988] D. Jerison and J. M. Lee, “Extremals for the Sobolev inequality on the Heisenberg group and the CR Yamabe problem”, *J. Amer. Math. Soc.* **1**:1 (1988), 1–13. MR 924699 Zbl 0634.32016

[Jerison and Lee 1989] D. Jerison and J. M. Lee, “Intrinsic CR normal coordinates and the CR Yamabe problem”, *J. Differential Geom.* **29**:2 (1989), 303–343. MR 982177 Zbl 0671.32016

[Jin 1988] Z. R. Jin, “A counterexample to the Yamabe problem for complete noncompact manifolds”, pp. 93–101 in *Partial differential equations* (Tianjin, 1986), edited by S. S. Chern, Lecture Notes in Math. **1306**, Springer, 1988. MR 1032773 Zbl 0648.53022

[Kazdan and Warner 1975] J. L. Kazdan and F. W. Warner, “Existence and conformal deformation of metrics with prescribed Gaussian and scalar curvatures”, *Ann. of Math.* (2) **101** (1975), 317–331. MR 0375153 Zbl 0297.53020

[Kim 1996] S. Kim, “Scalar curvature on noncompact complete Riemannian manifolds”, *Nonlinear Anal.* **26**:12 (1996), 1985–1993. MR 1386128 Zbl 0858.53029

[Kim 1997] S. Kim, “The Yamabe problem and applications on noncompact complete Riemannian manifolds”, *Geom. Dedicata* **64**:3 (1997), 373–381. MR 1440570 Zbl 0878.53037

[Kim 2000] S. Kim, “An obstruction to the conformal compactification of Riemannian manifolds”, *Proc. Amer. Math. Soc.* **128**:6 (2000), 1833–1838. MR 1646195 Zbl 0956.53032

[Lee and Parker 1987] J. M. Lee and T. H. Parker, “The Yamabe problem”, *Bull. Amer. Math. Soc.* (*N.S.*) **17**:1 (1987), 37–91. MR 888880 Zbl 0633.53062

[Lou 1998] Y. Lou, “Uniqueness and non-uniqueness of metrics with prescribed scalar curvature on compact manifolds”, *Indiana Univ. Math. J.* **47**:3 (1998), 1065–1081. MR 1665745 Zbl 0936.53028

[Schoen 1984] R. Schoen, “Conformal deformation of a Riemannian metric to constant scalar curvature”, *J. Differential Geom.* **20**:2 (1984), 479–495. MR 788292 Zbl 0576.53028

[Schwetlick and Struwe 2003] H. Schwetlick and M. Struwe, “Convergence of the Yamabe flow for “large” energies”, *J. Reine Angew. Math.* **562** (2003), 59–100. MR 2011332 Zbl 1079.53100

[Trudinger 1968] N. S. Trudinger, “Remarks concerning the conformal deformation of Riemannian structures on compact manifolds”, *Ann. Scuola Norm. Sup. Pisa* (3) **22** (1968), 265–274. MR 0240748 Zbl 0159.23801

[Ye 1994] R. Ye, “Global existence and convergence of Yamabe flow”, *J. Differential Geom.* **39**:1 (1994), 35–50. MR 1258912 Zbl 0846.53027

[Zhang 2003] Q. S. Zhang, “Finite energy solutions to the Yamabe equation”, *Geom. Dedicata* **101** (2003), 153–165. MR 2017900 Zbl 1058.58015

Received February 15, 2016. Revised April 18, 2016.

PAK TUNG HO
DEPARTMENT OF MATHEMATICS
SOGANG UNIVERSITY
SEOUL 121-742
SOUTH KOREA
ptho@sogang.ac.kr
paktungho@yahoo.com.hk

SEONGTAG KIM
DEPARTMENT OF MATHEMATICS EDUCATION
INHA UNIVERSITY
INCHEON 402-751
SOUTH KOREA
stkim@inha.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 285, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.285.393

ISOMETRY TYPES OF FRAME BUNDLES

WOUTER VAN LIMBEEK

We consider the oriented orthonormal frame bundle SO(M) of an oriented Riemannian manifold M. The Riemannian metric on M induces a canonical Riemannian metric on SO(M). We prove that for two closed oriented Riemannian n-manifolds M and N, the frame bundles SO(M) and SO(N) are isometric if and only if M and N are isometric, except possibly in dimensions 3, 4, and 8. This answers a question of Benson Farb except in dimensions 3, 4, and 8.

1. Introduction

Let M be an oriented Riemannian manifold, and let $X := \mathrm{SO}(M)$ be the oriented orthonormal frame bundle of M. The Riemannian structure g on M induces in a canonical way a Riemannian metric g_{SO} on $\mathrm{SO}(M)$. This construction was first carried out by O'Neill [1966] and independently by Mok [1978], and is very similar to Sasaki's [1958; 1962] construction of a metric on the unit tangent bundle of M, so we will henceforth refer to g_{SO} as the *Sasaki–Mok–O'Neill metric* on $\mathrm{SO}(M)$. Let us sketch the construction of g_{SO} and refer to Section 2 for the details. Consider the natural projection $\pi : \mathrm{SO}(M) \to M$. Each of the fibers of p is naturally equipped with a free and transitive $\mathrm{SO}(n)$-action, so that this fiber carries an $\mathrm{SO}(n)$-bi-invariant metric $g_{\mathcal{V}}$. The metric $g_{\mathcal{V}}$ is determined uniquely up to scaling. Further, the Levi-Civita connection on the tangent bundle $TM \to M$ induces a horizontal subbundle of TM. This in turn induces a horizontal subbundle $\mathcal{H}$ of $T\,\mathrm{SO}(M)$. We can pull

MSC2010: primary 53C24, 57S15; secondary 57R25, 55R10.
Keywords: frame bundle, Sasaki–Mok metric, large group action, symmetry degree.

back the metric on M along π to get a metric $g_{\mathcal{H}}$ on $\mathcal{H}$. The Sasaki–Mok–O'Neill metric on $\mathrm{SO}(M)$ is defined to be $g_{\mathrm{SO}} := g_{\mathcal{V}} \oplus g_{\mathcal{H}}$.

Note that g_{SO} is determined uniquely up to scaling of $g_{\mathcal{V}}$, and hence determined uniquely after fixing a bi-invariant metric on $\mathrm{SO}(n)$. The work of O'Neill [1966], Mok [1978], and later Takagi and Yawata [1991; 1994] has established many natural properties of Sasaki–Mok–O'Neill metrics and connections between the geometry of M and $\mathrm{SO}(M)$. The following natural question then arises, which was to my knowledge first posed by Benson Farb.

Question 1.1. Let M, N be Riemannian manifolds. If $\mathrm{SO}(M)$ is isometric to $\mathrm{SO}(N)$ (with respect to Sasaki–Mok–O'Neill metrics on each), is M isometric to N?

The purpose of this paper is to answer Question 1.1 except when $\dim M = 3, 4$ or 8. The question is a bit subtle, for it is not true in general that an isometry of $\mathrm{SO}(M)$ preserves the fibers of $\mathrm{SO}(M) \to M$, as shown by the following example.

Example 1.2. Let M be a constant curvature sphere S^n. Then $\mathrm{SO}(M)$ is diffeomorphic to $\mathrm{SO}(n+1)$. (To see this, identify S^n with the unit sphere in $\mathbb{R}^{n+1}$. If $p \in S^n$ and $v_1, \dots, v_n$ is a positively oriented orthonormal frame at p, then the matrix with columns $p, v_1, \dots, v_n$ belongs to $\mathrm{SO}(n+1)$.) There is a unique Sasaki–Mok–O'Neill metric that is isometric to the bi-invariant metric on $\mathrm{SO}(n+1)$. However, of course there are many isometries of $\mathrm{SO}(n+1)$ that do not preserve the fibers of $\mathrm{SO}(n+1) \to S^n$.

By differentiating the action of $\mathrm{SO}(n+1)$ in the above example, we obtain many Killing fields that do not preserve the fibers of $\mathrm{SO}(n+1) \to S^n$. However, by a theorem of Takagi and Yawata [1991], manifolds with constant positive curvature are the only Riemannian manifolds whose orthonormal frame bundles admit Killing fields that do not preserve the fibers. More examples of non-fiber-preserving isometries appear if we consider isometries that are not induced by Killing fields, as the following example shows.

Example 1.3. Let M be a flat 2-torus obtained as the quotient of $\mathbb{R}^2$ by the subgroup generated by translations by $(l_1, 0)$ and $(0, l_2)$ for some $l_1, l_2 > 0$. Further fix $l_3 > 0$ and equip $\mathrm{SO}(M)$ with the Sasaki–Mok–O'Neill metric associated to the scalar l_3. It is easy to see $\mathrm{SO}(M)$ is the flat 3-torus obtained as the quotient of $\mathbb{R}^3$ by the subgroup generated by translations by $(l_1, 0, 0)$, $(0, l_2, 0)$ and $(0, 0, l_3)$.

Now let N be the flat 2-torus obtained as the quotient of $\mathbb{R}^2$ by the subgroup generated by translations by $(l_1, 0)$ and $(0, l_3)$, and equip $\mathrm{SO}(N)$ with the Sasaki–Mok–O'Neill metric associated to the scalar l_2. Then $\mathrm{SO}(M)$ and $\mathrm{SO}(N)$ are isometric but if l_1, l_2, l_3 are distinct, M and N are not isometric.

On the other hand if $l_1 = l_3 \neq l_2$, then this construction produces an isometry $\mathrm{SO}(M) \to \mathrm{SO}(M)$ that is not a bundle map.

Example 1.3 produces counterexamples to Question 1.1. Note that we used different bi-invariant metrics $g_{\mathcal{V}}$ on the fibers. Therefore to give a positive answer to Question 1.1 we must normalize the volume of the fibers of $\mathrm{SO}(M) \to M$.

Our main theorem is that under the assumption of normalization, Question 1.1 has the following positive answer, except possibly in dimensions 3, 4 and 8.

Theorem A. *Let M, N be closed oriented connected Riemannian n-manifolds. Equip $\mathrm{SO}(M)$ and $\mathrm{SO}(N)$ with Sasaki–Mok–O'Neill metrics where the fibers of $\mathrm{SO}(M) \to M$ and $\mathrm{SO}(N) \to N$ have fixed volume $\nu > 0$. Assume $n \neq 3, 4, 8$. Then M, N are isometric if and only if $\mathrm{SO}(M)$ and $\mathrm{SO}(N)$ are isometric.*

We do not know if counterexamples to Question 1.1 exist in dimensions 3, 4, and 8.

Outline of proof. If $f : M \to N$ is an isometry, then the induced map

$$\mathrm{SO}(f) : \mathrm{SO}(M) \to \mathrm{SO}(N)$$

is also an isometry (see Proposition 2.5). This proves one direction of the theorem.

For the other direction, our strategy is to identify the fibers of the bundle $\mathrm{SO}(M) \to M$ using only the geometry of $\mathrm{SO}(M)$. To accomplish this, note that $X = \mathrm{SO}(M)$ carries an action of $\mathrm{SO}(n)$ by isometries, and the orbits of this action are exactly the fibers of $\mathrm{SO}(M) \to M$. This action gives rise to an algebra of Killing fields isomorphic to $\mathfrak{o}(n)$.

The full Lie algebra $i(X)$ of Killing fields on $X = \mathrm{SO}(M)$ has been computed by Takagi and Yawata [1994] except in dimensions 2, 3, 4 or 8, or when M has positive constant curvature. We show that if this computation applies, either $i(X)$ contains a unique copy of $\mathfrak{o}(n)$ or $\mathrm{Isom}(M)$ is extremely large or M is flat. If $i(X)$ contains a unique copy of $\mathfrak{o}(n)$, then the fibers of $X = \mathrm{SO}(M) \to M$ and $X = \mathrm{SO}(N) \to N$ coincide, and we deduce that M and N are isometric.

We are able to resolve the flat case separately. If $\mathrm{Isom}(M)$ is large we use classification theorems from the theory of compact transformation groups to prove that M and N are isometric.

Finally we prove the theorem in two situations where the computation of Takagi and Yawata does not apply, namely constant positive curvature and dimension 2. In these situations it is in general impossible to identify the fibers of $\mathrm{SO}(M) \to M$ using the geometry of $\mathrm{SO}(M)$ alone as shown by Examples 1.2 and 1.3. However, we are still able to obtain the main result using the scarcity of manifolds with a metric of constant positive curvature, and the classification of surfaces.

Outline of the paper. In Section 3 we will review preliminaries about actions of Lie groups G on a manifold M when $\dim G$ is large compared to $\dim M$. In Section 4 we will prove Theorem A except when M and N are surfaces or have metrics of constant positive curvature. The proof in the case that at least one of M or N has constant positive curvature will be given in Section 5. We prove Theorem A in the case that M and N are surfaces in Section 6.

2. Preliminaries

In this section we introduce the Sasaki–Mok–O'Neill metric, and we recall some basic properties. Then we discuss the classical relationship between isometries and Killing fields, and Takagi and Yawata's computations of Killing fields of Sasaki–Mok–O'Neill metrics. We end this section with a useful lemma for normalizing Sasaki–Mok–O'Neill metrics, and some general remarks about frame bundles of fiber bundles that will also be useful later.

Definition of the Sasaki–Mok–O'Neill metric. Our discussion here follows the construction of Mok [1978], where more details can be found. Let (M, g) be an oriented Riemannian manifold of dimension n, and let $X := \mathrm{SO}(M)$ be the oriented orthonormal frame bundle of M with natural projection map $\pi : \mathrm{SO}(M) \to M$. For $e \in \mathrm{SO}(M)$, the *vertical subspace* at e is defined to be $\mathcal{V}_e := \ker D_e\pi$. The collection of vertical subspaces forms a subbundle $\mathcal{V} \to TM$ of $T\,\mathrm{SO}(M) \to TM$.

Let ω be the Riemannian connection $\mathfrak{o}(n)$-valued 1-form associated to the Riemannian metric on M. Explicitly, if $p \in M$ and $e = (e_1, \dots, e_n)$ is a frame at p, we define for $X \in T_e\mathrm{SO}(M)$:

$$\omega_{ij}(X) := \theta_j(\nabla_X(e_i)) \quad (1 \le i, j \le n),$$

where θ_j is the form dual to e_j with respect to the Riemannian metric g.

We set $\mathcal{H}_e := \ker \omega_e$. We call $\mathcal{H}_e$ the *horizontal subspace* at e. We have a decomposition $T_e\mathrm{SO}(M) = \mathcal{V}_e \oplus \mathcal{H}_e$. Define an inner product on $T_e\mathrm{SO}(M)$ via

$$g_{\mathrm{SO}}(X, Y) = \langle \omega(X), \omega(Y) \rangle + g(\pi_* X, \pi_* Y),$$

where $\langle \cdot, \cdot \rangle$ is an $\mathrm{O}(n)$-invariant inner product on $\mathfrak{o}(n)$. Note that the choice of an $\mathrm{O}(n)$-invariant inner product on $\mathfrak{o}(n)$ is uniquely determined up to scaling by a positive number λ, so that we obtain a 1-parameter family of Sasaki–Mok–O'Neill metrics. Explicitly such an inner product is given by

$$\langle A, B \rangle_\lambda := -\lambda \operatorname{tr}(AB) = \lambda \sum_{i,j} A_{ij} B_{ij},$$

for $A, B \in \mathfrak{o}(n)$. We call $\langle \cdot, \cdot \rangle_1$ the *standard metric* on $\mathfrak{o}(n)$.

Remark 2.1. The oriented orthonormal frame bundle $\mathrm{SO}(M) \to M$ is an example of a $\mathrm{SO}(n)$-principal bundle of over M, and it has a natural connection form ω as defined above. For a principal G-bundle $E \to B$ with a principal connection form θ, one can construct a so called connection metric (see, e.g., [Ziller 2001, Section 1]).

The Sasaki–Mok–O'Neill metric is exactly this connection metric in the case of the principal $\mathrm{SO}(n)$-bundle $\mathrm{SO}(M) \to M$ with the connection form ω.

As mentioned above, the geometry of the above defined metric was first investigated by O'Neill and Mok. In particular they showed:

Proposition 2.2 [O'Neill 1966, p. 467, Mok 1978, Theorem 4.3]. *The fibers of* $\mathrm{SO}(M) \to M$ *are totally geodesic submanifolds of* $\mathrm{SO}(M)$ *with respect to any Sasaki–Mok–O'Neill metric.*

Vector fields on frame bundles. Let X be a vector field on $\mathrm{SO}(M)$. If $X_e \in \mathcal{V}_e$ for any $e \in \mathrm{SO}(M)$, we say X is *vertical*. If $X_e \in \mathcal{H}_e$ for any $e \in \mathrm{SO}(M)$, we say X is *horizontal*.

We will now discuss how to lift a vector field Y on M to a vector field X on $\mathrm{SO}(M)$ such that $\pi_* X = Y$. There are two useful constructions, called the *horizontal* and *complete* lift of Y. Both constructions start by considering the derivative of the bundle map $\pi : \mathrm{SO}(M) \to M$. For a frame $e \in \mathrm{SO}(M)$, we have a decomposition $T_e\mathrm{SO}(M) = \mathcal{V}_e \oplus \mathcal{H}_e$ as discussed above. Here $\mathcal{V}_e = \ker \pi_*$, and hence π_* restricts to an isomorphism $\mathcal{H}_e \to T_{\pi(e)}M$. Therefore for a vector field Y on M, there exists a unique horizontal vector field Y^H on M with $Y = \pi_* Y^H$. We call Y^H the *horizontal lift* of Y.

The *complete lift* Y^C of Y was first introduced in [Kobayashi and Nomizu 1963]. First observe that given a map $f : M \to M$, we can consider its induced map $\mathrm{SO}(f) : \mathrm{SO}(M) \to \mathrm{SO}(M)$ on frames. Then we can define Y^C as follows: Let φ_t be the 1-parameter family of diffeomorphisms of M obtained by integrating Y, so that $Y = \frac{d}{dt}\big|_{t=0} \varphi_t$. Then we define

$$Y^C := \frac{d}{dt}\bigg|_{t=0} \mathrm{SO}(\varphi_t).$$

Note that Y^C is in general neither vertical nor horizontal. Mok [1979, Section 3] has given a description of Y^C in terms of local coordinates.

Killing fields and isometries. Before considering the isometries of $\mathrm{SO}(M)$ equipped with a Sasaki–Mok–O'Neill metric g_{SO}, we will review some classical facts about the structure of the group of isometries $\mathrm{Isom}(M)$ of a Riemannian manifold M.

Myers and Steenrod [1939] have proved that $\mathrm{Isom}(M)$ of a Riemannian manifold is a Lie group. If $(h_t)_t$ is a 1-parameter group of isometries, then $Y := \frac{d}{dt}\big|_{t=0} h_t$ is a vector field on M. Differentiating the condition $h_t^* g = g$ gives the *Killing relation* for Y,

$$\mathcal{L}_Y g = 0, \tag{2-1}$$

where $\mathcal{L}$ is the Lie derivative. Any vector field Y satisfying equation (2-1) is called a Killing field. Given a Killing field Y on M, the 1-parameter group $(h_t)_t$ obtained by integrating Y consists of isometries. The Killing fields on M form a Lie algebra $i(M)$ of vector fields. We have:

Theorem 2.3. *Let M be a Riemannian manifold. Then* $\mathrm{Isom}(M)$ *is a Lie group* (*possibly not connected*), *with Lie algebra $i(M)$.*

The Takagi–Yawata theorem on Killing fields. We will now discuss a complete description due to Takagi and Yawata [1994] of the Killing fields on $\mathrm{SO}(M)$ in terms of the geometry of M for many manifolds M. Let us first discuss three constructions of Killing fields on $\mathrm{SO}(M)$.

For the first construction, recall that Sasaki [1958, Corollary 1] showed that whenever $f : M \to M$ is an isometry of M, the derivative $Df : TM \to TM$ is an isometry of TM (where TM is equipped with a Sasaki metric). Therefore if Y is a Killing field on M, then the complete lift of Y is a Killing field on TM. This is also true for frame bundles:

Proposition 2.4 [Mok 1978, Proposition 5.3]. *If Y is a Killing field on M, then Y^C is a Killing field on* $\mathrm{SO}(M)$ *with respect to any Sasaki–Mok–O'Neill metric.*

In fact the following more general statement is true:

Proposition 2.5. *Let M be a Riemannian manifold and $f : M \to M$ any isometry. Then the induced map* $\mathrm{SO}(f) : \mathrm{SO}(M) \to \mathrm{SO}(M)$ *is an isometry of* $\mathrm{SO}(M)$ *with respect to any Sasaki–Mok–O'Neill metric.*

Proof. Note that since the Riemannian connection form ω is canonically associated to the metric, we have $f^*\omega = \omega$. In particular $\mathrm{SO}(f)$ preserves the horizontal subbundle $\mathcal{H} := \ker \omega$. Also note that $\mathrm{SO}(f)$ is a bundle map of $\pi : \mathrm{SO}(M) \to M$ (i.e., we have $\mathrm{SO}(f) \circ \pi = \pi \circ f$), and in particular $\mathrm{SO}(f)$ preserves the vertical subbundle $\mathcal{V} := \ker \pi_*$. Using these facts it is easy to check $\mathrm{SO}(f)$ is an isometry. □

The second construction of Killing fields comes from the structure of $\mathrm{SO}(M) \to M$ as a principal $\mathrm{SO}(n)$-bundle. There is an action of $\mathrm{SO}(n)$ on the fibers of $\mathrm{SO}(M) \to M$, which is easily seen to be isometric with respect to any Sasaki–Mok–O'Neill metric. Differentiating any 1-parameter subgroup of $\mathrm{SO}(n)$ then gives a Killing field on $\mathrm{SO}(M)$. Explicitly, we can define these as follows: for $A \in \mathfrak{o}(n)$, define the vector field A^* on $\mathrm{SO}(M)$ via $\omega(A^*) = A$ and $\pi_*(A^*) = 0$, where ω is the connection form as above. Then A^* is a vertical Killing field. Write i_V^M for the Killing fields thus obtained. In particular $i_V^M \cong \mathfrak{o}(n)$ as Lie algebras.

Finally, here is the third construction of a Killing field on $\mathrm{SO}(M)$. Let φ be a 2-form on M, so that it defines a skew-symmetric bilinear form on every tangent space T_pM for $p \in M$. With respect to a frame e of T_pM, the skew-symmetric form φ_p can be represented as a skew-symmetric matrix $A_e \in \mathfrak{o}(n)$. We then define a vector field X^φ on $\mathrm{SO}(M)$ via $\omega_e(X_e^\varphi) := A_e$ and $\pi_*(X_e^\varphi) = 0$. Note that the latter condition just means that we define X^φ to be a vertical vector field. An explicit computation shows that if φ is parallel, then X^φ is a Killing field (see, e.g., [Takagi and Yawata 1991]). Denote by $(\Lambda^2 M)_0$ the Lie algebra of parallel 2-forms on M.

It is known that for many manifolds, these three constructions are the only ways of producing Killing fields on $\mathrm{SO}(M)$:

Theorem 2.6 [Takagi and Yawata 1994]. *Let M be a closed Riemannian manifold and equip $\mathrm{SO}(M)$ with the Sasaki–Mok–O'Neill metric corresponding to the standard inner product $\langle\,\cdot\,,\cdot\,\rangle_1$ on $\mathfrak{o}(n)$. Suppose M does not have constant curvature $\frac{1}{2}$ and* $\dim M \neq 2, 3, 4, 8$. *Then for any Killing field X on $\mathrm{SO}(M)$ there exist unique $Y \in i(M)$, $A \in \mathfrak{o}(n)$, and $\varphi \in (\Lambda^2 M)_0$ such that*

$$X = Y^C + A^* + X^\varphi.$$

Remark 2.7. Of course a version of the above result holds for different Sasaki–Mok–O'Neill metrics as well: If we use the inner product $\langle\,\cdot\,,\cdot\,\rangle_\lambda = \lambda\langle\,\cdot\,,\cdot\,\rangle_1$ on $\mathfrak{o}(n)$, the same conclusion holds except that we should now require that M does not have constant curvature $1/(2\sqrt{\lambda})$.

An explicit computation shows that if $Y \in i(M)$, $A \in \mathfrak{o}(n)$ and $\varphi \in (\Lambda^2 M)_0$, then the vector fields Y^C, A^* and X^φ pairwise commute. Combining this with Theorem 2.6, we obtain the following Lie algebra decomposition of Killing fields on $\mathrm{SO}(M)$:

Corollary 2.8 [Takagi and Yawata 1994]. *Let M be a Riemannian manifold satisfying the hypotheses of Theorem 2.6. Then there is a Lie algebra decomposition*

$$i(\mathrm{SO}(M)) = i(M) \oplus i_V^M \oplus (\Lambda^2 M)_0,$$

where $i(M)$ (resp. i_V^M, $(\Lambda^2 M)_0$) corresponds to the subalgebra of Killing fields consisting of Y^C (resp. A^*, X^φ) for $Y \in i(M)$ (resp. $A \in \mathfrak{o}(n)$, $\varphi \in (\Lambda^2 M)_0$).

Normalizing volume. Given a closed oriented Riemannian manifold M, we have previously obtained a 1-parameter family of Sasaki–Mok–O'Neill metrics on M. These can be parametrized by a choice of $O(n)$-invariant inner product on $\mathfrak{o}(n)$ (which is unique up to scaling), or, equivalently, by the volume of a fiber of $\mathrm{SO}(M) \to M$. The following easy lemma will be useful to us on multiple occasions in the rest of the paper.

Lemma 2.9. *Fix $v > 0$. Let M, N be closed orientable connected Riemannian n-manifolds and equip $\mathrm{SO}(M)$ and $\mathrm{SO}(N)$ with Sasaki–Mok–O'Neill metrics where the fibers of $\mathrm{SO}(M) \to M$ and $\mathrm{SO}(N) \to N$ have volume v. Suppose that $\mathrm{SO}(M)$ and $\mathrm{SO}(N)$ are isometric. Then* $\mathrm{vol}(M) = \mathrm{vol}(N)$.

Proof. Set $X := \mathrm{SO}(M) \cong \mathrm{SO}(N)$. Since the fiber bundle $X \to M$ has fibers with volume ν, we have $\mathrm{vol}(X) = \mathrm{vol}(M)/\nu$. Likewise we have $\mathrm{vol}(X) = \mathrm{vol}(N)/\nu$. Combining these we get $\mathrm{vol}(M) = \mathrm{vol}(N)$. □

3. High dimensional isometry groups of manifolds

In this section we review some known results about effective actions of a compact Lie group G on a closed n-manifold M when $\dim G$ is large compared to n. We will be especially interested in actions of $\mathrm{SO}(n)$ on an n-manifold M. First, there is the following classical upper bound for the dimension of a compact group acting smoothly on an n-manifold.

Theorem 3.1 [Kobayashi 1972, II.3.1]. *Let M be a closed n-manifold and G a compact group acting smoothly, effectively, and isometrically on M. Then* $\dim G \leq \frac{1}{2}n(n+1)$. *Further equality holds if and only if*

(i) *M is isometric to S^n with a metric of constant positive curvature, and we have $G = \mathrm{SO}(n+1)$ or $\mathrm{O}(n+1)$ acting on M in the standard way, or*

(ii) *M is isometric or $\mathbb{RP}^n$ with a metric of constant positive curvature, and $G = \mathrm{PSO}(n+1)$ or $\mathrm{PO}(n+1)$, acting on M in the standard way.*

Note that in the above case $G = \mathrm{Spin}(n+1)$ does not occur because there is no effective action on S^n or $\mathbb{RP}^n$. Theorem 3.1 leads us to study groups of dimension $< \frac{1}{2}n(n+1)$. First, there is the following remarkable "gap theorem" due to H. C. Wang.

Theorem 3.2 [Wang 1947]. *Let M be a closed n-manifold with $n \neq 4$. Then there is no compact group G acting effectively on M with*

$$\frac{n(n-1)}{2} + 1 < \dim G < \frac{n(n+1)}{2}.$$

Therefore the next case to consider is $\dim G = \frac{1}{2}n(n-1)+1$. The following characterization is independently due to Kuiper and Obata; see [Kobayashi 1972, II.3.3].

Theorem 3.3 (Kuiper, Obata). *Let M be a closed Riemannian n-manifold with $n > 4$ and G a connected compact group of dimension $\frac{1}{2}n(n-1)+1$ acting smoothly, effectively, and isometrically on M. Then M is isometric to $S^{n-1} \times S^1$ or $\mathbb{RP}^{n-1} \times S^1$ equipped with a product of a round metric on S^{n-1} or $\mathbb{RP}^{n-1}$ and the standard metric on S^1. Further $G = \mathrm{SO}(n) \times S^1$ or $\mathrm{PSO}(n) \times S^1$.*

After Theorem 3.3, the natural next case to consider is $\dim G = \frac{1}{2}n(n-1)$. There is a complete classification due to Kobayashi and Nagano.

Theorem 3.4 [Kobayashi and Nagano 1972]. *Let M be a closed Riemannian n-manifold with $n > 5$ and G a connected compact group of dimension $\frac{1}{2}n(n-1)$ acting smoothly, effectively, and isometrically on M. Then M must be one of the following.*

(1) *M is diffeomorphic to S^n or $\mathbb{RP}^n$ and $G = \mathrm{SO}(n)$ or $\mathrm{PSO}(n)$. In this case G has a fixed point on M. Every orbit is either a fixed point or has codimension* 1. *Regarding S^n as the solution set of $\sum_{i=0}^{n} x_i^2 = 1$ in $\mathbb{R}^{n+1}$, the metric on M (or its double cover if M is diffeomorphic to $\mathbb{RP}^n$) is of the form*

$$ds^2 = f(x_0) \sum_{i=0}^{n} dx_i^2$$

for a smooth positive function f on $[-1, 1]$.

(2) *M is diffeomorphic to a quotient $(L \times \mathbb{R})/\Gamma$ where $L = S^{n-1}$ or $L = \mathbb{RP}^{n-1}$ and $G = \mathrm{SO}(n)$ or $\mathrm{PSO}(n)$. Further, we have $\Gamma \cong \mathbb{Z}$. If $L = S^{n-1}$, then Γ is generated either by the map $(v, t) \mapsto (v, t+1)$ or by $(v, t) \mapsto (-v, t+1)$. If $L = \mathbb{RP}^{n-1}$, then Γ is generated by the map $(x, t) \mapsto (x, t+1)$. In all cases the projection on the second coordinate $S^{n-1} \times \mathbb{R} \to \mathbb{R}$ descends to a map $M \to S^1$ that is a fiber bundle with fibers diffeomorphic to L. The G-action preserves the fibers of $M \to S^1$ and restricts to an orthogonal action on each fiber.*

(3) *M is a quotient $(S^{n-1} \times \mathbb{R})/\Gamma$ where Γ is generated by*

$$(v, t) \mapsto (v, t+2) \quad \textit{and} \quad (v, t) \mapsto (-v, -t).$$

In this case $G = \mathrm{SO}(n)$ acts on $S^{n-1} \times \mathbb{R}$ by acting orthogonally on each copy $S^{n-1} \times \{t\}$. This action commutes with the action of Γ, so that the G-action descends to M. We have $M/G = [0, 1]$. The G-orbits lying over the endpoints 0, 1 *are isometric to round projective spaces $\mathbb{RP}^{n-1}$ and the G-orbits lying over points in* $(0, 1)$ *are round spheres.*

(4) *If $n = 6$ there is the additional case that $M \cong \mathbb{CP}^3$, equipped with the Fubini–Study metric and the standard action of $G = \mathrm{SO}(6) \cong \mathrm{SU}(4)/\{\pm \mathrm{id}\}$.*

(5) *If $n = 7$ there are the additional cases $M \cong \mathrm{Spin}(7)/G_2$ and $G = \mathrm{Spin}(7)$, or $M \cong \mathrm{SO}(7)/G_2$ and $G = \mathrm{SO}(7)$. In this case M is isometric to S^7 or $\mathbb{RP}^7$ with a constant curvature metric.*

Remark 3.5. Actually Kobayashi and Nagano prove a more general result that includes the possibility that M is noncompact, and there are more possibilities. Since we will not need the noncompact case, we have omitted these. In their formulation of case (4), M is a manifold of complex dimension 3 with constant holomorphic sectional curvature, and G is the largest connected group of holomorphic isometries.

Specializing to the compact case gives an explicit description of case (4) as follows. Hawley [1953] and Igusa [1954] independently proved that a simply connected complex n-manifold of constant holomorphic sectional curvature is isometric to either $\mathbb{C}^n$, $\mathbb{B}^n$ or $\mathbb{CP}^n$ (with standard metrics). Therefore in case (4) we obtain that M is isometric to $\mathbb{CP}^3$ (equipped with a scalar multiple of the Fubini–Study metric) and $G = \mathrm{SO}(6) \cong \mathrm{SU}(4)/\{\pm\,\mathrm{id}\}$.

Remark 3.6. If M admits the description in case (2) above, and is in addition assumed to be orientable, it follows that the bundle $M \to S^1$ is trivial. In particular M is diffeomorphic to $L \times S^1$.

To see this, note that the only other case to consider is that $M = (S^{n-1} \times \mathbb{R})/\Gamma$ where $\Gamma \cong \mathbb{Z}$ is generated by the map $(v, t) \mapsto (-v, t+1)$. This is a bundle with monodromy $-\mathrm{id} \in \mathrm{Diff}(S^{n-1})$. Two bundles over S^1 are equivalent if and only if their monodromies are isotopic (i.e., belong to the same component of $\mathrm{Diff}(S^{n-1})$. So let us check that $-\mathrm{id}$ is isotopic to the identity map: Indeed, because M is orientable, the map $(v, t) \mapsto (-v, t+1)$ is orientation preserving on $S^{n-1} \times \mathbb{R}$. It follows that n is even, so that $-\mathrm{id} \in \mathrm{SO}(n)$ and hence is clearly isotopic to the identity map.

Theorem 3.4 does not cover the case $n = 5$. In the following proposition we resolve this case for semisimple groups. We would like to thank an anonymous referee for the following statement and its proof, which improve upon those contained in an earlier version of this paper.

Proposition 3.7. *Let M be a closed oriented Riemannian 5-manifold and suppose G is a semisimple compact connected Lie group that acts on M smoothly, effectively and isometrically, and that* $\dim(G) = 10$. *Then M admits a description as in cases* (1), (2) *or* (3) *of Theorem 3.4.*

Proof. The proof of Theorem 3.4 (see [Kobayashi and Nagano 1972, Section 3]) shows that the assumption that $n > 5$ is only used to show that no G-orbit has codimension 2. We will show under the stated assumptions there are still no codimension 2 orbits, so that the rest of the proof of Theorem 3.4 applies.

Clearly we can assume that G is connected. Note that $\dim(G) = \mathrm{rk}(G) + 2k$, where k is the number of root spaces of G. Hence the rank of G is even. Any semisimple Lie group with rank ≥ 4 has dimension > 10, so that we must have that $\mathrm{rk}(G) = 2$ and therefore G is a quotient of $\mathrm{Spin}(5)$.

Suppose now that $x \in M$ and that the orbit $G(x)$ has codimension 2 in M. Let G_x be the stabilizer of x. Note that G_x has rank either 1 or 2, and since the orbit of x is codimension 2, we must have that $\dim G_x = 7$.

If G_x has rank 1, then it must be S^1 or $\mathrm{Spin}(3)$ (possibly up to a finite quotient), but then we see that $\dim G_x < 7$, so this is impossible.

On the other hand if $\mathrm{rk}(G_x)=2$, then the dimension of G_x is even, which is also a contradiction. □

4. Geometric characterization of the fibers of $\mathrm{SO}(M) \to M$

We will now start the proof of Theorem A. In this section we aim to prove the following theorem, which proves Theorem A in all cases except for round spheres and surfaces. The remaining cases are resolved in Section 5 (round spheres) and Section 6 (surfaces).

Theorem 4.1. *Let M, N be closed oriented connected Riemannian n-manifolds and fix $\lambda > 0$. Equip $\mathrm{SO}(M)$ and $\mathrm{SO}(N)$ with Sasaki–Mok–O'Neill metrics using the metric $\langle\cdot,\cdot\rangle_\lambda$ on $\mathfrak{o}(n)$. Assume that $n \neq 2, 3, 4, 8$ and that M does not have constant curvature $1/(2\sqrt{\lambda})$. Then M, N are isometric if and only if $\mathrm{SO}(M)$ and $\mathrm{SO}(N)$ are isometric.*

Proof. Write $X := \mathrm{SO}(M) \cong \mathrm{SO}(N)$, and let

$$\pi_M : X \to M \quad \text{and} \quad \pi_N : X \to N$$

be the natural projections. The strategy of the proof is to characterize the fibers of π_M and π_N just in terms of the geometry of X, except when M is flat or $\mathrm{Isom}(M)$ has dimension at least $\frac{1}{2}n(n-1)$. It automatically follows that in all but the exceptional cases the fibers of π_M and π_N must agree, and we will use this to show that M and N are isometric. Finally we will show that in the exceptional cases M and N also have to be isometric.

Note that the assumptions of Theorem 4.1 guarantee that we can use Takagi and Yawata's computation of the Lie algebra of Killing fields on X, so we can write (see Corollary 2.8)

$$i(X) = i(M) \oplus i_V^M \oplus (\Lambda^2 M)_0.$$

Here, as before, $i(M)$ denotes the space of Killing fields on M, and i_V^M consists of the Killing fields A^* for $A \in \mathfrak{o}(n)$ (in particular $i_V^M \cong \mathfrak{o}(n)$), and $(\Lambda^2 M)_0$ denotes the space of parallel 2-forms on M. On the other hand, the natural action of $\mathrm{SO}(n)$ on the fibers of π_N induces an embedding of $\mathrm{SO}(n)$ in $\mathrm{Isom}(X)$, hence an embedding of Lie algebras

$$\mathfrak{o}(n) \cong i_V^N \hookrightarrow i(X) = i_V^M \oplus (\Lambda^2 M)_0 \oplus i(M).$$

We identify i_V^N with its image throughout. Now consider the projections of i_V^N onto each of the factors of this decomposition. We have the following cases:

(1) $i_V^N = i_V^M$, or

(2) i_V^N projects nontrivially to $(\Lambda^2 M)_0$, or

(3) i_V^N projects trivially to $(\Lambda^2 M)_0$ but nontrivially to $i(M)$.

We will show below that these cases correspond to (1) the fibers of π_M coincide with the fibers of π_N, (2) M is flat, and (3) $\dim \operatorname{Isom}(M) \geq \frac{1}{2}n(n-1)$. We will complete the proof of Theorem 4.1 in each of these cases below.

***Case* 1** (vertical directions agree). Assume that $i_V^N = i_V^M$. For any $x \in X$, the values of i_V^M at x, i.e., the set of vectors

$$\{Z(x) \mid Z \in i_V^M\},$$

span the tangent space to the fiber of π_M through x. On the other hand, this set also spans the tangent space to the fiber of π_N through x. It follows that the fibers of π_M and π_N actually coincide. Hence we have a natural map $f : M \to N$ defined as follows: For $p \in M$, let $x \in \pi_M^{-1}(p)$ be any point in the fiber of π_M over p. Then set $f(p) := \pi_N(x)$. The fact that the fibers of π_M and π_N coincide proves that $f(p)$ does not depend on the choice of x.

We claim f is an isometry. Denote by $\mathcal{H}^M$ and $\mathcal{V}^M$ the horizontal and vertical subbundles with respect to $\pi_M : X \to M$. Because π_M is a Riemannian submersion, the metric on $T_x M$ coincides with the metric on the horizontal subbundle $\mathcal{H}_u^M$ at a point $u \in \pi_M^{-1}(x)$. We have

$$\mathcal{H}_u^M = (\mathcal{V}_u^M)^{\perp} = (\ker(\pi_M)_*)^{\perp} = (\ker(\pi_N)_*)^{\perp}.$$

Here the first identity is because by definition of the Sasaki–Mok–O'Neill metric on X, the horizontal and vertical subbundles are orthogonal. The last identity follows because we know the fibers of π_M and π_N agree. Finally, note that the space $(\ker(\pi_N)_*)^{\perp}$ is just the horizontal subbundle of $\pi_N : X \to N$. Since π_N is a Riemannian submersion, we conclude that the metric on $\mathcal{H}_u^M$ coincides with the metric on $T_{\pi_N(u)} N$. This proves the naturally induced map $f : M \to N$ is a local isometry. Since f is also injective, M and N are isometric.

***Case* 2** (many parallel forms). Assume that $i_V^N \cong \mathfrak{o}(n)$ projects nontrivially to $(\Lambda^2 M)_0$. Note that the kernel of the projection of i_V^N to $(\Lambda^2 M)_0$ is an ideal in i_V^N. On the other hand $i_V^N \cong \mathfrak{o}(n)$ is simple (because $n > 4$), so the projection $i_V^N \to (\Lambda^2 M)_0$ must be an isomorphism onto its image. Therefore

$$\dim(\Lambda^2 M)_0 \geq \dim \mathfrak{o}(n) = \frac{n(n-1)}{2}. \tag{4-1}$$

We claim that we actually have equality in equation (4-1). To see this, note that since a parallel form is invariant under parallel transport, it is determined by its values on a single tangent space, so that we have an embedding

$$(\Lambda^2 M)_0 \hookrightarrow \Lambda^2 T_x M. \tag{4-2}$$

Therefore $\dim(\Lambda^2 M)_0 \leq \frac{1}{2}n(n-1)$, and equality in equation (4-1) holds. Hence by a dimension count, the projection $i_V^N \to (\Lambda^2 M)_0$ is not only injective, but also surjective.

So we have $\mathfrak{o}(n) \cong (\Lambda^2 M)_0$, and M has the maximal amount of parallel forms it can possibly have (i.e., a space of dimension $\frac{1}{2}n(n-1)$). Note that a torus is an example of such a manifold. Motivated by these examples, we claim that M is a flat manifold.

To prove that M is flat, let us first show that for any $x \in M$, the holonomy group at x is trivial. Recall that the holonomy group consists of linear maps $T_x M \to T_x M$ obtained by parallel transport along a loop in M based at x. Therefore any holonomy map will fix parallel forms pointwise. Suppose now that $T : T_x M \to T_x M$ is a holonomy map at $x \in M$. We showed above that the evaluation at x is an isomorphism $(\Lambda^2 M)_0 \hookrightarrow \Lambda^2 T_x M$ (see equation (4-2)). Since T fixes parallel forms, it is therefore clear that $\Lambda^2 T = \mathrm{id}$ (i.e., T acts trivially on oriented planes in $T_x M$). Since $\dim(M) > 2$, it follows that $T = \mathrm{id}$.

So M has trivial holonomy. Since the holonomy algebra (i.e., the Lie algebra of the holonomy group) contains the Lie algebra generated by curvature operators $R(v, w)$ where $v, w \in T_x M$ (see, e.g., [Petersen 2006, Section 8.4]), it follows that $R(v, w) = 0$ for all $v, w \in T_x M$, so M is flat.

We will use that M is flat to obtain more information about the Killing fields $i(M)$ of M. Recall that the structure of flat manifolds is described by the Bieberbach theorems. Namely, any closed flat manifold is of the form $\mathbb{R}^n/\Gamma$ for some discrete torsion-free subgroup $\Gamma \subseteq \mathrm{Isom}(\mathbb{R}^n)$, and there is a finite index normal subgroup $\Lambda \subseteq \Gamma$ that consists of translations of $\mathbb{R}^n$ (so $\mathbb{R}^n/\Lambda$ is a torus). In particular the Killing fields on $\mathbb{R}^n/\Lambda$ are just obtained by translations of $\mathbb{R}^n$, so $i(\mathbb{R}^n/\Lambda) \cong \mathbb{R}^n$ as a Lie algebra.

The Killing fields on $M = \mathbb{R}^n/\Gamma$ are exactly those Killing fields of $\mathbb{R}^n/\Lambda$ invariant under the deck group Γ/Λ of the (regular) cover $\mathbb{R}^n/\Lambda \to M$. In particular $i(M)$ is a Lie subalgebra of $\mathbb{R}^n$.

Therefore $i(M)$ is abelian. Recall that we have

$$i(X) \cong i_V^M \oplus (\Lambda^2 M)_0 \oplus i(M).$$

We know that $i_V^N \cong \mathfrak{o}(n)$ has no abelian quotients, so we must have $i_V^N \subseteq i_V^M \oplus (\Lambda^2 M)_0$. Hence for any $x \in N$ and $\tilde{x} \in \pi_N^{-1}(x)$, we have

$$T_{\tilde{x}} \pi_N^{-1}(x) = i_V^N\big|_{\tilde{x}} \subseteq \big(i_V^M \oplus (\Lambda^2 M)_0\big)\big|_{\tilde{x}} \subseteq T_{\tilde{x}} \pi_M^{-1}(\pi_M(\tilde{x})),$$

where the last inclusion holds since the vector fields in $i_V^M \oplus (\Lambda^2 M)_0$ are vertical with respect to π_M (see page 398). Since $\pi_N^{-1}(x)$ and $\pi_M^{-1}(\pi_M(\tilde{x}))$ are connected submanifolds with the same dimension, we must have $\pi_N^{-1}(x) = \pi_M^{-1}(\pi_M(\tilde{x}))$.

Therefore the fibers of π_M and π_N agree. We conclude that M and N are isometric in the same way as case 1.

***Case* 3** (many Killing fields). Assume i_V^N projects nontrivially to $i(M)$. Again we use that $\mathfrak{o}(n)$ is a simple Lie algebra because we have $n > 4$. By assumption $i_V^N \cong \mathfrak{o}(n)$ projects nontrivially to $i(M)$, hence i_V^N projects isomorphically to $i(M)$. Let $\mathfrak{h}$ be the image of i_V^N in $i(M)$. At this point we would like to say that $i_V^N \subseteq i(M)$. We cannot in general establish this, but we have the following.

Claim 4.2. *Assume that* $\mathfrak{o}(n) \nsubseteq (\Lambda^2 M)_0$ *and that* $\mathfrak{o}(n) \nsubseteq (\Lambda^2 N)_0$. *Then*

(1) $i_V^N \subseteq i(M)$, *and*

(2) $i_V^M \subseteq i(N)$.

Therefore M and N have isometry groups of dimension $\geq \frac{1}{2}n(n-1)$.

Proof. Note that i_V^M and $\mathfrak{h}$ centralize each other and are isomorphic to $\mathfrak{o}(n)$. Consider the projection

$$p_1 : \mathfrak{h} \oplus i_V^M \subseteq i(X) \cong i_V^N \oplus (\Lambda^2 N)_0 \oplus i(N) \to i_V^N.$$

Note that $\dim(\mathfrak{h} \oplus i_V^M) = 2 \dim i_V^N$, so p_1 cannot be injective. If p_1 is trivial, then we have

$$\mathfrak{h} \oplus i_V^M \subseteq (\Lambda^2 N)_0 \oplus i(N).$$

Using again that $\mathfrak{o}(n)$ is simple, and since $(\Lambda^2 N)_0$ does not contain a copy of $\mathfrak{o}(n)$ by assumption, we must have that $\mathfrak{h} \oplus i_V^M$ projects isomorphically to $i(N)$. However note that $\dim i(N) \leq \frac{1}{2}n(n+1)$ by Theorem 3.1. Again by comparing dimensions we see that this is impossible. Therefore $\ker p_1$ is a proper ideal of $\mathfrak{h} \oplus i_V^M$, so $\ker p_1$ is either $\mathfrak{h}$ or i_V^M.

Now consider the projection

$$p_2 : \mathfrak{h} \oplus i_V^M \subseteq i(X) \cong i_V^N \oplus (\Lambda^2 N)_0 \oplus i(N) \to i(N).$$

As above we see that p_2 can be neither injective nor trivial. Hence we have that $\ker p_2$ is either $\mathfrak{h}$ or i_V^M.

If $\ker p_2 = i_V^M$, then we have $i_V^M = i_V^N$, but this contradicts the assumption that i_V^N projects nontrivially to $i(M)$. Therefore we must have $\ker p_1 = i_V^M$ and $\ker p_2 = \mathfrak{h}$. The latter implies $i_V^N = \mathfrak{h}$, which proves (1).

Since $\ker p_1 = i_V^M$, we have $i_V^M \subseteq (\Lambda^2 N)_0 \oplus i(N)$ and i_V^M projects trivially to $(\Lambda^2 N)_0$. Therefore we have $i_V^M \subseteq i(N)$, which proves (2). □

If $\mathfrak{o}(n) \subseteq (\Lambda^2 M)_0$ or $\mathfrak{o}(n) \subseteq (\Lambda^2 N)_0$, the proof is finished in case 2. Therefore we assume $i_V^N \subseteq i(M)$ and $i_V^M \subseteq i(N)$. Write $H_M := \exp(i_V^N)$ and $H_N := \exp(i_V^M)$, where exp is the exponential map on the Lie group $\mathrm{Isom}(X)$. Then H_M and H_N are subgroups of $\mathrm{Isom}(X)$, each isomorphic to $\mathrm{SO}(n)$, and $M = X/H_N$ and $N = X/H_M$.

Since H_M and H_N are commuting subgroups of $\mathrm{Isom}(X)$, the action of H_M on X descends to an action on $M = X/H_N$ with kernel $H_M \cap H_N$. We will write $\overline{H}_M := H_M/(H_M \cap H_N)$ for the group of isometries of M thus obtained. Similarly, H_N acts by isometries on $N = X/H_M$ with kernel $H_M \cap H_N$, and we will write $\overline{H}_N := H_N/(H_M \cap H_N)$ for this group of isometries.

Note that $H_M \cap H_N$ is discrete, since its Lie algebra is $i_V^M \cap i_V^N = 0$. In particular, since H_M and H_N are compact, it follows that $H_M \cap H_N$ is finite. Therefore the natural quotient map $H_M \to \overline{H}_M$ is a covering of finite degree, and $\overline{H}_M$ and H_M have the same Lie algebra. Similarly, $\overline{H}_N$ and H_N have the same Lie algebra. Therefore $\overline{H}_M$ and $\overline{H}_N$ are groups of isometries of closed n-manifolds with Lie algebras isomorphic to $\mathfrak{o}(n)$. The results of Section 3 exactly apply to such actions; these results will restrict the possibilities for M and N tremendously, as we will see below.

Motivated by the results of Section 3, we will now consider two cases: either one of $\overline{H}_M$ or $\overline{H}_N$ acts transitively, or neither acts transitively.

***Case* 3(*a*)** ($\overline{H}_M$ or $\overline{H}_N$ acts transitively). Suppose $\overline{H}_M$ acts transitively on M. Since $\overline{H}_M$ has Lie algebra $\mathfrak{o}(n)$ and $\dim M = n$, Theorem 3.4 and Proposition 3.7 give a classification of the possibilities for M and $\overline{H}_M$. Since in cases (1), (2), and (3) of Theorem 3.4 the group of isometries is not transitive, but by assumption $\overline{H}_M$ acts transitively on M, we know that either

- M is isometric to $S^7 \cong \mathrm{Spin}(7)/G_2$, equipped with a constant curvature metric, and $\overline{H}_M = \mathrm{Spin}(7)$, or
- M is isometric to $\mathbb{RP}^7 \cong \mathrm{SO}(7)/G_2$, equipped with a constant curvature metric, and $\overline{H}_M = \mathrm{SO}(7)$, or
- M is isometric to $\mathbb{CP}^3$, equipped with a metric of constant holomorphic sectional curvature, and $\overline{H}_M = \mathrm{SO}(6) \cong \mathrm{SU}(4)/\{\pm \mathrm{id}\}$.

We will show that the first case is impossible, and that in the other cases M and N are isometric.

Lemma 4.3. *M is not isometric to S^7.*

Proof. Since $\overline{H}_M = H_M/(H_M \cap H_N)$, we know that $\overline{H}_M$ is a quotient of $H_M \cong \mathrm{SO}(7)$. In particular, H_M is not simply connected. On the other hand, $\mathrm{Spin}(7)$ is simply connected. This is a contradiction. □

Lemma 4.4. *If M is isometric to $\mathbb{RP}^7$, then M and N are isometric.*

Proof. Suppose now M is isometric to $\mathbb{RP}^7$, and consider the action of H_N on N. From the classification in Theorem 3.4 and Remark 3.6, and using that $\dim(N) = \dim(M) = 7$, we see that N must be diffeomorphic to one of the following:

(1) $\mathbb{RP}^7$,

(2) S^7,

(3) $L_N \times S^1$ where L_N is S^6 or $\mathbb{RP}^6$, or

(4) $(S^6 \times \mathbb{R})/\Gamma$ where $\Gamma \cong D_\infty$ is generated by

$$(v, t) \mapsto (-v, -t) \quad \text{and} \quad (v, t) \mapsto (v, t+2).$$

Claim 4.5. *We must have that N is diffeomorphic to $\mathbb{RP}^7$ (and hence to M).*

Proof. We will show that we can distinguish the frame bundles of the manifolds appearing in cases (2), (3), and (4) from $\mathrm{SO}(\mathbb{RP}^7)$ by their fundamental group.

First, let us compute the fundamental group of $\mathrm{SO}(N) = \mathrm{SO}(\mathbb{RP}^7)$. Note that $\mathrm{SO}(S^7) \cong \mathrm{SO}(8)$ (see Example 1.2). It easily follows that $\mathrm{SO}(\mathbb{RP}^7) \cong \mathrm{SO}(8)/\{\pm \mathrm{id}\}$. In particular, $\pi_1 \mathrm{SO}(\mathbb{RP}^7)$ is obtained as an extension

$$1 \to \pi_1 \mathrm{SO}(8) \to \pi_1 \mathrm{SO}(\mathbb{RP}^7) \to \{\pm \mathrm{id}\} \to 1.$$

So $\pi_1(\mathrm{SO}(\mathbb{RP}^7))$ has order 4. So let us now show that in each of the cases (2), (3), and (4), π_1 does not have order 4.

- In case (2), note that $\pi_1 \mathrm{SO}(S^7) = \pi_1 \mathrm{SO}(8) \cong \mathbb{Z}/(2\mathbb{Z})$ has order 2.
- In case (3), $\pi_1 N$ is infinite. By the long exact sequence on homotopy groups for the fiber bundle $\mathrm{SO}(7) \to \mathrm{SO}(N) \to N$, we see that $\pi_1 \mathrm{SO}(N)$ surjects onto $\pi_1 N$. Therefore $\pi_1 \mathrm{SO}(N)$ is also infinite.
- In case (4), $\pi_1 N \cong D_\infty$ is infinite as well. The above argument for case (3) shows that $\pi_1 \mathrm{SO}(N)$ is infinite as well.

The only remaining possibility is that N is diffeomorphic to $\mathbb{RP}^7$ (and hence also to M). □

So we find that N is diffeomorphic to $\mathbb{RP}^7$. We will now determine the metric on N:

Claim 4.6. *N has constant curvature.*

Proof. Theorem 3.4 classifies the possible metrics on N. Namely, if $\overline{H}_N$ acts transitively on N, then N has constant curvature, as desired.

Suppose now that H_N does not act transitively on N. Identify the universal cover $\widetilde{N}$ of N (which is diffeomorphic to S^7) with the solution set of $\sum_{i=0}^{7} x_i^2 = 1$ in $\mathbb{R}^8$. Then by Theorem 3.4(3) the metric on $\widetilde{N}$ is of the form

$$ds^2_{\widetilde{N}} = f(x_0) \sum_{i=0}^{7} dx_i^2,$$

for some smooth positive function f on $[-1, 1]$. The function $|x_0|$ descends from $\widetilde{N}$ to N, and $\overline{H}_N$ acts isometrically and transitively on each level set

$$\left\{[x_0, \ldots, x_7] \in \mathbb{RP}^7 \;\middle|\; \sum_{i=1}^{7} x_i^2 = 1 - c^2\right\},$$

for $0 \leq c \leq 1$. For $c = 0$ this level set is a copy of $\mathbb{RP}^6$ (the image of the equator $S^6 \subseteq S^7 \cong \widetilde{N}$ in $\mathbb{RP}^7 \cong N$) and for $c = 1$ the level set consists of a single point (the image of the north and south pole). For $0 < c < 1$, the level set is a copy of S^6.

Let $x \in N$ be any point with $0 < x_0 < 1$, so that the $\overline{H}_N$-orbit of x is a copy of S^6. Since the metric on $\overline{H}_N x$ is given by $f(x_0) \sum_i dx_i^2$, we have

$$\mathrm{vol}(\overline{H}_N x) = (f(x_0))^{\frac{1}{2}n} \,\mathrm{vol}(S^6),$$

where on the right-hand side $\mathrm{vol}(S^6)$ is computed with respect to the standard metric $\sum_i dx_i^2$. Now consider the fiber bundle $\pi_N : \mathrm{SO}(N) \to N$. Recall that each fiber in $\mathrm{SO}(N)$ has a fixed volume $\nu > 0$, and is an H_M-orbit. Therefore for $e \in \pi_N^{-1}(x)$, we have

$$\mathrm{vol}(H_M H_N e) = \nu \,\mathrm{vol}(\overline{H}_N x) = \nu (f(x_0))^{\frac{6}{2}} \,\mathrm{vol}(S^6). \tag{4-3}$$

On the other hand, e is a frame at some point $y \in M$. Since the fibers of $\mathrm{SO}(M) \to M$ also have volume ν, it follows that

$$\mathrm{vol}(H_M H_N e) = \nu \,\mathrm{vol}(\overline{H}_M y).$$

Since $\overline{H}_M$ acts transitively on M, the right-hand side is just equal to $\nu \,\mathrm{vol}(M)$. In particular, the left-hand side does not depend on e. Using equation (4-3), we see that $f(x_0)$ does not depend on the point x chosen. Since the only requirements for x were that $-1 < x_0 < 1$ and $x_0 \neq 0$, we see that f is constant on $(-1, 1) \setminus \{0\}$. Since f is also continuous, it is in fact constant on $[-1, 1]$, so the metric on $\widetilde{N}$ is given by

$$ds^2_{\widetilde{N}} = c \sum_{i=0}^{7} dx_i^2,$$

for some $c > 0$. Therefore the metric is some multiple of the standard round metric, so N has constant curvature. □

So we have shown that both M and N are diffeomorphic to $\mathbb{RP}^7$ with constant curvature metrics. Since by Lemma 2.9, we also have that $\mathrm{vol}(M) = \mathrm{vol}(N)$, it follows that M and N have the same curvature, so that they are isometric, as desired. □

Lemma 4.7. *If M is isometric to $\mathbb{CP}^3$, equipped with a metric of constant holomorphic sectional curvature, then M and N are isometric.*

Proof. Again consider the action of $\overline{H}_N$ on N. From the classification in Theorem 3.4, and using that $\dim(N) = \dim(M) = 6$, we see that N must be one of the following:

(1) diffeomorphic to S^6 or $\mathbb{RP}^6$,

(2) diffeomorphic to $L_N \times S^1$ where L_N is S^5 or $\mathbb{RP}^5$,

(3) diffeomorphic to $(S^5 \times \mathbb{R})/\Gamma$ where $\Gamma \cong D_\infty$ is generated by $(v, t) \mapsto (-v, -t)$ and $(v, t) \mapsto (v, t+2)$, or

(4) isometric to $\mathbb{CP}^3$ with a metric of constant holomorphic sectional curvature.

We can rule out cases (1), (2), and (3) by computations of π_2. Namely, let us first compute $\pi_2(\mathrm{SO}(\mathbb{CP}^3))$. The long exact sequence on homotopy groups of the fibration $\mathrm{SO}(6) \to \mathrm{SO}(\mathbb{CP}^3) \to \mathbb{CP}^3$ gives

$$1 = \pi_2 \,\mathrm{SO}(6) \to \pi_2(\mathrm{SO}(\mathbb{CP}^3)) \to \pi_2(\mathbb{CP}^3) \to \pi_1(\mathrm{SO}(6)) = \mathbb{Z}/(2\mathbb{Z}).$$

Since $\pi_2(\mathbb{CP}^3) \cong \mathbb{Z}$ it follows that $\pi_2(\mathrm{SO}(\mathbb{CP}^3)) \cong \mathbb{Z}$. On the other hand, in case (1), we have $\pi_2(\mathrm{SO}(S^6)) = \pi_2(\mathrm{SO}(7)) = 1$ and similarly $\pi_2(\mathrm{SO}(\mathbb{RP}^6)) = 1$. In case (2), we have that $\pi_2 N \cong \pi_2 L_N$ since S^1 is aspherical. Since L_N is diffeomorphic to either S^5 or $\mathbb{RP}^5$, we have $\pi_2 L_N = 1$. Again by the long exact sequence on homotopy groups for the fibration $\mathrm{SO}(N) \to N$, we see that $\pi_2 \,\mathrm{SO}(N) = 1$. Finally in case (3) we have $\pi_2 N = \pi_2 S^5 = 1$. As in case (2) we have that $\pi_2 \,\mathrm{SO}(N) = 1$.

Therefore in cases (1), (2), and (3), we cannot have $\mathrm{SO}(N) \cong \mathrm{SO}(\mathbb{CP}^3)$, so we conclude that M and N are both isometric to $\mathbb{CP}^3$ with a metric of constant holomorphic sectional curvature.

A metric of constant holomorphic sectional curvature on $\mathbb{CP}^3$ is determined by a bi-invariant metric on $\mathrm{SU}(4)$, which is then induced on the quotient

$$\mathrm{SU}(4)/\,\mathrm{S}(\mathrm{U}(1) \times \mathrm{U}(3)) \cong \mathbb{CP}^3.$$

Hence the metrics on M and N differ only by scaling, so M and N are isometric if and only if $\mathrm{vol}(M) = \mathrm{vol}(N)$. By Lemma 2.9 we indeed have $\mathrm{vol}(M) = \mathrm{vol}(N)$ so M and N are isometric. □

Above we assumed that $\overline{H}_M$ acts transitively on M. If instead $\overline{H}_N$ acts transitively on N, the same proof applies verbatim.

***Case* 3(b)** (H_M and H_N do not act transitively)**.** Theorem 3.4 and Proposition 3.7 imply that M and N are of one of the following types:

(1) diffeomorphic to S^n or $\mathbb{RP}^n$ equipped with a metric as in Theorem 3.4(1),

(2) $L \times S^1$ where each copy $L \times \{z\}$ is an isometrically embedded round sphere or projective space, or

(3) $(S^{n-1} \times \mathbb{R})/\Gamma$ where $\Gamma \cong D_\infty$ is generated by

$$(v, t) \mapsto (v, t+2) \quad \text{and} \quad (v, t) \mapsto (-v, -t).$$

Claim 4.8. *M and N belong to the same types in the above classification.*

Proof. Again we will show that the different types can be distinguished by the fundamental group of the frame bundle. Since $\mathrm{SO}(M) = \mathrm{SO}(N)$, it must then follow that M and N belong to the same type.

The fundamental group of $X = \mathrm{SO}(M)$ can be computed using the long exact sequence on homotopy groups for the fiber bundle $X \to M$ (or $X \to N$). Namely, we have

$$\pi_2(M) \to \pi_1(\mathrm{SO}(n)) \to \pi_1(X) \to \pi_1(M) \to 1,$$

and likewise for N. Since $\pi_2(M) = \pi_2(N) = 1$ for all of the above types, we have a short exact sequence

$$1 \to \mathbb{Z}/2\mathbb{Z} \to \pi_1(X) \to \pi_1(M) \to 1,$$

and likewise for N. We see that $\pi_1(X) \cong \mathbb{Z}/2\mathbb{Z}$ precisely when M is diffeomorphic to S^n, and $\pi_1(X)$ has order 4 precisely when M is diffeomorphic to $\mathbb{RP}^n$. If $\pi_1(X)$ is infinite then M is of type (2) or (3). If the maximal finite subgroup of $\pi_1(X)$ has order 2 then M is of type (2), and if the maximal finite subgroup of $\pi_1(X)$ has order 4 then M is of type (3). Therefore we can distinguish all the possible cases by considering $\pi_1(X)$, so M and N are of the same type. □

We will now show that in each of these cases, M and N are isometric.

Case A **(*M* and *N* are of type (1)).** Identify S^n with the solution set of $\sum_{i=0}^n x_i^2 = 1$ in $\mathbb{R}^{n+1}$. By Theorem 3.4(1), the metric on M (or its double cover if M is diffeomorphic to $\mathbb{RP}^n$) is of the form

$$ds_M^2 = f_M(x_0) \sum_{i=0}^n dx_i^2. \tag{4-4}$$

Similarly the metric on N (or its double cover) can be written as

$$ds_N^2 = f_N(x_0) \sum_{i=0}^n dx_i^2. \tag{4-5}$$

We will now show that $f_M(x) = f_N(x)$ for all x. We will just do this in case M and N are diffeomorphic to S^n, since the proof for $\mathbb{RP}^n$ is similar (note that it is not possible that one of M and N is diffeomorphic to S^n, and the other to $\mathbb{RP}^n$, since $\mathrm{SO}(S^n)$ and $\mathrm{SO}(\mathbb{RP}^n)$ are not diffeomorphic). Theorem 3.4(1) also describes the action of $\overline{H}_N$ on N. Namely, $\overline{H}_N$ leaves the coordinate x_0 invariant and acts transitively on each level set of x_0. This yields an identification

$$N/\overline{H}_N \cong [-1, 1].$$

The $\overline{H}_N$-orbits lying over the points in $(-1, 1)$ are copies of S^{n-1}, and the orbits lying over ± 1 are fixed points (corresponding to the north and south pole). Similarly we can identify $M/\overline{H}_M$ with $[-1, 1]$. Of course we can also write $M = X/H_M$, and this yields an identification

$$X/(H_M H_N) = M/\overline{H}_M.$$

Let $-1 < x < 1$ and choose a lift $y_M \in M$ of x. Equation (4-4) shows that $\operatorname{vol}(H_M y_M) = f_M(x) \operatorname{vol}(S^{n-1})$ where S^{n-1} is equipped with the metric $\sum_{i=1}^n dx_i^2$. Similarly if y_N is a lift of x to N we have $\operatorname{vol}(H_N y_N) = f_N(x) \operatorname{vol}(S^{n-1})$. Now choose a common lift $\tilde{y}$ of y_M and y_N to X, i.e., $\tilde{y}$ is an oriented orthonormal frame at the point $y_M \in M$ and at the point $y_N \in N$. Recall that the volume of a fiber of $X \to M$ is a fixed constant $\nu > 0$. Hence we have

$$\operatorname{vol}(H_M H_N \tilde{y}) = \nu \operatorname{vol}(\overline{H}_M y_M) = \nu f_M(x) \operatorname{vol}(S^{n-1}).$$

Since the volume of a fiber of $X \to N$ is also equal to ν, we also have

$$\operatorname{vol}(H_M H_N \tilde{y}) = \nu \operatorname{vol}(H_N y_N) = \nu f_N(x) \operatorname{vol}(S^{n-1}).$$

It follows that $f_M(x) = f_N(x)$. Hence M and N are isometric.

***Case B* (*M* and *N* are of type (2)).** In this case M is diffeomorphic to $L_M \times S^1$ where each copy $L_M \times \{z\}$ of L_M is isometric to a round sphere or projective space. The group $\overline{H}_N$ acts orthogonally on each fiber. However, note that the metric on M is not assumed to be a product metric, but in this case it has to be:

Lemma 4.9. *M is isometric to a product $L_M \times S^1$ where L_M is either a round sphere or projective space.*

Proof. Let $q : M \to S^1$ be the projection onto the second coordinate. Of course the fibers of q are just the submanifolds $L_M \times \{z\}$ for $z \in S^1$, and form a foliation $\mathcal{L}$ of M. Fix an orientation of L_M and define $\mathrm{SO}_{\mathcal{L}}(M)$ to be the space of pairs (x, e) where $x \in M$ and e is a positively oriented frame for the tangent space at x of the leaf of $\mathcal{L}$ through x. There is a natural bundle map $p : \mathrm{SO}_{\mathcal{L}}(M) \to M$ defined by $p(x, e) := x$. Further because $\overline{H}_M$ acts isometrically on M preserving the leaves of $\mathcal{L}$, it follows that $\overline{H}_M$ acts on $\mathrm{SO}_{\mathcal{L}}(M)$.

Of course, explicitly we have $\mathrm{SO}_{\mathcal{L}}(M) \cong \mathrm{SO}(L_M) \times S^1$, and the bundle map $p : \mathrm{SO}_{\mathcal{L}}(M) \to M$ is given by applying the natural bundle map $\mathrm{SO}(L_M) \to L_M$ to the first coordinate. Next we can explicitly describe the action of $\overline{H}_M$ on $\mathrm{SO}_{\mathcal{L}}(M)$. Namely, the action of $\overline{H}_M$ on L_M is just the standard action of $\mathrm{SO}(n)$ on S^{n-1} (or the standard action of $\mathrm{PSO}(n)$ on $\mathbb{RP}^{n-1}$). Using that $\mathrm{SO}(L_M) \cong \mathrm{SO}(n)$ or $\mathrm{PSO}(n)$, we see that $\overline{H}_M$ just acts by left-translations on $\mathrm{SO}(L_M)$. Finally, the action of $\overline{H}_M$ on $\mathrm{SO}_{\mathcal{L}}(M) \cong \mathrm{SO}(L_M) \times S^1$ is just by left-translations on each copy $\mathrm{SO}(L_M) \times \{z\}$ of $\mathrm{SO}(L_M)$.

The advantage of initially defining $\mathrm{SO}_{\mathcal{L}}(M)$ more abstractly (in terms of frames for the fibers of q), is that we can define an embedding

$$j : \mathrm{SO}_{\mathcal{L}}(M) \hookrightarrow \mathrm{SO}(M)$$

in the following way. A point $(x, e) \in \mathrm{SO}_{\mathcal{L}}(M)$ consists of an oriented orthonormal frame e of the copy of L_M through x. Hence e can be extended to a frame for M at x by adding to e the unique unit vector $v \in T_x M$ such that (e, v) is a positively oriented orthonormal frame for M. We define $j(x, e) := (x, e, v)$. Using that $\overline{H}_M$ preserves each copy $L_M \times \{z\}$ of L_M, it is easy to see that $j(\mathrm{SO}_{\mathcal{L}}(M))$ is an H_M-invariant submanifold of $\mathrm{SO}(M)$.

We equip $\mathrm{SO}_{\mathcal{L}}(M)$ with the Riemannian metric on $j(\mathrm{SO}_{\mathcal{L}}(M))$ induced from $\mathrm{SO}(M)$. Since the H_M-orbits in $\mathrm{SO}(M)$ are the fibers of the map $\pi_N : X \to N$, the H_M-orbits are totally geodesic in $\mathrm{SO}_{\mathcal{L}}(M)$ (see Proposition 2.2). We conclude that the foliation $\mathcal{F}$ of $\mathrm{SO}_{\mathcal{L}}(M)$ by H_M-orbits is a totally geodesic codimension 1 foliation of $\mathrm{SO}_{\mathcal{L}}(M)$. Of course this is just the foliation of $\mathrm{SO}_{\mathcal{L}}(M) = \mathrm{SO}(L_M) \times S^1$ by copies $\mathrm{SO}(L_M) \times \{z\}$ for $z \in S^1$. Consider the horizontal foliation $\mathcal{F}^{\perp}$ of $\mathrm{SO}_{\mathcal{L}}(M)$. Since $\mathcal{F}^{\perp}$ is 1-dimensional, it is integrable.

Johnson and Whitt [1980, Theorem 1.6] proved that if the horizontal distribution associated to a totally geodesic foliation is integrable, then the horizontal distribution is also totally geodesic. Further they showed that a manifold with two orthogonal totally geodesic foliations is locally a Riemannian product [Johnson and Whitt 1980, Proposition 1.3]. Therefore $\mathrm{SO}_{\mathcal{L}}(M)$ is locally a Riemannian product $F \times U$ where F (resp. U) is an open neighborhood in a leaf of $\mathcal{F}$ (resp. $\mathcal{F}^{\perp}$).

Now we show the metric on M has to locally be a product. Recall that the map $p : \mathrm{SO}_{\mathcal{L}}(M) \to M$ is defined by $p(x, e) = x$. We have $p = \pi_M \circ j$, where $j : \mathrm{SO}_{\mathcal{L}}(M) \hookrightarrow \mathrm{SO}(M)$ is the isometric embedding defined above, and $\pi_M : \mathrm{SO}(M) \to M$ is the natural projection. Since j is an isometric embedding and π_M is a Riemannian submersion, it follows that p is also a Riemannian submersion.

Now let $x \in M$ be any point and choose $\tilde{x} \in \mathrm{SO}_{\mathcal{L}}(M)$ with $p(\tilde{x}) = x$. Since the metric on $\mathrm{SO}_{\mathcal{L}}(M)$ is locally a product, we can choose a neighborhood $\widetilde{U} \times \widetilde{V}$ of $\tilde{x}$ on which the metric is a product.

Now let $w = (u, v) \in T_x M \cong T_x \mathcal{L}_x \oplus T_{q(x)} S^1$, where $u \in T_x \mathcal{L}_x$ and $v \in T_{q(x)} S^1$. Let $\tilde{u}$ (resp. $\tilde{v}$) be a lift of u (resp. v) to $T_{\tilde{x}} \mathrm{SO}_{\mathcal{L}}(M)$ that is horizontal with respect to p. Set $\tilde{w} := (\tilde{u}, \tilde{v}) \in T_{\tilde{x}} \mathrm{SO}_{\mathcal{L}} M$, so that $\tilde{w}$ is a horizontal lift of w. Then we have

$$\|w\|^2 = \|\tilde{w}\|^2 = \|\tilde{u}\|^2 + \|\tilde{v}\|^2 = \|u\|^2 + \|v\|^2,$$

where in the first and last step we used that p is a Riemannian submersion, and in the second step we used that the metric on $\mathrm{SO}_{\mathcal{L}}(M)$ is locally a Riemannian product. This shows that the metric on M is locally a product.

It remains to show that the metric on M is globally a product. Recall that M is diffeomorphic to $L_M \times S^1$, and that each copy $L_M \times \{z\}$ (for $z \in S^1$) is isometric to a round sphere or projective space, say with curvature $\kappa(z)$. Therefore to show that the metric is globally a product, it suffices to show that κ is constant. This is immediate because the metric on M is locally a product. □

Of course, the same proof applies to N, and shows that N is also isometric to a product $L_N \times S^1$. Further the metrics on the constant curvature spheres or projective spaces L_M and L_N only depend on their curvatures.

Claim 4.10. *L_M and L_N have the same curvature.*

Proof. Recall that we normalized the Sasaki–Mok–O'Neill metrics on $\mathrm{SO}(M) \cong \mathrm{SO}(N)$ so that the fibers of $\mathrm{SO}(M) \to M$ and $\mathrm{SO}(N) \to N$ have volume ν. These fibers are exactly H_M and H_N-orbits in $\mathrm{SO}(M)$, and by definition of the Sasaki–Mok–O'Neill metric, the metric restricted to an H_M or H_N-orbit is bi-invariant. On the other hand, if we restrict $\pi_M : X \to M$ to the H_M-orbit of a point $x \in X$, we obtain a bundle

$$\pi_M : H_M x \to \overline{H}_M \pi_M(x) \cong \mathcal{L}_{\pi_M(x)}. \tag{4-6}$$

Here $H_M x$ is diffeomorphic to $\mathrm{SO}(n)$ (if the leaves of $\mathcal{L}$ are spheres) or $\mathrm{PSO}(n)$ (if the leaves of $\mathcal{L}$ are projective spaces), and the fiber of the bundle in equation (4-6) is diffeomorphic to $\mathrm{SO}(n-1)$.

Since the metric on $H_M x$ (viewed as a submanifold of $\mathrm{SO}(M)$) is a bi-invariant metric, the above bundle is isometric to a standard bundle

$$\mathrm{SO}(n) \to S^{n-1}(r_M) \quad \text{if } \mathcal{L}_{\pi_M(x)} \cong S^{n-1},$$

or

$$\mathrm{PSO}(n) \to \mathbb{RP}^{n-1} \quad \text{if } \mathcal{L}_{\pi_M(x)} \cong \mathbb{RP}^{n-1},$$

where the base is a round sphere or projective space of some radius r_M. It follows that the volume of $H_M x$ only depends on r_M. Likewise the volume of $H_N x$ will only depend on the radius r_N of L_N. On the other hand we know that $\mathrm{vol}(H_M x) = \mathrm{vol}(H_N x) = \nu$, so we must have that $r_M = r_N$, as desired. □

At this point we know that there are $r > 0$, $\ell_M > 0$ and $\ell_N > 0$ such that M is isometric to $S^n(r) \times S^1(\ell_M)$ (or $\mathbb{RP}^n(r) \times S^1(\ell_M)$) and N is isometric to $S^n(r) \times S^1(\ell_N)$ (or $\mathbb{RP}^n(r) \times S^1(\ell_N)$). It only remains to show that $\ell_M = \ell_N$.

To see this, we need only recall that by normalization of the Sasaki–Mok–O'Neill metrics, we have $\mathrm{vol}(M) = \mathrm{vol}(N)$ (see Lemma 2.9).

Case C **(*M* and *N* are of type (3)).** The unique torsion-free, index-2 subgroups of $\pi_1(M)$ and $\pi_1(N)$ give double covers M' and N'. We claim that the frame bundles $\mathrm{SO}(M')$ and $\mathrm{SO}(N')$ are also isometric. The fiber bundle $\mathrm{SO}(n) \to X \to M$ gives

$$1 \to \mathbb{Z}/2\mathbb{Z} \to \pi_1(X) \to D_\infty \to 1.$$

Now $\pi_1(\mathrm{SO}(M'))$ and $\pi_1(\mathrm{SO}(N'))$ are both index-2 subgroups of $\pi_1(X)$. Since M' and N' are diffeomorphic to $S^{n-1} \times S^1$ we see that $\pi_1(\mathrm{SO}(M')) \cong (\mathbb{Z}/2\mathbb{Z}) \times \mathbb{Z}$ and likewise for $\pi_1(\mathrm{SO}(N'))$. Therefore $\pi_1(\mathrm{SO}(M'))$ and $\pi_1(\mathrm{SO}(N'))$ correspond to the same index 2 subgroup of $\pi_1(X)$. It follows that $\mathrm{SO}(M')$ and $\mathrm{SO}(N')$ are also isometric.

Since M' and N' are diffeomorphic to $S^{n-1} \times S^1$ and H_M acts on S^{n-1} orthogonally, the argument from case B applies and yields that M' and N' are isometric to the same product $S^{n-1} \times S^1$. Then M and N are obtained as the quotient of $S^{n-1} \times S^1$ by the map $(v, z) \mapsto (-v, z^{-1})$. Hence M and N are isometric. □

5. Proof for *M* with positive constant curvature

In the previous section we have proved Theorem A in all cases except when M has constant curvature $1/(2\sqrt{\lambda})$ or M is a surface. We will resolve the latter case in the next section. In this section we will prove:

Theorem 5.1. *Let M, N be closed oriented connected Riemannian n-manifolds and assume M has constant curvature $1/(2\sqrt{\lambda})$ for some $\lambda > 0$. Equip* $\mathrm{SO}(M)$ *and* $\mathrm{SO}(N)$ *with Sasaki–Mok–O'Neill metrics using the invariant inner product $\langle\cdot,\cdot\rangle_\lambda$ on $\mathfrak{o}(n)$. Assume $n \neq 2, 3, 4, 8$. Then M, N are isometric if and only if* $\mathrm{SO}(M)$ *and* $\mathrm{SO}(N)$ *are isometric.*

Proof. By simultaneously rescaling the metrics on M and N we can assume that the universal cover of M is a round sphere of radius 1. (Note that in the rescaling, we should also rescale the inner product on $\mathfrak{o}(n)$ that is used in the definition of the Sasaki–Mok–O'Neill metric.)

Since M has positive constant curvature, M is a Riemannian quotient of S^n by a finite group of isometries. Since the group of orientation-preserving isometries of S^n is $\mathrm{SO}(n+1)$, we can write $M = S^n/\pi_1(M)$ for some (finite) group $\pi_1(M) \subseteq \mathrm{SO}(n+1)$.

Further we can write $S^n = \mathrm{SO}(n) \setminus \mathrm{SO}(n+1)$ where the quotient is on the left by the standard copy $\mathrm{SO}(n) \subseteq \mathrm{SO}(n+1)$. The action of $\mathrm{SO}(n+1)$ on S^n by isometries is then just the action of $\mathrm{SO}(n+1)$ by right-translations on $\mathrm{SO}(n) \setminus \mathrm{SO}(n+1)$, so that we have

$$M \cong \mathrm{SO}(n) \setminus \mathrm{SO}(n+1)/\pi_1(M).$$

Passing to the frame bundle, we obtain $X \cong \mathrm{SO}(n+1)/\pi_1(M)$, where the cover $\mathrm{SO}(n+1)$ is equipped with a bi-invariant metric. Further N is a quotient of X by a group $H_M \cong \mathrm{SO}(n)$ acting effectively and isometrically on X.

Consider now the cover $\mathrm{SO}(n+1) \to X$. The (effective) action of H_M on X lifts to an effective action of a unique connected cover $\widehat{H}_M$ of H_M on $\mathrm{SO}(n+1)$. Note that $\mathrm{SO}(n)$ has only one nontrivial connected cover, namely its universal cover $\mathrm{Spin}(n)$. Therefore we have either $\widehat{H}_M \cong \mathrm{SO}(n)$ or $\widehat{H}_M \cong \mathrm{Spin}(n)$. We can actually describe the action of $\widehat{H}_M$ on $\mathrm{SO}(n+1)$ precisely:

Claim 5.2. *$\widehat{H}_M$ is isomorphic to $\mathrm{SO}(n)$ and acts on $\mathrm{SO}(n+1)$ by either left- or right-translations.*

Proof. Consider the full isometry group of $\mathrm{SO}(n+1)$ (with respect to a bi-invariant metric), which has been computed by d'Atri and Ziller [1979]. Namely, they show that the isometry group of a simple compact Lie group G equipped with a bi-invariant metric is

$$\mathrm{Isom}(G) \cong G \rtimes \mathrm{Aut}(G),$$

where the copy of G acts by left-translations on G. We apply this to the group $G = \mathrm{SO}(n+1)$. Since $\widehat{H}_M$ is connected, it follows that the image of $\widehat{H}_M \hookrightarrow \mathrm{Isom}(G)$ is contained in the connected component $\mathrm{Isom}(G)^0$ of $\mathrm{Isom}(G)$ containing the identity. We can explicitly compute $\mathrm{Isom}(G)^0$. Namely, since $\mathrm{Out}(G)$ is discrete, $\mathrm{Isom}(G)^0$ is isomorphic to

$$G \rtimes \mathrm{Inn}(G) \cong (G \times G)/Z(G),$$

where $Z(G)$ is the center of G, and $Z(G) \hookrightarrow G \times G$ is the diagonal embedding. The two copies of G act by left- and right-translations on G.

It will be convenient to work with the product $G \times G$, rather than $(G \times G)/Z(G)$. Note that the preimage of $\widehat{H}_M$ under the natural projection

$$G \times G \to (G \times G)/Z(G)$$

is a (possibly disconnected) cover of $\widehat{H}_M$. Let $\widetilde{H}_M$ denote the connected component containing the identity (so $\widetilde{H}_M$ is a connected cover of H_M, and hence isomorphic to either $\mathrm{SO}(n)$ or $\mathrm{Spin}(n)$).

We will first show that $\widetilde{H}_M$ has to be contained in a single factor of $G \times G$. To see this, let $p_i : \widetilde{H}_M \to G$ be the projection to the i-th factor (where $i = 1, 2$). Since $\widetilde{H}_M$ is a simple connected Lie group, p_i either has finite kernel or is trivial.

Further at least one of the projections has to be faithful: First, if one of the projections is trivial, then $\widetilde{H}_M$ is contained in a single factor, so that the other projection is faithful. Therefore to show one of the projections has to be faithful, it suffices to consider the case where neither projection is trivial, so that both projections have finite kernel. Let K_i, $i = 1,2$, be the kernels of the projections of $\widetilde{H}_M$ onto the i-th factor. Then K_i is a discrete normal subgroup of $\widetilde{H}_M$, and hence central. As discussed above, the only possibilities for $\widetilde{H}_M$ are $\mathrm{SO}(n)$ and $\mathrm{Spin}(n)$. The center $Z(\widetilde{H}_M)$ of $\widetilde{H}_M$ is then

$$Z(\widetilde{H}_M) \cong \begin{cases} 1 & \text{if } \widetilde{H}_M \cong \mathrm{SO}(n),\ n \text{ is odd}, \\ \mathbb{Z}/(2\mathbb{Z}) & \text{if } \widetilde{H}_M \cong \mathrm{SO}(n),\ n \text{ is even}, \\ \mathbb{Z}/(2\mathbb{Z}) & \text{if } \widetilde{H}_M \cong \mathrm{Spin}(n),\ n \text{ is odd}, \\ \mathbb{Z}/(4\mathbb{Z}) & \text{if } \widetilde{H}_M \cong \mathrm{Spin}(n),\ n \text{ is even}. \end{cases}$$

Further, since no nontrivial element of $\widetilde{H}_M$ projects trivially to both factors (for such an element would be trivial in $G \times G$), we must have $K_1 \cap K_2 = 1$. On the other hand, none of the possibilities for $Z(\widetilde{H}_M)$ have two nontrivial subgroups that intersect trivially, so we conclude that K_1 or K_2 is trivial. Without loss of generality, we assume that $K_1 = 1$.

Therefore to prove the claim that $\widetilde{H}_M$ is contained in a single factor, we must show that $p_2(\widetilde{H}_M)$ is trivial. Suppose it is not. Then p_2 has finite kernel, so $p_2(\widetilde{H}_M)$ is a subgroup of $G = \mathrm{SO}(n+1)$ of dimension $\dim \widetilde{H}_M = \frac{1}{2}n(n-1)$. Fortunately, there are very few possibilities by the following fact:

Lemma 5.3 [Kobayashi 1972, Lemma 1 in II.3]. *Let H be a closed connected subgroup of* $\mathrm{SO}(n+1)$ *of dimension $\frac{1}{2}n(n-1)$ with $n+1 \neq 4$. Then either*

(1) *$H \cong \mathrm{SO}(n)$ and H fixes a line in $\mathbb{R}^{n+1}$, or*

(2) *$H \cong \mathrm{Spin}(7)$ (and hence $n+1 = 8$), and H is embedded in* $\mathrm{SO}(8)$ *via a spin representation.*

Here we say that a representation of $\mathrm{Spin}(n)$ is *spin* if it does not factor through the covering map $\mathrm{Spin}(n) \to \mathrm{SO}(n)$. To obtain the desired contradiction, we will now consider various cases depending on which of the above possibilities describe $p_1(\widetilde{H}_M)$ and $p_2(\widetilde{H}_M)$. For ease of notation we set $\widetilde{H}_i := p_i(\widetilde{H}_M)$ for $i = 1, 2$. Before considering each case separately, let us first make the following basic observation that underlies the argument in each case:

Recall that H_M acts freely on X. It follows that $\widetilde{H}_M/(Z(G) \cap \widetilde{H}_M)$ acts freely on G: namely, if $h \in \widetilde{H}_M$ fixes $x \in G$, then the image of h under $\widetilde{H}_M \to H_M$ fixes the image of x under the covering map $G \to X$. Since H_M acts freely on X, we see that h belongs to the kernel of $\widetilde{H}_M \to X$. Since the map $G \to X$ is equivariant with respect to the morphism $\widetilde{H}_M \to H_M$, it follows that for any $g \in G$, the points g and $h \cdot g$ of G have the same image in X. This exactly means that the action of h on G is a deck transformation of the covering $G \to X$. Since h fixes the point $x \in G$ and any deck transformation that fixes a point is trivial, h acts trivially on G. Since the kernel of the action of $G \times G$ on G is the center $Z(G)$, it follows that h is central, as desired.

Therefore if $h=(h_1,h_2)\in \widetilde{H}_M \subseteq G\times G$ fixes a point in G, then $h_1=h_2$ and h_i are central in G. Since $(h_1,h_2)\cdot g = h_1 g h_2^{-1}$, the stabilizer of $g\in G$ consists exactly of the elements of the form $(h_1, g h_1 g^{-1})$ where $h_1\in G$. Our strategy for obtaining a contradiction in each of the cases below is to find an element $h=(h_1, g h_1 g^{-1})\in \widetilde{H}_M$ but with $h_1\notin Z(G)$.

Case 1 ($\widetilde{\boldsymbol{H}}_{\mathbf{1}}$ and $\widetilde{\boldsymbol{H}}_{\mathbf{2}}$ are both of type (1) of Lemma 5.3)**.** By assumption, there are nonzero vectors v_1 and $v_2\in\mathbb{R}^{n+1}$ such that $\widetilde{H}_i\cong \mathrm{SO}(n)$ fixes v_i. The representation of $\widetilde{H}_i$ on $(\mathbb{R}v_i)^\perp$ is the standard representation of $\mathrm{SO}(n)$. Therefore there is some an intertwiner $T:(\mathbb{R}v_1)^\perp\to(\mathbb{R}v_2)^\perp$ of these representations. Recall that an irreducible representation leaves invariant at most one inner product up to positive scalars (for if Q_1 and Q_2 are linearly independent invariant bilinear forms, then a suitable linear combination $Q=\alpha Q_1+\beta Q_2$ is invariant and degenerate as a bilinear form; the kernel of Q is then a proper invariant subspace). It follows that after possibly replacing T by λT for some $\lambda>0$, the intertwiner T is orthogonal.

We can extend T to an intertwiner $\mathbb{R}^{n+1}\to\mathbb{R}^{n+1}$ between $\widetilde{H}_1$ and $\widetilde{H}_2$ by setting $Tv_1:=\mu v_2$ for some $\mu\neq 0$. We will denote the extension by T as well. By choosing μ suitably, we can arrange that T is orthogonal, and after possibly changing the sign of μ, we can also arrange that $\det T=1$.

The map T then belongs to $\mathrm{SO}(n+1)$, so that we have that

$$\widetilde{H}_M=\{(h, ThT^{-1}) \mid h\in \widetilde{H}_1\}.$$

As observed above, it follows that $\widetilde{H}_M$ does not act freely on X.

Case 2 (At least one of $\widetilde{\boldsymbol{H}}_{\mathbf{1}}$ and $\widetilde{\boldsymbol{H}}_{\mathbf{2}}$ is of type (2) of Lemma 5.3)**.** Note that it is not possible that $\widetilde{H}_1$ is of type (1) and $\widetilde{H}_2$ is of type (2). Namely, in this case we would have that $\widetilde{H}_M\cong\mathrm{SO}(n)$ (because $\widetilde{H}_M\cong\widetilde{H}_1$), but the map $\widetilde{H}_M\to\widetilde{H}_2$ would be a covering $\mathrm{SO}(n)\to\mathrm{Spin}(n)$, which is impossible (since the latter is simply connected but the former is not).

So we must have that $\widetilde{H}_1$ is of type (2). In particular we have $n=7$. Unfortunately, we cannot immediately apply the same argument as in case 1, because Spin(7) has multiple faithful representations of dimension 8. This difficulty is resolved by passing to a suitable subgroup of Spin(7): Namely given a spin representation of Spin(7), the stabilizer of any nonzero $v\in\mathbb{R}^8$ is isomorphic to the exceptional simple Lie group G_2.

For the rest of the proof we fix some nonzero $v\in\mathbb{R}^8$ and let L be the stabilizer in $\widetilde{H}_1$ of v. We have two representations of L on $\mathbb{R}^8$: On the one hand we have $L\subseteq\widetilde{H}_1$. On the other hand we can consider $p_2(p_1^{-1}(L))\subseteq\widetilde{H}_2$. We analyze these representations in turn and will show they are equivalent. Before doing so, it will be helpful to recall some classical facts about the representation theory of G_2 (see [Adams 1996, Chapter 5]) for (a) – (d) and [Helgason 1978, Table X.6.IV] for (e)):

(a) G_2 is obtained as the subgroup of matrices of SO(8) that preserve the product of the octonions $\mathbb{O}$,

(b) G_2 has no nontrivial representations of dimension less than 7,

(c) G_2 has a single representation of dimension 7 (the action on the purely imaginary octonions) that by fact (b) is necessarily irreducible,

(d) G_2 has no irreducible representation of dimension 8, and

(e) G_2 has trivial center.

We will write $\mathbb{1}$ for the trivial representation and $\mathrm{Im}(\mathbb{O})$ for the unique 7-dimensional faithful representation.

Let us now consider the first representation, obtained by considering L as a subgroup of $\widetilde{H}_1$. This representation is automatically faithful and has $\mathbb{R}v$ as a trivial summand. The summand $(\mathbb{R}v)^{\perp}$ is therefore a faithful 7-dimensional representation and by (c) equivalent to $\mathrm{Im}(\mathbb{O})$. Therefore the first representation is equivalent to $\mathbb{1} \oplus \mathrm{Im}\, \mathbb{O}$.

We turn to the second representation, obtained by the map $p_2 \circ p_1^{-1} : L \to \widetilde{H}_2$. This is a map with finite kernel (because p_2 has finite kernel and p_1 is an isomorphism), so that the kernel is contained in the center. Since G_2 has no center (see (e)), it follows that this representation is also faithful. Since G_2 has no irreducible representation of dimension 8 (see (d)), we must have that the second representation also decomposes as $\mathbb{1} \oplus \mathrm{Im}\, \mathbb{O}$. Therefore there is an intertwiner $T : \mathbb{R}^8 \to \mathbb{R}^8$ between these representations. The rest of the argument proceeds exactly as in case 1.

This concludes the proof that $\widetilde{H}_M$ is contained in one of the factors of $G \times G$. To complete the proof of the claim, we must show that $\widetilde{H}_M \cong \mathrm{SO}(n)$. By the dichotomy from Lemma 5.3, the only other possibility is that $n = 7$ and $\widetilde{H}_M$ is given by a spin representation of Spin(7).

In the latter case, we can see that N has constant positive curvature: Namely, since M has constant curvature, the metric on $X \cong \mathrm{SO}(8)/\pi_1(M)$ lifts to a bi-invariant metric on SO(8) and hence to a bi-invariant metric on Spin(8). On the other hand $N = X/H_M$ is finitely covered by $\mathrm{SO}(8)/\mathrm{Spin}(7)$, and hence also by $\mathrm{Spin}(8)/\mathrm{Spin}(7)$. It is well known that a bi-invariant metric on Spin(8) induces a metric of constant positive curvature on $S^7 \cong \mathrm{Spin}(8)/\mathrm{Spin}(7)$.

Since N has constant positive curvature, we can write $N = \mathrm{SO}(7) \backslash \mathrm{SO}(8)/\pi_1(N)$ for some finite subgroup $\pi_1(N) \subseteq \mathrm{SO}(8)$ acting by right-translations. The frame bundle of N is then $X = \mathrm{SO}(8)/\pi_1(N)$ with $H_M \cong \mathrm{SO}(7)$ acting by left-translations. This contradicts that $\widetilde{H}_M$ was given by a spin representation into SO(8), and hence finishes the proof of the claim. □

Since H_M acts by left- or right-translations on $\mathrm{SO}(n+1)$, we will identify H_M with a subgroup of $\mathrm{SO}(n+1)$. Then we can conjugate H_M to a standard copy of

$\mathrm{SO}(n)$ by an element of $\mathrm{SO}(n+1)$. Therefore without loss of generality we have $N \cong \mathrm{SO}(n) \setminus \mathrm{SO}(n+1)/\pi_1(N)$, and we have an isometry

$$f : \mathrm{SO}(n+1)/\pi_1(M) \cong \mathrm{SO}(M) \to \mathrm{SO}(N) \cong \mathrm{SO}(n+1)/\pi_1(N).$$

By composing with a left-translation of $\mathrm{SO}(n+1)$, we can also assume that $f(e\pi_1(M)) = e\pi_1(N)$. It remains to show there is an isometry

$$M \cong \mathrm{SO}(n) \setminus \mathrm{SO}(n+1)/\pi_1(M) \to \mathrm{SO}(n) \setminus \mathrm{SO}(n+1)/\pi_1(N) \cong N.$$

Claim 5.4. *f lifts to an isometry* $\mathrm{SO}(n+1) \to \mathrm{SO}(n+1)$.

Proof. The universal cover of $\mathrm{SO}(M)$ and $\mathrm{SO}(N)$ is $\mathrm{Spin}(n+1)$, so f lifts to a map

$$\tilde{f} : \mathrm{Spin}(n+1) \to \mathrm{Spin}(n+1).$$

We can choose the lift $\tilde{f}$ such that $\tilde{f}(e) = e$, where e is the identity element of $\mathrm{SO}(n+1)$. Note that since f is an isometry, $\tilde{f}$ is an isometry as well (with respect to a bi-invariant metric on $\mathrm{Spin}(n+1)$). As previously mentioned, d'Atri and Ziller [1979] computed the group of isometries of a connected compact semisimple Lie group G. Indeed, $\mathrm{Isom}(G) = G \rtimes \mathrm{Aut}(G)$, where the copy of G acts by left-translations. It immediately follows that any isometry fixing the identity element e is an automorphism. Therefore $\tilde{f}$ is an automorphism of $\mathrm{Spin}(n+1)$.

Recall that $\mathrm{Spin}(n+1)$ has a unique central element z of order 2, and we have $\mathrm{SO}(n+1) = \mathrm{Spin}(n+1)/\langle z\rangle$. Since z is the unique central element of order 2, we must have that $\tilde{f}(z) = z$. It follows that $\tilde{f}$ descends to an automorphism of $\mathrm{SO}(n+1)$, as desired. □

Let

$$\hat{f} : \mathrm{SO}(n+1) \to \mathrm{SO}(n+1)$$

denote a lift of f. As above, by choosing an appropriate lift, we can assume that $\hat{f}(e) = e$, and hence that $\hat{f}$ is an automorphism of $\mathrm{SO}(n+1)$ (here we again used the computation of d'Atri and Ziller of the isometry group of $\mathrm{SO}(n+1)$). Because $\hat{f}$ is a lift of f, we know that $\hat{f}$ restricts to an isomorphism $\pi_1 M \to \pi_1 N$.

Since $\hat{f}$ is an automorphism of $\mathrm{SO}(n+1)$, there is some $g \in \mathrm{SO}(n+1)$ such that $\hat{f}(\mathrm{SO}(n)) = g\,\mathrm{SO}(n)g^{-1}$. Here, as well as well as below, we identify $\mathrm{SO}(n)$ with a fixed standard copy in $\mathrm{SO}(n+1)$. Define a map

$$\hat{\varphi} : \mathrm{SO}(n+1) \to \mathrm{SO}(n+1)$$

by $\hat{\varphi}(x) := g^{-1}\hat{f}(x)$.

Claim 5.5. (1) *$\hat{\varphi}$ is an isometry,*

(2) *For any $x \in \mathrm{SO}(n+1)$, we have $\hat{\varphi}(\mathrm{SO}(n)x) = \mathrm{SO}(n)\hat{\varphi}(x)$.*

(3) *For any $x \in \mathrm{SO}(n+1)$, we have $\hat{\varphi}(x\pi_1(M)) = \hat{\varphi}(x)\pi_1(N)$.*

Proof. (1) Since left-translation by g is an isometry of $\mathrm{SO}(n+1)$, and $\hat{f}$ is also an isometry of $\mathrm{SO}(n+1)$, it follows that the map $\hat{\varphi}$ is an isometry.

(2) Let $x \in \mathrm{SO}(n+1)$. We have $\hat{\varphi}(\mathrm{SO}(n)x) = g^{-1}\hat{f}(\mathrm{SO}(n)x)$. Since $\hat{f}$ is an automorphism of $\mathrm{SO}(n+1)$, we then have

$$\hat{\varphi}(\mathrm{SO}(n)x) = g^{-1}\hat{f}(\mathrm{SO}(n))\hat{f}(x).$$

Using that $\hat{f}(\mathrm{SO}(n)) = g\,\mathrm{SO}(n)g^{-1}$, we see that

$$\hat{\varphi}(\mathrm{SO}(n)x) = \mathrm{SO}(n)g^{-1}\hat{f}(x) = \mathrm{SO}(n)\hat{\varphi}(x).$$

(3) Let $x \in \mathrm{SO}(n+1)$. This is similar to the proof of (2), but now using that $\hat{f}(\pi_1(M)) = \pi_1(N)$. We have

$$\hat{\varphi}(x\pi_1(M)) = g^{-1}\hat{f}(x\pi_1(M)) = g^{-1}\hat{f}(x)\hat{f}(\pi_1(M)) = \hat{\varphi}(x)\pi_1(N). \qquad \square$$

From properties (2) and (3) of Claim 5.5, it is immediate that $\hat{\varphi}$ descends to a map

$$\overline{\varphi} : \mathrm{SO}(n) \setminus \mathrm{SO}(n+1)/\pi_1(M) \to \mathrm{SO}(n) \setminus \mathrm{SO}(n+1)/\pi_1(N).$$

Claim 5.6. *$\overline{\varphi}$ is an isometry $M \to N$.*

Proof. Recall that at the end of case 1 of the proof of Theorem 4.1, we showed that an isometry $X \to X$ that maps the fibers of $\pi_M : X \to M$ to the fibers of $\pi_N : X \to N$, descends to an isometry $M \to N$.

In the current setting, the map $\hat{\varphi} : \mathrm{SO}(n+1) \to \mathrm{SO}(n+1)$ descends to a map

$$\varphi : X \cong \mathrm{SO}(n+1)/\pi_1(M) \to \mathrm{SO}(n+1)/\pi_1(N) \cong X$$

by property (3) of Claim 5.5. Since $\hat{\varphi}$ is an isometry and the maps

$$\mathrm{SO}(n+1) \to \mathrm{SO}(n+1)/\pi_1(M) \quad \text{and} \quad \mathrm{SO}(n+1) \to \mathrm{SO}(n+1)/\pi_1(N)$$

are Riemannian coverings, it follows that φ is an isometry.

Therefore to prove the claim, it suffices to show that φ maps fibers of $X \to M$ to fibers of $X \to N$. If we make the identifications $X \cong \mathrm{SO}(n+1)/\pi_1(M)$ and $M \cong \mathrm{SO}(n) \setminus \mathrm{SO}(n+1)/\pi_1(M)$, the map $X \to M$ is just the natural orbit map

$$\mathrm{SO}(n+1)/\pi_1(M) \to \mathrm{SO}(n) \setminus \mathrm{SO}(n+1)/\pi_1(M).$$

Therefore the fibers of $X \to M$ are exactly the $\mathrm{SO}(n)$-orbits in $\mathrm{SO}(n+1)/\pi_1(M)$ (under the action by left-translation). Likewise, the fibers of $X \to N$ are the $\mathrm{SO}(n)$-orbits in $\mathrm{SO}(n+1)/\pi_1(N)$ under the action by left-translation. It follows immediately from property (3) of Claim 5.5 that $\hat{\varphi}$ maps $\mathrm{SO}(n)$-orbits to $\mathrm{SO}(n)$-orbits, and hence so does φ. $\square$

We have shown that the map $\overline{\varphi}$ is an isometry $M \to N$, so that M and N are isometric, which finishes the proof of Theorem 5.1. $\square$

6. Proof of the main theorem for surfaces

In this section we prove Theorem A for surfaces. We cannot use the Takagi–Yawata theorem (Theorem 2.6) that computes $i(X)$ in this situation, but instead we use the classification of surfaces and Lie groups in low dimensions.

Let M and N be closed oriented surfaces with $\mathrm{SO}(M) \cong \mathrm{SO}(N)$. Therefore M and N are each diffeomorphic to one of Σ_g with $g \geq 2$, S^2, or T^2. We know that

- $\mathrm{SO}(S^2)$ is diffeomorphic to $\mathrm{SO}(3)$,
- $\mathrm{SO}(T^2)$ is diffeomorphic to T^3, and
- $\mathrm{SO}(\Sigma_g)$ is diffeomorphic to $T^1\Sigma_g = \mathrm{PSL}_2\,\mathbb{R}/\Gamma$ for a cocompact torsion-free lattice $\Gamma \subseteq \mathrm{PSL}_2\,\mathbb{R}$.

In particular the diffeomorphism type of the frame bundle of a surface determines the diffeomorphism type of the surface. It follows that M and N are diffeomorphic.

Consider the Lie algebra of Killing fields $i(X)$ of X. Then $i(X)$ contains the (1-dimensional) subalgebras i_V^M and i_V^N. If $i_V^M = i_V^N$, then we proceed as in case (1) in Section 4, and we find that M and N are isometric. Therefore we will assume that $i_V^M \neq i_V^N$. In particular we must have $\dim i(X) \geq 2$.

As before, let H_M (resp. H_N) be the subgroup of $\mathrm{Isom}(X)$ obtained by exponentiating the Lie algebra i_V^N (resp. i_V^M). Then H_M and H_N are closed subgroups of $\mathrm{Isom}(X)$ isomorphic to S^1.

We will now consider each of the possibilities of the diffeomorphism types of M and N, and prove that M and N have to be isometric.

***Case* 1 (M and N are diffeomorphic to Σ_g, $g \geq 2$).** Then $X = T^1\Sigma_g$ is a closed aspherical manifold. Conner and Raymond [1970] proved that if a compact connected Lie group G acts effectively on a closed aspherical manifold L, then G is a torus and $\dim G \leq \mathrm{rk}_{\mathbb{Z}}\, Z(\pi_1 L)$, where $Z(\pi_1 L)$ is the center of $\pi_1(L)$. In particular we find that $\dim i(X) \leq \mathrm{rk}_{\mathbb{Z}}\, Z(\pi_1 T^1\Sigma_g) = 1$. This contradicts our assumption that $\dim i(X) \geq 2$.

***Case* 2 (M and N are diffeomorphic to S^2).** Let G be the connected component of $\mathrm{Isom}(X)$ containing the identity. Then G is a compact connected Lie group acting effectively and isometrically on $X = \mathrm{SO}(3)$, and G contains H_M and H_N.

If $\dim G = 2$, then G is a 2-torus. In particular H_M and H_N centralize each other. Therefore H_N acts on $X/H_M = N$ and similarly H_M acts on M. The kernel of either of these actions is $H_M \cap H_N$, which is a finite subgroup of both H_M and H_N.

Since an S^1-action on S^2 has at least one fixed point (because $\chi(S^2) \neq 0$), we see that $N/H_N \cong [-1, 1] \cong M/H_M$. It is then straightforward to see that the metric on M (resp. N) is of the form

$$ds_M^2 = f_M(x_0)(dx_0^2 + dx_1^2 + dx_2^2)$$

(resp. $ds_N^2 = f_N(x_0)(dx_0^2 + dx_1^2 + dx_2^2)$) as in Theorem 3.4(1). We can apply the reasoning from case A of the proof of case 3(b) in Section 4 to show M and N are isometric.

Therefore we will assume $\dim G \geq 3$. In addition we know that $\dim G \leq 6$ by Theorem 3.1. Finally, we must have $\mathrm{rk}(G) \leq 2$: Namely let T be a maximal torus in G containing H_N. Since T centralizes H_N, the group T/H_N acts effectively on M. However, a torus of dimension ≥ 2 does not act effectively on S^2. (To see this, note that any 1-parameter subgroup H has a fixed point on S^2 because the Killing field generated by H has a zero on S^2. We can take H to be dense, so that the entire torus fixes a point p. The isotropy action on T_pM is a faithful 2-dimensional representation of the torus, which is impossible unless the torus is 1-dimensional.)

Therefore the only possibilities for the Lie algebra $\mathfrak{g}$ of G are

(a) $\mathfrak{g} \cong \mathfrak{o}(3)$,

(b) $\mathfrak{g} \cong \mathbb{R} \oplus \mathfrak{o}(3)$, and

(c) $\mathfrak{g} \cong \mathfrak{o}(3) \oplus \mathfrak{o}(3)$.

We will now consider each of these cases separately.

***Case* 2(a) ($\mathfrak{g} \cong \mathfrak{o}(3)$).** Since G has rank 1, H_M and H_N are both maximal tori of G. Since all maximal tori are conjugate, there is some element $g \in G$ so that $gH_Ng^{-1} = H_M$. Then g induces an obvious isometry $M \to N$.

***Case* 2(b) ($\mathfrak{g} \cong \mathbb{R} \oplus \mathfrak{o}(3)$).** We can conjugate H_N by an element $g \in G$ so that gH_Ng^{-1} and H_M centralize each other. Then either $gH_Ng^{-1} = H_M$, in which case g induces an isometry $M \to N$, or gH_Ng^{-1} and H_M generate a 2-torus. In the latter case the argument above in case $\dim G = 2$ shows that the metrics on M and N are of the form

$$ds^2 = f(x_0)(dx_0^2 + dx_1^2 + dx_2^2),$$

for some function f on $[-1, 1]$. Then the argument of case A of case 3(b) in Section 4 shows that M and N are isometric.

***Case* 2(c) ($\mathfrak{g} \cong \mathfrak{o}(3) \oplus \mathfrak{o}(3)$).** In this case $\dim \mathrm{Isom}(X) = 6$ is maximal. By Theorem 3.1 the metric on X has positive constant curvature. Therefore the metrics on M and N have positive constant curvature. Further by Lemma 2.9 we have $\mathrm{vol}(M) = \mathrm{vol}(N)$. It follows that M and N are isometric.

Case 3 (***M*** and ***N*** are diffeomorphic to $\boldsymbol{T^2}$). In this case X is diffeomorphic to T^3. Again by the theorem of Conner and Raymond [1970] on actions of compact Lie groups on aspherical manifolds, we know that a connected compact Lie group acting effectively on a torus is a torus. Therefore H_N and H_M centralize each other, so H_M and H_N generate a 2-torus. Further H_M acts on $M = X/H_N$ with finite kernel $H_M \cap H_N$. Again by [Conner and Raymond 1970], the action of $H_M/(H_M \cap H_N)$ on M is free, so that the map

$$M \to M/H_M \cong S^1$$

is a fiber bundle (with S^1 fibers). The argument of case B in case 3(b) of the proof of Theorem A constructs a (unit length) Killing field X_M on M that is orthogonal to the fibers of $M \to M/H_M$. It follows that M is a 2-torus equipped with a translation invariant metric. Any such metric is automatically flat: Namely, because the torus is abelian, the metric is automatically bi-invariant. Then we use the following general fact: on a Lie group H with a bi-invariant metric, the Lie structure and sectional curvature are tied by the identity (see, e.g., [Petersen 2006, Proposition 3.4.12])

$$K(X, Y) = \frac{1}{4}\|[X, Y]\|^2,$$

where X, Y are orthonormal vectors in $\mathfrak{h}$ (which is identified with T_eH in the usual way), and the bracket is the Lie bracket. Since T^2 is abelian, it follows that the sectional curvatures with respect to any invariant metric vanish.

We conclude that M is flat. By carrying out the same construction for N, we obtain a Killing field X_N on N that is orthogonal to the H_N-orbits, and we conclude that N is flat.

To show that M and N are isometric, recall that the isometry type of a flat 2-torus is specified by the length of two orthogonal curves that generate its fundamental group. For M we can consider the curves given by an H_M-orbit on M and an integral curve of X_M. Similarly for N we can consider an H_N-orbit on N and an integral curve of X_N.

For $x \in M$ and $\tilde{x} \in X$ lying over x, we have a covering

$$H_M\tilde{x} \to H_M x$$

of degree $|H_N \cap H_M|$. Recall that the H_M-orbits in X have a fixed volume ν, since we normalized the Sasaki–Mok–O'Neill metric on X in this way. Therefore

$$\ell(H_M x) = \frac{1}{|H_N \cap H_M|}\ell(H_M\tilde{x}) = \frac{\nu}{|H_N \cap H_M|}.$$

Combining this with a similar computation for the length of an H_N-orbit on N gives $\ell(H_M x) = \ell(H_N y)$ for every $x \in M$ and $y \in N$. Therefore we see that the length of an integral curve of X_M (resp. X_N) is $\mathrm{vol}(M)/(\ell(H_M \cdot x))$ for $x \in M$

(resp. $\mathrm{vol}(N)/(\ell(H_N \cdot y))$ for $y \in N$). Since $\mathrm{vol}(M) = \mathrm{vol}(N)$ by Lemma 2.9, it follows that M and N are isometric.

Acknowledgements

I am very grateful to Benson Farb, my thesis advisor, for posing Question 1.1 to me, his extensive comments on an earlier version of this paper, and his invaluable advice and enthusiasm during the completion of this work. I am indebted to an anonymous referee for many helpful suggestions. I would like to thank the University of Chicago for support.

References

[Adams 1996] J. F. Adams, *Lectures on exceptional Lie groups*, edited by Z. Mahmud and M. Mimura, University of Chicago Press, Chicago, IL, 1996. MR 1428422 Zbl 0866.22008

[Conner and Raymond 1970] P. E. Conner and F. Raymond, "Actions of compact Lie groups on aspherical manifolds", pp. 227–264 in *Topology of manifolds* (Athens, GA, 1969), Markham, Chicago, IL, 1970. MR 0271958 Zbl 0312.57025

[D'Atri and Ziller 1979] J. E. D'Atri and W. Ziller, *Naturally reductive metrics and Einstein metrics on compact Lie groups*, Mem. Amer. Math. Soc. **18**:215, 1979. MR 519928 Zbl 0404.53044

[Hawley 1953] N. S. Hawley, "Constant holomorphic curvature", *Canadian J. Math.* **5** (1953), 53–56. MR 0052869 Zbl 0050.16203

[Helgason 1978] S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*, Pure and Applied Mathematics **80**, Academic Press, New York, 1978. MR 514561 Zbl 0451.53038

[Igusa 1954] J.-i. Igusa, "On the structure of a certain class of Kaehler varieties", *Amer. J. Math.* **76**:3 (1954), 669–678. MR 0063740 Zbl 0058.37901

[Johnson and Whitt 1980] D. L. Johnson and L. B. Whitt, "Totally geodesic foliations", *J. Differential Geom.* **15**:2 (1980), 225–235. MR 614368 Zbl 0444.57017

[Kobayashi 1972] S. Kobayashi, *Transformation groups in differential geometry*, Ergebnisse der Mathematik (2) **70**, Springer, 1972. MR 0355886 Zbl 0246.53031

[Kobayashi and Nagano 1972] S. Kobayashi and T. Nagano, "Riemannian manifolds with abundant isometries", pp. 195–219 in *Differential geometry* (*in honor of Kentaro Yano*), Kinokuniya, Tokyo, 1972. MR 0358623 Zbl 0251.53036

[Kobayashi and Nomizu 1963] S. Kobayashi and K. Nomizu, *Foundations of differential geometry, I*, Interscience Tracts in Pure and Applied Mathematics **15**, Wiley, New York, 1963. MR 0152974 Zbl 0119.37502

[Mok 1978] K. P. Mok, "On the differential geometry of frame bundles of Riemannian manifolds", *J. Reine Angew. Math.* **302** (1978), 16–31. MR 511689 Zbl 0378.53016

[Mok 1979] K. P. Mok, "Complete lifts of tensor fields and connections to the frame bundle", *Proc. London Math. Soc.* (3) **38**:1 (1979), 72–88. MR 520973 Zbl 0395.53017

[Myers and Steenrod 1939] S. B. Myers and N. E. Steenrod, "The group of isometries of a Riemannian manifold", *Ann. of Math.* (2) **40**:2 (1939), 400–416. MR 1503467 Zbl 0021.06303

[O'Neill 1966] B. O'Neill, "The fundamental equations of a submersion", *Michigan Math. J.* **13**:4 (1966), 459–469. MR 0200865 Zbl 0145.18602

[Petersen 2006] P. Petersen, *Riemannian geometry*, 2nd ed., Graduate Texts in Mathematics **171**, Springer, 2006. MR 2243772 Zbl 1220.53002

[Sasaki 1958] S. Sasaki, "On the differential geometry of tangent bundles of Riemannian manifolds", *Tôhoku Math. J.* (2) **10**:3 (1958), 338–354. MR 0112152 Zbl 0086.15003

[Sasaki 1962] S. Sasaki, "On the differential geometry of tangent bundles of Riemannian manifolds, II", *Tôhoku Math. J.* (2) **14**:2 (1962), 146–155. MR 0145456 Zbl 0109.40505

[Takagi and Yawata 1991] H. Takagi and M. Yawata, "Infinitesimal isometries of frame bundles with natural Riemannian metric", *Tohoku Math. J.* (2) **43**:1 (1991), 103–115. MR 1088717 Zbl 0739.53017

[Takagi and Yawata 1994] H. Takagi and M. Yawata, "Infinitesimal isometries of frame bundles with a natural Riemannian metric, II", *Tohoku Math. J.* (2) **46**:3 (1994), 341–355. MR 1289183 Zbl 0835.53047

[Wang 1947] H.-C. Wang, "On Finsler spaces with completely integrable equations of Killing", *J. London Math. Soc.* **22** (1947), 5–9. MR 0022431 Zbl 0029.16604

[Ziller 2001] W. Ziller, "Fatness revisited", lecture notes, University of Pennsylvania, 2001, available at https://www.math.upenn.edu/~wziller/papers/Fat-09.pdf.

Received July 30, 2014. Revised August 13, 2016.

WOUTER VAN LIMBEEK
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MICHIGAN
2074 EAST HALL
530 CHURCH STREET
ANN ARBOR, MI 48109
UNITED STATES
wouterv@umich.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 285, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.285.427

BUNDLES OF SPECTRA AND ALGEBRAIC K-THEORY

JOHN A. LIND

A parametrized spectrum E is a family of spectra E_x continuously parametrized by the points $x \in X$ of a topological space. We take the point of view that a parametrized spectrum is a bundle-theoretic geometric object. When R is a ring spectrum, we consider parametrized R-module spectra and show that they give cocycles for the cohomology theory determined by the algebraic K-theory $K(R)$ of R in a manner analogous to the description of topological K-theory $K^0(X)$ as the Grothendieck group of vector bundles over X. We prove a classification theorem for parametrized spectra, showing that parametrized spectra over X whose fibers are equivalent to a fixed R-module M are classified by homotopy classes of maps from X to the classifying space $B\mathrm{Aut}_R M$ of the topological monoid of R-module equivalences from M to M.

1. Introduction

Contemporary algebraic topology features a vast array of generalized cohomology theories, but our knowledge of their geometric content remains limited to the examples of ordinary cohomology theories, topological K-theory and cobordism theories. In this paper we describe the geometry underlying the cohomology theory associated to the algebraic K-theory of a ring, or more generally a ring spectrum. The higher algebraic K-groups $K_n(R)$ of a ring spectrum R may be defined as the homotopy groups of the algebraic K-theory spectrum $K(R)$. By the geometry of $K(R)$-theory, we mean a geometric description of the cocycles whose equivalence classes form the cohomology groups $K(R)^*(X)$ associated

MSC2010: 19D99, 55P43, 55R15, 55R65, 55R70.
Keywords: algebraic K-theory, parametrized spectra, bundle theory, classifying spaces.

to the spectrum $K(R)$. Our methods only give a description of the degree zero cohomology group $K(R)^0(X)$ and the result is reminiscent of the description of topological K-theory $K^0(X)$ in terms of the Grothendieck group of vector bundles over X. The analogue of vector bundles for $K(R)$-theory are parametrized spectra that are modules over the ring spectrum R. We call these objects R-bundles. The main result is the following:

Theorem 1.1. *Let R be a connective ring spectrum and let $K(R)$ be the algebraic K-theory spectrum of R. Then for any finite CW complex X, there is a natural isomorphism*

$$K(R)^0(X) \cong \mathrm{Gr}[\textit{ lifted, free, finite rank } R\textit{-bundles over } X]$$

between the degree zero $K(R)$-cohomology classes of X and the Grothendieck group of the abelian monoid of equivalence classes of lifted R-bundles over X that are free and finite rank as parametrized R-modules.

We will give a precise meaning to all of the terms occurring in the statement of the theorem in §5 and §6, but for now we note that an R-bundle E over X is free of finite rank if every fiber E_x admits an equivalence of R-modules to the n-fold wedge $R^{\vee n}$ for some $n \geq 0$.

Our geometric description of $K(R)$-theory is inspired by previous work. For R a discrete ring, Karoubi [1987] gave a similar description of the cocycles for $K(R)$-theory in terms of fibrations of projective R-modules. For the case where R is the connective complex K-theory spectrum ku, Baas, Dundas, Richter and Rognes [Baas et al. 2004; 2011] interpreted the cocycles of $K(ku)$-theory as 2-vector bundles, which are a categorification of complex vector bundles.

By definition, $K(R)^0(X)$ is the group of homotopy classes of maps from X to the underlying infinite loop space of the algebraic K-theory spectrum, whose homotopy type can be described using Quillen's plus construction:

$$\Omega^\infty K(R) \simeq K_0(R) \times B\mathrm{GL}_\infty^+(R).$$

Here $K_0(R) = K_0^f(\pi_0 R)$ is the Grothendieck group of free modules over the discrete ring $\pi_0 R$ and $B\mathrm{GL}_\infty^+(R)$ is Quillen's plus construction applied to the H-space $B\mathrm{GL}_\infty(R) = \mathrm{colim}_n\, B\mathrm{GL}_n(R)$, where $B\mathrm{GL}_n(R)$ is the classifying space of the derived mapping space $\mathrm{GL}_n(R) = \mathbf{Aut}_R(R^{\vee n})$ of R-module equivalences $R^{\vee n} \to R^{\vee n}$.

One important point is that, unlike the case of vector bundles and complex K-theory, the plus construction can radically change the homotopy type. This forces the bundles that define cocycles for $K(R)$-theory to be *lifted* R-bundles over X, meaning R-bundles defined up to covers of X with homologically trivial fibers — see Section 6 for a precise definition.

The term "bundle" is perhaps a little naive: as one continuously varies the basepoint in X, the fibers of a parametrized spectrum are weak homotopy equivalent, but need not be strictly isomorphic. Put another way, to describe a parametrized spectrum in terms of cocycle data would require a derived or infinitely homotopy coherent descent condition. This point of view naturally leads to a description of parametrized objects as homotopy sheaves with values in a quasicategory, as developed by Ando, Blumberg, Gepner, Hopkins and Rezk [Ando et al. 2010; 2011; 2014a]. Rather than using quasicategories, we follow the foundations of parametrized stable homotopy theory developed by May and Sigurdsson [2006]. In their framework, parametrized spectra are defined in terms of a "total object" over X instead of cocycle data. Homotopical control of the fiber homotopy type of parametrized spectra is maintained via the framework of Quillen model categories.

Theorem 1.1 follows from a general classification theorem for parametrized R-module spectra. In this paper, a spectrum means an orthogonal spectrum, and we use the stable model structure on orthogonal ring and module spectra from Mandell, May, Schwede and Shipley [2001]. Given an R-module M, we say that a parametrized R-module spectrum E over X has fiber M if the fiber E_x of E over every point $x \in X$ admits a stable equivalence $E_x \simeq M$ of R-modules. We use the terms "R-bundle with fiber M" and "parametrized R-module with fiber M" interchangeably. Let $\mathbf{Aut}_R M$ be the derived mapping space of homotopy automorphisms of M as an R-module. In Section 5, we explain how to realize this homotopy type as a group-like topological monoid, so that we may form the classifying space $B\mathbf{Aut}_R M$, and prove the following classification theorem:

Theorem 1.2. *Let X be a CW complex, let R be a ring spectrum and let M be an R-module. There is a natural bijection between stable equivalence classes of R-bundles over X with fiber M and homotopy classes of maps* $[X, B\mathbf{Aut}_R M]$.

When $M = R$, Theorem 1.2 says that line R-bundles over X are classified by the classifying space $B\mathrm{GL}_1 R$ of the units of R. The construction of the line R-bundle associated to a map $f : X \to B\mathrm{GL}_1 R$ is the generalized Thom spectrum studied by Ando, Blumberg, Gepner, Hopkins and Rezk [2014a; 2014b]; see Remark 5.2. From another point of view, a parametrized spectrum with fiber M gives a twisted form of the cohomology theory M. We can then view Theorem 1.2 as giving a general classification theorem of the twists of M-theory.

Ando, Blumberg and Gepner, in their ∞-categorical approach to parametrized homotopy theories, proved in Theorem B.4 of [Ando et al. 2011] that the quasicategory of morphisms $\Pi_\infty X \to \mathscr{S}_\infty$ from the singular simplicial complex of a space X to the quasicategory of spectra $\mathscr{S}_\infty$ is equivalent to the quasicategory associated to the May-Sigurdsson model category of parametrized spectra over X. Variants of their arguments can be used to prove results in the same vein as Theorem 1.2. The proof in this paper is more concrete, using the pullback of a universal bundle to induce the equivalence instead of Lurie's straightening functor [Lurie 2009, §3.2.1].

In order to prove Theorem 1.2, we compare R-bundles with fiber M and principal $\mathrm{Aut}_R M$-fibrations, where $\mathrm{Aut}_R M$ is a point-set model for the derived mapping space of homotopy automorphisms constructed out of appropriate cofibrant and fibrant approximations. Much of the technical material in the paper goes into maintaining control of the fiberwise homotopy type of the principal fibration associated to an R-bundle with fiber M. By carefully intertwining a Quillen-type model structure and a Hurewicz-type model structure, we show that this construction induces a bijection of equivalence classes, and reduce the proof of the classification theorem for R-bundles with fiber M to the classification theorem for principal fibrations.

The classification theorem for R-bundles and the construction of the principal fibration associated to an R-bundle has recently been used by Cohen and Jones [2013a; 2013b] in their study of the gauge group of parametrized spectra and the K-theory of string topology.

Outline. In Section 2, we collect the necessary facts about model category structures on parametrized spaces, introduce a homotopical notion of a G-torsor and compare it to that of a principal G-fibration, where G is a topological monoid. The model category structures on parametrized spectra are recalled in Section 3, then in Section 4 we construct the principal $\mathrm{Aut}_R M$-fibration associated to a bundle with fiber M. We prove in Section 5 that this construction provides an inverse up to homotopy to the associated bundle construction

$$Y \mapsto M \wedge_{\Sigma^\infty_+ \mathrm{Aut}_R M} \Sigma^\infty_B Y,$$

and prove Theorem 1.2. The proof of Theorem 1.1 is given in Section 6.

Topological conventions. We will rely heavily on the foundations for parametrized homotopy theory developed by May and Sigurdsson [2006]. As explained there, it is advantageous to leave the category $\mathscr{U}$ of compactly generated spaces. By a "space" we mean a k-space as defined in [May and Sigurdsson 2006, Definition 1.1.1], and we denote the category of spaces by $\mathscr{K}$. We will always assume that the base space (denoted by B or X) is compactly generated. We assume throughout that the ring spectrum R is well-grounded, meaning that each constituent space is compactly generated and nondegenerately based.

2. Model category theory and principal fibrations

In this section, we recall some basic material on model category structures on the category of parametrized spaces from [May and Sigurdsson 2006]. We then introduce a homotopical notion of a G-torsor, where G is a topological monoid, and show that it is equivalent to that of a principal G-fibration.

The category $\mathscr{K}$ of k-spaces admits a compactly generated topological model structure with weak equivalences the weak homotopy equivalences, fibrations the Serre fibrations, and cofibrations the retracts of relative cell complexes. We refer to this model structure as the q-model structure, and use the terms q-equivalences, q-fibrations, and q-cofibrations for its weak equivalences, fibrations, and cofibrations. Let B be a compactly generated topological space. The category $\mathscr{K}/B$ of spaces $(X, p) = (p : X \to B)$ over B admits a model structure whose weak equivalences and fibrations are detected by the forgetful functor $(X, p) \mapsto X$ to the q-model structure on $\mathscr{K}$. An ex-space is a space (X, p) over B along with a map $s : B \to X$ such that $p \circ s = \mathrm{id}_B$. The category $\mathscr{K}_B$ of ex-spaces (X, p, s) also admits a model structure given by the forgetful functor to the q-model structure on $\mathscr{K}$. We refer to these model structures as the q-model structure on $\mathscr{K}/B$ and $\mathscr{K}_B$, respectively. While both of these model structures are compactly generated and topological, they are not well-grounded, in the sense of [May and Sigurdsson 2006, §5.3–5.6]. The problem is that the generating q-cofibrations and acyclic q-cofibrations do not satisfy the homotopy extension property defined in terms of fiberwise or fiberwise pointed homotopies in $\mathscr{K}/B$ or $\mathscr{K}_B$, as given by Definitions 5.1.7 and 5.1.8. of the same work. Instead, they are only Hurewicz cofibrations in the underlying category of spaces. As a result, applications of the gluing lemma that would allow standard inductive arguments over cell complexes built out of the generating sets fail for these model structures. In attempting to construct a stable model structure on parametrized spectra based on the q-model structure, the verification that relative cell complexes built out of the generating acyclic cofibrations are weak equivalences is unattainable.

As an alternative, May and Sigurdsson [2006, §6.1–6.2] develop the qf-model structure on $\mathscr{K}/B$ and $\mathscr{K}_B$. The qf-model structure also has the q-equivalences as weak equivalences, so that the associated homotopy category is still the homotopy category of spaces over B, but there are fewer qf-cofibrations than q-cofibrations. A qf-fibration need not be a Serre fibration but is a quasifibration. Here, we do not need the details of the definitions, only the fact that in each case the qf-model structure is a well-grounded compactly generated model category. We will work in the unsectioned context, building well-grounded compactly generated model structures on parametrized diagram spaces out of the qf-model structure on $\mathscr{K}/B$.

The category of spaces over B is tensored over the category of spaces via the cartesian product

$$\mathscr{K} \times \mathscr{K}/B \to \mathscr{K}/B, \quad (X, Y \xrightarrow{p} B) \mapsto (X \times Y \xrightarrow{p\circ\pi_2} B).$$

If G is a topological monoid, then we use this structure to define the notion of an object of $\mathscr{K}/B$ with a strictly associative and unital (left) action of G, which we call a G-space over B. The G-spaces over B form a category $G\mathscr{K}/B$ with morphisms the G-equivariant maps over B. Equivalently, the category $G\mathscr{K}/B$ is the comma category $(G\mathscr{K} \downarrow B)$ formed in the category of G-spaces, where we consider B to have a trivial action of G. We will also need the qf-model structure on the category of G-spaces over B.

Proposition 2.1. *Let G be a topological monoid. There is a well-grounded compactly generated model category structure on the category of $G\mathscr{K}/B$ of G-spaces over B with weak equivalences and fibrations created by the forgetful functor to the qf-model structure on spaces over B. If $f : A \to B$ is a map of spaces, then the pullback functor $f^* : G\mathscr{K}/B \to G\mathscr{K}/A$ and its left adjoint $f_!$ form a Quillen adjoint pair for the qf-model structure. If f is a q-equivalence of spaces, then $(f_!, f^*)$ is a Quillen equivalence.*

Proof. The corresponding statements when $G = *$ are Theorem 6.2.5 and Propositions 7.3.4 and 7.3.5 of [May and Sigurdsson 2006]. The generating cofibrations and acyclic cofibrations for the associated model structure on the category of parametrized G-spaces are obtained by applying the free G-space functor $G \times (-)$, defined in terms of the tensor of a space and a space over B, to the generating sets for the qf-model structure on $\mathscr{K}/B$. The result then follows by directly checking the criteria for compactly generated model structures in [May and Sigurdsson 2006, Theorem 5.5.1]. Note that the qf-model structure on $G\mathscr{K}/B$ inherits the property of being right proper from $\mathscr{K}/B$, so it is a well-grounded model structure, see Definition 5.5.4 of the same work. The claims about the adjunction follow directly from the case $G = *$. □

In particular, the fiber functor $i_b^* = (-)_b$ is right Quillen on the category of G-spaces over B. We let $\mathbf{F}_b = \mathbb{R}i_b^*$ denote its right derived functor. In other words, $\mathbf{F}_b Y$ is the object of the homotopy category of G-spaces determined by the fiber $(R^{qf}Y)_b$ of a qf-fibrant approximation of Y.

While the following terminology is nonstandard, it will be useful as an intermediary between the highly structured notion of a principal G-fibration and the model-theoretic fiber conditions on parametrized spectra.

Definition 2.2. A G-torsor over B is a G-space (Y, p) over B for which every derived fiber $\mathbf{F}_b Y$ admits a zigzag of q-equivalences of G-spaces to G, considered as a G-space via left multiplication. We write $\mathrm{Ho}(G\,\mathrm{Tor}/B)$ for the full subcategory of the homotopy category $\mathrm{Ho}(G\mathscr{K}/B)$ of G-spaces over B that is spanned by the G-torsors.

The notion of a G-torsor is native to the Quillen model structure. The following definition instead uses the Hurewicz model structure.

Definition 2.3. A principal G-fibration over B is a G-space (Y, p) over B for which

- the structure map $p : Y \to B$ is an h-fibration of G-spaces, meaning that it has the homotopy lifting property in the category of G-spaces,
- for every $b \in B$, there is a zigzag of weak equivalences of G-spaces $Y_b \simeq G$.

We now construct an approximation functor Γ in order to compare G-torsors with principal G-fibrations. Given a G-space (Y, p) over B, let $(\Gamma Y, \Gamma p)$ be the G-space over B defined by the mapping path-space construction

$$\Gamma p : \Gamma Y = B^I \times_B Y \xrightarrow{\mathrm{ev}_1} B, \quad (\gamma, y) \mapsto \gamma(1),$$

and note that the fiber $(\Gamma Y)_b$ of Γp is the homotopy fiber of Y at $b \in B$. Note that the map Γp is an h-fibration of G-spaces, since the lifting problem in the category of G-spaces

$$\begin{tikzcd} X \arrow[r, "{(\gamma, f)}"] \arrow[d, "i_0"'] & B^I \times_B Y \arrow[d, "\mathrm{ev}_1"] \\ X \times I \arrow[r, "h"] \arrow[ur, dashed, "\widetilde{h}"] & B \end{tikzcd}$$

has a solution given by $\widetilde{h}(x, t) = (\lambda_t(x), f(x))$, where $\lambda_t(x)$ is the path

$$\lambda_t(x)(s) = \begin{cases} \gamma(x)(s + st) & \text{for } 0 \le s \le 1/(1+t), \\ h(x, s + ts - 1) & \text{for } 1/(1+t) \le s \le 1, \end{cases}$$

and the map $\widetilde{h}$ is evidently G-equivariant.

The construction of mapping path-spaces is functorial, so that Γ defines an endofunctor of the category of G-spaces over B with the following easily verifiable properties:

Lemma 2.4. (i) *If p is a quasifibration and every fiber Y_b is q-equivalent to G, then $(\Gamma Y, \Gamma p)$ is a principal G-fibration over B.*

(ii) *Suppose that the map $(X, p) \to (Y, q)$ is a q-equivalence of G-spaces over B. Then the induced map $(\Gamma X, \Gamma p) \to (\Gamma Y, \Gamma q)$ is a q-equivalence of principal G-fibrations.*

(iii) *The map $(Y, p) \to (\Gamma Y, \Gamma p)$ defined by the inclusion into constant paths is a homotopy equivalence of G-spaces over B. If p is a quasifibration, then the map restricts to a q-equivalence on fibers.*

Proposition 2.5. *The functor Γ induces a natural isomorphism between the set of q-equivalence classes of G-torsors over B and the set of q-equivalence classes of principal G-fibrations over B.*

Proof. Let (Y, p) be a G-space over B. The inclusion of the fiber into the homotopy fiber for both (Y, p) and a qf-fibrant approximation $(R^{qf} Y, R^{qf} p)$ are related by the commutative diagram

$$
\begin{tikzcd}
Y_b \arrow[r] \arrow[d] & (R^{qf}Y)_b \arrow[d, "\simeq"] \\
(\Gamma Y)_b \arrow[r, "\simeq"] & (\Gamma R^{qf}Y)_b
\end{tikzcd}
\tag{2.6}
$$

induced by fibrant approximation and the inclusion of constant paths. Since the fibrant approximation is a q-equivalence of total spaces, it induces a q-equivalence of the homotopy fibers. The qf-fibration $R^{qf}p$ is in particular a quasifibration, which gives the other displayed q-equivalence. It follows that the derived fiber $\mathbf{F}_bY$ is canonically q-equivalent to the homotopy fiber $(\Gamma Y)_b$.

Thus Γ takes G-torsors to principal G-fibrations and preserves q-equivalences. Conversely, every principal G-fibration is a G-torsor. The map $\eta : Y \to \Gamma Y$ in Lemma 2.4.(iii) is a q-equivalence of G-spaces, so Γ is bijective on q-equivalence classes. $\square$

Using the proposition, the next theorem is a restatement of May's classification theorem [1975, Theorem 9.2] for principal G-fibrations.

Theorem 2.7. *Let G be a grouplike topological monoid with nondegenerate basepoint and let B be a CW complex. Taking the pullback of $\Gamma EG \to BG$ along a given map $B \to BG$ defines a natural bijective correspondence between the set of homotopy classes of maps $[B, BG]$ and the set of equivalence classes of G-torsors over B.*

3. Model categories of parametrized spectra

We now summarize what we need from the theory of parametrized spectra, following [May and Sigurdsson 2006, Chapters 11–12]. A spectrum over B is an orthogonal spectrum in the category of ex-spaces over B. That is, a parametrized spectrum X consists of an $O(V)$-equivariant ex-space $(X(V), p(V), s(V))$ for each finite-dimensional real inner product space V, along with compatible $(O(V) \times O(W))$-equivariant structure maps

$$\sigma : X(V) \wedge_B S^W_B \to X(V \oplus W)$$

over and under B. Here $S^V_B = r^*S^V = S^V \times B$ is the trivially twisted ex-space with fiber the one-point compactification S^V. The section of S^V_B is determined by the basepoint of S^V. The smash product $\wedge_B$ is the fiberwise smash product of ex-spaces. A map $f : X \to Y$ of spectra over B consists of an equivariant map $f(V) : X(V) \to Y(V)$ of ex-spaces for each indexing space V that are suitably compatible with the structure maps σ. For each point $b \in B$, the fiber of X over b is the spectrum $X_b = i_b^*X$ given by the pullback of X along the inclusion map

$i_b : \{b\} \to B$. The fiber spectrum is described levelwise in terms of the fibers of its constituent ex-spaces by the formula $X_b(V) = X(V)_b$.

The level model structure on the category $\mathscr{S}_B$ of spectra over B has as weak equivalences, respectively fibrations, those maps f such that each $f(V)$ is a q-equivalence, respectively qf-fibration, of ex-spaces. We refer to these maps as the levelwise q-equivalences and levelwise qf-fibrations, respectively. The homotopy groups of a levelwise qf-fibrant spectrum X over B are the homotopy groups $\pi_q X_b$ of all of the fibers of X. The homotopy groups of a spectrum X over B are the homotopy groups $\pi_q(R^l X)_b$ of the fibers of a levelwise qf-fibrant approximation $R^l X$ of X. We say that a map $X \to Y$ of spectra over B is a stable equivalence if it induces an isomorphism on all homotopy groups of all fibers. An Ω-spectrum over B is a level qf-fibrant spectrum X over B whose adjoint structure maps

$$\widetilde{\sigma} : X(V) \to \Omega_B^W X(V \oplus W)$$

are q-equivalences of ex-spaces over B.

Theorem 3.1 [May and Sigurdsson 2006, Theorem 12.3.10]. *The category $\mathscr{S}_B$ of spectra over B admits the structure of a well-grounded compactly generated model category whose weak equivalences are the stable equivalences. The fibrations and cofibrations are called the s-fibrations and the s-cofibrations, and the s-fibrant objects are the Ω-spectra over B. We refer to this model structure as the s-model structure (or stable model structure) on $\mathscr{S}_B$.*

In the case $B = *$, this coincides with the stable model structure on orthogonal spectra from Mandell, May, Schwede and Shipley [2001].

Parametrized spaces and parametrized spectra are related by suspension spectrum and underlying infinite loop space functors. If (Y, p) is a space over B, the fiberwise suspension spectrum $\Sigma_B^\infty Y$ is the spectrum over B defined by

$$(\Sigma_B^\infty Y)(V) = (Y, p)_+ \wedge_B S_B^V,$$

where

$$(Y, p)_+ = (Y \amalg B, p \amalg \mathrm{id}_B, \mathrm{id}_B)$$

is the ex-space over B obtained from (Y, p) by adjoining a disjoint section. The right adjoint Ω_B^∞ of Σ_B^∞ is defined by $\Omega_B^\infty X = X(0)$. By inspecting the definitions, we see that there are natural isomorphisms of fibers $(\Sigma_B^\infty Y)_b \cong \Sigma_+^\infty Y_b$ and $(\Omega_B^\infty X)_b \cong \Omega^\infty X_b$.

The category $\mathscr{S}_B$ of spectra over B is enriched and tensored over the category $\mathscr{S}$ of spectra with tensor the fiberwise smash product $\wedge$. We use this structure to define parametrized module spectra. Let R be a (nonparametrized) ring spectrum. We assume, once and for all, that R is well-grounded, meaning that each $R(V)$ is well-based and compactly generated. An R-module over B is a spectrum N over B with an associative and unital map of spectra $R \wedge N \to N$ over B.

Theorem 3.2 [May and Sigurdsson 2006, Theorem 14.1.7]. *The category $R\mathrm{Mod}_B$ of R-modules over B is a well-grounded compactly generated model category with weak equivalences and fibrations created by the forgetful functor to $\mathscr{S}_B$. We refer to this model structure as the s-model structure on $R\mathrm{Mod}_B$.*

If X is a space and Y is a space over B, then there is a natural isomorphism of parametrized spectra over B

$$\Sigma_B^\infty(X \times Y) \cong \Sigma_+^\infty X \wedge \Sigma_B^\infty Y$$

that satisfies the analogues of the associativity and unit diagrams for a monoidal natural transformation. Similarly, Ω_B^∞ preserves the monoidal structure up to a lax monoidal transformation, so that if G is a topological monoid, then the adjunction $(\Sigma_B^\infty, \Omega_B^\infty)$ restricts to give an adjunction between G-spaces over B and $\Sigma_+^\infty G$-module spectra over B.

Proposition 3.3. (i) *The adjoint pair $(\Sigma_B^\infty, \Omega_B^\infty)$ is a Quillen adjunction between the qf-model structure on spaces over B and the s-model structure on spectra over B.*

(ii) *Let G be a topological monoid. The adjoint pair $(\Sigma_B^\infty, \Omega_B^\infty)$ is a Quillen adjunction between the qf-model structure on G-spaces over B and the s-model structure on $\Sigma_+^\infty G$-modules over B.*

Proof. In both cases, this follows by examining the effect of Σ_B^∞ on generating cofibrations and acyclic cofibrations; since the s-model structure on $\mathscr{S}_B$ is a left Bousfield localization of the level qf-model structure, its generating sets contain all maps of the form $\Sigma_B^\infty i$, where i runs through the generating sets for the qf-model structure on $\mathscr{K}/B$. □

It is a formal consequence that the left Quillen functor Σ_B^∞ preserves weak equivalences between cofibrant objects. However, it will be useful to know that a stronger result is true.

Lemma 3.4. *The functor $\Sigma_B^\infty : \mathscr{K}/B \to \mathscr{S}_B$ preserves all weak equivalences.*

Proof. If $f : X \to Y$ is a weak homotopy equivalence of spaces over B, then each map of ex-spaces $f_+ \wedge_B S_B^V$ is a weak homotopy equivalence on total spaces. This means that $\Sigma_B^\infty f$ is a levelwise weak homotopy equivalence and thus a stable equivalence of parametrized spectra by [May and Sigurdsson 2006, Lemma 12.3.5]. □

We will work in the nonparametrized setting for a moment in order to fix notation on some constructions. Suppose that R and A are ring spectra. Consider the function spectrum $F^R(-,-)$ of R-modules. If P is an A-module, M is an (R, A)-bimodule and N is an R-module, then $F^R(M, N)$ is an A-module and we have the following adjunction:

$$\mathrm{Mod}_R(M \wedge_A P, N) \cong \mathrm{Mod}_A(P, F^R(M, N)). \tag{3.5}$$

It is a consequence of the fact that the category of R-modules is a spectrally enriched model category via the function spectra $F^R(-,-)$ that if M' is a cofibrant R-module, then the functor $F^R(M', -)$ preserves stable equivalences between fibrant R-modules, and similarly if N is a fibrant R-module, then the functor $F^R(-, N)$ preserves stable equivalences between cofibrant R-modules.

We will be interested in the generalization of the adjunction (3.5) where N and P are parametrized spectra. The smash product $M \wedge_A P$ occurring in the parametrized version of the adjunction is built out of the external smash product $\wedge : \mathscr{S} \times \mathscr{S}_B \to \mathscr{S}_B$, as described in [May and Sigurdsson 2006, §14.1]. In particular, there is never a need to internalize the smash product by taking the pullback Δ^* of a spectrum over $B \times B$ along the diagonal map. In this situation, we are able to maintain homotopical control of the smash product.

Lemma 3.6. *Let $i : X \to Y$ be an s-cofibration of R-modules and let $j : Z \to W$ be an s-cofibration of spectra over B. Then the pushout product*

$$i \,\square\, j : (Y \wedge Z) \cup_{X \wedge Z} (X \wedge W) \to Y \wedge W$$

is an s-cofibration of R-modules over B and a stable equivalence if either i or j is.

Proof. Since parametrized spectra and R-modules are well-grounded categories, we may induct up the cellular filtration of i and j, so it suffices to verify the result when i and j are generating cofibrations or generating acyclic cofibrations. This follows from the case when $R = S$ [May and Sigurdsson 2006, Proposition 12.6.5] because $R \wedge (-)$ takes s-cofibrations and acyclic s-cofibrations of spectra over B to s-cofibrations and acyclic s-cofibrations of R-modules over B. $\square$

The lemma has the following consequence:

Proposition 3.7. *Suppose that M is an (R, A)-bimodule that is cofibrant as an R-module. Then the adjunction*

$$(A\text{-modules over } B) \underset{F^R(M,-)}{\overset{M \wedge_A (-)}{\rightleftarrows}} (R\text{-modules over } B)$$

is a Quillen adjunction.

Proof. It follows from the lemma that the adjunction is Quillen when $A = S$. In particular, the functor $F^R(M, -)$ is right Quillen when we consider its codomain to be parametrized spectra. The general case then holds as well because s-fibrations and weak equivalences of A-modules over B are created by the forgetful functor to parametrized spectra. $\square$

4. The principal $\mathrm{Aut}_R M$-fibration associated to an R-bundle

Let R be a ring spectrum and let M be an R-module. In this section, we will define the topological monoid $\mathrm{Aut}_R M$ of autoequivalences of R-modules $M \to M$. We then describe the construction of an $\mathrm{Aut}_R M$-torsor from an R-bundle with fiber M.

Suppose that G is a topological monoid. While G may not be grouplike, there is a maximal grouplike submonoid $G^\times \subset G$ defined as the pullback

$$\begin{tikzcd} G^\times \arrow[r] \arrow[d] & G \arrow[d] \\ (\pi_0 G)^\times \arrow[r] & \pi_0 G \end{tikzcd} \tag{4.1}$$

where $(\pi_0 G)^\times \subset \pi_0 G$ is the subset of invertible elements of the monoid $\pi_0 G$. In other words, the inclusion $G^\times \to G$ is given by the inclusion of those path components that are invertible under the monoid multiplication. For example, if $G = \Omega^\infty R = R(0)$ is the multiplicative topological monoid underlying an s-fibrant ring spectrum R, then $G^\times = \mathrm{GL}_1 R$ is the space of units of R. A more delicate construction is required if R is commutative and one wants to keep control of the resulting E_∞-space structure on $\mathrm{GL}_1 R$ [Lind 2013; Schlichtkrull 2004; Sagave and Schlichtkrull 2013], but we will not need this for our purposes.

We assume for the rest of the section that R is an s-cofibrant ring spectrum and that M is an s-fibrant and s-cofibrant R-module. The function spectrum $F^R(M, M)$ is a ring spectrum under composition of maps and our assumptions guarantee that it is s-fibrant. Let

$$\mathrm{End}_R M = \Omega^\infty F^R(M, M) = F^R(M, M)(0)$$

be the underlying topological monoid. We define $\mathrm{Aut}_R M$ to be the units of the ring spectrum $F^R(M, M)$:

$$\mathrm{Aut}_R M = \mathrm{GL}_1 F^R(M, M) = (\Omega^\infty F^R(M, M))^\times.$$

We think of $\mathrm{Aut}_R M$ as the space of weak equivalences of R-modules $M \to M$, with monoid multiplication given by composition. The suspension spectrum of the monoid $\mathrm{Aut}_R M$ is a ring spectrum $\Sigma^\infty_+ \mathrm{Aut}_R M$. The R-module M also has the structure of a right $\Sigma^\infty_+ \mathrm{Aut}_R M$-module, with action map

$$M \wedge_S \Sigma^\infty_+ \mathrm{Aut}_R M \to M$$

the adjoint of the composite map of ring spectra

$$\Sigma^\infty_+ \mathrm{Aut}_R M \to \Sigma^\infty_+ \Omega^\infty F^R(M, M) \xrightarrow{\epsilon} F^R(M, M)$$

induced by the canonical inclusion $\mathrm{GL}_1 \to \Omega^\infty$ and the counit of the adjunction $(\Sigma^\infty_+, \Omega^\infty)$. Thus M is a $(R, \Sigma^\infty_+ \mathrm{Aut}_R M)$-bimodule.

We write $\mathbf{F}_b = \mathbb{R}i_b^*(-)$ for the right derived fiber functor. If N is an R-module over B, the derived fiber $\mathbf{F}_b R$ is the object of the homotopy category of R-modules determined by the fiber $i_b^* R^s N$ of an s-fibrant approximation of N as an R-module over B.

Definition 4.1. An R-bundle over B with fiber M is an R-module N over B such that every derived fiber $\mathbf{F}_b N$ of N admits a zigzag of stable equivalences of R-modules to M.

Let N be an R-bundle over B. The function spectrum $F^R(M, N)$ is a $\Sigma^\infty_+ \mathrm{End}_R M$-module over B. Applying Ω^∞_B we get an $\mathrm{End}_R M$-space $\Omega^\infty_B F^R(M, N)$ over B which is qf-fibrant when N is s-fibrant. The following lemma allows us to keep control of its fiber homotopy type. It is a direct consequence of the cofibrancy of M as an R-module.

Lemma 4.2. *Suppose that N is s-fibrant and fix a point $b \in B$. A stable equivalence of R-modules $N_b \simeq M$ determines*

(i) *a stable equivalence of $\Sigma^\infty_+ \mathrm{End}_R M$-modules $F^R(M, N)_b \simeq F^R(M, M)$, and*

(ii) *a q-equivalence of $\mathrm{End}_R M$-spaces $\Omega^\infty_B F^R(M, N)_b \simeq \Omega^\infty F^R(M, M)$.*

Notice that the second condition in the lemma implies that $\Omega^\infty_B F^R(M, N)$ is an $\mathrm{End}_R M$-torsor. We will now construct an $\mathrm{Aut}_R M$-torsor

$$E^R(M, N) \subset \Omega^\infty_B F^R(M, N).$$

The idea of the construction is to restrict to the subspace whose fiber over $b \in B$ consists of the stable equivalences of R-modules $M \to N_b$. To make this idea rigorous, we need to access the components $\pi_0 \Omega^\infty_B F^R(M, N)_b$ of each fiber in a way that remembers the topology of B.

To this end, we define the parametrized components $\pi_0^B X$ of a parametrized space $p : X \to B$. As a set, $\pi_0^B X$ consists of all components of all fibers of X:

$$\pi_0^B X = \bigcup_{b \in B} \pi_0 X_b.$$

Give $\pi_0^B X$ the quotient topology induced by the map $X \to \pi_0^B X$ that sends a point $x \in X$ to its component $[x] \in \pi_0 X_{p(x)}$. Since the quotient map is a map over B, the space $\pi_0^B X$ is a parametrized space over B.

Construction 4.3. We now define a fiberwise version of (4.1). Let G be a topological monoid and let (Y, p) be a G-torsor over B whose structure map $p : Y \to B$ is a quasifibration. A choice of q-equivalence of G-spaces $Y_b \simeq G$ gives an isomorphism of $\pi_0 G$-spaces $\pi_0 Y_b \cong \pi_0 G$. Define $\pi_0 Y_b^\times$ to be the subset of $\pi_0 Y_b$ corresponding to $\pi_0 G^\times$ under this isomorphism. Although the isomorphism $\pi_0 Y_b^\times \cong \pi_0 G^\times$ of $\pi_0 G^\times$-spaces depends on the choice of q-equivalence $Y_b \simeq G$, the subset $\pi_0 Y_b^\times$ does not. Let $\pi_0^B Y^\times \subset \pi_0^B Y$ be the subspace consisting of the sets $\pi_0 Y_b^\times$ in each fiber. Define the space $Y^\times$ over B to be the following pullback:

$$\begin{array}{ccc} Y^\times & \xrightarrow{\iota} & Y \\ \downarrow & & \downarrow \\ \pi_0^B Y^\times & \longrightarrow & \pi_0^B Y \end{array}$$

Notice that there is a canonical isomorphism $(Y^\times)_b \cong Y_b^\times$, and that a map $X \to Y$ of spaces over B factors through $Y^\times$ if and only if for every $b \in B$, the induced map $\pi_0 X_b \to \pi_0 Y_b$ has image lying in $\pi_0 Y_b^\times$.

It is straightforward to verify that the construction $Y \mapsto Y^\times$ is functorial for maps of G-spaces. We will at times write $\mu = (-)^\times$ for the resulting functor. The assumption that p is a quasifibration is the minimal hypothesis necessary for the construction to be possible. In practice, p will be either a qf-fibration or an h-fibration.

Lemma 4.4. *Suppose that the base space B is semilocally contractible and that (Y, p) is a principal G-fibration over B. Then $\iota : Y^\times \to Y$ is the inclusion of a subspace of path components.*

Proof. Let γ be a path in Y with $\gamma(0) \in Y^\times$. Assuming that $\gamma(1) \notin Y^\times$, let $t_0 = \inf\{t \in [0, 1] \mid \gamma(t) \notin Y^\times\}$. Set $b_0 = p(\gamma(t_0))$ and choose an open neighborhood U of b_0 along with a nullhomotopy $h : U \times I \to B$ of U in B. Consider the G-space h^*Y over $U \times I$ obtained from Y by pullback along h. The restriction $h^*Y|_{U\times\{0\}}$ is isomorphic to $Y|_U$, while the restriction $h^*Y|_{U\times\{1\}}$ is isomorphic to $U \times Y_{b_0}$. It follows that we may find a fiberwise homotopy equivalence of G-spaces $\rho : Y|_U \to U \times Y_{b_0}$ over U. Applying the functor $(-)^\times$ to ρ, we have a commutative diagram

$$\begin{array}{ccc} Y^\times|_U & \xrightarrow{\rho^\times} & U \times Y_{b_0}^\times \\ \downarrow & & \downarrow \\ Y|_U & \xrightarrow{\rho} & U \times Y_{b_0} \end{array}$$

which shows that in a neighborhood of t_0, the path $\rho \circ \gamma$ must lie in $U \times Y_{b_0}^\times$. Since ρ is a fiberwise homotopy equivalence, it follows that $\gamma(t) \in Y_{p(\gamma(t))}^\times$ for t near t_0, contradicting our initial assumption. □

Proposition 4.5. *Suppose that (Y, p) is a G-torsor over a semilocally contractible space B.*

(i) *The space $Y^\times$ is a $G^\times$-space over B and the canonical inclusion $\iota : Y^\times \to Y$ is a map of $G^\times$-spaces.*

(ii) *If the structure map $p : Y \to B$ is an h-fibration of G-spaces, so that Y is a principal G-fibration, then $p^\times : Y^\times \to B$ is a $G^\times$-torsor.*

(iii) *The functor $\mu : Y \mapsto Y^\times$ preserves q-equivalences between principal G-fibrations.*

Proof. Claim (i) is immediate from the definitions. For (ii), observe that by Lemma 4.4, $p^\times$ is an h-fibration of spaces. It follows that the natural map $Y_b^\times \to F_b Y^\times$ from the fiber to the homotopy fiber is a q-equivalence. A given chain of q-equivalences of G-spaces $Y_b \simeq G$ induces a chain of q-equivalences of $G^\times$-spaces $Y_b^\times \simeq G^\times$, so we conclude that $Y^\times$ is a $G^\times$-torsor.

For (iii), assume that $(Y, p) \to (Z, q)$ is a q-equivalence of G-torsors with p and q both h-fibrations of G-spaces. For any $b \in B$, the induced map of fibers $Y_b \to Z_b$ is a q-equivalence of G-spaces, and so the induced map of $G^\times$-spaces $Y_b^\times \to Z_b^\times$ is a q-equivalence. Since $p^\times$ and $q^\times$ are h-fibrations, it follows that $Y^\times \to Z^\times$ is a q-equivalence on total spaces. □

As a consequence of Proposition 4.5, we may define the derived functor of μ to be the functor from the homotopy category of G-torsors to the homotopy category of $G^\times$-torsors

$$\boldsymbol{\mu} : \mathrm{Ho}(G\,\mathrm{Tor}\,/B) \to \mathrm{Ho}(G^\times\,\mathrm{Tor}\,/B),$$
$$Y \mapsto \mu(\Gamma Y) = (\Gamma Y)^\times,$$

where Γ is the h-fibrant approximation functor from Section 2. Lemma 4.4 implies that when $p : Y \to B$ is an h-fibration, the fiber $Y_b^\times \cong (Y^\times)_b$ represents the derived fiber $\mathbf{F}_b Y^\times$ of $Y^\times$. In other words:

Lemma 4.6. *There is a canonical isomorphism of derived functors $\mathbf{F}_b \boldsymbol{\mu} \cong \boldsymbol{\mu} \mathbf{F}_b$.*

We will also need to know how to construct maps into $\boldsymbol{\mu}$.

Lemma 4.7. *A morphism $X \to Y$ in the homotopy category of $G^\times$-spaces over B factors through $\iota : \boldsymbol{\mu} Y \to Y$ if and only if for every $b \in B$, the induced map $\pi_0 \mathbf{F}_b X \to \pi_0 \mathbf{F}_b Y$ has image contained in the subset $\pi_0 \boldsymbol{\mu} \mathbf{F}_b Y$.*

Proof. First notice that the functor $\pi_0 \mathbf{F}_b$ is invariant under weak equivalences of spaces over B. We may represent a map in the homotopy category of $G^\times$-spaces over B by a zigzag of map where the wrong way maps are weak equivalences, and we assume without loss of generality that the final object in this zigzag is a h-fibrant $G^\times$-space over B. The result then follows by using the universal mapping property of the pullback of spaces defining μ. □

Definition 4.8. Let N be an R-bundle with fiber M and let R^sN be an s-fibrant approximation of N as an R-module over B. Since M is an s-cofibrant R-module, the $\mathrm{End}_R M$-torsor $\Omega_B^\infty F^R(M, R^sN)$ is qf-fibrant as an $\mathrm{End}_R M$-module. Applying Construction 4.3 defines an $\mathrm{Aut}_R M$-space over B

$$E^R(M, R^sN) = (\Omega_B^\infty F^R(M, R^sN))^\times,$$

which need not be an $\mathrm{Aut}_R M$-torsor. If we instead take the derived functor $\boldsymbol{\mu}$ by applying the h-fibration approximation functor Γ before $(-)^\times$, then the value of the associated derived functor

$$\mathbf{E}^R(M, N) = \boldsymbol{\mu}\Omega_B^\infty F^R(M, R^sN)$$

is our definition of the $\mathrm{Aut}_R M$-torsor associated to the R-bundle N. Since Ω_B^∞ and $F^R(M, -)$ are both right Quillen functors, we can summarize the definition by saying that

$$\mathbf{E} = \mathbf{E}^R(M, -) : \mathrm{Ho}(R\text{-bundles with fiber } M) \to \mathrm{Ho}(\mathrm{Aut}_R M\text{-torsors})$$

is the composite derived functor $\mathbf{E} = \boldsymbol{\mu} \circ \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is the right derived functor of $\Omega = \Omega_B^\infty F^R(M, -)$.

5. The classification of R-bundles

In the previous section we constructed an $\mathrm{Aut}_R M$-torsor from an R-bundle with fiber M. We now construct an R-bundle with fiber M from an $\mathrm{Aut}_R M$-torsor and show that the constructions are homotopy inverse to each other. At the end of the section, we complete the proof of Theorem 1.2. We assume that B is a CW complex, in particular semilocally contractible, so the functor μ from the previous section is well-behaved. We continue to assume that R is an s-cofibrant ring spectrum and that M is an s-bifibrant R-module.

For technical reasons, it will be useful to work with a q-cofibrant approximation $\mathrm{Aut}_R^c M \to \mathrm{Aut}_R M$ of $\mathrm{Aut}_R M$ as a topological monoid. By pullback along the approximation map, any $\mathrm{Aut}_R M$-torsor is also an $\mathrm{Aut}_R^c M$-torsor, so we consider the functor $\mathbf{E} = \boldsymbol{\mu} \circ \boldsymbol{\Omega}$ from the previous section as taking values in $\mathrm{Aut}_R^c M$-torsors. Similarly, the right $\Sigma_+^\infty \mathrm{Aut}_R M$-module structure of M pulls back to give a right $\Sigma_+^\infty \mathrm{Aut}_R^c M$-module structure on M.

Definition 5.1. If Y is an $\mathrm{Aut}_R^c M$-space over B, then the fiberwise suspension spectrum $\Sigma_B^\infty Y$ is a $\Sigma_+^\infty \mathrm{Aut}_R^c M$-module spectrum over B. The construction

$$T(Y) = M \wedge_{\Sigma_+^\infty \mathrm{Aut}_R^c M} \Sigma_B^\infty Y$$

defines a functor from $\mathrm{Aut}_R^c M$-spaces over B to R-module spectra over B which is left Quillen by Propositions 3.3 and 3.7. We let $\mathbf{T} = \mathbb{L}T$ denote its left derived functor. Note that $\mathbf{T}$ is left adjoint to the right derived functor $\mathbf{\Omega} = \Omega_B^\infty \mathbf{F}^R(M, -)$. In Proposition 5.6, we will prove that when Y is an $\mathrm{Aut}^c_R M$-torsor over B, then $\mathbf{T}Y$ is an R-bundle with fiber M, so that we have a functor

$$\mathbf{T} : \mathrm{Ho}(\mathrm{Aut}^c_R M\text{-torsors}) \to \mathrm{Ho}(R\text{-bundles with fiber } M).$$

Remark 5.2. In the case $M = R$, the definition recovers the construction of generalized Thom spectra from [Ando et al. 2014a, 2014b]. Given a map of spaces $f : B \to B\mathrm{GL}_1 R$, the classification of principal $\mathrm{GL}_1 R$-fibrations gives a principal $\mathrm{GL}_1 R$-fibration Y_f over B. Applying the functor T then gives a rank one R-bundle over B. The Thom spectrum associated to the map f is the (nonparametrized) R-module spectrum

$$\mathrm{M}f = r_! T Y_f \cong R \wedge_{\Sigma^\infty_+ \mathrm{GL}_1^c R} \Sigma^\infty_+ Y_f,$$

where $r_! : \mathscr{S}_B \to \mathscr{S}$ is left adjoint to the pullback functor $r^* : \mathscr{S} \to \mathscr{S}_B$.

The fiber functor $(-)_b = i_b^*$ is a left adjoint, but is not left Quillen for either the stable model structure on parametrized spectra or the qf-model structure on parametrized spaces. However, i_b^* is a right Quillen functor. On the other hand, $T = M \wedge_{\Sigma^\infty_+ \mathrm{Aut}^c_R M} \Sigma^\infty_B(-)$ is a left Quillen functor. There is a natural isomorphism of functors

$$(M \wedge_{\Sigma^\infty_+ \mathrm{Aut}^c_R M} \Sigma^\infty_B Y)_b \cong M \wedge_{\Sigma^\infty_+ \mathrm{Aut}^c_R M} \Sigma^\infty_+ Y_b$$

at the point-set level, but this does not imply an isomorphism of derived functors after passage to homotopy categories because we are composing left and right derived functors.

In order to prove the commutation of derived functors, we will make a slight modification to the functor T. By identifying $R \wedge_S (\Sigma^\infty_+ \mathrm{Aut}^c_R M)^{\mathrm{op}}$-modules with $(R, \Sigma^\infty_+ \mathrm{Aut}^c_R M)$-bimodules, the category of $(R, \Sigma^\infty_+ \mathrm{Aut}^c_R M)$-bimodules is a well-grounded compactly generated model category with weak equivalences and fibrations created in the s-model structure on spectra [Mandell et al. 2001, Theorem 12.1]. Let $M^\circ \to M$ be an s-cofibrant approximation of M as an $(R, \Sigma^\infty_+ \mathrm{Aut}^c_R M)$-bimodule and define

$$T^\circ(Y) = M^\circ \wedge_{\Sigma^\infty_+ \mathrm{Aut}^c_R M} \Sigma^\infty_B Y.$$

Note that since Σ^∞_+ is left Quillen, $\Sigma^\infty_+ \mathrm{Aut}^c_R M$ is s-cofibrant as a ring spectrum, and thus s-cofibrant as a spectrum. We record a basic consequence.

Lemma 5.3. *The underlying left R-module of M° is s-cofibrant. The underlying right $\Sigma^\infty_+ \mathrm{Aut}^c_R M$-module of M° is s-cofibrant.*

Proof. The right adjoint of the forgetful functor from $(R, \Sigma^\infty_+ \mathrm{Aut}^c_R M)$-bimodules to left R-modules is the function spectrum functor $F^S(\Sigma^\infty_+ \mathrm{Aut}^c_R M, -)$. This functor preserves fibrations and acyclic fibrations because $\Sigma^\infty_+ \mathrm{Aut}^c_R M$ is s-cofibrant.

Therefore its left adjoint, the forgetful functor, preserves cofibrations and acyclic cofibrations. This proves the first claim. The second claim follows using a similar argument and the fact that R is s-cofibrant. □

When $B = *$, the fact that M° is s-cofibrant as a $\Sigma^\infty_+ \mathrm{Aut}^c_R M$-module implies that the functor $M^\circ \wedge_{\Sigma^\infty_+ \mathrm{Aut}^c_R M} (-)$ preserves stable equivalences [Mandell et al. 2001, Proposition 12.7]. Along with Lemma 3.4, this shows that the functor T° takes q-equivalences to stable equivalences when the base is a point. The proof of the next result is inspired by Shulman's examples in [Shulman 2011, §9]. To improve clarity, we temporarily revert to the usual notation $\mathbb{L}$ and $\mathbb{R}$ for left and right derived functors.

Lemma 5.4. *Let $f : * \to B$ be the inclusion of a point. Then there is a natural isomorphism of derived functors $\mathbb{R}f^*\mathbb{L}T \cong \mathbb{L}T\mathbb{R}f^*$.*

Proof. The equivalence $M^\circ \to M$ induces an isomorphism of derived functors $\mathbb{R}F^R(M, -) \cong \mathbb{R}F^R(M^\circ, -)$ since M and M° are cofibrant R-modules. This determines an isomorphism of derived functors $\mathbb{L}T^\circ \cong \mathbb{L}T$, so it suffices to prove the result with T replaced by T°.

Suppose that X is a qf-bifibrant $\mathrm{Aut}^c_R M$-space over B, and consider the following natural transformation of R-modules:

$$T^\circ Q^{qf} f^* X \to T^\circ f^* X \xrightarrow{\cong} f^* T^\circ X \to f^* R^s T^\circ X, \tag{5.5}$$

where the first and third maps are induced by qf-cofibrant approximation and s-fibrant approximation, respectively. Since T° preserves all weak equivalences when the base is a point, the first map is a stable equivalence. The second map is the canonical isomorphism. It remains to show that f^* preserves the stable equivalence $T^\circ X \to R^s T^\circ X$.

Factor f as a q-equivalence followed by a q-fibration, and consider the two cases separately. In the first case, the Quillen adjunction $(f_!, f^*)$ is a Quillen equivalence both for parametrized $\mathrm{Aut}^c_R M$-spaces (Proposition 2.1) and parametrized R-modules (the case of $R = S$ is Proposition 12.6.7 in [May and Sigurdsson 2006] and the general case follows since stable equivalences and s-fibrations of R-modules are detected by the forgetful functor to parametrized spectra). It follows that the natural transformation of derived functors

$$\mathbb{L}T^\circ\mathbb{R}f^* \xrightarrow{\eta} \mathbb{R}f^*\mathbb{L}f_!\mathbb{L}T^\circ\mathbb{R}f^* \cong \mathbb{R}f^*\mathbb{L}T^\circ\mathbb{L}f_!\mathbb{R}f^* \xrightarrow{\epsilon} \mathbb{R}f^*\mathbb{L}T^\circ$$

is an isomorphism. As discussed in [Shulman 2011, §7], this isomorphism of derived functors is represented by the composite (5.5). In particular, $f^*T^\circ X \to f^*R^sT^\circ X$ is a stable equivalence in this case, since the map f is still the inclusion of a point.

When f is a q-fibration, we instead consider a levelwise qf-fibrant approximation $T^\circ X \to R^l T^\circ X$. There is a stable equivalence $R^l T^\circ X \to R^s T^\circ X$ under $T^\circ X$ [May and Sigurdsson 2006, Lemma 12.6.1] and the induced map $f^*R^lT^\circ X \to f^*R^sT^\circ X$ is a stable equivalence because f^* preserves stable equivalences between levelwise qf-fibrant spectra. Pullback along q-fibrations preserves weak homotopy equivalences of topological spaces, so $f^*T^\circ X \to f^*R^lT^\circ X$ is a levelwise q-equivalence, hence a stable equivalence. Therefore $f^*T^\circ X \to f^*R^sT^\circ X$ is also a stable equivalence. $\square$

We return to using boldface letters to denote derived functors: $\mathbf{T}$ is the left derived functor of T and $\mathbf{F}_b = \mathbb{R}i_b^*$ is the right derived fiber functor. Recall that the $\mathrm{Aut}_R^c M$-torsor associated to an R-bundle with fiber M is given by

$$\mathbf{E} = \mathbf{E}^R(M, -) = \boldsymbol{\mu} \circ \boldsymbol{\Omega},$$

where $\boldsymbol{\mu}$ is the derived functor of Construction 4.3 and $\boldsymbol{\Omega}$ is the right derived functor of $\Omega = \Omega_B^\infty F^R(M, -)$.

Proposition 5.6. *There are natural isomorphisms of derived functors* $\mathbf{F}_b\mathbf{T} \cong \mathbf{T}\mathbf{F}_b$ *and* $\mathbf{F}_b\mathbf{E} \cong \mathbf{E}\mathbf{F}_b$.

Proof. The first isomorphism is Lemma 5.4. For the second, observe that the canonical isomorphism $i_b^*\Omega \cong \Omega i_b^*$ descends to a canonical isomorphism of derived functors $\mathbf{F}_b\boldsymbol{\Omega} \cong \boldsymbol{\Omega}\mathbf{F}_b$ because i_b^* and Ω are both right Quillen. By Lemma 4.6, there is a natural isomorphism $\mathbf{F}_b\boldsymbol{\mu} \cong \boldsymbol{\mu}\mathbf{F}_b$, completing the proof. $\square$

In particular, the derived functor $\mathbf{T}$ takes $\mathrm{Aut}_R^c M$-torsors to R-bundles with fiber M, as promised in Definition 5.1. We are now ready to prove the main theorem of this section.

Theorem 5.7. *The pair of functors* $(\mathbf{T}, \mathbf{E})$ *defines a bijection between the set of q-equivalence classes of* $\mathrm{Aut}_R^c M$*-torsors over B and the set of stable equivalence classes of R-bundles with fiber M over B.*

Proof. We work in the homotopy categories of $\mathrm{Aut}_R^c M$-spaces over B and of R-modules over B. Suppose that Y is an $\mathrm{Aut}_R^c M$-torsor over B. We will construct a natural transformation of derived functors $\zeta : Y \to \mathbf{E}\mathbf{T}Y$ by showing that the unit of the adjunction $(\mathbf{T}, \boldsymbol{\Omega})$ factors through $\mathbf{E}^R(M, \mathbf{T}Y)$ as indicated in the following diagram:

$$\begin{tikzcd}
Y \arrow[r, "\eta"] \arrow[drr, dashed, "\zeta"'] & \Omega_B^\infty \Sigma_B^\infty Y \arrow[r, "\eta"] & \Omega_B^\infty \mathbf{F}^R(M, \mathbf{T}Y) \\
 & & \mathbf{E}^R(M, \mathbf{T}Y) \arrow[u, "\iota"']
\end{tikzcd} \tag{5.8}$$

By Lemma 4.7, it suffices to prove that if we apply $\pi_0\mathbf{F}_b$, then the unit map has image lying in the subset $\pi_0\boldsymbol{\mu}\mathbf{F}_b\boldsymbol{\Omega}\mathbf{T}Y$.

Apply $\mathbf{F}_b$ to diagram (5.8) and commute $\mathbf{F}_b$ past the constituent functors to the input variable Y. Now fix an isomorphism in the derived category $\mathbf{F}_b Y \cong \mathrm{Aut}^c_R M$ and consider the isomorphic diagram with $\mathbf{F}_b Y$ replaced by $\mathrm{Aut}^c_R M$. The composite of the two instances of η in this new diagram is the left vertical composite in the following commutative diagram:

$$\begin{tikzcd}
\mathrm{Aut}^c_R M \arrow[r] \arrow[d] & \mathrm{End}_R M \arrow[d] \\
\Omega^\infty \Sigma^\infty_+ \mathrm{Aut}^c_R M \arrow[r] \arrow[d] & \Omega^\infty \Sigma^\infty_+ \mathrm{End}_R M \arrow[d] \\
\Omega^\infty \mathbf{F}^R(M, M \wedge_{\Sigma^\infty_+ \mathrm{Aut}^c_R M} \Sigma^\infty_+ \mathrm{Aut}^c_R M) \arrow[r] \arrow[dr, "\cong"] & \Omega^\infty \mathbf{F}^R(M, M \wedge_{\Sigma^\infty_+ \mathrm{Aut}^c_R M} \Sigma^\infty_+ \mathrm{End}_R M) \arrow[d] \\
 & \Omega^\infty \mathbf{F}^R(M, M)
\end{tikzcd}$$

Here the horizontal maps are induced by the composite

$$\mathrm{Aut}^c_R M \to \mathrm{Aut}_R M \to \mathrm{End}_R M$$

of the cofibrant approximation map and the canonical inclusion. The diagonal map is induced by the action map for the right $\Sigma^\infty_+ \mathrm{Aut}^c_R M$-module structure on M and it is an isomorphism as indicated. Since M is bifibrant, we may choose to represent the value of the derived functor $\Omega^\infty \mathbf{F}^R(M, M)$ in the homotopy category by $\mathrm{End}_R M$. A diagram chase involving the triangle identities for the adjunctions shows that the right vertical composite is then the identity map. It follows that the left vertical composite factors through $\mathrm{Aut}_R M$ via the cofibrant approximation map, and so the map of components

$$\pi_0 \mathbf{F}_b Y \to \pi_0 \mathbf{F}_b \Omega^\infty_B \mathbf{F}^R(M, \mathbf{T}Y)$$

lands in the subset $\pi_0 \boldsymbol{\mu} \mathbf{F}_b \Omega^\infty_B \mathbf{F}^R(M, \mathbf{T}Y)$. This establishes the factorization in diagram (5.8), and so we have constructed the natural transformation $\zeta : Y \to \mathbf{ET}Y$.

As a consequence of the preceding argument, we see that up to natural isomorphisms in the domain and codomain, $\mathbf{F}_b \zeta$ may be identified with the cofibrant approximation map $\mathrm{Aut}^c_R M \to \mathrm{Aut}_R M$. It follows that ζ is a natural isomorphism of derived functors.

Now let N be an R-bundle with fiber M. Define $\xi : \mathbf{TE}N \to N$ to be the composite

$$\mathbf{TE}^R(M, N) \xrightarrow{\iota} \mathbf{T}\Omega^\infty_B \mathbf{F}^R(M, N) \xrightarrow{\epsilon} N$$

of the map induced by the inclusion $\iota : \mathbf{E}^R(M, N) \to \Omega^\infty_B \mathbf{F}^R(M, N)$ followed by the counit of the adjunction $(\mathbf{T}, \mathbf{\Omega})$. After applying the derived fiber functor $\mathbf{F}_b$, commuting it through to the variable N, and using a chosen equivalence $\mathbf{F}_b N \simeq M$, an argument similar to that just given for ζ proves that $\mathbf{F}_b \xi$ is a fiberwise equivalence. Hence ξ also induces a natural isomorphism of derived functors. □

Proof of Theorem 1.2. We return to the general situation of a well-grounded ring spectrum R and an R-module M. Take an s-cofibrant approximation R' of R as a ring spectrum and an s-bifibrant approximation M' of M as an R-module so that the material in the last two sections applies. The derived mapping space $\mathbf{Aut}_R M$ of homotopy automorphisms of M has a point-set model given by the space $\mathrm{Aut}^c_{R'} M'$. Theorem 2.7 and Theorem 5.7 combine to give that homotopy classes of maps $[X, B\mathbf{Aut}_R M]$ are in bijective correspondence with equivalence classes of R'-bundles with fiber M'. The homotopy category of parametrized R-modules and the homotopy category of parametrized R'-modules are equivalent by pullback along the approximation map [May and Sigurdsson 2006, Proposition 14.1.9], and the definition of an R-bundle with fiber M is invariant under stable equivalences in the entry M, so it follows that equivalence classes of R-bundles with fiber M are in bijective correspondence with equivalence classes of R'-bundles with fiber M'. This completes the proof. □

6. Lifted R-bundles and algebraic K-theory

In this section we will prove Theorem 1.1. The arguments are adapted from [Karoubi 1987; Baas et al. 2004]. Let X be a finite CW complex and let R be a connective ring spectrum. Let

$$\mathrm{GL}_n R = \mathbf{Aut}_R(R^{\vee n})$$

be the derived mapping space of homotopy automorphisms of the n-fold wedge sum $R^{\vee n}$ with the topological monoid structure coming from composition of maps. By Theorem 1.2, the classifying space $B\mathrm{GL}_n R$ classifies stable equivalence classes of R-bundles with fiber $R^{\vee n}$. Let $B\mathrm{GL}_\infty R = \mathrm{hocolim}_n\, B\mathrm{GL}_n R$. Recall the following description of the infinite loop space underlying the algebraic K-theory spectrum of R:

$$\Omega^\infty K(R) \simeq K_0 R \times B\mathrm{GL}_\infty R^+.$$

The group $K_0 R = K_0^f \pi_0 R$ is the algebraic K-theory of free $\pi_0 R$-modules, and the plus denotes Quillen's plus construction with respect to the commutator subgroup of $\pi_1 B\mathrm{GL}_\infty R$. Since the plus construction changes the homotopy type in general, we will need to work with lifted bundles, in the following sense:

Definition 6.1. A lifted R-bundle over X is the data of:

(i) An H_*-acyclic fibration $p : Y \to X$ of CW complexes, by which we mean a Serre fibration with $\widetilde{H}_*(\text{fiber}(p); \mathbf{Z}) = 0$.

(ii) An R-bundle E over Y.

We say that a lifted R-bundle (E, Y, p) over X is free if every fiber of E admits a stable equivalence of R-modules $E_y \simeq R^{\vee n}$ for some n.

Define a relation on lifted R-bundles over X by declaring $(E, Y, p) \sim (E', Y', p')$ if there exists a map $f : Y \to Y'$ over X such that the induced map of R-modules $E \to f^* E'$ over Y is a stable equivalence. This does *not* define an equivalence relation in general, so we work with the equivalence relation on lifted R-bundles over E generated by $\sim$. When convenient, we make the abbreviation $(E, Y) = (E, Y, p)$.

We assume from now on that X is a finite CW complex. Let $\Phi_R(X)$ be the set of equivalence classes of lifted free R-bundles over X. The set $\Phi_R(X)$ is an abelian monoid under the operation $(E_1, Y_1) \oplus (E_2, Y_2)$ taking a pair of lifted R-bundles over X to the lifted R-bundle

$$(g_1^* E_1 \vee_Z g_2^* E_2, Z),$$

where Z is the pullback

$$\begin{array}{ccc} Z & \xrightarrow{g_2} & Y_2 \\ {\scriptstyle g_1}\downarrow & & \downarrow \\ Y_1 & \longrightarrow & X \end{array}$$

The zero of $\Phi_R(X)$ is the trivial R-bundle $(*_X, X)$ over X. Let $\overline{K}_R(X)$ be the Grothendieck group of the monoid $\Phi_R(X)$.

We say that a lifted R-bundle is virtually trivial if there exist a space T such that $\widetilde{H}_*(T; \mathbf{Z}) = 0$, and a map $f : Y \to T$ (not necessarily over X) along with an R-bundle (E', T) over T, and a stable equivalence of R-bundles $E \simeq f^* E'$.

Lemma 6.2. *Let (E_1, Y_1) be a lifted free R-bundle over X. Then there exists a lifted free R-bundle (E_2, Y_2) over X such that $(E_1, Y_1) \oplus (E_2, Y_2)$ is virtually trivial.*

Proof. Let $f_1 : Y_1 \to B\mathrm{GL}_n R$ be a classifying map for E_1. Let P be the homotopy fiber of the H_*-acyclic fibration $Y_1 \to X$. By Proposition 1.3 in [Hausmann and Husemoller 1979], the kernel of $\pi_1 Y_1 \to \pi_1 X$ is the perfect normal subgroup $\mathrm{im}(\pi_1 P \to \pi_1 Y_1)$. This is annihilated by the following map to the plus construction:

$$\pi_1 P \to \pi_1 Y_1 \xrightarrow{f_1} \pi_1 B\mathrm{GL}_n R \to \pi_1 B\mathrm{GL}_n R^+.$$

By Proposition 3.1 in the same paper, f_1 descends to a map $g_1 : X \to B\mathrm{GL}_n R^+$. Use the grouplike H-space structure on $B\mathrm{GL}_\infty R^+$ to find $g_2 : X \to B\mathrm{GL}_m R^+$

such that $g_1 \oplus g_2 : X \to B\mathrm{GL}_{m+n} R^+$ is nullhomotopic. Define Y_2 as the following pullback:

$$\begin{array}{ccc} Y_2 & \xrightarrow{f_2} & B\mathrm{GL}_m R \\ \downarrow & & \downarrow \\ X & \xrightarrow{g_2} & B\mathrm{GL}_m R^+ \end{array}$$

We choose a model for the plus construction such that the right vertical map (and thus the left vertical map) is a q-fibration of CW complexes. Let E_2 be the free R-bundle over Y_2 classified by the map f_2. The sum $(E_1, Y_1) \oplus (E_2, Y_2)$ is a lifted R-bundle over the pullback $Y = Y_1 \times_X Y_2$ that is classified by a lift $f : Y \to B\mathrm{GL}_{m+n} R$ of $g_1 \oplus g_2$. Thus f is nullhomotopic, so it factors through the H_*-acyclic fiber of $B\mathrm{GL}_{m+n} R \to B\mathrm{GL}_{m+n} R^+$, proving that $(E_1, Y_1) \oplus (E_2, Y_2)$ is virtually trivial. $\square$

Given any space X, we generically write $r : X \to *$ for the canonical map to a point, so that the pullback $r^* M$ is the trivially twisted R-bundle with fiber M.

Lemma 6.3. *Suppose that (E, Y) is a virtually trivial lifted R-bundle over X. Then there exists a lifted R-bundle $(r^* M, Y')$ over X that is equivalent to (E, Y) as a lifted R-bundle*: $[(E, Y)] = [(r^* M, Y')]$ *in $\Phi_R(X)$. If E is a free R-bundle, then $M = R^{\vee n}$ for some n.*

Proof. We are given an H_*-acyclic fibration $p : Y \to X$, a map $f : Y \to T$ where $\widetilde{H}_*(T) = 0$ and a stable equivalence $E \simeq f^* E'$ where E' is an R-bundle over T. Choose a point $t : * \to T$. Consider the commutative diagram,

$$\begin{tikzcd} & Y \arrow[dl, "p"'] \arrow[d, "g"] \arrow[dr, "f"] & \\ X & T \times X \arrow[l, "\pi_2"'] \arrow[r, "\pi_1"] & T \\ & X \arrow[ul, "\mathrm{id}"] \arrow[u, "\chi"] \arrow[r, "r"'] & * \arrow[u, "t"'] \end{tikzcd}$$

where $g(y) = (f(y), p(y))$ and $\chi(x) = (t, x)$. The maps p, π_2 and id are all H_*-acyclic fibrations. Form the R-bundle $\pi_1^* E'$ over $T \times X$. Then we have a stable equivalence of R-bundles $g^* \pi_1^* E' = f^* E' \simeq E$ over Y. On the other hand $\chi^* \pi_1^* E' \cong r^* t^* E'$ is a trivial bundle over X with fiber $M = t^* E'$, since $t \circ r$ factors through a point. The two triangles on the left show that $(E, Y) \sim (\pi_1^* E', T \times X)$ and $(\pi_1^* E', T \times X) \sim (r^* M, X)$. $\square$

Consider the abelian group $K_0(R)$ as a discrete set and let $[X, K_0(R)]$ be the set of homotopy classes of maps from X, considered as an abelian group under the pointwise addition in the abelian group $K_0(R)$. Let $\psi : \Phi_R(X) \to [X, K_0(R)]$ be the function that takes a lifted free R-bundle (E, Y, p) to the map sending $x \in X$

to the equivalence class of the free R-module $E_x = (p \circ i)^* E$, where $i : * \to Y$ is a choice of a point lying in the fiber of p over x. Since the fibers of p are path-connected, different choices give the same equivalence class in $K_0(R)$ and it is easy to see that the definition depends only on the equivalence class of the lifted free R-bundle. Since the abelian group structure on $K_0(R)$ is induced by the wedge sum of free R-modules, the function ψ is a monoid homomorphism. Let $\overline{\psi} : \overline{K}_R(X) \to [X, K_0(R)]$ be the extension of ψ to the Grothendieck group. There is a natural splitting

$$\overline{K}_R(X) \cong \ker \overline{\psi} \oplus [X, K_0(R)]$$

induced by the section of ψ that takes an equivalence class $[R^{\vee n}] \in K_0(R)$ indexed by a path-component of X to the trivially twisted R-bundle $r^* R^{\vee n}$ over that component. Let $\Phi^n_R(X)$ be the set of equivalence classes of lifted R-bundles of rank n.

Proposition 6.4. *There is a natural isomorphism*

$$\ker \overline{\psi} \cong \operatorname{colim}_n \Phi^n_R(X).$$

Proof. Suppose $[E] - [F]$ is a formal difference of lifted free R-bundles in $\ker \overline{\psi}$. We associate to $[E] - [F]$ the element $[E \oplus F'] \in \operatorname{colim}_n \Phi^n_R(X)$ where F' is a lifted free R-bundle such that $F \oplus F'$ is virtually trivial (see Lemma 6.2). Conversely, to a class $[E] \in \Phi^n_R(X)$ we associate the formal difference $[E] - [T_n] \in \ker \overline{\psi}$, where $T_n = r^* R^{\vee n}$ is the trivial R-bundle of rank n. □

Proposition 6.5. *There is a natural isomorphism*

$$\operatorname{colim}_n \Phi^n_R(X) \cong [X, B\mathrm{GL}_\infty(R)^+].$$

Proof. Given the class of a lifted free R-bundle (E, Y) over X in $\operatorname{colim}_n \Phi^n_R(X)$, the arguments of Lemma 6.2 show that the classifying map f of E extends to a map g from X to the plus construction:

$$\begin{array}{ccc} Y & \xrightarrow{f} & B\mathrm{GL}_n R \\ {\scriptstyle p}\downarrow & & \downarrow \\ X & \xrightarrow{g} & B\mathrm{GL}_n R^+ \end{array}$$

Conversely, given a classifying map g define Y as the pullback displayed in the same diagram. Then p is an H_*-acyclic fibration and f classifies a lifted free R-bundle (E, Y) over X. □

All together, we have proved:

$$\overline{K}_R(X) \cong [X, K_0(R)] \oplus [X, B\mathrm{GL}_\infty(R)^+] \cong [X, \Omega^\infty K(R)].$$

This completes the proof of Theorem 1.1.

Acknowledgments. I thank Peter May for his support and interest in this project, as well as Matt Ando, Andrew Blumberg and Mike Shulman for stimulating conversations. The advice of an anonymous referee improved the quality of the manuscript, and I thank them for their efforts. This work was partially supported by the Deutsche Forschungsgemeinschaft (DFG) through SFB1085.

References

[Ando et al. 2010] M. Ando, A. J. Blumberg, and D. Gepner, "Twists of K-theory and TMF", pp. 27–63 in *Superstrings, geometry, topology, and C^*-algebras*, edited by R. S. Doran et al., Proc. Sympos. Pure Math. **81**, Amer. Math. Soc., Providence, RI, 2010. MR 2681757 Zbl 1210.81054

[Ando et al. 2011] M. Ando, A. J. Blumberg, and D. Gepner, "Parametrized spectra, multiplicative Thom spectra, and the twisted Umkehr map", preprint, 2011. arXiv 1112.2203

[Ando et al. 2014a] M. Ando, A. J. Blumberg, D. Gepner, M. J. Hopkins, and C. Rezk, "An ∞-categorical approach to R-line bundles, R-module Thom spectra, and twisted R-homology", *J. Topol.* **7**:3 (2014), 869–893. MR 3252967 Zbl 1312.55011

[Ando et al. 2014b] M. Ando, A. J. Blumberg, D. Gepner, M. J. Hopkins, and C. Rezk, "Units of ring spectra, orientations and Thom spectra via rigid infinite loop space theory", *J. Topol.* **7**:4 (2014), 1077–1117. MR 3286898 Zbl 1316.55005

[Baas et al. 2004] N. A. Baas, B. I. Dundas, and J. Rognes, "Two-vector bundles and forms of elliptic cohomology", pp. 18–45 in *Topology, geometry and quantum field theory*, edited by U. Tillmann, London Math. Soc. Lecture Note Ser. **308**, Cambridge Univ. Press, 2004. MR 2079370 Zbl 1106.55004

[Baas et al. 2011] N. A. Baas, B. I. Dundas, B. Richter, and J. Rognes, "Stable bundles over rig categories", *J. Topol.* **4**:3 (2011), 623–640. MR 2832571 Zbl 1227.19002

[Cohen and Jones 2013a] R. L. Cohen and J. D. S. Jones, "Gauge theory and string topology", preprint, 2013. arXiv 1304.0613

[Cohen and Jones 2013b] R. L. Cohen and J. D. S. Jones, "Homotopy automorphisms of R-module bundles, and the K-theory of string topology", preprint, 2013. arXiv 1310.4797

[Hausmann and Husemoller 1979] J.-C. Hausmann and D. Husemoller, "Acyclic maps", *Enseign. Math.* (2) **25**:1–2 (1979), 53–75. MR 543552 Zbl 0412.55008

[Karoubi 1987] M. Karoubi, *Homologie cyclique et K-théorie*, Astérisque **149**, Société Mathématique de France, Paris, 1987. MR 913964 Zbl 0648.18008

[Lind 2013] J. A. Lind, "Diagram spaces, diagram spectra and spectra of units", *Algebr. Geom. Topol.* **13**:4 (2013), 1857–1935. MR 3073903 Zbl 1271.55008

[Lurie 2009] J. Lurie, *Higher topos theory*, Annals of Mathematics Studies **170**, Princeton University Press, 2009. MR 2522659 Zbl 1175.18001

[Mandell et al. 2001] M. A. Mandell, J. P. May, S. Schwede, and B. Shipley, "Model categories of diagram spectra", *Proc. London Math. Soc.* (3) **82**:2 (2001), 441–512. MR 1806878 Zbl 1017.55004

[May 1975] J. P. May, *Classifying spaces and fibrations*, Mem. Amer. Math. Soc. **155**, American Mathematical Society, Providence, RI, 1975. MR 0370579 Zbl 0321.55033

[May and Sigurdsson 2006] J. P. May and J. Sigurdsson, *Parametrized homotopy theory*, Mathematical Surveys and Monographs **132**, American Mathematical Society, Providence, RI, 2006. MR 2271789 Zbl 1119.55001

[Sagave and Schlichtkrull 2013] S. Sagave and C. Schlichtkrull, “Group completion and units in $\mathscr{I}$-spaces”, *Algebr. Geom. Topol.* **13**:2 (2013), 625–686. MR 3044590 Zbl 1277.55004

[Schlichtkrull 2004] C. Schlichtkrull, “Units of ring spectra and their traces in algebraic K-theory”, *Geom. Topol.* **8** (2004), 645–673. MR 2057776 Zbl 1052.19001

[Shulman 2011] M. Shulman, “Comparing composites of left and right derived functors”, *New York J. Math.* **17** (2011), 75–125. MR 2781909 Zbl 1222.55017

Received November 25, 2015. Revised July 14, 2016.

JOHN A. LIND
REED COLLEGE
3203 SE WOODSTOCK BLVD.
PORTLAND, OREGON 97202
U.S.A.

john.alexander.lind@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 285, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.285.453

HIDDEN SYMMETRIES AND COMMENSURABILITY OF 2-BRIDGE LINK COMPLEMENTS

CHRISTIAN MILLICHAP AND WILLIAM WORDEN

In this paper, we show that any nonarithmetic hyperbolic 2-bridge link complement admits no hidden symmetries. As a corollary, we conclude that a hyperbolic 2-bridge link complement cannot irregularly cover a hyperbolic 3-manifold. By combining this corollary with the work of Boileau and Weidmann, we obtain a characterization of 3-manifolds with nontrivial JSJ-decomposition and rank-two fundamental groups. We also show that the only commensurable hyperbolic 2-bridge link complements are the figure-eight knot complement and the 6_2^2 link complement. Our work requires a careful analysis of the tilings of $\mathbb{R}^2$ that come from lifting the canonical triangulations of the cusps of hyperbolic 2-bridge link complements.

1. Introduction

Two manifolds are called *commensurable* if they share a common finite sheeted cover. Here, we focus on hyperbolic 3-manifolds, that is, $M = \mathbb{H}^3/\Gamma$ where Γ is a discrete, torsion-free subgroup of $\mathrm{Isom}(\mathbb{H}^3)$. We are interested in analyzing the set of all manifolds commensurable with M. Commensurability is a property of interest because it provides a method for organizing manifolds, and many topological properties are preserved within a commensurability class. For instance, Schwartz [1995] showed that two cusped hyperbolic 3-manifolds are commensurable if and only if their fundamental groups are quasi-isometric. In this paper, we restrict our attention to hyperbolic 2-bridge link complements; see Section 2 for the definition of a 2-bridge link. We use the word *link* to refer to a link in $\mathbb{S}^3$ with at least one component. We use the word *knot* to only mean a single component link.

A significant challenge in understanding the commensurability class of a hyperbolic 3-manifold $M = \mathbb{H}^3/\Gamma$ is determining whether or not M has any *hidden symmetries*. To understand hidden symmetries, we first need to introduce some terminology. The *commensurator* of Γ is

$$C(\Gamma) = \{g \in \mathrm{Isom}(\mathbb{H}^3) : |\Gamma : \Gamma \cap g\Gamma g^{-1}| < \infty\}.$$

MSC2010: 57M25, 57M50.
Keywords: 2-bridge links, hidden symmetries, commensurability.

It is a well known fact that two hyperbolic 3-manifolds are commensurable if and only if their corresponding commensurators are conjugate in Isom($\mathbb{H}^3$); see Lemma 2.3 of [Walsh 2011]. We denote by $C^+(\Gamma)$ the restriction of $C(\Gamma)$ to orientation-preserving isometries. We also denote by $N(\Gamma)$ the normalizer of Γ in Isom($\mathbb{H}^3$) and by $N^+(\Gamma)$ the restriction of $N(\Gamma)$ to orientation-preserving isometries. Note that $\Gamma \subset N(\Gamma) \subset C(\Gamma)$. A symmetry of M corresponds to an element of $N(\Gamma)/\Gamma$, and a hidden symmetry of M corresponds to an element of $C(\Gamma)$ that is not in $N(\Gamma)$. Geometrically, M admits a hidden symmetry if there exists a symmetry of a finite cover of M that is a not a lift of an isometry of M. See Sections 2 and 3 of [Walsh 2011] for more details on commensurators and hidden symmetries.

In this paper, we give a classification of the hidden symmetries of hyperbolic 2-bridge link complements. Reid and Walsh [2008] used algebraic methods to determine that hyperbolic 2-bridge knot complements (other than the figure-eight knot complement) have no hidden symmetries. However, their techniques do not apply to hyperbolic 2-bridge links with two components. Here, we use a geometric and combinatorial approach to prove the following theorem.

Theorem 1.1. *If $M = \mathbb{S}^3 \setminus K$ is a nonarithmetic hyperbolic 2-bridge link complement, then M admits no hidden symmetries (either orientation-preserving or orientation-reversing).*

The only arithmetic hyperbolic 2-bridge links are the figure-eight knot, the Whitehead link, the 6_2^2 link, and the 6_3^2 link. Though it will not be needed in what follows, we refer the interested reader to [Maclachlan and Reid 2003, Definition 8.2.1] for the definition of an arithmetic group $\Gamma \leq \mathrm{Isom}(\mathbb{H}^3)$.

We prove Theorem 1.1 by using the canonical triangulation $\mathcal{T}$ of a hyperbolic 2-bridge link complement, $M = \mathbb{H}^3/\Gamma = \mathbb{S}^3 \setminus K$. This triangulation was first described in [Sakuma and Weeks 1995]. Guéritaud in his thesis [2006a] proved that this triangulation is geometrically canonical, i.e., topologically dual to the Ford–Voronoi domain for equal volume cusp neighborhoods. In addition, Akiyoshi, Sakuma, Wada and Yamashita [2007] have announced a proof of this result where they analyze the triangulation $\mathcal{T}$ via cone deformations of M along the unknotting tunnel. Futer also showed that this triangulation is geometric by applying Rivin's volume maximization principle; see the appendix of [Guéritaud 2006b]. By [Goodman et al. 2008, Theorem 2.6], if any such M is nonarithmetic, then $C(\Gamma)$ can be identified with the group of symmetries of the tiling of $\mathbb{H}^3$ obtained by lifting $\mathcal{T}$, which we call $\widetilde{\mathcal{T}}$. We prove that any nonarithmetic hyperbolic 2-bridge link complement M does not admit hidden symmetries, by showing that any symmetry of $\widetilde{\mathcal{T}}$ actually corresponds to a composition of symmetries of M and deck transformations of M. In other words, $C(\Gamma) = N(\Gamma)$.

Rather than analyze this tiling of $\mathbb{H}^3$, we drop down a dimension and instead analyze the (canonical) cusp triangulation $\widetilde{T}$ of $\mathbb{R}^2$, induced by $\widetilde{\mathcal{T}}$. By intersecting a cusp cross-section of M with its canonical triangulation $\mathcal{T}$, we obtain a canonical triangulation T of the cusp(s). If K has two components, we still end up with the same canonical triangulation on both components of T since there is always a symmetry exchanging the two components, and we take equal volume cusp neighborhoods. We can lift T to a triangulation $\widetilde{T}$ of $\mathbb{R}^2$ (or two copies of $\mathbb{R}^2$ if K has two components). We also place edge labels on $\widetilde{T}$ which record edge valences of corresponding edges in the three-dimensional triangulation. This labeling provides us with enough rigid structure in $\widetilde{T}$ to rule out any hidden symmetries. Goodman, Heard and Hodgson [2008, Theorem 3.1] use a similar approach to prove that nonarithmetic hyperbolic punctured-torus bundles do not admit hidden symmetries.

If a hyperbolic 3-manifold M admits no hidden symmetries, then M can not irregularly cover any *hyperbolic* 3-*orbifolds*. A hyperbolic 3-orbifold is any $N = \mathbb{H}^3/\Gamma$, where Γ is a discrete subgroup of $\mathrm{Isom}(\mathbb{H}^3)$, possibly with torsion. All of the previous statements about commensurability of hyperbolic 3-manifolds and the commensurator of Γ also hold for hyperbolic 3-orbifolds. Theorem 1.1 quickly gives us the following corollary about coverings of hyperbolic 3-orbifolds by hyperbolic 2-bridge link complements. For the arithmetic cases, volume bounds are taken into consideration to rule out irregular covers of manifolds.

Corollary 1.2. *Let M be any hyperbolic* 2-*bridge link complement. If M is nonarithmetic*, *then M does not irregularly cover any hyperbolic* 3-*orbifolds* (*orientable or nonorientable*). *If M is arithmetic*, *then M does not irregularly cover any orientable hyperbolic* 3-*manifolds.*

By combining Corollary 1.2 with the work of Boileau and Weidmann [2005], we get the following characterization of 3-manifolds with nontrivial JSJ-decomposition and rank-two fundamental groups. For a more detailed description of this decomposition see page 478.

Corollary 1.3. *Let M be a compact*, *orientable*, *irreducible* 3-*manifold which has rank*$(\pi_1(M)) = 2$. *If M has a nontrivial JSJ-decomposition*, *then one of the following holds*:

(1) *M has Heegaard genus* 2.

(2) *M decomposes into a Seifert fibered* 3-*manifold and hyperbolic* 3-*manifold.*

(3) *M decomposes into two Seifert fibered* 3-*manifolds.*

The original characterization given by Boileau and Weidmann included a fourth possibility: a hyperbolic piece of M is irregularly covered by a 2-bridge link complement. Corollary 1.2 eliminates this possibility.

Ruling out hidden symmetries also plays an important role in analyzing the *commensurability class* of a hyperbolic 3-orbifold $M = \mathbb{H}^3/\,\Gamma$. By the commensurability class of a hyperbolic 3-orbifold (or manifold) N, we mean the set of all hyperbolic 3-orbifolds commensurable with N. A fundamental result of Margulis [1991] implies that $C(\Gamma)$ is discrete in $\mathrm{Isom}(\mathbb{H}^3)$ (and Γ is finite index in $C(\Gamma)$) if and only if Γ is nonarithmetic. Thus, in the arithmetic case, M will have infinitely many hidden symmetries. In the nonarithmetic case, this result implies that the hyperbolic 3-orbifold $\mathcal{O}^+ = \mathbb{H}^3/C^+(\Gamma)$ is the unique minimal (orientable) orbifold in the commensurability class of M. So, in the nonarithmetic case, M and M' are commensurable if and only if they cover a common minimal orbifold. Furthermore, when M admits no hidden symmetries, $C^+(\Gamma) = N^+(\Gamma)$, and so, $\mathcal{O}^+$ is just the quotient of M by its orientation-preserving symmetries.

By using Theorem 1.1 and thinking about commensurability in terms of covering a common minimal orbifold, we obtain the following result about commensurability classes of hyperbolic 2-bridge link complements.

Theorem 1.4. *The only pair of commensurable hyperbolic* 2*-bridge link complements are the figure-eight knot complement and the* 6^2_2 *link complement.*

We prove Theorem 1.4 by analyzing the cusp of each minimal (orientable) orbifold, $\mathcal{O}^+$, in the commensurability class of a nonarithmetic hyperbolic 2-bridge link complement. This orbifold always has one cusp since two component 2-bridge links always have a symmetry exchanging the components. The cusp of this orbifold inherits a canonical cellulation from the canonical triangulation T of the cusp(s) of M. By comparing minimal orbifold cusp cellulations, we establish this result.

We now describe the organization of this paper. In Section 2, we provide some background on 2-bridge links, including an algorithm for building any 2-bridge link from a word Ω in Ls and Rs. Section 3 describes how to build the canonical triangulation of a 2-bridge link complement and the corresponding cusp triangulation T based on this word Ω. In this section we also prove some essential combinatorial properties of $\widetilde{T}$, the lift of T to $\mathbb{R}^2$. Section 4 analyzes the possible symmetries of a 2-bridge link complement in terms of the word Ω, and describes the actions of these symmetries on $\widetilde{T}$. In Section 5, we prove Theorem 1.1, Corollary 1.2, and Corollary 1.3. In Section 6, we prove Theorem 1.4.

2. Background on 2-bridge links

In order to describe 2-bridge links, we first need to define *rational tangles*. First, a 2-*tangle* is a pair (B, t), where t is a pair of unoriented arcs embedded in the 3-ball B so that t only intersects the boundary of B in four specified marked points: SW, SE, NW, and NE (if we think of ∂B as the unit sphere centered at the origin in $\mathbb{R}^3$,

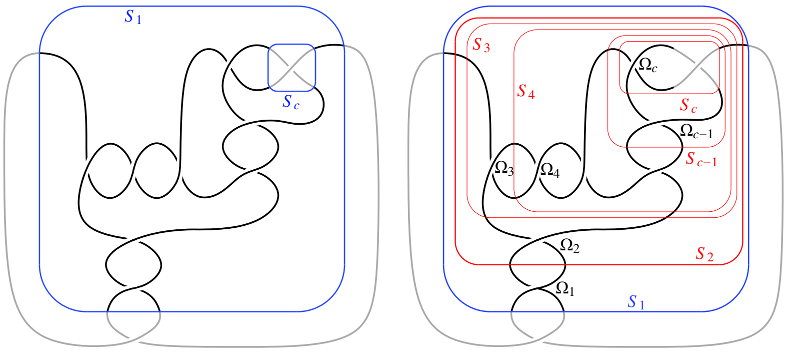


Figure 1. Left: the link $K(\Omega)$, where $\Omega = R^2L^3R^2L$, read from S_1 inward to S_c. Right: the same link, with crossings labeled and 4-punctured spheres S_i shown (note that S_5 and S_6 are omitted for readability).

then SW is the southwest corner $(-1/\sqrt{2}, -1/\sqrt{2}, 0)$, SE is the southeast corner $(1/\sqrt{2}, -1/\sqrt{2}, 0)$, etc.). Rational tangles are a special class of 2-tangles. The simplest rational tangles are the 0-tangle and the ∞-tangle. The 0-tangle consists of two arcs that don't twist about one another, with one arc connecting NW to NE, and the other arc connecting SW to SE. Similarly, the ∞-tangle consists of two unknotted arcs, with one arc connecting NE to SE and the other arc connecting NW to SW. Both of these tangles admit an obvious meridian curve contained on ∂B that bounds an embedded disk in the interior of B. A rational tangle is constructed by taking one of these trivial tangles and alternating between twisting about the western endpoints (NW and SW) and twisting about the southern endpoints (SW and SE). This twisting process maps the meridian of the 0-tangle (∞-tangle) to a closed curve with rational slope p/q, which determines this tangle, hence the name rational tangle. A 2-*bridge link* is constructed by taking a rational tangle, connecting its western endpoints by an unknotted strand, and connecting its eastern endpoints by an unknotted strand.

Here, we describe a 2-bridge link $K \subset \mathbb{S}^3$ in terms of a word Ω, which is a sequence of Ls and Rs: $\Omega = R^{\alpha_1}L^{\alpha_2}R^{\alpha_3}\cdots R^{\alpha_n}$, $\alpha_i \in \mathbb{N}$ (if n is odd and the starting letter is R). The sequence $[\alpha_1+1, \alpha_2, \ldots, \alpha_{n-1}, \alpha_n+1]$ gives the continued fraction expansion for the rational tangle p/q used to construct a 2-bridge link. Each L corresponds to performing a left-handed half-twist about the NW and SW endpoints of a 0-tangle and each R corresponds to performing a right-handed half-twist about the SW and SE endpoints of an ∞-tangle. Each *syllable*, i.e., each maximal subword L^{α_i} or R^{α_i}, corresponds to two strands wrapping around each other α_i times. This word Ω gives a procedure to construct an alternating 4-string braid between two 4-punctured spheres, S_1 and S_c, where S_1 is exterior to the braid and S_c is interior to the braid; see Figure 1. To construct a 2-bridge link, we add a single crossing to the outside of S_1, and we add a single crossing to the inside of S_c. There is a unique way to add these crossings so that the resulting link diagram is alternating. Any 2-bridge link can be constructed in this manner and we use the notation $K(\Omega)$ to designate the 2-bridge link constructed by the word Ω. The original source for this notation comes from the appendix of [Guéritaud 2006b], which contains more details of this construction.

The following are important facts about 2-bridge links that we will use. From now on, we will state results in terms of $K(\Omega)$ and we assume that any 2-bridge link has been constructed in the manner described above, unless otherwise noted.

- Given a 2-bridge link $K(\Omega)$, we obtain a mirror image of the same link (with orientations changed on $\mathbb{S}^3$) if we switch Ls and Rs in the word Ω. Since we will only be considering unoriented link complements, we consider such links equivalent.
- 2-bridge links (and their complements) are determined by the sequence of integers $\alpha_1, \ldots, \alpha_n$ up to inversion. Schubert [1956] gives this classification of 2-bridge knots and links, and Sakuma and Weeks [1995, Theorem II.3.1] give this classification of their complements by examining their (now known) canonical triangulations.
- A 2-bridge link $K(\Omega)$ is hyperbolic if and only if Ω has at least two syllables. This follows from Menasco's [1984] classification of alternating link complements.
- The *only* arithmetic hyperbolic 2-bridge links are those listed below. This classification was given by Gehring, Maclachlan and Martin [1998].
  - The figure-eight knot given by RL or LR,
  - The Whitehead link given by RLR or LRL,
  - The 6_2^2 link given by L^2R^2 or R^2L^2, and
  - The 6_3^2 link given by RL^2R or LR^2L.

  We care about distinguishing between nonarithmetic and arithmetic hyperbolic link complements because different techniques have to be used for analyzing hidden symmetries and commensurability classes.

Throughout this paper, we will always assume that $K(\Omega)$ is hyperbolic, i.e., Ω has at least two syllables. In Section 3, we will use the diagram of $K(\Omega)$ described above to build the canonical cusp triangulation of $\mathbb{S}^3 \setminus K(\Omega)$.

3. Cusp triangulations of 2-bridge link complements

Let $K = K(\Omega)$ be a 2-bridge link, defined as in Section 2, with Ω a word in R and L, and Ω_i its i^{th} letter. We may assume that $\Omega_1 = R$, as mentioned in Section 2. In this section we give a description of the construction of the triangulation $\mathcal{T}$

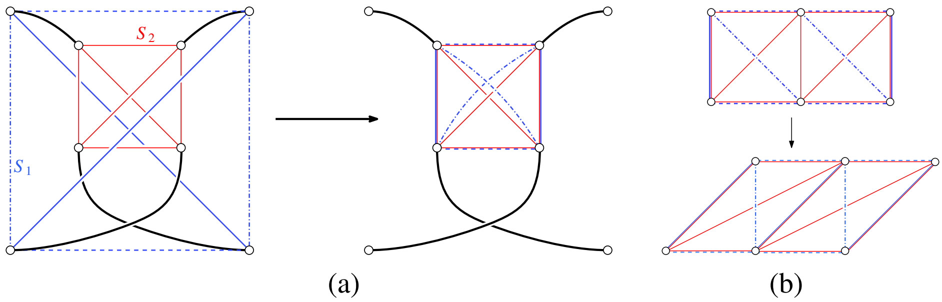


Figure 2. On the left (a), we see which edges of S_1 are identified to edges of S_2, and what the region between S_1 and S_2 looks like. In the right figure (b), it is a little easier to see, with S_1 and S_2 unfolded, that the region between them is a pair of tetrahedra.

of $\mathbb{S}^3 \setminus K$, and of the induced cusp triangulation T, and its lift $\widetilde{T}$ (if K has two components, then the two cusp triangulations are identical). We then describe an algorithmic approach for constructing $\widetilde{T}$, and prove some facts about simplicial homeomorphisms $f : \widetilde{T} \to \widetilde{T}$. Our description of these triangulations follows that of [Guéritaud 2006b, Appendix A] and [Sakuma and Weeks 1995, Chapter II], to which we refer the reader for further details.

To build the triangulation $\mathcal{T}$, we first place a 4-punctured sphere S_i at each crossing Ω_i corresponding to a letter of Ω, so that every crossing Ω_j for $j \geq i$ is on one side of S_i, and the remaining crossings are on the other side; see Figure 1 (right). We will start by focusing on S_1 and S_2. We triangulate both of them as shown in the first frame of Figure 2(a) (notice that the edge from the lower-left to upper-right puncture is in front for both). If we push S_1 along the link to the other side of the crossing Ω_1, we see that some of its edges coincide with edges of S_2 (in particular, the horizontal edges coincide, and the diagonal edges of S_1 become vertical in S_2, see Figure 2(a)). The vertical edges of S_1, however, get pushed to diagonal edges that cannot be identified to the diagonal edges of S_2. The top frame of Figure 2(b) shows S_1 and S_2 with appropriate edges identified, as seen lifted to $\mathbb{R}^2 \setminus \mathbb{Z}^2$ (i.e., cut along top, bottom, and left edges then unfold). If we lift S_1 to $\mathbb{R}^2 \setminus \mathbb{Z}^2$ in such a way that its triangulation has edge slopes $\frac{0}{1}, \frac{1}{1}, \frac{1}{0}$, this choice forces S_2 to have edge slopes $\frac{0}{1}$, $\frac{1}{2}$, $\frac{1}{1}$, as shown in the lower frame of Figure 2(b). This means that the triangulation of S_2 in $\mathbb{R}^2 \setminus \mathbb{Z}^2$ is obtained by applying the matrix $R = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ to the S_1 triangulation of $\mathbb{R}^2 \setminus \mathbb{Z}^2$. If the letter Ω_1 between S_1 and S_2 had been an L, we would have found by the same analysis that the matrix taking us from the triangulation of S_1 to the triangulation of S_2 must be $L = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$. This holds in general. If we know the edge slopes of the triangulation of S_i, we can apply the appropriate matrix, depending on whether Ω_i is an R or an L, to get the triangulation of S_{i+1} (see Figure 3).

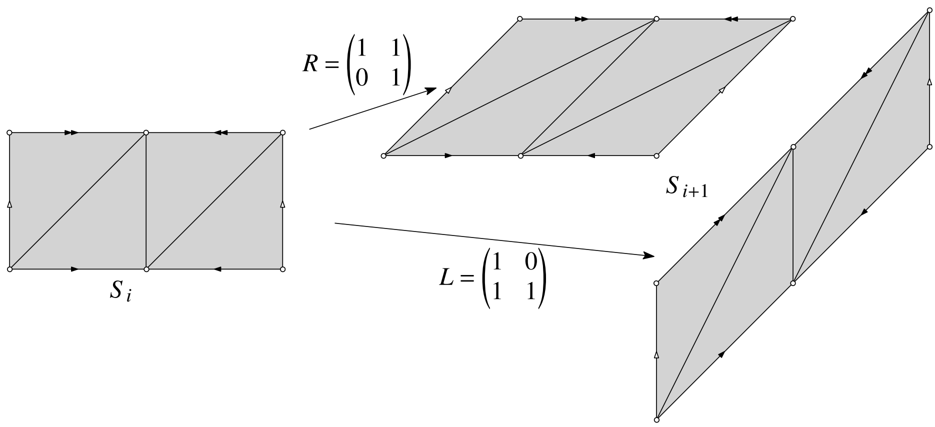


Figure 3. We apply the transformations R or L as shown, depending on whether Ω_i is an R or an L, to obtain S_{i+1} from S_i.

Remark 3.1. Though we do not use this fact in what follows, the word Ω can be viewed as a path in the Farey tessellation, with each letter corresponding to making a right (for R) or left (for L) turn from one Farey triangle to the next. In this case each 4-punctured sphere S_i corresponds to a Farey triangle, and its slopes are given by the vertices of that triangle. For details of this approach, we again direct the interested reader to [Guéritaud 2006b] and [Sakuma and Weeks 1995].

Coming back to S_1 and S_2, we see in Figure 2(b) that between the (red) triangulation of S_2 and the (blue) triangulation of S_1 is a layer of two tetrahedra, which we denote Δ_1. Similarly, between the 4-punctured spheres S_i and S_{i+1} we get a layer Δ_i of tetrahedra. This construction results in a "product region" $S \times I$, where $S \times \{0\} = S_1$ and $S \times \{1\} = S_c$. We use quotation marks here because $S \times I$ is not a true product for $\Omega \in \{RL^k, LR^k, RL^kR, LR^kL\}$, since there will be an edge shared by all the S_i.

To obtain $\mathbb{S}^3 \setminus K$ from $S \times I$, we first "clasp" S_1 by folding along edges with slope $\frac{1}{1}$ and identifying pairs of triangles adjacent to those edges, as shown in Figure 4. We clasp S_c in the same way, this time folding along either the edge with greatest slope or the edge with least slope, depending on whether the final letter of Ω is R or L, respectively.

To understand the induced triangulation T of a cusp cross section, we first consider a neighborhood of a single puncture P in $S \times I$. For each layer of tetrahedra Δ_i between S_i and S_{i+1}, we get a pair of triangles D_i and D_i' going once around the puncture, as in Figure 5(a). In this figure vertices of $D_i \cup D_i'$ are labeled according to the edges of Δ_i that they are contained in, and edges of $D_i \cup D_i'$ are labeled according to the edge of Δ_i that they are across a face from. Notice in Figure 5(a) that D_i has a vertex (c_-) meeting an edge of S_i but not meeting S_{i+1},

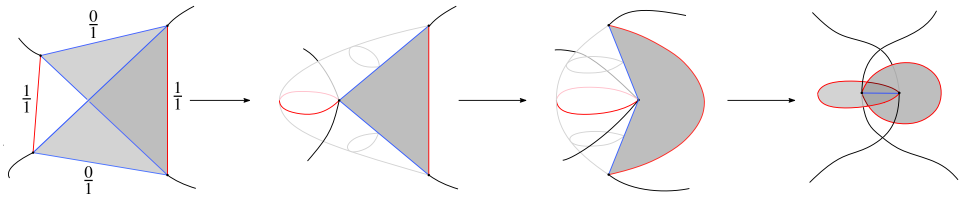


Figure 4. The clasping of S_1. The viewpoint of the reader is the "inside" of S_1, i.e., the side containing the braid in Figure 1.

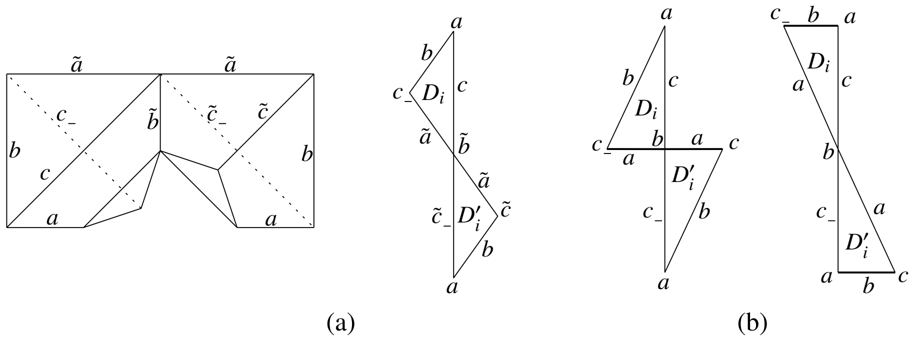


Figure 5. In (a), a layer Δ_i with a neighborhood of a cusp removed (left), and the triangles $D_i \cup D_i'$ that the layer Δ_i contributes to the cusp triangulation (right). Edges with the same slope have labels that differ by a $\sim$ decoration. Figure (b) shows D_i and D_i' after being adjusted as prescribed in Figure 6, with $\sim$ decorations removed so that edges with the same slope are labeled the same.

and D_i' has a vertex ($\tilde{c}$) meeting S_{i+1} but not meeting S_i. Thus D_i is distinguished from D_i'.

To see how $D_i \cup D_i'$ attaches to $D_{i-1} \cup D_{i-1}'$, we must consider how Δ_i attaches to Δ_{i-1}. Figure 6(a) shows Δ_i and Δ_{i-1} in $(\mathbb{R}^2 \setminus \mathbb{Z}^2) \times I$ (sandwiched between $S_{i-1} \cup S_i \cup S_{i+1}$) in the case where $\Omega_i = R$, and the corresponding triangles around the puncture. There is a unique edge e of $D_i \cup D_i'$, corresponding to an edge of S_i shared by both S_{i-1} and S_{i+1}, and with vertices $v_1 \in S_{i-1}$ and $v_2 \in S_{i+1}$. This means that the edge e moves us along the cusp cross-section in the longitudinal direction, so it will be part of a longitude in $\widetilde{T}$. It makes sense then to adjust these edges to be horizontal, as we build the triangulation $\widetilde{T}$ (see Figure 6(a)). Figure 6(b) shows the analogous adjustment when $\Omega_i = L$.

When we clasp S_1, an edge of D_1 is identified to an edge of D_1', and similarly for D_c' and D_c when S_c is clasped, as illustrated in Figure 7. We will call the triangles D_1 and D_c' *clasping triangles*. For $\Omega = R^2L^3R^2L$, the triangulation around a puncture before clasping and after clasping is shown in Figures 7(b) and 7(c), respectively.

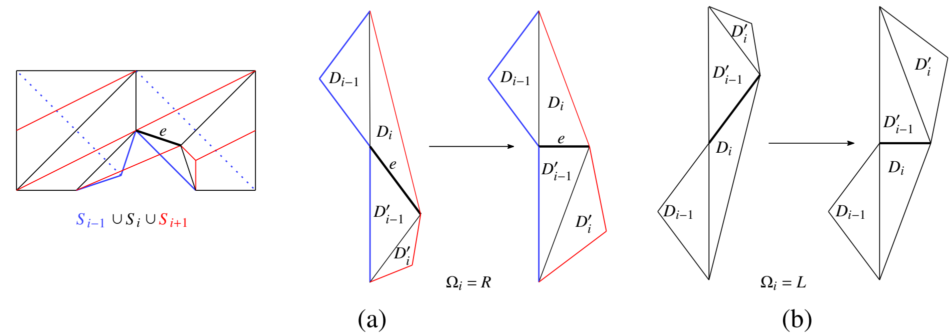


Figure 6. Building the cusp triangulation. In (a), the left frame shows three layers of 4-punctured spheres, with a truncated puncture. Note the special edge e on the truncated puncture, also shown in the right frame, which connects S_{i-1} to S_{i+1}. Note that in the two figures on the right, the top and bottom vertices are identified, and in (b) we have rotated (vertically) by π to make the picture more clear.

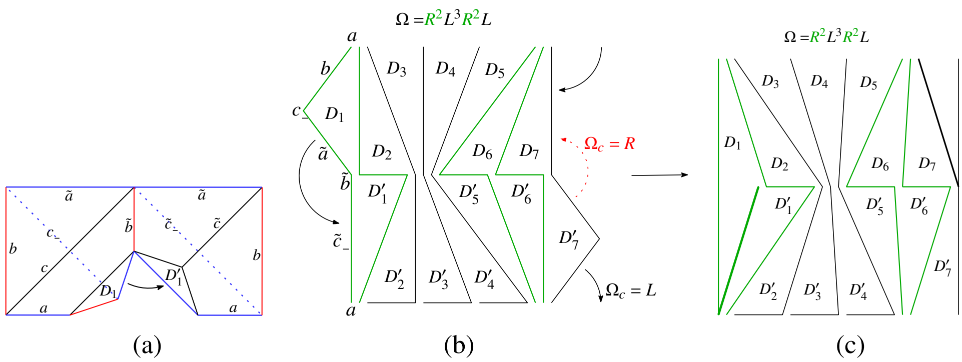


Figure 7. The effect of clasping on the triangulation around a puncture. (a) shows Δ_1, with S_1 below S_2, and edge colors of S_1 corresponding to colors in Figure 4. On the right, (b) and (c) show the effect of clasping as seen from the cusp cross-section.

Before clasping, it is clear from the construction that the combinatorics around each of the four punctures is identical. Clasping identifies the punctures on $S_1 = S \times \{0\}$ in pairs, and identifies the punctures on $S_c = S \times \{1\}$ in pairs, in an orientation-preserving way. This means that for a 2-component link, a cusp triangulation is obtained by gluing two puncture triangulations (as in Figure 7(c)) along their front edges, and along their back edges, in an orientation-preserving way. For a knot, the situation is similar, except that we glue all four puncture triangulations, always identifying front edges to front edges, and back to back, with orientation preserved. In both cases the lifted triangulation $\widetilde{T}$ of $\mathbb{R}^2$ is the same,

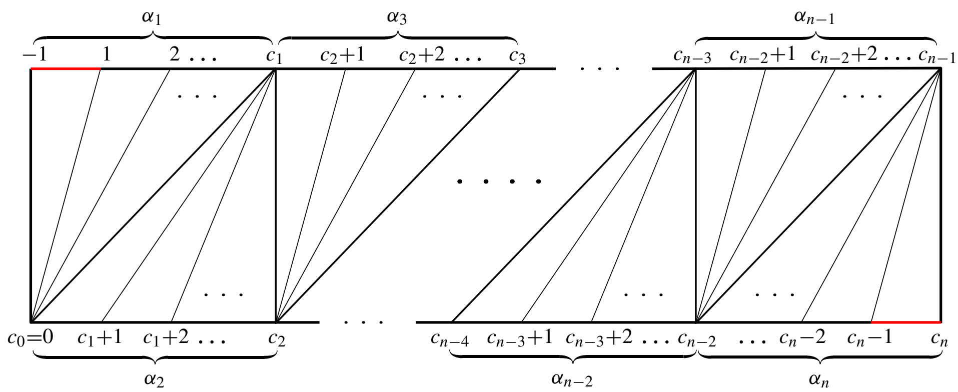


Figure 8. Triangulation of $D' = [0, 1] \times [0, 1] \subset \mathbb{R}^2$. The word $\Omega = R^{\alpha_1} L^{\alpha_2} \cdots L^{\alpha_n}$ can be read from the triangulation. Here, $c_j = \sum_{i=1}^{j} \alpha_i$.

except that the fundamental region for a knot is twice as large as for a link. Note that when $\Omega_1 \neq \Omega_c$, the clasping triangle on the right is offset vertically from the clasping triangle on the left (as in Figure 7(c)), whereas if $\Omega_1 = \Omega_c$ this will not be the case.

As a result of the above discussion, we can now give an algorithmic approach to constructing the lifted cusp triangulation $\widetilde{T}$ for an arbitrary word $\Omega = R^{\alpha_1} L^{\alpha_2} \cdots L^{\alpha_n}$ (we will assume the last letter is L for concreteness; the case where $\Omega_c = R$ is similar). This follows the approach in [Sakuma and Weeks 1995, Section II.4], with some changes of notation. We start with a rectangle $D' = [0, 1] \times [0, 1] \subset \mathbb{R}^2$ divided into $c = \sum_i \alpha_i$ triangles, each corresponding to a letter of Ω, as in Figure 8. Vertices of D' are labeled as shown, with $c_j = \sum_{i=1}^{j} \alpha_i$ for $1 \leq j \leq n$, and $c_0 = 0$. To fill out $\mathbb{R}^2$ we first reflect D' in its top edge to get its mirror D, so that $D \cup D'$ is a triangulation of a puncture (with triangles D_i in D and triangles D'_i in D'), as in Figure 7(c). We then rotate $D \cup D'$ by π about $(0, 1)$ (i.e., about the vertex labeled -1), and translate the resulting double of $D \cup D'$ vertically and horizontally to fill $\mathbb{R}^2$. Finally, we remove all edges $\overline{-1, 1}$ and $\overline{r, c_n}$, where $r = c_{n-2}$ if $\alpha_n = 1$, and $r = c_n - 1$ otherwise (i.e., all images of the red edges in Figure 8).

With this parametrization of the cusp triangulation in $\mathbb{R}^2$, deck transformations are generated by $(x, y) \mapsto (x, y + 2)$ and $(x, y) \mapsto (x + k, y)$, where $k = 2$ if $K = K(\Omega)$ has two components, and $k = 4$ if it has one component. We observe that the long edge of each clasping triangle goes all the way around the meridian of the cusp, and these edges are unique in this respect. For this reason we call these edges *meridional edges* (whether we are referring to them in T or $\widetilde{T}$), and we call each connected component of their union in $\widetilde{T}$ a *meridional line* (i.e., any line $x = c$, $c \in \mathbb{Z}$). A strip of adjacent nonclasping triangles that all meet the lines

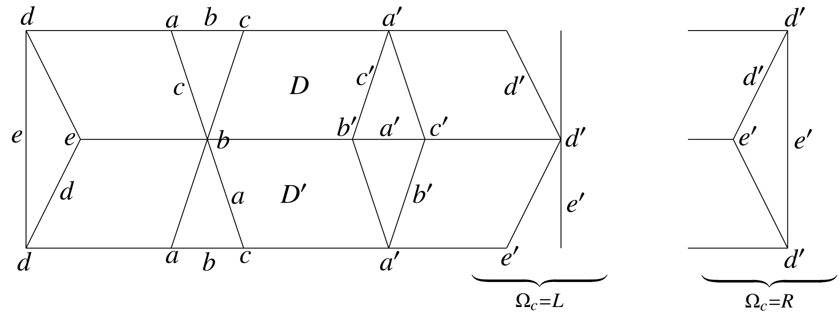


Figure 9. Edge/vertex correspondence in $\widetilde{T}$. Vertices and edges with the same slope (as edges in $\widetilde{\mathcal{T}}$), are labeled the same.

$y = m$ and $y = m+1$ (in an edge or vertex), for some $m \in \mathbb{Z}$, is called a *horizontal strip* (see Figure 12).

We will now describe a correspondence between edges and vertices of $\widetilde{T}$. Given an edge e in $\widetilde{T}$, meaning a truncated tip of an ideal triangle in $\widetilde{\mathcal{T}}$, we have a corresponding edge in $\widetilde{\mathcal{T}}$: this is just the edge of $\widetilde{\mathcal{T}}$ across from e in the ideal triangle, as in Figure 5(a). Similarly, a vertex of $\widetilde{T}$ corresponds to the edge in $\widetilde{\mathcal{T}}$ that it is contained in. We say that an edge e and a vertex v of $\widetilde{T}$ *correspond* if their corresponding edges in $\widetilde{\mathcal{T}}$ have the same slope (when viewed in $(\mathbb{R}^2 \setminus \mathbb{Z}^2) \times I$). Edge and vertex correspondence in $\widetilde{T}$, for edges and vertices that do not come from S_1 or S_c, can be read off Figure 5(b), which shows the cusp cross-section of a layer Δ_i with vertices and edges of the same slope labeled the same.

As for edges and vertices affected by clasping, we can easily read the correspondences off the labellings in Figure 7 for the clasping of S_1, and the S_c clasping works similarly. This gives edge/vertex correspondences for $D \cup D'$, as shown in Figure 9 (as usual, we assume $\Omega_1 = R$). A fundamental region of T is constructed by gluing together either two or four copies of $D \cup D'$ by orientation-reversing homeomorphisms $\{0\} \times [0, 1] \to \{0\} \times [0, 1]$ and $\{1\} \times [0, 1] \to \{1\} \times [0, 1]$, as previously discussed. Hence, the algorithmic construction of $\widetilde{T}$ by rotating $D \cup D'$ by π about $(0, 1)$ then translating to tile the plane respects edge valence, and so edge/vertex correspondence for all of $\widetilde{T}$ can be obtained in this way. From here forward we will consider the edges of $\widetilde{T}$ to be labeled by the valence of a corresponding vertex, and we will refer to this number as the *edge valence*.

We summarize the preceding discussion in the following lemma, part (d) of which corrects a minor error in the proof of Theorem II.3.1 in [Sakuma and Weeks 1995] (this error does not, however, affect the validity of their proof). Note that the relevant notation in [Sakuma and Weeks 1995] differs from ours in several ways: most importantly, what we call $val(i)$ they denote $d(i)$, and we follow a different indexing convention for vertices of $\widetilde{T}$.

Lemma 3.2. *The lifted cusp triangulation $\widetilde{T}$ for the link given by a word $\Omega = R^{\alpha_1} L^{\alpha_2} R^{\alpha_3} \cdots L^{\alpha_n}$ has the following description*:

(a) *$\widetilde{T}$ is obtained from the triangulated rectangle $D' = [0, 1] \times [0, 1] \subset \mathbb{R}^2$, described by* Figure 8, *as follows*: *reflect in* $[0, 1] \times \{1\}$ *to get D, then rotate $D \cup D'$ about* $(0, 1)$, *and translate the resulting two copies of $D \cup D'$ by $(x, y) \mapsto (x + 2k, y + 2m)$, where $k, m \in \mathbb{Z}$, to tile $\mathbb{R}^2$.*

(b) *The deck group of $\widetilde{T}$ is generated by* $(x, y) \mapsto (x, y+2)$ *and* $(x, y) \mapsto \left(x + \frac{4}{\epsilon}, y\right)$, *where $\epsilon \in \{1, 2\}$ is the number of components of the link $K(\Omega)$.*

(c) *Edge/vertex correspondence in $\widetilde{T}$ is as follows* (*see* Figure 9):

- *If e is horizontal or e is a meridional edge, then e corresponds to the vertices across the two triangles adjacent to it.*
- *If the lower endpoint of e meets the line $y = k$, and the upper endpoint meets $y = k + 1$, with k even* (*resp. odd*), *then e corresponds to the vertex across the triangle to the left* (*resp. right*) *of e.*

(d) *If $\Omega \notin \{R^2L^2, RL^m, RL^mR : m \geq 1\}$, then the vertices of $\widetilde{T}$, labeled as in* Figure 8, *have valence as follows* (*recall that $r = c_{n-2}$ if $\alpha_n = 1$, and $r = c_n - 1$ otherwise*):

$$val(c_i) = \begin{cases} 4\alpha_{i+1} + 4 & \text{for } i \in \{0, n-1\}, \\ 2\alpha_{i+1} + 4 & \text{for } 2 \leq i \leq n-3 \text{ or } i = 1, \alpha_1 > 1 \text{ or } i = n-2, \alpha_n > 1, \\ 2\alpha_{i+1} + 3 & \text{for } i = 1, \alpha_1 = 1 \text{ or } i = n-2, \alpha_n = 1, \end{cases}$$

$$val(1) = \begin{cases} 3 & \text{for } \alpha_1 > 1, \\ 2\alpha_2 + 3 & \text{for } \alpha_1 = 1, \end{cases}$$

$$val(r) = \begin{cases} 3 & \text{for } \alpha_n > 1, \\ 2\alpha_{n-1} + 3 & \text{for } \alpha_n = 1, \end{cases}$$

$$val(j) = 4 \text{ for } j \notin \{0, 1, c_1, c_2, \ldots, c_n, r\}.$$

In particular, note that for all $\Omega \notin \{R^2L^2, RL^m, RL^mR : m \geq 1\}$, $val(j)$ is odd if and only if $j \in \{1, r\}$. This fact is key to showing that nonarithmetic 2-bridge links cannot have hidden symmetries. Since a hidden symmetry restricts to an isometry of $\widetilde{\mathcal{T}}$, it is a *simplicial automorphism* of $\widetilde{\mathcal{T}}$ (i.e., a homeomorphism $\widetilde{\mathcal{T}} \to \widetilde{\mathcal{T}}$ preserving the simplicial structure) and hence it is a simplicial automorphism of $\widetilde{T}$ that preserves edge valence.

Definition 3.3. We denote by $\mathrm{Aut}_{ev}(\widetilde{T})$ the *group of simplicial automorphisms of $\widetilde{T}$ that preserve edge valence*. Note that if we identify $\widetilde{T}$ with the horoball centered at p, then there is a natural injection $\mathrm{Stab}_{\mathrm{Aut}(\widetilde{\mathcal{T}})}(p) \hookrightarrow \mathrm{Aut}_{ev}(\widetilde{T})$.

By analyzing $\mathrm{Aut}_{ev}(\widetilde{T})$, which must preserve these odd valence vertices, we learn about the possible isometries of $\widetilde{\mathcal{T}}$. The first step in this process is the following lemma:

Lemma 3.4. *If* $\Omega \notin \{RL,\ R^2L^2,\ RLR\}$, *then* $\mathrm{Aut}_{ev}(\widetilde{T})$ *preserves clasping triangles and meridional edges.*

Proof. By the symmetry of the problem, we need only show that any triangle $\triangle_{1,0,0}$ with vertex labels $\{1, 0, 0\}$ maps to a clasping triangle. Let $f \in \mathrm{Aut}_{ev}(\widetilde{T})$, and let $\triangle_{a,b,b'}$ be the image of a triangle $\triangle_{1,0,0}$ under f, so that $1 \mapsto a$.

Case 1: $\Omega \notin \{R^kL^m,\ RL^mR^k\}$. Since $val(j)$ is odd if and only if $j \in \{1, r\}$, we must have $a \in \{1, r\}$. We will assume that $a = 1$; the case $a = r$ is proved similarly. Then $b \in \{0, c_1, c_2, c_3\}$ since $val(0) = 4\alpha_1 + 4 \geq 8$ and all other vertices that could share an edge with 1 have valence 4.

If $val(1) = 3$ (i.e., $\alpha_1 > 1$), then $b \in \{0, c_1\}$, since in this case no vertex c_2 or c_3 is connected to 1 by an edge. If $b = c_1$, then we must have $\alpha_1 = 2$, so that $val(0) = 4\alpha_1 + 4 = 12 = val(c_1) = 2\alpha_2 + 4 \Longrightarrow \alpha_2 = 4$, which means that $c_1 + 1$ must have valence 4. But $b = c_1$ also implies that $c_1 + 1$ is the image of the valence 3 vertex of the clasping triangle that shares a meridional edge with $\triangle_{1,0,0}$, giving a contradiction. Thus $b = 0$, and by the same argument we must also have $b' = 0$.

If $val(1) \neq 3$, then $\alpha_1 = 1$ and $val(1) = val(c_1) = 2\alpha_2 + 3$, and we must have $b \in \{0, c_2, c_3\}$. Also, $val(0) = 4\alpha_1 + 4 = 8$.

If $b = c_2$, then $2\alpha_3 + 4 = val(c_2) = val(0) = 8$, so $\alpha_3 = 2$. This implies that $val(c_2 + 1) = 4 \neq 8$, so we must have $\triangle_{1,0,0} \mapsto \triangle_{1,c_2,0}$. This determines the image of the two nonclasping triangles $\triangle_{0,c_1,c_2}$ adjacent to $\triangle_{1,0,0}$, and we see that the c_2 vertex of one of these must be mapped to a $c_2 + 1$ vertex, which is impossible since $val(c_2 + 1) = 4 \neq 8 = val(c_2)$.

If $b = c_3$ then $1 = c_1$ and c_3 are connected by an edge, so $\alpha_3 = 1$, which forces the other 0-labeled vertex of $\triangle_{1,0,0}$ to map to c_2, which is impossible by the above argument. Hence $b = 0$, and by the same argument we have $b' = 0$.

Since $\Omega \notin \{R^kL^m,\ RL^mR^k\}$ implies that clasping triangles have a unique odd valence vertex (i.e., the vertex not meeting a meridional edge), that meridional edges map to meridional edges is immediate.

Case 2: $\Omega = R^kL^m$ and $\Omega \notin \{RL,\ R^2L^2\}$. If $k = 1$, then clasping triangles either have vertices with valences $8, 8, 4m + 2$ or $3, 4m + 2, 4m + 2$, and they are the only triangles in $\widetilde{T}$ with such a triple of valences. If $k \neq 1$ then clasping triangles either have vertices with valences $3, 4k + 4, 4k + 4$ or $3, 4m + 4, 4m + 4$, and they are the only triangles in $\widetilde{T}$ with such a triple of valences. Furthermore, in every case two of these vertices have equal valence and the third has distinct valence, so meridional edges must be preserved.

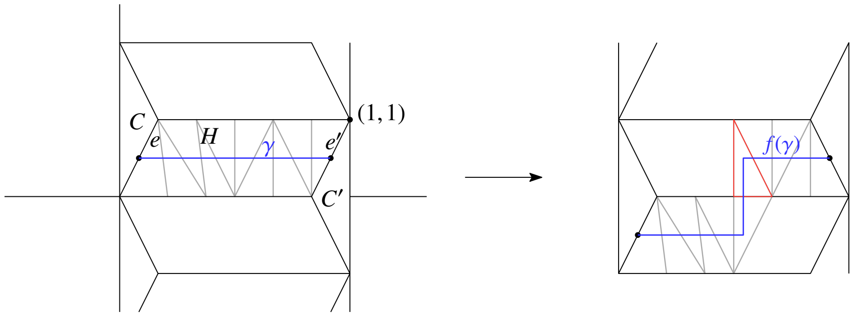


Figure 10. If H maps into more than one horizontal strip, then $f(\gamma)$ traverses more than $c-2$ triangles, which is impossible.

Case 3: $\Omega = RL^m R^k$, $\Omega \neq RLR$. Then $\alpha_1 = 1 \Longrightarrow val(0) = 8$. If $k > 1$ then $val(1) \neq val(r)$, so $1 \mapsto 1$ and we must have vertices labeled 0 mapping to vertices labeled 0 or $c_2 = c_{n-1}$. But $val(c_{n-1}) = 4k+4 \neq 8$, so $0 \mapsto 0$. If $k = 1$ then clasping triangles all have vertices with valences 8, 8, $2m+2$, and they are the only triangles in $\widetilde{T}$ with this triple of valences. Furthermore, meridional edges are preserved since even when $m = 3$ (so that $2m+2 = 8$), the vertices labeled $1 = r$ are combinatorially distinct from the vertices labeled 0 and c_{n-1}: vertices labeled 1 have four edges connecting them to valence 4 vertices, while vertices labeled 0 and c_{n-1} have only two such edges. □

Corollary 3.5. *If* $\Omega \notin \{RL, R^2L^2, RLR\}$, *then* $\mathrm{Aut}_{ev}(\widetilde{T})$ *preserves horizontal strips of* $\widetilde{T}$.

Proof. Let C be the clasping triangle in the first quadrant of $\mathbb{R}^2$ with a vertex at the origin. C is adjacent to two horizontal strips; let H be the one adjacent to the x-axis, and let C' be the other clasping triangle adjacent to H. Let γ be the path directly across H connecting the midpoints of the edges of adjacency with C and C'. Consider the image of γ under a simplicial automorphism $f : \widetilde{T} \to \widetilde{T}$. Since γ crosses exactly $c-2$ triangles, so must $f(\gamma)$. By Lemma 3.4, f maps e and e' to edges of clasping triangles, which are adjacent to distinct meridional lines since C and C' are, and f maps triangles crossed by γ to nonclasping triangles, so γ must be mapped into some number of vertically stacked horizontal strips. Since γ crosses all triangles transversely, if $f(\gamma)$ jumps from one horizontal strip to another the number of triangles it crosses must be one more that if it did not make the jump, as shown in Figure 10. Hence $f(\gamma)$ must be contained in one horizontal strip, the image of H. □

Recall that in our algorithmic construction of $\widetilde{T}$, we chose coordinates so that the rectangle D' shown in Figure 8 is identified with $[0, 1] \times [0, 1] \subset \mathbb{R}^2$.

Theorem 3.6. *If $\Omega \notin \{RL, R^2L^2, RLR\}$, then $\mathrm{Aut}_{ev}(\widetilde{T})$ is generated by the deck transformations and a subset of the following*:

- *Orientation-preserving: the rotations ρ_1, ρ_2, and ρ_3 about $(1,1)$, $(2,1)$, and $\left(\frac{1}{2},1\right)$, respectively, by an angle π.*
- *Orientation-reversing: the glide reflection g given by the reflection across $x=\frac{1}{2}$ composed with $(x, y) \mapsto (x, y+1)$.*

Further, we always have $\rho_1, \rho_2 \in \mathrm{Aut}^+_{ev}(\widetilde{T})$, and $\rho_3 \in \mathrm{Aut}^+_{ev}(\widetilde{T})$ (resp. $g \in \mathrm{Aut}_{ev}(\widetilde{T})$) if and only if ρ_3 (resp. g) is a simplicial automorphism.

Proof. Let $f \in \mathrm{Aut}_{ev}(\widetilde{T})$, and let E be the union of all edges of horizontal strips and clasping triangles, as shown in Figure 12. Since f maps clasping triangles to clasping triangles, and horizontal strips to horizontal strips, it must map E to itself. Since the simplicial structure of the triangulation within each horizontal strip must be preserved, and since we may assume all clasping triangles are congruent and triangles within each strip are uniformly sized, f is forced to be a Euclidean isometry of $\mathbb{R}^2$. Let ρ_4 be the rotation by π about the point $\left(\frac{1}{2}, \frac{1}{2}\right)$, and let r_y be the reflection about the line $y=1$. We first consider the possible Euclidean isometries preserving E:

Translations: Translations must preserve the integer lattice, so modulo deck transformations they have the form $\tau_{i,j} : (x, y) \mapsto (x+i, y+j)$, $i \in \{0, 1, 2, 3\}$, $j \in \{0, 1\}$. Since $\tau_{0,1}$, and $\tau_{2,1}$ do not preserve E, and $\tau_{0,0}$ is trivial, we are left with

$$\tau_{1,0} = \rho_1 \circ \rho_3; \quad \tau_{2,0} = \rho_2 \circ \rho_1; \quad \tau_{3,0} = \rho_2 \circ \rho_3; \quad \tau_{1,1} = \rho_1 \circ \rho_4; \quad \tau_{3,1} = \rho_2 \circ \rho_4,$$

and their inverses.

Rotations: Since meridional lines and integer lattice points must be preserved, any rotation must be by an angle π about a point $\left(\frac{k}{2}, \frac{m}{2}\right)$, $k, m \in \mathbb{Z}$. The rotations about $\left(1, \frac{1}{2}\right)$ and $\left(2, \frac{1}{2}\right)$ do not preserve clasping triangles, so modulo deck transformations we are left with ρ_1, ρ_2, ρ_3, ρ_4, and the rotations

$$\rho_4 \circ \rho_2 \circ \rho_1; \quad \rho_3 \circ \rho_2 \circ \rho_1,$$

about $\left(\frac{3}{2}, \frac{1}{2}\right)$ and $\left(\frac{3}{2}, 1\right)$, respectively.

Reflections: Reflections must preserve meridional lines and clasping triangles, so possible lines of reflection are $x = \frac{k}{2}$ or $y = k$, $k \in \mathbb{Z}$. Modulo deck transformations, we get the reflection r_y across $y = 1$, and the reflections r_i across the lines $x = i$, $i \in \left\{\frac{1}{2}, 1, \frac{3}{2}, 2\right\}$. We have

$$r_1 = r_y \circ \rho_1; \quad r_2 = r_y \circ \rho_2; \quad r_{\frac{1}{2}} = r_y \circ \rho_3; \quad r_{\frac{3}{2}} = r_{\frac{1}{2}} \circ \rho_2 \circ \rho_1.$$

Glide reflections: Since simplicial automorphisms preserve meridional lines and clasping triangles, the reflection component of the glide reflection must be across a line $x = \frac{k}{2}$ or $y = k$, $k \in \mathbb{Z}$. If the reflection is across $x = k \in \mathbb{Z}$, then the translation must be $(x, y) \mapsto (x, y + 2n)$, $n \in \mathbb{Z}$, so modulo deck transformations this is a pure reflection, and can be ruled out. Thus we are left with the glide reflection $g = \tau_{0,1} \circ r_{\frac{1}{2}}$, given by the reflection across $x = \frac{1}{2}$ followed by the translation $(x, y) \mapsto (x, y + 1)$, and the compositions

$$r_y \circ \tau_{1,0} = r_y \circ \rho_1 \circ \rho_3; \quad r_y \circ \tau_{2,0} = r_y \circ \rho_2 \circ \rho_1; \quad r_{\frac{3}{2}} \circ \tau_{0,1} = g \circ \rho_2 \circ \rho_1,$$

all others being obtained by composing with deck transformations.

We show that $r_y \notin \mathrm{Aut}_{ev}(\widetilde{T})$ by considering edge valences near a clasping triangle. Using the edge/vertex correspondences from Figure 9, we obtain the four pictures in Figure 11, which correspond to the cases $\alpha_1 \geq 3$, $\alpha_1 = 2$, $\alpha_1 = 1 \neq \alpha_2$, and $\alpha_1 = 1 = \alpha_2$, respectively (note that Ω nonarithmetic implies $\Omega \notin \{RL, RLR, R^2L^2\}$). For the first three pictures it is clear that r_y does not preserve edge valence. For the last picture, if $r_y \in \mathrm{Aut}_{ev}(\widetilde{T})$ then $c = d = 8$, so that $\alpha_3 = 2$, which implies $8 = d = a = 4$, a contradiction. Hence $r_y \notin \mathrm{Aut}_{ev}(\widetilde{T})$.

In order to rule out ρ_4 and the compositions above involving ρ_4 and r_y, we will first need to establish the last assertion of the theorem, namely that we always have $\rho_1, \rho_2 \in \mathrm{Aut}_{ev}(\widetilde{T})$, and ρ_3 and g are in $\mathrm{Aut}_{ev}(\widetilde{T})$ if and only if they are simplicial automorphisms of $\widetilde{T}$. To see this, first note that ρ_1 and ρ_2 are always simplicial automorphisms (by construction of $\widetilde{T}$). Thus we need only show that if any of g, ρ_1, ρ_2, or ρ_3 is a simplicial homeomorphism, then it is in $\mathrm{Aut}_{ev}(\widetilde{T})$. But this follows from the fact that each of g, ρ_1, ρ_2, and ρ_3 preserve the edge/vertex correspondence given in Lemma 3.2(c) (shown graphically in Figure 9). In particular, each of these maps switches the parity of k in part (c) of the lemma, but also exchanges right and left. Thus, if g is simplicial, it preserves vertex valence, and since it also preserves edge/vertex correspondence, it must preserve edge valence, that is, $g \in \mathrm{Aut}_{ev}(\widetilde{T})$. The same holds for ρ_1, ρ_2, and ρ_3, so the assertion is proved.

Now, suppose that $\rho_4 \in \mathrm{Aut}_{ev}(\widetilde{T})$. First, observe that

$$g = \tau_{0,1} \circ r_{\frac{1}{2}} = r_{\frac{1}{2}} \circ \tau_{0,1} = (\rho_3 \circ r_y) \circ (\rho_3 \circ \rho_4) = (\rho_3 \circ r_y \circ \rho_3) \circ \rho_4 = r_y \circ \rho_4.$$

Since r_y is always a simplicial automorphism (by construction of $\widetilde{T}$), $\rho_4 \in \mathrm{Aut}_{ev}(\widetilde{T})$ implies that g is a simplicial automorphism, so by the above paragraph, $g \in \mathrm{Aut}_{ev}(\widetilde{T})$. But $g, \rho_4 \in \mathrm{Aut}_{ev}(\widetilde{T})$ implies that $r_y \in \mathrm{Aut}_{ev}(\widetilde{T})$, a contradiction.

Thus we can rule out the compositions $\tau_{1,1}$, $\tau_{3,1}$, $\rho_4 \circ \rho_2 \circ \rho_1$, r_1, r_2, and $r_y \circ \tau_{2,0}$. For $r_{\frac{1}{2}}$ and $r_y \circ \tau_{1,0}$, since r_y is always a simplicial homeomorphism, the composition is simplicial if and only if ρ_3 is. But then by the above observation it follows that ρ_3 preserves edge valence, so the composition cannot preserve edge valence (because r_y does not). Last, $r_{\frac{3}{2}}$ can now be ruled out since $r_{\frac{1}{2}} \notin \mathrm{Aut}_{ev}(\widetilde{T})$.

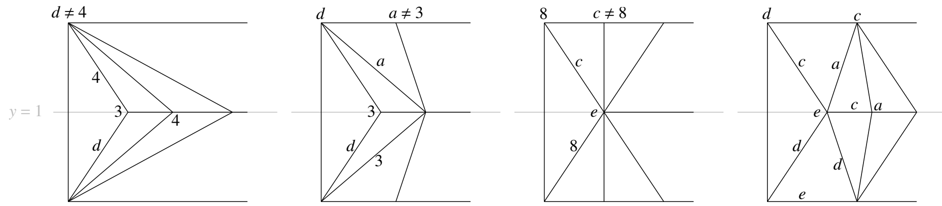


Figure 11. Reflecting by r_y about $y = 1$, for the cases $\alpha_1 \geq 3$, $\alpha_1 = 2$, $\alpha_1 = 1 \neq \alpha_2$, and $\alpha_1 = 1 = \alpha_2$, from left to right respectively.

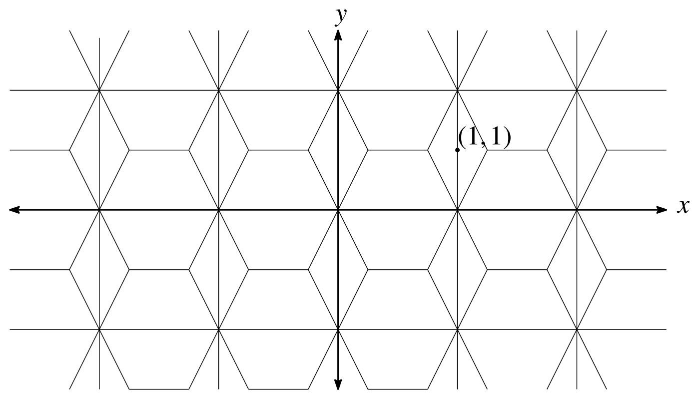


Figure 12. The union E of all edges of horizontal strips and clasping triangles, in the case where $\Omega_c = R$. If $\Omega_c = L$ then the clasping triangles adjacent to line $y = k$, k odd, will be shifted vertically by 1, and horizontal strips will be parallelograms.

Since the only compositions we have not ruled out are generated by ρ_1, ρ_2, ρ_3, and g, and since compositions involving ρ_3 (resp. g) are in $\mathrm{Aut}_{ev}(\widetilde{T})$ if and only if ρ_3 (resp. g) is, the result follows. $\square$

Remark 3.7. In Theorem 3.6 we have described a set *containing* the generators of $\mathrm{Aut}_{ev}(\widetilde{T})$, but we do not know whether they are all in fact generators. We will easily obtain in Section 5 a complete description of this group.

4. Symmetries of 2-bridge link complements

Let $M = \mathbb{S}^3 \setminus K(\Omega)$, and let $\mathrm{Sym}(M)$ denote the symmetries of M, i.e., $\mathrm{Sym}(M)$ is the group of self-homeomorphisms of M up to isotopy. Here, we describe the action of $\mathrm{Sym}(M)$ on the triangulation $\widetilde{T}$. First, Theorem 4.1 gives a classification of the symmetries of M in terms of the word Ω. This theorem comes from combining Theorem II.3.2 and Lemma II.3.3 in [Sakuma and Weeks 1995] and translating from $[a_1, a_2, \ldots, a_n]$ to the word Ω given by the following dictionary: $a_1 = \alpha_1 + 1$,

$a_i = \alpha_i$ for $i \le 2 \le n-1$, and $a_n = \alpha_n + 1$. In [Sakuma and Weeks 1995], these symmetries are called automorphisms of the triangulation $\mathcal{T}$ of M described in Section 3. Since by [Guéritaud 2006a] this triangulation is now known to coincide with the canonical triangulation of M, we know these automorphisms actually correspond to all of the symmetries of M.

We let $\mathrm{Sym}^+(M)$ denote the subgroup of $\mathrm{Sym}(M)$ consisting of orientation-preserving symmetries. We say that Ω is *palindromic* if $\alpha_i = \alpha_{n-i+1}$ for all $1 \le i \le n$.

Theorem 4.1 [Sakuma and Weeks 1995; Guéritaud 2006a]. *Let $M = \mathbb{S}^3 \setminus K(\Omega)$ be any hyperbolic 2-bridge link complement. Then $\mathrm{Sym}(M) = \mathrm{Sym}^+(M) \cong Z_2 \oplus Z_2$ if and only if Ω is not palindromic. When Ω is palindromic, then we have the following possibilities*:

- *If n is even, then* $\mathrm{Sym}(M) \cong D_4$ *and* $\mathrm{Sym}^+(M) \cong Z_2 \oplus Z_2$.
- *If n is odd and $\alpha_{\frac{n+1}{2}}$ is odd, then* $\mathrm{Sym}(M) = \mathrm{Sym}^+(M) \cong D_4$.
- *If n is odd and $\alpha_{\frac{n+1}{2}}$ is even, then* $\mathrm{Sym}(M) = \mathrm{Sym}^+(M) \cong Z_2 \oplus Z_2 \oplus Z_2$.

Note that the 2-bridge link complements with orientation-reversing symmetries are exactly those with n even and Ω palindromic.

We would like to understand how these symmetries act on $\widetilde{T}$. In order to accomplish this, we first show that $\mathrm{Sym}(M) = \mathrm{Sym}(\mathbb{S}^3, K(\Omega))$. Here, $\mathrm{Sym}(\mathbb{S}^3, K(\Omega))$ denotes the symmetries of $(\mathbb{S}^3, K(\Omega))$, that is, the group of self-homeomorphisms of the pair $(\mathbb{S}^3, K(\Omega))$ up to isotopy. Mostow–Prasad rigidity implies that $\mathrm{Sym}(M) \supseteq \mathrm{Sym}(\mathbb{S}^3, K)$ for any hyperbolic link K. In fact, if K is a hyperbolic knot, then by the knot complement theorem of Gordon and Luecke [1989], $\mathrm{Sym}(M) = \mathrm{Sym}(\mathbb{S}^3, K)$. However, here we do not rely on the knot complement theorem, and in addition, we prove the desired equality for both hyperbolic 2-bridge knots and hyperbolic 2-bridge links with two components. Once we have established this correspondence, we can determine how these symmetries act on the cusp triangulation, T. From here, we just lift this action of $\mathrm{Sym}(M)$ on T to the universal cover $\mathbb{R}^2$, to get the corresponding action on $\widetilde{T}$.

The following proposition is certainly known by the experts in the field. However, the authors were unable to find a reference in the literature.

Proposition 4.2. *Let $M = \mathbb{S}^3 \setminus K(\Omega)$ be a hyperbolic 2-bridge link complement. Then* $\mathrm{Sym}(M) = \mathrm{Sym}(\mathbb{S}^3, K(\Omega))$.

Proof. The work of Guéritaud [2006a] shows that $\mathcal{T}$ is in fact the canonical triangulation of any such hyperbolic 2-bridge link complement M. Thus, $\mathrm{Aut}(\mathcal{T})$, the group of combinatorial automorphisms of this triangulation, is isomorphic to $\mathrm{Sym}(M)$. The description of $\mathrm{Aut}(\mathcal{T})$ in [Sakuma and Weeks 1995, pp. 415-416] implies that it preserves the meridian(s) of $K(\Omega)$, and therefore extends to an

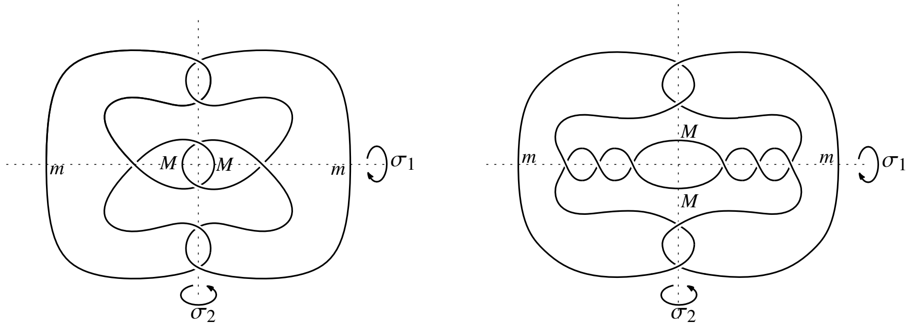


Figure 13. Trisymmetric projections of a 2-bridge link with two components (left) and a 2-bridge knot (right). The axes of symmetry for σ_1 and σ_2 are given in both projections. Maxima are labeled M and minima are labeled m.

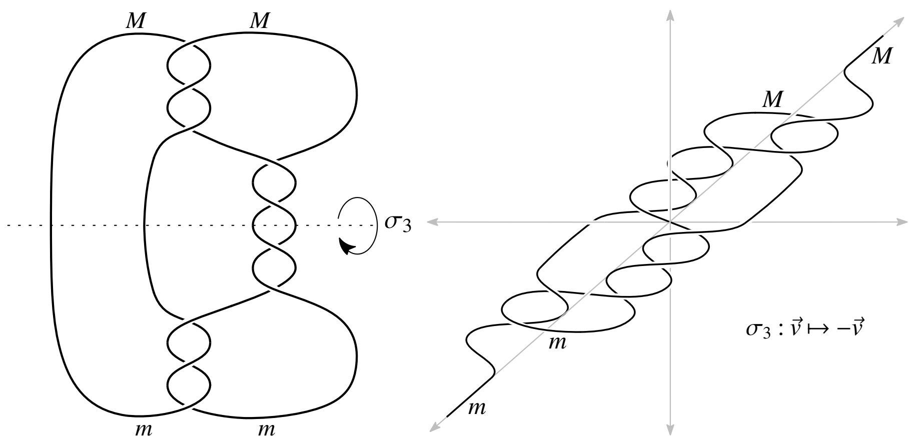


Figure 14. To the left is the standard projection of $K(\Omega)$ with Ω palindromic and n odd. To the right is a depiction of $K(\Omega)$ in $\mathbb{R}^3$ (with knot strands connecting at infinity) with Ω palindromic and n even. Both visuals show a symmetry σ_3 of $K(\Omega)$. Maxima are labeled M and minima are labeled m.

action on $(\mathbb{S}^3, K(\Omega))$. As a result, the natural inclusion from $\mathrm{Sym}(\mathbb{S}^3, K(\Omega))$ into $\mathrm{Sym}(M)$ is surjective, giving the desired isomorphism. $\square$

Since $\mathrm{Sym}(M)$ is isomorphic to $\mathrm{Sym}(\mathbb{S}^3, K(\Omega))$, we will no longer distinguish between symmetries of a hyperbolic 2-bridge link and its complement. Below, we provide visualizations of these symmetries, which will be useful in the proofs of Lemma 4.3 and Proposition 4.4. For more visualizations of 2-bridge link symmetries, see [Bleiler and Moriah 1988; Bonahon and Siebenmann 2010], and [Sakuma 1986].

Recall that any 2-bridge link $K(\Omega)$ can be isotoped so that its projection has exactly two maxima and two minima. In all four link diagrams given in Figure 13 and Figure 14 the corresponding maxima and minima are labeled. In what follows, we will examine how $\mathrm{Sym}(M)$ acts on these maxima and minima, and "meridional edges" of $\mathrm{Aut}(\mathcal{T})$ that wrap around them. For an arbitrary link $L \subset \mathbb{S}^3$, this would be an issue since the maxima and minima don't have to be preserved up to isotopy, and $\mathrm{Sym}(\mathbb{S}^3, L)$ is a group of homeomorphisms up to isotopy. However, for a 2-bridge link, from the work of Schubert [1956] we know that the set of maxima and minima will be preserved up to isotopy, and so, we are justified in using different projections of $K(\Omega)$ to analyze how symmetries act on the maxima and minima.

Lemma 4.3. *Each "meridional edge" of T wraps around a maximum or minimum of $K(\Omega)$. These meridional edges alternate between ones that wrap around maxima and minima.*

Proof. In all cases, $\mathcal{T}$, the canonical triangulation of $\mathbb{S}^3 \setminus K(\Omega)$, has exactly four meridional edges, and $K(\Omega)$ has exactly four extrema. These meridional edges of $\mathcal{T}$ result from clasping. See Section 3 for details on how clasping the innermost and outermost 4-punctured spheres, S_c and S_1, affects $\mathcal{T}$. Specifically, clasping S_1 introduces two meridional edges, each one going around one of the strands of the outermost crossing of $K(\Omega)$. We get the other two meridional edges from clasping S_c, each one going around one of the strands of the innermost crossing. See Figure 4 for how clasping forms these meridional edges. The two meridional edges coming from clasping S_1 each go around a maximum of $K(\Omega)$, while the two meridional edges coming from clasping S_c each go around a minimum of $K(\Omega)$. Since there are exactly four meridional edges in $\mathcal{T}$ and exactly four meridional edges in T, these sets must correspond with one another. Thus, each meridional edge of T wraps around a maximum or minimum of $K(\Omega)$. These meridional edges alternate between wrapping around maxima and minima since if we orient $K(\Omega)$, our path alternates between traversing maxima and minima. □

We now consider the lifts of the meridional edges of T to $\widetilde{T}$. In what follows, we shall call the lifts of meridional edges of T that wrap around a maximum of $K(\Omega)$ *maximal meridional edges.* Similarly, we shall call the lifts of the meridional edges of T that wrap around a minimum of $K(\Omega)$ *minimal meridional edges.*

We now describe how the symmetries of a hyperbolic 2-bridge link complement act on $\widetilde{T}$. Recall that n is the number of syllables in the word Ω. If $K(\Omega)$ is a two component link, then we say $\widetilde{T} = \widetilde{T}_1 \cup \widetilde{T}_2$, where $\widetilde{T}_1$ and $\widetilde{T}_2$ are identical triangulations of $\mathbb{R}^2$, coming from lifting an equal volume cusp cross-section of $\mathbb{S}^3 \setminus K(\Omega)$.

Recall that σ_1, σ_2, and σ_3 are the symmetries of $\mathrm{Sym}(\mathbb{S}^3, K(\Omega))$ described above and shown in Figure 13 and Figure 14.

Proposition 4.4. $\mathrm{Sym}(M) = \mathrm{Sym}(\mathbb{S}^3, K(\Omega))$ *acts on* $\widetilde{T}$ (*up to deck transformations*) *in the following manner*:

If $K(\Omega)$ *is a knot*, *then*

- σ_1 *acts as a rotation of* π *about* $(1, 1)$, *and*
- σ_2 *acts as a rotation of* π *about* $(2, 1)$.

If $K(\Omega)$ *is a two component link*, *then*

- σ_1 *acts as a rotation of* π *about* $(1, 1)$ *in both* $\widetilde{T}_1$ *and* $\widetilde{T}_2$, *and*
- σ_2 *exchanges* $(\mathbb{R}^2, \widetilde{T}_1)$ *and* $(\mathbb{R}^2, \widetilde{T}_2)$ *by the identity map.*

If Ω *is palindromic*, *then*

- *if* n *is odd*, σ_3 *acts as a rotation of* π *about* $\left(\frac{1}{2}, 1\right)$, *and*
- *if* n *is even*, σ_3 *acts as a glide reflection where we reflect across the line* $x = \frac{1}{2}$ *and translate by* $(x, y) \to (x, y + 1)$ (*possibly composed with the rotations* σ_1 *and* σ_2).

Proof. First, we claim that any symmetry of M acts on $(\mathbb{R}^2, \widetilde{T})$ by an isometry of $\mathbb{R}^2$. A priori, a symmetry of M gives rise only to an element f of $\mathrm{Aut}_{ev}(\widetilde{T})$ since this triangulation is metrically distorted in our construction. By Theorem 3.6, any such simplicial homeomorphism (that preserves edge valences) of $\widetilde{T}$ is a composition of deck transformations (which are specific translations) and a specific set of rotations, reflections, and glide reflections. Thus, any such f must be a Euclidean isometry.

First, we consider the symmetries σ_1 and σ_2 of M that generate a subgroup of $\mathrm{Sym}(M)$ isomorphic to $Z_2 \oplus Z_2$. By Theorem 4.1, these symmetries are always orientation-preserving, and so, we just need to consider rotations and translations of $\mathbb{R}^2$. We do this in three cases.

Case 1: $K(\Omega)$ is a knot. In this case, we note the following properties of σ_1 and σ_2. These properties come from examining the tri-symmetric projection given in Figure 13:

- σ_1 exchanges the maxima of $K(\Omega)$ while fixing the minima of $K(\Omega)$.
- σ_2 exchanges the minima of $K(\Omega)$ while fixing the maxima of $K(\Omega)$.
- σ_1 and σ_2 change the orientation of the longitude of $K(\Omega)$.

Since both σ_1 and σ_2 change the orientation of the longitude, they cannot be translations, and so, must be rotations. By Lemma 4.3, σ_1 must exchange the maximal meridional edges while fixing the two minimal meridional edges. Thus, up to deck transformations, σ_1 must be a rotation of π about $(1, 1)$. Similarly, up to deck transformations, σ_2 must be a rotation of π about $(2, 1)$.

Case 2: $K(\Omega)$ is a 2-component link. Here, we once again note several important features of σ_1 and σ_2 acting on $(\mathbb{S}^3, K(\Omega))$ which come from examining the tri-symmetric projection in Figure 13.

- σ_1 sends each component of $K(\Omega)$ to itself, with maxima mapping to maxima and minima mapping to minima.
- σ_2 exchanges the two link components, with maxima mapping to maxima and minima mapping to minima.
- σ_1 changes the orientations of both of the longitudes of $K(\Omega)$, while σ_2 preserves these orientations.

Since σ_1 is an orientation-preserving symmetry that switches the orientation of both of the longitudes, it must act as a rotation on both copies of $\mathbb{R}^2$. Up to deck transformations, the only possible rotation that maps the two maximal meridional edges to themselves and maps the two minimal meridional edges to themselves is a rotation of π about $(1, 1)$ in both $(\mathbb{R}^2, \widetilde{T}_1)$ and $(\mathbb{R}^2, \widetilde{T}_2)$. Since σ_2 interchanges the cusps and preserves orientations of the longitudes, it must take $\widetilde{T}_1$ to $\widetilde{T}_2$ by a translation. Since the minimal meridional edge of $\widetilde{T}_1$ must map to the minimal meridional edge of $\widetilde{T}_2$, σ_2 must be the identity map between these triangulations of $\mathbb{R}^2$, up to deck transformations.

Case 3: Ω is palindromic. Now, consider any additional symmetries of $\mathrm{Sym}(M)$, which occur only if Ω is palindromic. By examining the projections of $K(\Omega)$ given in Figure 14, we see that σ_3 has the following properties:

- σ_3 exchanges the maxima of $K(\Omega)$ with the minima of $K(\Omega)$.
- σ_3 changes the orientation of the longitude of $K(\Omega)$ (or both longitudes if $K(\Omega)$ is a two component link).

First, suppose that n is odd. By Theorem 4.1, σ_3 is an orientation-preserving symmetry, and since it changes the orientation of the longitude, it must be a rotation of $\mathbb{R}^2$. Since σ_3 must exchange maximal meridional edges with minimal meridional edges, it must act as a rotation about $\left(\frac{1}{2}, 1\right)$ on $(\mathbb{R}^2, \widetilde{T})$, or rotations about $\left(\frac{1}{2}, 1\right)$ in both $(\mathbb{R}^2, \widetilde{T}_1)$ and $(\mathbb{R}^2, \widetilde{T}_2)$, if $K(\Omega)$ has two components.

Now, suppose that n is even. By Theorem 4.1, σ_3 is an orientation-reversing symmetry of M, and so, σ_3's action on $\widetilde{T}$ is also orientation-reversing. Theorem 3.6 tells us that σ_3 must either correspond with the glide reflection g or a composition of g with the rotations ρ_1 and ρ_2 (up to deck transformation). This gives the desired description of σ_3. □

5. Hidden symmetries of 2-bridge link complements

Let the commensurator and normalizer of $M = \mathbb{H}^3/\Gamma = \mathbb{S}^3 \setminus K(\Omega)$, be $C(\Gamma)$ and $N(\Gamma)$, respectively, as defined in Section 1. Now that we understand the symmetries of M (Section 4), and the simplicial homeomorphisms of the canonical (lifted) cusp triangulation $\widetilde{T}$ (Section 3), we are ready to characterize the hidden symmetries of M, i.e., the elements of $C(\Gamma) \setminus N(\Gamma)$. Clearly, arithmetic links always have hidden symmetries, since in this case $C(\Gamma)$ is dense in $\mathrm{Isom}(\mathbb{H}^3)$. But hidden symmetries of arithmetic links will not necessarily be symmetries of the canonical cusp triangulation $\widetilde{T}$. We call a hidden symmetry *detectable* if it is also a symmetry of $\widetilde{T}$. For nonarithmetic links, all hidden symmetries are detectable.

Recall that $\mathrm{Aut}_{ev}(\widetilde{T})$ is the group of simplicial automorphisms of $\widetilde{T}$ preserving edge valence, so that $\mathrm{Aut}^+_{ev}(\widetilde{T})$ is the subgroup consisting of those that preserve orientation.

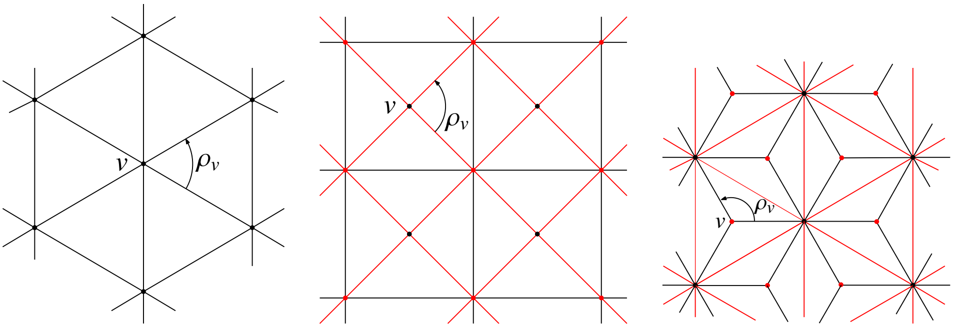
Orientation-preserving hidden symmetries.

Theorem 5.1. *If $M = \mathbb{S}^3 \setminus K(\Omega)$ is a hyperbolic 2-bridge link complement, then we have the following classification of orientation-preserving hidden symmetries:*

- *If M is nonarithmetic, then M admits no hidden symmetries.*
- *If M is the figure-eight knot complement, then M admits an order 6 detectable hidden symmetry.*
- *If M is the Whitehead link complement, then M admits an order 4 detectable hidden symmetry.*
- *If M is the 6^2_2 link complement, then M admits an order 3 detectable hidden symmetry.*
- *If M is the 6^2_3 link complement, then M does not admit any detectable hidden symmetries.*

Proof. Case 1: M is nonarithmetic. Since the triangulation $\mathcal{T}$ of M is canonical, it descends to a cellulation of the minimal (orientable) orbifold $\mathcal{O}^+ = \mathbb{H}^3/C^+(\Gamma)$, where $C^+(\Gamma)$ is the orientable commensurator of M. Hence any orientation-preserving symmetry or hidden symmetry $h \in C^+(\Gamma) \leq \mathrm{Isom}^+(\mathbb{H}^3)$ must preserve the lifted triangulation $\widetilde{\mathcal{T}}$, which we may assume has a vertex at $\infty \in S_\infty = \mathbb{R}^2 \cup \{\infty\}$. Since M either has one cusp or has a symmetry exchanging its cusps, $N^+(\Gamma)$ acts transitively on the set of vertices of $\widetilde{\mathcal{T}}$. Thus for some $g \in N^+(\Gamma)$, $h \circ g$ fixes $\infty \in S_\infty$. Since h is a symmetry of M if and only if $h \circ g$ is, we may assume that h fixes $\infty \in S_\infty$. Identifying $\widetilde{T}$ with a horosphere about ∞, we see then that h restricts to a simplicial automorphism of $\widetilde{T}$, and this restriction determines h (if K has two components, we understand $\widetilde{T}$ to mean a component of $\widetilde{T}_1 \cup \widetilde{T}_2$). It is enough, then, to show that any element of $\mathrm{Aut}^+_{ev}(\widetilde{T})$ comes from a symmetry of M (possibly composed with deck transformations of $\widetilde{T}$).

Let $G = \mathbb{Z} \oplus \mathbb{Z}$ be the deck group of $\widetilde{T}$. By Theorem 3.6, $\mathrm{Aut}^+_{ev}(\widetilde{T})/G$ is generated by $\{\rho_1, \rho_2, \rho_3\}$ if ρ_3 is a simplicial automorphism, and is generated by $\{\rho_1, \rho_2\}$ if ρ_3



$\Omega = RL$ $\qquad$ $\Omega = RLR$ $\qquad$ $\Omega = R^2L^2$

Figure 15. Lifted cusp triangulation $\widetilde{T}$ for the figure-eight knot, Whitehead link, and 6^2_2 link complements, from left to right. Edges/vertices with the same coloring (within each figure) have the same valence.

is not simplicial. Let σ_1, σ_2, σ_3 be the symmetries described in Proposition 4.4, and let H be the horizontal strip in the first quadrant with a vertex at the origin.

We first observe that $\rho_1 = \sigma_1$, and ρ_2 is either σ_2, or σ_1 composed with a deck transformation, depending on whether K has one or two components. Hence ρ_1 and ρ_2 come from symmetries of M in both cases, and so for the case where ρ_3 is not simplicial, M cannot have hidden symmetries. If ρ_3 *is* simplicial, then since the reflection r_y across $y = 1$ is always a simplicial automorphism (by construction of $\widetilde{T}$), the reflection $\rho_3 \circ r_y$ across $x = \frac{1}{2}$ is also simplicial. Hence in this case H is symmetric about the line $x = \frac{1}{2}$, and so Ω is palindromic with $\Omega_c = R$, and it follows that ρ_3 comes from the symmetry σ_3 of M. Again, we conclude that M has no hidden symmetries.

Case 2: M is arithmetic. There are exactly four arithmetic 2-bridge links: the figure-eight knot ($\Omega = RL$), the Whitehead link ($\Omega = RLR$), the 6^2_2 link ($\Omega = R^2L^2$), and the 6^2_3 link ($\Omega = RL^2R$).

Since $\Omega = RL^2R$ is not an excluded case in Lemma 3.4 and its corollaries, the arguments in Case 1 above show that, if M is the 6^2_3 link complement, then every $h \in \mathrm{Aut}^+_{ev}(\widetilde{T})$ that preserves edge valence comes from a symmetry of M, i.e., M admits no detectable orientation-preserving hidden symmetries.

If M is the figure-eight knot, the Whitehead link complement, or the 6^2_2 link complement, then we can see by edge/vertex (valence) correspondences in $\widetilde{T}$ that if e and e' are two edges of a tetrahedron in $\mathcal{T}$ which are opposite each other (i.e., they do not share a vertex), then $val(e) = val(e')$. This is evident in $\widetilde{T}$ by the fact that any edge and vertex of $\widetilde{T}$ that are across from each other (i.e., their convex hull is a single triangle of $\widetilde{T}$) have the same valence. This makes it easy to identify the (unique) hyperbolic structure on $\mathcal{T}$. If an edge of a tetrahedron has valence k, then we make the dihedral angle at that edge $2\pi/k$. We just need to make sure that this gives a Euclidean structure to the cusp cross-sections, but this is confirmed by Figure 15. It follows that the depictions of $\widetilde{T}$ in Figure 15 are actually metrically correct (up to scaling), so the rotations ρ_v indicated are isometries of $\widetilde{T}$. Next we check that ρ_v extends to an isometry of the three-dimensional triangulation $\widetilde{\mathcal{T}}$. Viewing $\widetilde{T}$ as a horosphere about ∞ in the upper half-space model of $\mathbb{H}^3$, the vertex v about which ρ_v rotates $\widetilde{T}$ corresponds to some edge e_v of $\widetilde{\mathcal{T}}$ connecting ∞ to a point $p_v \in \partial\mathbb{H}^3 \setminus \{\infty\}$. The rotation of $\mathbb{H}^3$ about e_v that agrees with ρ_v on $\widetilde{T}$ induces a rotation of the lift $\widetilde{T}_v$ of T centered at p_v, which is an isometry since $\widetilde{T}$ and $\widetilde{T}_v$ are isometric and e_v appears in both as a vertex of the same valence. If v_1 is some other vertex of $\widetilde{T}$, and $\rho_v(v_1) = v_2$, then since ρ_v differs from ρ_{v_1} by composition with symmetries of M and deck transformations of $\widetilde{T}$, the rotation of $\mathbb{H}^3$ induced by ρ_v takes $\widetilde{T}_{v_1}$ to $\widetilde{T}_{v_2}$ isometrically. It follows that ρ_v induces an isometry on $\widetilde{\mathcal{T}}$, of the order indicated in the statement of the theorem. □

Orientation-reversing hidden symmetries.

Theorem 5.2. *If $M = \mathbb{S}^3 \setminus K(\Omega)$ is a hyperbolic 2-bridge link complement, then we have the following classification of orientation-reversing hidden symmetries:*

- *If M is nonarithmetic, then M admits no orientation-reversing hidden symmetries.*
- *If M is the 6_3^2 link complement, then M admits no detectable orientation-reversing hidden symmetry.*
- *If M is the figure-eight knot complement, the Whitehead link complement, or the 6_2^2 link complement, then M admits an order 2 orientation-reversing hidden symmetry.*

Proof. Case 1: M is nonarithmetic. The proof will be analogous to the orientation-preserving case. As in that case, we need only show that any $h \in \mathrm{Aut}_{ev}(\widetilde{T})$ is in fact a symmetry of M. By Theorem 3.6, h must be a composition of ρ_1, ρ_2, ρ_3, and g, where ρ_1, ρ_2, and ρ_3 are the rotations by π about $(1, 1)$, $(2, 1)$, and $\left(\frac{1}{2}, 1\right)$, respectively, and g is the glide reflection given by the composition of $r_{\frac{1}{2}}$ with $(x, y) \mapsto (x, y + 1)$. If $g \notin \mathrm{Aut}_{ev}(\widetilde{T})$, then $\mathrm{Aut}_{ev}(\widetilde{T}) = \mathrm{Aut}^+_{ev}(\widetilde{T})$, and we are done. If $g \in \mathrm{Aut}_{ev}(\widetilde{T})$, then it is clear from the construction of $\widetilde{T}$ that we must have $\Omega_c = L$, and Ω must be palindromic. In this case, though, g corresponds to the symmetry σ_3 in the notation of Proposition 4.4, so the nonarithmetic case is proved.

Case 2: M is arithmetic. The proof is analogous to the orientation-preserving case. □

Irregular coverings by hyperbolic 2-bridge link complements. Theorem 5.1 and Theorem 5.2 give us the following corollary about irregular coverings of 3-manifolds.

Corollary 5.3. *Let N be any hyperbolic 2-bridge link complement. If N is nonarithmetic, then N does not irregularly cover any hyperbolic 3-orbifolds (orientable or nonorientable). If N is arithmetic, then N does not irregularly cover any orientable hyperbolic 3-manifolds.*

Proof. By Theorem 5.1 and Theorem 5.2, any nonarithmetic hyperbolic 2-bridge link complement N does not have any hidden symmetries (orientation-preserving or orientation-reversing). Thus, if any such N covers a hyperbolic 3-orbifold, it must be a regular cover.

If N is arithmetic, then N is the complement of either the figure-eight knot, the Whitehead link, the 6_2^2 link, or the 6_3^2 link. If N irregularly covers some hyperbolic 3-manifold N', then it must be at least a degree 3 covering. Here, we get a volume contradiction. Cao and Meyerhoff [2001] showed that the figure-eight knot complement and its sister are the orientable cusped hyperbolic 3-manifolds of minimal volume, with volume ≥ 2.029. Therefore, $vol(N') \geq 2.029$, and so, $vol(N) \geq 3(2.029) = 6.087$. However, the volumes of any of the four arithmetic hyperbolic 2-bridge link complements are strictly smaller than 6.087. Thus, we can't have any such irregular coverings in the arithmetic case. □

Boileau and Weidmann [2005] give a characterization of 3-manifolds that admit a nontrivial JSJ-decomposition and whose fundamental groups are generated by two elements. Their work shows that there are four possibilities for such manifolds, one of which is that the hyperbolic part of the JSJ decomposition admits a finite-sheeted irregular covering by a hyperbolic 2-bridge link complement. Corollary 5.3 immediately eliminates this possibility, giving the following revised characterization of such manifolds. In the following corollary, D stands for a disk, A for an annulus, and Mb for a Möbius band. For an orbifold, cone points are listed in parentheses after the topological type of the orbifold is given.

Corollary 5.4. *Let M be a compact, orientable, irreducible 3-manifold which has $rank(\pi_1(M)) = 2$. If M has a nontrivial JSJ-decomposition, then one of the following holds:*

(1) *M has Heegaard genus 2.*

(2) *$M = S \cup_T H$ where S is a Seifert manifold with basis $D(p, q)$ or $A(p)$, H is a hyperbolic manifold and $\pi_1(H)$ is generated by a pair of elements with a single parabolic element. The gluing map identifies the fiber of S with the curve corresponding to the parabolic generator of $\pi_1(H)$.*

(3) *$M = S_1 \cup_T S_2$ where S_1 is a Seifert manifold over Mb or $Mb(p)$ and S_2 is a Seifert manifold over $D(2, 2l + 1)$. The gluing map identifies the fiber of S_1 with a curve on the boundary of S_2 that has intersection number one with the fiber of S_2.*

6. Commensurability of 2-bridge link complements

In this section, we show that there is only one pair of commensurable hyperbolic 2-bridge link complements. We accomplish this by analyzing the cusp of the unique minimal orbifold in the commensurability class of a nonarithmetic hyperbolic 2-bridge link complement.

Let $M = \mathbb{S}^3 \setminus K(\Omega) = \mathbb{H}^3/\Gamma$ be any nonarithmetic hyperbolic 2-bridge link complement. By a theorem of Margulis [1991], there exists a unique minimal (orientable) orbifold in the commensurability class of M, specifically, $\mathcal{O}^+ = \mathbb{H}^3/C^+(\Gamma)$. By Theorem 5.1 we know that M admits no hidden symmetries, and therefore, $C^+(\Gamma) = N^+(\Gamma)$. Since $N^+(\Gamma)/\Gamma = \mathrm{Sym}^+(M)$, we only have to quotient M by its orientation-preserving symmetries to obtain $\mathcal{O}^+$.

We will analyze the commensurability class of M by considering the cusp of $\mathcal{O}^+$. Recall that every 2-bridge link is either a knot or a link with two components. If K has two components, then there always exists a symmetry exchanging those components; see Section 4. Thus, the orbifold $\mathcal{O}^+$ admits a single cusp, C. If we quotient the cusp(s) of M along with the cusp triangulation T by the symmetries of M, then we obtain the cusp C of $\mathcal{O}^+$, along with a canonical cellulation, T_C. Technically, T_C is not a triangulation, but just a quotient of a triangulation (hence we call it a cellulation). If M and M' are commensurable, then their corresponding minimal orbifolds must admit isometric cusps that have identical cusp triangulations. In this case, we say that the corresponding cusp cellulations, T_C and $T_{C'}$, are *equivalent*. We wish to determine when these cusps are equivalent. The following two lemmas take care of this classification.

Lemma 6.1. *Let $M = \mathbb{S}^3 \setminus K(\Omega)$ be a nonarithmetic hyperbolic 2-bridge link complement. Suppose Ω is not palindromic or n is even. Then $C \cong S^2(2, 2, 2, 2)$ and T_C determines the word Ω up to inversion and switching Ls and Rs.*

Proof. By Theorem 4.1, $\mathrm{Sym}^+(M) \cong Z_2 \oplus Z_2$, and Proposition 4.4 tells us exactly how $\mathrm{Sym}^+(M)$ acts on T and $\widetilde{T}$. First, assume $K(\Omega)$ is a knot. Here, we choose the rectangle $[0, 4] \times [0, 2]$ in $\widetilde{T}$ as a fundamental domain for the torus T. In this case, $\sigma_1 \circ \sigma_2$ acts as a translation of $\widetilde{T}$ by $(x, y) \to (x+2, y)$. When we quotient our fundamental domain by the symmetry $\sigma_1 \circ \sigma_2$, we produce a fundamental domain for a torus given by the rectangle $[0, 2] \times [0, 2]$, with opposite sides identified. If $K(\Omega)$ is a link with two components, then our fundamental domain for T is given by two copies of $[0, 2] \times [0, 2]$. When we quotient by σ_2, we just exchange the cusps. This again produces a fundamental domain for a (single) torus of the form $[0, 2] \times [0, 2]$ in $\widetilde{T}$. In either case (a knot or a two component link), we just need to quotient by σ_1, which acts as a rotation about $(1, 1)$, to obtain C along with T_C. This gives us a fundamental domain of the form $[0, 1] \times [0, 2]$, with identifications given in Figure 16. We can see that this resulting cusp is $S^2(2, 2, 2, 2)$.

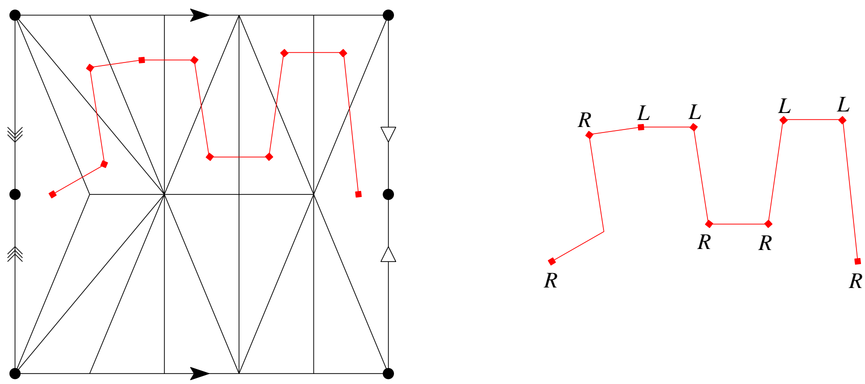


Figure 16. This shows the cusp triangulation T_C for the word $\Omega = R^2L^3R^2L^2R$. The order two singularities are marked by solid black circles. The red line segment gives l_C.

To each such T_C we associate a labeled line segment, l_C, in the following manner, depicted in Figure 16. The two endpoints of this line segment come from vertices placed in the centers of the two clasping triangles of the fundamental domain of T_C. We also place a vertex in the center of each triangle in the top half of the triangulation of the fundamental domain for T_C. We connect two vertices by an edge if and only if the corresponding triangles in T_C share an edge. We label each vertex of l_C (including the endpoints) by L or R corresponding to the label of the triangle in T_C. We say that l_C is equivalent to another labeled line segment $l_{C'}$ if there exists a simplicial homeomorphism between the two that preserves labelings or switches Ls and Rs between labelings.

Now, T_C is equivalent to $T_{C'}$ if and only if l_C is equivalent to $l_{C'}$. This holds because l_C tells you exactly how to build T_C and vice versa. However, there are only two possibilities for how l_C can be equivalent to $l_{C'}$: either the left endpoint maps to the left endpoint, or the left endpoint maps to the right endpoint. In the first case, Ω must be the same as Ω'. In the second case, Ω' must be an inversion of Ω. $\square$

Lemma 6.2. *Let $M = \mathbb{S}^3 \setminus K(\Omega)$ be a nonarithmetic hyperbolic 2-bridge link complement. Suppose Ω is palindromic and n is odd. Then $C \cong S^2(2,2,2,2)$ and T_C determines the word Ω up to inversion and switching Ls and Rs.*

Proof. By Theorem 4.1, either $\mathrm{Sym}^+(M) \cong Z_2 \oplus Z_2 \oplus Z_2$, or $\mathrm{Sym}^+(M) \cong D_4$. Just as in the previous lemma, we can first quotient a fundamental domain for T in $\widetilde{T}$ by the $Z_2 \oplus Z_2$ subgroup of $\mathrm{Sym}^+(M)$ to obtain a single $S^2(2,2,2,2)$ cusp. To obtain C and T_C, we also quotient by the action of σ_3, which is a rotation about $\left(\frac{1}{2}, 1\right)$ in $\widetilde{T}$ by Proposition 4.4; see Figure 17.

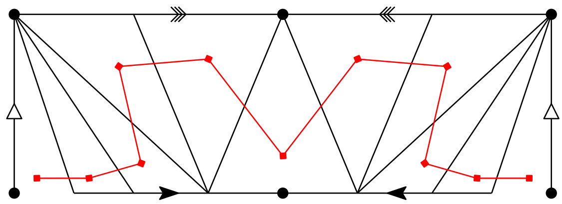


Figure 17. This shows the cusp triangulation T_C for the word $\Omega = R^3L^2RL^2R^3$. The order two singularities are marked by solid black circles. The red line segment gives l_C.

Similar to Lemma 6.1, we can associate a marked line segment l_C to each cusp T_C, as depicted in Figure 17. Once again, we see that this marked line segment determines T_C up to inversions and switching Ls and Rs. We leave the details for the reader. □

Corollary 6.3. *Let $M = \mathbb{S}^3 \setminus K(\Omega)$ be a nonarithmetic hyperbolic* 2*-bridge link complement. Then $C \cong S^2(2, 2, 2, 2)$ and T_C is determined by the word Ω up to inversion and switching Ls and Rs.*

Proof. We claim that the two types of cusp cellulations coming from Lemma 6.1 and Lemma 6.2 can not be equivalent. First, note that the tiling T_C for an $S^2(2, 2, 2, 2)$ from Lemma 6.1 always has singularities located at vertices. Furthermore, any of these vertices with singularities have valence $\neq 2$. Now, the tiling coming from Lemma 6.2 either has a singularity that is not located at a vertex (this happens if $\alpha_{\frac{n+1}{2}}$ is odd) or it has a singularity located at a vertex of valence 2 (this happens if $\alpha_{\frac{n+1}{2}}$ is even). Thus, these two types of cusp cellulations can not be equivalent, and so, the previous two lemmas imply that any such T_C is determined by the word Ω up to inversion and switching Ls and Rs. □

We can now prove our main theorem.

Theorem 6.4. *The only commensurable hyperbolic* 2*-bridge link complements are the figure-eight knot complement and the 6^2_2 link complement.*

Proof. It is a well known fact that cusped, arithmetic hyperbolic 3-manifolds are commensurable if and only if they have the same invariant trace field; see [Maclachlan and Reid 2003] for details. The figure-eight knot complement and the 6^2_2 link complement both have invariant trace field $\mathbb{Q}(\sqrt{-3})$, while the Whitehead link complement has $\mathbb{Q}(\sqrt{-1})$ and the 6^2_3 link complement has $\mathbb{Q}(\sqrt{-7})$. Thus, among hyperbolic arithmetic 2-bridge link complements, only the figure-eight knot complement and the 6^2_2 link complement are commensurable. Now, a nonarithmetic hyperbolic 2-bridge link complement can not be commensurable with an arithmetic hyperbolic 2-bridge link complement. This is because their commensurators determine their commensurability classes, and by a theorem of Margulis [1991], the commensurator of a hyperbolic 3-manifold is discrete if and only if it is nonarithmetic.

It remains to check that nonarithmetic hyperbolic 2-bridge link complements are pairwise incommensurable. Let $M = \mathbb{S}^3 \setminus K(\Omega)$ and $M' = \mathbb{S}^3 \setminus K(\Omega')$ be any two such manifolds. We use T_C and $T_{C'}$ to denote the cusp cellulations of the minimal orbifolds in the commensurability classes of M and M' respectively. Recall that if T_C is not equivalent to $T_{C'}$, then M and M' are not commensurable. By Corollary 6.3, T_C and $T_{C'}$ are equivalent only if Ω and Ω' differ by inversion or switching Ls and Rs. As noted in Section 2, both of these possibilities result in M and M' being isometric. Thus, M and M' are commensurable only if they are isometric, as desired. □

Acknowledgements

The authors would like to thank David Futer for helpful conversations and guidance on this work. We would also like to thank the referee for making a number of helpful suggestions.

References

[Akiyoshi et al. 2007] H. Akiyoshi, M. Sakuma, M. Wada, and Y. Yamashita, *Punctured torus groups and 2-bridge knot groups, I*, Lecture Notes in Mathematics **1909**, Springer, 2007. MR 2330319 Zbl 1132.57001

[Bleiler and Moriah 1988] S. A. Bleiler and Y. Moriah, "Heegaard splittings and branched coverings of B^3", *Math. Ann.* **281**:4 (1988), 531–543. MR 958258 Zbl 0627.57007

[Boileau and Weidmann 2005] M. Boileau and R. Weidmann, "The structure of 3-manifolds with two-generated fundamental group", *Topology* **44**:2 (2005), 283–320. MR 2114709 Zbl 1082.57014

[Bonahon and Siebenmann 2010] F. Bonahon and L. Siebenmann, "New geometric splittings of classical knots, and the classification and symmetries of arborescent knots", preprint, 2010, http://www-bcf.usc.edu/~fbonahon/Research/Preprints/BonSieb.pdf.

[Cao and Meyerhoff 2001] C. Cao and G. R. Meyerhoff, "The orientable cusped hyperbolic 3-manifolds of minimum volume", *Invent. Math.* **146**:3 (2001), 451–478. MR 1869847 Zbl 1028.57010

[Gehring et al. 1998] F. W. Gehring, C. Maclachlan, and G. J. Martin, "Two-generator arithmetic Kleinian groups, II", *Bull. London Math. Soc.* **30**:3 (1998), 258–266. MR 1608106 Zbl 0930.30037

[Goodman et al. 2008] O. Goodman, D. Heard, and C. Hodgson, "Commensurators of cusped hyperbolic manifolds", *Experiment. Math.* **17**:3 (2008), 283–306. MR 2455701 Zbl 1338.57016

[Gordon and Luecke 1989] C. M. Gordon and J. Luecke, "Knots are determined by their complements", *J. Amer. Math. Soc.* **2**:2 (1989), 371–415. MR 965210 Zbl 0678.57005

[Guéritaud 2006a] F. Guéritaud, *Géométrie hyperbolique effective et triangulations idéales canoniques en dimension trois*, Ph.D. thesis, Université Paris Sud, 2006, https://tel.archives-ouvertes.fr/tel-00119465.

[Guéritaud 2006b] F. Guéritaud, "On canonical triangulations of once-punctured torus bundles and two-bridge link complements", *Geom. Topol.* **10** (2006), 1239–1284. MR 2255497 Zbl 1130.57024

[Maclachlan and Reid 2003] C. Maclachlan and A. W. Reid, *The arithmetic of hyperbolic 3-manifolds*, Graduate Texts in Mathematics **219**, Springer, 2003. MR 1937957 Zbl 1025.57001

[Margulis 1991] G. A. Margulis, *Discrete subgroups of semisimple Lie groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **17**, Springer, 1991. MR 1090825 Zbl 0732.22008

[Menasco 1984] W. Menasco, "Closed incompressible surfaces in alternating knot and link complements", *Topology* **23**:1 (1984), 37–44. MR 721450 Zbl 0525.57003

[Reid and Walsh 2008] A. W. Reid and G. S. Walsh, "Commensurability classes of 2-bridge knot complements", *Algebr. Geom. Topol.* **8**:2 (2008), 1031–1057. MR 2443107 Zbl 1154.57001

[Sakuma 1986] M. Sakuma, "On strongly invertible knots", pp. 176–196 in *Algebraic and topological theories* (Kinosaki, 1984), edited by M. Nagata et al., Kinokuniya, Tokyo, 1986. MR 1102258 Zbl 0800.57001

[Sakuma and Weeks 1995] M. Sakuma and J. Weeks, "Examples of canonical decompositions of hyperbolic link complements", *Japan. J. Math.* (*N.S.*) **21**:2 (1995), 393–439. MR 1364387 Zbl 0858.57021

[Schubert 1956] H. Schubert, "Knoten mit zwei Brücken", *Math. Z.* **65** (1956), 133–170. MR 0082104 Zbl 0071.39002

[Schwartz 1995] R. E. Schwartz, "The quasi-isometry classification of rank one lattices", *Inst. Hautes Études Sci. Publ. Math.* 82 (1995), 133–168. MR 1383215 Zbl 0852.22010

[Walsh 2011] G. S. Walsh, "Orbifolds and commensurability", pp. 221–231 in *Interactions between hyperbolic geometry, quantum topology and number theory*, edited by E. Loubeau and S. Montaldo, Contemp. Math. **541**, Amer. Math. Soc., Providence, RI, 2011. MR 2796635 Zbl 1231.57017

Received January 5, 2016. Revised July 5, 2016.

CHRISTIAN MILLICHAP
DEPARTMENT OF MATHEMATICS
LINFIELD COLLEGE
MCMINNVILLE, OR 97128
UNITED STATES
cmillich@linfield.edu

WILLIAM WORDEN
DEPARTMENT OF MATHEMATICS
TEMPLE UNIVERSITY
PHILADELPHIA, PA 19122
UNITED STATES
william.worden@temple.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 285, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.285.485

ON SEAWEED SUBALGEBRAS AND MEANDER GRAPHS IN TYPE C

DMITRI I. PANYUSHEV AND OKSANA S. YAKIMOVA

In 2000, Dergachev and Kirillov introduced subalgebras of "seaweed type" in $\mathfrak{gl}_n$ (or $\mathfrak{sl}_n$) and computed their index using certain graphs. In this article, those graphs are called type-A meander graphs. Then the subalgebras of seaweed type, or just "seaweeds", were defined by Panyushev (2001) for arbitrary simple Lie algebras. Namely, if $\mathfrak{p}_1, \mathfrak{p}_2 \subset \mathfrak{g}$ are parabolic subalgebras such that $\mathfrak{p}_1 + \mathfrak{p}_2 = \mathfrak{g}$, then $\mathfrak{q} = \mathfrak{p}_1 \cap \mathfrak{p}_2$ is a seaweed in $\mathfrak{g}$. If $\mathfrak{p}_1$ and $\mathfrak{p}_2$ are "adapted" to a fixed triangular decomposition of $\mathfrak{g}$, then $\mathfrak{q}$ is said to be standard. The number of standard seaweeds is finite. A general algebraic formula for the index of seaweeds was proposed by Tauvel and Yu (2004) and then proved by Joseph (2006).

In this paper, elaborating on the "graphical" approach of Dergachev and Kirillov, we introduce the type-C meander graphs, i.e., the graphs associated with the standard seaweed subalgebras of $\mathfrak{sp}_{2n}$, and give a formula for the index in terms of these graphs. We also note that the very same graphs can be used in the case of the odd orthogonal Lie algebras.

Recall that $\mathfrak{q}$ is called Frobenius if the index of $\mathfrak{q}$ equals 0. We provide several applications of our formula to Frobenius seaweeds in $\mathfrak{sp}_{2n}$. In particular, using a natural partition of the set $\mathcal{F}_n$ of standard Frobenius seaweeds, we prove that $\#\mathcal{F}_n$ strictly increases for the passage from n to $n+1$. The similar monotonicity question is open for the standard Frobenius seaweeds in $\mathfrak{sl}_n$, even for the passage from n to $n+2$.

1. Introduction

The index of an (algebraic) Lie algebra $\mathfrak{q}$, $\operatorname{ind}\mathfrak{q}$, is the minimal dimension of the stabilisers for the coadjoint representation of $\mathfrak{q}$. It can be regarded as a generalisation of the notion of rank. That is, $\operatorname{ind}\mathfrak{q}$ equals the rank of $\mathfrak{q}$ if $\mathfrak{q}$ is reductive. In [Dergachev and Kirillov 2000], the index of the subalgebras of "seaweed type"

Panyushev gratefully acknowledges the hospitality of MPIM (Bonn) during the preparation of this article. Yakimova is partially supported by the DFG priority programme SPP 1388 (Darstellungstheorie) and by the Graduiertenkolleg GRK 1523 (Quanten- und Gravitationsfelder).
MSC2010: primary 17B08; secondary 17B20.
Keywords: index of Lie algebra, Frobenius Lie algebra.

in $\mathfrak{gl}_n$ (or $\mathfrak{sl}_n$) were computed using certain graphs. In this article, those graphs are called *type-*A *meander graphs*. Then the subalgebras of seaweed type, or just *seaweeds*, were defined and studied for an arbitrary simple Lie algebra $\mathfrak{g}$ [Panyushev 2001]. Namely, if $\mathfrak{p}_1, \mathfrak{p}_2 \subset \mathfrak{g}$ are parabolic subalgebras such that $\mathfrak{p}_1 + \mathfrak{p}_2 = \mathfrak{g}$, then $\mathfrak{q} = \mathfrak{p}_1 \cap \mathfrak{p}_2$ is a seaweed in $\mathfrak{g}$. If $\mathfrak{p}_1$ and $\mathfrak{p}_2$ are "adapted" to a fixed triangular decomposition of $\mathfrak{g}$, then $\mathfrak{q}$ is said to be standard; see Section 2 for details. A general algebraic formula for the index of seaweeds was proposed in [Tauvel and Yu 2004, Conjecture 4.7] and then proved in [Joseph 2006, Section 8].

In this paper, elaborating on the "graphical" approach of [Dergachev and Kirillov 2000], we introduce the *type-*C *meander graphs*, i.e., the graphs associated with the standard seaweed subalgebras of $\mathfrak{sp}_{2n}$, and give a formula for the index in terms of these graphs. Although the seaweeds in $\mathfrak{sp}_{2n}$ are our primary object in Sections 2–4, we note that the very same graphs can be used in the case of the odd orthogonal Lie algebras; see Section 5.

Recall that $\mathfrak{q}$ is called *Frobenius* if $\operatorname{ind}\mathfrak{q} = 0$. Frobenius Lie algebras are very important in mathematics because of their connection with the Yang–Baxter equation. We provide some applications of our formula to Frobenius seaweeds in $\mathfrak{sp}_{2n}$. Let $\mathcal{F}_n$ denote the set of standard Frobenius seaweeds of $\mathfrak{sp}_{2n}$. For a natural partition

$$\mathcal{F}_n = \bigsqcup_{k=1}^{n} \mathcal{F}_{n,k}$$

(see Section 4 for details), we construct the embeddings $\mathcal{F}_{n,k} \hookrightarrow \mathcal{F}_{n+1,k+1}$ for all $n, k \geqslant 1$. Since $\mathcal{F}_{n+1,1}$ does not meet the image of the induced embedding $\mathcal{F}_n \hookrightarrow \mathcal{F}_{n+1}$ and $\#(\mathcal{F}_{n+1,1}) > 0$, this implies that $\#(\mathcal{F}_n) < \#(\mathcal{F}_{n+1})$. The similar monotonicity question is open for the standard Frobenius seaweeds in $\mathfrak{sl}_n$, even for the passage from n to $n+2$. We also show that $\mathcal{F}_{n,1}$ and $\mathcal{F}_{n,2}$ are related to certain Frobenius seaweeds in $\mathfrak{sl}_n$.

The ground field is algebraically closed and of characteristic zero.

2. Generalities on seaweed subalgebras and meander graphs

Let $\mathfrak{p}_1$ and $\mathfrak{p}_2$ be two parabolic subalgebras of a simple Lie algebra $\mathfrak{g}$. If $\mathfrak{p}_1+\mathfrak{p}_2 = \mathfrak{g}$, then $\mathfrak{p}_1 \cap \mathfrak{p}_2$ is called a *seaweed subalgebra* or just a *seaweed* in $\mathfrak{g}$ (see [Panyushev 2001]). The set of seaweeds includes all parabolics (if $\mathfrak{p}_2 = \mathfrak{g}$), all Levi subalgebras (if $\mathfrak{p}_1$ and $\mathfrak{p}_2$ are opposite), and many interesting nonreductive subalgebras. We assume that $\mathfrak{g}$ is equipped with a fixed triangular decomposition, so that there are two opposite Borel subalgebras $\mathfrak{b}$ and $\mathfrak{b}^-$, and a Cartan subalgebra $\mathfrak{t} = \mathfrak{b} \cap \mathfrak{b}^-$. Without loss of generality, we may also assume that $\mathfrak{p}_1 \supset \mathfrak{b}$ (i.e., $\mathfrak{p}_1$ is standard) and $\mathfrak{p}_2 = \mathfrak{p}_2^- \supset \mathfrak{b}^-$ (i.e., $\mathfrak{p}_2$ is opposite-standard). Then the seaweed $\mathfrak{q} = \mathfrak{p}_1 \cap \mathfrak{p}_2^-$ is said to be *standard*, too. Either of these parabolics is determined by a subset of Π, the set of simple roots associated with $(\mathfrak{b}, \mathfrak{t})$. Therefore, a standard seaweed is determined by two arbitrary subsets of Π; see [Panyushev 2001, Section 2] for details.

For classical Lie algebras $\mathfrak{sl}_n$ and $\mathfrak{sp}_{2n}$, we exploit the usual numbering of Π, which allows us to identify the standard and opposite-standard parabolic subalgebras with certain compositions related to n. It is also more convenient to deal with $\mathfrak{gl}_n$ in place of $\mathfrak{sl}_n$.

I. $\mathfrak{g} = \mathfrak{gl}_n$. We work with the obvious triangular decomposition of $\mathfrak{gl}_n$, where $\mathfrak{b}$ consists of the upper-triangular matrices. If $\mathfrak{p}_1 \supset \mathfrak{b}$ and the standard Levi subalgebra of $\mathfrak{p}_1$ is $\mathfrak{gl}_{a_1} \oplus \cdots \oplus \mathfrak{gl}_{a_s}$, then we set $\mathfrak{p}_1 = \mathfrak{p}(\underline{a})$, where $\underline{a} = (a_1, a_2, \ldots, a_s)$. Note that $a_1 + \cdots + a_s = n$ and all $a_i \geqslant 1$. Likewise, if $\mathfrak{p}_2^- \supset \mathfrak{b}^-$ is represented by a composition $\underline{b} = (b_1, \ldots, b_t)$ with $\sum b_j = n$, then the standard seaweed $\mathfrak{p}_1 \cap \mathfrak{p}_2^- \subset \mathfrak{gl}_n$ is denoted by $\mathfrak{q}^{\mathsf{A}}(\underline{a} \,|\, \underline{b})$. The corresponding type-A meander graph $\Gamma = \Gamma^{\mathsf{A}}(\underline{a} \,|\, \underline{b})$ is defined by the following rules:

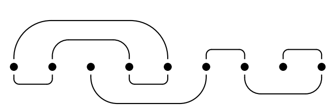
- Γ has n consecutive vertices on a horizontal line numbered from 1 to n.
- The parts of $\underline{a}$ determine the set of pairwise disjoint arcs (edges) that are drawn *above* the horizontal line. Namely, part a_1 determines $\left[\frac{1}{2}a_1\right]$ consecutively embedded arcs above the nodes $1, \ldots, a_1$, where the widest arc joins vertices 1 and a_1, the following joins 2 and $a_1 - 1$, etc. If a_1 is odd, then the middle vertex $\frac{1}{2}(a_1 + 1)$ acquires no arc at all. Next, part a_2 determines $\left[\frac{1}{2}a_2\right]$ embedded arcs above the nodes $a_1 + 1, \ldots, a_1 + a_2$, etc.
- The arcs corresponding to $\underline{b}$ are drawn following the same rules, but *below* the horizontal line.

It follows that the degree of each vertex in Γ is at most 2 and each connected component of Γ is homeomorphic to either a circle or a segment. (An isolated vertex is also a segment!) By [Dergachev and Kirillov 2000], the index of $\mathfrak{q}^{\mathsf{A}}(\underline{a} \,|\, \underline{b})$ can be computed via $\Gamma = \Gamma^{\mathsf{A}}(\underline{a} \,|\, \underline{b})$ as follows:

(2-1) $\quad \operatorname{ind} \mathfrak{q}^{\mathsf{A}}(\underline{a} \,|\, \underline{b}) = 2(\text{number of cycles in } \Gamma) + (\text{number of segments in } \Gamma)$.

Remark 2.1. Formula (2-1) gives the index of a seaweed in $\mathfrak{gl}_n$, not in $\mathfrak{sl}_n$. However, if $\mathfrak{q} \subset \mathfrak{gl}_n$ is a seaweed, then $\mathfrak{q} \cap \mathfrak{sl}_n$ is a seaweed in $\mathfrak{sl}_n$ and the respective mapping $\mathfrak{q} \mapsto \mathfrak{q} \cap \mathfrak{sl}_n$ is a bijection. Here $\mathfrak{q} = (\mathfrak{q} \cap \mathfrak{sl}_n) \oplus (1\text{-dim centre of } \mathfrak{gl}_n)$; hence $\operatorname{ind}(\mathfrak{q} \cap \mathfrak{sl}_n) = \operatorname{ind} \mathfrak{q} - 1$. Since $\operatorname{ind} \mathfrak{q}^{\mathsf{A}}(\underline{a} \,|\, \underline{b}) \geqslant 1$ and the minimal value 1 is achieved if and only if Γ is a sole segment, we also obtain a characterisation of the Frobenius seaweeds in $\mathfrak{sl}_n$.

Example 2.2. We have

$\Gamma^{\mathsf{A}}(5, 2, 2 \,|\, 2, 4, 3) =$ 

and the index of the corresponding seaweed in $\mathfrak{gl}_9$ (resp. $\mathfrak{sl}_9$) equals 3 (resp. 2).

II. $\mathfrak{g} = \mathfrak{sp}_{2n}$. We use the embedding $\mathfrak{sp}_{2n} \subset \mathfrak{gl}_{2n}$ such that

$$\mathfrak{sp}_{2n} = \left\{ \begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & -\hat{\mathcal{A}} \end{pmatrix} \middle| \, \mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathfrak{gl}_n,\ \mathcal{B} = \hat{\mathcal{B}},\ \mathcal{C} = \hat{\mathcal{C}} \right\},$$

where $\mathcal{A} \mapsto \hat{\mathcal{A}}$ is the transpose with respect to the antidiagonal. If $\tilde{\mathfrak{b}} \subset \mathfrak{gl}_{2n}$ and $\tilde{\mathfrak{b}}^-$ are the sets of upper-triangular and lower-triangular matrices, respectively, then $\mathfrak{b} = \tilde{\mathfrak{b}} \cap \mathfrak{sp}_{2n}$ and $\mathfrak{b}^- = \tilde{\mathfrak{b}}^- \cap \mathfrak{sp}_{2n}$ are our fixed Borel subalgebras of $\mathfrak{g} = \mathfrak{sp}_{2n}$. If $\mathfrak{p}_1 \supset \mathfrak{b}$, then the standard Levi subalgebra of $\mathfrak{p}$ is $\mathfrak{gl}_{a_1} \oplus \cdots \oplus \mathfrak{gl}_{a_s} \oplus \mathfrak{sp}_{2d}$, where $a_1 + \cdots + a_s + d = n$, all $a_i \geqslant 1$, and $d \geqslant 0$. Since d is determined by n and the '$\mathfrak{gl}$' parts, $\mathfrak{p}_1$ can be represented by n and the composition $\underline{a} = (a_1, \ldots, a_s)$. We write $\mathfrak{p}_n(\underline{a})$ for it. Likewise, if $\mathfrak{p}_2^-$ is represented by another composition $\underline{b} = (b_1, \ldots, b_t)$ with $\sum b_j \leqslant n$, then $\mathfrak{p}_1 \cap \mathfrak{p}_2^-$ is denoted by $\mathfrak{q}_n^{\mathsf{C}}(\underline{a} \,|\, \underline{b})$. To a standard parabolic $\mathfrak{p}_1 = \mathfrak{p}_n(\underline{a}) \subset \mathfrak{sp}_{2n}$, one can associate the parabolic subalgebra $\tilde{\mathfrak{p}}_1 \subset \mathfrak{gl}_{2n}$ that is represented by the symmetric composition $\underline{\tilde{a}} = (a_1, \ldots, a_s, 2d, a_s, \ldots, a_1)$ of $2n$. In the matrix form, the standard Levi subalgebra of $\tilde{\mathfrak{p}}_1$ has the consecutive diagonal blocks $\mathfrak{gl}_{a_1}, \ldots, \mathfrak{gl}_{a_s}, \mathfrak{gl}_{2d}, \mathfrak{gl}_{a_s}, \ldots, \mathfrak{gl}_{a_1}$ and, for the above embedding $\mathfrak{sp}_{2n} \subset \mathfrak{gl}_{2n}$ and compatible triangular decompositions, one has $\mathfrak{p}_1 = \tilde{\mathfrak{p}}_1 \cap \mathfrak{sp}_{2n}$ (and likewise for $\mathfrak{p}_2^- \subset \mathfrak{sp}_{2n}$ and $\tilde{\mathfrak{p}}_2^- \subset \mathfrak{gl}_{2n}$); see [Panyushev 2001, Section 5] for details. If $\underline{\tilde{a}}$ and $\underline{\tilde{b}}$ are symmetric compositions of $2n$, then the seaweed $\mathfrak{q}^{\mathsf{A}}(\underline{\tilde{a}} \,|\, \underline{\tilde{b}}) \subset \mathfrak{gl}_{2n}$ is said to be *symmetric*, too. The above construction provides a bijection between the standard seaweeds in $\mathfrak{sp}_{2n}$ and the symmetric standard seaweeds in $\mathfrak{gl}_{2n}$ (or $\mathfrak{sl}_{2n}$).

We define the *type-*C *meander graph* $\Gamma_n^{\mathsf{C}}(\underline{a} \,|\, \underline{b})$ for $\mathfrak{q}_n^{\mathsf{C}}(\underline{a} \,|\, \underline{b})$ to be the type-A meander graph of the corresponding symmetric seaweed $\tilde{\mathfrak{q}} = \tilde{\mathfrak{p}}_1 \cap \tilde{\mathfrak{p}}_2^- \subset \mathfrak{gl}_{2n}$. Formally,

$$\Gamma_n^{\mathsf{C}}(\underline{a} \,|\, \underline{b}) = \Gamma^{\mathsf{A}}(\underline{\tilde{a}} \,|\, \underline{\tilde{b}}).$$

We indicate below new features of these graphs.

- $\Gamma_n^{\mathsf{C}}(\underline{a} \,|\, \underline{b})$ has $2n$ consecutive vertices on a horizontal line numbered from 1 to $2n$.
- Part a_1 determines $\left[\frac{1}{2}a_1\right]$ embedded arcs above the nodes $1, \ldots, a_1$. By symmetry, the same set of arcs appears above the vertices $2n - a_1 + 1, \ldots, 2n$. Next, part a_2 determines $\left[\frac{1}{2}a_2\right]$ embedded arcs above the nodes $a_1 + 1, \ldots, a_1 + a_2$ and also the symmetric set of arcs above the nodes $2n - a_1 - a_2 + 1, \ldots 2n - a_1$, etc.
- If $d = n - \sum a_i > 0$, then there are $2d$ unused vertices in the middle, and we draw d embedded arcs above them. This corresponds to part $2d$ that occurs in the middle of $\underline{\tilde{a}}$. The arcs corresponding to $\underline{b}$ are depicted by the same rules, but *below* the horizontal line.
- A type-C meander graph is symmetric with respect to the vertical line between the n-th and $(n+1)$-th vertices, and the symmetry with respect to this line is

$\Gamma^{\mathsf{C}}_7(2,3\,|\,\varnothing)$

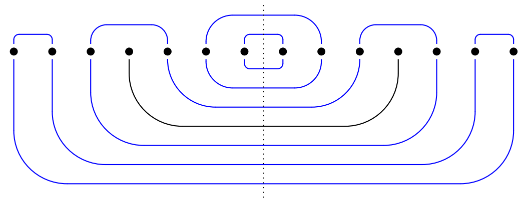


Figure 1. The meander graph for a parabolic subalgebra of $\mathfrak{sp}_{14}$.

denoted by σ. We also say that this line is the σ*-mirror*. The arcs crossing the σ-mirror are said to be *central*. These are exactly the arcs corresponding to $d = n - \sum a_i$ and $d' = n - \sum b_j$.

Our main result is the following formula for the index in terms of the connected components of $\Gamma^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b})$:

(2-2) $$\operatorname{ind} \mathfrak{q}^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b}) = (\text{number of cycles}) + \tfrac{1}{2}(\text{number of segments that are not } \sigma\text{-stable}).$$

To illustrate this formula, we recall that, for the parabolic subalgebra $\mathfrak{p}$ with Levi part $\mathfrak{gl}_{a_1} \oplus \cdots \oplus \mathfrak{gl}_{a_s} \oplus \mathfrak{sp}_{2d}$, we have $\operatorname{ind} \mathfrak{p} = \left[\frac{1}{2}a_1\right] + \cdots + \left[\frac{1}{2}a_s\right] + d$; see [Panyushev 2001, Theorem 5.5]. Here $\mathfrak{p}_2^- = \mathfrak{sp}_{2n}$ and the composition $\underline{b}$ is empty. On the other hand, the graph $\Gamma^{\mathsf{C}}_n(\underline{a}\,|\,\varnothing)$ has n central arcs below the horizontal line corresponding to $\underline{b} = \varnothing$. Hence each part a_i gives rise to $\left[\frac{1}{2}a_i\right]$ cycles and, if a_i is odd, to one additional segment, which is σ-invariant. The middle part corresponding to $\mathfrak{sp}_{2d}$ gives rise to d cycles. This clearly yields the same answer; cf. Example 2.3. Hence we already know that (2-2) is correct if $\mathfrak{q}$ is a parabolic subalgebra, i.e., if $\underline{a} = \varnothing$ or $\underline{b} = \varnothing$. Note also that $\operatorname{ind} \mathfrak{p} = 0$ if and only if $d = 0$ and all $a_i = 1$, i.e., if $\mathfrak{p} = \mathfrak{b}$.

Example 2.3. See Figure 1. Here $\underline{a} = (2,3)$ and $n = 7$ (hence $d = 2$), and the σ-mirror is represented by the vertical dotted line. It is easily seen that the only segment here is σ-stable and the total number of circles is 4. (The circles are depicted by blue arcs). Hence $\operatorname{ind} \mathfrak{p} = 4$.

Remark 2.4. (1) For both $\mathfrak{gl}_n$ and $\mathfrak{sp}_{2n}$, one has $\mathfrak{q}^*(\underline{a}\,|\,\underline{b}) \simeq \mathfrak{q}^*(\underline{b}\,|\,\underline{a})$. Hence one can freely choose what composition is going to appear first.

(2) Moreover, $\mathfrak{q}^*(\underline{a}\,|\,\underline{b})$ is reductive (i.e., a Levi subalgebra) if and only if $\underline{a} = \underline{b}$.

Convention. If $\mathfrak{q}$ is a seaweed in either $\mathfrak{sp}_{2n}$ or $\mathfrak{gl}_{2n}$, and the corresponding compositions are not specified, then the respective meander graph is denoted by $\Gamma^{\mathsf{C}}(\mathfrak{q})$ or $\Gamma^{\mathsf{A}}(\mathfrak{q})$.

Remark 2.5. Let $\mathfrak{q}$ be a seaweed in $\mathfrak{sp}_{2n}$ or $\mathfrak{gl}_n$. Then there is a point $\gamma \in \mathfrak{q}^*$ such that the stabiliser $\mathfrak{q}_\gamma \subset \mathfrak{q}$ is a reductive subalgebra; see [Panyushev 2005]. A Lie algebra possessing such a point in the dual space is said to be (*strongly*) *quasi-reductive* [Duflo et al. 2012]; see also [Moreau and Yakimova 2012, Definition 2.1]. One of the main results of [Duflo et al. 2012] states that if a Lie algebra $\mathfrak{q} = \operatorname{Lie} Q$ is strongly quasi-reductive, then there is a reductive stabiliser Q_γ (with $\gamma \in \mathfrak{q}^*$) such that any other reductive stabiliser Q_β (with $\beta \in \mathfrak{q}^*$) is contained in Q_γ up to conjugation. In [Moreau and Yakimova 2012] this subgroup Q_γ is a called a *maximal reductive stabiliser*, MRS for short. For a seaweed $\mathfrak{q} = \mathfrak{q}^{\mathsf{A}}(\underline{a}\,|\,\underline{b})$, an MRS of $\mathfrak{q}$ can be described in terms of $\Gamma^{\mathsf{A}}(\underline{a}\,|\,\underline{b})$ [Moreau and Yakimova 2012, Theorem 5.3]. A similar description is possible in type C if we use $\Gamma^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b})$. It will appear elsewhere.

3. Symplectic meander graphs and the index of seaweed subalgebras

In this section, we prove formula (2-2) on the index of the seaweed subalgebras of type C.

Let us recall the inductive procedure for computing the index of seaweeds in a symplectic Lie algebra introduced by the first author [Panyushev 2001]. Suppose that $\underline{a} = (a_1, \dots, a_s)$ and $\underline{b} = (b_1, \dots, b_t)$ are two compositions with $\sum a_i \leqslant n$ and $\sum b_j \leqslant n$. Then we consider the standard seaweed $\mathfrak{q}^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b}) \subset \mathfrak{sp}_{2n}$.

Inductive procedure:

(1) If either $\underline{a}$ or $\underline{b}$ is empty, then $\mathfrak{q}^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b})$ is a parabolic subalgebra and the index is computed using [Panyushev 2001, Theorem 5.5] (see also the Introduction).

(2) Suppose that both $\underline{a}$ and $\underline{b}$ are nonempty. Without loss of generality, we can assume that $a_1 \leqslant b_1$. By [Panyushev 2001, Theorem 5.2], $\operatorname{ind} \mathfrak{q}^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b})$ can inductively be computed as follows:

(i) If $a_1 = b_1$, then $\mathfrak{q}^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b}) \simeq \mathfrak{gl}_{a_1} \oplus \mathfrak{q}^{\mathsf{C}}_{n-a_1}(a_2, \dots, a_s\,|\,b_2, \dots, b_t)$; hence

$$\operatorname{ind} \mathfrak{q}^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b}) = a_1 + \operatorname{ind} \mathfrak{q}^{\mathsf{C}}_{n-a_1}(a_2, \dots, a_s\,|\,b_2, \dots, b_t).$$

(ii) If $a_1 < b_1$, then

$$\operatorname{ind} \mathfrak{q}^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b}) = \begin{cases} \operatorname{ind} \mathfrak{q}^{\mathsf{C}}_{n-a_1}(a_2, \dots, a_s\,|\,b_1 - 2a_1, a_1, b_2, \dots, b_t) & \text{if } a_1 \leqslant \frac{1}{2} b_1, \\ \operatorname{ind} \mathfrak{q}^{\mathsf{C}}_{n-b_1+a_1}(2a_1 - b_1, a_2, \dots, a_s\,|\,a_1, b_2, \dots, b_t) & \text{if } a_1 > \frac{1}{2} b_1. \end{cases}$$

(iii) Step 2 terminates when one of the compositions becomes empty, i.e., one obtains a parabolic subalgebra in a smaller symplectic Lie algebra, where Step 1 applies.

Remark 3.1. Iterating transformations of the form 2(ii) yields a formula that does not require considering cases; see [Panyushev 2001, Theorem 5.3]. Namely, if $a_1 < b_1$, then $\operatorname{ind} \mathfrak{q}^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b}) = \operatorname{ind} \mathfrak{q}^{\mathsf{C}}_{n-a_1}(\underline{a}'\,|\,\underline{b}')$, where $\underline{a}' = (a_2, \dots, a_s)$, $\underline{b}' = (b_1', b_1'', b_2, \dots, b_t)$, and b_1' and b_1'' are defined as follows. Let p be the unique integer such that

$$\frac{p}{p+1} < \frac{a_1}{b_1} \leqslant \frac{p+1}{p+2}.$$

Then $b'_1 = (p+1)b_1 - (p+2)a_1 \geqslant 0$ and $b'_2 = (p+1)a_1 - pb_1 > 0$. (If $b'_1 = 0$, then it has to be omitted.)

Theorem 3.2. *Let $\mathfrak{q} = \mathfrak{q}_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b})$ be a seaweed in $\mathfrak{sp}_{2n}$ and $\Gamma^{\mathsf{C}}(\mathfrak{q}) = \Gamma_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b})$ the type-C meander graph associated with $\mathfrak{q}$. Then*

$$\begin{aligned}&\operatorname{ind} \mathfrak{q}_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b})\\ &\qquad = \#\{\textit{cycles of } \Gamma_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b})\} + \tfrac{1}{2}\#\{\textit{segments of } \Gamma_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b}) \textit{ that are not } \sigma\textit{-stable}\}.\end{aligned}$$

Proof. Our argument exploits the above *inductive procedure*. Let us temporarily write $\mathcal{T}_n(\underline{a}\,|\,\underline{b})$ for the topological quantity in the right-hand side of the formula. Let us prove that for the pairs of seaweeds occurring in either 2(i) or 2(ii) of the inductive procedure, the required topological quantity behaves accordingly.

If $a_1 = b_1$ and $\mathfrak{gl}_{a_1}$ is a direct summand of $\mathfrak{q}$, then $\operatorname{ind} \mathfrak{q}_{n-a_1}^{\mathsf{C}}(a_2, \dots, a_s\,|\,b_2, \dots, b_t)$ decreases by a_1; on the other hand, $\Gamma_{n-a_1}^{\mathsf{C}}(a_2, \dots, a_s\,|\,b_2, \dots, b_t)$ is obtained from $\Gamma^{\mathsf{C}}(\mathfrak{q})$ by deleting $2\left[\frac{1}{2}a_1\right]$ cycles (and two segments, which are not σ-invariant in case a_1 is odd). This is in perfect agreement with the formula.

If $a_1 < b_1$, then one step of $\mathfrak{sp}$-reduction for $\mathfrak{q}$ is equivalent to two steps of $\mathfrak{gl}$-reduction for the meander graph of $\Gamma^{\mathsf{A}}(\tilde{\mathfrak{q}})$, where $\tilde{\mathfrak{q}}$ is the corresponding symmetric seaweed in $\mathfrak{gl}_{2n}$. These two "symmetric" steps are applied one after another to the left and right sides of $\Gamma^{\mathsf{A}}(\tilde{\mathfrak{q}}) = \Gamma^{\mathsf{C}}(\mathfrak{q})$. According to [Moreau and Yakimova 2012, Lemma 5.4(i)], the $\mathfrak{gl}$-reduction does not change the topological structure of the graph. Hence $\mathcal{T}_n(\underline{a}\,|\,\underline{b}) = \mathcal{T}_{n-a_1}(\underline{a}'\,|\,\underline{b}')$.

Since we have already observed (in Section 2) that our formula holds for the parabolic subalgebras, the result follows. □

Example 3.3. For the seaweed $\mathfrak{q}_{10}(3, 3|4, 5)$ in $\mathfrak{sp}_{20}$, the recursive formula of Remark 3.1 yields the following chain of reductions:

$$\mathfrak{q} = \mathfrak{q}_{10}^{\mathsf{C}}(3, 3|4, 5) \rightsquigarrow \mathfrak{q}_7^{\mathsf{C}}(3|1, 5) \rightsquigarrow \mathfrak{q}_6^{\mathsf{C}}(1, 1|5) \rightsquigarrow \mathfrak{q}_5^{\mathsf{C}}(1|3, 1) \rightsquigarrow \mathfrak{q}_4^{\mathsf{C}}(\varnothing|1, 1, 1).$$

The last term represents the minimal parabolic subalgebra of $\mathfrak{sp}_8$ corresponding to the unique long simple root. The respective graphs are gathered in Figure 2. It is readily seen that both ends of the graphs undergo the symmetric transformations on each step; also all the segments are σ-stable and the total number of cycles equals 1. Thus, $\operatorname{ind} \mathfrak{q} = 1$.

One can notice that each reduction step consists of contracting certain arcs starting from some end vertices of a meander graph. Clearly, such a procedure does not change the topological structure of the graph, and this is exactly how Lemma 5.4(i) in [Moreau and Yakimova 2012] was proved.

Example 3.4. In Figure 3, one finds the graph of a seaweed in $\mathfrak{sp}_{16}$ of index 1. The segments that are not σ-stable are depicted by red arcs.

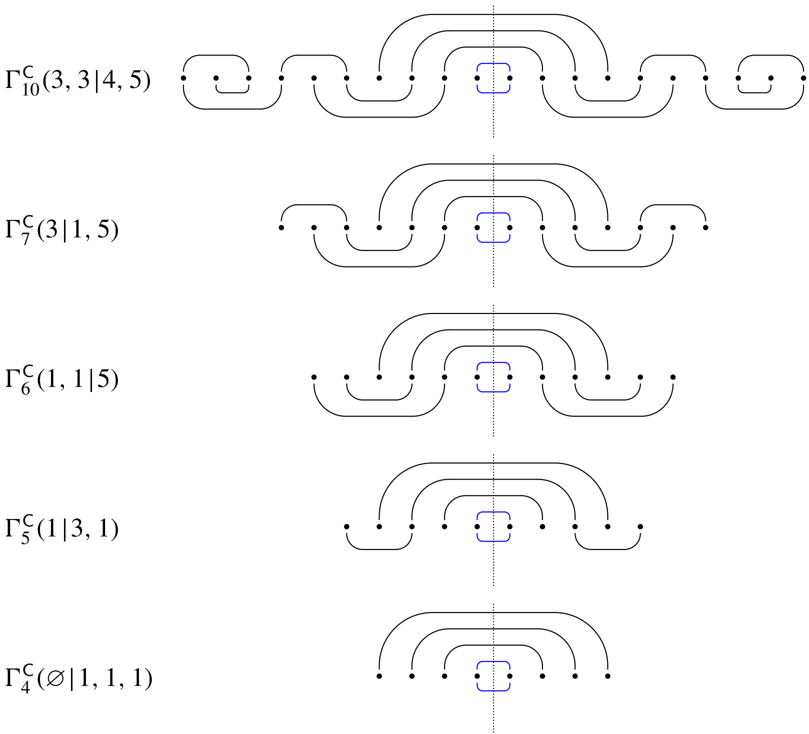


Figure 2. The reduction steps for a seaweed subalgebra of $\mathfrak{sp}_{20}$.

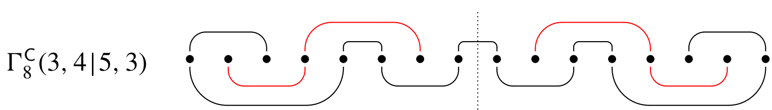


Figure 3. A seaweed subalgebra of $\mathfrak{sp}_{16}$ with index 1.

4. Applications of symplectic meander graphs

In this section, we present some applications of Theorem 3.2. We begin with a simple property of the index.

Lemma 4.1. *If* $\sum a_i < n$ *and* $\sum b_j < n$, *then* $\operatorname{ind}\mathfrak{q}^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b}) = (n-n') + \operatorname{ind}\mathfrak{q}^{\mathsf{C}}_{n'}(\underline{a}\,|\,\underline{b})$, *where* $n' = \max\{\sum a_i, \sum b_j\}$.

Proof. Here $\Gamma^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b})$ contains $n-n'$ arcs crossing the σ-mirror on *both* sides of the horizontal line. They form $n-n'$ central circles, and removing these circles reduces the index by $n-n'$ and yields the graph $\Gamma^{\mathsf{C}}_{n'}(\underline{a}\,|\,\underline{b})$. □

Recall that a Lie algebra $\mathfrak{q}$ is *Frobenius* if $\operatorname{ind}\mathfrak{q} = 0$. In the rest of the section, we apply Theorem 3.2 to studying Frobenius seaweeds. Clearly, if $\mathfrak{q}^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b})$ is Frobenius,

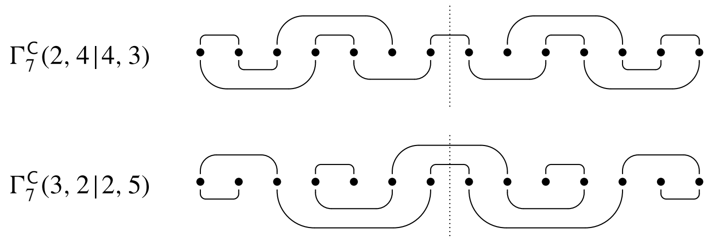


Figure 4. Frobenius seaweed subalgebras of $\mathfrak{sp}_{14}$.

then $\Gamma_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b})$ has only σ-stable segments and no cycles. Another consequence of Theorem 3.2 is the following necessary condition.

Lemma 4.2. *If $\mathfrak{q}_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b})$ is Frobenius, then either $\sum a_i < n$ and $\sum b_j = n$ or vice versa.*

Proof. If $\sum a_i < n$ and $\sum b_j < n$, then the index is positive in view of Lemma 4.1. If $\sum a_i = \sum b_j = n$, then there are no arcs crossing the σ-mirror. Therefore $\Gamma_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b})$ consists of two disjoint σ-symmetric parts, and the topological quantity of Theorem 3.2 cannot be equal to 0. (More precisely, in the second case $\mathfrak{q}_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b})$ is isomorphic to the seaweed $\mathfrak{q}^{\mathsf{A}}(\underline{a}\,|\,\underline{b})$ in $\mathfrak{gl}_n$, and $\operatorname{ind}\mathfrak{q} \geqslant 1$ for all seaweeds $\mathfrak{q} \subset \mathfrak{gl}_n$; see Remark 2.1.) $\square$

Graphically, Lemma 4.2 means that, for a Frobenius seaweed, one must have some central arcs (= arcs crossing the σ-mirror) on one side of the horizontal line in the meander graph, and then there must be no central arcs on the other side. The number of central arcs can vary from 1 to n (the last possibility represents the case in which one of the parabolics is the Borel subalgebra). Let $\mathcal{F}_{n,k}$ denote the set of standard Frobenius seaweeds whose meander graph contains k central arcs. Then $\mathcal{F}_n = \bigsqcup_{k=1}^{n} \mathcal{F}_{n,k}$ is the set of all standard Frobenius seaweeds in $\mathfrak{sp}_{2n}$. If $\mathfrak{q}_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b})$ lies in $\mathcal{F}_{n,k}$, then so does $\mathfrak{q}_n^{\mathsf{C}}(\underline{b}\,|\,\underline{a})$. As we are interested in essentially different meander graphs, we will not distinguish graphs and algebras corresponding to $(\underline{a}\,|\,\underline{b})$ and $(\underline{b}\,|\,\underline{a})$. Set $\boldsymbol{F}_{n,k} = \#(\mathcal{F}_{n,k}/{\sim})$ and $\boldsymbol{F}_n = \#(\mathcal{F}_n/{\sim})$, where $\sim$ is the corresponding equivalence relation. Then

$$\boldsymbol{F}_{n,n} = 1; \qquad \boldsymbol{F}_{n,n-1} = \begin{cases} 1, & n = 2, \\ 2, & n \geqslant 3; \end{cases} \qquad \boldsymbol{F}_{n,n-2} = \begin{cases} 2, & n = 3, \\ 4, & n = 4, \\ 5, & n \geqslant 5. \end{cases}$$

It follows from Lemma 4.2 that if $\mathfrak{q}_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b}) \in \mathcal{F}_n$ and $\sum b_j = n$, then the integer k such that $\mathfrak{q}_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b}) \in \mathcal{F}_{n,k}$ is determined as $k = n - \sum a_i$.

In Figure 4, one finds the meander graphs of Frobenius seaweeds in $\mathfrak{sp}_{14}$ with $k = 1$ and 2.

Lemma 4.3. *If* $\mathfrak{q} \in \mathcal{F}_{n,k}$, *then* $\Gamma^{\mathsf{C}}(\mathfrak{q})$ *has exactly k connected components* (σ*-stable segments*) *corresponding to the central arcs. Furthermore*, *the total number of arcs in* $\Gamma^{\mathsf{C}}(\mathfrak{q})$ *equals* $2n - k$.

Proof. (1) Let $\mathcal{A}_i$ be the i-th central arc and Γ_i the connected component of $\Gamma^{\mathsf{C}}(\mathfrak{q})$ that contains $\mathcal{A}_i$. Each Γ_i is a σ-stable segment.

- If $\Gamma_i = \Gamma_j$ for $i \neq j$, then continuations of $\mathcal{A}_i$ and $\mathcal{A}_j$ meet somewhere in the left half of $\Gamma^{\mathsf{C}}(\mathfrak{q})$. By symmetry, the same happens in the right half, which produces a cycle. Hence the connected components $\Gamma_1, \dots, \Gamma_k$ must be different.
- Assume that there exists yet another connected component Γ_{k+1}. Then it belongs to only one half of $\Gamma^{\mathsf{C}}(\mathfrak{q})$. By symmetry, there is also the "same" component Γ_{k+2} in the other half of $\Gamma^{\mathsf{C}}(\mathfrak{q})$. This would imply that $\operatorname{ind}\mathfrak{q} > 0$.

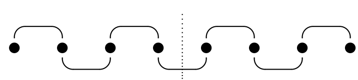
(2) Since the graph $\Gamma^{\mathsf{C}}(\mathfrak{q})$ has $2n$ vertices and is a disjoint union of k trees, the number of edges (arcs) must be $2n - k$. □

Lemma 4.4. *For any* $k \geqslant 1$, *there is an injective map* $\mathcal{F}_{n,k} \to \mathcal{F}_{n+1,k+1}$. *Moreover*, $\boldsymbol{F}_{n+1} > \boldsymbol{F}_n$; *that is*, *the total number of Frobenius seaweeds strictly increases under the passage from* n *to* $n+1$.

Proof. For any $\mathfrak{q} \in \mathcal{F}_{n,k}$ ($k \geqslant 1$), we can add two new vertices in the middle of $\Gamma^{\mathsf{C}}(\mathfrak{q})$ and connect them by an arc (on the appropriate side!). This yields an injective mapping $\mathcal{F}_{n,k} \to \mathcal{F}_{n+1,k+1}$ for any $k \geqslant 1$ and thereby an injection $i_n : \mathcal{F}_n \hookrightarrow \mathcal{F}_{n+1}$.

Since $\mathcal{F}_{n+1,1}$ does not intersect the image of i_n, the second assertion follows from the fact that $\boldsymbol{F}_{n+1,1} > 0$ for any $n \geqslant 0$; see the example below. □

Example. We point out an explicit element $\mathfrak{q}_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b}) \in \mathcal{F}_{n,1}$. For $n = 2k$, one takes $\underline{a} = (2^k)$ and $\underline{b} = (1, 2^{k-1})$. For $n = 2k+1$, one takes $\underline{a} = (2^k)$ and $\underline{b} = (1, 2^k)$. For $n = 4$, the meander graph is



Proposition 4.5. (i) *For a fixed* $m \in \mathbb{N}$, *the numbers* $\boldsymbol{F}_{n,n-m}$ *stabilise for* $n \geqslant 2m+1$. *In other words*, $\boldsymbol{F}_{n,n-m} = \boldsymbol{F}_{2m+1,m+1}$ *for all* $n \geqslant 2m+1$.

(ii) *Furthermore*, $\boldsymbol{F}_{2m+1,m+1} = \boldsymbol{F}_{2m,m} + 1$.

Proof. (i) Let $\mathfrak{q} = \mathfrak{q}_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b}) \in \mathcal{F}_{n,n-m}$. Then $\sum_{i=1}^{s} a_i = m$ and $\sum_{j=1}^{t} b_j = n$. Consider the n-th vertex of the graph (the one that is closest to the σ-mirror). We are interested in b_t, the size of the last part of $\underline{b}$, i.e., the part that contains the n-th vertex. By the assumption, we have $n - m$ central arcs over the horizontal line. Therefore, if $n \geqslant 2m+2$ and $b_t \geqslant 2$, then the smallest arc corresponding to b_t hits two vertices covered by central arcs above the line. And this produces a cycle in the graph! This contradiction shows that the only possibility is $b_t = 1$. Then one can safely remove two central vertices from the graph and conclude that $\boldsymbol{F}_{n,n-m} = \boldsymbol{F}_{n-1,n-1-m}$ as long as $n \geqslant 2m+2$. (The last step is opposite to one that is used in the proof of Lemma 4.4.)

(ii) Again, for $\mathfrak{q} = \mathfrak{q}^{\mathsf{C}}_{2m+1}(\underline{a}\,|\,\underline{b}) \in \mathcal{F}_{2m+1,m+1}$, we consider b_t, the last coordinate of $\underline{b}$. If $b_t = 1$, then the central pair of vertices in $\Gamma^{\mathsf{C}}(\mathfrak{q})$ can be removed, which yields a seaweed in $\mathcal{F}_{2m,m}$. Next, it is easily seen that if $b_t \in \{2, 3, \dots, 2m\}$, then $\Gamma^{\mathsf{C}}(\mathfrak{q})$ contains a cycle. Hence this is impossible. While for $b_t = 2m+1$, one obtains a unique admissible possibility

$$\underline{a} = (\underbrace{1, 1, \dots, 1}_{m}). \qquad \square$$

Remark. Using a similar analysis, one obtains $\boldsymbol{F}_{2m,m} = \boldsymbol{F}_{2m-1,m-1} + 3$ if $m \geqslant 3$.

Remark 4.6. Our stabilisation result for $\boldsymbol{F}_{n,n-m}$ can be compared with [Duflo and Yu 2015], where Duflo and Yu consider a partition of the set of standard Frobenius seaweeds in $\mathfrak{sl}_n$ into classes and study the asymptotic behaviour of the cardinality of these classes as n tends to infinity. Let $p(\underline{a})$ be the number of nonzero parts of the composition $\underline{a}$ and let $\tilde{F}_{n,p}$ be the number of the standard Frobenius seaweeds $\mathfrak{q}^{\mathsf{A}}(\underline{a}\,|\,\underline{b}) \cap \mathfrak{sl}_n$ such that $p(\underline{a}) + p(\underline{b}) = p$. By [Duflo and Yu 2015, Theorem 1.1(b)], if n is sufficiently large, then $\tilde{F}_{n,n+1-t}$ is a polynomial in n of degree $\left[\frac{1}{2}t\right]$, with positive rational coefficients.

It seems that $\mathcal{F}_{n,1}$ is the most interesting part of the symplectic Frobenius seaweeds. Recall from Section 2 that to any standard seaweed $\mathfrak{q} \subset \mathfrak{sp}_{2n}$ one can associate a "symmetric" seaweed $\tilde{\mathfrak{q}} \subset \mathfrak{gl}_{2n}$ such that $\mathfrak{q} = \tilde{\mathfrak{q}} \cap \mathfrak{sp}_{2n}$. In this context, we also set $\tilde{\mathfrak{q}}_0 = \tilde{\mathfrak{q}} \cap \mathfrak{sl}_{2n}$.

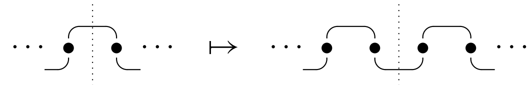
Proposition 4.7.

(i) *If* $\mathfrak{q} \in \mathcal{F}_{n,1}$, *then* $\operatorname{ind} \tilde{\mathfrak{q}} = 1$, *hence* $\tilde{\mathfrak{q}}_0$ *is a Frobenius seaweed in* $\mathfrak{sl}_{2n}$.

(ii) *There is an injective map* $\mathcal{F}_{n,1} \to \mathcal{F}_{n+1,1}$, *which is not onto if* $n \geqslant 2$.

Proof. (i) If $\mathfrak{q} \in \mathcal{F}_{n,1}$, then $\Gamma^{\mathsf{C}}(\mathfrak{q})$ and thereby $\Gamma^{\mathsf{A}}(\tilde{\mathfrak{q}})$ consists of a sole segment (Lemma 4.3). By (2-1), we have $\operatorname{ind} \tilde{\mathfrak{q}} = 1$ and therefore $\operatorname{ind} \tilde{\mathfrak{q}}_0 = \operatorname{ind} \tilde{\mathfrak{q}} - 1 = 0$.

(ii) If $\mathfrak{q} = \mathfrak{q}^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b}) \in \mathcal{F}_{n,1}$, then $\sum_{i=1}^{s} a_i = n-1$ and $\sum_{j=1}^{t} b_j = n$. We associate to it a seaweed $\hat{\mathfrak{q}} \in \mathcal{F}_{n+1,1}$ as follows. Set $\hat{\mathfrak{q}} = \mathfrak{q}^{\mathsf{C}}_{n+1}(\hat{\underline{a}}\,|\,\underline{b})$, where $\hat{\underline{a}} = (a_1, \dots, a_s, 2)$. Note that $\Gamma^{\mathsf{C}}_n(\underline{a}\,|\,\underline{b})$ has one central arc above the horizontal line, while $\Gamma^{\mathsf{C}}_{n+1}(\hat{\underline{a}}\,|\,\underline{b})$ has one central arc below. The following is a graphical illustration of the transform $\mathfrak{q} \mapsto \hat{\mathfrak{q}}$:



$n\downarrow$ $k\rightarrow$	1	2	3	4	5	6	7	$\Sigma=\boldsymbol{F}_n$
1	1	-	-	-	-	-	-	1
2	1	1	-	-	-	-	-	2
3	2	2	1	-	-	-	-	5
4	4	4	2	1	-	-	-	11
5	8	10	5	2	1	-	-	26
6	15	20	13	5	2	1	-	56
7	28	44	28	14	5	2	1	122

Table 1. The numbers $\boldsymbol{F}_{n,k}$ for $n \leqslant 7$.

This provides a bijection between $\mathcal{F}_{n,1}$ and the seaweeds in $\mathcal{F}_{n+1,1}$ whose last part of the composition that sums to $n+1$ equals 2. If $n+1 \geqslant 3$, then there are seaweeds in $\mathcal{F}_{n+1,1}$ such that the above-mentioned last part is bigger than 2. Hence $\boldsymbol{F}_{n,1} < \boldsymbol{F}_{n+1,1}$. □

Remark 4.8. Another curious observation is that $\mathcal{F}_{n,1}$ and $\mathcal{F}_{n,2}$ are related to certain Frobenius seaweeds in $\mathfrak{sl}_n$:

(i) Suppose that $\mathfrak{q} \in \mathcal{F}_{n,1}$. Let us remove the only central arc in $\Gamma^{\mathsf{C}}(\mathfrak{q})$ and take the remaining left half of the graph as it is. It is a *connected* type-A meander graph with n vertices. Therefore, it represents a seaweed of index 1 in $\mathfrak{gl}_n$ (= Frobenius seaweed in $\mathfrak{sl}_n$). Formally, if $\mathfrak{q} = \mathfrak{q}_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b})$, with $\sum a_i = n-1$ and $\sum b_j = n$, then we set $\mathfrak{q}' = \mathfrak{q}^{\mathsf{A}}(\underline{a}'\,|\,\underline{b}) \subset \mathfrak{sl}_n$, where $\underline{a}' = (\underline{a}, 1)$. This yields a bijection between $\mathcal{F}_{n,1}$ and the Frobenius seaweeds of $\mathfrak{sl}_n$ such that the last part of $\underline{a}'$ equals 1.

(ii) Suppose that $\mathfrak{q} \in \mathcal{F}_{n,2}$. Let us remove the two central arcs and take the remaining left half. We obtain a graph with n vertices and two connected components (segments). Joining the last two "lonely" vertices by an arc, we get a *connected* type-A meander graph. Formally, if $\mathfrak{q} = \mathfrak{q}_n^{\mathsf{C}}(\underline{a}\,|\,\underline{b})$, with $\sum a_i = n-2$ and $\sum b_j = n$, then we set $\mathfrak{q}' = \mathfrak{q}^{\mathsf{A}}(\underline{a}'\,|\,\underline{b}) \subset \mathfrak{sl}_n$, where $\underline{a}' = (\underline{a}, 2)$. Again, this yields a bijection between $\mathcal{F}_{n,2}$ and the Frobenius seaweeds of $\mathfrak{sl}_n$ such that the last part of $\underline{a}'$ equals 2.

Unfortunately, such a nice relationship does not extend to $\mathcal{F}_{n,3}$.

In Table 1 we present the numbers $\boldsymbol{F}_{n,k}$ for $n \leqslant 7$. Note that the values 14, 5, 2, 1 in the seventh row are stable in the sense of Proposition 4.5(i). Using the preceding information, we can also compute the next stable value:

$$\boldsymbol{F}_{9,5} = \boldsymbol{F}_{8,4} + 1 = (\boldsymbol{F}_{7,3} + 3) + 1 = 32.$$

5. On meander graphs for the odd orthogonal Lie algebras

As in the case of $\mathfrak{sp}_{2n}$, the standard parabolic subalgebras of $\mathfrak{so}_{2n+1}$ are parametrised by the compositions $\underline{a} = (a_1, \dots, a_s)$ such that $\sum a_i \leqslant n$. For instance, if $\mathfrak{p}_n^{\mathsf{B}}(\underline{a})$ is the standard parabolic subalgebra corresponding to $\underline{a}$, then a Levi subalgebra of it is of the form $\mathfrak{gl}_{a_1} \oplus \cdots \oplus \mathfrak{gl}_{a_s} \oplus \mathfrak{so}_{2(n-\sum a_i)+1}$. Therefore, the standard seaweed subalgebras of $\mathfrak{so}_{2n+1}$ are also parametrised by the pairs of compositions $\underline{a}$, $\underline{b}$ such that $\sum a_i \leqslant n$ and $\sum b_j \leqslant n$; see [Panyushev 2001, Section 5]. Furthermore, the inductive procedure for computing the index of standard seaweeds (see Section 3, Step 2), which reduces the case of arbitrary seaweeds to parabolic subalgebras, also remains the same [Panyushev 2001, Theorem 5.2].

This means that if the formula for the index of parabolic subalgebras of $\mathfrak{so}_{2n+1}$ in terms of $\underline{a}$ also remains the "same" as in the symplectic case, then one can use our type-C meander graphs in type B_n as well. Although there are only partial results on the index of parabolic subalgebras of $\mathfrak{so}_{2n+1}$ in [Panyushev 2001, Section 6], one can use the general Tauvel–Yu–Joseph formula; see [Tauvel and Yu 2004, Conjecture 4.7; Joseph 2006, Section 8]. Namely, if $\mathfrak{q} = \mathfrak{q}(S, T)$ is the seaweed corresponding to the subsets $S, T \subset \Pi$, then

$$\text{(5-1)} \qquad \operatorname{ind} \mathfrak{q} = \operatorname{rk} \mathfrak{g} + \dim E_S + \dim E_T - 2 \dim(E_S + E_T).$$

Here $\dim E_T = \#\mathcal{K}(T)$ is the cardinality of the cascade of strongly orthogonal roots in the Levi subalgebra of $\mathfrak{g}$ corresponding to T; see [Tauvel and Yu 2004] for the details. Our observation is that it easily implies that, for any composition $\underline{a}$, one has

$$\text{(5-2)} \qquad \operatorname{ind} \mathfrak{p}_n^{\mathsf{B}}(\underline{a}) = \left[\tfrac{1}{2}a_1\right] + \cdots + \left[\tfrac{1}{2}a_s\right] + \left(n - \sum_{i=1}^{s} a_i\right) = \operatorname{ind} \mathfrak{p}_n^{\mathsf{C}}(\underline{a}).$$

Indeed, for the parabolic subalgebras, we may assume that $S = \Pi$, and since $\operatorname{ind} \mathfrak{b} = 0$ for the series B_n, we have $\dim E_\Pi = \operatorname{rk} \mathfrak{g}$. Therefore, $\operatorname{ind} \mathfrak{p}_n^{\mathsf{B}}(\underline{a}) = \dim E_T = \#\mathcal{K}(T)$. As we noticed before, for $\mathfrak{p}_n^{\mathsf{B}}(\underline{a})$, we have $\mathfrak{l} = \mathfrak{gl}_{a_1} \oplus \cdots \oplus \mathfrak{gl}_{a_s} \oplus \mathfrak{so}_{2(n-\sum a_i)+1}$. As is well known, the cardinality of the cascade of strongly orthogonal roots in $\mathfrak{gl}_a$ (resp. $\mathfrak{so}_{2n+1}$) equals $\left[\tfrac{1}{2}a\right]$ (resp. n); see [Joseph 1977, Section 2]. Therefore, the cardinality of the cascade in the above $\mathfrak{l}$ is given by the middle term in (5-2).

There is another interesting formula for the index of a parabolic subalgebra, which generalises the above observation.

Theorem 5.1. *Let* $\mathfrak{g}$ *be a simple Lie algebra such that* $\operatorname{ind} \mathfrak{b} = 0$. *Let* $\mathfrak{p} \subset \mathfrak{g}$ *be a parabolic subalgebra, with a Levi subalgebra* $\mathfrak{l}$. *If* $\mathfrak{b}(\mathfrak{l})$ *is a Borel subalgebra of* $\mathfrak{l}$ *and* $\mathfrak{u}(\mathfrak{l}) = [\mathfrak{b}(\mathfrak{l}), \mathfrak{b}(\mathfrak{l})]$, *then*

$$\text{(5-3)} \qquad \operatorname{ind} \mathfrak{p} = \operatorname{ind} \mathfrak{u}(\mathfrak{l}) = \operatorname{rk} \mathfrak{l} - \operatorname{ind} \mathfrak{b}(\mathfrak{l}) = \operatorname{rk} \mathfrak{g} - \operatorname{ind} \mathfrak{b}(\mathfrak{l}).$$

In particular, $\operatorname{ind} \mathfrak{p} = 0$ *if and only if* $\mathfrak{u}(\mathfrak{l}) = 0$, *i.e.*, $\mathfrak{p} = \mathfrak{b}$.

Outline of the proof. Again, under the assumption that $\operatorname{ind} \mathfrak{b} = 0$, we have $S = \Pi$, $\dim E_\Pi = \operatorname{rk} \mathfrak{g}$, and $\mathfrak{l}$ is determined by T. Hence (5-1) implies that $\operatorname{ind} \mathfrak{p} = \dim E_T = \#\mathcal{K}(T)$. It is implicit in [Joseph 1977, Section 2.6] that $\#\mathcal{K}(T) = \operatorname{ind} \mathfrak{u}(\mathfrak{l})$, and the second equality in (5-3) is a consequence of the fact that $\mathsf{rk}\,\mathfrak{l} = \mathsf{ind}\,\mathfrak{b}(\mathfrak{l}) + \mathsf{ind}\,\mathfrak{u}(\mathfrak{l})$ for any reductive Lie algebra $\mathfrak{l}$. A more detailed explanation and some applications of the theorem will appear elsewhere. □

Recall that, for a simple Lie algebra $\mathfrak{g}$, $\mathsf{ind}\,\mathfrak{b} = 0$ if and only if $\mathfrak{g} \neq \mathsf{A}_n, \mathsf{D}_{2n+1}, \mathsf{E}_6$.

Conclusion. (1) Given a standard seaweed $\mathfrak{q} = \mathfrak{q}_n^{\mathsf{B}}(\underline{a}\,|\,\underline{b}) \subset \mathfrak{so}_{2n+1}$, we can draw exactly the same meander graph as in type C (with $2n$ vertices) and use exactly the same topological formula (Theorem 3.2) to compute the index of $\mathfrak{q}$.

(2) Using our type-C meander graphs, we can establish a bijection between the standard Frobenius seaweeds for the symplectic and odd orthogonal Lie algebras of the same rank. It would be very interesting to realise whether there is a deeper reason for such a bijection.

(3) For the even-dimensional orthogonal Lie algebras (type D_n), there is a similar inductive procedure that reduces the problem of computing the index of arbitrary seaweeds to parabolic subalgebras. However, $\mathsf{ind}\,\mathfrak{b} = 1$ for D_{2n+1} and Theorem 5.1 does not apply. Furthermore, although $\mathsf{ind}\,\mathfrak{b} = 0$ for D_{2n}, the general formula for the index of parabolic subalgebras cannot be expressed nicely in terms of compositions. Of course, the reason is that the Dynkin diagram has a branching node!

References

[Dergachev and Kirillov 2000] V. Dergachev and A. Kirillov, "Index of Lie algebras of seaweed type", *J. Lie Theory* **10**:2 (2000), 331–343. MR 1774864 Zbl 0980.17001

[Duflo and Yu 2015] M. Duflo and R. W. T. Yu, "On compositions associated to Frobenius parabolic and seaweed subalgebras of $\mathrm{sl}_n(\Bbbk)$", *J. Lie Theory* **25**:4 (2015), 1191–1213. MR 3350093 Zbl 06556746

[Duflo et al. 2012] M. Duflo, M. S. Khalgui, and P. Torasso, "Algèbres de Lie quasi-réductives", *Transform. Groups* **17**:2 (2012), 417–470. MR 2921072 Zbl 1307.17012

[Joseph 1977] A. Joseph, "A preparation theorem for the prime spectrum of a semisimple Lie algebra", *J. Algebra* **48**:2 (1977), 241–289. MR 0453829 Zbl 0405.17007

[Joseph 2006] A. Joseph, "On semi-invariants and index for biparabolic (seaweed) algebras, I", *J. Algebra* **305**:1 (2006), 487–515. MR 2264140 Zbl 1183.17009

[Moreau and Yakimova 2012] A. Moreau and O. Yakimova, "Coadjoint orbits of reductive type of parabolic and seaweed Lie subalgebras", *Int. Math. Res. Not.* **2012**:19 (2012), 4475–4519. MR 2981717 Zbl 1315.17006

[Panyushev 2001] D. I. Panyushev, "Inductive formulas for the index of seaweed Lie algebras", *Mosc. Math. J.* **1**:2 (2001), 221–241, 303. MR 1878277 Zbl 0998.17008

[Panyushev 2005] D. I. Panyushev, "An extension of Raïs' theorem and seaweed subalgebras of simple Lie algebras", *Ann. Inst. Fourier* (*Grenoble*) **55**:3 (2005), 693–715. MR 2149399 Zbl 1151.17305

[Tauvel and Yu 2004] P. Tauvel and R. W. T. Yu, "Sur l'indice de certaines algèbres de Lie", *Ann. Inst. Fourier* (*Grenoble*) **54**:6 (2004), 1793–1810. MR 2134224 Zbl 1137.17300

Received December 29, 2015. Revised May 5, 2016.

DMITRI I. PANYUSHEV
DOBRUSHIN MATHEMATICS LABORATORY
INSTITUTE FOR INFORMATION TRANSMISSION PROBLEMS
RUSSIAN ACADEMY OF SCIENCES
BOLSHOI KARETNYI PER. 19
MOSCOW
127051
RUSSIA
panyushev@iitp.ru

OKSANA S. YAKIMOVA
INSTITUT FÜR MATHEMATIK
FRIEDRICH-SCHILLER-UNIVERSITÄT JENA
D-07737 JENA
GERMANY
oksana.yakimova@uni-jena.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 285, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.285.501

THE GENUS FILTRATION IN THE SMOOTH CONCORDANCE GROUP

SHIDA WANG

We define a filtration of the smooth concordance group based on the genus of representative knots. We use the Heegaard Floer ε- and Υ-invariants to prove the quotient groups with respect to this filtration are infinitely generated. Results are applied to three infinite families of topologically slice knots.

1. Introduction

Let $\mathcal{C}$ be the smooth concordance group. Let $\mathcal{G}_k$ denote the subgroup of $\mathcal{C}$ generated by knots of genus not greater than k. Clearly $\mathcal{G}_0 \subseteq \mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \cdots \subseteq \mathcal{G}_k \subseteq \cdots \subseteq \mathcal{C}$ and $\bigcup_{k=1}^{\infty} \mathcal{G}_k = \mathcal{C}$. This gives a filtration of $\mathcal{C}$. We call it the *genus filtration*.

There is another way to understand $\mathcal{G}_k$. Recall that the concordance genus g_c of a knot K is defined to be the minimal genus of a knot K' concordant to K. It is obvious that $g_c(K) = \min\{k \mid K \text{ is concordant to } K' \text{ and } g(K') \leq k\}$. This motivates the following definition.

Definition 1.1. The *splitting concordance genus* of a knot K is

$$g_{\mathrm{sp}}(K) := \min\{k \mid K \text{ is concordant to } K_1\#\cdots\#K_m \text{ for some } m \text{ and } g(K_1), \ldots, g(K_m) \leq k\}.$$

That is to say, $g_{\mathrm{sp}}(K)$ is the filtration level of K in $\mathcal{G}_0 \subseteq \mathcal{G}_1 \subseteq \cdots \subseteq \mathcal{G}_k \subseteq \cdots \subseteq \mathcal{C}$. By [Endo 1995], $\mathcal{G}_1$ contains a $\mathbb{Z}^{\infty}$ subgroup whose elements are topologically slice.

Let $\mathcal{C}_{TS} \subseteq \mathcal{C}$ be the subgroup of topologically slice knots. Recently several results have appeared which reveal that the group $\mathcal{C}_{TS}$ is quite large. For example, in [Hom 2015a; Ozsvath et al. 2014] it is shown that $\mathcal{C}_{TS}$ contains $\mathbb{Z}^{\infty}$ as a direct summand. In [Hedden et al. 2012] it is shown that $\mathcal{C}_{TS}$ contains $\mathbb{Z}^{\infty}$ as a subgroup whose nonzero elements are not concordant to knots of Alexander polynomial one. In [Hedden et al. 2016] it is shown that $\mathcal{C}_{TS}$ contains $\mathbb{Z}_2^{\infty}$ as a subgroup whose nonzero elements are not concordant to knots of Alexander polynomial one.

MSC2010: primary 57N70; secondary 57M25.
Keywords: knot concordance, genus, Heegaard Floer homology.

We will show that $\mathcal{C}_{TS}$ is large in another sense. We will prove $\mathcal{C}_{TS} \not\subseteq \mathcal{G}_k$ for any k. Moreover, the difference between $\mathcal{C}_{TS}$ and $\mathcal{G}_k$ is large. Corollary 1.4 states that $\mathcal{C}_{TS}/(\mathcal{C}_{TS} \cap \mathcal{G}_k)$ contains a direct summand isomorphic to $\mathbb{Z}^\infty$.

Our examples will be built from those of [Hom 2014a; Ozsvath et al. 2014]. Let $\mathrm{Wh}(K)$ denote the untwisted Whitehead double of a knot K. Additionally let $K_{p,q}$ denote the (p,q)-cable of K, let $J_n = (\mathrm{Wh}(T_{2,3}))_{n,n+1} \# -T_{n,n+1}$, and let $J'_n = (\mathrm{Wh}(T_{2,3}))_{n,2n-1} \# -T_{n,2n-1}$. These knots are topologically slice and used to prove the following theorems.

Theorem 1 [Hom 2015a, Theorem 1]. *The group $\mathcal{C}_{TS}$ contains a summand which is isomorphic to $\mathbb{Z}^\infty$ and generated by $\{J_n\}_{n=2}^\infty$.*

Theorem 2 [Ozsvath et al. 2014, Theorem 1.20]. *The topologically slice knots $\{J'_n\}_{n=2}^\infty$ form a basis for a free direct summand of $\mathcal{C}_{TS}$.*

We will prove the following results.

Theorem 1.2. *$\{J_n\}_{n=k}^\infty$ forms a basis for a $\mathbb{Z}^\infty$ summand of $\mathcal{C}_{TS}/(\mathcal{C}_{TS} \cap \mathcal{G}_{\lfloor k/2 \rfloor})$ for any $k \geq 2$.*

Theorem 1.3. *$\{J'_n\}_{n=k}^\infty$ forms a basis for a $\mathbb{Z}^\infty$ summand of $\mathcal{C}_{TS}/(\mathcal{C}_{TS} \cap \mathcal{G}_{k-1})$ for any $k \geq 2$.*

Hence we have the following consequence.

Corollary 1.4. *For any $k \in \mathbb{N}$ we have $\mathcal{C}_{TS} \not\subseteq \mathcal{G}_k$. Moreover, the quotient group $\mathcal{C}_{TS}/(\mathcal{C}_{TS} \cap \mathcal{G}_k)$ contains a direct summand isomorphic to $\mathbb{Z}^\infty$.*

One can define another subgroup $\mathcal{H}_k$ of $\mathcal{C}$ generated by knots of 4-genus not greater than k. Clearly $\mathcal{G}_k \subseteq \mathcal{H}_k$. It is natural to ask whether $\mathcal{H}_k/\mathcal{G}_k$ is infinitely generated. We show the answer is affirmative by proving the following:

Theorem 1.5. *The quotient group $\mathcal{C}_{TS}/(\mathcal{C}_{TS} \cap \mathcal{G}_k)$ contains a subgroup isomorphic to $\mathbb{Z}^\infty$ whose basis elements have slice genus 1 for any $k \geq 2$.*

Conjecture 1.6. (1) For any $k \in \mathbb{N}$, the quotient $(\mathcal{C}_{TS} \cap \mathcal{G}_{k+1})/(\mathcal{C}_{TS} \cap \mathcal{G}_k)$ contains a direct summand isomorphic to $\mathbb{Z}^\infty$ whose basis elements have slice genus 1.

(2) For any $k \in \mathbb{N}$, the group $\mathcal{C}/\mathcal{H}_n$ is nontrivial.

This paper is organized as follows. In Section 2 we use Alexander polynomials to prove the splitting concordance genus can be arbitrarily large. In Section 3 we review Hom's ε-invariant and develop an obstruction, which is used to prove Theorems 1.2 and 1.5. In Section 4 we use the Υ-invariant to develop an obstruction and prove Theorem 1.3.

2. A first glance at the genus filtration

Given a knot K, let $\Delta_K(t)$ be its Alexander polynomial, and breadth$(\Delta_K(t))$ be the maximal exponent of $\Delta_K(t)$ minus the minimal exponent of $\Delta_K(t)$. Recall that for any K, breadth$(\Delta_K(t)) \leq 2g(K)$. Moreover, if K is slice, recall that it must satisfy the Fox–Milnor condition, factoring as $t^{\pm n} f(t) f(t^{-1})$. Based on these facts, we can prove the following theorem, generalizing [Livingston 2004, Theorem 2.2].

Proposition 2.1. *For any knot K, if $p(t)$ appears an odd number of times in the irreducible factorization of $\Delta_K(t)$ in $\mathbb{Z}[t, t^{-1}]$, then*

$$g_{\mathrm{sp}}(K) \geq \tfrac{1}{2}\,\mathrm{breadth}(p(t)).$$

Proof. By definition, we can choose knots $K_1, \ldots, K_m$ such that K is concordant to $K_1 \# \cdots \# K_m$ and $g(K_i) \leq g_{\mathrm{sp}}(K)$ for each $1 \leq i \leq m$. Thus $K \# -K_1 \# \cdots \# -K_m$ is a slice knot and its Alexander polynomial $\Delta_K(t)\Delta_{K_1}(t) \cdots \Delta_{K_m}(t)$ must factor as $t^{\pm n} f(t) f(t^{-1})$ for some $f \in \mathbb{Z}[t, t^{-1}]$. If some $p(t)$ appears an odd number of times in the irreducible factorization of $\Delta_K(t)$, it must appear in the irreducible factorization of one of $\Delta_{K_1}(t), \ldots, \Delta_{K_m}(t)$. Since $2g_{\mathrm{sp}}(K) \geq \mathrm{breadth}(\Delta_{K_i}(t))$ for each $1 \leq i \leq m$, we conclude that $2g_{\mathrm{sp}}(K) \geq \mathrm{breadth}(p(t))$. □

Example 2.2. The Alexander polynomial of the torus knot $T_{p,q}$ is

$$\Delta_{T_{p,q}}(t) = ((t^{pq} - 1)(t - 1))/((t^p - 1)(t^q - 1)),$$

in whose irreducible factorization the cyclotomic polynomial Φ_{pq} appears exactly once. Hence $g_{\mathrm{sp}}(T_{p,q}) \geq \varphi(pq)/2$, where φ is Euler's totient function. If p and q are prime, we have $g_{\mathrm{sp}}(T_{p,q}) \geq ((p-1)(q-1))/2$. This is actually an equality, because $g(T_{p,q}) = ((p-1)(q-1))/2$.

Corollary 2.3. *$\mathcal{C}/\mathcal{G}_k$ is nontrivial for any $k \in \mathbb{N}$.*

Working a little harder, we can show the following.

Proposition 2.4. *$\mathcal{C}/\mathcal{G}_k$ contains an infinitely generated free subgroup for any $k \in \mathbb{N}$.*

Proof. Let $\{p_n\}_{n=1}^{\infty}$ be a sequence of strictly increasing prime numbers with $p_1 > k$. We will prove that the torus knots $\{T_{p_{2n-1},p_{2n}}\}_{n=1}^{\infty}$ are linearly independent in $\mathcal{C}/\mathcal{G}_k$.

Suppose towards a contradiction that $\#_{i=1}^{l} c_i T_{p_{2n_i-1},p_{2n_i}}$, where $0 < n_1 < \cdots < n_l$ and $c_1, \ldots, c_l$ are nonzero integers, is concordant to $K_1 \# \cdots \# K_m$ with $g(K_j) \leq k$ for $1 \leq j \leq m$. Notice that $\Delta_{T_{p_{2n_i-1},p_{2n_i}}}(t) = \Phi_{p_{2n_i-1}p_{2n_i}}$, where $\Phi_{p_{2n_i-1}p_{2n_i}}$ is the cyclotomic polynomial, which is irreducible of degree $(p_{2n_i-1} - 1)(p_{2n_i} - 1)$. By a combinatorial formula [Litherland 1979, Proposition 1] for the Tristram–Levine signature functions of torus knots, $\sigma_\omega(T_{p_{2n_i-1},p_{2n_i}})$ jumps by ± 2 at the primitive $(p_{2n_i-1}p_{2n_i})$-th roots of unity. Since the products $p_{2n_i-1}p_{2n_i}$ are distinct for $i = 1, \ldots, l$, we know $\sigma_\omega(\#_{i=1}^{l} c_i T_{p_{2n_i-1},p_{2n_i}})$ has a jump discontinuity at a primitive $(p_{2n_1-1}p_{2n_1})$-th root of unity. Hence $\sigma_\omega(K_1\#\cdots\#K_m)$ also has a jump discontinuity at a primitive $(p_{2n_1-1}p_{2n_1})$-th root of unity, and so does one of $\sigma_\omega(K_1),\ldots,\sigma_\omega(K_m)$. Without loss of generality, assume that $\sigma_\omega(K_1)$ has a jump discontinuity at a primitive $(p_{2n_1-1}p_{2n_1})$-th root of unity. Since jump discontinuities of the Tristram–Levine signature function can only appear at roots of the Alexander polynomial, it follows that $\Delta_{K_1}(t)$ has a root at a primitive $(p_{2n_1-1}p_{2n_1})$-th root of unity and thus is divisible by $\Phi_{p_{2n_i-1}p_{2n_i}}$, but this is impossible because

$$\deg\Delta_{K_1}(t)\leq 2g(K_1)\leq 2k<(p_{2n_i-1}-1)(p_{2n_i}-1). \qquad \square$$

3. Obstruction by ε-invariant

We assume the reader is familiar with knot Floer homology defined by Ozsváth and Szabó [2004b] and independently Rasmussen [2003] and the ε-invariant defined by Hom [2014a]. We briefly recall some of their properties for later use.

The knot Floer complex and ε-invariant. The knot Floer complex associates to a knot $K\subset S^3$ a doubly filtered, free, finitely generated chain complex over $\mathbb{F}[U,U^{-1}]$, denoted by $CFK^\infty(K)$, where $\mathbb{F}$ is the field with two elements. The two filtrations are called the algebraic and Alexander filtrations and the grading of the chain complex is called the homological or Maslov grading. Multiplication by U shifts each filtration down by one and lowers the homological grading by two. $CFK^\infty(K)$ is an invariant of K up to filtered chain homotopy equivalence. Furthermore, up to filtered chain homotopy equivalence, one can assume the differential strictly lowers at least one of the filtrations [Rasmussen 2003].

A quick corollary from [Ozsváth and Szabó 2004a, Theorem 1.2] is the following.

Proposition 3.1. *If K has genus g, then there exists a representative of the filtered chain homotopy equivalence class of $CFK^\infty(K)$ all of whose elements have filtration levels (i,j) such that $-g\leq i-j\leq g$.*

For a subset $S\subseteq\mathbb{Z}\oplus\mathbb{Z}$ that is downward closed under the standard product partial order on $\mathbb{Z}\oplus\mathbb{Z}$, let $C\{S\}$ denote the subcomplex of $CFK^\infty(K)$ generated by elements with filtration levels in S. If S is the difference of two such subsets, let $C\{S\}$ denote the corresponding subquotient complex of $CFK^\infty(K)$. For example, $C\{i=0\}=C\{i\leq 0\}/C\{i<0\}=CFK^\infty(K)\{i\leq 0\}/CFK^\infty(K)\{i<0\}$. The invariant

$$\tau(K)=\min\{s\mid \text{the inclusion map } C\{i=0,\ j\leq s\}\to C\{i=0\} \text{ induces a nontrivial map on homology}\}$$

is proven to be a smooth concordance invariant in [Ozsváth and Szabó 2003].

For any knot K, Hom [2014a] defines an invariant called ε taking on values -1, 0 or 1, which has the following properties.

Proposition 3.2 [Hom 2014a, Proposition 3.6]. *The invariant ε satisfies the following properties*:

(1) *If K is smoothly slice, then* $\varepsilon(K) = 0$.

(2) $\varepsilon(-K) = -\varepsilon(K)$.

(3) *If* $\varepsilon(K) = \varepsilon(K')$, *then* $\varepsilon(K \# K') = \varepsilon(K) = \varepsilon(K')$.

(4) *If* $\varepsilon(K) = 0$, *then* $\varepsilon(K \# K') = \varepsilon(K')$.

Thus the relation $\sim$, defined by $K \sim K' \Leftrightarrow \varepsilon(K \# - K') = 0$, is an equivalence relation coarser than smooth concordance. It gives an equivalence relation on $\mathcal{C}$ called ε-equivalence. The ε-equivalence class of K is denoted by $[\![K]\!]$. The set of all ε-equivalence classes forms a group $\mathcal{F}$ (also denoted by $\mathcal{CFK}$ in [Hom 2015a]), which is a quotient group of $\mathcal{C}$. The kernel of the natural homomorphism from $\mathcal{C}$ to $\mathcal{F}$ is $\{[K] \in \mathcal{C} \mid \varepsilon(K) = 0\}$, where $[K]$ denotes the concordance class of K.

According to [Hom 2014b, Proposition 4.1], ε induces a total order on $\mathcal{F}$. The proof uses Proposition 3.2. The total order is defined by

$$[\![K]\!] > [\![K']\!] \Leftrightarrow \varepsilon(K \# - K') = 1.$$

Moreover, this order respects the addition operation on $\mathcal{F}$. Therefore there is a quotient homomorphism from $\mathcal{C}$ to the totally ordered abelian group $\mathcal{F}$, which can be used to show linear independence in $\mathcal{C}$.

Some facts about totally ordered abelian groups. Let G be a totally ordered abelian group, that is an abelian group with a total order respecting the addition operation. Denote its identity element by 0.

The absolute value of an element $a \in G$ is defined to be

$$|a| = \begin{cases} a & \text{if } a \geq 0, \\ -a & \text{if } a < 0. \end{cases}$$

Definition 3.3. Two nonzero elements a and b of G are *Archimedean equivalent*, denoted by $a \sim_A b$, if there exists a natural number N such that $N \cdot |a| > |b|$ and $N \cdot |b| > |a|$. If a and b are not Archimedean equivalent and $|a| < |b|$, we say that b dominates a. We write $a \ll b$ if $a > 0$, $b > 0$ and b dominates a.

Property A. An element $a \in G$ *satisfies Property A* if for every $b \in G$ such that $b \sim_A a$, we have that $b = ka + c$, where k is an integer and c is dominated by a.

We have the following two facts:

Lemma 3.4 [Hom 2014b, Lemma 4.7]. *If* $0 < a_1 \ll a_2 \ll a_3 \ll \cdots$ *in* G, *then* $a_1, a_2, a_3, \ldots$ *are linearly independent in* G.

Lemma 3.5 [Hom 2015a, Proposition 1.3]. *If $0 < a_1 \ll a_2 \ll a_3 \ll \cdots$ in G and each a_i satisfies Property A, then $a_1, a_2, a_3, \ldots$ generate (as a basis) a direct summand isomorphic to $\mathbb{Z}^\infty$ in G.*

The following lemmas are proven in [Hom 2014b] and [Hom 2015a] respectively.

Lemma 3.6 [Hom 2014b, Remark 4.9]. *We have $0 < [\![J_n]\!] \ll [\![J_{n+1}]\!]$ for any $n \geq 2$.*

Lemma 3.7 [Hom 2015a, Proposition 4.1, Lemmas 5.2 and 5.3]. *The class $[\![J_n]\!]$ satisfies Property A for any $n \geq 2$.*

It is straightforward to check that $\{a : |a| \ll x\}$ is a subgroup of G for any $x > 0$ in G. Denote this subgroup by G_x. Let φ_x be the quotient homomorphism. Define a relation $<$ in G/G_x by $\varphi_x(a) < \varphi_x(b)$ if and only if $a < b$ and $b - a \notin G_x$.

Proposition 3.8. *The relation $<$ makes G/G_x into a totally ordered abelian group with the following properties: If $0 < a \ll b$ in G and $b \notin G_x$, then $0 \leq \varphi_x(a) \ll \varphi_x(b)$ in G/G_x. If a satisfies Property A in G, then $\varphi_x(a)$ satisfies Property A in G/G_x.*

Proof. First we check that the relation $<$ in G/G_x is well defined. Suppose $\varphi_x(a) < \varphi_x(b)$. Let $c \in G_x$. We must show $\varphi_x(a+c) < \varphi_x(b)$ and $\varphi_x(a) < \varphi_x(b+c)$. Since $b - a > 0$ and $b - a \notin G_x$ it is easy to verify that $b - a \gg |y|$ for any $y \in G_x$. Thus $b - a \pm c > 0$. Additionally $b - a \notin G_x$ implies $b - a \pm c \notin G_x$. Hence $\varphi_x(a + c) < \varphi_x(b)$ and $\varphi_x(a) < \varphi_x(b + c)$, which means the definition does not depend on the choices of a and b.

Next we verify $<$ is a strict total order on G/G_x that respects the addition operation. For trichotomy, let $\varphi_x(a)$ and $\varphi_x(b)$ be two distinct elements in G/G_x. Then $b - a \notin G_x$. Thus $b - a \neq 0$ and exactly one of $a < b$ and $b < a$ is true. Hence exactly one of $\varphi_x(a) < \varphi_x(b)$ and $\varphi_x(b) < \varphi_x(a)$ is true by definition. For transitivity, let $\varphi_x(a), \varphi_x(b), \varphi_x(c) \in G/G_x$ satisfy $\varphi_x(a) < \varphi_x(b)$ and $\varphi_x(b) < \varphi_x(c)$. Then $a < b$, $b < c$ and $b - a, c - b \notin G_x$. Immediately $a < c$. Suppose towards a contradiction that $c - a \in G_x$. Then the fact that $b - a \gg |y|$ for any $y \in G_x$ implies $b - a - (c - a) > 0$, which contradicts $b < c$. Hence $c - a \notin G_x$ and $\varphi_x(a) < \varphi_x(c)$ by definition. For consistency with the addition operation, let $\varphi_x(a)$, $\varphi_x(b)$, $\varphi_x(c) \in G/G_x$ and $\varphi_x(a) < \varphi_x(b)$. Then $a < b$ and $b - a \notin G_x$. Thus $a + c < b + c$ and $(b + c) - (a + c) \notin G_x$. Hence $\varphi_x(a) + \varphi_x(c) = \varphi_x(a + c) < \varphi_x(b + c) = \varphi_x(b) + \varphi_x(c)$ by definition.

Next, we show that if b dominates a in G and $b \notin G_x$, then $\varphi_x(b)$ dominates $\varphi_x(a)$. Suppose $0 < a \ll b$ in G and $b \notin G_x$. Then $0 < Na < b$ for any $N \in \mathbb{N}$. Additionally, the fact that $b \gg |y|$ for any $y \in G_x$ implies $Na + y < b$, $\forall y \in G_x$. It follows that $b - Na > 0$ and that $b - Na \notin G_x$. Hence $0 \leq \varphi_x(a) \ll \varphi_x(b)$ in G/G_x by definition.

Finally we show that if a has Property A in G, then $\varphi_x(a)$ has Property A in G/G_x. Suppose a satisfies Property A in G, that is, if $b \sim_A a$ in G then $b = ka + c$ for some integer k and some $c \in G$ dominated by a. Without loss of generality we assume $\varphi_x(a) \neq 0$. Let $\varphi_x(b) \sim_A \varphi_x(a)$, so $b \sim_A a$ in G. Otherwise either $|a| \ll |b|$ or $|b| \ll |a|$, which would imply $|\varphi_x(a)| \ll |\varphi_x(b)|$ or $|\varphi_x(b)| \ll |\varphi_x(a)|$. Thus $b = ka + c$ for some integer k and some $c \in G$ dominated by a. Thus $\varphi_x(b) = k\varphi_x(a) + \varphi_x(c)$. Since c is dominated by a, we know $\varphi_x(c)$ is dominated by $\varphi_x(a)$. Hence $\varphi_x(a)$ satisfies Property A in G/G_x. □

Restriction on the Archimedean equivalence class by genus. Given a knot K with $\varepsilon(K) = 1$, Hom [2015a, Section 3] defines a tuple of numerical invariants $\boldsymbol{a}^+(K) = (a_1(K), \dots, a_n(K))$. Here each $a_i(K)$ is a positive integer, and the number n depends on K. It is shown that $\boldsymbol{a}^+(K)$ is an invariant of the ε-equivalence class $[\![K]\!]$ (see [Hom 2015a, Proposition 3.1]).

Computations in [Hom 2014b] show the following result.

Lemma 3.9 [Hom 2014b, p.568]. *We have* $\boldsymbol{a}^+(J_p) = (1, p, \dots)$.

The integers a_1 and a_2 are useful in determining domination.

Lemma 3.10 [Hom 2014b, Lemmas 6.3 and 6.4]. *If* $\boldsymbol{a}^+(K) = (a_1(K), \dots)$ *and* $\boldsymbol{a}^+(K') = (a_1(K'), \dots)$ *with* $a_1(K) > a_1(K') > 0$, *then* $[\![K]\!] \ll [\![K']\!]$.

Additionally, if $\boldsymbol{a}^+(K) = (a_1(K), a_2(K), \dots)$ *and* $\boldsymbol{a}^+(K') = (a_1(K'), a_2(K'), \dots)$ *with* $a_1(K) = a_1(K') > 0$ *and* $a_2(K) > a_2(K') > 0$, *then* $[\![K]\!] \gg [\![K']\!]$.

Based on Proposition 3.1, the following is shown.

Lemma 3.11 ([Hom 2015b, Theorem 1.2 and Lemma 2.3]). *Suppose that* $\varepsilon(K) = 1$, *and* $a_2(K)$ *is defined, then* $|\tau(K) - a_1(K) - a_2(K)| \leq g(K)$.

Next we prove our obstruction theorem.

Proposition 3.12. *Suppose J is a knot with $\boldsymbol{a}^+(J) = (1, b, \dots)$ with $b \geq 2n$ for some positive integer n. Then for any knot $K \in \mathcal{G}_n$, we have $|[\![K]\!]| \ll [\![J]\!]$.*

Proof. Before proving the proposition for $K \in \mathcal{G}_n$, first consider the case $g(K) \leq n$. We may further assume that $[\![K]\!] > 0$, since $[\![-K]\!] > 0$ if $[\![K]\!] < 0$ and the proposition is trivial if $[\![K]\!] = 0$. Notice that $a_1(K)$ is always defined [Hom 2014b, §6]. If $a_1(K) > 1$, then $[\![K]\!] \ll [\![J]\!]$ by Lemma 3.10. If $a_1(K) = 1$, then $a_2(K)$ is defined [Hom 2015a, Lemma 3.7]. Observe that $\tau(K) - a_1(K) - a_2(K) \geq -g(K)$ by Lemma 3.11. Combining this with $\tau(K) \leq g_4(K) \leq g(K)$, it follows that $g(K) - a_1(K) - a_2(K) \geq -g(K)$. This implies $a_2(K) \leq 2n - 1$, if $a_1(K) = 1$. Hence $|[\![K]\!]| \ll [\![J]\!]$ by Lemma 3.10.

Generally, let $K \in \mathcal{G}_n$. Then $K = K_1 + \cdots + K_m$, where $g(K_i) \leq n$ for $i = 1, \dots, m$. Since $[\![K]\!] = [\![K_1]\!] + \cdots + [\![K_m]\!]$, we know $|[\![K]\!]| \leq |[\![K_1]\!]| + \cdots + |[\![K_m]\!]|$. Then the conclusion follows from the last paragraph. □

Applying the obstruction to concrete families of knots.

Proof of Theorem 1.2. Fix an integer $k \geq 2$. Under the quotient homomorphism from $\mathcal{C}$ to $\mathcal{F}$, the image of $\mathcal{G}_{\lfloor k/2 \rfloor}$ is included in $\mathcal{F}_{[\![J_k]\!]} = \{[\![K]\!] : |[\![K]\!]| \ll [\![J_k]\!]\}$ by Proposition 3.12 and Lemma 3.9. This gives a homomorphism from $\mathcal{C}/\mathcal{G}_{\lfloor k/2 \rfloor}$ to $\mathcal{F}/\mathcal{F}_{[\![J_k]\!]}$. By Lemma 3.6 and Proposition 3.8, the family $\{J_n\}_{n=k}^{\infty}$ maps to a family of elements with Property A and each term is dominated by the next. Hence $\{J_n\}_{n=k}^{\infty}$ forms a basis of a direct summand isomorphic to $\mathbb{Z}^{\infty}$ by Lemma 3.5. Note that since the J_n are topologically slice, the above argument can be restricted to the subgroup $\mathcal{C}_{TS}/(\mathcal{C}_{TS} \cap \mathcal{G}_{\lfloor k/2 \rfloor})$ of $\mathcal{C}/\mathcal{G}_{\lfloor k/2 \rfloor}$ to complete the proof. □

Proof of Theorem 1.5. Instead of $\{J_n\}$, we use another family of topologically slice knots $\{L_n\}$, where $L_n = (\mathrm{Wh}(T_{2,3}))_{n,1} \# - (\mathrm{Wh}(T_{2,3}))_{n-1,1}$. These knots have slice genus 1 [Hom 2015b, Lemma 3.1]. Additionally, Hom [2015b] computes that $a_1(L_n) = 1$ and $a_2(L_n) = n$. By the same argument as the above proof, except for applying Lemma 3.4 rather than Lemma 3.5, we immediately know $\{L_n\}_{n=2k}^{\infty}$ are linearly independent in $\mathcal{C}_{TS}/(\mathcal{C}_{TS} \cap \mathcal{G}_k)$. □

4. Obstruction by Υ-invariant

Ozsvath et al. [2014] introduced a new family of knot invariants, $\Upsilon_K(t)$. We refer the reader to their construction, and confine ourselves to recalling the basic properties of the Υ-invariant.

For any knot K, the invariant $\Upsilon_K(t)$ is a piecewise linear function on [0, 2] whose derivative has finitely many discontinuities [Ozsvath et al. 2014, Proposition 1.4]. Thus, one can define $\Delta\Upsilon'_K(t_0) = \lim_{t \to t_0^+} \Upsilon'_K(t) - \lim_{t \to t_0^-} \Upsilon'_K(t)$ for any $t_0 \in (0, 2)$.

As an example, the authors of [Ozsvath et al. 2014] compute the family $\{J'_n\}$:

$$\Delta\Upsilon'_{J'_n}(t) = \begin{cases} 0 & \text{for } t < 2/(2n-1), \\ 2n-1 & \text{for } t = 2/(2n-1). \end{cases}$$

In [Ozsvath et al. 2014, Corollary 1.12] it is shown that Υ gives a homomorphism from $\mathcal{C}$ to the vector space of continuous functions on [0, 2]. Additionally,

$$K \mapsto \begin{cases} (1/q)\Delta\Upsilon'_K(p/q) & \text{if } p \text{ is even}, \\ (1/2q)\Delta\Upsilon'_K(p/q) & \text{if } p \text{ is odd}, \end{cases}$$

gives a homomorphism from $\mathcal{C}$ to $\mathbb{Z}$ for any $p/q \in (0, 2) \cap \mathbb{Q}$.

The location of singularities of Υ is related to the genus of the knot, as in the following proposition. The proof of this proposition, much like that of Lemma 3.11, is based on the fact in Proposition 3.1.

Proposition 4.1 [Livingston 2015, Theorem 8.2]. *Suppose that $\Delta\Upsilon'_K(t)$ is nonzero at $t = p/q$ with* $\gcd(p, q) = 1$. *Then $q \leq g(K)$ if p is odd, and $q \leq 2g(K)$ if p is even.*

With this proposition, we can easily prove our obstruction theorem.

Proposition 4.2. *Suppose $K \in \mathcal{G}_n$ for some positive integer n. Then $\Delta\Upsilon'_K(t) = 0$ for $t \in (0, 1/n) \cap \mathbb{Q}$.*

Proof. Before proving the proposition for $K \in \mathcal{G}_n$, first consider the case $g(K) \leq n$. If $\Upsilon_K(t)$ has a singularity at a rational number p/q with $\gcd(p, q) = 1$, then Proposition 4.1 implies $p/q \geq 1/n$.

Generally, let $K \in \mathcal{G}_n$. Then $K = K_1 + \cdots + K_m$, where $g(K_i) \leq n$ for $i = 1, \ldots, m$. If $\Upsilon_K(t)$ has a singularity at a rational number p/q, then so does one of $\Upsilon_{K_1}(t), \ldots, \Upsilon_{K_m}(t)$, since Υ is a homomorphism. The conclusion follows from the last paragraph. □

Proof of Theorem 1.3. Fix an integer $k \geq 2$. If $K \in \mathcal{G}_{k-1}$, then $\Upsilon_K(t)$ has no singularities on $(0, 1/(k-1)) \cap \mathbb{Q}$. Thus $\{K \mapsto 1/(2n-1)\Delta\Upsilon'_K(2/(2n-1))\}_{n=k}^{\infty}$ gives a homomorphism from $\mathcal{C}/\mathcal{G}_{k-1}$ to $\mathbb{Z}^{\infty}$. Hence $\{J'_n\}_{n=k}^{\infty}$ form a basis for a $\mathbb{Z}^{\infty}$ summand of $\mathcal{C}/\mathcal{G}_{k-1}$. Note that since the J'_n are topologically slice, the above argument can be restricted to the subgroup $\mathcal{C}_{TS}/(\mathcal{C}_{TS} \cap \mathcal{G}_{k-1})$ of $\mathcal{C}/\mathcal{G}_{k-1}$ to complete the proof. □

Acknowledgements

The author wishes to express sincere thanks to Professor Charles Livingston for proposing this study and carefully reading a draft of this paper. Thanks also to Jennifer Hom for valuable comments on a draft and to the referee for very detailed suggestions on writing style and grammar.

References

[Endo 1995] H. Endo, "Linear independence of topologically slice knots in the smooth cobordism group", *Topology Appl.* **63**:3 (1995), 257–262. MR 1334309

[Hedden et al. 2012] M. Hedden, C. Livingston, and D. Ruberman, "Topologically slice knots with nontrivial Alexander polynomial", *Adv. Math.* **231**:2 (2012), 913–939. MR 2955197

[Hedden et al. 2016] M. Hedden, S.-G. Kim, and C. Livingston, "Topologically slice knots of smooth concordance order two", *J. Differential Geom.* **102**:3 (2016), 353–393. MR 3466802

[Hom 2014a] J. Hom, "Bordered Heegaard Floer homology and the tau-invariant of cable knots", *J. Topol.* **7**:2 (2014), 287–326. MR 3217622

[Hom 2014b] J. Hom, "The knot Floer complex and the smooth concordance group", *Comment. Math. Helv.* **89**:3 (2014), 537–570. MR 3260841

[Hom 2015a] J. Hom, "An infinite-rank summand of topologically slice knots", *Geom. Topol.* **19**:2 (2015), 1063–1110. MR 3336278

[Hom 2015b] J. Hom, "On the concordance genus of topologically slice knots", *Int. Math. Res. Not.* **2015**:5 (2015), 1295–1314. MR 3340357

[Litherland 1979] R. A. Litherland, "Signatures of iterated torus knots", pp. 71–84 in *Topology of low-dimensional manifolds* (Proc. Second Sussex Conf., Chelwood Gate, 1977), edited by R. A. Fenn, Lecture Notes in Math. **722**, Springer, Berlin, 1979. MR 547456

[Livingston 2004] C. Livingston, "The concordance genus of knots", *Algebr. Geom. Topol.* **4** (2004), 1–22. MR 2031909

[Livingston 2015] C. Livingston, "Notes on the knot concordance invariant Upsilon", preprint, 2015. arXiv 1412.0254v2

[Ozsváth and Szabó 2003] P. Ozsváth and Z. Szabó, "Knot Floer homology and the four-ball genus", *Geom. Topol.* **7** (2003), 615–639. MR 2026543

[Ozsváth and Szabó 2004a] P. Ozsváth and Z. Szabó, "Holomorphic disks and genus bounds", *Geom. Topol.* **8** (2004), 311–334. MR 2023281

[Ozsváth and Szabó 2004b] P. Ozsváth and Z. Szabó, "Holomorphic disks and knot invariants", *Adv. Math.* **186**:1 (2004), 58–116. MR 2065507

[Ozsvath et al. 2014] P. Ozsvath, A. Stipsicz, and Z. Szabo, "Concordance homomorphisms from knot Floer homology", preprint, 2014. arXiv 1407.1795v2

[Rasmussen 2003] J. A. Rasmussen, *Floer homology and knot complements*, Ph.D. Thesis, Harvard University, 2003, Available at http://search.proquest.com/docview/305332635. MR 2704683

Received October 26, 2015. Revised June 23, 2016.

SHIDA WANG
DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, IN 47405
UNITED STATES
shidawang@gmail.com

CONTENTS

Volume 285, no. 1 and no. 2

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 285 No. 2 December 2016

0030-8730(201612)285:2;1-W