Pacific Journal of Mathematics

Volume 282 No. 2 June 2016

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

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

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. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.257

EXHAUSTING CURVE COMPLEXES BY FINITE RIGID SETS

JAVIER ARAMAYONA AND CHRISTOPHER J. LEININGER

Let S be a connected orientable surface of finite topological type. We prove that there is an exhaustion of the curve complex $\mathscr{C}(S)$ by a sequence of finite rigid sets.

1. Introduction

The curve complex $\mathscr{C}(S)$ of a surface S is a simplicial complex whose k-simplices correspond to sets of $k+1$ distinct isotopy classes of essential simple closed curves on S with pairwise disjoint representatives. The extended mapping class group $\mathrm{Mod}^{\pm}(S)$ of S acts on $\mathscr{C}(S)$ by simplicial automorphisms, and a well-known theorem due to Ivanov [1997], Korkmaz [1999] and Luo [2000] asserts that $\mathscr{C}(S)$ is *simplicially rigid* for $S \neq S_{1,2}$. More concretely, the natural homomorphism

$$\mathrm{Mod}^{\pm}(S) \to \mathrm{Aut}(\mathscr{C}(S))$$

is surjective unless $S = S_{1,2}$; in the case $S = S_{1,2}$ there is an automorphism of $\mathscr{C}(S)$ that sends a separating curve on S to a nonseparating one and thus cannot be induced by an element of $\mathrm{Mod}^{\pm}(S)$ (see [Luo 2000]).

In [Aramayona and Leininger 2013], henceforth abbreviated [AL], we extended this picture and showed that curve complexes are *finitely rigid*. Specifically, for $S \neq S_{1,2}$ we identified a finite subcomplex $\mathfrak{X}(S) \subset \mathscr{C}(S)$ with the property that every locally injective map $\mathfrak{X}(S) \to \mathscr{C}(S)$ is the restriction of an element of $\mathrm{Mod}^{\pm}(S)$; in the case of $S_{1,2}$ a similar statement can be made, this time using the group $\mathrm{Aut}(\mathscr{C}(S))$ instead of $\mathrm{Mod}^{\pm}(S)$. We refer to such a subset $\mathfrak{X}(S)$ as a *rigid* set.

The rigid sets constructed in [AL] enjoy some curious properties. For instance, if $S = S_{0,n}$ is a sphere with n punctures then $\mathfrak{X}(S)$ is homeomorphic to an $(n-4)$-dimensional sphere. Since $\mathscr{C}(S)$ has dimension $n-4$, it follows that $\mathfrak{X}(S)$ represents a nontrivial element of $H_{n-4}(\mathscr{C}(S), \mathbb{Z})$ which, by a result of Harer [1986], is the only nontrivial homology group of $\mathscr{C}(S)$. In fact, Broaddus [2012] and Birman, Broaddus and Menasco [Birman et al. 2015] have recently proved that $\mathfrak{X}(S)$ is a

Aramayona was supported by BQR (Toulouse) and Campus Iberus grants. Leininger was supported by NSF grants DMS 0905748 and 1207183.
MSC2010: 57M99.
Keywords: curve complex, mapping class group, rigidity.

generator of $H_{n-4}(\mathscr{C}(S), \mathbb{Z})$ when viewed as a $\mathrm{Mod}^{\pm}(S)$-module; in the case when S has genus ≥ 2 and at least one puncture, they prove that $\mathfrak{X}(S)$ *contains* a generator for the homology of $\mathscr{C}(S)$.

The rigid sets identified in [AL] all have diameter 2 in $\mathscr{C}(S)$, and a natural question is whether there exist finite rigid sets in $\mathscr{C}(S)$ of arbitrarily large diameter; see Question 1 of that work. In this paper we prove that, in fact, there exists an exhaustion of $\mathscr{C}(S)$ by finite rigid sets:

Theorem 1.1. *Let $S \neq S_{1,2}$ be a connected orientable surface of finite topological type. There exists a sequence $\mathfrak{X}_1 \subset \mathfrak{X}_2 \subset \cdots \subset \mathscr{C}(S)$ such that*

(1) *$\mathfrak{X}_i$ is a finite rigid set for all $i \geq 1$,*

(2) *$\mathfrak{X}_i$ has trivial pointwise stabilizer in* $\mathrm{Mod}^{\pm}(S)$, *for all $i \geq 1$, and*

(3) $\bigcup_{i\geq 1} \mathfrak{X}_i = \mathscr{C}(S)$.

Remarks. (i) A similar statement can be made for $S = S_{1,2}$, by replacing $\mathrm{Mod}^{\pm}(S)$ by $\mathrm{Aut}(\mathscr{C}(S))$ in the definition of rigid set above.

(ii) We stress that Theorem 1.1 above does not follow from the main result in [AL]. Indeed, a subset of $\mathscr{C}(S)$ containing a rigid set need not itself be rigid; compare with Proposition 3.2 below.

(iii) The combination of a recent theorem of J. Hernández (as yet unpublished) and the main result in [AL] gives an alternate proof of Theorem 1.1 in the case when S has genus ≥ 3; compare with the remark on page 262 below.

As a consequence of Theorem 1.1 we will obtain a "finitistic" proof of the aforementioned result of [Ivanov 1997; Korkmaz 1999; Luo 2000] on the simplicial rigidity of the curve complex. In fact, we will deduce the following stronger form due to Shackleton [2007].

Corollary 1.2. *Let $S \neq S_{1,2}$ be a connected orientable surface of finite topological type. If $\phi : \mathscr{C}(S) \to \mathscr{C}(S)$ is a locally injective simplicial map, then there exists $h \in$* $\mathrm{Mod}^{\pm}(S)$ *such that $h = \phi$.*

The first author and Souto [2013] proved that if $\mathfrak{X} \subset \mathscr{C}(S)$ is a rigid set satisfying some extra conditions, then every (weakly) injective homomorphism from the right-angled Artin group $\mathbb{A}(\mathfrak{X})$ into $\mathrm{Mod}^{\pm}(S)$ is obtained, up to conjugation, by taking powers of roots of Dehn twists in the vertices of $\mathfrak{X}$. Since the finite rigid sets $\mathfrak{X}_i$ of Theorem 1.1 all satisfy the conditions of [Aramayona and Souto 2013], we obtain the following result; here, T_γ denotes the Dehn twist about γ.

Corollary 1.3. *Let $S \neq S_{1,2}$ be a connected orientable surface of finite topological type, and consider the sequence $\mathfrak{X}_1 \subset \mathfrak{X}_2 \subset \cdots \subset \mathscr{C}(S)$ of finite rigid sets given by*

Theorem 1.1. If $\rho_i : \mathbb{A}(\mathfrak{X}_i) \to \mathrm{Mod}^{\pm}(S)$ is an injective homomorphism, then there exist functions $a, b : \mathscr{C}^{(0)}(S) \to \mathbb{Z} \setminus \{0\}$ and $f_i \in \mathrm{Mod}^{\pm}(S)$ such that

$$\rho_i(\gamma^{a(\gamma)}) = f_i T_\gamma^{b(\gamma)} f_i^{-1},$$

for every vertex γ of $\mathfrak{X}_i$.

We remark that Kim and Koberda [2013] have previously shown the existence of injective homomorphisms $\mathbb{A}(Y_i) \to \mathrm{Mod}^{\pm}(S)$ for sequences $Y_1 \subset Y_2 \subset \cdots$ of subsets of $\mathscr{C}(S)$. Such homomorphisms may in fact be obtained by sending a generator of $\mathbb{A}(Y_i)$ to a sufficiently high power of a Dehn *multitwist*, see [Kim and Koberda 2015].

Plan of the paper. In Section 2 we recall some necessary definitions and basic results from our previous paper [AL]. Section 3 deals with the problem of enlarging a rigid set in such a way that it remains rigid. As was the case in [AL], the techniques used in the proof of our main result differ depending on the genus of S. As a result, we prove Theorem 1.1 for surfaces of genus $g = 0$, $g \geq 2$ and $g = 1$ in Sections 4, 5 and 6, respectively.

2. Definitions

Let $S = S_{g,n}$ be an orientable surface of genus g with n punctures and/or marked points. We define the *complexity* of S as $\xi(S) = 3g - 3 + n$. We say that a simple closed curve on S is *essential* if it does not bound a disk or a once-punctured disk on S. An *essential subsurface* of S is a properly embedded subsurface $N \subset S$ for which each boundary component is an essential curve in S.

The *curve complex* $\mathscr{C}(S)$ of S is a simplicial complex whose k-simplices correspond to sets of $k + 1$ isotopy classes of essential simple closed curves on S with pairwise disjoint representatives. In order to simplify the notation, a set of isotopy classes of simple closed curves will be confused with its representative curves, the corresponding vertices of $\mathscr{C}(S)$, and the subcomplex of $\mathscr{C}(S)$ spanned by the vertices. We also assume that representatives of isotopy classes of curves and subsurfaces intersect *minimally* (that is, transversely and in the minimal number of components), and denote by $i(\alpha, \beta)$ their intersection number.

If $\xi(S) > 1$, then $\mathscr{C}(S)$ is a connected complex of dimension $\xi(S) - 1$. If $\xi(S) \leq 0$ and $S \neq S_{1,0}$, then $\mathscr{C}(S)$ is empty. If $\xi(S) = 1$ or $S = S_{1,0}$, then $\mathscr{C}(S)$ is a countable set of vertices; in order to obtain a connected complex, we modify the definition of $\mathscr{C}(S)$ by declaring $\alpha, \beta \in \mathscr{C}^{(0)}(S)$ to be adjacent in $\mathscr{C}(S)$ whenever $i(\alpha, \beta) = 1$ if $S = S_{1,1}$ or $S = S_{1,0}$, and whenever $i(\alpha, \beta) = 2$ if $S = S_{0,4}$. Furthermore, we add triangles to make $\mathscr{C}(S)$ into a flag complex. In all three cases, the complex $\mathscr{C}(S)$ so obtained is isomorphic to the well-known *Farey complex*.

We recall some definitions and results from [AL] that we will need later.

Definition 2.1 (detectable intersection). Let S be a surface and $Y \subset \mathscr{C}(S)$ a subcomplex. If $\alpha, \beta \in Y$ are curves with $i(\alpha, \beta) \neq 0$, then we say that their intersection is *Y-detectable* (or simply *detectable* if Y is understood) if there are two pants decompositions $P_\alpha, P_\beta \subset Y$ such that

$$\alpha \in P_\alpha, \quad \beta \in P_\beta \quad \text{and} \quad P_\alpha - \alpha = P_\beta - \beta. \tag{1}$$

We note that if α, β have detectable intersection, then they must fill a $\xi = 1$ (essential) subsurface, which we denote $N(\alpha \cup \beta) \subset S$. For notational purposes, we call $P = P_\alpha - \alpha = P_\beta - \beta$ a *pants decomposition of $S - N(\alpha \cup \beta)$*, even though it includes the boundary components of $N(\alpha \cup \beta)$. The following lemma is Lemma 2.3 in [AL].

Lemma 2.2. *Let $Y \subset \mathscr{C}(S)$ be a subcomplex, and $\alpha, \beta \in Y$ intersecting curves with Y-detectable intersection. If $\phi : Y \to \mathscr{C}(S)$ is a locally injective simplicial map, then $\phi(\alpha), \phi(\beta)$ have $\phi(Y)$-detectable intersection, and hence fill a $\xi = 1$ subsurface.*

Farey neighbors. A large part of our arguments will rely on being able to recognize when two curves are *Farey neighbors*, which we now define.

Definition 2.3 (Farey neighbors). Let α and β be curves on S which fill a $\xi = 1$ subsurface $N \subset S$. We say α and β are *Farey neighbors* if they are adjacent in $\mathscr{C}(N)$.

The following result is a useful tool for recognizing Farey neighbors, and is a rephrasing of Lemma 2.4 in [AL] (see also the comment immediately after it):

Lemma 2.4. *Suppose $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are curves on S such that*

(1) *α_2, α_3 together fill a $\xi = 1$ subsurface $N \subset S$,*
(2) *$i(\alpha_i, \alpha_j) = 0 \Leftrightarrow |i - j| > 1$ for all $i \neq j$, and*
(3) *α_1 and α_4 have nonzero intersection number with exactly one component of ∂N.*

Then α_2 and α_3 are Farey neighbors.

3. Enlarging rigid sets

In this section we discuss the problem of enlarging rigid sets of the curve complex. We recall the definition of rigid set from [AL].

Definition 3.1 (rigid set). Suppose $S \neq S_{1,2}$. We say that $Y \subset \mathscr{C}(S)$ is *rigid* if for every locally injective simplicial map $\phi : Y \to \mathscr{C}(S)$ there exists $h \in \mathrm{Mod}^{\pm}(S)$ with $h|_Y = \phi$, unique up to the pointwise stabilizer of Y in $\mathrm{Mod}^{\pm}(S)$.

Remark. The definition above may seem somewhat different from the one used in [AL], where we used the group $\mathrm{Aut}(\mathscr{C}(S))$ instead of $\mathrm{Mod}^{\pm}(S)$. Nevertheless, in light of the results of Ivanov [1997], Korkmaz [1999] and Luo [2000] mentioned in the introduction, the two definitions are essentially the same as $S \neq S_{1,2}$. For

$S = S_{1,2}$, however, we will use the group $\mathrm{Aut}(\mathscr{C}(S))$ instead of $\mathrm{Mod}^{\pm}(S)$, due to the existence of *nongeometric* automorphisms of $\mathscr{C}(S)$.

The main step in the proof of Theorem 1.1 is to enlarge the rigid sets constructed in [AL] in a way that the sets we obtain remain rigid. As we mentioned in the introduction, while one might be tempted to guess that a set that contains a rigid set is necessarily rigid, this is far from true, as the next result shows.

Proposition 3.2. *Let $S = S_{0,n}$, with $n \geq 5$, and $\mathfrak{X}$ the finite rigid set identified in [AL] (defined in Section 4). For every curve $\alpha \in \mathscr{C}(S) \setminus \mathfrak{X}$, the set $\mathfrak{X} \cup \{\alpha\}$ is not rigid.*

Proof. Let S_α be the smallest subsurface of S containing all the curves in $\mathfrak{X}$ which are disjoint from α. Observe that since $\mathfrak{X}$ is rigid, it is also filling and therefore S_α is a proper subsurface of S. Let S'_α be the connected component of $S \setminus S_\alpha$ that contains α; from the construction in [AL], every component of $\partial S'_\alpha$ which is essential in S is an element of $\mathfrak{X}$. We claim that there exists $f \in \mathrm{Mod}(S)$ with the following two properties:

(1) The restriction of f to S_α is the identity map.

(2) For every $\beta \in \mathfrak{X}$ with $i(\alpha, \beta) \neq 0$, we have $i(f(\alpha), \beta) \neq 0$.

In order to construct such an f, one can for instance consider an element $h \in \mathrm{Mod}(S)$ that is pseudo-Anosov on S'_α and the identity on S_α; any sufficiently high power of h will satisfy the two conditions above.

At this point, define a map $\phi : \mathfrak{X} \cup \{\alpha\} \to \mathscr{C}(S)$ by $\phi(\beta) = \beta$ for all $\beta \neq \alpha$, and $\phi(\alpha) = f(\alpha)$. By construction, the map ϕ is locally injective and simplicial, but cannot be the restriction of an element of $\mathrm{Mod}^{\pm}(S)$. □

While Proposition 3.2 serves to highlight the obstacles for enlarging a rigid set to a set that is also rigid, we now explain two procedures for doing so. First, we recall the following definition from [AL].

Definition 3.3. Let A be a set of curves in S.

(1) A is *almost filling (in S)* if the set

$$B = \{\beta \in \mathscr{C}^{(0)}(S) \setminus A \mid i(\alpha, \beta) = 0 \ \forall \alpha \in A\}$$

is finite. In this case, we call B the *set of curves determined by A*.

(2) If A is almost filling (in S), and $B = \{\beta\}$ is a single curve, then we say that β is uniquely determined by A.

An immediate consequence of the definition is the following.

Lemma 3.4. *Let Y be a rigid set of curves, and $A \subset Y$ an almost filling set in S. If β is uniquely determined by A, then $Y \cup \{\beta\}$ is rigid.*

Proof. Given any locally injective simplicial map $\phi : Y \cup \{\beta\} \to \mathscr{C}(S)$, we let $f \in \mathrm{Mod}^{\pm}(\mathscr{C}(S))$ be such that $f|_Y = \phi$. Then $f(\beta)$ is the unique curve determined by $f(A) = \phi(A)$. On the other hand, $\phi(\beta)$ is connected by an edge to every vertex in $\phi(A)$, since ϕ is simplicial. Since ϕ is injective on the star of β, it is injective on $\beta \cup A$, and so $\phi(\beta) \notin A$. It follows that $\phi(\beta)$ is the curve uniquely determined by $\phi(A)$, and hence $f(\beta) = \phi(\beta)$. □

In particular, this gives rise to one method for enlarging a rigid set which we formalize as follows. Given a subset $Y \subset \mathscr{C}(S)$, define

$$Y' = Y \cup \{\beta \mid \beta \text{ is uniquely determined by some almost filling set } A \subset Y\}.$$

From this we recursively define $Y = Y^0$ and $Y^r = (Y^{r-1})'$ for all $r > 0$. Observe that, as an immediate consequence of Lemma 3.4, we obtain:

Proposition 3.5. *If $Y \subset \mathscr{C}(S)$ is a rigid set, then so is Y^r for all $r \geq 0$.*

Remark. J. Hernández has recently proved that for every surface S of genus ≥ 3, there exists an explicit finite subcomplex $Y \subset \mathscr{C}(S)$ such that

$$\bigcup_{r \geq 0} Y^r = \mathscr{C}(S).$$

As a corollary of this result, and using the main result in [AL], he provides an alternate proof of Theorem 1.1; compare with remark (iii) on page 258 above.

Next, we give a sufficient condition for the union of two rigid sets to be rigid. Before doing so, we need the following definition.

Definition 3.6 (weakly rigid set). We say that a set $Y \subset \mathscr{C}(S)$ is *weakly rigid* if, whenever $h, h' \in \mathrm{Mod}^{\pm}(S)$ satisfy $h|_Y = h'|_Y$, then $h = h'$.

Alternatively, Y is weakly rigid if the pointwise stabilizer in $\mathrm{Mod}^{\pm}(S)$ is trivial. Note that if Y is a weakly rigid set, then so is every set containing Y.

Lemma 3.7. *Let $Y_1, Y_2 \subset \mathscr{C}(S)$ be rigid sets. If $Y_1 \cap Y_2$ is weakly rigid then $Y_1 \cup Y_2$ is rigid.*

Proof. Let $\phi : Y_1 \cup Y_2 \to \mathscr{C}(S)$ be a locally injective simplicial map. Since Y_i is rigid and has trivial pointwise stabilizer in $\mathrm{Mod}^{\pm}(S)$ (because $Y_1 \cap Y_2$ does), there exists a unique $h_i \in \mathrm{Mod}^{\pm}(S)$ such that $h_i|_{Y_i} = \phi|_{Y_i}$. Finally, since $Y_1 \cap Y_2$ is weakly rigid we have $h_1 = h_2 = h$. Therefore $h|_{Y_1 \cup Y_2} = \phi$, and the result follows. □

We now proceed to describe our second method for enlarging a rigid set. We start with some definitions and notation. We write T_α for the Dehn twist along a curve α. Recall that the half-twist H_α about a curve α is defined if and only if the curve cuts off a pair of pants containing two punctures of S. Furthermore, there is exactly one half-twist about α if in addition S is not a four-holed sphere.

Definition 3.8. Farey neighbors α and β are *twistable* if either

(1) $N(\alpha \cup \beta)$ is a one-holed torus, or

(2) $N(\alpha \cup \beta)$ is a four-holed sphere and H_α, H_β are both defined and unique.

In this situation we define $f_\alpha = T_\alpha$ and $f_\beta = T_\beta$ in the first case and $f_\alpha = H_\alpha$ and $f_\beta = H_\beta$ in the second. We call f_α, f_β the *twisting pair* for α, β.

In case (1) we call α and β *toroidal* and in case (2) we call them *spherical*.

We note that whether twistable Farey neighbors α and β are toroidal or spherical can be distinguished

(i) by $i(\alpha, \beta)$ (whether it is 1 or 2),

(ii) by the homeomorphism types of α and β (whether they are nonseparating curves or they cut off a pair of pants), or

(iii) by the homeomorphism type of $N(\alpha \cup \beta)$ (whether it is a one-holed torus or a four-holed sphere).

The following well-known fact describes the common feature of these two situations.

Proposition 3.9. *Suppose α, β are twistable Farey neighbors and that f_α, f_β is their twisting pair. Then*

$$f_\alpha(\beta) = f_\beta^{-1}(\alpha) \quad \textit{and} \quad f_\alpha^{-1}(\beta) = f_\beta(\alpha),$$

and these are the unique common Farey neighbors of both α and β.

Sets of twistable Farey neighbors which interact with each other frequently occur in our rigid sets. We distinguish one particular type of such sets in the following definition.

Definition 3.10. Suppose Y is a rigid subset of $\mathscr{C}(S)$ and $A = \{\alpha_1, \dots, \alpha_k\} \subset Y$. We say that A is a *closed string of Farey neighbors* in Y provided the following conditions are satisfied, counting indices modulo k:

(1) The curves α_i, α_{i+1} are twistable Farey neighbors with twisting pair f_{α_i}, $f_{\alpha_{i+1}}$.

(2) $i(\alpha_i, \alpha_{i+1}) \neq 0$ is Y-detectable.

(3) $i(\alpha_i, \alpha_j) = 0$ if $i - j \neq \pm 1$ modulo k.

(4) $\alpha_i, \alpha_{i+1}, \alpha_{i+2}, \alpha_{i+3}$ satisfy the hypothesis of Lemma 2.4.

Given a closed string of twistable Farey neighbors $A \subset Y$, we define

$$Y_A = Y \cup \{f_{\alpha_i}^{\pm 1}(\alpha_j)\}_{i,j=1}^k.$$

Remark. Two comments are in order:

(1) There is a priori some ambiguity in the notation as f_{α_i} can be defined as part of the twisting pair for α_i, α_{i+1} as well as for α_{i-1}, α_i. However, if α_i is part of two pairs of different twistable Farey neighbors in Y, then they must both be toroidal or both spherical as this is determined by the homeomorphism type of α_i. Consequently, the mapping class f_{α_i} is independent of what twistable pair it is included in.

(2) Given condition (3) of Definition 3.10, the set Y_A has a more descriptive definition. Namely,

$$Y_A = Y \cup \{f_{\alpha_i}^{\pm 1}(\alpha_j) \mid i - j = \pm 1 \text{ modulo } k\}.$$

See Figure 1 for an example of a closed string of twistable Farey neighbors and two of their images under the twisting pair.

The situation in the next proposition arises in multiple settings, and provides a way to extend a rigid set to a larger set which is nearly rigid.

Proposition 3.11. *Let Y be a rigid subset of $\mathscr{C}(S)$ and $A = \{\alpha_1, \dots, \alpha_k\} \subset Y$ a closed string of twistable Farey neighbors in Y. Then, counting indices modulo k:*

(1) $f_{\alpha_i}^{\pm 1}(\alpha_{i+1}) = f_{\alpha_{i+1}}^{\mp 1}(\alpha_i)$ *are the unique common Farey neighbors of α_i and α_{i+1}.*

(2) $i(f_{\alpha_i}^{\pm 1}(\alpha_j), \alpha_{j'}) \neq 0$ *for all i and all*

$$(j, j') \in \{(i+1, i), (i+1, i+1), (i-1, i), (i-1, i-1)\}.$$

Furthermore, these intersections are Y_A-detectable.

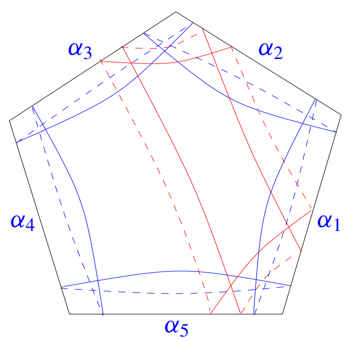

Figure 1. The set $Y = A = \{\alpha_1, \dots, \alpha_5\}$ is the rigid set $\mathfrak{X}(S_{0,5})$ identified in [Luo 2000] and [AL], and is a closed string of twistable Farey neighbors. The red curves in the picture are $f_{\alpha_1}(\alpha_2)$, $f_{\alpha_2}(\alpha_1)$, for the twistable pair α_1, α_2. The automorphism group of Y_A that fixes Y pointwise is generated by an orientation-reversing involution $\sigma : S_{0,5} \to S_{0,5}$ that fixes α_i and interchanges f_{α_i} and $f_{\alpha_{i+1}}$, for all i (mod 5).

(3) *For any locally injective simplicial map* $\phi : Y_A \to \mathscr{C}(S)$,

$$\phi(f_{\alpha_i}(\alpha_{i+1})) = \phi(f_{\alpha_{i+1}}^{-1}(\alpha_i)) \quad \text{and} \quad \phi(f_{\alpha_i}^{-1}(\alpha_{i+1})) = \phi(f_{\alpha_{i+1}}(\alpha_i))$$

are the unique Farey neighbors of $\phi(\alpha_i)$ *and* $\phi(\alpha_{i+1})$.

Proof. Conclusion (1) follows immediately from Definition 3.10 part (1) and Proposition 3.9.

Next we prove conclusion (2). Fix (j, j') as in the proposition. Then since $i(\alpha_i, \alpha_j) \neq 0$, it follows that $f_{\alpha_i}(\alpha_j)$ nontrivially intersects both α_i and α_j. Since $\alpha_{j'}$ is one of these latter two curves, the first statement follows. By part (2) of Definition 3.10, $i(\alpha_i, \alpha_j) \neq 0$ is Y-detectable. Let $P_{\alpha_i}, P_{\alpha_j} \subset Y$ be pants decompositions containing α_i and α_j, respectively, as in Definition 2.1, and set $P = P_{\alpha_i} - \alpha_i = P_{\alpha_j} - \alpha_j$. Then since f_{α_i} is supported in $N(\alpha_i \cup \alpha_j)$ which is contained in the complement of P, we can define two more pants decompositions

$$P_{f_{\alpha_i}^{\pm 1}(\alpha_j)} = P \cup f_{\alpha_i}^{\pm 1}(\alpha_j) \subset Y_A.$$

Together with P_{α_i} and P_{α_j} these are sufficient to detect all the intersections claimed. In all cases, $P \subset Y$ is the pants decomposition of the complement of $N(\alpha_i \cup \alpha_j)$, as required.

For conclusion (3), we explain why $\phi(f_{\alpha_i}(\alpha_{i+1}))$ and $\phi(\alpha_i)$ are Farey neighbors. The other three cases are similar. For this, we consider the set

$$\{\phi(f_{\alpha_i}(\alpha_{i-1})),\ \phi(\alpha_i) = \phi(f_{\alpha_i}(\alpha_i)),\ \phi(f_{\alpha_i}(\alpha_{i+1})),\ \phi(\alpha_{i+2}) = \phi(f_{\alpha_i}(\alpha_{i+2}))\}.$$

The equalities here follow from the disjointness property (3) of Definition 3.10 since a Dehn twist or half-twist has no effect on a curve that is disjoint from the curve supporting the twist. The goal is to prove that all three conditions of Lemma 2.4 are satisfied.

By part (2) of the proposition and Lemma 2.2 it follows that any two consecutive curves in this set have $\phi(Y_A)$-detectable intersections and fill a $\xi = 1$ subsurface. Therefore condition (1) of Lemma 2.4 is satisfied for this set of curves. Since $\alpha_{i-1}, \alpha_i, \alpha_{i+1}, \alpha_{i+2}$ satisfy condition (2) of Lemma 2.4 and the given set is the image of these under the simplicial map $\phi \circ f_{\alpha_i}$, these curves also satisfy condition (2) of Lemma 2.4.

Finally, we wish to verify that condition (3) of Lemma 2.4 is satisfied. Since Y is rigid, there exists $f \in \mathrm{Mod}^{\pm 1}(S)$ inducing $\phi|_Y$. We also note that

$$N = N(\alpha_i \cup \alpha_{i+1}) = N(\alpha_i \cup f_{\alpha_i}(\alpha_{i+1}))$$

has only one boundary component — all other holes of this subsurface (if any) must be punctures of S. Since the pants decomposition of the complement of N is

$$P = P_{\alpha_i} - \alpha_i = P_{\alpha_{i+1}} - \alpha_{i+1} = P_{f_{\alpha_i}(\alpha_{i+1})} - f_{\alpha_i}(\alpha_{i+1})$$

and is contained in Y, we see that $\phi(P) = f(P)$. Because this is used in the $\phi(Y_A)$-detection of both $i(\phi(\alpha_i), \phi(\alpha_{i+1})) \neq 0$ and $i(\phi(\alpha_i), \phi(f_{\alpha_i}(\alpha_{i+1}))) \neq 0$, we have

$$f(N) = N(f(\alpha_i) \cup f(\alpha_{i+1})) = N(\phi(\alpha_i) \cup \phi(\alpha_{i+1})) = N(\phi(\alpha_i) \cup \phi(f_{\alpha_i}(\alpha_{i+1}))).$$

Consequently, this surface has only one boundary component, and so condition (3) of Lemma 2.4 is satisfied. □

Proposition 3.12. *If Y is a rigid subset of $\mathscr{C}(S)$ and $A = \{\alpha_1, \ldots, \alpha_k\}$ is a closed string of twistable Farey neighbors in Y, then any locally injective simplicial map $\phi : Y_A \to \mathscr{C}(S)$ which is the identity on Y satisfies $\phi(Y_A) = Y_A$. Furthermore, the subgroup of the automorphism group of Y_A fixing Y pointwise has order at most 2. If this subgroup is nontrivial, then it is generated by the involution $\sigma : Y_A \to Y_A$ given by $\sigma(f_{\alpha_i}(\alpha_j)) = f_{\alpha_i}^{-1}(\alpha_j)$ for all i, j (or equivalently, for all i, j with $i - j = \pm 1$ (modulo k)).*

Proof. Since $f_{\alpha_i}^{\pm 1}(\alpha_{i+1})$ is the unique pair of common Farey neighbors of α_i, α_{i+1}, and since $\phi(\alpha_i) = \alpha_i$, $\phi(\alpha_{i+1}) = \alpha_{i+1}$, Proposition 3.11 implies that for every i and j with $i - j = \pm 1$ (modulo k), we have

$$\{\phi(f_{\alpha_i}(\alpha_j)), \phi(f_{\alpha_i}^{-1}(\alpha_j))\} = \{f_{\alpha_i}(\alpha_j), f_{\alpha_i}^{-1}(\alpha_j)\},$$

and so the first claim of the proposition follows.

Next we suppose ϕ is any automorphism of Y that restricts to the identity on Y. We claim that if there is some i, j with $i - j = \pm 1$ (modulo k) so that $\phi(f_{\alpha_i}(\alpha_j)) = f_{\alpha_i}^{-1}(\alpha_j)$, then this is true for every i, j with $i - j = \pm 1$ (modulo k). To this end, suppose that $\phi(f_{\alpha_i}(\alpha_{i+1})) = f_{\alpha_i}^{-1}(\alpha_{i+1})$ for some index i (the case $\phi(f_{\alpha_i}(\alpha_{i-1})) = f_{\alpha_i}^{-1}(\alpha_{i-1})$ is similar). Then note that

$$i(f_{\alpha_i}(\alpha_{i-1}), f_{\alpha_i}(\alpha_{i+1})) = 0 = i(f_{\alpha_i}^{-1}(\alpha_{i-1}), f_{\alpha_i}^{-1}(\alpha_{i+1}))$$

while

$$i(f_{\alpha_i}(\alpha_{i-1}), f_{\alpha_i}^{-1}(\alpha_{i+1})) \neq 0 \neq i(f_{\alpha_i}^{-1}(\alpha_{i-1}), f_{\alpha_i}(\alpha_{i+1})).$$

Since ϕ is simplicial and locally injective, we must have $\phi(f_{\alpha_i}(\alpha_{i-1})) = f_{\alpha_i}^{-1}(\alpha_{i-1})$ and $\phi(f_{\alpha_i}^{-1}(\alpha_{i-1})) = f_{\alpha_i}(\alpha_{i-1})$. Consequently,

$$\phi(f_{\alpha_{i-1}}(\alpha_i)) = \phi(f_{\alpha_i}^{-1}(\alpha_{i-1})) = f_{\alpha_i}(\alpha_{i-1}) = f_{\alpha_{i-1}}^{-1}(\alpha_i).$$

Repeating this argument again, it follows that $\phi(f_{\alpha_i}(\alpha_{i+1})) = f_{\alpha_i}^{-1}(\alpha_{i+1})$ for all i, as required. Thus, in this case, ϕ is given by σ as in the statement of the proposition.

If we are not in the situation of the previous paragraph, then it follows that ϕ is the identity, completing the proof. □

After this discussion we are in a position to explain how to obtain an exhaustion of $\mathscr{C}(S)$ by finite rigid sets. Here, $\mathrm{Mod}(S)$ denotes the index 2 subgroup of $\mathrm{Mod}^{\pm}(S)$ consisting of those mapping classes that preserve orientation.

Proposition 3.13. *Let $Y \subset \mathscr{C}(S)$ be a finite rigid set such that* $\mathrm{Mod}(S) \cdot Y = \mathscr{C}(S)$. *Suppose there exists $G \subset Y$ such that*

(1) *the set $\{f_\alpha \mid \alpha \in G\}$ generates* $\mathrm{Mod}(S)$, *and*

(2) *$Y \cap f_\alpha(Y)$ is weakly rigid for all $\alpha \in G$.*

Then there exists a sequence $Y = Y_1 \subset Y_2 \subset \cdots \subset Y_n \subset \cdots$ such that Y_i is a finite rigid set, Y_i has trivial pointwise stabilizer in $\mathrm{Mod}^{\pm}(S)$ *for all i, and*

$$\bigcup_{i \in \mathbb{N}} Y_i = \mathscr{C}(S).$$

Proof. First, the fact that Y is rigid implies that $f_\alpha(Y)$ is rigid for all $\alpha \in Y$. Therefore, the set $Y_2 := Y \cup f_G(Y)$ is also rigid by assumption (2) and repeated application of Lemma 3.7. We now define, for all $n \geq 2$,

$$Y_{n+1} := Y_n \cup f_G(Y_n).$$

By induction, we see that Y_n is rigid for all n and so the first claim follows. Next, the pointwise stabilizer of Y in $\mathrm{Mod}^{\pm}(S)$ is trivial because $Y \cap f_\alpha(Y)$ is weakly rigid. Therefore, Y_n has trivial pointwise stabilizer in $\mathrm{Mod}^{\pm}(S)$, as $Y \subset Y_n$ for all n. Finally, since $\{f_\alpha \mid \alpha \in G\}$ generates $\mathrm{Mod}(S)$ and $\mathrm{Mod}(S) \cdot Y = \mathscr{C}(S)$, it follows that

$$\bigcup_{i \in \mathbb{N}} Y_i = \mathscr{C}(S),$$

which completes the proof. □

We end this section by explaining how Theorem 1.1 implies that curve complexes are simplicially rigid

Proof of Corollary 1.2. Let $S \neq S_{1,2}$, and let $\phi : \mathscr{C}(S) \to \mathscr{C}(S)$ be a locally injective simplicial map. Let $\mathfrak{X}_1 \subset \mathfrak{X}_2 \subset \cdots$ be the exhaustion of $\mathscr{C}(S)$ provided by Theorem 1.1. Since $\mathfrak{X}_i$ is rigid and has trivial pointwise stabilizer in $\mathrm{Mod}^{\pm}(S)$, there exists a unique mapping class $h_i \in \mathrm{Mod}^{\pm}(S)$ such that $h_i|_{\mathfrak{X}_i} = \phi|_{\mathfrak{X}_i}$. Finally, Lemma 3.7 implies that $h_i = h_j$ for all i, j, and thus the result follows. □

4. Punctured spheres

In this section we prove Theorem 1.1 for $S = S_{0,n}$. If $n \leq 3$ then $\mathscr{C}(S)$ is empty and thus the result is trivially true. The case $n = 4$ is dealt with at the end of this section, as it needs special treatment. Thus, from now on we assume that $n \geq 5$. As in [AL] we represent S as the double of an n-gon Δ with vertices removed, and define $\mathfrak{X}$ as

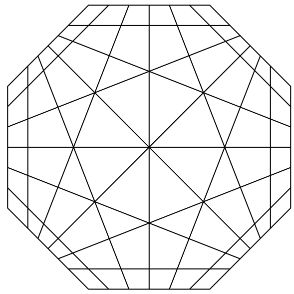

Figure 2. Octagon and arcs for $S_{0,8}$.

the set of curves on S obtained by connecting every nonadjacent pair of sides of Δ by a straight line segment and then doubling; see Figure 2 for the case $n = 8$.

Note that the pointwise stabilizer of $\mathfrak{X}$ in $\mathrm{Mod}^{\pm}(S)$ has order two, and is generated by an orientation-reversing involution $i : S \to S$ that interchanges the two copies of Δ. The rigidity of the set $\mathfrak{X}$, which was established in [AL], may be rephrased as follows:

Theorem 4.1 [AL]. *For any locally injective simplicial map $\phi : \mathfrak{X} \to \mathscr{C}(S)$, there exists a unique $h \in \mathrm{Mod}(S)$ such that $h|_{\mathfrak{X}} = \phi$, unique up to precomposing with i.*

We are going to enlarge the set $\mathfrak{X}$ in the fashion described in Section 3. We number the sides of Δ in a cyclic order, and denote by α_j the curve defined by the arc on Δ that connects the sides with labels j and $j+2 \bmod n$. Let $A = \{\alpha_1, \dots, \alpha_n\}$; in the terminology of [AL], A is the set of *chain curves* of $\mathfrak{X}$. Observe that every element of A bounds a disk containing exactly two punctures of S, and that if two elements of A have nonzero intersection number then they are Farey neighbors in $\mathfrak{X}$. Thus we see that A is a closed string of n twistable Farey neighbors, and may consider the set $\mathfrak{X}_A$ from Definition 3.10. As a first step towards proving Theorem 1.1 for $S_{0,n}$, we show that $\mathfrak{X}_A$ is rigid. Since the pointwise stabilizer of $\mathfrak{X}_A$ is trivial, this amounts to the following statement:

Theorem 4.2. *For any locally injective simplicial map $\phi : \mathfrak{X}_A \to \mathscr{C}(S)$, there exists a unique $g \in \mathrm{Mod}^{\pm}(S)$ such that $g|_{\mathfrak{X}_A} = \phi$.*

Proof. Let $\phi : \mathfrak{X}_A \to \mathscr{C}(S)$ be a locally injective simplicial map. By Theorem 4.1, there exists $h \in \mathrm{Mod}^{\pm}(S)$ such that $h|_{\mathfrak{X}} = \phi|_{\mathfrak{X}}$, unique up to precomposing with the involution i. Since i fixes every element of $\mathfrak{X}$, after precomposing ϕ with h^{-1} we may assume that $\phi|_{\mathfrak{X}}$ is the identity map. We have $\phi(\mathfrak{X}_A) = \mathfrak{X}_A$ by Proposition 3.12; moreover, the automorphism group of $\mathfrak{X}_A$ fixing $\mathfrak{X}$ pointwise has order two, generated by the involution $\sigma : \mathfrak{X}_A \to \mathfrak{X}_A$ that interchanges $f_{\alpha_i}(\alpha_{i+1})$ and $f^{-1}_{\alpha_{i+1}}(\alpha_i)$ for all i. Since $i|_{\mathfrak{X}_A} = \sigma$, up to precomposing ϕ with i, we deduce that $\phi|_{\mathfrak{X}_A}$ is the identity, as we wanted to prove. $\square$

We now prove Theorem 1.1 for spheres with punctures:

Proof of Theorem 1.1 for $S = S_{0,n}$, $n \geq 5$. Let $\mathfrak{X}_A$ be the set constructed above, which is rigid and has trivial pointwise stabilizer in $\mathrm{Mod}^{\pm}(S)$, by Theorem 4.2. The set $\{H_\alpha \mid \alpha \in A\}$ generates $\mathrm{Mod}(S)$; see, for instance, Corollary 4.15 of [Farb and Margalit 2012]. In addition, $\mathfrak{X}_A \cap H_\alpha(\mathfrak{X}_A)$ is weakly rigid for all $\alpha \in A$, as it contains A and $H_{\alpha_i}(\alpha_j)$ for any α_i, α_j disjoint from α. Finally, by inspection we see $\mathrm{Mod}(S) \cdot \mathfrak{X}_A = \mathscr{C}(S)$. Therefore, we may apply Proposition 3.13 to the sets $Y = \mathfrak{X}_A$ and $G = A$ to obtain the desired sequence $\mathfrak{X}_A = Y_1 \subset Y_2 \subset \cdots \subset Y_n \subset \cdots$ of finite rigid sets. □

Proof of Theorem 1.1 for $S = S_{0,4}$. As mentioned in the introduction, in this case $\mathscr{C}(S)$ is isomorphic to the Farey complex. It is easy to see, and is otherwise explicitly stated in [AL], that any triangle in $\mathscr{C}(S)$ is rigid. From this, plus the fact that any edge in $\mathscr{C}(S)$ is contained in exactly two triangles, it follows that any subcomplex of $\mathscr{C}(S)$ that is homeomorphic to a disk is also rigid. Consider the dual graph of $\mathscr{C}(S)$ (which is in fact a trivalent tree T), equipped with the natural path metric. Let Y_1 be a triangle in $\mathscr{C}(S)$, and define Y_n to be the union of all triangles of $\mathscr{C}(S)$ whose corresponding vertices in T are at distance at most n from the vertex corresponding to Y_1. Then the sequence $(Y_n)_{n\in\mathbb{N}}$ gives the desired exhaustion of $\mathscr{C}(S)$. □

5. Closed and punctured surfaces of genus $g \geq 2$

In this section we consider the case of a surface S of genus $g \geq 2$ with $n \geq 0$ marked points. First observe that if $g = 2$ and $n = 0$, then since $\mathscr{C}(S_{2,0}) \cong \mathscr{C}(S_{0,6})$ [Luo 2000], the main theorem for $S_{2,0}$ follows from the case $S_{0,6}$, already proved in Section 4. We therefore assume that $n \geq 1$ if $g = 2$. Once we have recalled some properties of $\mathfrak{X} \subset \mathscr{C}(S)$ from [AL], we sketch the proof of Theorem 1.1 for closed surfaces as it is simpler.

We let $\mathfrak{X} \subset \mathscr{C}(S)$ denote the finite rigid set constructed in [AL]. The definition of the set $\mathfrak{X}$ is somewhat involved and we will not recall it in full detail. Instead, we first note that $\mathfrak{X}$ contains the set of *chain curves*

$$\mathfrak{C} = \{\alpha_0^0, \ldots, \alpha_0^n, \alpha_1, \ldots, \alpha_{2g+1}\}$$

depicted in Figure 3. For notational purposes we also write $\alpha_0 = \alpha_0^1$ (or in case $n = 0$, $\alpha_0 = \alpha_0^0$). In addition to these curves, $\mathfrak{X}$ contains every curve which occurs as the boundary component of a subsurface of S filled by a subset $A \subset \mathfrak{C}$, provided its union is connected in S and has one of the following forms:

(1) $A = \{\alpha_0^i, \alpha_0^j, \alpha_k\}$ where $0 \leq i \leq j \leq n$ and $k = 1$ or $2g+1$.

(2) $A = \{\alpha_0^i, \alpha_0^j, \alpha_k, \alpha_{k+1}\}$ where $0 \leq i \leq j \leq n$ and $k = 1$ or $2g$.

(3) $A = \{\alpha_i \mid i \in I\}$ where $I \subset \{0, \dots, 2g+1\}$ is an interval (modulo $2g+2$). If $n > 0$ and A has an odd number of curves, then we additionally require that the first and last numbers in the interval I be even.

See Figure 4 for some key examples.

The pointwise stabilizer of $\mathfrak{X}$ in $\mathrm{Mod}^{\pm}(S)$ is trivial. Thus the rigidity of the set $\mathfrak{X}$, established in [AL], may be rephrased as follows.

Theorem 5.1 [AL]. *Let $S = S_{g,n}$ with $g \geq 2$ and $n \geq 0$ (and $n \geq 1$ if $g = 2$). For any locally injective simplicial map $\phi : \mathfrak{X} \to \mathscr{C}(S)$, there exists a unique $h \in \mathrm{Mod}^{\pm}(S)$ such that $h|_{\mathfrak{X}} = \phi$.*

Sketch of Theorem 1.1 for closed surfaces. Since the closed case avoids some of the technicalities that arise in the general case, we sketch the proof here. We begin by noting that in [AL] it is shown that $\mathfrak{X}$ contains every curve which occurs as the boundary component of a subsurface of S filled by a subset $A \subset \mathfrak{C}$, provided its union is connected in S, without any further qualifications on the set A.

We enlarge $\mathfrak{X}$ to $\mathfrak{X}^2 = (\mathfrak{X}')'$ and consider the set $\mathfrak{X}^2 \cup T_{\mathfrak{C}}(\mathfrak{C})$. The set $\mathfrak{C}$ is a closed string of twistable Farey neighbors (that the nonzero intersections are $\mathfrak{X}^2$-detectable follows from their $\mathfrak{X}$-detectability proved in [AL]). By Proposition 3.12, this set will be rigid if we can rule out the potential order two symmetry. It thus suffices to show that one of the curves in $T_{\mathfrak{C}}(\mathfrak{C})$ is already in $\mathfrak{X}^2$. This is illustrated for $T_{\alpha_{2g}}(\alpha_{2g-1})$ in Figure 8 (the pictures for a closed surface are obtained by ignoring punctures and any curves which subsequently become trivial, and identifying pairs

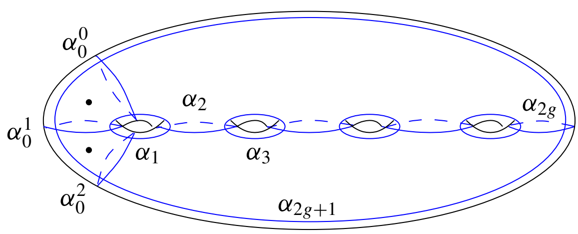

Figure 3. Chain curves $\mathfrak{C}$ on a genus 4 surface with 2 marked points.

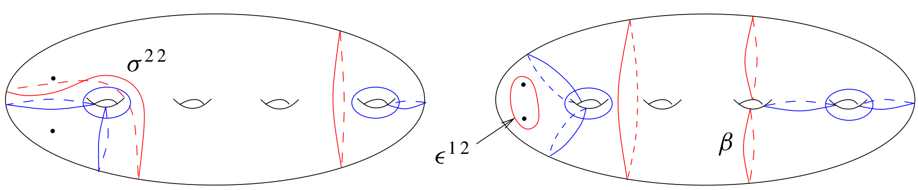

Figure 4. Examples of subsets of $\mathfrak{C}$ (in blue), together with the boundary components (in red) of the subsurface filled by them. The red curves are in $\mathfrak{X}$.

that become isotopic). Therefore, $\mathfrak{X}^2 \cup \mathcal{T}_{\mathfrak{C}}(\mathfrak{C})$ is indeed rigid. One can also find an appropriate closed string of twistable Farey neighbors containing the curve β shown in Figure 4 (see Lemma 5.5 below), and so it follows that $\mathfrak{X}^2 \cup T_{\mathfrak{C}}(\mathfrak{C}) \cup T_\beta(\mathfrak{C})$ is also rigid. Since $T_{\alpha_0}, \ldots, T_{\alpha_{2g+1}}, T_\beta$ generate $\mathrm{Mod}(S)$ (see, e.g., [Farb and Margalit 2012]), Theorem 1.1 follows from Proposition 3.13.

The general case. It will be necessary to refer to some of the curves in $\mathfrak{X}$ by name, so we describe the naming convention briefly in those cases, along the lines of [AL]. We have already described the names of the elements of $\mathfrak{C}$. For $0 < i < j \leq n$ we let ϵ^{ij} be the boundary component of the subsurface $N(\alpha_1 \cup \alpha_0^{i-1} \cup \alpha_0^j)$ that also bounds a $(j-i+1)$-punctured disk in S (containing the i-th through j-th punctures). We call the curves ϵ^{ij} *outer curves*; see Figure 4. For $0 < i \leq j \leq n$, we also consider the other boundary component of $N(\alpha_1 \cup \alpha_0^{i-1} \cup \alpha_0^j)$; this is a separating curve dividing the surface into two (punctured) subsurfaces of genus 1 and $g-1$ respectively. We denote this curve σ^{ij}. One more curve in $\mathfrak{X}$ that we refer to as β is shown in Figure 4, and is a component of the boundary of the subsurface $N(\alpha_{2g-2} \cup \alpha_{2g-1} \cup \alpha_{2g})$.

The strategy for proving Theorem 1.1 for surfaces of genus $g \geq 2$ is similar in spirit to the one for punctured spheres, although considerably more involved. The main idea is to produce successive rigid enlargements of the rigid set $\mathfrak{X}$ identified in [AL], until we are in a position to apply Proposition 3.13. We begin by replacing $\mathfrak{X}$ with $\mathfrak{X}'$, which is rigid by Proposition 3.5. For every $0 < j \leq n$, let

$$A_j = \{\sigma^{ij} \mid 0 < i \leq j\} \cup \{\sigma^{ji} \mid j \leq i \leq n\} \cup \{\alpha_1, \alpha_3, \alpha_4, \alpha_5, \ldots, \alpha_{2g+1}\}.$$

The set A_j is almost filling and uniquely determines a curve denoted α_1^j; see Figure 5. The naming is suggestive, as all α_1^j are homotopic to α_1 upon filling in the punctures.

We can similarly find a subset A_0 (shown in the left of Figure 6) which is almost filling and uniquely determines a curve denoted α_1^0 (shown on the right of Figure 6), which bounds a disk enclosing every puncture of S. Consequently, $\alpha_1^j \in \mathfrak{X}'$ for all $j = 0, \ldots, n$.

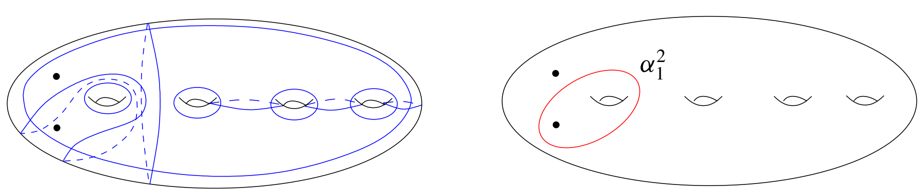

Figure 5. The surface on the left contains the set A_2 which uniquely determines α_1^2.

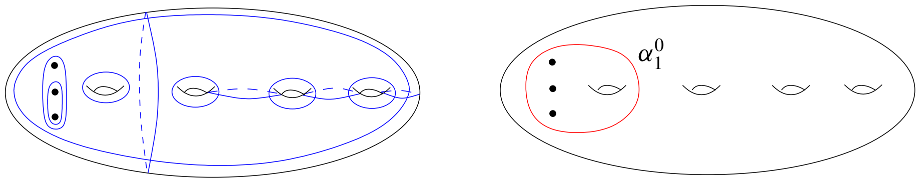

Figure 6. The curves $A_0 \subset \mathfrak{X}$ (left) and the curve $\alpha_1^0 \in \mathfrak{X}'$ (right).

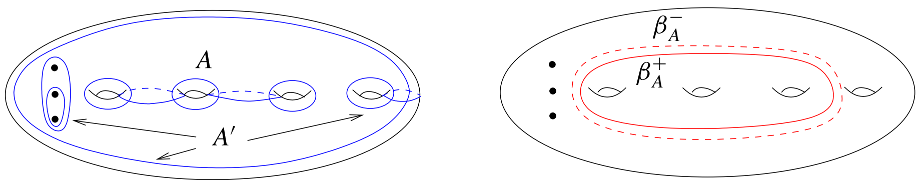

Figure 7. The sets $A = \{\alpha_1, \dots, \alpha_5\} \subset \mathfrak{C}$ and $A' \subset \mathfrak{X}$ (left) and the curves $\beta_A^\pm \in \mathfrak{X}'$ determined by $A \cup A'$ (right).

Punctured surface promotion. One issue that arises only in the case $n > 0$ is that for intervals $I \subset \{0, \dots, 2g+1\}$ (modulo $2g+2$) of odd length, the boundary curves of the neighborhood of the subsurface filled by $A = \{\alpha_i \mid i \in I\}$ are only contained in $\mathfrak{X}$ when I starts and ends with even indexed curves. Passing to the set $\mathfrak{X}'$ allows us to easily enlarge further to a set which rectifies this problem.

Specifically, we define $\mathfrak{X}_1$ to be the union of $\mathfrak{X}'$ together with boundary components of subsurfaces filled by sets $A = \{\alpha_i, \alpha_{i+1}, \dots, \alpha_j\}$ where $0 < i \le j \le 2g-1$ and i, j are both odd. See Figure 7 for examples. Let $\mathfrak{B}_o$ be the set of all curves defined by such sets A.

Before we proceed, we describe this set in more detail. Cutting S open along $\alpha_1 \cup \alpha_3 \cup \cdots \cup \alpha_{2g-1} \cup \alpha_{2g+1}$ we obtain two components Θ_o^+ and Θ_o^-. These are each spheres with holes: Θ_o^+ is the sphere in "front" in Figure 3, which is a $(g+n+1)$-holed sphere containing the n punctures of S, while Θ_o^- is the $(g+1)$-holed sphere in the "back" in Figure 3. For every $A = \{\alpha_i, \alpha_{i+1}, \dots, \alpha_j\}$ where $0 < i < j \le 2g-1$ and i, j are both odd, the boundary of the subsurface filled by A has exactly two components $\beta_A^\pm$ with $\beta_A^+ \subset \Theta_o^+$ and $\beta_A^- \subset \Theta_o^-$ (possibly peripheral in $\Theta_o^\pm$ depending on A). Furthermore, for every such set A, there is a "complementary" set $A' \subset \mathfrak{X}$ such that $A \cup A'$ is almost filling, and such that $\{\beta_A^\pm\}$ is the set determined by $A \cup A'$. See Figure 7.

Lemma 5.2. *For all $g \ge 2$ and $n \ge 1$, the set $\mathfrak{X}_1$ is rigid and has trivial pointwise stabilizer in* $\mathrm{Mod}^\pm(S_{g,n})$.

Proof. First, $\mathfrak{X}_1$ has trivial pointwise stabilizer since $\mathfrak{X}$ does. Given any locally injective simplicial map $\phi : \mathfrak{X}_1 \to \mathscr{C}(S)$, there exists a unique $h \in \mathrm{Mod}^\pm(S)$ such that $\phi = h|_{\mathfrak{X}'}$, by Theorem 5.1 and Proposition 3.5. Composing with the inverse of h if necessary, we can assume ϕ is the identity on $\mathfrak{X}'$. So we need only show that $\phi(\gamma) = \gamma$ for all $\gamma \in \mathfrak{X}_1 - \mathfrak{X}'$. With respect to the notation above, any such curve is $\beta_A^{\pm}$ for $A = \{\alpha_i, \alpha_{i+1}, \ldots, \alpha_j\}$, where $0 < i \le j \le 2g-1$ and i, j are both odd. Since $A \cup A'$ is almost filling, $\phi(\{\beta_A^{\pm}\}) = \{\beta_A^{\pm}\}$. Now, for $A = \{\alpha_1, \alpha_2, \alpha_3\}$, we have $i(\beta_A^+, \alpha_0^1) \neq 0$ and $i(\beta_A^-, \alpha_0^1) = 0$; here, α_0^1 is the curve depicted in Figure 6. Therefore $\phi(\beta_A^+) = \beta_A^+$, as ϕ is locally injective and simplicial. Finally, an easy connectivity argument involving the set of curves $\{\beta_A^{\pm}\}_A$ yields the desired result. $\square$

Half the proof and the case of one or fewer punctures. We now enlarge the set $\mathfrak{X}_1 \subset \mathscr{C}(S)$ from Lemma 5.2 to $\mathfrak{X}_1^2 = (\mathfrak{X}_1')' \subset \mathscr{C}(S)$. According to Proposition 3.5, $\mathfrak{X}_1^2$ is rigid, and since the pointwise stabilizer of $\mathfrak{X}$ is trivial, so is the pointwise stabilizer of $\mathfrak{X}_1^2$. We will need the following lemma; see Figure 3 for the labeling of the curves.

Lemma 5.3. *For any $g \ge 2$ and $n \ge 0$ (with $n \ge 1$ if $g = 2$), we have $T_{\alpha_{2g}}(\alpha_{2g-1}) \in \mathfrak{X}_1^2$.*

Proof. This requires a series of pictures, slightly different for the case $g \ge 3$ and for $g = 2$.

Case 1: $g \ge 3$. We refer the reader to Figure 8; although we have only drawn the figures for $g = 3$ and $n = 2$, it is straightforward to extend them to all $g \ge 3$ and $n \ge 0$. The upper left figure shows an almost filling set of curves contained in $\mathfrak{X}_1$, determining uniquely the curve on the upper right figure, which is thus in $\mathfrak{X}_1'$. This curve is then used to produce an almost filling set, depicted on the lower left hand figure, that uniquely determines $T_{\alpha_{2g}}(\alpha_{2g-1})$, shown on the right. Thus we see that $T_{\alpha_{2g}}(\alpha_{2g-1}) \in \mathfrak{X}_1^2$, as claimed.

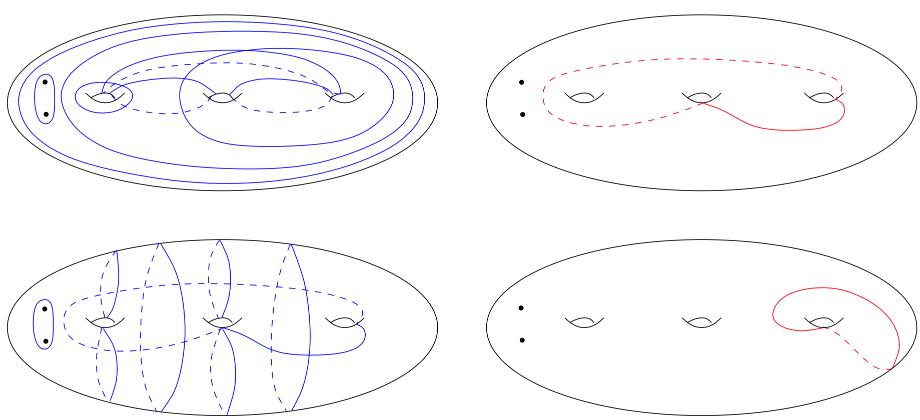

Figure 8. Illustrating $T_{\alpha_{2g}}(\alpha_{2g-1})$ in $\mathfrak{X}_1^2$, when $g = 3$. The almost filling set on the left (blue) uniquely determines the curve in the right figure (red).

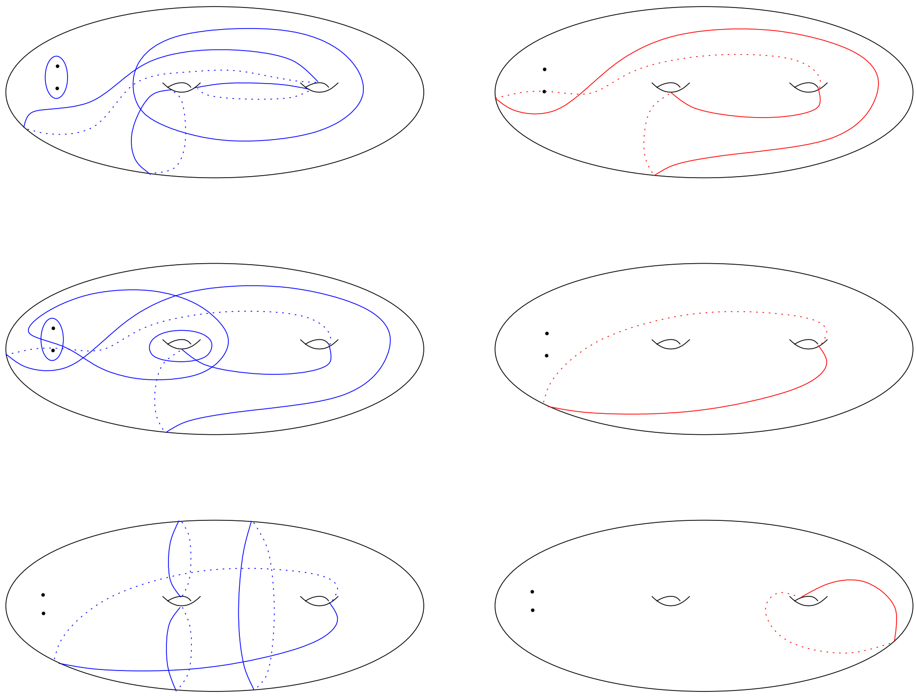

Figure 9. Illustrating $T_{\alpha_{2g}}(\alpha_{2g-1})$ in $\mathfrak{X}_1^2$, when $g = 2$ and $n \geq 1$. The almost filling set on the left (blue) uniquely determines the curve in the right figure (red).

Case 2: $g = 2$ and $n \geq 1$. In this case a different set of pictures is required; see Figure 9. The upper left hand figure shows an almost filling set of curves that is contained in $\mathfrak{X}_1$ and uniquely determines the curve shown on the upper right. This curve is then used to produce an almost filling set, depicted in the middle left picture, which is contained in $\mathfrak{X}_1'$ and uniquely determines the curve in the middle right figure. We now make use of this new curve to produce an almost filling set (lower left) that is contained in $\mathfrak{X}_1^2$ and uniquely determines $T_{\alpha_{2g}}(\alpha_{2g-1})$ (lower right). $\square$

We claim that the set $\mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C})$ is rigid. More concretely:

Lemma 5.4. *Let $\phi : \mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C}) \to \mathscr{C}(S)$ be a locally injective simplicial map. Then there exists a unique $h \in \mathrm{Mod}^{\pm}(S)$ such that $h|_{\mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C})} = \phi$.*

Proof. Let $\phi : \mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C})$ be a locally injective simplicial map. Since $\mathfrak{X}_1^2$ is rigid and its pointwise stabilizer in $\mathrm{Mod}^{\pm}(S)$ is trivial, there exists a unique $h \in \mathrm{Mod}^{\pm}(S)$ such that $h|_{\mathfrak{X}_1^2} = \phi|_{\mathfrak{X}_1^2}$. Precomposing ϕ with h^{-1}, we may assume that in fact $\phi|_{\mathfrak{X}_1^2}$ is the identity map.

For $i = 0, \ldots, n$, $\mathfrak{C}_i = \{\alpha_0^i, \alpha_1, \ldots, \alpha_{2g+1}\}$ is a closed string of twistable Farey neighbors in $\mathfrak{X}_1^2$ (the fact that the nonzero intersection numbers between these curves is $\mathfrak{X}$-detectable, hence $\mathfrak{X}_1^2$-detectable, is shown in the proofs of Theorem 5.1 and 6.1 in [AL]). Consider the set $\mathfrak{X}_1^2 \cup T_{\mathfrak{C}_i}(\mathfrak{C}_i)$, and observe that, in the terminology of Definition 3.10, it equals Y_A for $Y = \mathfrak{X}_1^2$ and $A = \mathfrak{C}_i$. By Proposition 3.12,

$$\phi(\mathfrak{X}_1^2 \cup T_{\mathfrak{C}_i}(\mathfrak{C}_i)) = \mathfrak{X}_1^2 \cup T_{\mathfrak{C}_i}(\mathfrak{C}_i);$$

moreover, the automorphism group of $\mathfrak{X}_1^2 \cup T_{\mathfrak{C}_i}(\mathfrak{C}_i)$ fixing $\mathfrak{X}_1^2$ pointwise has order at most two. But, by Lemma 5.3, $T_{\alpha_{2g}}(\alpha_{2g-1}) \in \mathfrak{X}_1^2$ and thus this group is trivial. In other words, we have shown that the set $\mathfrak{X}_1^2 \cup T_{\mathfrak{C}_i}(\mathfrak{C}_i)$ is rigid.

Now, $\mathfrak{X}_1^2 \cup T_{\mathfrak{C}_0}(\mathfrak{C}_0) \cup T_{\mathfrak{C}_1}(\mathfrak{C}_1)$ is also rigid by Lemma 3.7, since

$$(\mathfrak{X}_1^2 \cup T_{\mathfrak{C}_0}(\mathfrak{C}_0)) \cap (\mathfrak{X}_1^2 \cup T_{\mathfrak{C}_1}(\mathfrak{C}_1))$$

is weakly rigid as it contains $\mathfrak{X}_1^2$. Since $T_{\mathfrak{C}}(\mathfrak{C}) = \bigcup_{i=0}^n T_{\mathfrak{C}_i}(\mathfrak{C}_i)$, we may repeat essentially this same argument $n-1$ more times to conclude $\mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C})$ is rigid, as required. □

Next, we provide a further enlargement of our rigid set. Let β be the curve depicted in Figure 4, which is one of the boundary components of the surface $N(\alpha_{2g-2} \cup \alpha_{2g-1} \cup \alpha_{2g})$. We claim:

Lemma 5.5. *The set $\mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C}) \cup T_\beta(\mathfrak{C})$ is rigid.*

Proof. Let $\phi : \mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C}) \cup T_\beta(\mathfrak{C}) \to \mathscr{C}(S)$ be a locally injective simplicial map. By Lemma 5.4, $\mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C})$ is rigid and thus, up to precomposing ϕ with an element of $\mathrm{Mod}^{\pm}(S)$, we may assume that $\phi|_{\mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C})}$ is the identity. The set

$$A = \{\alpha_{2g}, \alpha_{2g-1}, \alpha_{2g-2}, \beta, \alpha_{2g+1}\} \subset \mathfrak{X}_1^2$$

is a closed string of twistable Farey neighbors in $\mathfrak{X}_1^2$ (again, detectability of the nonzero intersection numbers is shown in [AL]). Therefore, we may apply Proposition 3.12 to $\mathfrak{X} = \mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C})$ and A to deduce that $\phi(\mathfrak{X}_A) = \mathfrak{X}_A$; observe that $\mathfrak{X}_A = \mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C}) \cup T_\beta(\mathfrak{C})$. Moreover, the automorphism group of $\mathfrak{X}_A$ fixing $\mathfrak{X}$ pointwise is trivial, by Lemma 5.4, and thus the result follows. □

Proof of Theorem 1.1 for $g \geq 2$ and $n \leq 1$. Let $Y = \mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C}) \cup T_\beta(\mathfrak{C})$. When S is closed or has one puncture, the Dehn twists about chain curves and the Dehn twist about the curve β generate $\mathrm{Mod}(S)$; see, for example, Corollary 4.15 of [Farb and Margalit 2012]. For $\gamma \in \mathfrak{C} \cup \{\beta\}$, the set

$$T_\gamma(Y) \cap (Y)$$

contains $\mathfrak{C}$, together with $T_\alpha(\alpha')$ for any $\alpha, \alpha' \in \mathfrak{C}$ which are disjoint from γ. In particular, this set is weakly rigid. By inspection, the $\mathrm{Mod}(S)$-orbit of Y is all of $\mathscr{C}(S)$, and so by Proposition 3.13, this set suffices to prove the theorem. □

Multiple punctures. When $S_{g,n}$ has $n \geq 2$ (and $g \geq 2$), the twists in the curves $\mathfrak{C}$ and $\{\beta\}$ do not generate the entire mapping class group. In this case, one needs to add the set of half-twists about the outer curves $\epsilon^{i\,(i+1)}$ bounding twice-punctured disks; see again Corollary 4.15 of [Farb and Margalit 2012]. Because of this, and in light of Proposition 3.13, when $n \geq 2$ we would like to enlarge our rigid set from the previous subsection by adding half-twists of chain curves about outer curves $\epsilon^{i\,(i+1)}$. In fact, denoting this set of outer curves by $\mathfrak{O}_P = \{\epsilon^{i\,(i+1)}\}_{i=1}^{n-1}$ we shall show that these curves are already in $\mathfrak{X}_1^2$. Specifically, we prove:

Lemma 5.6. *We have $H_{\mathfrak{O}_P}(\mathfrak{C}) \subset \mathfrak{X}_1^2$.*

Proof. If $\alpha \in \mathfrak{C}$ and $\epsilon^{j\,(j+1)} \in \mathfrak{O}_P$, then we must show that $H_{\epsilon^{j\,(j+1)}}(\alpha) \in \mathfrak{X}_1^2$ for each $j = 1, \ldots, n-1$. This is clear if $i(\alpha, \epsilon^{j\,(j+1)}) = 0$, since then $H_{\epsilon^{j\,(j+1)}}(\alpha) = \alpha$. The intersection number is nonzero only when $\alpha = \alpha_0^j$, so it suffices to consider only this case.

To prove $H_{\epsilon^{j\,(j+1)}}(\alpha_0^j) \in \mathfrak{X}_1^2$, we need only exhibit the almost filling sets from $\mathfrak{X}_1'$ uniquely determining this curve. This in turn requires an almost filling set from $\mathfrak{X}_1$. As before, we provide the necessary curves in a sequence of two figures. First, the almost filling set on the left of Figure 10 is contained in $\mathfrak{X}'$, and hence in $\mathfrak{X}_1$ (compare with Figure 5), and uniquely determines the curve γ_1 depicted on the right of the same figure. Therefore, $\gamma_1 \in \mathfrak{X}_1'$. Figure 11 is then an almost filling set in $\mathfrak{X}_1'$, and uniquely determines the curve on the right of the same figure. This curve is $H_{\epsilon^{j\,(j+1)}}(\alpha_0^j)$, and so completes the proof. $\square$

We are finally in a position to prove Theorem 1.1 for surfaces of genus $g \geq 2$ and $n \geq 2$.

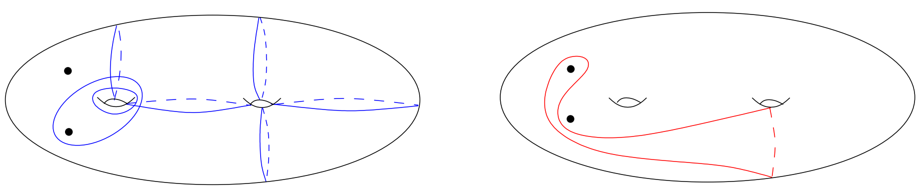

Figure 10. Determining the curve γ_1.

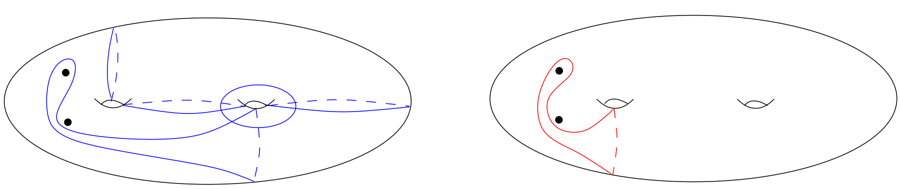

Figure 11. Determining the curve $H_{\epsilon^{i\,(i+1)}}(\alpha_0^i)$.

Proof of Theorem 1.1 for $S = S_{g,n}$, $g \geq 2$, $n \geq 2$.. The set $Y = \mathfrak{X}_1^2 \cup T_{\mathfrak{C}}(\mathfrak{C}) \cup T_\beta(\mathfrak{C})$ is rigid by Lemma 5.6, and has trivial pointwise stabilizer in $\mathrm{Mod}^{\pm}(S)$ since $\mathfrak{X}$ does. Moreover, $\mathrm{Mod}(S)\cdot Y = \mathscr{C}(S)$ by inspection. Consider the subset $G = \mathfrak{C} \cup \{\beta\} \cup \mathfrak{D}_P$; as mentioned before, the (half-)twists about elements of G generate $\mathrm{Mod}(S)$. In addition, for every $\alpha \in G$, $Y \cap f_\alpha(Y)$ is weakly rigid. Thus we can apply Proposition 3.13 to Y and G, hence obtaining the desired exhaustion of $\mathscr{C}(S)$. □

6. Tori

In this section we will prove Theorem 1.1 for $S = S_{1,n}$, for $n \geq 0$. First, if $n \leq 1$ then $\mathscr{C}(S)$ is isomorphic to the Farey complex, and thus the result follows as in the case of $S_{0,4}$; see Section 4. For $n = 2$, Theorem 1.1 is not true as stated due to the existence of *nongeometric* automorphisms of $\mathscr{C}(S)$, as mentioned in the introduction. However, in light of the isomorphism $\mathscr{C}(S_{0,5}) \cong \mathscr{C}(S_{1,2})$ [Luo 2000], the same statement holds after replacing the group $\mathrm{Mod}^{\pm}(S)$ by $\mathrm{Aut}(\mathscr{C}(S))$ in the definition of rigid set, by the results of Section 4.

Therefore, from now on we assume $n \geq 3$. In [AL], we constructed a finite rigid set $\mathfrak{X}$ described as follows. View $S_{1,n}$ as a unit square with n punctures along the horizontal midline and the sides identified. The set $\mathfrak{X}$ contains a subset $\mathfrak{C} \subset \mathfrak{X}$ of $n+1$ *chain curves*

$$\mathfrak{C} = \{\alpha_1, \ldots, \alpha_n\} \cup \{\beta\}$$

where $\alpha_1, \ldots, \alpha_n$ are distinct curves which appear as vertical lines in the square and β is the curve which appears as a horizontal line; see Figure 12. We assume that the indices on the α_i are ordered cyclically around the torus, and that the punctures are labeled so that the i-th puncture lies between α_i and α_{i+1}. The boundaries of the subsurfaces filled by connected unions of these chain curves form a collection

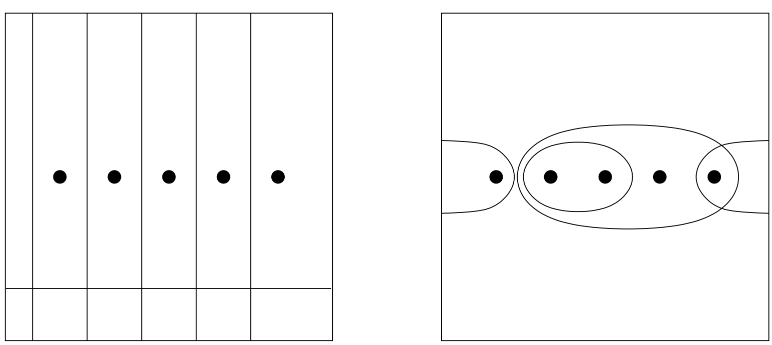

Figure 12. Chain curves on the left, and some examples of outer curves on the right, in $S_{1,5}$.

of curves, denoted $\mathfrak{O}$ which we refer to as *outer curves*. Then

$$\mathfrak{X} = \mathfrak{C} \cup \mathfrak{O}.$$

This set has a nontrivial pointwise stabilizer in $\mathrm{Mod}^{\pm}(S_{1,n})$, which can be realized as the (descent to $S_{1,n}$ of the) horizontal reflection of the square through the midline containing the punctures. Denoting this involution $r : S_{1,n} \to S_{1,n}$, we summarize the result of [AL] in the following theorem.

Theorem 6.1 [AL]. *For any locally injective simplicial map $\phi : \mathfrak{X} \to \mathscr{C}(S_{1,n})$ there exists $h \in \mathrm{Mod}^{\pm}(S_{1,n})$ such that $h|_{\mathfrak{X}} = \phi$. Moreover, h is unique up to precomposing with r.*

The strategy of proof is again similar to that of previous sections, although the technicalities are different, and boils down to producing an enlargement of the set $\mathfrak{X}$ so that Proposition 3.13 can be applied.

We begin by enlarging the set $\mathfrak{X}$ as follows. We let δ_i be the curve coming from the vertical line through the i-th puncture in the square. For every $1 \le i \le n$, let β_i^+ be the curve obtained from β by pushing it up over the i-th puncture. More precisely, we consider the point-pushing homeomorphism $f_i : S_{1,n} \to S_{1,n}$ that pushes the i-th puncture up and around δ_i, and then let $\beta_i^+ = f_i(\beta)$. We similarly define $\beta_i^- = f_i^{-1}(\beta)$, and set $\beta^{\pm}_{i\,(i+1)} = f^{\pm 1}_{i+1} f^{\pm 1}_i(\beta)$, where the subscripts are taken modulo n. See Figure 13.

Let

$$\mathfrak{X}_1 = \mathfrak{X} \cup \{\beta_i^{\pm} \mid 1 \le i \le n\} \cup \{\beta^{\pm}_{i\,(i+1)} \mid 1 \le i \le n\}$$

with indices in the last set taken modulo n. We first prove that this set is rigid; since the pointwise stabilizer of $\mathfrak{X}_1$ in $\mathrm{Mod}^{\pm}(S_{1,n})$ is trivial, this amounts to the following proposition.

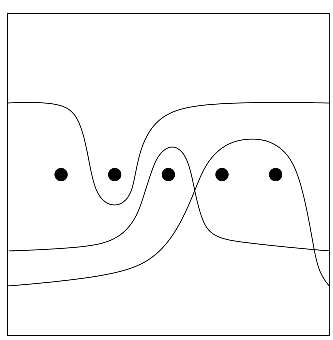

Figure 13. Curves β_2^-, β_3^+, and $\beta_{4\,5}^+$ on $S_{1,5}$.

Proposition 6.2. *For any locally injective simplicial map $\phi : \mathfrak{X}_1 \to \mathscr{C}(S_{1,n})$, there exists a unique $h \in \mathrm{Mod}^{\pm}(S_{1,n})$ so that $h|_{\mathfrak{X}_1} = \phi$.*

The proof of this proposition will require a repeated application of Lemma 2.4, and as such, we must verify that certain quadruples of curves satisfy the hypotheses of that lemma. We will need to refer to the outer curves by name. To this end, note that since any outer curve surrounds a set of (cyclically) consecutive punctures, we can determine an outer curve by specifying the first and last puncture surrounded. Consequently, we let ϵ^{ij} denote the outer curve surrounding all punctures from the i-th to the j-th, with all indices taken modulo n. Observe that since the set of punctures is cyclically ordered, we do not need to assume that $i < j$ in the definition of ϵ^{ij}. We will need the following lemma.

Lemma 6.3. *For each $1 \le i \le n$, consider the following four quadruples of curves in $\mathfrak{X}_1$, with indices taken modulo n:*

- $\beta^{\pm}_{(i-1)\,i}, \epsilon^{(i+1)\,i}, \beta^{\pm}_{i}, \epsilon^{(i-1)\,i}$,
- $\beta^{\pm}_{i\,(i+1)}, \epsilon^{i\,(i-1)}, \beta^{\pm}_{i}, \epsilon^{i\,(i+1)}$.

Each of these satisfies the hypothesis of Lemma 2.4. Furthermore, the nonzero intersections are all $\mathfrak{X}_1$-detectable. Consequently, $\epsilon^{(i+1)\,i}$ and $\epsilon^{i\,(i-1)}$ are the unique Farey neighbors of β^{-}_{i} and β^{+}_{i}.

Proof. The fact that the four quadruples of curves each satisfy the hypothesis of Lemma 2.4 is clear by inspection. See the left side of Figure 14 for the case

$$\beta^{-}_{(i-1)\,i}, \epsilon^{(i+1)\,i}, \beta^{-}_{i}, \epsilon^{(i-1)\,i}.$$

The four-holed sphere N filled by the Farey neighbors $\epsilon^{(i+1)\,i}$, $\beta^{\pm}_{i}$ and $\epsilon^{i\,(i-1)}$, $\beta^{\pm}_{i}$ has holes corresponding to the i-th puncture and the curves β and $\epsilon^{(i+1)\,(i-1)}$; see the right side of Figure 14. Only $\epsilon^{(i+1)\,(i-1)}$ intersects $\beta^{\pm}_{(i-1)\,i}, \beta^{\pm}_{i\,(i+1)}, \epsilon^{(i-1)\,i}, \epsilon^{i\,(i+1)}$ nontrivially, as required for Lemma 2.4.

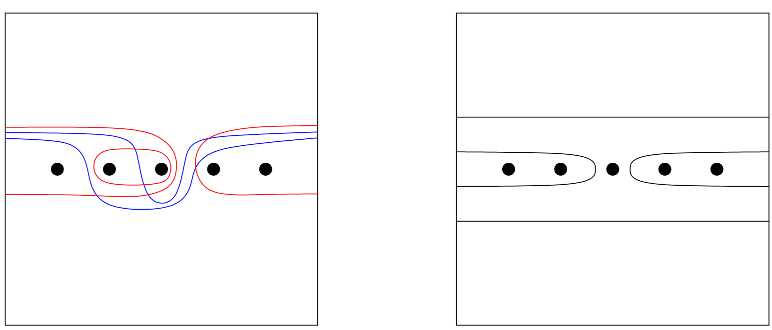

Figure 14. The curves $\beta^{-}_{2\,3}, \epsilon^{4\,3}, \beta^{-}_{3}, \epsilon^{2\,3}$ on the left. The four-holed sphere filled by $\epsilon^{4\,3}$ and $\beta^{\pm}_{3}$ on the right.

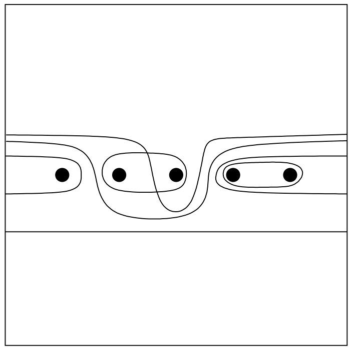

Figure 15. We use $\{\beta, \beta_{23}^-, \epsilon^{45}, \epsilon^{41}\}$ to detect $i(\beta_3^-, \epsilon^{23}) \neq 0$.

To see that all the intersections are $\mathfrak{X}_1$-detectable, we need only exhibit the necessary curves in $\mathfrak{X}_1$ determining a pants decomposition of $S - N$. See Figure 15 for the curves necessary to detect $i(\beta_i^-, \epsilon^{(i-1)\,i}) \neq 0$. We leave the other cases to the reader. $\square$

We are now in a position to prove Proposition 6.2.

Proof of Proposition 6.2. Let $\phi : \mathfrak{X}_1 \to \mathscr{C}(S_{1,n})$ be a locally injective simplicial map. By Theorem 6.1, there exists $f \in \mathrm{Mod}^{\pm}(S_{1,n})$ such that $f|_{\mathfrak{X}} = \phi|_{\mathfrak{X}}$, unique up to precomposing with r. In fact, after precomposing ϕ with f^{-1} we may as well assume that $\phi|_{\mathfrak{X}}$ is the identity.

According to Lemma 6.3, for all i, $\phi(\epsilon^{(i+1)\,i}) = \epsilon^{(i+1)\,i}$ and $\phi(\epsilon^{i\,(i-1)}) = \epsilon^{i\,(i-1)}$ are the unique Farey neighbors of $\phi(\beta_i^-)$ and $\phi(\beta_i^+)$ (with indices taken modulo n). Consequently, $\phi(\{\beta_i^{\pm}\}) = \{\beta_i^{\pm}\}$ for all i. Notice that $i(\beta_i^+, \beta_j^-) = 0$ for all i, j, while $i(\beta_i^+, \beta_j^+) = i(\beta_i^-, \beta_j^-) = 2$ for all i, j. It follows that if $\phi(\beta_i^-) = \beta_i^+$ for some i, then this is true for all i. Composing with r if necessary, we deduce that $\phi(\beta_i^{\pm}) = \beta_i^{\pm}$ for all i. All that remains is to see that $\phi(\beta_{i\,(i+1)}^{\pm}) = \beta_{i\,(i+1)}^{\pm}$ for all i.

To prove this we need only show that

$$\beta_{i\,(i+1)}^{\pm} \in (\mathfrak{X} \cup \{\beta_j^{\pm} \mid 1 \le j \le n\})',$$

and then we can apply Proposition 3.5. First note that if $n = 3$, then $\beta_{i\,(i+1)}^{\pm} = \beta_{i+2}^{\mp}$, so there is nothing to prove in this case. In general, one readily checks that $\beta_{i\,(i+1)}^{+}$ is uniquely determined by the almost filling set

$$\{\beta, \beta_1^-, \beta_2^-, \ldots, \beta_n^-\} \setminus \{\beta_i^-, \beta_{i+1}^-\}.$$

This completes the proof. $\square$

Let $\mathfrak{O}_P = \{\epsilon^{i\,(i+1)}\}_{i=1}^n$, counting indices modulo n. For $n \geq 5$, this is a closed string of twistable Farey neighbors in $\mathfrak{X}_1$, and we could appeal to Proposition 3.12

to add the half-twists about curves in $\mathfrak{O}_P$ in this case. However, we can provide a single argument for all $n \geq 3$.

Lemma 6.4. *For all $\epsilon, \epsilon' \in \mathfrak{O}_P$, $H_\epsilon^{\pm 1}(\epsilon') \in \mathfrak{X}'$. Consequently, $H_\epsilon^{\pm 1}(\mathfrak{X}'_1) \cup \mathfrak{X}'_1$ is rigid.*

Proof. We start with the proof of the first statement. If $i(\epsilon, \epsilon') = 0$, then there is nothing to prove. Otherwise, up to a homeomorphism we may assume that $\epsilon = \epsilon^{i\,(i+1)}$ and $\epsilon' = \epsilon^{(i+1)\,(i+2)}$. Then we note that $H_{\epsilon^{i\,(i+1)}}(\epsilon^{(i+1)\,(i+2)})$ is the curve uniquely determined by the almost filling set of curves

$$\{\beta_{i+1}^+\} \cup \{\alpha_1, \dots, \alpha_n\} \setminus \{\alpha_{i+1}, \alpha_{i+2}\},$$

completing the proof of the first statement.

For the second statement, we note that $H_{\epsilon^{i\,(i+1)}}(\mathfrak{X}'_1) \cap \mathfrak{X}'_1$ contains the weakly rigid set $\mathfrak{O}_P \cup \{\beta_{i+2}^+\}$, for example. Therefore, since $\mathfrak{X}_1$ is rigid by Proposition 6.2, so is $\mathfrak{X}'_1$ by Proposition 3.5, and hence by Lemma 3.7 it follows that $H_{\epsilon^{i\,(i+1)}}(\mathfrak{X}'_1) \cup \mathfrak{X}'_1$ is rigid, as required. A similar argument proves the statement for $H^{-1}_{\epsilon^{i\,(i+1)}}$. □

We also need to consider Dehn twists in α_i and β. To deal with these, we first define $\mathfrak{X}_2 = \mathfrak{X}'_1 \cup H_{\mathfrak{O}_P}(\mathfrak{X}'_1)$, where $H_{\mathfrak{O}_P}(\mathfrak{X}'_1)$ is the union of $H_\epsilon^{\pm 1}(\mathfrak{X}'_1)$ over all $\epsilon \in \mathfrak{O}_P$. By Lemma 6.4, $\mathfrak{X}_2$ is rigid.

Lemma 6.5. *For all $i = 1, \dots, n$, we have $T_{\alpha_i}^{\pm 1}(\beta) = T_\beta^{\mp 1}(\alpha_i) \in \mathfrak{X}_1^2 \subset \mathfrak{X}_2^2$. Consequently, $T_{\alpha_i}^{\pm 1}(\mathfrak{X}_2^2) \cup \mathfrak{X}_2^2$ and $T_\beta^{\pm 1}(\mathfrak{X}_2^2) \cup \mathfrak{X}_2^2$ are rigid.*

Proof. As in previous arguments, we exhibit a series of pictures that will yield the desired result; see Figure 16. It is straightforward to modify such pictures to treat the case of an arbitrary $n \geq 3$. The top left picture shows an almost filling set in $\mathfrak{X}_1$ that uniquely determines a curve in $\mathfrak{X}'_1$ on the top right. Then the lower left is an almost filling set in $\mathfrak{X}'_1$ that uniquely determines the curve in $(\mathfrak{X}'_1)' = \mathfrak{X}_1^2$. This curve is precisely $T_\beta(\alpha_i) = T_{\alpha_i}^{-1}(\beta)$. Similarly, $T_{\alpha_i}(\beta) = T_\beta^{-1}(\alpha_i) \in \mathfrak{X}_1^2$.

Finally, we easily observe that $\mathfrak{X}_2^2 \cap T_{\alpha_i}(\mathfrak{X}_2^2)$ is weakly rigid, as it contains $\mathfrak{C} \cup H_{\epsilon^{(i-1)\,i}}(\alpha_{i-1})$, which is weakly rigid. Appealing to Lemma 3.7, it follows that $\mathfrak{X}_2^2 \cup T_{\alpha_i}(\mathfrak{X}_2^2)$ is rigid. The other cases follow similarly. □

Finally, we prove our main result for surfaces of genus 1.

Proof of Theorem 1.1 for $S = S_{1,n}$. Since $\mathfrak{X}_2$ is rigid, by Propositions 3.5, the set $Y = \mathfrak{X}_2^2$ is rigid. Moreover, $\mathrm{Mod}(S_{1,n}) \cdot Y = \mathscr{C}(S_{1,n})$, by inspection. A generating set for $\mathrm{Mod}(S_{1,n})$ is given by the Dehn twists f_α about the elements $\alpha \in \mathfrak{C}$ and the half-twists f_ϵ about the elements $\epsilon \in A = \{\epsilon^{i\,(i+1)}\}$ (see Section 4.4 of [Farb and Margalit 2012], for instance). Let $G = \mathfrak{C} \cup A$ and note that, for each $\alpha \in G$, the set $Y \cup f_\alpha Y$ is rigid. Therefore, we may apply Proposition 3.13 to obtain the desired exhaustion of $\mathscr{C}(S_{1,n})$ by finite rigid sets. □

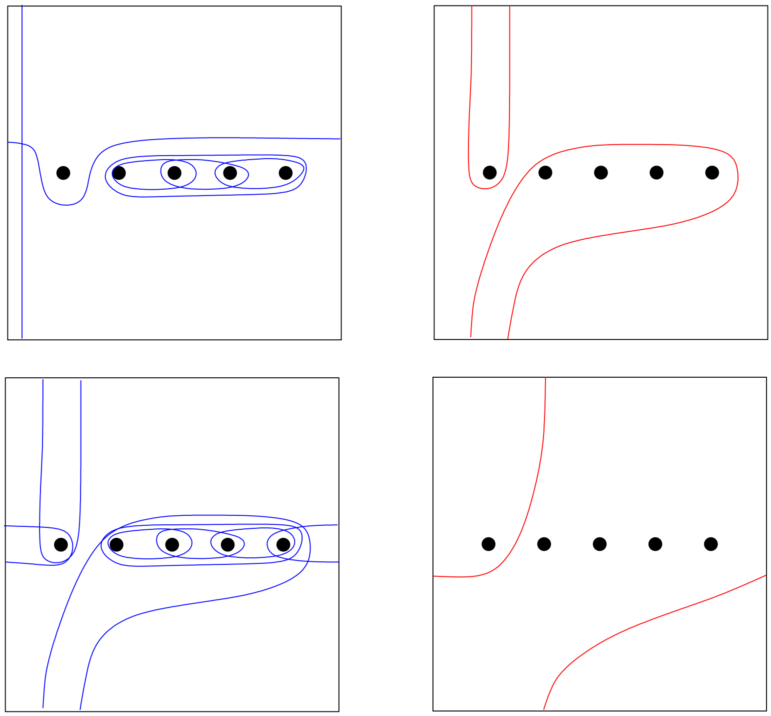

Figure 16. Illustrating $T_\beta(\alpha_2) \in \mathfrak{X}_2^2$ on $S_{1,5}$.

Acknowledgements

We thank Brian Bowditch for conversations and for his continued interest in this work. We also thank Juan Souto for conversations. Finally, we thank the referee for her/his comments. Parts of this paper were completed during the conference "Mapping class groups and Teichmüller Theory"; we would like to express our gratitude to Michah Sageev and the Technion for their hospitality and financial support.

References

[Aramayona and Leininger 2013] J. Aramayona and C. J. Leininger, "Finite rigid sets in curve complexes", *J. Topol. Anal.* **5**:2 (2013), 183–203. MR 3062946 Zbl 1277.57017

[Aramayona and Souto 2013] J. Aramayona and J. Souto, "A remark on homomorphisms from right-angled Artin groups to mapping class groups", *C. R. Math. Acad. Sci. Paris* **351**:19-20 (2013), 713–717. MR 3125410 Zbl 1283.57005

[Birman et al. 2015] J. S. Birman, N. Broaddus, and W. W. Menasco, "Finite rigid sets and homologically nontrivial spheres in the curve complex of a surface", *J. Topol. Anal.* **7**:1 (2015), 47–71. MR 3284389 Zbl 1308.57009

[Broaddus 2012] N. Broaddus, "Homology of the curve complex and the Steinberg module of the mapping class group", *Duke Math. J.* **161**:10 (2012), 1943–1969. MR 2954621 Zbl 1250.57032

[Farb and Margalit 2012] B. Farb and D. Margalit, *A primer on mapping class groups*, Princeton Mathematical Series **49**, Princeton University Press, 2012. MR 2012h:57032 Zbl 1245.57002

[Harer 1986] J. L. Harer, "The virtual cohomological dimension of the mapping class group of an orientable surface", *Invent. Math.* **84**:1 (1986), 157–176. MR 87c:32030 Zbl 0592.57009

[Ivanov 1997] N. V. Ivanov, "Automorphism of complexes of curves and of Teichmüller spaces", *Int. Math. Res. Notices* **1997**:14 (1997), 651–666. MR 98j:57023 Zbl 0890.57018

[Kim and Koberda 2013] S.-H. Kim and T. Koberda, "Embedability between right-angled Artin groups", *Geom. Topol.* **17**:1 (2013), 493–530. MR 3039768 Zbl 1278.20049

[Kim and Koberda 2015] S.-H. Kim and T. Koberda, "Right-angled Artin groups and finite subgraphs of curve graphs", *Osaka J. Math.* (online publication May 2015).

[Korkmaz 1999] M. Korkmaz, "Automorphisms of complexes of curves on punctured spheres and on punctured tori", *Topology Appl.* **95**:2 (1999), 85–111. MR 2000d:57025 Zbl 0926.57012

[Luo 2000] F. Luo, "Automorphisms of the complex of curves", *Topology* **39**:2 (2000), 283–298. MR 2000j:57045 Zbl 0951.32012

[Shackleton 2007] K. J. Shackleton, "Combinatorial rigidity in curve complexes and mapping class groups", *Pacific J. Math.* **230**:1 (2007), 217–232. MR 2008g:57018 Zbl 1165.57017

Received October 24, 2014. Revised September 11, 2015.

JAVIER ARAMAYONA
DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD AUTÓNOMA DE MADRID
CIUDAD UNIVERSITARIA DE CANTOBLANCO
28049 MADRID
SPAIN

aramayona@gmail.com

CHRISTOPHER J. LEININGER
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS AT URBANA–CHAMPAIGN
1409 W. GREEN STREET
URBANA, IL 61801
UNITED STATES

c.j.leininger95@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.285

A VARIATIONAL CHARACTERIZATION OF FLAT SPACES IN DIMENSION THREE

GIOVANNI CATINO, PAOLO MASTROLIA AND DARIO D. MONTICELLI

We prove that, in dimension three, flat metrics are the only complete metrics with nonnegative scalar curvature which are critical for the σ_2-curvature functional.

1. Introduction

Let (M^n, g) be a Riemannian manifold of dimension $n \geq 3$. To fix the notation, we recall the decomposition of the Riemann curvature tensor of a metric g into the Weyl, Ricci, and scalar curvature components:

$$\mathrm{Rm} = W + \frac{1}{n-2}\mathrm{Ric} \owedge g - \frac{1}{(n-1)(n-2)} R g \owedge g,$$

where $\owedge$ denotes the Kulkarni–Nomizu product. It is well known [Hilbert 1915] that Einstein metrics are critical points for the Einstein–Hilbert functional

$$\mathcal{H} = \int R \, dV$$

on the space of unit volume metrics $\mathcal{M}_1(M^n)$. From this perspective, it is natural to study canonical metrics which arise as solutions of the Euler–Lagrange equations for more general curvature functionals. Berger [1970] commenced the study of Riemannian functionals which are quadratic in the curvature (see [Besse 2008, Chapter 4] and [Smolentsev 2005] for surveys). A basis for the space of quadratic curvature functionals is given by

$$\mathcal{W} = \int |W|^2 \, dV, \qquad \rho = \int |\mathrm{Ric}|^2 \, dV, \qquad \mathcal{S} = \int R^2 \, dV.$$

All such functionals, which also naturally arise as total actions in certain gravitational field theories in physics, have been deeply studied in recent years by many authors, in particular on compact Riemannian manifolds with normalized volume (for instance, see [Berger 1970; Besse 2008; Lamontagne 1994; 1998; Anderson 1997; Gursky and Viaclovsky 2001; 2015; 2013; Catino 2015] and references therein).

MSC2010: primary 53C24; secondary 53C25.
Keywords: critical metrics, quadratic functionals, rigidity results.

On the other hand, the study of critical metrics for quadratic curvature functionals also has a lot of interest in the noncompact setting. For instance, Anderson [2001] proved that every complete three-dimensional critical metric for the Ricci functional ρ with nonnegative scalar curvature is flat; whereas, Catino [2014] showed a characterization of complete critical metrics for $\mathcal{S}$ with nonnegative scalar curvature in every dimension.

In this paper we focus our attention on the three-dimensional case and consider the σ_2-curvature functional

$$\mathscr{F}_2 = \int \sigma_2(A)\, dV,$$

where $\sigma_2(A)$ denotes the second elementary symmetric function of the eigenvalues of the Schouten tensor $A = \mathrm{Ric} - \frac{1}{4} R\, g$. This functional was first considered by Gursky and Viaclovsky in the compact three-dimensional case. In [2001] they proved a beautiful characterization theorem of space forms as critical metrics for $\mathscr{F}_2$ on $\mathscr{M}_1(M^3)$ with nonnegative energy $\mathscr{F}_2 \geq 0$.

The main result of this paper is the following variational characterization of three-dimensional flat spaces.

Theorem 1.1. *Let (M^3, g) be a complete critical metric for $\mathscr{F}_2$ with nonnegative scalar curvature. Then (M^3, g) is flat.*

We remark the fact that the nonnegativity condition on the scalar curvature cannot be removed. This is clear from the example in [loc. cit.] where the authors exhibit an explicit family of critical metrics for $\mathscr{F}_2$ on $\mathbb{R}^3$. For instance, the metric given in standard coordinates by

$$g = dx^2 + dy^2 + (1 + x^2 + y^2)^2 dz^2$$

is complete, critical and has strictly negative scalar curvature

$$R = -\frac{8}{1 + x^2 + y^2}.$$

2. The Euler–Lagrange equation for $\mathcal{F}_t$

In this section we will compute the Euler–Lagrange equation satisfied by critical metrics for $\mathscr{F}_2$. To begin, we observe that, in dimension $n \geq 3$, the second elementary symmetric function of the eigenvalues of the Schouten tensor

$$A = \frac{1}{n-2}\Big(\mathrm{Ric} - \frac{1}{2(n-1)}\, R\, g\Big)$$

can be written as

$$\sigma_2(A) = -\frac{1}{2(n-2)^2}|\mathrm{Ric}|^2 + \frac{n}{8(n-1)(n-2)^2} R^2.$$

In particular, the functional $\mathscr{F}_2$ is proportional to a general quadratic functional of the form

$$\mathcal{F}_t = \int |\mathrm{Ric}|^2 \, dV + t \int R^2 \, dV,$$

with the choice $t = -n/4(n-1)$; see also [Gursky and Viaclovsky 2015; Catino 2015]. The gradients of the functionals ρ and $\mathcal{S}$, computed using compactly supported variations, are given by [Besse 2008, Proposition 4.66]

$$(\nabla\rho)_{ij} = -\Delta R_{ij} - 2R_{ikjl}R_{kl} + \nabla^2_{ij}R - \tfrac{1}{2}(\Delta R)g_{ij} + \tfrac{1}{2}|\mathrm{Ric}|^2 g_{ij}$$

and

$$(\nabla\mathcal{S})_{ij} = 2\nabla^2_{ij}R - 2(\Delta R)g_{ij} - 2RR_{ij} + \tfrac{1}{2}R^2 g_{ij}.$$

Hence, the gradient of $\mathcal{F}_t$ reads

$$\begin{aligned}(\nabla\mathcal{F}_t)_{ij} = -\Delta R_{ij} + (1+2t)\nabla^2_{ij}R - \tfrac{1}{2}(1+4t)(\Delta R)g_{ij} \\ + \tfrac{1}{2}\big(|\mathrm{Ric}|^2 + tR^2\big)g_{ij} - 2R_{ikjl}R_{kl} - 2tRR_{ij}.\end{aligned}$$

Tracing the equation $(\nabla\mathcal{F}_t) = 0$, we obtain

$$\big(n + 4(n-1)t\big)\Delta R = (n-4)\big(|\mathrm{Ric}|^2 + tR^2\big).$$

Defining the tensor E to be the traceless Ricci tensor, $E_{ij} = R_{ij} - \frac{1}{n}Rg_{ij}$, we obtain the Euler–Lagrange equation of critical metrics for $\mathcal{F}_t$.

Proposition 2.1. *Let M^n be a complete manifold of dimension $n \geq 3$. A metric g is critical for $\mathcal{F}_t$ if and only if it satisfies*

$$\begin{aligned}\Delta E_{ij} = (1+2t)\nabla^2_{ij}R - \frac{n+2+4nt}{2n}(\Delta R)g_{ij} \\ - 2R_{ikjl}E_{kl} - \frac{2+2nt}{n}RE_{ij} + \frac{1}{2}\Big(|\mathrm{Ric}|^2 - \frac{4-n(n-4)t}{n^2}R^2\Big)g_{ij}\end{aligned}$$

and

$$\big(n + 4(n-1)t\big)\Delta R = (n-4)\big(|\mathrm{Ric}|^2 + tR^2\big).$$

In dimension three we recall the decomposition of the Riemann curvature tensor

$$R_{ikjl} = E_{ij}g_{kl} - E_{il}g_{jk} + E_{kl}g_{ij} - E_{kj}g_{il} + \tfrac{1}{6}R(g_{ij}g_{kl} - g_{il}g_{jk}).$$

In particular,

$$R_{ikjl}E_{kl} = -2E_{ip}E_{jp} - \tfrac{1}{6}RE_{ij} + |E|^2 g_{ij}.$$

Hence, if $n = 3$ and $t = -n/4(n-1) = -3/8$, one has

$$\mathscr{F}_2 = -\tfrac{1}{2}\mathcal{F}_{-3/8},$$

and the following formulas hold.

Proposition 2.2. *Let M^3 be a complete manifold of dimension three. A metric g is critical for $\mathfrak{F}_2$ if and only if it satisfies*

$$\Delta E_{ij} = \tfrac{1}{4}\nabla^2_{ij}R - \tfrac{1}{12}(\Delta R)g_{ij} + 4E_{ip}E_{jp} + \tfrac{5}{12}RE_{ij} - \tfrac{1}{2}\big(3|E|^2 - \tfrac{1}{72}R^2\big)g_{ij} \tag{2-1}$$

and

$$-2\sigma_2(A) = |\mathrm{Ric}|^2 - \tfrac{3}{8}R^2 = |E|^2 - \tfrac{1}{24}R^2 = 0. \tag{2-2}$$

Now, contracting (2-1) with E, we obtain the following Weitzenböck formula.

Corollary 2.3. *Let M^3 be a complete manifold of dimension three. If g is a critical metric for $\mathfrak{F}_2$, then the following formula holds*

$$\tfrac{1}{2}\Delta|E|^2 = |\nabla E|^2 + \tfrac{1}{4}E_{ij}\nabla^2_{ij}R + 4E_{ip}E_{jp}E_{ij} + \tfrac{5}{12}R|E|^2. \tag{2-3}$$

3. Proof of Theorem 1.1

In this section we will prove Theorem 1.1. We assume that (M^3, g) is a critical metric for $\mathfrak{F}_2$ with nonnegative scalar curvature $R \geq 0$. In particular, g has zero σ_2-curvature, i.e., $|E|^2 = \frac{1}{24}R^2$ and we obtain

$$\tfrac{1}{2}\Delta|E|^2 = \tfrac{1}{48}\Delta R^2 = \tfrac{1}{24}R\Delta R + \tfrac{1}{24}|\nabla R|^2.$$

Putting together this equation with (2-3), we obtain that the scalar curvature R satisfies the PDE

$$\tfrac{1}{24}\big(Rg_{ij} - 6E_{ij}\big)\nabla^2_{ij}R = |\nabla E|^2 - \tfrac{1}{24}|\nabla R|^2 + 4E_{ip}E_{jp}E_{ij} + \tfrac{5}{12}R|E|^2. \tag{3-1}$$

To begin, we need the following purely algebraic lemmas.

Lemma 3.1. *Let (M^3, g) be a Riemannian manifold with $R \geq 0$ and $\sigma_2(A) \geq 0$. Then,*

$$Rg_{ij} \geq 6E_{ij}$$

and g has nonnegative sectional curvature.

Proof. Let $\lambda_1 \leq \lambda_2 \leq \lambda_3$ be the eigenvalues of the Schouten tensor $A = E + \frac{1}{12}Rg$ at some point. Then, by the assumptions, we have

$$4R = \mathrm{tr}(A) = \lambda_1 + \lambda_2 + \lambda_3 \geq 0 \quad \text{and} \quad \sigma_2(A) = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 \geq 0.$$

We want to show that $E \leq \frac{1}{6}Rg$ or, equivalently, that

$$A \leq \tfrac{1}{4}Rg = \mathrm{tr}(A)g.$$

Hence, it suffices to prove that $\lambda_3 \leq \mathrm{tr}(A) = \lambda_1 + \lambda_2 + \lambda_3$, i.e., that $\lambda_1 + \lambda_2 \geq 0$. But this follows by

$$0 \leq \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 = (\lambda_1 + \lambda_2)\,\mathrm{tr}(A) - (\lambda_1^2 + \lambda_2^2 + \lambda_1\lambda_2) \leq (\lambda_1 + \lambda_2)\,\mathrm{tr}(A).$$

The fact that g has nonnegative sectional curvature follows from the decomposition of the Riemann tensor in dimension three and the curvature condition $\mathrm{Ric} \leq \frac{1}{2} Rg$ (for instance see [Hamilton 1982, Corollary 8.2]). $\square$

Lemma 3.2. *Let (M^3, g) be a Riemannian manifold with $R \geq 0$ and $\sigma_2(A) = \mathrm{const} \geq 0$. Then,*

$$|\nabla E|^2 \geq \tfrac{1}{24} |\nabla R|^2.$$

Proof. We will follow the proof in [Gursky and Viaclovsky 2001, Lemma 4.1]. Let p be a point in M^3. If $R(p) = 0$, then $\nabla R = 0$ and the lemma follows. So we can assume that $R(p) > 0$. Since $-2\sigma_2(A) = |E|^2 - \frac{1}{24} R^2 = \mathrm{const}$,

$$|E|^2 \big|\nabla |E|\big|^2 = \tfrac{1}{576} R^2 |\nabla R|^2. \tag{3-2}$$

By Kato's inequality $\big|\nabla |E|\big|^2 \leq |\nabla E|^2$ and the fact that $|E|^2 \leq \frac{1}{24} R^2$,

$$|E|^2 |\nabla E|^2 \geq \tfrac{1}{576} R^2 |\nabla R|^2 \geq \tfrac{1}{24} |E|^2 |\nabla R|^2.$$

By dividing by $|E|^2(p) \neq 0$, the result follows; otherwise, if $|E|(p) = 0$, then $(\nabla R)(p) = 0$ from (3-2), and we conclude. $\square$

Lemma 3.3. *Let (M^3, g) be a Riemannian manifold. Then,*

$$E_{ip} E_{jp} E_{ij} \geq -\frac{1}{\sqrt{6}} |E|^3.$$

Proof. For a proof of this lemma, for instance, see [op. cit., Lemma 4.2]. $\square$

Corollary 3.4. *Let (M^3, g) be a complete critical metric for $\mathcal{F}_2$ with nonnegative scalar curvature. Then, $Rg_{ij} \geq 6E_{ij}$, g has nonnegative sectional curvature, and the scalar curvature satisfies the differential inequality*

$$\big(Rg_{ij} - 6E_{ij}\big) \nabla^2_{ij} R \geq \tfrac{1}{12} R^3.$$

Proof. From (3-1), combining Lemmas 3.1, 3.2, and 3.3, we obtain

$$\frac{1}{24}\big(Rg_{ij} - 6E_{ij}\big)\nabla^2_{ij} R \geq \frac{5}{12} R|E|^2 - \frac{4}{\sqrt{6}} |E|^3 = |E|^2 \Big(\frac{5}{12} R - \frac{4}{\sqrt{6}} |E|\Big) = \frac{1}{288} R^3,$$

where in the last equality we have used the fact that $|E|^2 = \frac{1}{24} R^2$. $\square$

Now we can prove Theorem 1.1. Clearly, if M^3 is compact, from Corollary 3.4, at a maximum point of R we obtain $R \leq 0$. Hence, $R \equiv 0$ on M^3, and from (2-2), $\mathrm{Ric} \equiv 0$ and the metric is flat. So, from now on, we will assume the manifold M^3 to be noncompact.

Choose now $\phi = \phi(r)$ to be a function of the distance r to a fixed point $O \in M^3$ and let $B_s(O)$ be a geodesic ball of radius $s > 0$. We denote by C_O the cut locus at the point O and we choose ϕ satisfying the following properties: $\phi = 1$ on $B_s(O)$, $\phi = 0$ on $M^3 \setminus B_{2s}(O)$,

$$-\frac{c}{s}\phi^{3/4} \le \phi' \le 0 \quad \text{and} \quad |\phi''| \le \frac{c}{s^2}\phi^{1/2}$$

on $B_{2s}(O) \setminus B_s(O)$ for some positive constant $c > 0$. In particular, ϕ is C^3 in $M^3 \setminus C_O$. Let $u := R\phi$ and $a_{ij} := (Rg_{ij} - 6E_{ij})$. From Corollary 3.4, we know that $a_{ij} \ge 0$ and we obtain

$$\begin{aligned} a_{ij}\nabla^2_{ij}u &= a_{ij}\big(\phi\nabla^2_{ij}R + R\nabla^2_{ij}\phi + 2\nabla_i R\nabla_j\phi\big) \\ &\ge \tfrac{1}{12}R^3\phi + R\phi' a_{ij}\nabla^2_{ij}r + R\phi'' a(\nabla r, \nabla r) + 2a(\nabla R, \nabla\phi). \end{aligned} \tag{3-3}$$

Now, let p_0 be a maximum point of u and assume that $p_0 \notin C_O$. If $\phi(p_0) = 0$, then $u \equiv 0$ and then $R \equiv 0$ on $B_{2s}(O)$. Hence, from now on we will assume $\phi(p_0) > 0$. Then, at p_0, we have $\nabla u(p_0) = 0$ and $\nabla^2_{ij}u(p_0) \le 0$. In particular, at p_0,

$$\nabla R(p_0) = -\frac{R(p_0)}{\phi(p_0)}\nabla\phi(p_0).$$

Moreover, since $a_{ij} \ge 0$, for every vector field X, $a(X, X) \le \mathrm{tr}(a)|X|^2 = 3R|X|^2$. On the other hand, from the standard Hessian comparison theorem, since g has nonnegative sectional curvature, we know that on $M^3 \setminus C_O$, one has $\nabla^2_{ij}r \le \frac{1}{r}g_{ij}$. Thus, from (3-3), at p_0, we get

$$\begin{aligned} 0 &\ge \frac{1}{12}R^3\phi + R\phi' a_{ij}\nabla^2_{ij}r + R\phi'' a(\nabla r, \nabla r) - 2\frac{R}{\phi}a(\nabla\phi, \nabla\phi) \\ &\ge \frac{1}{12}R^3\phi - \Big(\frac{|\phi'|}{r} + |\phi''| + 2\frac{(\phi')^2}{\phi}\Big)R\,\mathrm{tr}(a) \\ &\ge \frac{1}{12}R^3\phi - 3\Big(\frac{|\phi'|}{s} + |\phi''| + 2\frac{(\phi')^2}{\phi}\Big)R^2, \end{aligned}$$

where, in the last inequality, we have used the fact that $r \ge s$ on $B_{2s}(O) \setminus B_s(O)$, i.e., where $\phi' \ne 0$. From the assumptions on the cut-off function ϕ, we obtain, at the maximum point p_0,

$$0 \ge \frac{1}{12}R^2\phi^{1/2}\Big(R\phi^{1/2} - \frac{c'}{s^2}\Big)$$

for some positive constant $c' > 0$. Thus, we have proved that, if $p_0 \notin C_O$, then for every $p \in B_{2s}(O)$

$$u(p) \le u(p_0) = R(p_0)\phi(p_0) \le \frac{c'}{s^2}.$$

If $p_0 \in C_O$ we argue as follows (this trick is usually referred to Calabi). Let $\gamma : [0, L] \to M^3$, where $L = d(p_0, O)$, be a minimal geodesic joining O to p_0, the maximum point of u. Let $p_\varepsilon = \gamma(\varepsilon)$ for some $\varepsilon > 0$. Define now

$$u_\varepsilon(x) = R(x)\phi\big(d(x, p_\varepsilon) + \varepsilon\big).$$

Since $d(x, p_\varepsilon) + \varepsilon \geq d(x, O)$ and $d(p_0, p_\varepsilon) + \varepsilon = d(p_0, O)$, it is easy to see that $u_\varepsilon(p_0) = u(p_0)$ and

$$u_\varepsilon(x) \leq u(x) \quad \text{for all } x \in M^3,$$

since $\phi' \leq 0$. Hence p_0 is also a maximum point for u_ε. Moreover, $p_0 \notin C_{p_\varepsilon}$, so the function $d(x, p_\varepsilon)$ is smooth in a neighborhood of p_0 and we can apply the maximum principle argument as before to obtain an estimate for $u_\varepsilon(p_0)$ which depends on ε. Taking the limit as $\varepsilon \to 0$, we obtain the desired estimate on u.

By letting $s \to +\infty$ we obtain $u \equiv 0$, so $R \equiv 0$. From (2-2) we have $E \equiv 0$ and so $\mathrm{Ric} \equiv 0$ and Theorem 1.1 follows.

Acknowledgments

The authors are members of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM). Catino is supported by the GNAMPA project "Equazioni di evoluzione geometriche e strutture di tipo Einstein". Mastrolia and Monticelli are supported by the GNAMPA project "Analisi Globale ed Operatori Degeneri". Monticelli is partially supported by FSE, Regione Lombardia.

References

[Anderson 1997] M. T. Anderson, "Extrema of curvature functionals on the space of metrics on 3-manifolds", *Calc. Var. Partial Differential Equations* **5**:3 (1997), 199–269. MR 98a:58041 Zbl 0889.58027

[Anderson 2001] M. T. Anderson, "Extrema of curvature functionals on the space of metrics on 3-manifolds, II", *Calc. Var. Partial Differential Equations* **12**:1 (2001), 1–58. MR 2002a:58012 Zbl 1018.53020

[Berger 1970] M. Berger, "Quelques formules de variation pour une structure Riemannienne", *Ann. Sci. École Norm. Sup.* (4) **3** (1970), 285–294. MR 43 #3969 Zbl 0204.54802

[Besse 2008] A. L. Besse, *Einstein manifolds*, Springer, 2008. MR 2008k:53084 Zbl 1147.53001

[Catino 2014] G. Catino, "Critical metrics of the L^2-norm of the scalar curvature", *Proc. Amer. Math. Soc.* **142**:11 (2014), 3981–3986. MR 3251738 Zbl 1300.53044

[Catino 2015] G. Catino, "Some rigidity results on critical metrics for quadratic functionals", *Calc. Var. Partial Differential Equations* **54**:3 (2015), 2921–2937.

[Gursky and Viaclovsky 2001] M. J. Gursky and J. A. Viaclovsky, "A new variational characterization of three-dimensional space forms", *Invent. Math.* **145**:2 (2001), 251–278. MR 2002j:53039 Zbl 1006.58008

[Gursky and Viaclovsky 2013] M. J. Gursky and J. A. Viaclovsky, "Critical metrics on connected sums of Einstein four-manifolds", preprint, 2013. arXiv 1303.0827

[Gursky and Viaclovsky 2015] M. J. Gursky and J. A. Viaclovsky, "Rigidity and stability of Einstein metrics for quadratic curvature functionals", *J. Reine Angew. Math.* **700** (2015), 37–91. MR 3318510 Zbl 06412911

[Hamilton 1982] R. S. Hamilton, "Three-manifolds with positive Ricci curvature", *J. Differential Geom.* **17**:2 (1982), 255–306. MR 84a:53050 Zbl 0504.53034

[Hilbert 1915] D. Hilbert, "Die Grundlagen der Physik (erste Mitteilung)", *Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl.* **1915** (1915), 395–407. JFM 45.1111.01

[Lamontagne 1994] F. Lamontagne, "Une remarque sur la norme L^2 du tenseur de courbure", *C. R. Acad. Sci. Paris Sér. I Math.* **319**:3 (1994), 237–240. MR 95i:58048 Zbl 0806.53052

[Lamontagne 1998] F. Lamontagne, "A critical metric for the L^2-norm of the curvature tensor on S^3", *Proc. Amer. Math. Soc.* **126**:2 (1998), 589–593. MR 98d:58035 Zbl 0904.58010

[Smolentsev 2005] N. K. Smolentsev, "Spaces of Riemannian metrics", *Sovrem. Mat. Prilozh.* **31** (2005), 69–147. In Russian; translated in *J. Math. Sci.* (*New York*) **142**:5 (2007), 2436–2519. MR 2010a:58015 Zbl 1130.58006

Received March 15, 2015.

GIOVANNI CATINO
DIPARTIMENTO DI MATEMATICA
POLITECNICO DI MILANO
PIAZZA LEONARDO DA VINCI 32
I-20133 MILANO
ITALY
giovanni.catino@polimi.it

PAOLO MASTROLIA
DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DEGLI STUDI DI MILANO
VIA CESARE SALDINI 50
I-20133 MILANO
ITALY
paolo.mastrolia@unimi.it

DARIO D. MONTICELLI
DIPARTIMENTO DI MATEMATICA
POLITECNICO DI MILANO
PIAZZA LEONARDO DA VINCI 32
I-20133 MILANO
ITALY
dario.monticelli@polimi.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.293

ESTIMATES OF THE GAPS BETWEEN CONSECUTIVE EIGENVALUES OF LAPLACIAN

DAGUANG CHEN, TAO ZHENG AND HONGCANG YANG

For the eigenvalue problem of the Dirichlet Laplacian on a bounded domain in Euclidean space $\mathbb{R}^n$, we obtain estimates for the upper bounds of the gaps between consecutive eigenvalues which are the best possible in terms of the orders of the eigenvalues. Therefore, it is reasonable to conjecture that this type of estimate also holds for the eigenvalue problem on a Riemannian manifold. We give some particular examples.

1. Introduction

Let Ω be a bounded domain in an n-dimensional complete Riemannian manifold M with boundary (possible empty). Then the Dirichlet eigenvalue problem of the Laplacian on Ω is given by

$$\begin{cases} \Delta u = -\lambda u & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{1-1}$$

where Δ is the Laplacian on M. It is well known that the spectrum of (1-1) has the real and purely discrete eigenvalues

$$0 < \lambda_1 < \lambda_2 \le \lambda_3 \le \cdots \nearrow \infty, \tag{1-2}$$

where each λ_i has finite multiplicity and is repeated according to its multiplicity. The corresponding orthonormal basis of real eigenfunctions will be denoted $\{u_j\}_{j=1}^{\infty}$. We go forward under the assumption that $L^2(\Omega)$ represents the real Hilbert space of real-valued L^2 functions on Ω. We put $\lambda_0 = 0$ if $\partial\Omega = \varnothing$.

An important aspect of estimating higher eigenvalues is to estimate as precisely as possible the gaps between consecutive eigenvalues of (1-1). In this regard, we will review some important results on the estimates of eigenvalue problem (1-1).

The work of Chen was partially supported by NSFC grants No. 11271213 and No. 11471180. The work of Zheng was partially supported by the fundamental fund of Beijing Institute of Technology No. 20131742009 and No. 20141742002, postdoctoral fund No. 2014M550620 and No. 2015T80040, and NSFC grants No. 11401023 and No. 11471180. The work of Yang was partially supported by NSFC grant No. 11471180 and SF of CAS.

MSC2010: primary 35P15, 58C40; secondary 58J50.

Keywords: Laplacian, consecutive eigenvalues, test function, Riemannian manifold, hyperbolic space.

For the upper bound of the gap between consecutive eigenvalues of (1-1), when Ω is a bounded domain in a 2-dimensional Euclidean space $\mathbb{R}^2$, Payne, Pólya and Weinberger (see [Payne et al. 1955; 1956]) proved

$$\lambda_{k+1}-\lambda_k \le \frac{2}{k}\sum_{i=1}^{k}\lambda_i. \tag{1-3}$$

C. J. Thompson [1969] extended (1-3) to the n-dimensional case and obtained

$$\lambda_{k+1}-\lambda_k \le \frac{4}{nk}\sum_{i=1}^{k}\lambda_i. \tag{1-4}$$

Hile and Protter [1980] improved (1-4) to

$$\sum_{i=1}^{k}\frac{\lambda_i}{\lambda_{k+1}-\lambda_i} \ge \frac{nk}{4}. \tag{1-5}$$

Yang (see [Yang 1991] and more recently [Cheng and Yang 2007]) has obtained a sharp inequality:

$$\sum_{i=1}^{k}(\lambda_{k+1}-\lambda_i)\Big(\lambda_{k+1}-\Big(1+\frac{4}{n}\Big)\lambda_i\Big) \le 0. \tag{1-6}$$

From (1-6), one can infer

$$\lambda_{k+1} \le \frac{1}{k}\Big(1+\frac{4}{n}\Big)\sum_{i=1}^{k}\lambda_i. \tag{1-7}$$

The inequalities (1-6) and (1-7) are called Yang's first inequality and second inequality, respectively (see [Ashbaugh 1999; 2002; Ashbaugh and Benguria 1996; Harrell and Stubbe 1997]). Also we note that Ashbaugh and Benguria gave an optimal estimate for $k=1$ (see [Ashbaugh and Benguria 1991; 1992a; 1992b]). From Chebyshev's inequality, it is easy to prove that

$$\text{(1-6)} \Longrightarrow \text{(1-7)} \Longrightarrow \text{(1-5)} \Longrightarrow \text{(1-4)}.$$

From (1-6), Cheng and Yang [2005] obtained

$$\lambda_{k+1}-\lambda_k \le 2\Bigg(\Big(\frac{2}{n}\frac{1}{k}\sum_{i=1}^{k}\lambda_i\Big)^2-\Big(1+\frac{4}{n}\Big)\frac{1}{k}\sum_{i=1}^{k}\Big(\lambda_i-\frac{1}{k}\sum_{j=1}^{k}\lambda_j\Big)^2\Bigg)^{\frac{1}{2}}. \tag{1-8}$$

Cheng and Yang [2007], using their recursive formula, obtained

$$\lambda_{k+1} \le C_0(n)k^{2/n}\lambda_1, \tag{1-9}$$

where $C_0(n) \leq 1 + 4/n$ is a constant. From Weyl's asymptotic formula (see [Weyl 1912]), we know that the upper bound (1-9) is best possible in terms of the order on k.

For a complete Riemannian manifold M, from Nash's theorem [1956], there exists an isometric immersion

$$\psi : M \to \mathbb{R}^N,$$

where $\mathbb{R}^N$ is Euclidean space. The mean curvature of the immersion ψ is denoted by H and $|H|$ denotes its norm. Define

$$\Phi = \{\psi \mid \psi \text{ is an isometric immersion from } M \text{ into Euclidean space}\}.$$

When Ω is a bounded domain of a complete Riemannian manifold M, isometrically immersed into a Euclidean space $\mathbb{R}^N$, Cheng and the first author [Chen and Cheng 2008] (see also [El Soufi et al. 2009; Harrell 2007]) obtained

$$\sum_{i=1}^{k}(\lambda_{k+1} - \lambda_i)^2 \leq \frac{4}{n}\sum_{i=1}^{k}(\lambda_{k+1} - \lambda_i)\left(\lambda_i + \tfrac{1}{4}n^2 H_0^2\right), \tag{1-10}$$

where

$$H_0^2 = \inf_{\psi\in\Phi} \sup_{\Omega} |H|^2. \tag{1-11}$$

In the same paper, using the recursive formula in [Cheng and Yang 2007], Cheng and Chen also deduced

$$\lambda_{k+1} + \tfrac{1}{4}n^2 H_0^2 \leq C_0(n) k^{2/n}\left(\lambda_1 + \tfrac{1}{4}n^2 H_0^2\right), \tag{1-12}$$

where H_0^2 and $C_0(n)$ are given by (1-11) and (1-9), respectively.

From (1-10), we can get the gaps between the consecutive eigenvalues of the Laplacian:

$$\lambda_{k+1} - \lambda_k \leq 2\left(\left(\frac{2}{n}\frac{1}{k}\sum_{i=1}^{k}\lambda_i + \tfrac{1}{2}nH_0^2\right)^2 - \left(1 + \frac{4}{n}\right)\frac{1}{k}\sum_{i=1}^{k}\left(\lambda_i - \frac{1}{k}\sum_{j=1}^{k}\lambda_j\right)^2\right)^{\frac{1}{2}}. \tag{1-13}$$

Remark 1.1. When Ω is an n-dimensional compact homogeneous Riemannian manifold, a compact minimal submanifold without boundary and a connected bounded domain in the standard unit sphere $\mathbb{S}^N(1)$, and a connected bounded domain and a compact complex hypersurface without boundary of the complex projective space $\mathbb{CP}^n(4)$ with holomorphic sectional curvature 4, many mathematicians have studied the universal inequalities for eigenvalues and the difference of the consecutive eigenvalues (see [Cheng and Yang 2005; 2006; 2009; Harrell 1993;

Harrell and Michel 1994; Harrell and Stubbe 1997; Li 1980; Yang and Yau 1980; Leung 1991; Sun et al. 2008; Chen et al. 2012]).

Remark 1.2. Another problem is the lower bound of the gap between the first two eigenvalues. In general, there exists the famous fundamental gap conjecture for the Dirichlet eigenvalue problem of the Schrödinger operator (see [Ashbaugh and Benguria 1989; van den Berg 1983; Singer et al. 1985; Yau 1986; Yu and Zhong 1986] and the references therein). The fundamental gap conjecture was solved by B. Andrews and J. Clutterbuck [2011].

From (1-8) and (1-13), it is not difficult to see that both Yang's estimate for the gap between consecutive eigenvalues of (1-1) implicit in [Yang 1991] and the estimate from [Chen and Cheng 2008] are on the order of $k^{3/(2n)}$. However, by the calculation of the gap between the consecutive eigenvalues of $\mathbb{S}^n$ with the standard metric and Weyl's asymptotic formula, the order of the upper bound of this gap is $k^{1/n}$. Therefore, we make a conjecture:

Conjecture 1.3. *Let Ω be a bounded domain in an n-dimensional complete Riemannian manifold M. For the Dirichlet problem* (1-1), *the upper bound for the gap between consecutive eigenvalues of the Laplacian should be*

$$\lambda_{k+1} - \lambda_k \leq C_{n,\Omega} k^{1/n}, \quad k > 1, \tag{1-14}$$

where $C_{n,\Omega}$ is a constant dependent on Ω itself and the dimension n.

Remark 1.4. The famous Payne–Pólya–Weinberger conjecture (see [Payne et al. 1955; 1956; Thompson 1969; Ashbaugh and Benguria 1993a; 1993b]) states that, when $M = \mathbb{R}^n$, for the Dirichlet eigenvalue problem (1-1), one should have

$$\frac{\lambda_{k+1}}{\lambda_k} \leq \frac{\lambda_2}{\lambda_1}\bigg|_{\mathbb{B}^n} = \left(\frac{j_{n/2,1}}{j_{n/2-1,1}}\right)^2, \tag{1-15}$$

where $\mathbb{B}^n$ is the n-dimensional unit ball in $\mathbb{R}^n$, and $j_{p,k}$ is the k-th positive zero of the Bessel function $J_p(t)$. From Weyl's asymptotic formula and (1-15), the order of the upper bound of the consecutive eigenvalues of (1-1) is $k^{2/n}$. Therefore, Conjecture 1.3 reflects the distribution of eigenvalues from another point of view. From the order of the upper bound of the gap between the consecutive eigenvalues of $\mathbb{S}^n$, the estimate in (1-14) is best possible in terms of the order on k.

In the following, the constants $C_{n,\Omega}$ are allowed to be different in different cases.

When Ω is a bounded domain in $\mathbb{R}^n$, for the Dirichlet eigenvalue problem (1-1), we give an affirmative answer to Conjecture 1.3.

Theorem 1.5. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain in Euclidean space $\mathbb{R}^n$ and λ_k be the k-th* ($k > 1$) *eigenvalue of the Dirichlet eigenvalue problem* (1-1). *Then we have*

$$\lambda_{k+1} - \lambda_k \leq C_{n,\Omega} k^{1/n}, \tag{1-16}$$

where $C_{n,\Omega} = 4\lambda_1\sqrt{C_0(n)/n}$ *and* $C_0(n)$ *is given by* (1-9).

It is reasonable to conjecture that this type of estimate also holds on a Riemannian manifold. We give some particular examples as follows.

Corollary 1.6. *Let* $\Omega \subset \mathbb{H}^n(-1)$ *be a bounded domain in hyperbolic space* $\mathbb{H}^n(-1)$, *and* λ_k *be the* k*-th* ($k > 1$) *eigenvalue of the Dirichlet eigenvalue problem* (1-1). *Then we have*

$$\lambda_{k+1} - \lambda_k \le C_{n,\Omega} k^{1/n}, \tag{1-17}$$

where $C_{n,\Omega}$ *depends on* Ω *and the dimension* n *and is given by*

$$C_{n,\Omega} = 4\big(C_0(n)\big(\lambda_1 - \tfrac{1}{4}(n-1)^2\big)\big(\lambda_1 + \tfrac{1}{4}n^2H_0^2\big)\big)^{1/2}, \tag{1-18}$$

where $C_0(n)$ *and* H_0^2 *are the same as in* (1-12).

In fact, by the comparison theorem for the distance function in a Riemannian manifold, we have:

Corollary 1.7. *Let* M *be an* n*-dimensional* ($n \ge 3$) *simply connected complete noncompact Riemannian manifold with sectional curvature* Sec *satisfying*

$$-a^2 \le \mathrm{Sec} \le -b^2,$$

where a *and* b *are constants with* $0 \le b \le a$. *Let* $\Omega \subset M$ *be a bounded domain of* M *and* λ_k *be the* k*-th* ($k > 1$) *eigenvalue of* (1-1). *Then we have*

$$\lambda_{k+1} - \lambda_k \le C_{n,\Omega} k^{1/n}, \tag{1-19}$$

where $C_{n,\Omega}$ *depends on* Ω *and the dimension* n *and is given by*

$$C_{n,\Omega} = 4\big(C_0(n)\big(\lambda_1 - \tfrac{1}{4}(n-1)^2b^2 + \tfrac{1}{4}(a^2-b^2)\big)\big(\lambda_1 + \tfrac{1}{4}n^2H_0^2\big)\big)^{1/2}, \tag{1-20}$$

where $C_0(n)$ *and* H_0^2 *are the same as in* (1-12).

2. Preliminaries

In this section, we first recall some basic concepts and a theorem of Chapter 10 in [Kolmogorov and Fomin 1960], and then we prove a theorem which will be used in the next section.

Define

$$\mathcal{H}^\infty = \left\{ x = (x_j)_{j=1}^\infty \;\middle|\; x_j \in \mathbb{R},\ \left(\sum_{j=1}^\infty x_j^2\right)^{\frac{1}{2}} < +\infty \right\}$$

and

$$\mathcal{H}^2 = \{x = (x_1, x_2) \mid x_1, x_2 \in \mathbb{R},\ (x_1^2 + x_2^2)^{1/2} < +\infty\}.$$

The inner product $\langle\,\cdot\,,\,\cdot\,\rangle_\infty$ on $\mathcal{H}^\infty$ is defined by

$$\langle x, y\rangle_\infty = \sum_{j=1}^{\infty} x_j y_j, \quad \forall\, x = (x_j)_{j=1}^{\infty},\ y = (y_j)_{j=1}^{\infty}.$$

The inner product $\langle\,\cdot\,,\,\cdot\,\rangle_2$ on $\mathcal{H}^2$ can be defined similarly. Obviously, both $\mathcal{H}^\infty$ and $\mathcal{H}^2$ are Hilbert spaces. The dual space of $\mathcal{H}^2$ is denoted by $(\mathcal{H}^2)^*$. It is well known that $(\mathcal{H}^2)^*$ is isomorphic to $\mathcal{H}^2$ itself.

In order to prove our theorem, we need the following Lagrange multiplier theorem for real Banach spaces (see Chapter 10, Section 3, paragraph 3 in [Kolmogorov and Fomin 1960] or page 270 in [Zeidler 1995]).

Theorem 2.1. *Let X and Y be real Banach spaces. Assume that $F : x_0 \in U \subset X \to \mathbb{R}$ and $\Phi : x_0 \in U \subset X \to Y$ are continuously Fréchet differentiable on an open neighborhood of x_0, where $x_0 \in \Phi^{-1}(0) = \{x \in U \mid \Phi(x) = 0 \in Y\}$. If the set $\{\Phi'(x_0)(x) \in Y \mid x \in X\}$ is closed and x_0 is an extremum* (*maximum or minimum*) *of F on $\Phi^{-1}(0)$, then there exists $\lambda_0 \in \mathbb{R}$ and a linear functional $y^* \in Y^*$, where*

$$\lambda_0^2 + \|y^*\|^2 \neq 0,$$

such that

$$\lambda_0 F'(x_0) + (\Phi'(x_0))^*(y^*) = 0. \tag{2-1}$$

Moreover, if $\{\Phi'(x_0)(x) \in Y \mid x \in X\} = Y$, then we can take $\lambda_0 = 1$.

Theorem 2.2. *Assume that $\{\mu_j\}_{j=1}^{\infty}$ is a nondecreasing sequence, i.e.,*

$$0 < \mu_1 \le \mu_2 \le \cdots \le \mu_k \le \cdots \nearrow \infty,$$

where each μ_i has finite multiplicity m_i and is repeated according to its multiplicity.

Define

$$\begin{aligned} B &= \sum_{j=1}^{\infty} x_j^2 > 0, \\ A &= \sum_{j=1}^{\infty} \mu_j^2 x_j^2, \quad x = (x_j)_{j=1}^{\infty} \in \mathcal{H}^\infty. \end{aligned} \tag{2-2}$$

If $x_{m_1} \neq 0$ and $\sum_{j=1}^{\infty} \mu_j x_j^2 < \sqrt{AB}$, under the conditions in (2-2), *we have*

$$\sum_{j=1}^{\infty} \mu_j x_j^2 \le \frac{A + \mu_{m_1}\mu_{m_1+1} B}{\mu_{m_1} + \mu_{m_1+1}}. \tag{2-3}$$

Proof. First, assume that $\{\mu_j\}_{j=1}^{\infty}$ is a strictly increasing sequence, i.e.,

$$0 < \mu_1 < \mu_2 < \cdots < \mu_k < \cdots \nearrow \infty.$$

Suppose

$$F(x) = \sum_{j=1}^{\infty} \mu_j x_j^2,$$

$$\Psi(x) = \left(\sum_{j=1}^{\infty} x_j^2 - B,\ \sum_{j=1}^{\infty} \mu_j^2 x_j^2 - A\right) \in \mathcal{H}^2, \quad x \in \mathcal{H}^{\infty}.$$

Let $x_0 = (a_j)_{j=1}^{\infty}$ be an extremum of $F(x)$ on $\Phi^{-1}(0)$. Since $\forall\, h = (h_j)_{j=1}^{\infty} \in \mathcal{H}^{\infty}$,

$$F'(x_0)h = 2\sum_{j=1}^{\infty} \mu_j x_j h_j,$$

$$\Psi'(x_0)h = \left(2\sum_{j=1}^{\infty} x_j h_j,\ 2\sum_{j=1}^{\infty} \mu_j^2 x_j h_j\right),$$

and

$$\Psi'(x_0)(\mathcal{H}^{\infty}) = \mathcal{H}^2,$$

there exists $y^* \in (\mathcal{H}^2)^*$ such that

$$F'(x_0)h + (\Psi'(x_0))^*(y^*)h = 0. \tag{2-4}$$

Since $\mathcal{H}^2 = (\mathcal{H}^2)^*$, we can use some unique vector $(\mu, \lambda) \in \mathcal{H}^2$ to rewrite (2-4) as

$$\sum_{j=1}^{\infty} \mu_j a_j h_j + \mu \sum_{j=1}^{\infty} a_j h_j + \lambda \sum_{j=1}^{\infty} \mu_j^2 a_j h_j = 0. \tag{2-5}$$

Choosing

$$h_j = \delta_{jk}, \quad j = 1, 2, \ldots,$$

from (2-5), we obtain a system of equations

$$\mu_k a_k + \mu a_k + \lambda \mu_k^2 a_k = 0, \quad k = 1, 2, \ldots. \tag{2-6}$$

Since $\{\mu_k\}$ is a strictly increasing sequence, and there are only two varieties μ and λ, there are only two cases for x_0.

Case 1. There is only one $a_k \neq 0$, whether $k = 1$ or not. In this case, the critical value of $F(x)$ is given by

$$F(x_0) = \sqrt{AB},$$

which contradicts the assumption of the theorem.

Case 2. There are only two nonzero components of x_0, say a_k and a_l (without loss of generality, set $k < l$). In this case, we have

$$A = \mu_k^2 a_k^2 + \mu_l^2 a_l^2,\qquad B = a_k^2 + a_l^2. \tag{2-7}$$

From (2-7), we have

$$F(x_0) = \frac{A + \mu_k \mu_l B}{\mu_k + \mu_l}.$$

Since

$$A = \mu_k^2 a_k^2 + \mu_l^2 a_l^2 > \mu_k^2 (a_k^2 + a_l^2) = \mu_k^2 B,$$

we have

$$\mu_k < \sqrt{A/B}. \tag{2-8}$$

Similarly, we can also deduce

$$\mu_l > \sqrt{A/B}. \tag{2-9}$$

Hence, we have

$$F(x_0) - \sqrt{AB} = \frac{B\big(\mu_k - \sqrt{A/B}\big)\big(\mu_l - \sqrt{A/B}\big)}{\mu_k + \mu_l} < 0. \tag{2-10}$$

Since $\{\mu_i\}$ is a strictly increasing sequence, for μ_k fixed, from (2-8) and (2-9), we know that the right side of (2-10) is strictly decreasing in μ_l, i.e.,

$$\frac{B\big(\mu_k - \sqrt{A/B}\big)\big(\mu_{k+1} - \sqrt{A/B}\big)}{\mu_k + \mu_{k+1}} > \frac{B\big(\mu_k - \sqrt{A/B}\big)\big(\mu_{k+2} - \sqrt{A/B}\big)}{\mu_k + \mu_{k+2}} > \cdots .$$

Hence, we know that

$$\frac{A + \mu_k \mu_{k+1} B}{\mu_k + \mu_{k+1}}, \quad k = 1, 2, \ldots,$$

are local maximal values of $F(x)$.

Since $x_{m_1} = x_1 \neq 0$, k must be equal to $m_1 = 1$ only. Finally, we have the global maximum of $F(x)$

$$\frac{A + \mu_1 \mu_2 B}{\mu_1 + \mu_2}.$$

Second, assume that $\{\mu_j\}_{j=1}^{\infty}$ is an increasing sequence, i.e.,

$$0 < \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k \leq \cdots \nearrow \infty,$$

where each μ_i has finite multiplicity m_i and is repeated according to its multiplicity.

Replacing (2-7) by

$$A = m_k \mu_k^2 a_k^2 + m_l \mu_l^2 a_l^2,$$
$$B = m_k a_k^2 + m_l a_l^2,$$

and following the above steps almost word for word, we deduce that the local maximal value of $F(x)$ is

$$\frac{A + \mu_{m_k}\mu_{m_k+1} B}{\mu_{m_k} + \mu_{m_k+1}}$$

and

$$\mu_{m_k} < \sqrt{A/B}, \quad \mu_{m_k+1} > \sqrt{A/B}.$$

Since $x_{m_1} \neq 0$, m_k must be equal to m_1 and the local maximal value of $F(x)$ is the global maximum. Since

$$\frac{A + \mu_{m_1}\mu_{m_1+1} B}{\mu_{m_1} + \mu_{m_1+1}} - \sqrt{AB} = \frac{B\big(\mu_{m_1} - \sqrt{A/B}\big)\big(\mu_{m_k+1} - \sqrt{A/B}\big)}{\mu_{m_1} + \mu_{m_1+1}} < 0,$$

we can obtain (2-3). □

3. Proofs of main results

In this section, we will give the proof of Theorem 1.5. In order to prove our main results, we need the following key lemma and related corollaries of Theorem 2.2.

Lemma 3.1. *For the Dirichlet eigenvalue problem* (1-1), *let u_k be the orthonormal eigenfunction corresponding to the k-th eigenvalue λ_k, i.e.,*

$$\begin{cases} \Delta u_k = -\lambda_k u_k & \text{in } \Omega, \\ u_k = 0 & \text{on } \partial\Omega, \\ \int_\Omega u_i u_j = \delta_{ij}. \end{cases}$$

Then for any complex-valued function $g \in C^3(\Omega) \cap C^2(\overline{\Omega})$ such that gu_i is not the $\mathbb{C}$-linear combination of

$$u_1, \ldots, u_{k+1},$$

and such that

$$a_{k+1} = \int_\Omega g u_i u_{k+1} \neq 0,$$

with $\lambda_i < \lambda_{k+1} < \lambda_{k+2},\ k, i \in \mathbb{Z}^+,\ i \geq 1$, we have

$$\begin{aligned} &\big((\lambda_{k+1} - \lambda_i) + (\lambda_{k+2} - \lambda_i)\big) \int_\Omega |\nabla g|^2 u_i^2 \\ &\qquad \leq \int_\Omega |2\nabla g \cdot \nabla u_i + u_i \Delta g|^2 + (\lambda_{k+1} - \lambda_i)(\lambda_{k+2} - \lambda_i) \int_\Omega |g u_i|^2. \end{aligned} \tag{3-1}$$

Proof. Define

$$a_{ij} = \int_\Omega g u_i u_j,$$
$$b_{ij} = \int_\Omega \big(\nabla u_i \cdot \nabla g + \tfrac{1}{2} u_i \Delta g\big) u_j,$$

where ∇ denotes the gradient operator. Obviously,

$$a_{ij} = a_{ji}. \tag{3-2}$$

Then, from Stokes' theorem, we get

$$\begin{aligned}\lambda_j a_{ij} &= \int_\Omega g u_i(-\Delta u_j)\\ &= -\int_\Omega (u_i \Delta g + g\Delta u_i + 2\nabla g \cdot \nabla u_i) u_j\\ &= \lambda_i \int_\Omega g u_i u_j - 2\int_\Omega \big(\nabla u_i \cdot \nabla g + \tfrac{1}{2} u_i \Delta g\big) u_j,\end{aligned}$$

i.e.,

$$2b_{ij} = (\lambda_i - \lambda_j) a_{ij}. \tag{3-3}$$

From Stokes' theorem, we have

$$\int_\Omega |\nabla g|^2 u_i^2 = -2\int_\Omega g u_i \big(\nabla \bar{g} \cdot \nabla u_i + \tfrac{1}{2} u_i \Delta \bar{g}\big). \tag{3-4}$$

Since $\{u_k\}_{k=1}^\infty$ consists of a complete orthonormal basis of $L^2(\Omega)$, by the definition of a_{ij} and b_{ij}, from (3-3), (3-4) and Parseval's identity, we obtain

$$\int_\Omega |g u_i|^2 = \sum_{j=1}^\infty |a_{ij}|^2, \tag{3-5}$$

$$\int_\Omega |\nabla g|^2 u_i^2 = 2\sum_{j=1}^\infty a_{ij}\overline{b_{ij}} = \sum_{j=1}^\infty (\lambda_j - \lambda_i)|a_{ij}|^2, \tag{3-6}$$

$$\int_\Omega |2\nabla \bar{g} \cdot \nabla u_i + u_i \Delta \bar{g}|^2 = 4\sum_{j=1}^\infty |b_{ij}|^2 = \sum_{j=1}^\infty (\lambda_j - \lambda_i)^2 |a_{ij}|^2. \tag{3-7}$$

From the Cauchy–Schwarz inequality, we have

$$\Big(\sum_{j=k+1}^\infty (\lambda_j - \lambda_i)|a_{ij}|^2\Big)^2 \le \sum_{j=k+1}^\infty (\lambda_j - \lambda_i)^2 |a_{ij}|^2 \sum_{j=k+1}^\infty |a_{ij}|^2. \tag{3-8}$$

From (3-5), (3-6), (3-7) and (3-8), we can deduce

$$\left(\int_\Omega |\nabla g|^2 u_i^2 - \sum_{j=1}^{k} (\lambda_j - \lambda_i)|a_{ij}|^2\right)^2 \leq \left(\int_\Omega |g u_i|^2 - \sum_{j=1}^{k} |a_{ij}|^2\right)\left(\int_\Omega |2\nabla g \cdot \nabla u_i + u_i \Delta g|^2 - \sum_{j=1}^{k} (\lambda_j - \lambda_i)^2 |a_{ij}|^2\right). \tag{3-9}$$

Define

$$\widetilde{B}(i) = \int_\Omega |g u_i|^2 - \sum_{j=1}^{k} |a_{ij}|^2 = \sum_{j=k+1}^{\infty} |a_{ij}|^2 > 0, \quad \text{since } \int_\Omega g u_i u_{k+1} \neq 0,$$

$$\widetilde{A}(i) = \int_\Omega |2\nabla g \cdot \nabla u_i + u_i \Delta g|^2 - \sum_{j=1}^{k} (\lambda_j - \lambda_i)^2 |a_{ij}|^2 = \sum_{j=k+1}^{\infty} (\lambda_j - \lambda_i)^2 |a_{ij}|^2 \geq 0,$$

$$\widetilde{C}(i) = \int_\Omega |\nabla g|^2 u_i^2 - \sum_{j=1}^{k} (\lambda_j - \lambda_i)|a_{ij}|^2 = \sum_{j=k+1}^{\infty} (\lambda_j - \lambda_i)|a_{ij}|^2.$$

Since $g u_i$ is not the $\mathbb{C}$-linear combination of

$$u_1, \ldots, u_{k+1},$$

there exists some $l > k+1$ such that

$$a_l = \int_\Omega g u_i u_l \neq 0.$$

Since

$$\lambda_i < \lambda_{k+1} < \lambda_{k+2} \leq \lambda_l,$$

the vector

$$(|a_{ij}|)_{j=k+1}^{\infty}$$

is not proportional to

$$\big((\lambda_j - \lambda_i)^2 |a_{ij}|\big)_{j=k+1}^{\infty}.$$

From the Cauchy–Schwarz inequality, we have

$$\widetilde{C}(i) < \sqrt{\widetilde{A}(i)\widetilde{B}(i)}. \tag{3-10}$$

Since $a_{k+1} \neq 0$, from (3-10) and Theorem 2.2, we have

$$\widetilde{C}(i) \leq \frac{\widetilde{A}(i) + (\lambda_{k+2} - \lambda_i)(\lambda_{k+1} - \lambda_i)\widetilde{B}(i)}{(\lambda_{k+2} - \lambda_i) + (\lambda_{k+1} - \lambda_i)}. \tag{3-11}$$

From (3-11) and the definition of $\widetilde{A}(i)$, $\widetilde{B}(i)$ and $\widetilde{C}(i)$, we obtain

$$
\begin{aligned}
(3\text{-}12)\quad & \big((\lambda_{k+2}-\lambda_i)+(\lambda_{k+1}-\lambda_i)\big)\int_\Omega |\nabla g|^2 u_i^2 \\
&\le \int_\Omega |2\nabla g\cdot\nabla u_i + u_i\Delta g|^2 + (\lambda_{k+1}-\lambda_i)(\lambda_{k+2}-\lambda_i)\int_\Omega |gu_i|^2 \\
&\qquad\qquad - \sum_{j=1}^{k}(\lambda_{k+1}-\lambda_j)(\lambda_{k+2}-\lambda_j)|a_{ij}|^2 \\
&\le \int_\Omega |2\nabla g\cdot\nabla u_i + u_i\Delta g|^2 + (\lambda_{k+1}-\lambda_i)(\lambda_{k+2}-\lambda_i)\int_\Omega |gu_i|^2. \qquad \square
\end{aligned}
$$

Corollary 3.2. *Under the assumption of Lemma 3.1, for any nonconstant real-valued function $f \in C^3(\Omega)\cap C^2(\overline{\Omega})$, we have*

$$
\begin{aligned}
(3\text{-}13)\quad & \big((\lambda_{k+2}-\lambda_i)+(\lambda_{k+1}-\lambda_i)\big)\int_\Omega |\nabla f|^2 u_i^2 \\
&\le 2\sqrt{\big((\lambda_{k+2}-\lambda_i)(\lambda_{k+1}-\lambda_i)\big)\int_\Omega |\nabla f|^4 u^2} + \int_\Omega (2\nabla f\cdot\nabla u_i + u_i\Delta f)^2.
\end{aligned}
$$

Proof. Taking $g=\exp\big(\sqrt{-1}\alpha f\big)$, $\alpha\in\mathbb{R}\setminus\{0\}$, in (3-1), we have

$$
\begin{aligned}
(3\text{-}14)\quad & \alpha^2\big((\lambda_{k+1}-\lambda_i)+(\lambda_{k+2}-\lambda_i)\big)\int_\Omega |\nabla f|^2 u_i^2 \\
&\le \alpha^4\int_\Omega |\nabla f|^4 u_i^2 + \alpha^2\int_\Omega |2\nabla f\cdot\nabla u_i + u_i\Delta f|^2 + (\lambda_{k+1}-\lambda_i)(\lambda_{k+2}-\lambda_i).
\end{aligned}
$$

From (3-14), we have

$$
\begin{aligned}
(3\text{-}15)\quad & \big((\lambda_{k+1}-\lambda_i)+(\lambda_{k+2}-\lambda_i)\big)\int_\Omega |\nabla f|^2 u_i^2 \\
&\le \alpha^2\int_\Omega |\nabla f|^4 u_i^2 + \frac{1}{\alpha^2}(\lambda_{k+1}-\lambda_i)(\lambda_{k+2}-\lambda_i) + \int_\Omega |2\nabla f\cdot\nabla u_i + u_i\Delta f|^2.
\end{aligned}
$$

Since the inequality (3-15) is valid for any $\alpha\neq 0$ and

$$
(\lambda_{k+1}-\lambda_i)(\lambda_{k+2}-\lambda_i)\neq 0, \qquad \int_\Omega |\nabla f|^4 u_i^2 \neq 0,
$$

we can choose

$$
\alpha^2 = \left(\frac{(\lambda_{k+1}-\lambda_i)(\lambda_{k+2}-\lambda_i)}{\int_\Omega |\nabla f|^4 u_i^2}\right)^{\frac{1}{2}}
$$

to have (3-13). $\square$

Corollary 3.3. *Under the assumption of Lemma 3.1, for any real-valued function $f \in C^3(\Omega) \cap C^2(\overline{\Omega})$ with $|\nabla f|^2 = 1$, we have*

$$(\lambda_{k+2} - \lambda_{k+1})^2 \le 16\biggl(\int_\Omega (\nabla f \cdot \nabla u_i)^2 - \frac{1}{4}\int_\Omega (\Delta f)^2 u_i^2 - \frac{1}{2}\int_\Omega (\nabla(\Delta f) \cdot \nabla f) u_i^2\biggr)\lambda_{k+2}. \tag{3-16}$$

Furthermore, we have

$$\lambda_{k+2} - \lambda_{k+1} \le 4\biggl(\lambda_i - \frac{1}{4}\int_\Omega (\Delta f)^2 u_i^2 - \frac{1}{2}\int_\Omega (\nabla(\Delta f) \cdot \nabla f) u_i^2\biggr)^{\frac{1}{2}} \sqrt{\lambda_{k+2}}. \tag{3-17}$$

Proof. From Corollary 3.2 and $|\nabla f|^2 = 1$, we have

$$\bigl((\lambda_{k+2} - \lambda_i) + (\lambda_{k+1} - \lambda_i)\bigr) - 2\sqrt{(\lambda_{k+2} - \lambda_i)(\lambda_{k+1} - \lambda_i)} \le \int_\Omega (2\nabla f \cdot \nabla u_i + u_i \Delta f)^2,$$

that is,

$$\bigl(\sqrt{\lambda_{k+2} - \lambda_i} - \sqrt{\lambda_{k+1} - \lambda_i}\bigr)^2 \le \int_\Omega (2\nabla f \cdot \nabla u_i + u_i \Delta f)^2.$$

By integration by parts, we have

$$\int_\Omega (2\nabla f \cdot \nabla u_i + u_i \Delta f)^2 = 4\int_\Omega (\nabla f \cdot \nabla u_i)^2 - \int_\Omega (\Delta f)^2 u_i^2 - 2\int_\Omega (\nabla(\Delta f) \cdot \nabla f) u_i^2.$$

Hence, we have

$$\bigl(\sqrt{\lambda_{k+2} - \lambda_i} - \sqrt{\lambda_{k+1} - \lambda_i}\bigr)^2 \le 4\int_\Omega (\nabla f \cdot \nabla u_i)^2 - \int_\Omega (\Delta f)^2 u_i^2 - 2\int_\Omega (\nabla(\Delta f) \cdot \nabla f) u_i^2. \tag{3-18}$$

Multiplying (3-18) by $\bigl(\sqrt{\lambda_{k+2} - \lambda_i} + \sqrt{\lambda_{k+1} - \lambda_i}\bigr)^2$ on both sides, we can get

$$\begin{aligned}(\lambda_{k+2} - \lambda_{k+1})^2 &\le 4\biggl(\int_\Omega (\nabla f \cdot \nabla u_i)^2 - \frac{1}{4}\int_\Omega (\Delta f)^2 u_i^2 - \frac{1}{2}\int_\Omega (\nabla(\Delta f) \cdot \nabla f) u_i^2\biggr) \\ &\qquad \times \bigl(\sqrt{\lambda_{k+2} - \lambda_i} + \sqrt{\lambda_{k+1} - \lambda_i}\bigr)^2 \\ &\le 16\biggl(\int_\Omega (\nabla f \cdot \nabla u_i)^2 - \frac{1}{4}\int_\Omega (\Delta f)^2 u_i^2 - \frac{1}{2}\int_\Omega (\nabla(\Delta f) \cdot \nabla f) u_i^2\biggr)\lambda_{k+2},\end{aligned}$$

which is the inequality (3-16).

Since $|\nabla f|^2 = 1$, from (3-16), the Cauchy–Schwarz inequality and integration by parts, we obtain

$$\begin{aligned}(\lambda_{k+2}-\lambda_{k+1})^2 &\le 16\biggl(\int_\Omega |\nabla u_i|^2 - \frac{1}{4}\int_\Omega (\Delta f)^2 u_i^2 - \frac{1}{2}\int_\Omega (\nabla(\Delta f)\cdot\nabla f)u_i^2\biggr)\lambda_{k+2}\\ &= 16\biggl(\lambda_i - \frac{1}{4}\int_\Omega (\Delta f)^2 u_i^2 - \frac{1}{2}\int_\Omega (\nabla(\Delta f)\cdot\nabla f)u_i^2\biggr)\lambda_{k+2}. \qquad \square\end{aligned}$$

Remark 3.4. If $\lambda_{k+1} = \lambda_{k+2}$, (3-17) also holds trivially. Hence, under the conditions in Corollary 3.3, when $i = 1$, (3-17) holds for any $k > 1$.

Proof of Theorem 1.5. Since the inequality (3-20) always holds for $\lambda_{k+1} = \lambda_{k+2}$, without loss of generality, we assume that $\lambda_{k+1} < \lambda_{k+2}$ in the following discussion.

Let $\{x_1, x_2, \dots, x_n\}$ be the standard coordinate functions in $\mathbb{R}^n$. Taking

$$i = 1 \qquad \text{and} \qquad f = x_l, \quad l = 1, \dots, n,$$

in (3-16) and summing over l from 1 to n, we have

$$n(\lambda_{k+2}-\lambda_{k+1})^2 \le 16\lambda_{k+2}\int_\Omega \sum_{l=1}^n \left(\frac{\partial u_1}{\partial x_l}\right)^2 = 16\lambda_1\lambda_{k+2}, \tag{3-19}$$

where we use $|\nabla x_l| = 1,\ l = 1, \dots, n$.

From Theorem 3.1 in [Cheng and Yang 2007] (see also (1-9)) and from (3-19), we deduce

$$\lambda_{k+2}-\lambda_{k+1} \le 4\sqrt{\frac{\lambda_1}{n}}\sqrt{\lambda_{k+2}} \le 4\lambda_1\sqrt{\frac{C_0(n)}{n}}(k+1)^{1/2} = C_{n,\Omega}(k+1)^{1/2}, \tag{3-20}$$

where $C_{n,\Omega} = 4\lambda_1\sqrt{C_0(n)/n}$ and $C_0(n)$ is given by (1-9).

Therefore, (3-20) holds for arbitrary $k > 1$. $\square$

Proof of Corollary 1.6. For convenience, we will use the upper-half-plane model of hyperbolic space, i.e.,

$$\mathbb{H}^n(-1) = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_n > 0\}$$

with the standard metric

$$\mathrm{d}s^2 = \frac{(\mathrm{d}x_1)^2 + \cdots + (\mathrm{d}x_n)^2}{(x_n)^2}.$$

Taking $r = \log x_n$, we have

$$\mathrm{d}s^2 = (\mathrm{d}r)^2 + \mathrm{e}^{-2r}\sum_{i=1}^{n-1}(\mathrm{d}x_i)^2.$$

Without loss of generality, we assume that $\lambda_{k+1} < \lambda_{k+2}$. Taking $f = r$ and $i = 1$ in (3-17), we have

$$\begin{aligned} \lambda_{k+2} - \lambda_{k+1} &\le 4\biggl(\lambda_1 - \frac{1}{4}\int_\Omega (\Delta r)^2 u_i^2 - \frac{1}{2}\int_\Omega (\nabla(\Delta r)\cdot\nabla r) u_1^2\biggr)^{\frac{1}{2}} \sqrt{\lambda_{k+2}} \\ &= 4\bigl(\lambda_1 - \tfrac{1}{4}(n-1)^2\bigr)^{1/2}\sqrt{\lambda_{k+2}}, \end{aligned} \tag{3-21}$$

where $|\nabla r| = 1$ and $\Delta r = -(n-1)$ are used.

By (1-12) and (3-21), we have

$$\begin{aligned} \lambda_{k+2} - \lambda_{k+1} &\le 4\bigl(\lambda_1 - \tfrac{1}{4}(n-1)^2\bigr)^{1/2}\sqrt{C_0(n)\bigl(\lambda_1 + \tfrac{1}{4}n^2 H_0^2\bigr)}(k+1)^{1/n} \\ &= C_{n,\Omega}(k+1)^{1/n}, \end{aligned} \tag{3-22}$$

where $C_{n,\Omega}$ is defined by (1-18). Therefore, we can deduce (3-22) for any $k > 1$. □

4. Proof of Corollary 1.7

Assume that (M, g) is an n-dimensional complete noncompact Riemannian manifold with sectional curvature Sec satisfying $-a^2 \le \mathrm{Sec} \le -b^2$, where a and b are constants with $0 \le b \le a$. Let Ω be a bounded domain of M. For a fixed point $p \notin \overline{\Omega}$, the distance function $\rho(x)$ is defined by $\rho(x) = \mathrm{distance}(x, p)$. From $|\nabla\rho| = 1$ and Proposition 2.2 of [Schoen and Yau 1994], we have

$$\nabla\rho\cdot\nabla(\Delta\rho) = -|\mathrm{Hess}\,\rho|^2 - \mathrm{Ric}(\nabla\rho, \nabla\rho). \tag{4-1}$$

Assume that $h_1, \dots, h_{n-1}$, with $0 \le h_1 \le \cdots \le h_{n-1}$, are the eigenvalues of $\mathrm{Hess}\,\rho$. We have

$$\begin{aligned} 2|\mathrm{Hess}\,\rho|^2 - (\Delta\rho)^2 &= 2\sum_{i=1}^{n-1} h_i^2 - \biggl(\sum_{i=1}^{n-1} h_i\biggr)^2 \\ &= \sum_{i=1}^{n-1} h_i^2 - \sum_{i\ne j} h_i h_j \\ &\le h_{n-1}^2 + h_1h_2 + \cdots + h_{n-2}h_{n-1} - \sum_{i\ne j} h_i h_j \\ &= h_{n-1}^2 - h_1h_2 - \cdots - h_{n-2}h_{n-1} - \sum_{\substack{i\ne j\\ i,j\le n-2}} h_i h_j \\ &\le h_{n-1}^2 - (n-2)^2 h_1^2. \end{aligned} \tag{4-2}$$

From the Hessian comparison theorem (see [Wu et al. 1989]), we have

$$a\frac{\cosh a\rho}{\sinh a\rho} \ge h_{n-1} \ge \cdots \ge h_1 \ge b\frac{\cosh b\rho}{\sinh b\rho}. \tag{4-3}$$

Since $n \geq 3$ and $a^2/(\sinh^2 a\rho)$ is a decreasing function of a, from (4-2) and (4-3), we have

$$
\begin{aligned}
(4\text{-}4)\quad & 2|\mathrm{Hess}\,\rho|^2 + 2\,\mathrm{Ric}(\nabla\rho, \nabla\rho) - (\Delta\rho)^2 \\
&\leq a^2 \frac{\cosh^2 a\rho}{\sinh^2 a\rho} - (n-2)^2 b^2 \frac{\cosh^2 b\rho}{\sinh^2 b\rho} - 2(n-1)b^2 \\
&= a^2 + \frac{a^2}{\sinh^2 a\rho} - (n-2)^2 b^2 - (n-2)^2 \frac{b^2}{\sinh^2 b\rho} - 2(n-1)b^2 \\
&\leq -(n-1)^2 b^2 + (a^2 - b^2) + \frac{b^2}{\sinh^2 b\rho} - (n-2)^2 \frac{b^2}{\sinh^2 b\rho} \\
&\leq -(n-1)^2 b^2 + (a^2 - b^2).
\end{aligned}
$$

Without loss of generality, we assume $\lambda_{k+1} < \lambda_{k+2}$. By taking $f = \rho$ and $i = 1$ in (3-17), we have

$$
(4\text{-}5)\qquad \lambda_{k+2} - \lambda_{k+1} \leq 4\left(\lambda_1 - \frac{1}{4}\int_\Omega (\Delta\rho)^2 u_1^2 - \frac{1}{2}\int_\Omega (\nabla(\Delta\rho)\cdot\nabla\rho)u_1^2\right)^{\frac{1}{2}} \sqrt{\lambda_{k+2}}.
$$

From (4-1) and (4-4), we obtain

$$
\begin{aligned}
(4\text{-}6)\quad & \lambda_1 - \frac{1}{4}\int_\Omega (\Delta\rho)^2 u_1^2 - \frac{1}{2}\int_\Omega (\nabla(\Delta\rho)\cdot\nabla\rho)u_1^2 \\
&= \lambda_1 + \frac{1}{4}\int_\Omega \big(2|\mathrm{Hess}\,\rho|^2 + 2\,\mathrm{Ric}(\nabla\rho, \nabla\rho) - (\Delta\rho)^2\big)u_1^2 \\
&\leq \lambda_1 - \tfrac{1}{4}(n-1)^2 b^2 + \tfrac{1}{4}(a^2 - b^2).
\end{aligned}
$$

By (1-12), (4-5) and (4-6), we have

$$
\begin{aligned}
(4\text{-}7)\quad & \lambda_{k+2} - \lambda_{k+1} \\
&\leq 4\big(\lambda_1 - \tfrac{1}{4}(n-1)^2 b^2 + \tfrac{1}{4}(a^2 - b^2)\big)^{1/2} \sqrt{C_0(n)\big(\lambda_1 + \tfrac{1}{4}n^2 H_0^2\big)}(k+1)^{1/n} \\
&\leq C_{n,\Omega}(k+1)^{1/n},
\end{aligned}
$$

where $C_{n,\Omega}$ is defined by (1-20). Therefore, we can deduce (4-7) for any $k > 1$. □

References

[Andrews and Clutterbuck 2011] B. Andrews and J. Clutterbuck, "Proof of the fundamental gap conjecture", *J. Amer. Math. Soc.* **24**:3 (2011), 899–916. MR 2012d:35051 Zbl 1222.35130

[Ashbaugh 1999] M. S. Ashbaugh, "Isoperimetric and universal inequalities for eigenvalues", pp. 95–139 in *Spectral theory and geometry* (Edinburgh, 1998), edited by B. Davies and Y. Safarov, London Mathematical Society Lecture Note Series **273**, Cambridge University Press, 1999. MR 2001a:35131 Zbl 0937.35114

[Ashbaugh 2002] M. S. Ashbaugh, “The universal eigenvalue bounds of Payne–Pólya–Weinberger, Hile–Protter, and H. C. Yang”, *Proc. Indian Acad. Sci. Math. Sci.* **112**:1 (2002), 3–30. MR 2004c:35302 Zbl 1199.35261

[Ashbaugh and Benguria 1989] M. S. Ashbaugh and R. D. Benguria, “Optimal lower bound for the gap between the first two eigenvalues of one-dimensional Schrödinger operators with symmetric single-well potentials”, *Proc. Amer. Math. Soc.* **105**:2 (1989), 419–424. MR 89f:81028 Zbl 0668.34023

[Ashbaugh and Benguria 1991] M. S. Ashbaugh and R. D. Benguria, “Proof of the Payne–Pólya–Weinberger conjecture”, *Bull. Amer. Math. Soc.* (*N.S.*) **25**:1 (1991), 19–29. MR 91m:35173 Zbl 0736.35075

[Ashbaugh and Benguria 1992a] M. S. Ashbaugh and R. D. Benguria, “A second proof of the Payne–Pólya–Weinberger conjecture”, *Comm. Math. Phys.* **147**:1 (1992), 181–190. MR 93k:33002 Zbl 0758.34075

[Ashbaugh and Benguria 1992b] M. S. Ashbaugh and R. D. Benguria, “A sharp bound for the ratio of the first two eigenvalues of Dirichlet Laplacians and extensions”, *Ann. of Math.* (2) **135**:3 (1992), 601–628. MR 93d:35105 Zbl 0757.35052

[Ashbaugh and Benguria 1993a] M. S. Ashbaugh and R. D. Benguria, “Isoperimetric bounds for higher eigenvalue ratios for the n-dimensional fixed membrane problem”, *Proc. Roy. Soc. Edinburgh Sect. A* **123**:6 (1993), 977–985. MR 95b:35162 Zbl 0819.35104

[Ashbaugh and Benguria 1993b] M. S. Ashbaugh and R. D. Benguria, “More bounds on eigenvalue ratios for Dirichlet Laplacians in n dimensions”, *SIAM J. Math. Anal.* **24**:6 (1993), 1622–1651. MR 94i:35139 Zbl 0809.35067

[Ashbaugh and Benguria 1996] M. S. Ashbaugh and R. D. Benguria, “Bounds for ratios of the first, second, and third membrane eigenvalues”, pp. 30–42 in *Nonlinear problems in applied mathematics*, edited by T. S. Angell et al., Society of Industrial and Applied Mathematics, Philadelphia, 1996. MR 2410595 Zbl 0882.35085

[van den Berg 1983] M. van den Berg, “On condensation in the free-boson gas and the spectrum of the Laplacian”, *J. Statist. Phys.* **31**:3 (1983), 623–637. MR 85i:82022

[Chen and Cheng 2008] D. Chen and Q.-M. Cheng, “Extrinsic estimates for eigenvalues of the Laplace operator”, *J. Math. Soc. Japan* **60**:2 (2008), 325–339. MR 2010b:35323 Zbl 1147.35060

[Chen et al. 2012] D. Chen, T. Zheng, and M. Lu, “Eigenvalue estimates on domains in complete noncompact Riemannian manifolds”, *Pacific J. Math.* **255**:1 (2012), 41–54. MR 2923693 Zbl 1236.35097

[Cheng and Yang 2005] Q.-M. Cheng and H. Yang, “Estimates on eigenvalues of Laplacian”, *Math. Ann.* **331**:2 (2005), 445–460. MR 2005i:58038 Zbl 1122.35086

[Cheng and Yang 2006] Q.-M. Cheng and H. Yang, “Inequalities for eigenvalues of Laplacian on domains and compact complex hypersurfaces in complex projective spaces”, *J. Math. Soc. Japan* **58**:2 (2006), 545–561. MR 2007k:58051 Zbl 1127.35026

[Cheng and Yang 2007] Q.-M. Cheng and H. Yang, “Bounds on eigenvalues of Dirichlet Laplacian”, *Math. Ann.* **337**:1 (2007), 159–175. MR 2007k:35064 Zbl 1110.35052

[Cheng and Yang 2009] Q.-M. Cheng and H. Yang, “Estimates for eigenvalues on Riemannian manifolds”, *J. Differential Equations* **247**:8 (2009), 2270–2281. MR 2010j:58066 Zbl 1180.35390

[El Soufi et al. 2009] A. El Soufi, E. M. Harrell, II, and S. Ilias, “Universal inequalities for the eigenvalues of Laplace and Schrödinger operators on submanifolds”, *Trans. Amer. Math. Soc.* **361**:5 (2009), 2337–2350. MR 2010e:58032 Zbl 1162.58009

[Harrell 1993] E. M. Harrell, II, “Some geometric bounds on eigenvalue gaps”, *Comm. Partial Differential Equations* **18**:1-2 (1993), 179–198. MR 94c:35135 Zbl 0810.35067

[Harrell 2007] E. M. Harrell, II, "Commutators, eigenvalue gaps, and mean curvature in the theory of Schrödinger operators", *Comm. Partial Differential Equations* **32**:1-3 (2007), 401–413. MR 2008i:35041 Zbl 05150097

[Harrell and Michel 1994] E. M. Harrell, II and P. L. Michel, "Commutator bounds for eigenvalues, with applications to spectral geometry", *Comm. Partial Differential Equations* **19**:11-12 (1994), 2037–2055. MR 95i:58182 Zbl 0815.35078

[Harrell and Stubbe 1997] E. M. Harrell, II and J. Stubbe, "On trace identities and universal eigenvalue estimates for some partial differential operators", *Trans. Amer. Math. Soc.* **349**:5 (1997), 1797–1809. MR 97i:35129 Zbl 0887.35111

[Hile and Protter 1980] G. N. Hile and M. H. Protter, "Inequalities for eigenvalues of the Laplacian", *Indiana Univ. Math. J.* **29**:4 (1980), 523–538. MR 82c:35052 Zbl 0454.35064

[Kolmogorov and Fomin 1960] A. N. Kolmogorov and S. V. Fomin, Елементы теории функций и функционального анализа, 2: Мера, интеграл Лебега, Гилбертого пространство, Izdat. Moskov. Univ., Moscow, 1960. Translated as *Elements of the theory of functions and functional analysis, 2: Measure. The Lebesgue integral. Hilbert space*, Graylock Press, Albany, NY, 1961; reprinted by Dover, Mineola, NY, 1999; Chinese translation, Higher Education Press, Beijing, 2006. MR 0118795 Zbl 0090.08702

[Leung 1991] P.-F. Leung, "On the consecutive eigenvalues of the Laplacian of a compact minimal submanifold in a sphere", *J. Austral. Math. Soc. Ser. A* **50**:3 (1991), 409–416. MR 92d:58212 Zbl 0728.53035

[Li 1980] P. Li, "Eigenvalue estimates on homogeneous manifolds", *Comment. Math. Helv.* **55**:3 (1980), 347–363. MR 81k:58067 Zbl 0451.53036

[Nash 1956] J. Nash, "The imbedding problem for Riemannian manifolds", *Ann. of Math.* (2) **63** (1956), 20–63. MR 17,782b Zbl 0070.38603

[Payne et al. 1955] L. E. Payne, G. Pólya, and H. F. Weinberger, "Sur le quotient de deux fréquences propres consécutives", *C. R. Acad. Sci. Paris* **241** (1955), 917–919. MR 17,372d Zbl 0065.08801

[Payne et al. 1956] L. E. Payne, G. Pólya, and H. F. Weinberger, "On the ratio of consecutive eigenvalues", *J. Math. and Phys.* **35** (1956), 289–298. MR 18,905c Zbl 0073.08203

[Schoen and Yau 1994] R. Schoen and S.-T. Yau, *Lectures on differential geometry*, Conference Proceedings and Lecture Notes in Geometry and Topology **1**, International Press, Cambridge, MA, 1994. MR 97d:53001 Zbl 0830.53001

[Singer et al. 1985] I. M. Singer, B. Wong, S.-T. Yau, and S. S.-T. Yau, "An estimate of the gap of the first two eigenvalues in the Schrödinger operator", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **12**:2 (1985), 319–333. MR 87j:35280 Zbl 0603.35070

[Sun et al. 2008] H. Sun, Q.-M. Cheng, and H. Yang, "Lower order eigenvalues of Dirichlet Laplacian", *Manuscripta Math.* **125**:2 (2008), 139–156. MR 2009i:58042 Zbl 1137.35050

[Thompson 1969] C. J. Thompson, "On the ratio of consecutive eigenvalues in N-dimensions", *Studies in Appl. Math.* **48** (1969), 281–283. MR 41 #2242 Zbl 0183.11005

[Weyl 1912] H. Weyl, "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)", *Math. Ann.* **71**:4 (1912), 441–479. MR 1511670 JFM 43.0436.01

[Wu et al. 1989] H. Wu, L. C. Shen, and Y. L. Yu, *Introduction to Riemannian geometry*, Peking University Press, Beijing, 1989. In Chinese.

[Yang 1991] H. Yang, "An estimate of the difference between consecutive eigenvalues", preprint, International Centre for Theoretical Physics, Trieste, 1991, available at https://inis.iaea.org/search/search.aspx?orig_q=RN:23015356. Report IC/91/60.

[Yang and Yau 1980] P. C. Yang and S.-T. Yau, "Eigenvalues of the Laplacian of compact Riemann surfaces and minimal submanifolds", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **7**:1 (1980), 55–63. MR 81m:58084 Zbl 0446.58017

[Yau 1986] S.-T. Yau, *Nonlinear analysis in geometry*, Monographies de L'Enseignement Mathématique **33**, L'Enseignement Mathématique, Geneva, 1986. MR 88e:53001 Zbl 0631.53003

[Yu and Zhong 1986] Q. H. Yu and J. Q. Zhong, "Lower bounds of the gap between the first and second eigenvalues of the Schrödinger operator", *Trans. Amer. Math. Soc.* **294**:1 (1986), 341–349. MR 87f:35179 Zbl 0593.53030

[Zeidler 1995] E. Zeidler, *Applied functional analysis: main principles and their applications*, Applied Mathematical Sciences **109**, Springer, New York, 1995. MR 96i:00006 Zbl 0834.46003

Received November 17, 2014.

DAGUANG CHEN
DEPARTMENT OF MATHEMATICAL SCIENCES
TSINGHUA UNIVERSITY
BEIJING 100084
CHINA
dgchen@math.tsinghua.edu.cn

TAO ZHENG
SCHOOL OF MATHEMATICS AND STATISTICS
BEIJING INSTITUTE OF TECHNOLOGY
BEIJING 100081
CHINA
zhengtao08@amss.ac.cn

HONGCANG YANG
HUA LOO-KENG KEY LABORATORY OF MATHEMATICS
CHINESE ACADEMY OF SCIENCES
BEIJING 100080
CHINA
yanghc2@netease.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.313

LIOUVILLE TYPE THEOREMS FOR THE p-HARMONIC FUNCTIONS ON CERTAIN MANIFOLDS

JINGYI CHEN AND YUE WANG

We show that for a certain range of $p > n$, the Dirichlet problem at infinity is unsolvable for the p-Laplace equation for any nonconstant continuous boundary data on an n-dimensional Cartan–Hadamard manifold constructed from a complete noncompact shrinking gradient Ricci soliton. Using the steady gradient Ricci soliton, we find an incomplete Riemannian metric on $\mathbb{R}^2$ with positive Gauss curvature such that every positive p-harmonic function must be constant for $p \geq 4$.

1. Introduction

In this article, we study two questions about the p-Laplace equation on Riemannian manifolds. The first one is the solvability of the Dirichlet problem at infinity on a negatively curved complete noncompact manifold, and the second one is the Liouville property for positive solutions on $\mathbb{R}^2$ equipped with an incomplete metric with positive Gauss curvature. In both cases, the n-dimensional manifold M under consideration is equipped with a Riemannian metric $e^{2f/(p-n)}g$ where (M, g, f) is a complete gradient Ricci soliton which is shrinking for the first case and steady for the second case.

On a Riemannian manifold, for a constant $p > 1$, a function v in $W^{1,p}_{\mathrm{loc}} \cap L^{\infty}_{\mathrm{loc}}$ is p-harmonic if it is a weak solution to the p-Laplacian equation

$$\operatorname{div}(|\nabla v|^{p-2}\nabla v) = 0. \tag{1-1}$$

It is known that p-harmonic functions are in $C^{1,\alpha}$ (see [Tolksdorf 1984] and the references therein).

Chen is partially supported by NSERC (RGPIN 203199-1). Wang is supported in part by the National Natural Science Foundation of China (No. 10901147) and Zhejiang Provincial Natural Science Foundation of China (No. LY12A01028), and she is grateful to PIMS/UBC for hosting her visit during which part of this work was carried out.
MSC2010: 53C21, 58J05.
Keywords: Liouville type theorems, p-harmonic functions, gradient Ricci solitons.

The behavior of harmonic and, more generally, p-harmonic functions depends on the sign of the curvature of the manifold in an essential way. Therefore, we must treat negatively curved and nonnegatively curved manifolds separately.

A Cartan–Hadamard manifold is a complete simply connected Riemannian manifold with nonpositive sectional curvature everywhere. It is well-known that a Cartan–Hadamard manifold M can be compactified by attaching a sphere $M(\infty)$ at infinity. In the cone topology, the compactification is homeomorphic to a closed Euclidean n-ball [Eberlein and O'Neill 1973]. The Dirichlet problem at infinity for p-harmonic functions is to solve the p-Laplace equation (1-1) on M such that v agrees with a given continuous function φ on $M(\infty)$. For $p = 2$, the Dirichlet problem at infinity for harmonic functions is solvable if there are suitable lower and upper bounds for the sectional curvature [Anderson 1983; Anderson and Schoen 1985; Choi 1984; Hsu 2003; Sullivan 1983]. Ancona [1994] constructed an example showing that the Dirichlet problem is unsolvable if only a negative constant upper bound is imposed. For $p \in (1, \infty)$, the Dirichlet problem at infinity is solvable under similar curvature assumptions like those in the case $p = 2$; in particular, it is solvable if the sectional curvature is bounded by

$$-r^{2\alpha-4-\epsilon} \le K \le -\frac{\alpha(\alpha-1)}{r^2} \tag{1-2}$$

near $M(\infty)$ where $\epsilon > 0$ and $\alpha > 1$, where r is the distance to a fixed point, and for $p \in (1, 1+(n-1)\alpha)$ [Holopainen 2002; Holopainen and Vähäkangas 2007; Pansu 1989].

Our first result is to show the unsolvability of the Dirichlet problem at infinity on certain Cartan–Hadamard manifolds constructed from shrinking gradient Ricci solitons, for a certain range of $p > n$. In particular, the unsolvability holds for the shrinking Gaussian soliton $(\mathbb{R}^n, dx^2, |x|^2/4)$ for every $p > n$. It is interesting to observe that the sectional curvature of the complete negatively curved metric $e^{|x|^2/(2(p-n))}dx^2$ is not bounded above by $-\alpha(\alpha-1)/r^2$, for any constant $\alpha > 1$, at certain sections for sufficiently large r (see remark on page 319). This indicates the upper bound in (1-2) is sharp in some sense for the solvability of the Dirichlet problem at infinity.

Theorem 1.1. *Suppose that (M, g, f) is a simply connected n-dimensional complete noncompact shrinking gradient Ricci soliton whose sectional curvatures are bounded above by a constant K_0 with $0 < K_0 < 1/(2(n-1))$. Then the Dirichlet problem at infinity for the p-Laplace equation on $(M, e^{2f/(p-n)}g)$ is unsolvable for any nonconstant continuous boundary value φ and $n < p < \frac{1}{K_0} + 2 - n$.*

The proof relies on a Liouville type property (Proposition 2.1) for positive solutions to the p-Laplace equation on $(M, e^{-2f/(n-p)}g)$ for every $p > 1$, where Cao and Zhou's [2010] estimates on f and on the volume growth for gradient shrinking Ricci solitons are crucial as they imply that e^{-f} is integrable on (M, g). The advantage for considering the range $p > n$ is that, under the conformal change of metric, it yields a complete metric $\tilde{g}$ and it guarantees the negativity of the curvature of $\tilde{g}$ under the curvature assumption $K \leq K_0$, while one does not have such flexibility for $p = 2$.

However, the integration argument in the proof of Proposition 2.1 is no longer valid for steady gradient Ricci solitons due to different behavior of f (typically f tends to $-\infty$ along a sequence of points x_k that go to infinity [Munteanu and Sesum 2013; Wu 2013]). Alternatively, a powerful way to prove Liouville type theorems for positive harmonic functions on complete manifolds with nonnegative Ricci curvature is via Yau's gradient estimate [1975]. The p-harmonic version of Yau's estimate is established by Wang and Zhang [2011] (see [Sung and Wang 2014] for a sharp form of the estimate). For a positive p-harmonic function u in the conformally changed metric $\tilde{g} = e^{-2f/(n-p)}g$, we first derive a maximum principle for $|\nabla \log u|$ for steady (or shrinking) gradient Ricci solitons, via a Bochner type formula. However, the required assumption on Ricci curvature for the gradient estimates cannot hold globally for steady gradient Ricci solitons if $\dim M > 2$ because it would imply that the scalar curvature of g possesses a positive constant lower bound. But this is impossible as shown in [Munteanu and Sesum 2013; Wu 2013]. In dimension 2, we can combine the maximum principle (Proposition 3.3) and the gradient estimate to prove a Liouville type result on the 2-plane with a positively curved *incomplete* metric.

Theorem 1.2. *Let* $(\mathbb{R}^2, g, f)$ *be Hamilton's cigar soliton. Then there does not exist any nonconstant positive* p*-harmonic function on* $(\mathbb{R}^2, \tilde{g})$ *for* $p \geq 4$.

Harmonic functions on the complete gradient Ricci solitons have been studied by Munteanu and Sesum [2013] and Munteanu and Wang [2012] with applications to the geometry and topology of the solitons. Moser [2007] observed an interesting connection between the inverse mean curvature flow formulated as level sets in $\mathbb{R}^n$ and 1-harmonic functions. Kotschwar and Ni [2009] generalize this to Riemannian ambient manifolds. There is also recent work on gradient estimates for weighted p-harmonic functions and the first p-eigenfunctions [Dung and Dat 2015].

2. The Dirichlet problem at infinity

In this section, the triple (M, g, f) is assumed to be a complete noncompact shrinking gradient Ricci soliton. We first establish the following Liouville property for positive p-harmonic functions for $p > 1$ with no additional curvature assumption.

An n-dimensional Riemannian manifold (M, g) is a gradient Ricci soliton if

$$\mathrm{Ric} + \nabla\nabla f + \varepsilon g = 0 \tag{2-1}$$

for some smooth function f and $\varepsilon = -\frac{1}{2}, 0, \frac{1}{2}$. Corresponding to the three values of ε, the gradient Ricci soliton (M, g, f) is shrinking, steady, or expanding [Chow et al. 2006; Hamilton 1995].

Proposition 2.1. *Let (M, g, f) be a complete noncompact gradient shrinking Ricci soliton. Then there is no nonconstant positive p-harmonic function on $(M, e^{-2f/(n-p)}g)$ for $p > 1$.*

Proof. Since u is a p-harmonic function on $(M, \tilde{g})$ where $\tilde{g} = e^{-2f/(n-p)}g$,

$$\operatorname{div}_{\tilde{g}}\big(|\widetilde{\nabla} w|_{\tilde{g}}^{p-2}\widetilde{\nabla} w\big) = |\widetilde{\nabla} w|_{\tilde{g}}^{p} \tag{2-2}$$

holds for $w = -(p-1)\log u$. For any smooth cut-off function $\phi \in C_0^\infty(M)$, in the complete metric g, we require

$$\begin{cases} \phi = 1 & \text{on } B_{x_0}(\rho, g), \\ \phi = 0 & \text{on } M \setminus B_{x_0}(2\rho, g), \\ 0 \le \phi \le 1 & \text{on } M, \\ |\nabla\phi|^2 \le C/\rho^2 & \text{on } M. \end{cases}$$

Here $B_{x_0}(r, g)$ stands for the geodesic ball centered at x_0 with radius r in the metric g in M. Multiplying (2-2) by ϕ^2, then integrating and applying Stokes' theorem, we have

$$\begin{aligned} \int_M |\widetilde{\nabla} w|_{\tilde{g}}^{p}\,\phi^2\,d\mu_{\tilde{g}} &= -2\int_M \phi|\widetilde{\nabla} w|_{\tilde{g}}^{p-2}\widetilde{\nabla} w\widetilde{\nabla}\phi\,d\mu_{\tilde{g}} \\ &\le 2\Big(\int_M \phi^2|\widetilde{\nabla} w|_{\tilde{g}}^{p}\,d\mu_{\tilde{g}}\Big)^{(p-1)/p}\Big(\int_M \phi^2|\widetilde{\nabla}\phi|_{\tilde{g}}^{p}\,d\mu_{\tilde{g}}\Big)^{1/p} \end{aligned}$$

by the Cauchy–Schwarz inequality ($p > 1$). Therefore, we have

$$\int_M \phi^2|\widetilde{\nabla} w|_{\tilde{g}}^{p}\,d\mu_{\tilde{g}} \le 2^p\int_M \phi^2|\widetilde{\nabla}\phi|_{\tilde{g}}^{p}\,d\mu_{\tilde{g}}.$$

Converting back to the metric g, we are led to

$$\int_M \phi^2|\nabla w|^p e^{-f}\,d\mu_g \le 2^p\int_M \phi^2|\nabla\phi|^p e^{-f}\,d\mu_g. \tag{2-3}$$

By Theorem 1.1 in [Cao and Zhou 2010], the potential function f for a shrinking gradient Ricci soliton satisfies the pointwise estimate

$$\tfrac{1}{4}(r(x)-c)^2 \le f(x) \le \tfrac{1}{4}(r(x)+c)^2 \tag{2-4}$$

for $x \in M \setminus B_{x_0}(1, g)$, where $r(x)$ is the distance from x to a fixed point x_0 in M and c is a positive constant.

Therefore, by (2-3) and (2-4),

$$\begin{aligned}\int_{B(x_0,\rho)} |\nabla w|^p e^{-(r+c)^2/4}\, d\mu_g &\le \int_M \phi^2 |\nabla w|^p e^{-f}\, d\mu_g \\ &\le \frac{2^p C e^{-(\rho-c)^2/4}}{\rho^p} \int_{B_{x_0}(2\rho,g)\setminus B_{x_0}(\rho,g)} d\mu_g \\ &\le \frac{2^p C e^{-(\rho-c)^2/4}}{\rho^p} \rho^n\end{aligned}$$

where the last inequality follows from the volume growth estimate (Theorem 1.2 in [Cao and Zhou 2010]) on shrinking gradient Ricci solitons:

$$\mathrm{Vol}(B_{x_0}(\rho, g)) \le C\rho^n$$

for sufficiently large ρ and uniform constant C. Now letting $\rho \to \infty$, we conclude $|\nabla w| \equiv 0$ on M, so u is a constant. □

Next, we show that $(M, \tilde{g})$ can be turned into a negatively curved manifold under suitable assumptions on p and the sectional curvature K of (M, g).

Proposition 2.2. *Let (M, g, f) be a simply connected n-dimensional complete noncompact shrinking gradient Ricci soliton whose sectional curvature is bounded above by a constant K_0 with $0 < K_0 < 1/(2(n-1))$. Then $(M, e^{-2f/(n-p)}g)$ is a Cartan–Hadamard manifold for $n < p \le \frac{1}{K_0} + 2 - n$.*

Proof. When $p > n$, the metric $\tilde{g} = e^{-2f/(n-p)}g$ is complete since

$$-\frac{2f(x)}{n-p} = \frac{2f(x)}{p-n} \ge \frac{(r-c)^2}{2(p-n)}$$

by [Cao and Zhou 2010] and completeness of g.

We use the conventions in [Chow et al. 2006] for curvatures. The Riemann curvature tensor is written as

$$R\Big(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\Big)\frac{\partial}{\partial x^k} = R_{ijk}^l \frac{\partial}{\partial x^l}$$

$$R_{ijkl} = \Big\langle R\Big(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\Big)\frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^l}\Big\rangle$$

and if $\partial/\partial x^1, \dots, \partial/\partial x^n$ is orthonormal at $x_0 \in M$, then the sectional curvature of the plane P_{ij} spanned by $\partial/\partial x^i$, $\partial/\partial x^j$ at x_0 is

$$K(P_{ij}) = R_{ijji}$$

and the Ricci curvature at x_0 is

$$R_{jk} = \sum_{i=1}^n R_{ijk}^i.$$

Under the conformal change of metric $\tilde{g} = e^{2f/(p-n)}g$, the sectional curvature at x_0 becomes

$$\begin{aligned}
\text{(2-5)} \qquad \widetilde{K}(P_{ij}) &= \frac{\tilde{g}\big(\widetilde{R}^s_{ijj}\frac{\partial}{\partial x^s}, \frac{\partial}{\partial x^i}\big)}{\tilde{g}_{ii}\tilde{g}_{jj} - \tilde{g}_{ij}^2} \\
&= e^{4f/(n-p)}\widetilde{R}_{ijji} \\
&= e^{4f/(n-p)} \cdot e^{2f/(p-n)}\bigg(R_{ijji} - \frac{f_{ii}+f_{jj}}{p-n} - \frac{|\nabla f|^2 - f_i^2 - f_j^2}{(p-n)^2}\bigg) \\
&= e^{2f/(n-p)}\bigg(K(P_{ij}) - \frac{f_{ii}+f_{jj}}{p-n} - \frac{|\nabla f|^2 - f_i^2 - f_j^2}{(p-n)^2}\bigg)
\end{aligned}$$

(see p. 27 in [Chow et al. 2006]). On the gradient shrinking Ricci soliton, we therefore have

$$\widetilde{K}(P_{ij}) \le e^{2f/(n-p)}\Big(K(P_{ij}) + \frac{R_{ii}+R_{jj}-1}{p-n}\Big)$$

by using the defining equation for shrinking gradient Ricci solitons and dropping the last term above that is nonpositive for $i \neq j$.

From the assumption on K_0 and $p > n$, it follows that

$$\begin{aligned}
K(P_{ij}) + \frac{R_{ii}+R_{jj}-1}{p-n} &= K(P_{ij}) + \frac{\sum_{s\neq i} K(P_{is}) + \sum_{s\neq j} K(P_{sj}) - 1}{p-n} \\
&\le \Big(1 + \frac{2(n-1)}{p-n}\Big)K_0 - \frac{1}{p-n} \\
&\le \frac{1}{p-n}((p+n-2)K_0 - 1).
\end{aligned}$$

Therefore, we conclude that the sectional curvature $\widetilde{K}$ of $(M, e^{2f/(p-n)}g)$ is nonpositive since $p+n-2 \le \frac{1}{K_0}$. □

Proof of Theorem 1.1. Suppose there is a solution u to the Dirichlet problem at infinity and $u = \varphi$ on $M(\infty)$ for some nonconstant function $\varphi \in C^0(M(\infty))$. Then u is continuous on $M \cup M(\infty)$, hence it is bounded. Then $u - \inf_M u + 1$ is a positive solution to the p-Laplace equation on $(M, \tilde{g})$, therefore it must be constant from Proposition 2.1. Thus, u is constant on M and φ must be constant on $M(\infty)$. The contradiction concludes the proof. □

When $\mathbb{R}^n$ is viewed as a shrinking gradient Ricci soliton with $f(x) = |x|^2/4$, we can take $K_0 = 0$ and obtain the following corollary.

Corollary 2.3. *The Dirichlet problem at infinity for the p-Laplace equation is unsolvable on* $(\mathbb{R}^n, e^{|x|^2/(2(p-n))}dx^2)$ *for every $p > n$.*

Remark. The sectional curvature of $\tilde{g} = e^{2|x|^2/(4(p-n))}dx^2$ can be computed from (2-5):

$$\widetilde{K}(P_{ij})(x) = -e^{-|x|^2/(2(p-n))}\Big(\frac{1}{p-n} + \frac{|x|^2-(x^i)^2-(x^j)^2}{4(p-n)^2}\Big)$$

where $P_{ij}(x)$ is the plane spanned by $\{\partial/\partial x^i, \partial/\partial x^j\}$ at $x \in \mathbb{R}^n$. The Riemannian distance from x to the origin is

$$r(x) = \int_0^{|x|} e^{s^2/(4(p-n))}\,ds.$$

If we take $x = (0, \dots, 0, x^i, 0, \dots, 0)$, then $|x|^2 - (x^i)^2 - (x^j)^2 = 0$ and

$$\begin{aligned}\lim_{|x|\to\infty} -\widetilde{K}(P_{ij}(x))r^2(x) &= \lim_{|x|\to\infty} \frac{\big(\int_0^{|x|} e^{s^2/(4(p-n))}\,ds\big)^2}{(p-n)e^{|x|^2/(2(p-n))}} \\ &= \frac{1}{p-n}\Big(\lim_{|x|\to\infty} \frac{2(p-n)}{|x|}\Big)^2 = 0\end{aligned}$$

by l'Hôpital's rule. This in particular shows that there does not exist a constant $\alpha > 1$ for which

$$K(x) \le -\frac{\alpha(\alpha-1)}{r^2(x)}$$

for all sections at x for large $r(x)$.

3. A Liouville theorem on $\mathbb{R}^2$ with an incomplete metric with positive curvature

In this section, we consider the p-Laplace equation weighted by a smooth function f on a manifold (M, g), which is equivalent to the p-Laplace equation on $(M, e^{-2f/(n-p)}g)$, and derive a Bochner formula for its solutions. Specialized to the shrinking or steady gradient Ricci solitons, the Bochner formula yields a maximum principle, and this is applied to Hamilton's cigar soliton.

A Bochner type formula for the weighted p-Laplace equation. Let g be a Riemannian metric on an n-dimensional manifold M, and let f be a smooth real-valued function on M. Consider the equation

$$\operatorname{div}(|\nabla u|^{p-2}\nabla u) - |\nabla u|^{p-2}\langle\nabla f, \nabla u\rangle = 0 \tag{3-1}$$

on M. This equation has a variational structure; in fact, it is the Euler–Lagrange equation of the weighted p-energy functional

$$E_{p,f}(u) = \int_M |\nabla u|^p e^{-f}\,d\mu_g.$$

We call (3-1) the f-weighted p-Laplacian equation on (M, g).

Proposition 3.1. *Under a conformal change $\tilde{g} = e^{-2f/(n-p)}g$, u is a solution to* (3-1) *on (M, g) if and only if u is a solution to the p-Laplace equation* (1-1) *on $(M, \tilde{g})$.*

Proof. We write ∇ for ∇_g and $\widetilde{\nabla}$ for $\nabla_{\tilde{g}}$. For any $\varphi \in C_0^\infty(M)$,

$$\begin{aligned}
\int_M \langle \widetilde{\nabla}\varphi, |\widetilde{\nabla}u|_{\tilde{g}}^{p-2}\widetilde{\nabla}u\rangle_{\tilde{g}}\, d\mu_{\tilde{g}} \\
&= \int_M |\widetilde{\nabla}u|_{\tilde{g}}^{p-2}\langle \widetilde{\nabla}\varphi, \widetilde{\nabla}u\rangle_{\tilde{g}}\, d\mu_{\tilde{g}} \\
&= \int_M \big(e^{(p-2)f/(n-p)}|\nabla u|_g^{p-2}\big)e^{2f/(n-p)}\langle \nabla\varphi, \nabla u\rangle_g\, e^{-nf/(n-p)}\, d\mu_g \\
&= \int_M \langle \nabla\varphi, |\nabla u|_g^{p-2}\nabla u\rangle_g\, e^{-f}\, d\mu_g.
\end{aligned}$$

This shows that any weak solution to (3-1) on (M, g) is also a weak solution to (1-1) on $(M, \tilde{g})$ and vice versa. □

Suppose $u(x, t)$ is a positive solution of (3-1). Define

$$w = -(p-1)\log u,$$
$$h = |\nabla w|^2.$$

We consider the symmetric $n \times n$ matrix

$$A = \mathrm{id} + (p-2)\frac{\nabla w \otimes \nabla w}{h}.$$

Note that A is well defined whenever $h > 0$ and is positive definite for $p > 1$. Arising from the linearized operator of the nonlinear p-harmonic equations, this matrix was first introduced in [Moser 2007] and was used in [Kotschwar and Ni 2009; Wang and Zhang 2011] to study positive p-harmonic functions.

For the f-weighted p-Laplace equation (3-1), the linearized operator is

$$\mathcal{L}(\psi) = \mathrm{div}\big(h^{\frac{p}{2}-1}A(\nabla\psi)\big) - h^{\frac{p}{2}-1}\langle \nabla f, A(\nabla\psi)\rangle - ph^{\frac{p}{2}-1}\langle \nabla w, \nabla\psi\rangle$$

for smooth functions ψ on M, and the following Bochner type formula holds.

Proposition 3.2. *Let u be a positive smooth solution to* (3-1) *in an open subset U in M and assume $h > 0$ on U. Then*

$$\begin{aligned}
\text{(3-2)}\quad & \mathrm{div}\big(h^{\frac{p}{2}-1}A(\nabla h)\big) - h^{\frac{p}{2}-1}\langle \nabla f, A(\nabla h)\rangle - ph^{\frac{p}{2}-1}\langle \nabla w, \nabla h\rangle \\
&= \big(\tfrac{p}{2}-1\big)|\nabla h|^2 h^{\frac{p}{2}-2} + 2h^{\frac{p}{2}-1}\big(|\nabla\nabla w|^2 + \mathrm{Ric}(\nabla w, \nabla w) + \nabla\nabla f(\nabla w, \nabla w)\big).
\end{aligned}$$

Proof. Using (3-1), we first observe

$$\begin{aligned}\text{(3-3)}\quad \operatorname{div}(|\nabla w|^{p-2}\nabla w) - |\nabla w|^p &= -(p-1)^{p-1}\operatorname{div}\Big(\frac{|\nabla u|^{p-2}\nabla u}{u^{p-1}}\Big) - (p-1)^p\frac{|\nabla u|^p}{u^p}\\ &= -(p-1)^{p-1}\frac{|\nabla u|^{p-2}\langle\nabla f,\nabla u\rangle}{u^{p-1}}\\ &= |\nabla w|^{p-2}\langle\nabla f,\nabla w\rangle.\end{aligned}$$

Then we calculate directly

$$\begin{aligned}&\operatorname{div}\big(h^{\frac{p}{2}-1}A(\nabla h)\big)\\ &\quad= \big(\tfrac{p}{2}-1\big)h^{\frac{p}{2}-2}|\nabla h|^2 + h^{\frac{p}{2}-1}\Delta h + \big(\tfrac{p}{2}-2\big)(p-2)h^{\frac{p}{2}-3}\langle\nabla w,\nabla h\rangle^2\\ &\qquad + (p-2)h^{\frac{p}{2}-2}\langle\nabla w,\nabla h\rangle\Delta w + (p-2)h^{\frac{p}{2}-2}\langle\nabla\langle\nabla w,\nabla h\rangle,\nabla w\rangle.\end{aligned}$$

Using the standard Bochner type formula for $h = |\nabla w|^2$, namely

$$\Delta h = 2|\nabla\nabla w|^2 + 2\operatorname{Ric}(\nabla w,\nabla w) + 2\langle\nabla\Delta w,\nabla w\rangle,$$

we have

$$\begin{aligned}\text{(3-4)}\quad &\operatorname{div}\big(h^{\frac{p}{2}-1}A(\nabla h)\big)\\ &= \big(\tfrac{p}{2}-1\big)h^{\frac{p}{2}-2}|\nabla h|^2 + 2h^{\frac{p}{2}-1}\big(|\nabla\nabla w|^2 + \operatorname{Ric}(\nabla w,\nabla w) + \langle\nabla\Delta w,\nabla w\rangle\big)\\ &\quad + \big(\tfrac{p}{2}-2\big)(p-2)h^{\frac{p}{2}-3}\langle\nabla w,\nabla h\rangle^2 + (p-2)h^{\frac{p}{2}-2}\langle\nabla w,\nabla h\rangle\Delta w\\ &\quad + (p-2)h^{\frac{p}{2}-2}\langle\nabla\langle\nabla w,\nabla h\rangle,\nabla w\rangle.\end{aligned}$$

Rewrite (3-3) by using $h = |\nabla w|^2$ as

$$\text{(3-5)}\qquad h^{\frac{p}{2}-1}\Delta w + \big(\tfrac{p}{2}-1\big)h^{\frac{p}{2}-2}\langle\nabla h,\nabla w\rangle - h^{\frac{p}{2}} = h^{\frac{p}{2}-1}\langle\nabla f,\nabla w\rangle.$$

Taking the gradient of both sides of (3-5) and then taking the product with ∇w, we are led to

$$\begin{aligned}\text{(3-6)}\quad &\big(\tfrac{p}{2}-1\big)\big(\tfrac{p}{2}-2\big)h^{\frac{p}{2}-3}\langle\nabla w,\nabla h\rangle^2 + \big(\tfrac{p}{2}-1\big)h^{\frac{p}{2}-2}\langle\nabla\langle\nabla w,\nabla h\rangle,\nabla w\rangle\\ &\quad + h^{\frac{p}{2}-1}\langle\nabla\Delta w,\nabla w\rangle + \big(\tfrac{p}{2}-1\big)h^{\frac{p}{2}-2}\langle\nabla h,\nabla w\rangle\Delta w - \tfrac{p}{2}h^{\frac{p}{2}-1}\langle\nabla h,\nabla w\rangle\\ &\quad = \big(\tfrac{p}{2}-1\big)h^{\frac{p}{2}-2}\langle\nabla f,\nabla w\rangle\langle\nabla h,\nabla w\rangle + h^{\frac{p}{2}-1}\langle\nabla\langle\nabla f,\nabla w\rangle,\nabla w\rangle.\end{aligned}$$

Adding (3-4) and twice (3-6) together and then simplifying, we have

$$\begin{aligned}\text{(3-7)}\quad &\operatorname{div}\big(h^{\frac{p}{2}-1}A(\nabla h)\big) - ph^{\frac{p}{2}-1}\langle\nabla h,\nabla w\rangle\\ &= \big(\tfrac{p}{2}-1\big)h^{\frac{p}{2}-2}|\nabla h|^2 + 2h^{\frac{p}{2}-1}|\nabla\nabla w|^2 + 2h^{\frac{p}{2}-1}\operatorname{Ric}(\nabla w,\nabla w)\\ &\quad + (p-2)h^{\frac{p}{2}-2}\langle\nabla f,\nabla w\rangle\langle\nabla h,\nabla w\rangle + 2h^{\frac{p}{2}-1}\langle\nabla\langle\nabla f,\nabla w\rangle,\nabla w\rangle.\end{aligned}$$

We also have

$$\begin{aligned}(3\text{-}8)\quad 2h^{\frac{p}{2}-1}\langle\nabla\langle\nabla f,\nabla w\rangle,\nabla w\rangle\\ &= 2h^{\frac{p}{2}-1}(\nabla\nabla f)(\nabla w,\nabla w)+2h^{\frac{p}{2}-1}(\nabla\nabla w)(\nabla f,\nabla w)\\ &= 2h^{\frac{p}{2}-1}(\nabla\nabla f)(\nabla w,\nabla w)+h^{\frac{p}{2}-1}\langle\nabla f,\nabla|\nabla w|^2\rangle\\ &= 2h^{\frac{p}{2}-1}(\nabla\nabla f)(\nabla w,\nabla w)+h^{\frac{p}{2}-1}\langle\nabla f,\nabla h\rangle.\end{aligned}$$

Moreover,

$$\begin{aligned}(3\text{-}9)\quad h^{\frac{p}{2}-1}\langle\nabla f,A(\nabla h)\rangle &= h^{\frac{p}{2}-1}\langle\nabla f,\nabla h\rangle+(p-2)h^{\frac{p}{2}-2}\langle\nabla f,(\nabla w\otimes\nabla w)\nabla h\rangle\\ &= h^{\frac{p}{2}-1}\langle\nabla f,\nabla h\rangle+(p-2)h^{\frac{p}{2}-2}\langle\nabla f,\nabla w\rangle\langle\nabla h,\nabla w\rangle.\end{aligned}$$

Now, (3-7) − (3-9) + (3-8) yields the desired result. □

A maximum principle. When the triple (M, g, f) is either shrinking or steady, Proposition 3.2 can be used to prove the following maximum principle.

Proposition 3.3. *Let u be a positive smooth solution to* (3-1) *in a bounded connected open subset U in M with smooth boundary ∂U, $p > 1$. Suppose (M, g, f) is a shrinking or steady gradient Ricci soliton. Then $|\nabla u|/u$ attains its maximum on ∂U.*

Proof. Let $h = (p-1)^2|\nabla u|^2/u^2$. Assume $\max_{\overline{U}} h > \max_{\partial U} h$. Then there exists $x_0 \in U$ such that $h(x_0) = \max_{\overline{U}} h > 0$. Since $u \in C^{1,\alpha}$ and $u > 0$, h is continuous. Let

$$V = \{x \in U : h(x) = h(x_0)\}.$$

By the continuity of h, V is a closed subset of U and V does not intersect ∂U. In fact, h is positive and hence smooth in a neighborhood of V. There exists a point $x_1 \in V$ such that for some r_0 the geodesic ball $B_{x_1}(r, g) \subset U$ is not contained in V for any $0 < r < r_0$, i.e., x_1 is a boundary point of V. By the continuity of h again, there is a geodesic ball $B_{x_1}(r_1, g)$ in U on which h is positive. Observe that

$$\begin{aligned}\text{RHS of (3-2)} &= \tfrac{p-2}{2}|\nabla h|^2h^{\frac{p}{2}-2}+2h^{\frac{p}{2}-1}|\nabla\nabla w|^2+2h^{\frac{p}{2}-1}(\mathrm{Ric}+\nabla\nabla f)(\nabla w,\nabla w)\\ &\geq 2h^{\frac{p}{2}-1}(\mathrm{Ric}+\nabla\nabla f)(\nabla w,\nabla w)\\ &= \begin{cases}2h^{\frac{p}{2}-1}|\nabla w|^2\geq 0 & \text{if } (M,g,f) \text{ is a shrinking soliton,}\\ 0 & \text{if } (M,g,f) \text{ is a steady soliton,}\end{cases}\end{aligned}$$

where for the first inequality, we argue as

$$\begin{aligned}4h|\nabla\nabla w|^2+(p-2)|\nabla h|^2 &\geq 4|\nabla w|^2|\nabla\nabla w|^2-|\nabla|\nabla w|^2|^2\\ &= 4|\nabla w|^2(|\nabla\nabla w|^2-|\nabla|\nabla w||^2)\\ &\geq 0\end{aligned}$$

by Kato's inequality and $p \geq 1$. Then it follows that the linear differential operator $\mathcal{L}$ satisfies $\mathcal{L}(h) \geq 0$ on U. Next, since A is positive definite and symmetric on $B_{x_1}(r_1, g)$, so is $h^{\frac{p}{2}-1}A$; therefore, $\mathcal{L}$ is uniformly elliptic on $B_{x_1}(r_1, g)$. By Hopf's strong maximum principle (see Theorem 3.5 in [Gilbarg and Trudinger 1998]), h must be a constant on $B_{x_1}(r_1, g)$ since it attains its maximum at the interior point x_1. But this contradicts the maximality of V as $B_{x_1}(r_1, g)$ contains points not in V. □

Gradient estimates. Let us first recall the following gradient estimate:

Theorem 3.4 [Wang and Zhang 2011]. *Let (M^n, g) be a complete Riemannian manifold with* $\operatorname{Ric} \geq -(n-1)\kappa$ *for some positive constant κ. Assume that v is a positive p-harmonic function on the geodesic ball $B_{x_0}(R, g) \subset M$. Then*

$$\frac{|\nabla v|}{v} \leq C(p, n)\Big(\frac{1}{R} + \sqrt{\kappa}\Big)$$

on $B_{x_0}\big(\frac{R}{2}, g\big)$ for some constant $C(p, n)$.

We now prove a gradient estimate for the f-weighted p-Laplacian equation.

Proposition 3.5. *Let (M^n, g, f) be a complete gradient Ricci soliton with*

$$\left(\frac{2-p}{n-p}\right)\operatorname{Ric} \geq -(n-1)\kappa e^{-2f/(n-p)} g - \frac{2\varepsilon g}{n-p} - \frac{Sg}{n-p} - (df \otimes df - |\nabla f|^2 g)\frac{n-2}{(n-p)^2}, \tag{3-10}$$

where S is the scalar curvature of (M, g). Assume that u is a positive solution of equation (3-1). *Then there exists a constant $C(p, n)$ such that*

$$\frac{|\nabla u(x)|}{u(x)} \leq C(p, n)\Big(\frac{1}{R} + \sqrt{\kappa}\Big)e^{-f(x)/(n-p)}$$

for $x \in B_{x_0}\big(\frac{R}{2}, e^{-2f/(n-p)} g\big)$.

Proof. For a smooth function f, let ∇f be the gradient, Δf the Laplacian, and $\nabla\nabla f$ the Hessian with respect to g. For the conformal change of metrics $\tilde{g} = e^{-2f/(n-p)} g$, the Ricci tensors of $\tilde{g}$ and g are related by

$$\widetilde{\operatorname{Ric}} = \operatorname{Ric} - (n-2)\Big(-\frac{\nabla\nabla f}{n-p} - \frac{df \otimes df}{(n-p)^2}\Big) + \Big(-\frac{\Delta f}{n-p} - \frac{n-2}{(n-p)^2}|\nabla f|^2\Big) g \tag{3-11}$$

(see [Anderson and Schoen 1985, p. 59]).

From the gradient Ricci soliton equation (2-1), the scalar curvature S of M satisfies the two equations

$$S + \Delta f - n\varepsilon = 0, \tag{3-12}$$

$$S + |\nabla f|^2 + \varepsilon f = 0 \tag{3-13}$$

(see [Besse 1987]).

Putting (2-1) and (3-12) into (3-11), we have

$$\begin{aligned}\widetilde{\mathrm{Ric}} &= \mathrm{Ric} + (n-2)\Big(\frac{-\mathrm{Ric} - \varepsilon g}{n-p} + \frac{df \otimes df}{(n-p)^2}\Big) + \Big(\frac{S + n\varepsilon}{n-p} - \frac{n-2}{(n-p)^2}|\nabla f|^2\Big)g \\ &= \frac{2-p}{n-p}\mathrm{Ric} + \frac{2\varepsilon g}{n-p} + \frac{Sg}{n-p} + (df \otimes df - |\nabla f|^2 g)\frac{n-2}{(n-p)^2}.\end{aligned}$$

Therefore, the curvature assumption in Proposition 3.5 implies

$$\widetilde{\mathrm{Ric}} \geq -(n-1)\kappa.$$

By Proposition 3.1, we know that u is also a positive solution to (1-1) for the metric $\tilde{g}$, hence by Theorem 3.4 we have

$$\frac{|\nabla u|_{\tilde{g}}}{u} \leq C(p,n)\Big(\frac{1}{R} + \sqrt{\kappa}\Big)$$

on $B_{x_0}\big(\frac{R}{2}, \tilde{g}\big)$. This is equivalent to

$$\frac{|\nabla u(x)|}{u(x)} \leq C(p,n)\Big(\frac{1}{R} + \sqrt{\kappa}\Big)e^{-f(x)/(n-p)}$$

for $x \in B_{x_0}\big(\frac{R}{2}, \tilde{g}\big)$. □

A Liouville type theorem for the p-Laplace equation in dimension 2. For a steady gradient Ricci soliton, the condition (3-10) on the Ricci curvature in Proposition 3.5 cannot hold globally when $n \geq 3$ because it would imply, by taking the trace, that the scalar curvature is bounded below by a positive constant, which is impossible. However, the condition (3-10) is satisfied when $n = 2$ for $p \geq 4$ or $1 < p < 2$ because

$$\mathrm{Ric} = \tfrac{1}{2}Sg \geq \tfrac{1}{p-2}Sg,$$

since $S \geq 0$ for any steady gradient Ricci soliton [Chen 2009] and $\kappa = 0$.

Note that Hamilton's cigar soliton is the unique 2-dimensional nonflat complete noncompact steady gradient Ricci soliton. The cigar soliton is $\mathbb{R}^2$ equipped with the complete metric

$$g = \frac{dx^2 + dy^2}{1 + x^2 + y^2}$$

(see [Chow et al. 2006]) and the potential function

$$f(x,y) = -\log(1 + x^2 + y^2).$$

The conformally altered metric is

$$\tilde{g} = e^{2\log(1+x^2+y^2)/(2-p)}g = (1 + x^2 + y^2)^{p/(2-p)}(dx^2 + dy^2).$$

In particular, $\tilde{g}$ is complete if $1 < p < 2$ and incomplete if $p > 2$. However, to use the gradient estimate in proving a Liouville type result, we will need $p \geq 4$. It is straightforward to compute the Gauss curvature of $\tilde{g}$:

$$\begin{aligned}\widetilde{K} &= -\frac{1}{2}(1+r^2)^{p/(p-2)}\Big(\partial_{rr}^2 + \frac{1}{r}\partial_r\Big)\log(1+r^2)^{-p/(p-2)} \\ &= \frac{2p}{p-2}(1+r^2)^{(p/(p-2))-2} \\ &= \frac{2p}{p-2}(1+r^2)^{-(p-4)/(p-2)}\end{aligned}$$

which is positive and tends to 0 as $r \to \infty$ if $p > 4$. When $p = 4$, the incomplete metric $(1+x^2+y^2)^{-2}(dx^2+dy^2)$ has constant curvature $\widetilde{K} = 4$.

Theorem 3.6. *Let* $(\mathbb{R}^2, g, f)$ *be Hamilton's cigar soliton. Then there does not exist any nonconstant positive p-harmonic function on* $(\mathbb{R}^2, \tilde{g})$ *for* $p \geq 4$.

Proof. Let u be a positive solution to (3-1). For any point $x_0 \in M$, the maximum principle (Proposition 3.3) asserts

$$\frac{|\nabla u(x_0)|}{u(x_0)} \leq \max_{x \in \partial B_0(R,g)} \frac{|\nabla u(x)|}{u(x)} = \frac{|\nabla u(x_R)|}{u(x_R)}$$

for some $x_R \in \partial B_0(R, g)$ where $x_0 \in B_0(R, g)$ and $r(x_0, 0) < R$. From the discussion above, when $n = 2$ and $p \geq 4$, the Ricci curvature condition (3-10) in Proposition 3.5 is satisfied. The diameter of $(\mathbb{R}^2, \tilde{g})$ is

$$2R_0 = 2\int_0^\infty \frac{dr}{(1+r^2)^{p/(2(p-2))}} < \infty.$$

It is clear that $r(x_R, 0) \to \infty$ if and only if $\tilde{r}(x_R, 0) \to R_0$, where $\tilde{r}$ denotes the distance function for the metric $\tilde{g}$. Let

$$r_R = \int_R^\infty \frac{dr}{(1+r^2)^{p/(2(p-2))}}.$$

It follows from Proposition 3.5, applied on the ball $B_{x_R}(r_R, \tilde{g})$, that

$$\begin{aligned}\frac{|\nabla u(x_R)|}{u(x_R)} &\leq C(n,p)\Big(\frac{r_{x_R}}{2}\Big)^{-1} e^{-2\log(1+|x_R|^2)/(p-2)} \\ &= 2C(n,p)\bigg(\int_R^\infty \frac{dr}{(1+r^2)^{p/(2(p-2))}}(1+R^2)^{2/(p-2)}\bigg)^{-1} \\ &\leq 2C(n,p)\bigg((1+R^2)^{2/(p-2)}\int_R^\infty \frac{dr}{r^{p/(p-2)}}\bigg)^{-1} \\ &= 2C(n,p)\Big(\frac{p-2}{2}(1+R^2)^{2/(p-2)}R^{-2/(p-2)}\Big)^{-1}.\end{aligned}$$

Since $p > 2$, letting $R \to 0$ we conclude $|\nabla u(x_0)| = 0$, hence u is constant as x_0 is arbitrary. □

Acknowledgement

Chen would like to thank Jiaping Wang for valuable discussions and his interest in this work.

References

[Ancona 1994] A. Ancona, "Convexity at infinity and Brownian motion on manifolds with unbounded negative curvature", *Rev. Mat. Iberoamericana* **10**:1 (1994), 189–220. MR 95a:58132 Zbl 0804.58056

[Anderson 1983] M. T. Anderson, "The Dirichlet problem at infinity for manifolds of negative curvature", *J. Differential Geom.* **18**:4 (1983), 701–721. MR 85m:58178 Zbl 0541.53036

[Anderson and Schoen 1985] M. T. Anderson and R. Schoen, "Positive harmonic functions on complete manifolds of negative curvature", *Ann. of Math.* (2) **121**:3 (1985), 429–461. MR 87a:58151 Zbl 0587.53045

[Besse 1987] A. L. Besse, *Einstein manifolds*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **10**, Springer, Berlin, 1987. MR 88f:53087 Zbl 0613.53001

[Cao and Zhou 2010] H.-D. Cao and D. Zhou, "On complete gradient shrinking Ricci solitons", *J. Differential Geom.* **85**:2 (2010), 175–186. MR 2011k:53040 Zbl 1246.53051

[Chen 2009] B.-L. Chen, "Strong uniqueness of the Ricci flow", *J. Differential Geom.* **82**:2 (2009), 363–382. MR 2010h:53095 Zbl 1177.53036

[Choi 1984] H. I. Choi, "Asymptotic Dirichlet problems for harmonic functions on Riemannian manifolds", *Trans. Amer. Math. Soc.* **281**:2 (1984), 691–716. MR 85b:53040 Zbl 0541.53035

[Chow et al. 2006] B. Chow, P. Lu, and L. Ni, *Hamilton's Ricci flow*, Graduate Studies in Mathematics **77**, American Mathematical Society, Providence, RI, 2006. MR 2008a:53068 Zbl 1118.53001

[Dung and Dat 2015] N. T. Dung and N. D. Dat, "Local and global sharp gradient estimates for weighted p-harmonic functions", preprint, Vietnam Institute for Advanced Study in Mathematics, Ha Noi, 2015. arXiv 1505.07623

[Eberlein and O'Neill 1973] P. Eberlein and B. O'Neill, "Visibility manifolds", *Pacific J. Math.* **46** (1973), 45–109. MR 49 #1421 Zbl 0264.53026

[Gilbarg and Trudinger 1998] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, revised 2nd ed., Grundlehren der Mathematischen Wissenschaften **224**, Springer, Berlin, 1998. MR 2001k:35004 Zbl 1042.35002

[Hamilton 1995] R. S. Hamilton, "The formation of singularities in the Ricci flow", pp. 7–136 in *Proceedings of the Conference on Geometry and Topology* (Cambridge, MA, 1993), edited by C.-C. Hsiung and S.-T. Yau, Surveys in Differential Geometry **2**, International Press, Cambridge, MA, 1995. MR 97e:53075 Zbl 0867.53030

[Holopainen 2002] I. Holopainen, "Asymptotic Dirichlet problem for the p-Laplacian on Cartan–Hadamard manifolds", *Proc. Amer. Math. Soc.* **130**:11 (2002), 3393–3400. MR 2003d:58055 Zbl 1016.58021

[Holopainen and Vähäkangas 2007] I. Holopainen and A. Vähäkangas, "Asymptotic Dirichlet problem on negatively curved spaces", *J. Anal.* **15** (2007), 63–110. MR 2010j:58056 Zbl 1202.58017

[Hsu 2003] E. P. Hsu, "Brownian motion and Dirichlet problems at infinity", *Ann. Probab.* **31**:3 (2003), 1305–1319. MR 2004e:58063 Zbl 1054.58027

[Kotschwar and Ni 2009] B. Kotschwar and L. Ni, "Local gradient estimates of p-harmonic functions, $1/H$-flow, and an entropy formula", *Ann. Sci. École Norm. Sup.* (4) **42**:1 (2009), 1–36. MR 2010g:53121 Zbl 1182.53060

[Moser 2007] R. Moser, "The inverse mean curvature flow and p-harmonic functions", *J. Eur. Math. Soc.* **9**:1 (2007), 77–83. MR 2007h:53102 Zbl 1116.53040

[Munteanu and Sesum 2013] O. Munteanu and N. Sesum, "On gradient Ricci solitons", *J. Geom. Anal.* **23**:2 (2013), 539–561. MR 3023848 Zbl 1275.53061

[Munteanu and Wang 2012] O. Munteanu and J. Wang, "Analysis of weighted Laplacian and applications to Ricci solitons", *Comm. Anal. Geom.* **20**:1 (2012), 55–94. MR 2903101 Zbl 1245.53039

[Pansu 1989] P. Pansu, "Cohomologie L^p des variétés à courbure négative, cas du degré 1", *Rend. Sem. Mat. Univ. Politec. Torino* Special Issue (1989), 95–120. MR 92e:58200 Zbl 0723.53023

[Sullivan 1983] D. Sullivan, "The Dirichlet problem at infinity for a negatively curved manifold", *J. Differential Geom.* **18**:4 (1983), 723–732. MR 85m:58177 Zbl 0541.53037

[Sung and Wang 2014] C.-J. A. Sung and J. Wang, "Sharp gradient estimate and spectral rigidity for p-Laplacian", *Math. Res. Lett.* **21**:4 (2014), 885–904. MR 3275651 Zbl 1304.58018

[Tolksdorf 1984] P. Tolksdorf, "Regularity for a more general class of quasilinear elliptic equations", *J. Differential Equations* **51**:1 (1984), 126–150. MR 85g:35047 Zbl 0488.35017

[Wang and Zhang 2011] X. Wang and L. Zhang, "Local gradient estimate for p-harmonic functions on Riemannian manifolds", *Comm. Anal. Geom.* **19**:4 (2011), 759–771. MR 2880214 Zbl 1244.58007

[Wu 2013] P. Wu, "On the potential function of gradient steady Ricci solitons", *J. Geom. Anal.* **23**:1 (2013), 221–228. MR 3010278 Zbl 1269.53051

[Yau 1975] S.-T. Yau, "Harmonic functions on complete Riemannian manifolds", *Comm. Pure Appl. Math.* **28** (1975), 201–228. MR 55 #4042 Zbl 0291.31002

Received November 13, 2014. Revised September 3, 2015.

JINGYI CHEN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
ROOM 121, 1984 MATHEMATICS ROAD
VANCOUVER BC V6T 1Z2
CANADA

jychen@math.ubc.ca

YUE WANG
DEPARTMENT OF MATHEMATICS
CHINA JILIANG UNIVERSITY
HANGZHOU, ZHEJIANG 310018
CHINA

kellywong@cjlu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.329

CARTAN–FUBINI TYPE RIGIDITY OF DOUBLE COVERING MORPHISMS OF QUADRATIC MANIFOLDS

HOSUNG KIM

Let $Z \subset \mathbb{P}^N$ be a Fano manifold whose Picard group is generated by the hyperplane section class. Assume that Z is covered by lines and $i(Z) \geq 3$. Let $\phi : X^Z \to Z$ be a double cover, branched along a smooth hypersurface section of degree $2m$, $1 \leq m \leq i(Z) - 2$. We describe the defining ideal of *the variety of minimal rational tangents* at a general point. As an application, we show that if $Z \subset \mathbb{P}^N$ is defined by quadratic equations and $2 \leq m \leq i(Z) - 2$, then the morphism ϕ satisfies the Cartan–Fubini type rigidity property.

1. Introduction

Throughout the paper, we will work over the field of complex numbers. Let X be a Fano manifold of Picard number 1. The index of X is the integer $i(X)$ such that $-K_X = i(X)L$ where L is the ample generator of the Picard group of X. For a general point $x \in X$, a rational curve through x is called a *minimal rational curve* if it has minimal K_X^{-1}-degree among all rational curves through x. Denote by $\mathcal{K}_x$ the normalized space of minimal rational curves through x. It is known (e.g., [Kollár 1996, II.3.11.5]) that $\mathcal{K}_x$ is a disjoint union of finitely many nonsingular projective varieties of dimension $i(X) - 2$. The rational morphism $\mathcal{K}_x \dashrightarrow \mathbb{P}T_x(X)$, sending a member of $\mathcal{K}_x$ which is smooth at x to its tangent direction, can be extended to a birational morphism $\tau_x : \mathcal{K}_x \to \mathbb{P}T_x(X)$; see [Hwang and Mok 2004; Kebekus 2002]. We denote the image of τ_x by $\mathcal{C}_x$ and call it *the variety of minimal rational tangents (VMRT)* at x. The projective geometry of $\mathcal{C}_x \subset \mathbb{P}T_x(X)$ helps us to understand the geometry of X. This is the motivation for the study of the VMRT for various examples of X. For example, let $\phi : X \to \mathbb{P}^n$ be a double cover branched on a smooth hypersurface of degree $2m$, $2 \leq m \leq n - 1$. Then for a general point $x \in X$, the VMRT $\mathcal{C}_x \subset \mathbb{P}T_x(X)$ is a complete intersection of multidegree $(m+1, \ldots, 2m)$, and this description implies a certain rigidity property of ϕ [Hwang and Kim 2013].

This work was supported by the National Institute for Mathematical Sciences (NIMS) grant funded by the Korea government (No. C21501).
MSC2010: 14J45.
Keywords: double covers of Fano manifolds, varieties of minimal rational tangents, Cartan–Fubini type rigidity.

Let $Z \subset \mathbb{P}^N$ be a Fano manifold whose Picard group is generated by the hyperplane section class. For each point $y \in Z$, we denote by $\mathcal{L}_y(Z) \subset \mathbb{P}T_y(Z)$ the space of tangent directions of lines in Z through y. We say that $Z \subset \mathbb{P}^N$ is covered by lines if $\mathcal{L}_y(Z)$ is nonempty for each $y \in Z$. If $Z \subset \mathbb{P}^N$ is covered by lines, then minimal rational curves on Z are lines in $\mathbb{P}^N$ contained in Z, and for general $y \in Z$, $\mathcal{L}_y(Z) \subset \mathbb{P}T_y(Z)$ coincides with the VMRT $\mathcal{C}_y \subset \mathbb{P}T_y(Z)$, which is smooth of dimension $i(Z)-2$; see [Hwang 2001, Proposition 1.5].

Our first result is the following theorem.

Theorem 1.1. *Let $Z \subset \mathbb{P}^N$ be a Fano manifold whose Picard group is generated by the hyperplane section class. Assume that Z is covered by lines and $i(Z) \geq 3$. Let $Y \subset \mathbb{P}^N$ be a hypersurface of degree $2m$, $1 \leq m \leq i(Z)-2$, with smooth intersection $Y \cap Z$. Let $\phi : X^Z \to Z$ be a double cover branched along $Y \cap Z$. Then for a general point $x \in X^Z$, the VMRT $\mathcal{C}_x$ is smooth of dimension $i(Z)-m-2$ and the differential $d\phi_x : \mathbb{P}T_x(X^Z) \to \mathbb{P}T_{\phi(x)}(Z)$ sends the VMRT $\mathcal{C}_x \subset \mathbb{P}T_x(X^Z)$ isomorphically to an intersection of $\mathcal{L}_{\phi(x)}(Z) \subset \mathbb{P}T_{\phi(x)}(Z)$ and m hypersurfaces in $\mathbb{P}T_{\phi(x)}(Z)$ of degrees $m+1, \dots, 2m$ respectively.*

In order to prove the above theorem, we first show that for a certain choice of Y, the statements in Theorem 1.1 hold by identifying minimal rational curves on X^Z with ECO (even contact order) lines with respect to Y (see Definition 2.4) contained in Z. For arbitrary Y, we use a flatness argument.

We are going to show an application of Theorem 1.1. First, let us introduce the definition of Cartan–Fubini type rigidity (CF-rigidity) which was initially defined by Jun-Muk Hwang.

Definition 1.2 (Cartan–Fubini type rigidity). Let X_1 and X_2 be Fano manifolds of Picard number 1 such that $2 \leq \dim X_1 \leq \dim X_2$, and for general $x_1 \in X_1$ and $x_2 \in X_2$, $0 \leq \dim \mathcal{K}_{x_1} \leq \dim \mathcal{K}_{x_2}$. We say that a morphism $\phi : X_1 \to X_2$ is *CF-rigid* if for any connected open subset (in classical topology) U of X_1 and any biholomorphic immersion $\psi : U \to X_2$ such that for any member C of $\mathcal{K}_x$, $x \in U$, $\psi(C \cap U)$ is contained in a minimal rational curve of X_2, then there exists $\Gamma \in \mathrm{Aut}(X_2)$ such that $\psi = \Gamma \circ \phi|_U$.

The next theorem is on the CF-rigidity of the identity morphism, which was essentially proved in [Hwang and Mok 2001].

Theorem 1.3 (Cartan–Fubini type extension theorem). *Let X be a Fano manifold of Picard number* 1 *and suppose that* $\dim \mathcal{K}_x \geq 1$ *for general $x \in X$. Then the identity morphism on X is CF-rigid*, *i.e.*, *for any connected open subset* (*in the standard topology*) *U of X and any biholomorphic immersion $\psi : U \to X$ such that for any member C of $\mathcal{K}_x$, $x \in U$, $\psi(C \cap U)$ is contained in a minimal rational curve of X, then there exists* $\Gamma \in \mathrm{Aut}(X)$ *such that* $\psi = \Gamma|_U$.

Let X be a Fano manifold of Picard number 1 and let $\phi : X \to \mathbb{P}^n$ is a surjective holomorphic map sending minimal rational curves on X to lines in $\mathbb{P}^n$. [Hwang and Kim 2013, Theorem 5.4] says that if the VMRT $\mathcal{C}_x \subset \mathbb{P}T_x(X)$ is not contained in a hyperquadric, then $\phi : X \to \mathbb{P}^n$ is CF-rigid. For example, the double covering morphism $\phi : X \to \mathbb{P}^n$ branched along a smooth hypersurface of degree $2m$, with $2 \le m \le n-1$, is CF-rigid.

Next is an application of Theorem 1.1 on CF-rigidity.

Theorem 1.4. *In the setting of* Theorem 1.1, *assume that $Z \subset \mathbb{P}^N$ is a quadratic manifold* (*i.e.*, *scheme theoretically defined by quadratic equations*) *with $i(Z) \ge 4$, and $2 \le m \le i(Z) - 2$. Then $\phi : X^Z \to Z$ is CF-rigid. In other words*, *for any connected open subset* (*in classical topology*) *$U \subset X^Z$ and any biholomorphic immersion $\psi : U \to Z$ such that for any member C of $\mathcal{K}_x$, $x \in U$, the image $\psi(C \cap U) \subset Z$ is contained in a line in $\mathbb{P}^N$, there exists $\Gamma \in \mathrm{Aut}(Z)$ such that $\psi = \Gamma \circ \phi|_U$.*

The key point is that for $Z \subset \mathbb{P}^N$ in Theorem 1.4, its VMRT $\mathcal{L}_y(Z) \subset \mathbb{P}T_y(Z)$ is also a quadratic manifold for a general point $y \in Z$; see [Ionescu and Russo 2013, Theorem 2.4]. Using this observation and Theorem 1.3, we shall prove Theorem 1.4.

The organization of this paper is as follows. In Section 2, we will review some basic facts on ECO lines. In Section 3, we will study minimal rational curves on double covers of certain Fano manifolds covered by lines. Theorem 1.1 will be proved in Section 4. In the final section, we will prove Theorem 1.4 and present its applications.

2. ECO lines

The aim of this section is to give a brief review of basic facts on ECO lines which will be used in the proof of Theorem 1.1. For more details, see [Hwang and Kim 2013].

Definition 2.1. A homogeneous polynomial of degree $2m$, $m \ge 1$, in the polynomial ring $\mathbb{C}[s, t]$ with two variables s and t is an *ECO (even contact order) polynomial* if it can be written as the square of a homogeneous polynomial of degree m in $\mathbb{C}[s, t]$.

Proposition 2.2 [op. cit., Proposition 3.3]. *For each $m > 0$, there exist m unique polynomials in the variables $t_1, \dots, t_m$,*

$$A_k(t_1, \dots, t_m) \in \mathbb{C}[t_1, \dots, t_m], \quad m+1 \le k \le 2m$$

with the following properties:

(i) *$A_k(t_1, \dots, t_m)$ is weighted homogeneous of degree k with respect to* $\mathrm{wt}(t_i) = i$ *for each $i = 1, \dots, m$;*

(ii) *the polynomial in two variables* (s, t)

$$s^{2d} + a_1 s^{2d-1} t + \cdots + a_{2d-1} s t^{2d-1} + a_{2d} t^{2d}$$

is an ECO polynomial if and only if $a_k = A_k(a_1, \ldots, a_d)$ *for each* $d+1 \le k \le 2d$. *In particular, the polynomial*

$$s^{2d} + a_{d+1} s^{d-1} t^{d+1} + \cdots + a_{2d-1} s t^{2d-1} + a_{2d} t^{2d}$$

is an ECO polynomial if and only if $a_{d+1} = \cdots = a_{2d} = 0$.

Definition 2.3. Let $f(t_0, \ldots, t_N)$ be a homogeneous polynomial of degree $2m$ in variables $t_0, \ldots, t_N$. Write

$$f(1, y_1 + \lambda z_1, \ldots, y_N + \lambda z_N) = a_0^f(y; z) + a_1^f(y; z)\lambda + \cdots + a_{2m}^f(y; z)\lambda^{2m},$$

where each $a_k^f(y; z) = a_k^f(y_1, \ldots, y_N; z_1, \ldots, z_N)$ is a polynomial in $2N$ variables $y_1, \ldots, y_N, z_1, \ldots, z_N$. Let A_k be as in Proposition 2.2 and set

$$B_k^f(y; z) := \frac{a_k^f(y; z)}{a_0^f(y; z)} - A_k\left(\frac{a_1^f(y; z)}{a_0^f(y; z)}, \ldots, \frac{a_m^f(y; z)}{a_0^f(y; z)}\right).$$

We remark that for a fixed y, $a_k^f(y; z)$ is a homogeneous polynomial in variables $z_1, \ldots, z_N$ of degree k. Furthermore for a fixed y with

$$a_0^f(y; z) = f(1, y_1, \ldots, y_N) \neq 0,$$

each $B_k^f(y; z)$ is a homogeneous polynomial of degree k in variables $z_1, \ldots, z_N$.

Definition 2.4. Let $Y \subset \mathbb{P}^N$ be a hypersurface of even degree $2m$. A line $\ell \subset \mathbb{P}^N$ is called an *ECO (even contact order) line* with respect to Y if $\ell \not\subset Y$ and the local intersection number at each point of $\ell \cap Y$ is even. For each point $y \in \mathbb{P}^N \setminus Y$, we denote by $\mathcal{E}_y^Y \subset \mathbb{P}T_y(\mathbb{P}^N)$ the space of tangent directions of ECO lines with respect to Y passing through y.

Proposition 2.5 [Hwang and Kim 2013, Proposition 3.8]. *Choose a homogeneous coordinate system* $t_0, \ldots, t_N$ *on* $\mathbb{P}^N$. *We denote by* $\mathbb{P}_\infty^{N-1} \subset \mathbb{P}^N$ *the hyperplane defined by* $t_0 = 0$ *and choose a homogeneous coordinate system* $z_1, \ldots, z_N$ *on* $\mathbb{P}_\infty^{N-1}$ *given by the restrictions of* $t_1, \ldots, t_N$ *respectively. Let* $f(t_0, \ldots, t_N)$ *be a homogeneous polynomial of degree* $2m$ *and let* $Y \subset \mathbb{P}^N$ *be its associated hypersurface. For each point* $y = [1 : y_1 : \cdots : y_N] \in \mathbb{P}^N \setminus Y \cup \mathbb{P}_\infty^{N-1}$, *define the projective isomorphism*

$$\upsilon_y : \mathbb{P}_\infty^{N-1} \to \mathbb{P}T_y(\mathbb{P}^N)$$

by sending $[z_1 : \cdots : z_N] \in \mathbb{P}_\infty^{N-1}$ *to the tangent direction of the line*

$$\{(y_1 + \lambda z_1, \ldots, y_N + \lambda z_N) \mid \lambda \in \mathbb{C}\}$$

at the point y. *Then the variety* $\upsilon_y^{-1}(\mathcal{E}_y^Y) \subset \mathbb{P}_\infty^{N-1}$ *is set-theoretically the intersection of* m *hypersurfaces in* $\mathbb{P}_\infty^{N-1}$ *defined by polynomials in* $\{B_k^f(y; z) \mid m+1 \leq k \leq 2m\}$.

3. Minimal rational curves on double covers of prime Fano manifolds

Definition 3.1. Let $Z \subset \mathbb{P}^N$ be a Fano manifold whose Picard group is generated by the hyperplane section class H. Assume that Z is covered by lines and $i(Z) \geq 3$. Let $Y \subset \mathbb{P}^N$ be a hypersurface of degree $2m$, $1 \leq m \leq i(Z) - 2$, defining smooth hypersurface section $B := Y \cap Z \subset Z$. Let

$$\phi : X^Z \to Z$$

be a double cover branched along B. From the adjunction formula

$$K_{X^Z} = \phi^*\big(K_Z + \tfrac{1}{2}(B)\big) = \phi^*\big((-i(Z)+m)H\big),$$

X^Z is a Fano manifold of index $i(X^Z) = i(Z) - m$ and its Picard group is generated by $\phi^*(H)$.

Proposition 3.2. *In the setting of* Definition 3.1, *an irreducible reduced curve* $C \subset X^Z$ *with* $\phi(C) \not\subset B$ *is a minimal rational curve if and only if its image curve* $\phi(C)$ *is an ECO line with respect to* Y. *Moreover for any minimal rational curve* $C \subset X^Z$ *with* $\phi(C) \not\subset B$, $\phi|_C : C \to \phi(C)$ *is an isomorphism.*

Proof. We fist observe:

Claim: An irreducible reduced curve $C \subset X^Z$ with $\phi(C) \not\subset B$ has ϕ^*H-degree 1 if and only if its image curve $\phi(C) \subset Z$ is an ECO line with respect to Y. Moreover for any ϕ^*H-degree 1 curve $C \subset X^Z$ with $\phi(C) \not\subset B$, $\phi|_C : C \to \phi(C)$ is an isomorphism.

Proof of the claim. Let $C \subset X^Z$ be an irreducible reduced curve such that the image $\phi(C)$ is an ECO line with respect to Y. Suppose that $\phi|_C : C \to \phi(C)$ is not birational. For a point $z \in \phi(C) \cap Y$, let t be a local uniformizing parameter on $\phi(C)$ at z and let r_z be the local intersection number of $\phi(C)$ and Y at z. Then C is analytically defined by the equation $s^2 = t^{r_z}$. Since r_z is even for any choice of $z \in \phi(C) \cap Y$, the composition of the normalization morphism $\tilde{C} \to C$ and the covering morphism $\phi|_C : C \to \phi(C)$ induces a morphism $\tilde{C} \to \phi(C)$ of degree 2 without ramification point, a contradiction. Thus $\phi|_C : C \to \phi(C)$ is birational and C has ϕ^*H-degree 1.

Conversely, if C is an irreducible reduced curve of ϕ^*H-degree 1, then $\phi(C) \subset Z$ with $\phi(C) \not\subset B$ and $\phi|_C : C \to \phi(C)$ must be birational. Thus $\phi^{-1}(\phi(C))$ has an irreducible component C' different from C with $\phi(C \cap C') = \phi(C) \cap B$. By the same argument as before, if the local intersection number r_z at $z \in \phi(C) \cap Y$ is odd, the germ of $\phi^{-1}(\phi(C))$ over z, defined by $s^2 = t^{r_z}$, is irreducible, a contradiction. Thus

r_z is even for all $z \in \phi(C) \cap Y$ which implies that $\phi(C)$ is an ECO line. Moreover, C must be smooth and the morphism $\phi|_C : C \to \phi(C)$ is an isomorphism. □

Let us go back to the proof of our proposition. By the claim above, we only need to show that for a general point $x \in X^Z$, there exists a ϕ^*H-degree 1 curve through x. From $\dim \mathcal{L}_{\phi(x)}(Z) = i(Z) - 2 \geq m$ and Proposition 2.5, it is induced that there exists an ECO line ℓ with respect to Y through $\phi(x)$ and contained in Z. Take one such ECO line ℓ. The claim above shows that the inverse image $\phi^{-1}(\ell)$ consists of two smooth rational curves of degree 1 with respect to ϕ^*H. Clearly, one of those two curves passes through x. □

4. Defining equations of VMRT

In order to find the defining ideal of the VMRT $\mathcal{C}_x \subset \mathbb{P}T_x(X^Z)$, we proceed in a manner analogous to [Hwang and Kim 2013].

Notation 4.1. Let $Y \subset \mathbb{P}^N$ be a hypersurface of even degree $2m$, $m \geq 1$, and let $Z \subset \mathbb{P}^N$ be a projective submanifold which is not contained in Y. For each point $y \in Z \setminus Z \cap Y$, we denote by $\mathcal{E}^Y_y(Z) \subset \mathbb{P}T_y(Z)$ the space of tangent directions of ECO lines with respect to Y contained in Z.

Proposition 4.2. *In the setting of* Definition 3.1, *for a general point $x \in X^Z$, the tangent morphism $\tau_x : \mathcal{K}_x \to \mathbb{P}T_x(X^Z)$, sending each member of $\mathcal{K}_x$ to its tangent direction, is an embedding. In particular the VMRT $\mathcal{C}_x = \mathrm{Im}(\tau_x) \subset \mathbb{P}T_x(X^Z)$ is a nonsingular projective variety with finitely many components of dimension $i(Z) - m - 2$, isomorphic to $\mathcal{E}^Y_{\phi(x)}(Z) \subset \mathbb{P}T_{\phi(x)}(Z)$ via the differential morphism $d\phi_x : \mathbb{P}T_x(X^Z) \to \mathbb{P}T_{\phi(x)}(Z)$.*

Proof. From Proposition 3.2, the differential $d\phi_x : \mathbb{P}T_x(X^Z) \to \mathbb{P}T_{\phi(x)}(Z)$ sends the VMRT $\mathcal{C}_x \subset \mathbb{P}T_x(X^Z)$ isomorphically to the variety $\mathcal{E}^Y_{\phi(x)}(Z) \subset \mathbb{P}T_{\phi(x)}(Z)$.

We note that $\tau_x : \mathcal{K}_x \to \mathbb{P}T_x(X^Z)$ is the normalization morphism of its image, which is equal to $\mathcal{C}_x$. Thus we only need to show that τ_x is an embedding because $\mathcal{K}_x$ is a smooth projective variety of dimension $i(X) - 2 = i(Z) - m - 2$.

Assume that there are two distinct members C_1 and C_2 of $\mathcal{K}_x$ such that $\tau_x([C_1]) = \tau_x([C_2])$. Thus $\phi(C_1)$ and $\phi(C_2)$ are lines on $\mathbb{P}^N$ passing through $\phi(x)$ with the same tangent direction at $\phi(x)$, which implies that $\phi(C_1) = \phi(C_2)$ is a line; denote it by ℓ. Therefore $\phi^{-1}(\ell) = C_1 \cup C_2$, and hence C_2 and C_2 meets only over the points on $\ell \cap B$, a contradiction because $x \in C_1 \cap C_2$ but $\phi(x) \notin B$ by the general condition on x. Thus we have shown that τ_x is injective.

Since we know that $\mathcal{K}_x$ is nonsingular, to prove that τ_x is an embedding, it remains to show that τ_x is an immersion. By [Hwang 2001, Proposition 1.4], this is equivalent to showing that for any member $C \subset X^Z$ of $\mathcal{K}_x$, the normal bundle N_{C/X^Z} satisfies

$$N_{C/X^Z} = O_{\mathbb{P}^1}(1)^{i(Z)-m-2} \oplus \mathcal{O}_{\mathbb{P}^1}^{n+m-i(Z)+1}.$$

By the generality of x, we can write

$$N_{C/X^Z} = \mathcal{O}_{\mathbb{P}^1}(a_1) \oplus \cdots \oplus \mathcal{O}_{\mathbb{P}^1}(a_{n-1})$$

for integers $a_1 \geq \cdots \geq a_{n-1} \geq 0$ satisfying $\sum_i a_i = i(Z) - m - 2$. Since ϕ is unramified at general points of C and $\phi|_C : C \to \ell := \phi(C)$ is an isomorphism, we have an injective sheaf homomorphism

$$\phi_* : N_{C/X^Z} \to N_{\ell/\mathbb{P}^N} = \mathcal{O}(1)^{N-1}.$$

Thus $a_1 \leq 1$; hence, $a_1 = \cdots = a_{i(Z)-m-2} = 1$ and $a_{i(Z)-m-1} = \cdots = a_{n-1} = 0$. $\square$

Proposition 4.3. *Let $Z \subset \mathbb{P}^N$ be a Fano manifold whose Picard group is generated by the hyperplane section class. Assume that Z is covered by lines and $i(Z) \geq 3$. Then for each m with $1 \leq m \leq i(Z) - 2$, there exists a hypersurface Y of degree $2m$ with smooth $Y \cap Z$ such that for a general point $y \in Z$, $\mathcal{E}^Y_y(Z) \subset \mathbb{P}T_y(Z)$ is scheme-theoretically the intersection of $\mathcal{L}_y(Z) \subset \mathbb{P}T_y(Z)$ with m hypersurfaces in $\mathbb{P}T_y(Z)$ of degrees $m+1, \ldots, 2m$, respectively.*

Proof. Take a point $\hat{y} \in Z$ such that $\mathcal{L}_{\hat{y}}(Z)$ is smooth of dimension $i(Z)-2$. Choose a homogeneous coordinate system $t_0, \ldots, t_N$ so that $\hat{y} = [1 : 0 : \cdots : 0] \in Z$ and the hyperplane section $Z \cap (t_0 = 0)$ is smooth. Choose homogeneous polynomials $\{b_k(t_1, \ldots, t_N) \mid m+1 \leq k \leq 2m\}$ with $\deg b_k = k$ so that each of the following is smooth:

(i) the intersection of $Z \cap (t_0 \neq 0)$ with the hypersurface in $\mathbb{P}^N$ defined by

$$1 + b_{m+1}(t_1, \ldots, t_N) + \cdots + b_{2m}(t_1, \ldots, t_N) = 0,$$

(ii) the intersection of $Z \cap (t_0 = 0)$ with the hypersurface in $\mathbb{P}^N$ defined by

$$b_{2m}(t_1, \ldots, t_N) = 0.$$

Set

$$\begin{aligned} f(t_0, t_1, \ldots, t_N) := t_0^{2m} + t_0^{m-1} b_{m+1}(t_1, \ldots, t_N) \\ + \cdots + t_0 b_{2m-1}(t_1, \ldots, t_N) + b_{2m}(t_1, \ldots, t_N). \end{aligned}$$

The assumptions (i) and (ii) imply that the hypersurface section of Z defined by $f(t_0, \ldots, t_N)$ is smooth. From Proposition 2.2(ii) we obtain the equalities

$$B^f_k(y; z) = b_k(z_1, \ldots, z_N), \quad m+1 \leq k \leq 2m.$$

Since $\upsilon_y^{-1}(\mathcal{L}_y(Z))$ is smooth of dimension $i(Z) - 2 \geq m$, it follows that for general $\{b_k(z_1, \ldots, z_N) \mid m+1 \leq k \leq 2m\}$, the scheme-theoretical intersection of $\upsilon_y^{-1}(\mathcal{L}_y(Z))$ and the m hypersurfaces defined by $B^f_k(y; z)$, $m+1 \leq k \leq 2m$, is smooth of dimension $i(Z) - m - 2$. Therefore our proposition comes from Proposition 2.5. □

Proof of Theorem 1.1. By Propositions 4.2 and 4.3, the theorem holds for a general hypersurface $Y \subset \mathbb{P}^N$ with smooth intersection $Y \cap Z$. In order to prove it for arbitrary hypersurface $Y \subset \mathbb{P}^N$ with smooth intersection $Y \cap Z$, choose a deformation $\{Y_t \mid |t| < \epsilon\}$ of $Y = Y_0$ with smooth $Y_t \cap Z$ such that for a Zariski open subset $U_t \subset Z \setminus Y_t \cap Z$, the varieties $\mathcal{E}_y^{Y_t}(Z) \subset \mathbb{P}T_y(Z)$, $y \in U_t$, are

(i) smooth of dimension $i(Z) - m - 2$ for any t, and

(ii) the intersection of m hypersurface sections of $\mathcal{L}_y(Z) \subset \mathbb{P}T_y(Z)$ defined by hypersurfaces in $\mathbb{P}T_y(Z)$ of degrees $m + 1, \ldots, 2m$, respectively. (The intersection is scheme-theoretic for any $t \neq 0$ and set-theoretic for $t = 0$.)

By shrinking ϵ if necessary, the intersection $\bigcap_t U_t$ is nonempty. Let V be a Zariski open subset of Z such that the variety $\mathcal{L}_y(Z)$ is smooth of dimension $i(Z) - 2$. Pick a point $y \in \left(\bigcap_t U_t\right) \cap V$. We can construct a smooth family $\{\phi_t : X_t^Z \to Z \mid |t| < \epsilon\}$ of double covers of Z branched along the $Z \cap Y_t$. Choose $x_t \in \phi_t^{-1}(y)$ in a continuous way. The family $\{\mathcal{K}_{x_t} \mid |t| < \epsilon\}$ is a flat family of nonsingular projective subvarieties; see, e.g., [Kollár 1996, II.3.11.5]. Via Proposition 4.2, this implies that $\{\mathcal{E}_y^{Y_t}(Z) \mid |t| < \epsilon\}$ is a flat family of nonsingular projective subvarieties of $\mathcal{L}_y(Z) \subset \mathbb{P}T_y(Z)$. From condition (ii) and the flatness, we conclude that $\mathcal{E}_y^{Y_0}(Z)$ is also scheme-theoretically the intersection of $\mathcal{L}_y(Z) \subset \mathbb{P}T_y(Z)$ with m hypersurfaces in $\mathbb{P}T_y(Z)$ of degrees $m + 1, \ldots, 2m$, respectively. □

5. Rigidity and Extension

The following fact is obvious, but plays an important role in the proof of Theorem 1.4.

Lemma 5.1. *Let R be the polynomial ring $\mathbb{C}[z_1, \ldots, z_N]$ in variables $z_1, \ldots, z_N$. Consider R as a graded ring with* $\deg z_i = 1$. *Let I, J, and K be homogeneous ideals of R such that I and K are generated by homogeneous polynomials of degree* 2, *and J is generated by homogeneous polynomials of degrees* ≥ 3. *If $h : R \to R$ is an automorphism of the graded ring R with $h(K) \subset I + J$, then $h(K) \subset I$.*

Proof of Theorem 1.4. By shrinking U if necessary, we may assume that

- the restriction $\phi|_U : U \to Z$ is an embedding,
- for any $x \in U$, $\mathcal{L}_{\phi(x)}(Z) \subset \mathbb{P}T_{\phi(x)}(Z)$ is a smooth quadratic manifold of dimension $i(Z) - 2$; see [Ionescu and Russo 2013, Theorem 2.4],
- for any $x \in U$, the space $\mathcal{E}_{\phi(x)}^Y \subset \mathbb{P}T_{\phi(x)}(Z)$ is scheme-theoretically the intersection of $\mathcal{L}_{\phi(x)}(Z) \subset \mathbb{P}T_{\phi(x)}(Z)$ and m hypersurfaces in $\mathbb{P}T_{\phi(x)}(Z)$ of degrees $m + 1, \ldots, 2m$, respectively.

Set $U_1 = \phi(U)$ and $U_2 = \psi(U)$. Then we get a biholomorphism

$$\gamma := \psi \circ \phi|_U^{-1} : U_1 \to U_2.$$

We note that for any $y \in U_1$, the differential $d\gamma_y : \mathbb{P}T_y(Z) \to \mathbb{P}T_{\gamma(y)}(Z)$ is an isomorphism sending $\mathcal{E}_y^Y(Z) \subset \mathbb{P}T_y(Z)$ into $\mathcal{L}_{\gamma(y)}(Z) \subset \mathbb{P}T_{\gamma(y)}(Z)$. From Lemma 5.1 and Theorem 1.1, it follows that $d_y\gamma(\mathcal{L}_y(Z)) \subset \mathcal{L}_{\gamma(y)}(Z)$. By shrinking U if necessary, we may assume that $\mathcal{L}_{\gamma(y)} \subset \mathbb{P}T_{\gamma(y)}(Z)$ is also a smooth quadratic manifold of dimension $i(Z)-2$. Therefore it follows that $d_y\gamma(\mathcal{L}_y(Z)) = \mathcal{L}_{\gamma(y)}(Z)$. We finish the proof by applying [Hwang 2001, Theorem 3.2]. $\square$

The next corollary is an algebraic version of Theorem 1.4

Corollary 5.2. *In the setting of* Theorem 1.4, *let $\hat{X}$ be a projective variety with generically finite surjective morphisms $g : \hat{X} \to X^Z$ and $h : \hat{X} \to Z$ such that for a minimal rational curve C through a general point of X^Z, there exists an irreducible component C' of $g^{-1}(C)$ whose image $h(C') \subset Z \subset \mathbb{P}^N$ is a line. Then there exists an automorphism $\Gamma : Z \to Z$ such that $h = \Gamma \circ \phi \circ g$.*

Next, Theorems 5.3 and 5.4 can be proved by the same arguments as in the proof of Theorems 1.7 and 1.9 in [Hwang and Kim 2013], respectively. We include their proof for the reader's convenience.

Theorem 5.3. *Let $Z \subset \mathbb{P}^N$ be a quadratic Fano manifold such that its Picard group is generated by the hyperplane section class and $i(Z) \geq 4$. Let $Y_1, Y_2 \subset \mathbb{P}^N$, $N \geq 3$, be two hypersurfaces of degree $2(i(Z)-2)$ with smooth intersections $Y_1 \cap Z$ and $Y_2 \cap Z$. Let $\phi_1 : X_1 \to Z$ and $\phi_2 : X_2 \to Z$ be double covers of Z branched along $Y_1 \cap Z$ and $Y_2 \cap Z$, respectively. Suppose there exists a finite morphism $f : X_1 \to X_2$. Then f is an isomorphism.*

Proof. Put $m = i(Z)-2$ in the proof of Proposition 4.2. Then minimal rational curves on X_i, $i = 1, 2$, have trivial normal bundles and rational curves through general points with trivial normal bundles are minimal rational curves. By [Hwang and Mok 2003, Proposition 6], for a general minimal rational curve $C \subset X_2$, each irreducible component of $f^{-1}(C)$ is a minimal rational curve in X_1. In other words, f sends minimal rational curves of X_1 through a general point to those of X_2. Putting

$$\hat{X} = X_1, \quad X = X_2, \quad g = f, \quad \phi = \phi_2, \quad \text{and} \quad h = \phi_1$$

in Corollary 5.2, we see that $\phi_1 = \Gamma \circ \phi_2 \circ f$ for some automorphism Γ of Z. Thus f must be birational, and hence an isomorphism. $\square$

The next theorem is a stronger version of Theorem 1.3.

Theorem 5.4. *Let $Z \subset \mathbb{P}^N$ be a quadratic Fano manifold such that its Picard group is generated by the hyperplane section class and $i(Z) \geq 4$. Let $Y_1, Y_2 \subset \mathbb{P}^N$ be two* *hypersurfaces of degree* $2m$, $2 \le m \le i(Z)-2$, *with smooth* $Y_1 \cap Z$ *and* $Y_2 \cap Z$. *Let* $\phi_1 : X_1^Z \to Z$ *and* $\phi_2 : X_2^Z \to Z$ *be double covers of* Z *branched along* $Y_1 \cap Z$ *and* $Y_2 \cap Z$ *respectively, and let* $U_1 \subset X_1^Z$ *and* $U_2 \subset X_2^Z$ *be two connected open subsets. Suppose that we are given a biholomorphic map* $\gamma : U_1 \to U_2$ *such that for any minimal rational curve* $C_1 \subset X_1^Z$, *there exists a minimal rational curve* $C_2 \subset X_2^Z$ *with* $\gamma(U_1 \cap C_1) = U_2 \cap C_2$. *Then we can find a biregular morphism* $\Gamma : X_1^Z \to X_2^Z$ *with* $\Gamma|_{U_1} = \gamma$.

Sketch of the proof. Applying Theorem 1.4 to $\psi := \phi_2 \circ \gamma : U_1 \to \phi_2(U_2) \subset Z$ and $\phi := \phi_1$, we have $\Gamma' \in \mathrm{Aut}(Z)$ such that $\Gamma' \circ \phi_1|_{U_1} = \phi_2 \circ \gamma$. By the assumption on γ and Proposition 3.2, for a general point $y \in \phi_1(U_1)$, we have $d\Gamma'(\mathcal{E}_y^{Y_1}) = \mathcal{E}_{\Gamma'(y)}^{Y_2}$, which implies that a general ECO line with respect to Y_2 contained in Z should be an ECO line with respect to Y'.

Since the Picard group of Z is generated by the hyperplane section class and $\Gamma' \in \mathrm{Aut}(Z)$, there exists a hypersurface $Y' \subset \mathbb{P}^N$ of degree $2m$ such that $\Gamma'(Y_1 \cap Z) = Y' \cap Z$. Suppose $Y' \cap Z \neq Y_2 \cap Z$. By the similar arguments in [Hwang and Kim 2013, Proposition 2.5], we can show that a general ECO line with respect to Y_2 contained in Z cannot be an ECO line with respect to Y', a contradiction. Therefore $Y' \cap Z = Y_2 \cap Z$.

Thus replacing $Y_1 \cap Z$ by $\Gamma(Y_1 \cap Z)$ and ϕ_1 by $\Gamma' \circ \phi_1$, we may assume that $Y_1 \cap Z = Y_2 \cap Z$ and $\phi_1(U_1) = \phi_2(U_2)$. By the uniqueness of double covering, it follows that there exists a biregular morphism $\Gamma : X_1^Z \to X_2^Z$ with $\Phi|_{U_1} = \gamma$. □

Acknowledgments

The author would like to thank Jun-Muk Hwang for a number of discussions on this subject. She is also grateful to the referee for valuable comments to modify the original version.

References

[Hwang 2001] J.-M. Hwang, "Geometry of minimal rational curves on Fano manifolds", pp. 335–393 in *School on Vanishing Theorems and Effective Results in Algebraic Geometry* (Trieste, 2000), edited by J. P. Demailly et al., ICTP Lecture Notes **6**, Abdus Salam Int. Cent. Theoret. Phys., Trieste, 2001. MR 2003g:14054 Zbl 1086.14506

[Hwang and Kim 2013] J.-M. Hwang and H. Kim, "Varieties of minimal rational tangents on double covers of projective space", *Math. Z.* **275**:1–2 (2013), 109–125. MR 3101799 Zbl 1282.14070

[Hwang and Mok 2001] J.-M. Hwang and N. Mok, "Cartan–Fubini type extension of holomorphic maps for Fano manifolds of Picard number 1", *J. Math. Pures Appl.* (9) **80**:6 (2001), 563–575. MR 2002j:32023 Zbl 1033.32013

[Hwang and Mok 2003] J.-M. Hwang and N. Mok, "Finite morphisms onto Fano manifolds of Picard number 1 which have rational curves with trivial normal bundles", *J. Algebraic Geom.* **12**:4 (2003), 627–651. MR 2005f:14086 Zbl 1038.14018

[Hwang and Mok 2004] J.-M. Hwang and N. Mok, “Birationality of the tangent map for minimal rational curves”, *Asian J. Math.* **8**:1 (2004), 51–63. MR 2006c:14065 Zbl 1072.14015

[Ionescu and Russo 2013] P. Ionescu and F. Russo, “Manifolds covered by lines and the Hartshorne conjecture for quadratic manifolds”, *Amer. J. Math.* **135**:2 (2013), 349–360. MR 3038714 Zbl 1271.14055

[Kebekus 2002] S. Kebekus, “Families of singular rational curves”, *J. Algebraic Geom.* **11**:2 (2002), 245–256. MR 2003e:14011 Zbl 1054.14035

[Kollár 1996] J. Kollár, *Rational curves on algebraic varieties*, Ergebnisse der Mathematik (3) **32**, Springer, Berlin, 1996. MR 98c:14001

Received November 27, 2014. Revised August 22, 2015.

HOSUNG KIM
NATIONAL INSTITUTE FOR MATHEMATICAL SCIENCE
DAEJEON 34047
REPUBLIC OF KOREA
hosung@nims.re.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.341

ON THE UNIFORM SQUEEZING PROPERTY OF BOUNDED CONVEX DOMAINS IN $\mathbb{C}^n$

KANG-TAE KIM AND LIYOU ZHANG

We prove that the bounded convex domains and the C^2-smoothly bounded strongly pseudoconvex domains in $\mathbb{C}^n$ admit the uniform squeezing property. Moreover, we prove by the scaling method that the squeezing function approaches 1 near the strongly pseudoconvex boundary points.

1. Introduction

The notion of *holomorphic homogeneous regular*, or equivalently, *uniformly squeezing* for complex manifolds has been introduced in [Liu et al. 2004; 2005] and [Yeung 2009], respectively. This concept is essential for the estimation of several invariant metrics. See the above cited papers for details.

Let Ω be a complex manifold of dimension n. The *squeezing function* $\sigma_\Omega : \Omega \to \mathbb{R}$ of Ω is defined in [Deng et al. 2012] as follows. For each $p \in \Omega$, let

$$\mathcal{F}(p, \Omega) := \{f : \Omega \to \mathbb{B}^n : f \text{ is 1-1 holomorphic, } f(p) = 0\},$$

where

- $\mathbb{B}^n(p; r) = \{z \in \mathbb{C}^n : \|z - p\| < r\}$, and
- $\mathbb{B}^n = \mathbb{B}^n(\mathbf{0}; 1) = \mathbb{B}^n((0, \ldots, 0); 1)$.

Then define

$$\sigma_\Omega(p) = \sup\{r : \mathbb{B}^n(\mathbf{0}; r) \subset f(\Omega) \text{ for some } f \in \mathcal{F}(p, \Omega)\}.$$

Furthermore, the *squeezing constant* $\hat{\sigma}_\Omega$ for Ω is defined by

$$\hat{\sigma}_\Omega := \inf_{p \in \Omega} \sigma_\Omega(p).$$

Definition ([Liu et al. 2004; 2005; Yeung 2009]). A complex manifold Ω is called *holomorphic homogeneous regular* (HHR), or equivalently *uniformly squeezing*, if $\hat{\sigma}_\Omega > 0$.

MSC2010: primary 32F45; secondary 32A25.
Keywords: squeezing function, convex domains, scaling method.

Notice that the HHR property is preserved by biholomorphisms. The squeezing function and squeezing constants are also biholomorphic invariants [Deng et al. 2012].

These concepts have been developed for the study of completeness and other geometric properties such as the metric equivalence of the invariant metrics, including the Carathéodory, Kobayashi–Royden, Teichmüller, Bergman, and Kähler–Einstein metrics. It is obvious that the examples of HHR manifolds include bounded homogeneous domains. In case the manifold is biholomorphic to a bounded domain and the holomorphic automorphism orbits accumulate at every boundary point, such as in the case of the Bers embedding of the Teichmüller space, again the HHR property holds. A somewhat less obvious example are the bounded strongly convex domains (as the majority of them do not possess any holomorphic automorphisms except the identity map), proved by S.-K. Yeung [2009]. But there, some of the most standard examples, such as the bounded convex domains and the bounded strongly pseudoconvex domains, were left untouched.

Indeed, the starting point of this article is to show:

Theorem 1.1. *All bounded convex domains in* $\mathbb{C}^n$ $(n \geq 1)$ *are HHR.*

Note that we do not assume any additional conditions such as boundary smoothness or "finite type" in the sense of D'Angelo in the above theorem. Nevertheless, the concept of squeezing function σ_Ω defined above plays an important role, and moreover, it appeals to us that further investigations of this function would be worthwhile. One immediate observation is that if $\sigma_\Omega(p) = 1$ for some $p \in \Omega$, then Ω is biholomorphic to the unit open ball. In light of studies on the asymptotic behavior of several invariant metrics of strongly pseudoconvex domains, perhaps it is natural to ask, for a bounded strongly pseudoconvex domain Ω in $\mathbb{C}^n$, whether $\lim_{\Omega \ni q \to p} \sigma_\Omega(q) = 1$ holds for every boundary point $p \in \partial\Omega$.

It was proved in [Deng et al. 2015] that the HHR property holds for all bounded strongly pseudoconvex domains, using an improvement of the method in [Fridman and Ma 1995]. In the present paper, by using a different approach — the scaling method — we will prove:

Theorem 1.2. *If* Ω *is a bounded domain in* $\mathbb{C}^n$ *with a* C^2 *strongly convex boundary, then* $\lim_{\Omega \ni q \to p} \sigma_\Omega(q) = 1$ *for every* $p \in \partial\Omega$.

Actually, we have a more general conclusion in Theorem 3.1, which implies Theorem 1.2. The question posed above follows quickly from Theorem 3.1 and the following remarkable theorem of Diederich, Fornaess, and Wold [2014].

Theorem 1.3 [Diederich et al. 2014, Theorem 1.1]. *Let* $\Omega \subset \mathbb{C}^n$ *be a bounded domain which is locally convexifiable and has finite type* $2k$ *near a point* $p \in \partial\Omega$. *Assume further that* $\partial\Omega$ *is* C^∞*-smooth near* p, *and that* $\overline{\Omega}$ *has a Stein neighborhood* *basis. Then there exists a holomorphic embedding* $f : \overline{\Omega} \to \overline{B}^n_k$, *where*

$$B^n_k = \{z \in \mathbb{C}^2 : |z_n|^2 + |z'|^{2k} < 1\},$$

such that $f(p) = (0, \dots, 0, 1)$ *and* $\{z \in \overline{\Omega} : f(z) \in \partial B^n_k\} = \{p\}$.

In particular, if $\partial\Omega$ *is strongly pseudoconvex near* p (*i.e.*, $k = 1$), *it is enough to assume that* $\partial\Omega$ *is* C^2*-smooth near* p.

We mention here that the proof of Theorem 1.2 is of interest in its own right, and also clarifies and simplifies some previously known theorems. These are mentioned in the final section.

2. Bounded convex domains are HHR manifolds

The aim of this section is to establish Theorem 1.1 stated above. Not only does this theorem cover the case left untreated in [Yeung 2009], but our method is different (see also [Deng et al. 2012] on this matter). Our method uses a version of the "scaling method in several complex variables" initiated by S. Pinchuk [1991]. In fact, we use the version presented in [Kim 1992], modified for the purpose of studying the asymptotic boundary behavior of holomorphic invariants.

Proof of Theorem 1.1. We proceed in 5 steps.

Step 1. *Setup.* Let Ω be a bounded convex domain in $\mathbb{C}^n$. Suppose that Ω is not HHR. Then there exists a sequence $\{q_j\}$ in Ω converging to a boundary point, say $q \in \partial\Omega$, such that

$$\lim_{j\to\infty} \sigma_\Omega(q_j) = 0.$$

Needless to say, it suffices to show that such a sequence cannot exist.

Step 2. *The* j*-th orthonormal frame.* Let $\langle\, ,\, \rangle$ represent the standard Hermitian inner product of $\mathbb{C}^n$, and let $\|v\| = \sqrt{\langle v, v\rangle}$. For every $q \in \mathbb{C}^n$ and complex linear subspace V of $\mathbb{C}^n$, denote by

$$B^V(q, r) = \{p \in \mathbb{C}^n : p - q \in V \text{ and } \|p - q\| < r\}.$$

Now let $q \in \Omega$ and define the positive number $\lambda(q, V)$ by

$$\lambda(q, V) = \max\{r > 0 : B^V(q, r) \subset \Omega\}.$$

This number is finite for each (q, V), whenever $\dim V > 0$, since Ω is Kobayashi hyperbolic.

Fix the index j momentarily. Then we choose an orthonormal basis for $\mathbb{C}^n$, with respect to the standard Hermitian inner product $\langle\, ,\, \rangle$. First consider

$$\lambda^1_j := \lambda(q_j, \mathbb{C}^n).$$

There exists $q_j^{1*} \in \partial\Omega$ such that $\|q_j^{1*} - q_j\| = \lambda_j^1$. Let

$$e_j^1 = \frac{q_j^{1*} - q_j}{\|q_j^{1*} - q_j\|}.$$

Then consider the complex span $\mathrm{Span}_{\mathbb{C}}\{e_j^1\}$, and let V^1 be its orthogonal complement in $\mathbb{C}^n$. Take

$$\lambda_j^2 := \lambda(q_j, V^1)$$

and $q_j^{2*} \in \partial\Omega$ such that $q_j^{2*} - q_j \in V^1$ and $\|q_j^{2*} - q_j\| = \lambda_j^2$. Then let

$$e_j^2 := \frac{q_j^{2*} - q_j}{\|q_j^{2*} - q_j\|}.$$

With $e_j^1, e_j^2, \dots, e_j^{\ell *}$ and $\lambda_j^1, \lambda_j^2, \dots, \lambda_j^\ell$ chosen, the next elements $e_j^{\ell+1*}$ and $\lambda_j^{\ell+1}$ are selected as follows. Denote by V^ℓ the complex orthogonal complement of $\mathrm{Span}_{\mathbb{C}}\{e_j^1, e_j^2, \dots, e_j^\ell\}$. Then

$$\lambda_j^{\ell+1} := \lambda(q_j, V^\ell)$$

and $q_j^{\ell+1*} \in \partial\Omega$ are such that $q_j^{\ell+1*} - q_j \in V^\ell$ and $\|q_j^{\ell+1*} - q_j\| = \lambda_j^{\ell+1}$. Let

$$e_j^{\ell+1} := \frac{q_j^{\ell+1*} - q_j}{\|q_j^{\ell+1*} - q_j\|}.$$

By induction, this process yields an orthonormal set $e_j^1, \dots, e_j^n$ for $\mathbb{C}^n$ and the positive numbers $\lambda_j^1, \dots, \lambda_j^n$.

Step 3. *Stretching complex linear maps*. Let $\hat{e}^1, \dots, \hat{e}^n$ denote the standard orthonormal basis for $\mathbb{C}^n$, i.e.,

$$\hat{e}^1 = (1, 0, \dots, 0),\ \hat{e}^2 = (0, 1, 0, \dots, 0),\ \dots,\ \hat{e}^n = (0, \dots, 0, 1).$$

Define the *stretching linear map* $L_j : \mathbb{C}^n \to \mathbb{C}^n$ by

$$L_j(z) = \sum_{k=1}^n \frac{\langle z - q_j, e_j^k \rangle}{\lambda_j^k}\, \hat{e}^k$$

for every $z \in \mathbb{C}^n$. Note that for each j, L_j maps Ω biholomorphically onto its image.

Step 4. *Supporting hyperplanes*. Notice that

$$L_j(q_j) = \mathbf{0} = (0, \dots, 0),\ L_j(q_j^{1*}) = \hat{e}^1,\ \dots,\ L_j(q_j^{n*}) = \hat{e}^n.$$

We shall consider the supporting hyperplanes, say Π_j^k for $k = 1, \dots, n$, of $L_j(\Omega)$ at points $L_j(q_j^{k*})$ for $k = 1, \dots, n$, respectively.

Substep 4.1. The supporting hyperplane Π_j^1. As noted above, $L_j(q_j^{1*}) = \hat{e}^1$. Due to the choice of q_j^{1*}, the supporting hyperplane of Ω at q_j^{1*} must also support the sphere tangent to the boundary $\partial\Omega$. Consequently the supporting hyperplane Π_j^1 of $L_j(\Omega)$ must support a smooth surface (an ellipsoid) tangent to $L_j(\partial\Omega)$ at $\hat{e}^1$. Thus, the equation for this hyperplane Π_j^1 is

$$\operatorname{Re}(z_1 - 1) = 0$$

(independently of j, being perpendicular to $\hat{e}^1$). We also note that

$$L_j(\Omega) \subset \{(z_1, \dots, z_n) \in \mathbb{C}^n : \operatorname{Re} z_1 < 1\}.$$

Substep 4.2. The rest of the supporting hyperplanes Π_j^k, *for* $k \geq 2$. First consider the case $k = 2$. Then the supporting hyperplane Π_j^2 passes through $L_j(q_j^{2*}) = \hat{e}^2$. Since the restriction of Ω to V^1 contains the sphere in V^1 tangent to the restriction of $\partial\Omega$ at the point $\hat{e}^2$, the supporting hyperplane Π_j^2 restricted to $L_j(V^1)$ takes the equation $\{(z_2, \dots, z_n) \in \mathbb{C}^n : \operatorname{Re}(z_2 - 1) = 0\}$. Hence

$$\Pi_j^2 = \{(z_1, \dots, z_n) \in \mathbb{C}^n : \operatorname{Re}(a_j^{2,1} z_1 + a_j^{2,2}(z_2 - 1)) = 0\}$$

for some $\left(a_j^{2,1}, a_j^{2,1}\right) \in \mathbb{C}^2$ with $\left|a_j^{2,1}\right|^2 + \left|a_j^{2,2}\right|^2 = 1$ and $a_j^{2,2} > 0$. We also have that

$$L_j(\Omega) \subset \{(z_1, \dots, z_n) \in \mathbb{C}^n : \operatorname{Re}(a_j^{2,1} z_1 + a_j^{2,2}(z_2 - 1)) < 0\}.$$

For $k \in \{3, \dots, n\}$, one deduces inductively that the supporting hyperplane Π_j^k passes through the point $\hat{e}^k$, and that

$$\Pi_j^k = \{(z_1, \dots, z_n) \in \mathbb{C}^n : \operatorname{Re}(a_j^{k,1} z_1 + \cdots + a_j^{k,k-1} z_{k-1} + a_j^{k,k}(z_k - 1)) = 0\},$$

with $a_j^{k,k} > 0$ and $\sum_{\ell=1}^k \left|a_j^{k,\ell}\right|^2 = 1$. Also,

$$L_j(\Omega) \subset \{(z_1, \dots, z_n) \in \mathbb{C}^n : \operatorname{Re}(a_j^{k,1} z_1 + \cdots + a_j^{k,k-1} z_{k-1} + a_j^{k,k}(z_k - 1)) < 0\}.$$

Substep 4.3. Polygonal envelopes. We add this small substep for convenience. From the discussion so far in this step, we have the j-th polygonal envelope (of $L_j(\Omega)$)

$$\begin{aligned} \Sigma_j := \{(z_1, \dots, z_n) \in \mathbb{C}^n : {} & \operatorname{Re} z_1 < 1, \\ & \operatorname{Re}(a_j^{2,1} z_1 + a_j^{2,2}(z_2 - 1)) < 0, \\ & \vdots \\ & \operatorname{Re}(a_j^{n,1} z_1 + \cdots + a_j^{n,n-1} z_{n-1} + a_j^{n,n}(z_n - 1)) < 0\}. \end{aligned}$$

Step 5. *Bounded realization.* Notice that, for every $k \in \{1, \dots, n\}$, the disc

$$D_j^k := \{z = (z_1, \dots, z_n) \in \mathbb{C}^n : \|z - q_j\| < \lambda_j^k \text{ and } \forall \ell \neq k, \langle z - q_j, e_j^\ell \rangle = 0\}$$

is contained in Ω. Hence, every $L_j(\Omega)$ contains the discs $D^k := \{\zeta \hat{e}^k : \zeta \in \mathbb{C}, |\zeta| < 1\}$ for every $k = 1, \dots, n$. Since Ω is convex and L_j is linear, $L_j(\Omega)$ is also convex. Therefore, the "unit acorn"

$$A := \{(z_1, \dots, z_n) \in \mathbb{C}^n : |z_1| + \cdots + |z_n| < 1\}$$

is contained in $L_j(\Omega)$. This restricts the unit normal vectors

$$n_j^k := (a_j^{k,1}, \dots, a_j^{k,k}, 0, \dots, 0) \in \mathbb{C}^n$$

for every $k = 2, \dots, n$. Namely, there is a positive constant $\delta > 0$ independent of j and k such that $a_j^{k,k} \geq \delta$ for every j, k.

Now taking a subsequence (of q_j), we may assume that the sequence of unit vectors $\{n_j^k\}_{j=1}^{\infty}$ converges for every $k \in \{2, \dots, n\}$. Let us write

$$\lim_{j \to \infty} n_j^k = n^k = (a^{k,1}, \dots, a^{k,k}, 0, \dots, 0)$$

for each $k = 1, 2, \dots, n$.

Consider the maps

$$B_j(z_1, \dots, z_n) = (\zeta_1, \dots, \zeta_n)$$

defined by

$$\zeta_1 = z_1,\ \zeta_2 = a_j^{2,1} z_1 + a_j^{2,2} z_2,\ \dots,\ \zeta_n = a_j^{n,1} z_1 + \cdots + a_j^{n,n} z_n.$$

Then it follows that

$$\begin{aligned} B_j \circ L_j(\Omega) &\subset B_j(\Sigma_j) \\ &= \{(\zeta_1, \dots, \zeta_n) \in \mathbb{C}^n : \operatorname{Re} \zeta_1 < 1, \operatorname{Re} \zeta_2 < a_j^{2,2}, \dots, \operatorname{Re} \zeta_n < a_j^{n,n}\}. \end{aligned}$$

Now, for each j, we consider the Cayley transformation

$$\Phi_j(z_1, \dots, z_n) = \left(\frac{z_1}{2 - z_1}, \frac{z_2}{2a_j^{2,2} - z_2}, \dots, \frac{z_n}{2a_j^{n,n} - z_n} \right).$$

Then $\Phi_j \circ B_j(\Sigma_j) \subset D^n$, where D^n denotes the unit polydisc in $\mathbb{C}^n$ centered at the origin. Also, there exists a positive constant $\delta' \in (0, \delta)$ such that $\Phi_j \circ B_j(\Sigma_j) \subset D^n$ contains the ball of radius δ' centered at the origin $\mathbf{0}$.

Since $\Phi_j \circ B_j \circ L_j(q_j) = (0, \dots, 0)$ for every j, we now conclude that the squeezing function satisfies

$$\sigma_\Omega(q_j) \geq \frac{\delta'}{\sqrt{n}}.$$

This estimate, which holds for every sequence q_j approaching the boundary, yields the desired contradiction at last. Thus the proof is complete. $\square$

3. Boundary behavior of the squeezing function on strongly convex domains

Definition. Let Ω be a domain in $\mathbb{C}^n$. A boundary point $p \in \partial\Omega$ is said to be *spherically extreme* if

(1) the boundary $\partial\Omega$ is C^2-smooth in an open neighborhood of p, and

(2) there exists a ball $\mathbb{B}^n(c(p); R)$ in $\mathbb{C}^n$ of some radius R, centered at some point $c(p)$ such that $\Omega \subset \mathbb{B}^n(c(p); R)$ and $p \in \partial\Omega \cap \partial\mathbb{B}^n(c(p); R)$.

The main goal of this section is to establish the following theorem.

Theorem 3.1. *If a domain Ω in $\mathbb{C}^n$ admits a spherically extreme boundary point p in a neighborhood of which the boundary $\partial\Omega$ is C^2-smooth, then*

$$\lim_{\Omega \ni q \to p} \sigma_\Omega(q) = 1.$$

Since every boundary point of a C^2 strongly convex bounded domain is spherically extreme, this theorem implies Theorem 1.2. The rest of this section is devoted to the proof of Theorem 3.1.

Proof. The proof proceeds in seven steps.

Step 1. *Sphere envelopes.* Let Ω be a bounded domain in $\mathbb{C}^n$ with a boundary point $p \in \partial\Omega$ such that

(i) $\partial\Omega \cap \mathbb{B}^n(p; r_0)$ is C^2-smooth for some $r_0 > 0$, and

(ii) p is a spherically extreme boundary point of Ω.

Then there exist positive constants r_1, r_2, and R with $r_0 > r_1 > r_2$ such that every $q \in \Omega \cap \mathbb{B}^n(p; r_2)$ admits points $b(q) \in \partial\Omega \cap \mathbb{B}^n(p; r_1)$ and $c(q) \in \mathbb{C}^n$ satisfying the conditions

(iii) $\|q - b(q)\| < \|q - z\|$ for any $z \in \partial\Omega - \{b(q)\}$, and

(iv) $\|c(q) - b(q)\| = R$ and $\Omega \subset \mathbb{B}^n(c(q); R)$.

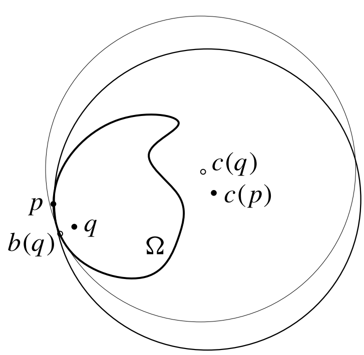

Figure 1. Sphere envelopes.

See Figure 1. Notice that (iii) says that $b(q)$ is the unique boundary point that is the closest to q, and that the constant R in (iv) is independent of the choice of $q \in \mathbb{B}^n(p; r_2)$.

Step 2. *Centering.* In the following we shall use the familiar notation

$$\begin{aligned} z &= (z_1, \dots, z_n), & z' &= (z_2, \dots, z_n), \\ u &= \operatorname{Re} z_1, & v &= \operatorname{Im} z_1. \end{aligned} \tag{3-1}$$

For each $q \in \Omega \cap \mathbb{B}^n(p; r_2)$, choose a unitary transform U_q of $\mathbb{C}^n$ such that the map $A_q(z) := U_q(z - b(q))$, depicted in Figure 2, satisfies the following conditions:

$$A_q(q) = (\lambda_q, 0, \dots, 0) \tag{3-2}$$

for some $\lambda_q > 0$, and

$$A_q(\Omega) \subset \mathbb{B}^n((R, 0, \dots, 0); R) = \{z \in \mathbb{C}^n : |z_1 - R|^2 + \|z'\|^2 < R^2\}. \tag{3-3}$$

Then there exists a positive constant $r_3 < r_2$ such that

$$\begin{aligned} &z \in A_q(\Omega) \cap \mathbb{B}^n(\mathbf{0}; r_3) \\ &\qquad \iff \|z\| < r_3 \text{ and } 2u > H_{b(q)}(z') + \mathcal{K}_{b(q)}(v, z') + \mathcal{R}_{b(q)}(v, z'), \end{aligned} \tag{3-4}$$

where

- $H_{b(q)}$ is a quadratic positive-definite Hermitian form such that there exists a constant $c_0 > 0$, independent of q, satisfying

$$H_{b(q)}(z') \geq c_0 \|z'\|^2, \tag{3-5}$$

  and

- there exists a constant $C > 0$, independent of $q \in \mathbb{B}^n(p; r_3) \cap \Omega$, such that

$$|\mathcal{K}_{b(q)}(v, z')| \leq C(|v|^2 + |v|\|z'\|) \tag{3-6}$$

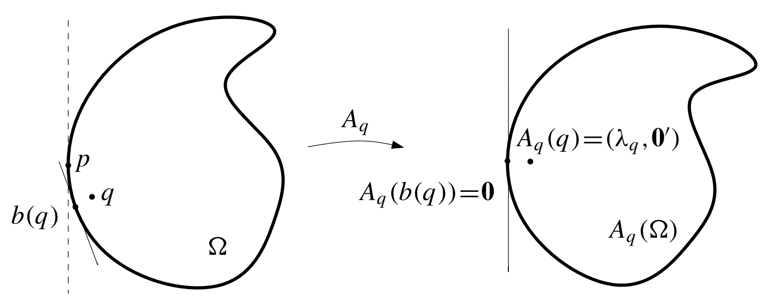

Figure 2. The centering process.

whenever $z \in \mathbb{B}^n(\mathbf{0}; r_3)$. Furthermore, we have

$$|\mathcal{R}_{b(q)}(v, z')| = o(|v|^2 + \|z'\|^2).$$

In particular, the choice of r_3 can allow us the estimate

$$|\mathcal{R}_{b(q)}(v, z')| \le \frac{c_0}{2}(|v|^2 + \|z'\|^2).$$

Notice that

$$\lim_{\Omega \ni q \to p} b(q) = p, \qquad \lim_{\Omega \ni q \to p} H_{b(q)}(z') = H_p(z'),$$

and

$$\lim_{\Omega \ni q \to p} A_q = I \text{ (the identity map).}$$

The latter limit and an inductive construction yield that for each integer $m > 2$ there exists a strictly increasing integer-valued function $k(m)$ such that

$$\mathbb{B}^n\big(\mathbf{0}; r_3/(2k(m))\big) \subset A_q\big(\mathbb{B}^n(p; r_3/k(m))\big) \subset \mathbb{B}^n(\mathbf{0}; r_3/m) \tag{3-7}$$

whenever $q \in \mathbb{B}^n\big(p; r_3/(2k(m))\big)$.

Step 3. *The Cayley transform.* The *Cayley transform* considered here is the map

$$\kappa(z) := \Big(\frac{1-z_1}{1+z_1}, \frac{\sqrt{2}z_2}{1+z_1}, \dots, \frac{\sqrt{2}z_n}{1+z_1}\Big), \tag{3-8}$$

well-defined except at points of $Z = \{z \in \mathbb{C}^n : z_1 = -1\}$. Notice that this transform maps the open unit ball $\mathbb{B}^n(\mathbf{0}; 1)$ biholomorphically onto the Siegel half-space

$$\mathcal{S}_0 := \{z \in \mathbb{C}^n : 2\operatorname{Re} z_1 > \|z'\|^2\}. \tag{3-9}$$

Moreover, $\kappa \circ \kappa = 1$ and consequently, $\kappa(\mathcal{S}_0) = \mathbb{B}^n(\mathbf{0}; 1)$. Notice also that, for $\mathbf{1} = (1, 0, \dots)$ and $-\mathbf{1} = (-1, 0, \dots)$, we have $\kappa(\mathbf{1}) = (0, \dots, 0)$, $\kappa((0, \dots, 0)) = \mathbf{1}$, $\kappa(-\mathbf{1}) = \infty$, and $\kappa(\infty) = -\mathbf{1}$.

Step 4. *Stretching.* Let $q \in \Omega \cap \mathbb{B}^n\big(p; r_3/(2k(m))\big)$. If we let m tend to infinity, then of course $A_q(q) = (\lambda_q, 0, \dots, 0)$ approaches $A_q(b(q)) = (0, \dots, 0)$, and so λ_q approaches zero. For simplicity, we write $\lambda = \lambda_q$, suppressing the q but keeping in mind that λ is still dependent upon q. Note that

$$A_q(\mathbb{B}^n(c(q); R)) = \{z \in \mathbb{C}^n : 2R \cdot \operatorname{Re} z_1 > \|z\|^2\}. \tag{3-10}$$

Define the *stretching map* $\Lambda_\lambda : \mathbb{C}^n \to \mathbb{C}^n$, first introduced in [Pinchuk 1991], by

$$\Lambda_\lambda(z) := \Big(\frac{z_1}{\lambda}, \frac{z_2}{\sqrt{\lambda}}, \dots, \frac{z_n}{\sqrt{\lambda}}\Big). \tag{3-11}$$

Recall (3-6). The stretching map transforms $A_q(\Omega)\cap\mathbb{B}^n(\mathbf{0}; r_3/3)$ to the domain $\Lambda_\lambda\big(A_q(\Omega)\cap\mathbb{B}^n(\mathbf{0}; r_3/3)\big)$ so that

$$\begin{aligned}(3\text{-}12)\quad & z\in\Lambda_\lambda\circ A_q(\Omega)\cap\mathbb{B}^n\Big(\mathbf{0}; \frac{r_3}{\sqrt{\lambda}k(3)}\Big)\\ &\iff \|z\|<\frac{r_3}{\sqrt{\lambda}k(3)} \text{ and}\\ &\qquad 2u>H_{b(q)}(z')+\frac{1}{\lambda}K_{b(q)}\big(\lambda v, \sqrt{\lambda}z'\big)+\frac{1}{\lambda}\mathcal{R}_{b(q)}\big(\lambda v, \sqrt{\lambda}z'\big).\end{aligned}$$

On the other hand, notice that

$$\Big\|\frac{1}{\lambda}K_{b(q)}\big(\lambda v, \sqrt{\lambda}z'\big)\Big\|\le C\sqrt{\lambda}\big(\sqrt{\lambda}|v|^2+|v|\|z'\|\big)$$

and that

$$\Big\|\frac{1}{\lambda}\mathcal{R}_{b(q)}\big(\lambda v, \sqrt{\lambda}z'\big)\Big\|\le\frac{1}{\lambda}o\big((|\lambda v|^2+\|\sqrt{\lambda}z'\|^2)\big)=\frac{1}{\lambda}o(\lambda)$$

on $\mathbb{B}^n(\mathbf{0};\rho)$ for any fixed constant $\rho>0$. Notice that both terms approach zero as λ tends to zero. Thus, these terms can become sufficiently small if we limit q to being contained in $\mathbb{B}^n\big(p; r_3/(2k(m))\big)$ for some sufficiently large m.

Step 5. *Set-convergence.* This step is in part heuristic; the heuristics appearing in this step, especially those which concern set convergences, are not used in the proof, strictly speaking. We include this step because they seem to help us to grasp the logical structure of the proof. On the other hand, the constructions in (3-13)–(3-15) shall be used in the remainder of the proof, especially in Step 7.

The main role of the stretching map Λ_λ, as $\lambda\searrow 0$, is to rescale the domains successively, letting them converge to the set limits.

For instance, if one considers

$$\Lambda_\lambda(A_q(\Omega)\cap\mathbb{B}^n(\mathbf{0}; r_3)),$$

then one can see that $\Lambda_\lambda(\mathbb{B}^n(\mathbf{0}; r_3))$ contains $\mathbb{B}^n\big(\mathbf{0}; r_3/\sqrt{\lambda}\big)$, a very large ball which exhausts $\mathbb{C}^n$ as λ approaches zero. In the meantime, within that large ball, $\Lambda_\lambda(A_q(\Omega))$ is restricted only by the inequality

$$2u>H_{b(q)}(z')+\widetilde{K}_\lambda(v, z'),$$

where $\widetilde{K}_\lambda=o(\lambda)$ is small enough to be negligible. One can imagine that indeed the "limit domain" of this procedure should be

$$\widehat{\Omega}:=\{z\in\mathbb{C}^n : 2u>H_p(z')\}. \tag{3-13}$$

Here, of course, $H_p(z')$ is the quadratic positive-definite Hermitian form which appears in the defining inequality of Ω about the boundary point p (understood as the origin):

$$2\operatorname{Re} z_1 > H_p(z') + o(|\operatorname{Im} z_1| + \|z'\|^2).$$

Notice that

$$\kappa(\widehat{\Omega}) = \{z \in \mathbb{C}^n : |z_1|^2 + H_p(z') < 1\},$$

and hence there is a $\mathbb{C}$-linear isomorphism

$$L : \mathbb{C}^n \to \mathbb{C}^n \tag{3-14}$$

that maps $\kappa(\widehat{\Omega})$ biholomorphically onto the unit ball $\mathbb{B}^n(\mathbf{0}; 1)$ with $L(\mathbf{1}) = \mathbf{1}$.

Before moving on to the next step we remark that, since $\Omega \subset \mathbb{B}^n(c(q); R)$ whenever $q \in \mathbb{B}^n(p; r_2)$,

$$A_q(\Omega) \subset A_q(\mathbb{B}^n(c(q); R)) = \mathbb{B}^n((R, 0, \dots, 0); R).$$

This in turn implies that

$$\begin{aligned} \Lambda_\lambda \circ A_q(\Omega) &\subset \Lambda_\lambda\big(\mathbb{B}^n((R, 0, \dots, 0); R)\big) \\ &\subset \mathcal{E} := \{z \in \mathbb{C}^n : 2R \cdot \operatorname{Re} z_1 > \|z'\|^2\}. \end{aligned} \tag{3-15}$$

The last inclusion follows by (3-10).

Step 6. *Auxiliary domains.* Let $\delta > 0$ be given. Consider the domains

$$\mathcal{G}_\delta := \{z \in \mathbb{C}^n : 2u > -\delta|v| + (1-\delta)H_{b(q)}(z')\}, \tag{3-16}$$

$$\mathcal{F}_\delta := \{z \in \mathbb{C}^n : 2u > \delta|v| + (1+\delta)H_{b(q)}(z')\}, \text{ and} \tag{3-17}$$

$$\mathcal{H}_q := \{z \in \mathbb{C}^n : 2u > H_{b(q)}(z')\}, \tag{3-18}$$

in addition to $\widehat{\Omega}$ and $\mathcal{E}$ introduced in (3-13) and (3-15).

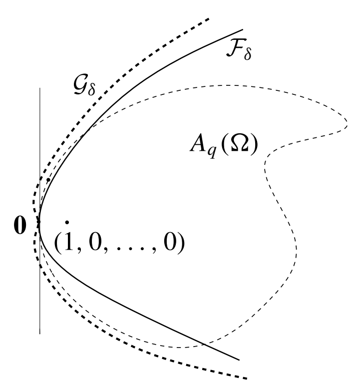

Figure 3. Auxiliary domains $\mathcal{G}_\delta$ and $\mathcal{F}_\delta$.

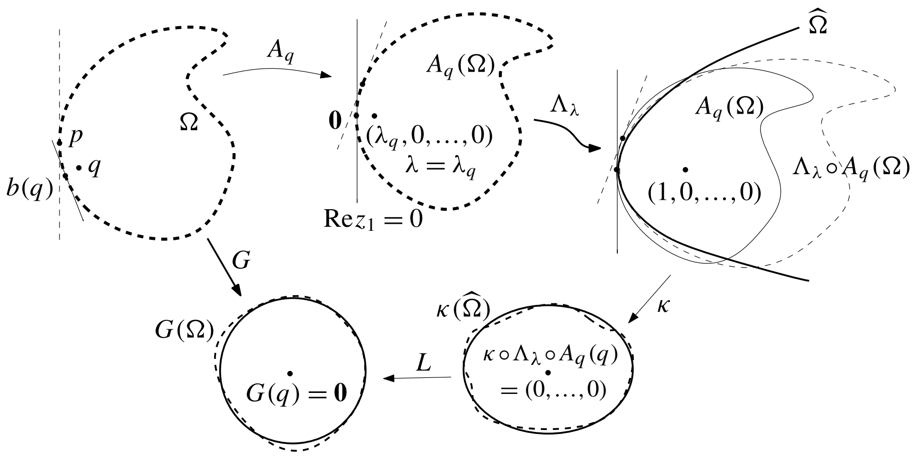

Figure 4. $G(\Omega) = L \circ \kappa \circ \Lambda_\lambda \circ A_q(\Omega)$ for $q \sim p$.

A straightforward computation checks that the image $\kappa(\mathcal{G}_\delta)$ of $\mathcal{G}_\delta$ via the Cayley transform κ introduced earlier is

$$\kappa(\mathcal{G}_\delta) = \left\{ z \in \mathbb{C}^n : |z_1|^2 - \frac{\delta}{2}|z_1 - \bar{z}_1| + (1-\delta) H_{b(q)}(z') < 1 \right\}. \tag{3-19}$$

Hence, there exists $\delta_0 > 0$ such that, for every δ with $0 < \delta < \delta_0$, $\kappa(\mathcal{G}_\delta)$ is a bounded domain. Notice also that this domain is arbitrarily close to the domain $\kappa(\mathcal{H}_{b(q)})$ as δ_0 becomes arbitrarily small. It follows therefore that, for every $\epsilon > 0$, there exists $\delta_0 > 0$ such that

$$L \circ \kappa(\mathcal{G}_\delta) \subset \mathbb{B}^n(\mathbf{0}; 1+\epsilon) \tag{3-20}$$

whenever $0 < \delta < \delta_0$. Moreover, observe that the stretching map Λ_λ preserves all such domains as

$$\mathcal{F}_\delta, \quad \mathcal{G}_\delta, \quad \mathcal{H}_q, \quad \widehat{\Omega}, \quad \text{and} \quad \mathcal{E}.$$

Let us now define the expression

$$G(z) := L \circ \kappa \circ \Lambda_\lambda \circ A_q(z) \tag{3-21}$$

for $z \in \mathbb{C}^n - (\Lambda_\lambda \circ A_q)^{-1}(Z)$. (The set Z is defined in (3-8). Notice that this expression G depends upon $q \in \mathbb{B}^n(\mathbf{0}; r_2)$; see Figure 3 for an illustration.) In particular, this G maps Ω onto its image $G(\Omega)$ biholomorphically. See also Figure 4.

Step 7. *Proof of Theorem 3.1.* Our final goal is to establish the following claim.

Claim. *For any ϵ with $0 < \epsilon < 1/2$, there exists an integer $m > 0$ such that*

$$\mathbb{B}^n(\mathbf{0}; 1-\epsilon) \subset G(\Omega) \subset \mathbb{B}^n(\mathbf{0}; 1+\epsilon) \tag{3-22}$$

whenever $q \in \Omega \cap \mathbb{B}^n\big(p; r_3/(2k(m))\big)$.

Since $G(q) = \mathbf{0}$, this implies that the squeezing function σ_Ω satisfies

$$\sigma_\Omega(q) \geq \frac{1-\epsilon}{1+\epsilon}.$$

Notice that this completes the proof of Theorem 3.1. Therefore it remains only to prove the claim.

Proof. Start with $\mathbb{B}^n(\mathbf{0}; 1-\epsilon)$. Notice first, by the definition of $\mathcal{F}_\delta$, that for every $\delta > 0$ there exists $m_1 > 0$ such that

$$\mathcal{F}_\delta \cap \mathbb{B}^n(\mathbf{0}; r_2/m) \subset A_q(\Omega) \cap \mathbb{B}^n(\mathbf{0}; r_2/m)$$

for any $m > m_1$.

Also,

$$\kappa^{-1} \circ L^{-1}(\mathbb{B}^n(\mathbf{0}; 1-\epsilon)) \Subset \kappa^{-1} \circ L^{-1}(\mathbb{B}^n(\mathbf{0}; 1)) = \widehat{\Omega}.$$

As discussed in (3-4)–(3-7), $L \circ \kappa(\mathcal{H}_q)$ is sufficiently close to $L \circ \kappa(\widehat{\Omega})$, which is the unit ball, whenever $q \in \mathbb{B}^n\big(p; r_3/(2k(m))\big)$ and m is sufficiently large. Therefore there exists an integer $m_2 > m_1$ such that $(L \circ \kappa)^{-1}(\mathbb{B}^n(\mathbf{0}; 1-\epsilon)) \Subset \mathcal{H}_q$ whenever $q \in \mathbb{B}^n(p; r_3/m_2)$.

As in (3-19), a direct computation yields

$$\kappa(\mathcal{F}_\delta) = \big\{z \in \mathbb{C}^n : |z_1|^2 + \tfrac{1}{2}\delta|z_1 - \bar{z}_1| + (1+\delta)H_{b(q)}(z') < 1\big\}. \tag{3-23}$$

Now, consider the set $L \circ \kappa \circ \Lambda_\lambda(\mathcal{F}_\delta)$ for each $\delta > 0$. (Recall that $\Lambda_\lambda(\mathcal{F}_\delta) = \mathcal{F}_\delta$ as remarked in the line below (3-20).) These domains increase monotonically as $\delta \searrow 0$ (since the $\mathcal{F}_\delta$'s do) in such a way that the union $\bigcup_{0<\delta<\delta_0} L \circ \kappa \circ (\mathcal{F}_\delta)$ becomes arbitrarily close to $\mathbb{B}^n(\mathbf{0}; 1)$ for m sufficiently large. Consequently there exists a constant $\delta > 0$ such that $\mathbb{B}^n(\mathbf{0}; 1-\epsilon) \Subset L \circ \kappa \circ (\mathcal{F}_\delta)$. Moreover there is an integer $m_3 > m_2$ such that

$$\Lambda_\lambda^{-1}\big(\kappa^{-1} \circ L^{-1}(\mathbb{B}^n(\mathbf{0}; 1-\epsilon)\big) \subset \mathbb{B}^n(\mathbf{0}; r_3/k(m_1)), \tag{3-24}$$

as Λ_λ^{-1} scales down the compact subsets (for $\lambda < r_3/m_2$ sufficiently small) to a small set near the origin. Hence, we have

$$\Lambda_\lambda^{-1}\big(\kappa^{-1} \circ L^{-1}(\mathbb{B}^n(\mathbf{0}; 1-\epsilon))\big) \subset \mathcal{F}_\delta \cap \mathbb{B}^n(\mathbf{0}; r_3/k(m_1)) \subset \Omega.$$

Consequently, as long as $q \in \mathbb{B}^n\big(p; r_3/(2k(m_3))\big)$,

$$\begin{aligned} \mathbb{B}^n(\mathbf{0}; 1-\epsilon) &\subset L \circ \kappa \circ \Lambda_\lambda\big(\mathcal{F}_\delta \cap \mathbb{B}^n(\mathbf{0}; r_3/k(m_1))\big) \\ &\subset L \circ \kappa \circ \Lambda_\lambda(A_q(\Omega)) \\ &= G(\Omega). \end{aligned} \tag{3-25}$$

See Figure 5.

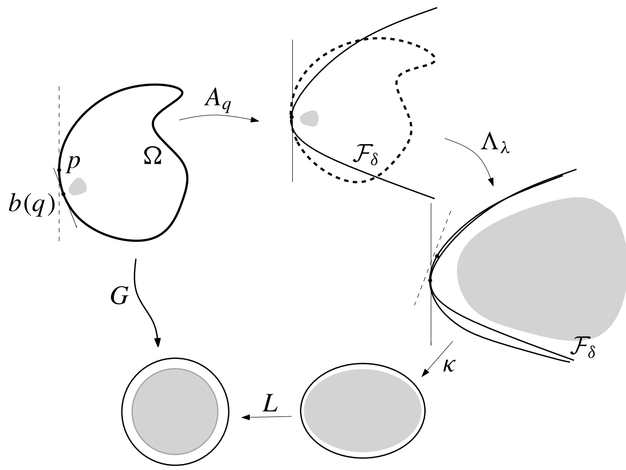

Figure 5. $\mathbb{B}^n(\mathbf{0};\, 1-\epsilon) \subset G(\Omega)$.

Now we show that $G(\Omega) \subset \mathbb{B}^n(\mathbf{0};\, 1+\epsilon)$. Consider

$$\Omega' := \Omega - \mathbb{B}^n(p;\, r_2).$$

Notice that there exists an integer $\ell \gg 1$ such that

$$A_q(\Omega') \subset A_q(\Omega) - \mathbb{B}^n(\mathbf{0};\, r_2/\ell) \subset \mathcal{E} - \mathbb{B}^n(\mathbf{0};\, r_2/\ell). \tag{3-26}$$

Now, there is an integer $m_4 > m_3$ such that, if $m > m_4$ and $q \in \mathbb{B}^n\big(p;\, r_3/(2k(m))\big)$, then

$$\Lambda_\lambda(\mathcal{E} - \mathbb{B}^n(\mathbf{0};\, r_2/k)) \subset \left\{ z \in \mathcal{E} : \operatorname{Re} z_1 > \frac{r_2}{r_3} \cdot \frac{m_4}{\ell} \right\}.$$

This implies that there exists m_4 such that

$$G(\Omega') \subset L \circ \kappa \left(\left\{ z \in \mathcal{E} : \operatorname{Re} z_1 > \frac{r_2}{r_3} \cdot \frac{m_4}{\ell} \right\} \right) \subset \mathbb{B}^n(-\mathbf{1};\, \rho(m_4))$$

for some $\rho(m)$ which approaches zero as m tends to infinity; a direct computation with the Cayley transform and the choice of L (see (3-14)) verify this immediately. Therefore, choosing m_4 sufficiently large, we arrive at

$$G(\Omega') \subset \mathbb{B}^n(-\mathbf{1};\, \epsilon), \tag{3-27}$$

as in Figure 6. For the ϵ given above, there exists δ such that

$$L \circ \kappa(\mathcal{G}_\delta) \subset \mathbb{B}^n(\mathbf{0};\, 1+\epsilon). \tag{3-28}$$

Fix this δ, and recall how the auxiliary domain $\mathcal{G}_\delta$ was defined in (3-16). Given any $\delta > 0$, according to (3-4)–(3-6), there exists $\rho > 0$ such that

$$A_q(\Omega) \cap \mathbb{B}^n(\mathbf{0};\, \rho) \subset \mathcal{G}_\delta.$$

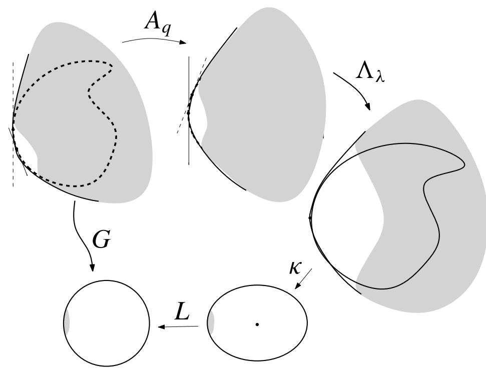

Figure 6. $G(\Omega') \subset \mathbb{B}^n(-\mathbf{1}; \epsilon)$.

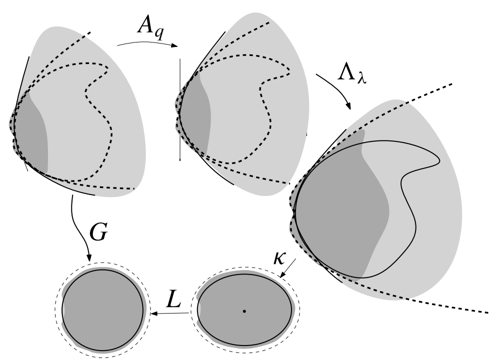

Figure 7. $G(\Omega) \subset \mathbb{B}^n(\mathbf{0}; 1+\epsilon)$.

On the other hand, we can go back to (3-26) and require that $r_2/\ell < \rho/2$. Then we have

$$A_q(\Omega) \cap \mathbb{B}^n(\mathbf{0}; 2r_2/\ell) \subset \mathcal{G}_\delta. \tag{3-29}$$

Since there exists an integer $m_5 > 0$ such that $A_q(\mathbb{B}^n(p; r_2/\ell)) \subset \mathbb{B}^n(\mathbf{0}; 2r_2/\ell)$, we have that

$$G(\Omega - \Omega') \subset L \circ \kappa \circ \Lambda_\lambda\big(A_q(\Omega) \cap \mathbb{B}^n(\mathbf{0}; 2r_2/\ell)\big).$$

This implies

$$\begin{aligned} G(\Omega - \Omega') &\subset L \circ \kappa \circ \Lambda_\lambda\big(A_q(\Omega) \cap \mathbb{B}^n(\mathbf{0}; 2r_2/\ell)\big) \\ &\subset L \circ \kappa \circ \Lambda_\lambda(\mathcal{G}_\delta) && \text{by (3-29)} \\ &\subset L \circ \kappa(\mathcal{G}_\delta) && \text{by the sentence following (3-20)} \\ &\subset \mathbb{B}^n(\mathbf{0}; 1+\epsilon). \end{aligned} \tag{3-30}$$

By (3-27) and (3-30), as in Figure 7 we have that

$$G(\Omega) \subset \mathbb{B}^n(\mathbf{0}; 1+\epsilon).$$

This completes the proof of the claim, and therefore of Theorem 3.1. □

4. Remarks

In this final section we present several remarks.

On the spherically extreme points. Pertaining to the question in the introduction, one of the naturally arising questions would be whether one may re-embed (the closure of) the bounded strongly pseudoconvex domain so that the preselected boundary point becomes spherically extreme. This question was answered affirmatively in by Diederich, Fornaess, and Wold in [Diederich et al. 2014]. Owing to this new result, Theorem 3.1 now implies the following.

Theorem 4.1. *If Ω is a bounded domain in $\mathbb{C}^n$ with a C^2-smooth strongly pseudoconvex boundary*, *then* $\lim_{\Omega \ni z \to \partial\Omega} \sigma_\Omega(z) = 1$.

On the other hand, a more ambitious goal may be to re-embed the domain using the automorphisms of $\mathbb{C}^n$ to achieve the same result. But this cannot work, as shown by the following counterexample.

Example. Consider the domain U which is the open $1/10$-tubular neighborhood of the circle $S := \{(e^{it}, 0) \in \mathbb{C}^2 : t \in \mathbb{R}\}$. This domain is strongly pseudoconvex. Let $p = (9/10, 0)$. Clearly $p \in \partial U$. If there were $\psi \in \mathrm{Aut}(\mathbb{C}^2)$ which makes $\psi(p)$ spherically extreme for $\psi(U)$, then consider the analytic disc $\Sigma := \psi(\Delta)$ where $\Delta := \{(z, 0) : |z| \le 1\}$). Since Δ crosses ∂U transversally at $\psi(p)$, Σ crosses the sphere envelope at $\psi(p)$ and extends to the exterior of the sphere. On the other hand, the boundary of Σ remains inside $\psi(U)$ and hence inside the sphere. Now let the sphere expand radially from its center, stopping at the radius beyond which it cannot intersect the holomorphic disc Σ. Then the sphere is tangent to a point to Σ at an interior point, keeping the whole disc inside the sphere. The maximum principle now implies that Σ should be entirely on the sphere. But the boundary of Σ is strictly inside the sphere, which is a contradiction. This implies that p cannot be made spherically extreme via any re-embedding by an automorphism of $\mathbb{C}^n$.

Acknowledgement: This example was obtained after a valuable discussion between Kim and Josip Globevnik. Kim would like to express his thanks to Josip Globevnik for pointing out such a possibility.

On the exhaustion theorem by Fridman–Ma. The main theorem by Buma Fridman and Daowei Ma [1995] obtained the conclusion of Theorem 3.1 in the special case $\Omega \ni q \to p$ *transversely* to the boundary $\partial\Omega$. However, that is not sufficient to prove Theorem 3.1; it is indeed necessary to consider all possible sequences approaching the boundary. Fridman and Ma [1995] did not need to consider the point sequences approaching the boundary tangentially, as their interest was only on the holomorphic exhaustion of the ball by the biholomorphic images of a bounded strongly pseudoconvex domain. On the other hand, our proof of Theorem 3.1 gives a proof to their theorem as well; we only need to use $(1+\epsilon)^{-1}G(z)$ instead of G. (Recall that G depends upon q. Letting q converge to p and ϵ tend to zero, one gets a sequence of maps that exhausts the unit ball holomorphically.)

Plane domain cases. For domains in $\mathbb{C}$, several theorems have been obtained by F. Deng, Q. Guan, and L. Zhang [Deng et al. 2012]. Theorem 3.1 obviously includes many of those results, as every boundary point of a plain domain with C^2-smooth boundary is spherically extreme.

Acknowledgements

This research is supported in part by SRC-GaiA (Center for Geometry and its Applications) and the Grant 2011-0030044 from the Ministry of Education. The research of Kim is also supported in part by National Research Foundation Grant 2011-0007831 of South Korea.

References

[Deng et al. 2012] F. Deng, Q. Guan, and L. Zhang, "Some properties of squeezing functions on bounded domains", *Pacific J. Math.* **257**:2 (2012), 319–341. MR 2972468 Zbl 1254.32015

[Deng et al. 2015] F. Deng, Q. Guan, and L. Zhang, "Properties of squeezing functions and global transformations of bounded domains", *Trans. Amer. Math. Soc.* (online publication August 2015).

[Diederich et al. 2014] K. Diederich, J. E. Fornæss, and E. F. Wold, "Exposing points on the boundary of a strictly pseudoconvex or a locally convexifiable domain of finite 1-type", *J. Geom. Anal.* **24**:4 (2014), 2124–2134. MR 3261733 Zbl 1312.32006

[Fridman and Ma 1995] B. L. Fridman and D. Ma, "On exhaustion of domains", *Indiana Univ. Math. J.* **44**:2 (1995), 385–395. MR 96h:32028 Zbl 0843.32014

[Kim 1992] K.-T. Kim, "Asymptotic behavior of the curvature of the Bergman metric of the thin domains", *Pacific J. Math.* **155**:1 (1992), 99–110. MR 93f:32025 Zbl 0723.32009

[Liu et al. 2004] K. Liu, X. Sun, and S.-T. Yau, "Canonical metrics on the moduli space of Riemann surfaces I", *J. Differential Geom.* **68**:3 (2004), 571–637. MR 2007g:32009 Zbl 1078.30038

[Liu et al. 2005] K. Liu, X. Sun, and S.-T. Yau, "Canonical metrics on the moduli space of Riemann surfaces II", *J. Differential Geom.* **69**:1 (2005), 163–216. MR 2007g:32010 Zbl 1086.32011

[Pinchuk 1991] S. Pinchuk, "The scaling method and holomorphic mappings", pp. 151–161 in *Several complex variables and complex geometry, Part 1* (Santa Cruz, CA, 1989), edited by E. Bedford et al., Proceedings of Symposia in Pure Mathematics **52**, American Mathematical Society, Providence, RI, 1991. MR 92i:32031 Zbl 0744.32013

[Yeung 2009] S.-K. Yeung, "Geometry of domains with the uniform squeezing property", *Adv. Math.* **221**:2 (2009), 547–569. MR 2010b:32034 Zbl 1165.32004

Received October 8, 2014. Revised October 13, 2015.

KANG-TAE KIM
CENTER FOR GEOMETRY AND ITS APPLICATIONS AND DEPARTMENT OF MATHEMATICS
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY
POHANG 790-784
SOUTH KOREA
kimkt@postech.ac.kr

LIYOU ZHANG
SCHOOL OF MATHEMATICAL SCIENCE
CAPITAL NORMAL UNIVERSITY
BEIJING, 100048
CHINA
zhangly@cnu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.359

LEFSCHETZ PENCILS AND FINITELY PRESENTED GROUPS

RYOMA KOBAYASHI AND NAOYUKI MONDEN

From the works of Gompf and Donaldson, it is known that every finitely presented group can be realized as the fundamental group of the total space of a Lefschetz pencil. We give an alternative proof of this fact by providing the monodromy explicitly. In the proof, we give an alternative construction of the monodromy of Gurtas' fibration and a lift of that to the mapping class group of a surface with two boundary components.

1. Introduction

There exist Lefschetz pencils (fibrations over S^2 with (-1)-sections) whose total spaces have a prescribed fundamental group. This follows as a corollary of the results of Gompf [1995], who showed that every finitely presented group is realized as the fundamental group of some closed symplectic 4-manifold, and of Donaldson [1999], who showed that every closed symplectic 4-manifold admits a Lefschetz pencil. Note that since we obtain a Lefschetz fibration with (-1)-sections by blowing up the base locus of a Lefschetz pencil, and blowing up has no effect on the fundamental groups of 4-manifolds, the above claim for Lefschetz fibrations with (-1)-sections follows. Conversely, a 4-manifold admitting a Lefschetz pencil (fibration with fiber genus greater than one) is symplectic (cf. [Gompf and Stipsicz 1999]).

Let Σ_g^b be a compact oriented surface of genus g with b boundary components $\delta_1, \dots, \delta_b$, and let Mod_g^b be the mapping class group of Σ_g^b. We denote by t_c the right-handed Dehn twist along a simple closed curve c in Σ_g^b. Then a relation $\prod_{j=1}^{b} t_{\delta_j} = \prod_{i=1}^{m} t_{v_i}$ provides a genus-g Lefschetz pencil/fibration with b base points/(-1)-sections. Conversely, given any Lefschetz pencil (fibration with (-1)-sections), we obtain such a relation. However, the relations corresponding to the above Lefschetz pencils/fibrations constructed based on the results of [Gompf 1995] and [Donaldson 1999] are implicit. Our purpose is to provide the relation of such a genus-g Lefschetz pencil explicitly, so this gives an alternative proof of the above corollary using mapping class group arguments. To state our main result, we need to introduce some notation.

MSC2010: primary 57R17; secondary 20F34.
Keywords: Lefschetz pencil, Lefschetz fibration, fundamental group, mapping class group.

Definition 1.1. Let $\Gamma = \langle x_1, x_2, \ldots, x_n \mid r_1, r_2, \ldots, r_k \rangle$ be a finitely presented group with n generators and k relations. For $w \in \Gamma$, we define $l(w)$, called the *syllable length* of w, to be

$$l(w) = \min\{s \mid w = x_{i_1}^{m_1} x_{i_2}^{m_2} \cdots x_{i_s}^{m_s} \text{ for } 1 \leq i_j \leq n \text{ and } m_j \in \mathbb{Z}\}.$$

Define $l = \max\{l(r_i) \mid 1 \leq i \leq k\}$. If $k = 0$, we define $l = 1$ (note that l depends on the presentation and that our definition of l differs from that of [Korkmaz 2009]). We always assume that the relators r_i are cyclically reduced.

In Section 5A, we give a relation $t_{\delta_1} t_{\delta_2} = W_2^g(1, \psi_k)$ in Mod_g^2 using certain substitution techniques, where $W_2^g(1, \psi_k)$ is a product of right-handed Dehn twists. Our main result is the following:

Theorem 1.2. *If $k \geq 1$ (resp. $k = 0$), then, for $g \geq 4(n + l - 1) + k$ (resp. $g \geq 4n + 2$), there exists a genus-g Lefschetz pencil/fibration with two base points/(-1)-sections on a closed symplectic 4-manifold X such that $t_{\delta_1} t_{\delta_2} = W_2^g(1, \psi_k)$ is the corresponding relation and $\pi_1(X)$ is isomorphic to Γ.*

Theorem 1.2 gives an upper bound for the minimum g, denoted by $g_P(\Gamma)$, for which there exists a genus-g Lefschetz pencil on X such that $\pi_1(X)$ is isomorphic to Γ. We describe it in Section 8. To give a better upper bound on $g_P(\Gamma)$, we construct a lift of Gurtas' positive relator (see [Gurtas 2004]), denoted by θ^2, to Mod_g^2 in Section 6 by combining a lift of a hyperelliptic involution and the relation given in [Korkmaz 2009] to Mod_g^2. On the other hand, Gurtas showed that the positive word θ^2 given in [Gurtas 2004] is a positive relator by checking the images of certain cycles on Σ_g under θ. In this sense, our construction of the monodromy of Gurtas' fibration is different from that in [Gurtas 2004].

Here, we explain why we focus on Lefschetz fibrations with (-1)-sections. A section of a Lefschetz fibration over S^2 plays important roles in the total space. The existence of a section σ of a Lefschetz fibration $f : X \to S^2$ with a fiber F is required to compute the fundamental group of X and to decide whether X is spin or not (see [Gompf and Stipsicz 1999; Stipsicz 2001b]). In addition, the complement of a regular neighborhood of $F \cup \sigma$ is a Stein filling of its boundary equipped with the induced tight contact structure (see [Akbulut and Ozbagci 2002; Etnyre and Honda 2002; Loi and Piergallini 2001]). Especially, a (-1)-section is important in Lefschetz fibrations in the following senses.

(i) Blowing up of the base locus of a Lefschetz pencil yields a Lefschetz fibration with (-1)-sections. Conversely, we can obtain a Lefschetz pencil by blowing down of (-1)-sections of a Lefschetz fibration.

(ii) From given Lefschetz fibrations, we can construct a new Lefschetz fibration by fiber summing them. If a Lefschetz fibration admits a (-1)-section, it cannot be decomposed as any nontrivial fiber sum (see [Stipsicz 2001a; Smith 2001]).

For these reasons, we can regard Lefschetz fibrations with (-1)-sections as "fundamental" and "prime" ones.

Note that we can express Gompf's result in terms of Lefschetz fibrations over S^2. The article [Amorós et al. 2000] gave a construction of Lefschetz fibrations whose total spaces have a given fundamental group without using Donaldson's result. However, their monodromies are implicit. The explicit monodromies of such fibrations were given by Korkmaz [2009]. Akhmedov and Ozbagci [2013] gave a new construction of such fibrations, and the first author [Kobayashi 2015] improved the result of [Korkmaz 2009]. For technical reasons, the fibrations in [Korkmaz 2009; Akhmedov and Ozbagci 2013; Kobayashi 2015] have no (-1)-sections (see Section 8), so we would like to emphasize that our result is different from the above four results.

Here is an outline of this paper. In Section 2, we fix notation. In Section 3, we introduce a substitution technique and the relation constructed by Korkmaz. Section 4 reviews some standard facts on Lefschetz fibrations and pencils. In Section 5, we prove the main results. In Section 6, we give an alternative construction of the monodromy of Gurtas' fibration and provide a lift of that to the mapping class group of a surface with two boundary components. In Section 7, we introduce the construction of a loop which is needed for the proof of Theorem 1.2. In Section 8, we give an upper bound of $g_P(\Gamma)$ and some remarks.

2. Notation

Let Σ_g be the closed oriented surface of genus g standardly embedded in 3-space as shown in Figure 1. We use the symbols $a_1, b_1, \dots, a_g, b_g$ to denote the standard generators of the fundamental group $\pi_1(\Sigma_g)$ of Σ_g. For a and b in $\pi_1(\Sigma_g)$, the notation ab means that we first apply a then b.

Let $c_0, c_1, c_2, \dots, c_g, a_0, a_{g+1}, a'_0, a'_{g+1}$ be the simple loops in Σ_g depicted in Figure 1. Note that in $\pi_1(\Sigma_g)$, up to conjugation,

(1) $$c_i = b_i^{-1} \cdots b_1^{-1}(a_1 b_1 a_1^{-1}) \cdots (a_i b_i a_i^{-1}) \quad \text{for each } 1 \le i \le g;$$

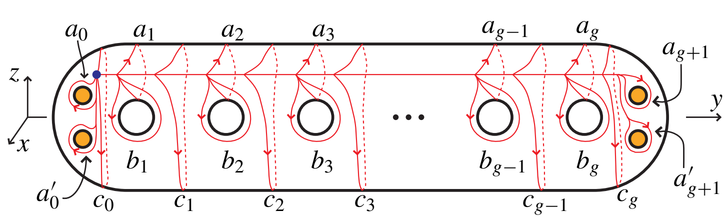

Figure 1. Generators a_j, b_j of the fundamental group and loops c_j, a'_0, a'_{g+1}.

as well as

$$c_0 = c_g = 1, \tag{2}$$

$$a_0 = a_{g+1} = a'_0 = a'_{g+1} = 1. \tag{3}$$

Then the fundamental group $\pi_1(\Sigma_g)$ has the presentation

$$\pi_1(\Sigma_g) = \langle a_1, b_1, \dots, a_g, b_g \mid c_g \rangle.$$

Let $B_0, B_1, B_2, \dots, B_g, a'_1, \dots, a'_g$ be the simple closed curves in Σ_g shown in Figure 2. Suppose that $g = 2r$. Then it is easy to check that, up to conjugation, the following equalities hold in $\pi_1(\Sigma_g)$:

$$B_{2k-1} = a_k b_k b_{k+1} \cdots b_{g+1-k} c_{g+1-k} a_{g+1-k} \quad \text{for } 1 \le k \le r; \tag{4}$$

$$B_{2k} = a_k b_{k+1} b_{k+2} \cdots b_{g-k} c_{g-k} a_{g+1-k} \quad \text{for } 0 \le k \le r; \tag{5}$$

$$a'_{k+1} = c_k a_{k+1} \quad \text{for } 0 \le k \le g-1. \tag{6}$$

If $g = 2r+1$, then B_{2k-1} satisfies the equality (4) for $1 \le k \le r+1$.

Let $A_1, \dots, A_{2g+1}$ be the simple closed curves on Σ_g shown in Figure 3. It is easily seen that, up to conjugation, the following equalities hold in $\pi_1(\Sigma_g)$:

$$A_{2k} = b_k \quad \text{for } 1 \le k \le g; \tag{7}$$

$$A_{2k+1} = a_k a_{k+1}^{-1} \quad \text{for } 0 \le k \le g. \tag{8}$$

Moreover, when we denote by $D_0, D_1, D_2, \dots, D_{2h_1}$ and E_{h_1} the simple closed curves on Σ_g indicated in Figure 3, it is immediate that, up to conjugation, the

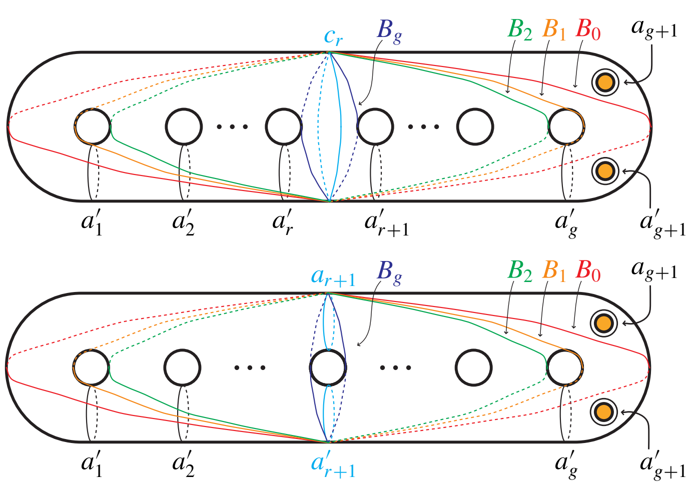

Figure 2. The curves $B_0, B_1, B_2, \dots, B_g, a'_1, \dots, a'_g$.

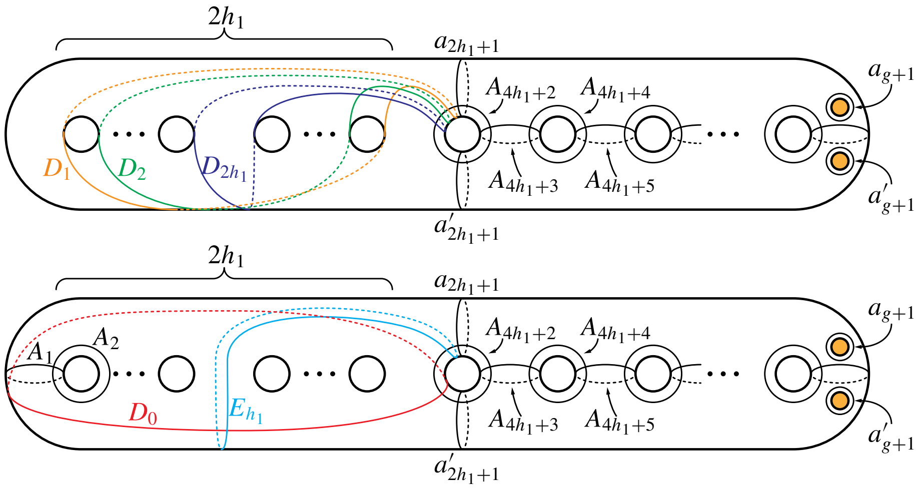

Figure 3. The curves $A_1, A_2, \dots, A_{2g+1}, D_0, D_1, \dots, D_{2h_1}$ and E_{h_1}.

following equalities hold in $\pi_1(\Sigma_g)$:

(9) $$D_0 = b_1 b_2 \cdots b_{2h_1} a_{2h_1+1}^{-1};$$

(10) $$D_{2k-1} = a_k b_k b_{k+1} \cdots b_{2h_1+1-k} c_{2h_1+1-k} a_{2h_1+1-k} a_{2h_1+1}^{-1} \quad \text{for } 1 \le k \le h_1;$$

(11) $$D_{2k} = a_k b_{k+1} b_{k+2} \cdots b_{2h_1-k} c_{2h_1-k} a_{2h_1+1-k} a_{2h_1+1}^{-1} \quad \text{for } 1 \le k \le h_1;$$

(12) $$E_{h_1} = c_{h_1} a_{2h_1+1}.$$

Note that we can modify Σ_g and $D_0, D_1, D_2, \dots, D_{2h_1}, E_{h_1}$ by isotopy as in Figure 4.

Throughout this paper, we use the same symbol for a loop and its homotopy class. Similarly, we use the same symbol for a diffeomorphism and its isotopy class, or a simple closed curve and its isotopy class. A simple loop and a simple closed curve will even be denoted by the same symbol. It will cause no confusion as it will be clear from the context which one we mean.

3. Mapping class groups

3A. ***Substitution techniques.*** Let Σ_g^b be a compact oriented surface of genus g with b boundary components. The *mapping class group* of Σ_g^b, which we denote by Mod_g^b, is the group of isotopy classes of orientation preserving self-diffeomorphisms of Σ_g^b. We assume that diffeomorphisms and isotopies fix the points of the boundary. To simplify notation, we write $\Sigma_g = \Sigma_g^0$ and $\mathrm{Mod}_g = \mathrm{Mod}_g^0$. For ϕ_1 and ϕ_2 in Mod_g^b, the notation $\phi_1\phi_2$ means that we first apply ϕ_2 then ϕ_1 (Our notation differs from that of [Korkmaz 2009].) Let t_c be the Dehn twist about

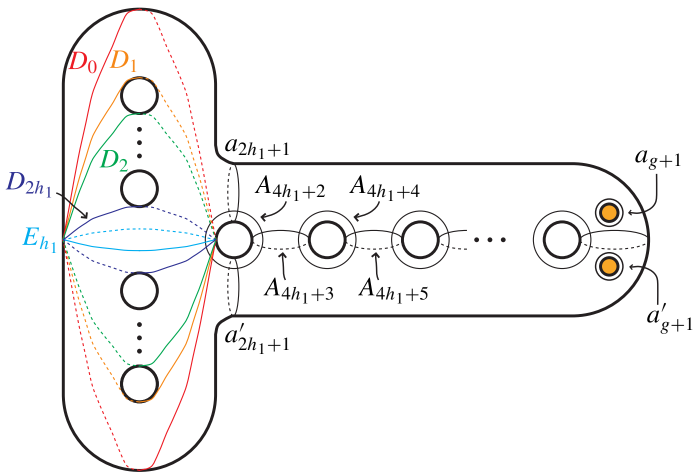

Figure 4. Modified surface Σ_g and modified curves $D_0, D_1, \ldots, D_{2h_1}$ and E_{h_1}.

a simple closed curve c in Σ_g^b. Note that $t_{\phi(c)} = \phi t_c \phi^{-1}$ for an element ϕ in Mod_g^b and $t_c t_d = t_d t_c$ if c is disjoint from d.

Definition 3.1. A word $\varrho := t_{c_1} t_{c_2} \cdots t_{c_n}$ in Mod_g is called a *positive relator* if ϱ satisfies $\varrho = 1$.

We introduce a primary technique to construct new products of right-handed Dehn twists in Mod_g^b from old ones.

Definition 3.2. Let ϕ be an element in Mod_g^b. Write

$$W = t_{c_1} t_{c_2} \cdots t_{c_k}, \qquad W^{\phi} = t_{\phi(c_1)} t_{\phi(c_2)} \cdots t_{\phi(c_k)}, \qquad V = t_{d_1} t_{d_2} \cdots t_{d_l}.$$

If the relation $V = W$ holds in Mod_g^b and $\phi(d_i) = d_i$ for all i, then by $t_{\phi(c)} = \phi t_c \phi^{-1}$ we obtain the relation

$$V = W^{\phi}.$$

in Mod_g^b. Let ϱ be a product of right-handed Dehn twists which includes V as a subword:

$$\varrho := U_1 \cdot V \cdot U_2,$$

where U_1 and U_2 are products of right-handed Dehn twists. Then we get a new product $\varsigma(\phi)$ of right-handed Dehn twists

$$\varsigma(\phi) := U_1 \cdot W^{\phi} \cdot U_2,$$

and $\varsigma(\phi)$ is said to be obtained by applying a W^{ϕ}-*substitution* of V to ϱ.

Remark 3.3. Fuller introduced the above operation for $\phi = \mathrm{id}$. Auroux [2006b; 2006a] introduced the operation to obtain $\varsigma(\phi)$ from $\varsigma(\mathrm{id})$, called a "partial conjugation" by ϕ. In a previous paper, we call the operation in Definition 3.2 a "twisted substitution". As B. Ozbagci and R. I. Baykur kindly pointed out to us, the twisted substitution is a combination of these two operations.

3B. ***The word W_2^g.*** In this section, we introduce a word W_2^g in Mod_g^2. We denote by Σ_g^2 the surface of genus g with two boundary components obtained from Σ_g by removing two disjoint open disks bounded by a_{g+1} and a'_{g+1} (cf. Figure 1 and 2), so a_{g+1} and a'_{g+1} are the boundary curves of Σ_g^2. Set

$$W_2^g := \begin{cases} (t_{B_0}t_{B_1}t_{B_2}\cdots t_{B_g}t_{c_r})^2 & \text{if } g = 2r, \\ \left(t_{B_0}t_{B_1}t_{B_2}\cdots t_{B_g}t^2_{a_{r+1}}t^2_{a'_{r+1}}\right)^2 & \text{if } g = 2r+1. \end{cases}$$

Korkmaz [2009] gave the following relation:

Lemma 3.4 [Korkmaz 2009]. *We have $t_{a_{g+1}}t_{a'_{g+1}} = W_2^g$ in Mod_g^2.*

Although Korkmaz does not prove Lemma 3.4, we can prove it by applying the same argument as in Section 2 of [Korkmaz 2001]. In Section 6A, we give a very short outline of the proof. Since the simple closed curves a_{g+1} and a'_{g+1} are null-homotopic in Σ_g, it follows that $t_{a_{g+1}} = t_{a'_{g+1}} = 1$ in Mod_g. Therefore, the word W_2^g in Mod_g is a positive relator. This positive relator for $g = 2$ was discovered by Matsumoto [1996], and its generalization was constructed independently by Cadavid [1998] and Korkmaz [2001].

4. Lefschetz pencils and fibrations

We recall the definition and basic properties of Lefschetz pencils and fibrations. More details can be found in [Gompf and Stipsicz 1999].

Definition 4.1. Let X be a closed, connected, oriented smooth 4-manifold, and let $B = \{b_1, \ldots, b_m\}$ and $C = \{p_1, \ldots, p_n\}$ be finite, disjoint subsets of X.

Let $f : X \setminus B \to S^2$ be a smooth map satisfying the following three conditions:

(a) For each $b_i \in B$, called the *base point*, there are orientation-preserving complex coordinate charts on which f is of the form $f(z_1, z_2) = z_1/z_2$.

(b) C is the set of critical points of f, and for each p_i and $f(p_i)$, there are complex local coordinate charts agreeing with the orientations of X and S^2 on which f is of the form $f(z_1, z_2) = z_1z_2$.

(c) For $q \in S^2 - f(C)$, the set $f^{-1}(q) \cup B \subset X$ is diffeomorphic to Σ_g.

Then f is called a genus-g *Lefschetz pencil* if B is a nonempty set, and f is called a genus-g *Lefschetz fibration* if B is the empty set.

The set B is called the *base locus*, and for each $q \in S^2$, the set $f(q)^{-1} \cup B$ is called the *fiber* of f. We assume that f is injective on C and that f is relatively minimal (i.e., no fiber contains a sphere with self-intersection number -1). A fiber containing a critical point is called a *singular fiber*. Each singular fiber is obtained by collapsing a simple closed curve, called the *vanishing cycle*, in the regular fiber to a point.

Once we fix an identification of Σ_g with the fiber over a base point of $S^2 - f(C)$, we can characterize the Lefschetz fibration $f : X \to S^2$ by its *monodromy representation* $\pi_1(S^2 - f(C)) \to \mathrm{Mod}_g$. Note that in this paper, this map is an antihomomorphism. Let $\gamma_1, \dots, \gamma_n$ be an ordered system of generating loops for $\pi_1(S^2 - f(C))$, such that each γ_i encircles only $f(p_i)$ and $\gamma_1\gamma_2 \cdots \gamma_n$ is homotopically trivial. Thus, since the monodromy of the fibration along each of the loops γ_i is a right-handed Dehn twist along the corresponding vanishing cycle, the monodromy of f comprises a positive relator

$$t_{v_n} \cdots t_{v_2} t_{v_1} = 1 \in \mathrm{Mod}_g,$$

where the v_i are the corresponding vanishing cycles of the singular fibers. Conversely, for any positive relator $\varrho \in \mathrm{Mod}_g$, we can construct a genus-g Lefschetz fibration over S^2 whose monodromy is ϱ. Therefore, we denote a genus-g Lefschetz fibration associated to a positive relator ϱ in Mod_g by $f_\varrho : X_\varrho \to S^2$.

Definition 4.2. For a Lefschetz fibration $f : X \to S^2$, a map $\sigma : S^2 \to X$ is called a *k-section* of f if $f \circ \sigma = \mathrm{id}_{S^2}$ and the self-intersection number of the homology class $[\sigma(S^2)]$ in $H_2(X; \mathbb{Z})$ is equal to k.

When a Lefschetz fibration $X \to S^2$ admits a section, we can compute the fundamental group of X as follows.

Lemma 4.3 (cf. [Gompf and Stipsicz 1999]). *Let ϱ be a positive relator given by $t_{v_n} \cdots t_{v_2} t_{v_1} = 1$ in Mod_g. Suppose that a genus-g Lefschetz fibration $f_\varrho : X_\varrho \to S^2$ admits a section σ. Then the fundamental group $\pi_1(X_\varrho)$ is isomorphic to the quotient of $\pi_1(\Sigma_g)$ by the normal subgroup generated by $v_1, \dots, v_n$.*

From the definitions of Lefschetz fibrations and pencils, blowing up all points of $B = \{q_1, \dots, q_b\}$ of a genus-g Lefschetz pencil yields a genus-g Lefschetz fibration with b disjoint (-1)-sections. Let $\delta_1, \delta_2, \dots, \delta_b$ be b boundary curves of Σ_g^b. Then a lift of a positive relator ϱ in Mod_g, namely $t_{v_n} \cdots t_{v_2} t_{v_1} = 1$, to Mod_g^b as

$$t_{v_n'} \cdots t_{v_2'} t_{v_1'} = t_{\delta_1} t_{\delta_2} \cdots t_{\delta_b}$$

shows the existence of b disjoint (-1)-sections of f_ϱ. Here, v_i' is a simple closed curve mapped to v_i under $\Sigma_g^b \to \Sigma_g$. Conversely, such a relation determines a genus-g Lefschetz fibration with m disjoint (-1)-sections and a genus-g Lefschetz pencil by blowing these sections down.

5. Proof of Theorem 1.2

For a finitely presented group $\Gamma = \langle x_1, x_2, \dots, x_n \mid r_1, r_2, \dots, r_k \rangle$ with n generators and k relators, let $l = \max\{l(r_i) \mid 1 \le i \le k\}$, where $l(r_i)$ is the syllable length of r_i. In this section, we denote by h_1 and h_2 two integers satisfying $h_1 \ge n + l - 1$ and $2(h_2 - 1) \ge k$, respectively.

5A. *Construction of a word $W_2^g(1, \psi_i)$.* In this subsection, we construct a key relation in Mod_g^2.

Let us consider Σ_g^2 obtained from Σ_g by removing two disjoint open disks surrounded by a_{g+1} and a'_{g+1} (see Section 2 and Figures 1–3). Write $r = 2h_1 + h_2 - 1$ and $g = 2r$ or $2r + 1$. For $h_2 - 1 \ge 1$, we set

$$\begin{aligned} X &= t_{A_{4h_1+2}} t_{A_{4h_1+3}} \cdots t_{A_{2r}}, \\ \overline{X} &= t_{A_{2r}} \cdots t_{A_{4h_1+3}} t_{A_{4h_1+2}}, \\ Y &= (t_{D_0} t_{D_1} \cdots t_{D_{2h_1}})^2. \end{aligned}$$

Moreover, we define words V_1 and V_2 to be

$$\begin{aligned} V_1 &= t_{E_{h_1}} X t_{a_r} t_{a'_r} \overline{X} t_{E_{h_1}} t_{a'_r} \overline{X} Y X t_{a'_r}, \\ V_2 &= t_{E_{h_1}} X t_{a_r} t_{a'_r} \overline{X} t_{E_{h_1}} t_{A_{2r+1}} \overline{X} Y X t_{A_{2r+1}}. \end{aligned}$$

Then we obtain the relations in the following proposition.

Proposition 5.1. *We have $t_{c_r} = V_1$ and $t_{a_{r+1}} t_{a'_{r+1}} = V_2$ in Mod_g^2.*

We postpone the proof of Proposition 5.1 until Section 6 (see Proposition 6.1).

Let $h_1 \ge n + l - 1$ and $2(h_2 - 1) \ge k$. The next proposition is needed to prove Theorem 1.2.

Proposition 5.2. *Let F_n be the subgroup of $\pi_1(\Sigma_g)$ generated by the generators $a_1, \dots, a_n$, i.e., F_n is a free group of rank n. Let $r_1, \dots, r_k$ be k elements in F_n represented as words in $a_1, \dots, a_n$. Let $l = \max_{1 \le i \le k}\{l(r_i)\}$, where $l(r_i)$ is the syllable length of r_i. Then there are simple loops $R_1, \dots, R_k$ in Σ_g (see Figure 5) with the property that, for $4h_1 + 2 \le j \le 4h_1 + 2h_2 - 2$ and $1 \le i \le k$,*

(a) *R_i is disjoint from $A_{2h_1+1}, \dots, A_{4h_1}, c_{2h_1+h_2-1} (= c_r)$.*

(b) *R_1 intersects $a_{2h_1+h_2-1}$ at one point and does not intersect A_j for any j.*

(c) *R_i intersects $A_{4h_1+2h_2-i}$ at one point and intersects neither $a_{2h_1+h_2-1}$ nor A_j for any $j \ne 4h_1 + 2h_2 - i$ and $i \ge 2$.*

(d) *$\Phi([R_i]) = r_i$, where $[R_i] \in \pi_1(\Sigma_g)$ is the homotopy class of the loop R_i, and $\Phi : \pi_1(\Sigma_g) \to \pi_1(\Sigma_n)$ is the map defined by $\Phi(a_m) = a_m$ for $1 \le m \le n$ and $\Phi(\alpha) = 1$ for $\alpha \in \{a_{n+1}, \dots, a_g, b_1, \dots, b_g\}$.*

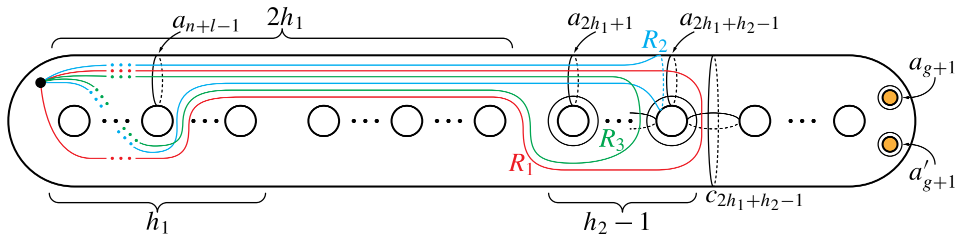

Figure 5. Curves $R_1, \dots, R_k$ in Σ_g.

In Section 7, we prove Proposition 5.2 by constructing simple loops $R_1, \dots, R_k$ explicitly. We also consider the loops $R_1, \dots, R_k$ as simple loops on Σ_g^2 by removing two disjoint open disks surrounded by a_{g+1}, a'_{g+1} from Σ_g (see Figure 5).

For $i = 0, 1, \dots, k$, we define an element ψ_i in Mod_g^2 to be

$$\psi_0 = t_{a_{h_1}} t_{b_{h_1+1}} t_{b_{h_1+2}} \cdots t_{b_{2h_1}},$$
$$\psi_i = t_{R_{k+1-i}} t_{R_{k+2-i}} \cdots t_{R_k} \psi_0,$$

where the R_i are the loops on Σ_g^2 described above. From Proposition 5.2, for each i, we see that $\psi_i(c_r) = c_r$ if $g = 2r$, while $\psi_1(a_{r+1}) = a_{r+1}$ and $\psi_1(a'_{r+1}) = a'_{r+1}$ if $g = 2r+1$.

If $g = 2r$, then we can find two t_{c_r} in the word W_2^g. By Proposition 5.1, we can apply V_1^{id}-substitution for one t_{c_r} and $V_1^{\psi_i}$-substitution for the other.

If $g = 2r+1$, then since $t_{a_{r+1}}^2 t_{a'_{r+1}}^2 = (t_{a_{r+1}} t_{a'_{r+1}})^2$, we can find four $t_{a_{r+1}} t_{t_{a'_{r+1}}}$ in the word W_2^g. By Proposition 5.1, we can apply V_2^{id}-substitution for one $t_{a_{r+1}} t_{a'_{r+1}}$ and $V_2^{\psi_i}$-substitution for the other.

If we set

$$W_2^g(1, \psi_i) := (t_{B_0} t_{B_1} t_{B_2} \cdots t_{B_g} V_1)(t_{B_0} t_{B_1} t_{B_2} \cdots t_{B_g} V_1^{\psi_i})$$

if $g = 2r$, and

$$W_2^g(1, \psi_i) := (t_{B_0} t_{B_1} t_{B_2} \cdots t_{B_g} t_{a_{r+1}} t_{a'_{r+1}} V_2)(t_{B_0} t_{B_1} t_{B_2} \cdots t_{B_g} t_{a_{r+1}} t_{a'_{r+1}} V_2^{\psi_i})$$

if $g = 2r+1$, then we get the next lemma.

Lemma 5.3. *We have* $t_{a_{g+1}} t_{a'_{g+1}} = W_2^g(1, \psi_i)$ *in* Mod_g^2.

Since $t_{a_{g+1}} = 1$ and $t_{a'_{g+1}} = 1$ in Mod_g, the word $W_2^g(1, \psi_i)$ in Mod_g is a positive relator. Therefore, we obtain a genus-g Lefschetz fibration $f_{W_2^g(1,\psi_i)}$ with two disjoint (-1)-sections (and genus-g Lefschetz pencil with two base points corresponding to $W_2^g(1, \psi_i)$). Then, we have the following results which we prove in Section 5B and in Section 5C.

Theorem 5.4. *Suppose that $k = 0$. We denote by F_n a free group of rank n. If $g \geq 2(2n+1)$, then we have*

$$\pi_1(X_{W_2^g(1,\psi_0)}) \cong F_n.$$

Theorem 5.5. *Suppose that $k > 0$. If $g \geq 4(n+l-1)+k$, then we have*

$$\pi_1(X_{W_2^g(1,\psi_k)}) \cong \Gamma.$$

Combining Theorem 5.4 and 5.5, we obtain Theorem 1.2.

5B. ***Proof of Theorem 5.4.*** In this section, we prove Theorem 5.4. We begin with a lemma.

Lemma 5.6. *Let $r = 2h_1 + h_2 - 1$. Let $\langle S \rangle$ be the normal closure of the elements of the set S of simple closed curves on Σ_g defined by*

$$S = \{B_0, B_1, \ldots, B_g, D_0, D_1, \ldots, D_{2h_1}, E_{h_1}, A_{4h_1+2}, \ldots, A_{2r}, a_r, a_r'\}$$

if $g = 2r$, and by

$$S = \{B_0, B_1, \ldots, B_g, a_{r+1}, a_{r+1}', D_0, D_1, \ldots, D_{2h_1}, E_{h_1}, A_{4h_1+2}, \ldots, A_{2r+1}, a_r, a_r'\}$$

if $g = 2r+1$. Then $\pi_1(\Sigma_g)/\langle S \rangle$ has a presentation with generators $a_1, b_1, \ldots, a_g, b_g$ and with relations

$$\begin{aligned} a_i a_{g+1-i} = b_i a_{g+1-i} b_{g+1-i} a_{g+1-i}^{-1} &= 1 \quad \text{for } 1 \leq i \leq r, \\ a_{2h_1+k} = b_{2h_1+k} &= 1 \quad \text{for } 1 \leq k \leq h_2 - 1, \\ a_j a_{2h_1+1-j} = b_j a_{2h_1+1-j} b_{2h_1+1-j} a_{2h_1+1-j}^{-1} &= 1 \quad \text{for } 1 \leq j \leq h_1, \\ c_{h_1} &= 1 \end{aligned}$$

if $g = 2r$, and

$$\begin{aligned} a_i a_{g+1-i} = b_i a_{g+1-i} b_{g+1-i} a_{g+1-i}^{-1} &= 1 \quad \text{for } 1 \leq i \leq r, \\ a_{2h_1+k} = b_{2h_1+k} &= 1 \quad \text{for } 1 \leq k \leq h_2 - 1, \\ a_j a_{2h_1+1-j} = b_j a_{2h_1+1-j} b_{2h_1+1-j} a_{2h_1+1-j}^{-1} &= 1 \quad \text{for } 1 \leq j \leq h_1, \\ a_{r+1} = c_{h_1} &= 1 \end{aligned}$$

if $g = 2r+1$.

Proof. Suppose that $g = 2r$. From the equalities (4) and (5) in Section 2, in $\pi_1(\Sigma_g)/\langle S \rangle$ we have

$$a_i a_{g+1-i} = 1. \tag{13}$$

This gives

$$
\begin{aligned}
1 = B_{2i-1} &= b_i b_{i+1} \cdots b_{g+1-i} c_{g+1-i} &&\text{for } 1 \le i \le r,\\
1 = B_{2i} &= b_{i+1} b_{i+2} \cdots b_{g-i} c_{g-i} &&\text{for } 1 \le i \le r
\end{aligned}
$$

in $\pi_1(\Sigma_g)/\langle S\rangle$. From these two equalities, we have $b_i c_{g-i}^{-1} b_{g+1-i} c_{g+1-i} = 1$ for each $1 \le i \le r$ and

$$c_r = 1. \tag{14}$$

Note that $c_{g+1-i} = b_{g+1-i}^{-1} c_{g-i} (a_{g+1-i} b_{g+1-i} a_{g+1-i}^{-1})$ from the equality (1). Therefore, by $b_i c_{g-i}^{-1} b_{g+1-i} c_{g+1-i} = 1$, we obtain

$$b_k a_{g+1-i} b_{g+1-i} a_{g+1-i}^{-1} = 1. \tag{15}$$

From $a_r = 1$, $A_l = 1$ for $4h_1 + 2 \le l \le 2r$ and the equalities (7) and (8), we obtain

$$a_{2h_1+k} = b_{2h_1+k} = 1 \tag{16}$$

for $1 \le k \le h_2 - 1$. From $a'_r = 1$ and the equalities (6), (14), (1) and (16), we have

$$c_{r-1} = c_{2h_1} = 1. \tag{17}$$

By $a_{2h_1+1} = 1$, $c_{2h_1} = 1$ and the equalities (9), (10) and (11), an argument similar to the proofs of the relations (13) and (15) gives

$$a_j a_{2h_1+1-j} = b_j a_{2h_1+1-j} b_{2h_1+1-j} a_{2h_1+1-j}^{-1} = 1 \quad \text{and} \quad c_{h_1} = 1 \tag{18}$$

for $1 \le j \le 2h_1$.

From the equalities (13), (14), (15), (16), (17) and (18), we see that $\pi_1(\Sigma_g)/\langle S\rangle$ has a presentation with generators $a_1, b_1, \dots, a_g, b_g$ and with relations

$$
\begin{aligned}
a_i a_{g+1-i} = b_i a_{g+1-i} b_{g+1-i} a_{g+1-i}^{-1} &= 1 &&\text{for } 1 \le i \le r,\\
a_{2h_1+k} = b_{2h_1+k} &= 1 &&\text{for } 1 \le k \le h_2 - 1,\\
a_j a_{2h_1+1-j} = b_j a_{2h_1+1-j} b_{2h_1+1-j} a_{2h_1+1-j}^{-1} &= 1 &&\text{for } 1 \le j \le h_1,\\
c_g = c_r = c_{r-1} = c_{2h_1} = c_{h_1} &= 1.
\end{aligned}
$$

Then by the equalities (1), (16) and (18), we can delete from the above the relations $c_g = c_r = c_{r-1} = c_{2h_1} = 1$. This is our claim.

Suppose now that $g = 2r + 1$. Since $a_{r+1} = a'_{r+1} = 1$ and $a'_{r+1} = c_r a_{r+1}$, we have $c_r = 1$. A similar argument as in the case $g = 2r$ shows that $\pi_1(\Sigma_g)/\langle S\rangle$ has the desired presentation. This completes the proof. □

We can now prove Theorem 5.4.

Proof of Theorem 5.4. Let $h_1 \geq n$ and $h_2 - 1 \geq 1$. For simplicity of notation, we write G instead of $\pi_1(X_{W_2^g(1,\psi_0)})$.

Suppose that $g = 2(2h_1 + h_2 - 1)$ and let $r = 2h_1 + h_2 - 1$. Note that G has a presentation with generators $a_1, b_1, \dots, a_g, b_g$ and with relations

$$\begin{aligned}
c_g &= 1, \\
B_i &= 1 \quad \text{for } 0 \leq i \leq g, \\
a_r = a_r' = E_{h_1} &= 1, \\
D_j = A_k &= 1 \quad \text{for } 0 \leq j \leq 2h_1,\ 4h_1 + 2 \leq k \leq 4h_1 + 2h_2 - 2, \\
\psi_0(a_r) = \psi_0(a_r') = \psi_0(E_{h_1}) &= 1, \\
\psi_0(D_j) = \psi_0(A_k) &= 1 \quad \text{for } 0 \leq j \leq 2h_1,\ 4h_1 + 2 \leq k \leq 4h_1 + 2h_2 - 2.
\end{aligned}$$

It is easily seen that, up to conjugation, we have the equalities

$$\begin{aligned}
\psi_0(D_0) &= a_{h_1} \cdots a_{n+2} a_{n+1} D_0, \\
\psi_0(D_{2l-1}) &= b_{2h_1-l+1}^{-1} a_{h_1} \cdots a_{n+2} a_{n+1} D_{2l-1} \quad \text{for } 1 \leq l \leq n, \\
\psi_0(D_{2l}) &= b_{2h_1-l+1}^{-1} a_{h_1} \cdots a_{n+2} a_{n+1} D_{2l} \qquad \text{for } 1 \leq l \leq n
\end{aligned}$$

in $\pi_1(\Sigma_g)$. Thus, by $D_0 = \psi_0(D_0) = D_j = \psi_0(D_j) = 1$ for $1 \leq j \leq 2h_1$, we obtain

$$b_{2h_1-l+1} = 1 \quad \text{for } 1 \leq l \leq n.$$

Similarly, we have the following equalities (up to conjugation) in $\pi_1(\Sigma_g)$:

$$\begin{aligned}
\psi_0(D_{2l-1}) &= b_{2h_1-l+1}^{-1} a_{h_1} \cdots a_{l+1} a_l D_{2l-1} \qquad \text{for } n+1 \leq l \leq r-1, \\
\psi_0(D_{2l}) &= b_{2h_1-l+1}^{-1} a_{h_1} \cdots a_{l+2} a_{l+1} D_{2l-1} \quad \text{for } n+1 \leq l \leq r-1, \\
\psi_0(D_{2h_1-1}) &= b_{h_1+1}^{-1} a_{h_1} D_{2h_1-1}, \\
\psi_0(D_{2h_1}) &= b_{h_1+1}^{-1} B_{2h_1}.
\end{aligned}$$

By $D_j = 1$ for $1 \leq j \leq 2h_1$ and $\psi_0(D_{2l-1}) = \psi_0(D_{2l}) = 1$ for $n+1 \leq l \leq h_1$, we obtain

$$a_l = 1 \quad \text{for } n+1 \leq l \leq h_1.$$

Moreover, by $\psi_0(D_{2l}) = \psi_0(D_{2l+1}) = \psi_0(D_{2h_1}) = 1$ for $n+1 \leq l \leq h_1 - 1$, we have

$$b_{2h_1-l+1} = 1 \quad \text{for } n+1 \leq l \leq h_1.$$

Here, since $\psi_0(a_r) = a_r$, $\psi_0(a_r') = a_r'$, $\psi_0(E_{h_1}) = E_{h_1}$ and $\psi_0(A_k) = A_k$ in $\pi_1(\Sigma_g)$ for each $4h_1 + 2 \leq k \leq 4h_1 + 2h_2 - 2$, we can delete the relations $\psi_0(a_r) = 1$, $\psi_0(a_r') = 1$, $\psi_0(E_{h_1}) = 1$ and $\psi_0(A_K) = 1$ from the above presentation of G.

From the above arguments and Lemma 5.6, we see that G has a presentation with generators $a_1, b_1, \dots, a_g, b_g$ and with relations

$$\begin{aligned} a_i a_{g+1-i} &= b_i a_{g+1-i} b_{g+1-i} a_{g+1-i}^{-1} && \text{for } 1 \le i \le r, \\ a_{2h_1+k} &= b_{2h_1+k} = 1 && \text{for } 1 \le k \le h_2 - 1, \\ a_j a_{2h_1+1-j} &= b_j a_{2h_1+1-j} b_{2h_1+1-j} a_{2h_1+1-j}^{-1} = 1 && \text{for } 1 \le j \le h_1, \\ c_{h_1} &= 1, \\ a_{n+1} &= a_{n+2} = \cdots = a_{h_1} = 1, \\ b_{h_1} &= b_{h_1+1} = \cdots = b_{2h_1} = 1. \end{aligned}$$

It is easily shown that this is a presentation of the free group of rank n with free basis $a_1, \dots, a_n$, that is, G is isomorphic to F_n.

The proof for $g = 2r + 1$ is similar. This completes the proof of Theorem 5.4. □

5C. *Proof of Theorem 5.5.* We now prove Theorem 5.5. The proof is inspired by [Korkmaz 2009] and that of Proposition 13 in [Akhmedov and Ozbagci 2013]. For simplicity, we write G' instead of $\pi_1(X_{W_2^g(1,\psi_1)})$.

Proof of Theorem 5.5. Suppose that $g = 2(2h_1 + h_2 - 1)$. Since R_1 intersects $a_{2h_1+h_2-1}$ at one point and does not intersect A_j for $j = 4h_1+2, \dots, 4h_1+2h_2-2$, and $a_{2h_1+h_2-1}$ is disjoint from $a_{n+1}, \dots, a_{h_1}, b_{h_1+1}, \dots, b_{2h_1}$ and $R_2, \dots, R_k$, we see that in $\pi_1(\Sigma_g)$, up to conjugation,

$$\psi_k(a_{2h_1+h_2-1}) = t_{R_1}(a_{2h_1+h_2-1}) = a_{2h_1+h_2-1} R_1^{\epsilon},$$

where ϵ is equal to 1 or -1. Since $a_{2h_1+h_2-1} = 1$ in G', we may replace the relator $\psi_k(a_{2h_1+h_2-1}) = 1$ by $R_1 = 1$.

Let c be an element of the set of the vanishing cycles of $f_{W_2^g(1,\psi_k)}$. If R_1 is disjoint from $\psi_{k-1}(c)$, then we have $\psi_k(c) = t_{R_1}(\psi_{k-1}(c)) = \psi_{k-1}(c)$. If R_1 intersects $\psi_{k-1}(c)$ at t points, then it is easily seen that there are elements $x_1, \dots, x_{t+1}$ in $\pi_1(\Sigma_g)$ such that $\psi_{k-1}(c) = x_1 x_2 \cdots x_{t+1}$ and that

$$t_{R_1}(\psi_{k-1}(c)) = x_1 R_1^{\zeta_1} x_2 R_1^{\zeta_2} \cdots x_t R_1^{\zeta_t} x_{t+1}$$

(up to conjugacy), where each ζ_s is equal to 1 or -1. From $R_1 = 1$, we obtain $\psi_k(c) = t_{R_1}(\psi_{k-1}(c)) = \psi_{k-1}(c)$ in G'. Therefore, we may replace the relator $\psi_k(c) = 1$ by $\psi_{k-1}(c) = 1$.

By repeating this argument for each $i = k-1, \dots, 1$, we see that we may replace the relators $\psi_k(A_{4h_1+2h_2-(k+1-i)}) = 1$ and $\psi_k(c) = 1$ by $R_{k+1-i} = 1$ and $\psi_0(c) = 1$, respectively. In particular, since for each $j = 4h_1+2, \dots, 4h_1+2h_2-2$, $a_{2h_1+h_2-1} = 1$ and $A_j = 1$ in G' and $a_{2h_1+h_2-1} = \psi_0(a_{2h_1+h_2-1})$ and $A_j = \psi_0(A_j)$ in $\pi_1(\Sigma_g)$ (up to conjugation), we can delete the relators $\psi_k(a_{2h_1+h_2-1}) = 1$ and

$\psi_k(A_j) = 1$. Therefore, from the proof of Theorem 5.4, we see that G' has a presentation with generators $a_1, b_1, \dots, a_g, b_g$ and with relations

$$\begin{aligned} a_i a_{g+1-i} &= b_i a_{g+1-i} b_{g+1-i} a_{g+1-i}^{-1} && \text{for } 1 \le i \le r, \\ a_{2h_1+k} &= b_{2h_1+k} = 1 && \text{for } 1 \le k \le h_2 - 1, \\ a_j a_{2h_1+1-j} &= b_j a_{2h_1+1-j} b_{2h_1+1-j} a_{2h_1+1-j}^{-1} = 1 && \text{for } 1 \le j \le h_1, \\ c_{h_1} &= 1, \\ a_{n+1} &= a_{n+2} = \cdots = a_{h_1} = 1, \\ b_{h_1} &= b_{h_1+1} = \cdots = b_{2h_1} = 1, \\ R_1 &= R_2 = \cdots = R_k = 1. \end{aligned}$$

We note that the element $[R_i] \in \pi_1(\Sigma_g)$ is contained in the subgroup generated by $a_1, b_1, \dots, a_{h_1}, b_{h_1}$ and $a_{2h_1+1}, b_{2h_1+1}, \dots, a_{2h_1+h_2-1}, b_{2h_1+h_2-1}$. Since from this presentation, we see that $a_s = 1$ for $s = n+1, \dots, h_1, 2h_1+1, \dots, 2h_1+h_2-1$ and $b_j = 1$ for $j = 1, \dots, h_1, 2h_1+1, \dots, 2h_1+h_2-1$, we get a word representing the element r_i by Proposition 5.2. Therefore, G' is isomorphic to Γ.

A similar argument works for $g = 2(2h_1+h_2-1)+1$. Since $f_{W_2^g(1,\psi_k)}$ has at least two disjoint (-1)-sections, by blowing down one of them we obtain the required genus-g Lefschetz pencil. This completes the proof of Theorem 5.5 and therefore, as discussed in Section 5A, also of Theorem 1.2. $\square$

6. Construction of a lift of Gurtas' positive relator

In this section, we prove Proposition 5.1 and give a lift to Mod_g^2 of the positive relator in Mod_g given by Gurtas [2004].

6A. ***Outline of the proof of Lemma 3.4.*** We now give an outline of the proof of Lemma 3.4, which is needed to prove Proposition 5.1.

Outline of the proof of Lemma 3.4. We define $\Delta_0 = \overline{\Delta}_0 = 1$. Moreover, for each $k = 1, \dots, 2g+1$, we define Δ_k and $\overline{\Delta}_k$ to be the words

$$\Delta_k = t_{A_1} t_{A_2} \cdots t_{A_k} \quad \text{and} \quad \overline{\Delta}_k = t_{A_k} \cdots t_{A_2} t_{A_1}.$$

For each $k = 0, 1, \dots, g$, the words β_k and β are defined by

$$\beta_k = \overline{\Delta}_k \Delta_{2g+1-k} \Delta_{2g-k}^{-1} \overline{\Delta}_k^{-1} \quad \text{and} \quad \beta = \overline{\Delta}_g^{g+1}.$$

Then by applying the argument from Section 2 of [Korkmaz 2001] with σ_i (which is the standard generator of the braid group B_{2g+2} on $2g+2$ strings) replaced by t_{A_i}, we have the relation

$$\beta_0 \beta_1 \beta_2 \cdots \beta_g \beta^2 = \Delta_{2g+1} \Delta_{2g} \cdots \Delta_3 \Delta_2 \Delta_1. \tag{19}$$

It is easy to check that $\overline{\Delta}_k\Delta_{2g-k}(A_{2g+1-k}) = B_k$. This gives

$$t_{B_k} = (\overline{\Delta}_k\Delta_{2g-k})t_{A_{2g+1-k}}(\overline{\Delta}_k\Delta_{2g-k})^{-1} = \overline{\Delta}_k\Delta_{2g+1-k}\Delta_{2g-k}^{-1}\overline{\Delta}_k^{-1} = \beta_k.$$

Therefore, from the relation (19), we have

$$t_{B_0}t_{B_1}t_{B_2}\cdots t_{B_g}(\overline{\Delta}_g)^{2g+2} = \Delta_{2g+1}\Delta_{2g}\cdots\Delta_3\Delta_2\Delta_1.$$

Using the chain relations $\overline{\Delta}_g^{2g+2} = t_{c_r}$ when $g = 2r$ and $\overline{\Delta}_g^{g+1} = t_{a_{r+1}}t_{a'_{r+1}}$ when $g = 2r+1$, we have

$$\Delta_{2g+1}\Delta_{2g}\cdots\Delta_3\Delta_2\Delta_1 = \begin{cases} t_{B_0}t_{B_1}t_{B_2}\cdots t_{B_g}t_{c_r} & \text{for } g = 2r, \\ t_{B_0}t_{B_1}t_{B_2}\cdots t_{B_g}t_{a_{r+1}}t_{a'_{r+1}} & \text{for } g = 2r+1. \end{cases} \tag{20}$$

If we prove that $t_{a_{g+1}}t_{a'_{g+1}} = (\Delta_{2g+1}\Delta_{2g}\cdots\Delta_3\Delta_2\Delta_1)^2$ in Mod_g^2, the assertion follows. Note that *the chain relation* $\Delta_{2g+1}^{2g+2} = t_{a_{g+1}}t_{a'_{g+1}}$, and $t_{A_k}\Delta_m = \Delta_m t_{A_{k-1}}$ if $1 < k \le m$ (see [Korkmaz 2001, Lemma 2.1(a)]), hold in Mod_g^2. Then we have

$$\begin{aligned}
\Delta_{2g+1}^{2g+2} &= \Delta_{2g+1}\Delta_{2g}t_{A_{2g+1}}\Delta_{2g+1}\Delta_{2g+1}^{2g-1} \\
&= \Delta_{2g+1}\Delta_{2g}\Delta_{2g+1}t_{A_{2g}}\Delta_{2g+1}^{2g-1} \\
&= \Delta_{2g+1}\Delta_{2g}\Delta_{2g-1}(t_{A_{2g}}t_{A_{2g+1}})t_{A_{2g}}\Delta_{2g+1}^{2g-1} \\
&= \Delta_{2g+1}\Delta_{2g}\Delta_{2g-1}\Delta_{2g+1}(t_{A_{2g-1}}t_{A_{2g}})t_{A_{2g-1}}\Delta_{2g+1}^{2g-2} \\
&= \Delta_{2g+1}\Delta_{2g}\Delta_{2g-1}\Delta_{2g-2}(t_{A_{2g-1}}t_{A_{2g}}t_{A_{2g+1}})(t_{A_{2g-1}}t_{A_{2g}})t_{A_{2g-1}}\Delta_{2g+1}^{2g-2} \\
&\;\;\vdots \\
&= \Delta_{2g+1}\Delta_{2g}\cdots\Delta_1(t_{A_2}t_{A_3}\cdots t_{A_{2g+1}})(t_{A_2}t_{A_3}\cdots t_{A_{2g}})\cdots(t_{A_2}t_{A_3})t_{A_2}\Delta_{2g+1} \\
&= \Delta_{2g+1}\Delta_{2g}\cdots\Delta_1\Delta_{2g+1}\Delta_{2g}\cdots\Delta_1,
\end{aligned}$$

and the proof is complete. □

6B. *Proof of Proposition 5.1.* In this section, we prove Proposition 6.1 instead of Proposition 5.1. Note that if we set $g = r$ in the notation of Proposition 6.1 and consider an embedding $\Sigma_r^2 \hookrightarrow \Sigma_g^2$ (resp. $\Sigma_r^1 \hookrightarrow \Sigma_g^2$) mapping (a_{r+1}, a'_{r+1}) (resp. a_{r+1}) in Proposition 6.1 to (a_{r+1}, a'_{r+1}) (resp. c_r) in Proposition 5.1, then we get Proposition 5.1. Therefore, it is sufficient to prove Proposition 6.1.

Proposition 6.1. *Let Σ_g^2 (resp. Σ_g^1) be the compact oriented surface of genus g with two boundary components (resp. one boundary component) obtained from Σ_g by removing two disjoint open disks (resp. one open disk). Let $a_{g+1}, a'_{g+1} = c_g a_{g+1}$ (resp. a_{g+1}) be the boundary curves of Σ_g^2 (resp. the boundary curve of Σ_g^1). Then* *the relations*

$$t_{a_{g+1}}t_{a'_{g+1}} = t_{E_{h_1}}t_{A_{4h_1+2}}\cdots t_{A_{2g}}t_{a_g}t_{a'_g}t_{A_{2g}}\cdots t_{A_{4h_1+2}}t_{E_{h_1}} \cdot t_{A_{2g+1}}t_{A_{2g}}\cdots t_{A_{4h_1+2}}\cdot(t_{D_0}t_{D_1}\cdots t_{D_{2h_1}})^2\cdot t_{A_{4h_1+2}}\cdots t_{A_{2g}}t_{A_{2g+1}}, \tag{21}$$

$$t_{a_{g+1}} = t_{E_{h_1}}t_{A_{4h_1+2}}\cdots t_{A_{2g}}t_{a_g}t_{a'_g}t_{A_{2g}}\cdots t_{A_{4h_1+2}}t_{E_{h_1}} \cdot t_{a'_g}t_{A_{2g}}\cdots t_{A_{4h_1+2}}\cdot(t_{D_0}t_{D_1}\cdots t_{D_{2h_1}})^2\cdot t_{A_{4h_1+2}}\cdots t_{A_{2g}}t_{a'_g} \tag{22}$$

hold in Mod_g^2 *and* Mod_g^1, *respectively.*

In order to prove Proposition 6.1, we prepare Lemma 6.2 and Proposition 6.3.

Lemma 6.2. *Suppose that $g = 2r$. In the notation of Lemma 3.4, let c'_r be the separating simple closed curve defined by $a_{g+1}(b_{r+1}\cdots b_g)a'_{g+1}(b_{r+1}\cdots b_g)^{-1}c_r$ (cf. Figure 6(a)). We modify Σ_g^2 and $B_0,\dots,B_g$, c_r, c'_r by isotopy as shown in Figure 6(b) and (c). Then in Mod_g^2, the following relation holds*:

$$t_{a_{g+1}}t_{a'_{g+1}} = t_{c_r}t_{c'_r}(t_{B_0}t_{B_1}\cdots t_{B_g})^2.$$

Proof. It is easily seen that for each $i = 1,\dots,g$, we have

$$\Delta_{2g+1}\cdots\Delta_2\Delta_1(A_i) = A_{2g+2-i}.$$

This gives the relation

$$\Delta_{2g+1}\cdots\Delta_2\Delta_1 t_{A_i} = t_{A_{2g+i}}\Delta_{2g+1}\cdots\Delta_2\Delta_1$$

for each $i = 1,\dots,2r$. Therefore, we have

$$\Delta_{2g+1}\cdots\Delta_2\Delta_1(\overline{\Delta}_g)^{-(2g+2)} = (t_{A_{g+2}}\cdots t_{A_{2g+1}})^{-(2g+2)}\Delta_{2g+1}\cdots\Delta_2\Delta_1.$$

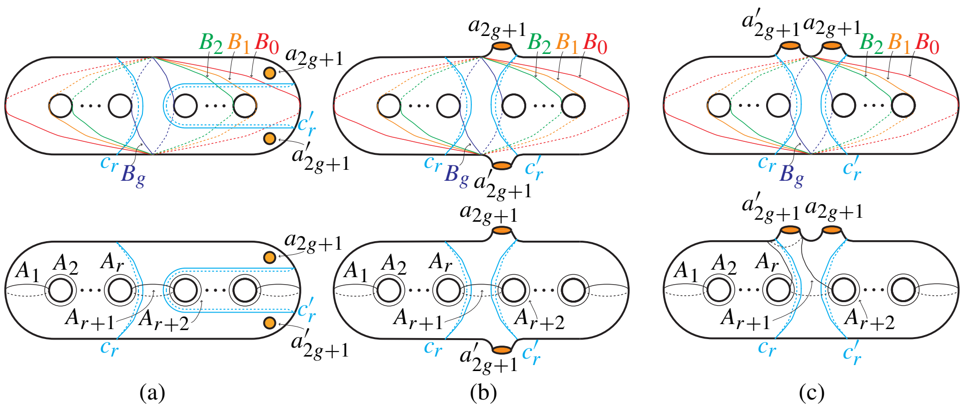

Figure 6. Modified surface Σ_g^2 and curves $B_0,\dots,B_g$, c_r, c'_r.

Since

$$t_{B_0}t_{B_1}t_{B_2}\cdots t_{B_g}(\overline{\Delta}_g)^{2g+2}=\Delta_{2g+1}\cdots\Delta_2\Delta_1\ (=t_{B_0}t_{B_1}t_{B_2}\cdots t_{B_g}t_{c_r})$$

from the proof of Lemma 3.4, we have

$$(t_{A_{g+2}}\cdots t_{A_{2g+1}})^{2g+2}t_{B_0}t_{B_1}t_{B_2}\cdots t_{B_g}=\Delta_{2g+1}\cdots\Delta_2\Delta_1 \\ (=t_{B_0}t_{B_1}t_{B_2}\cdots t_{B_g}t_{c_r}).$$

By the chain relation, we obtain $t_{c'_r}=(t_{A_{g+2}}\cdots t_{A_{2g+1}})^{2g+2}$. Therefore,

$$t_{a_{g+1}}t_{a'_{g+1}}=t_{c'_r}t_{B_0}t_{B_1}\cdots t_{B_g}\cdot t_{B_0}t_{B_1}\cdots t_{B_g}t_{c_r}$$

follows by Lemma 3.4. By conjugation by t_{c_r}, we have

$$t_{a_{g+1}}t_{a'_{g+1}}=t_{c_r}t_{c'_r}(t_{B_0}t_{B_1}\cdots t_{B_g})^2. \qquad \square$$

Proposition 6.3 was shown by Hamada [≥ 2016] based on the argument of [Tanaka 2012]. Its statement concerns a'_0, a null-homotopic simple closed curve in Σ_g defined by $a'_0=c_0a_0$.

Proposition 6.3 [Hamada ≥ 2016]. *Let Σ_g^4 be the compact oriented surface of genus g with four boundary components obtained from Σ_g by removing four disjoint open disks surrounded by a_0, a'_0, a_{g+1} and a'_{g+1}. Then the following relation in* Mod_g^4 *holds*:

$$t_{a_0}t_{a'_0}t_{a_{g+1}}t_{a'_{g+1}}=t_{A_{2g+1}}\cdots t_{A_2}t_{a_1}t_{a'_1}t_{A_2}\cdots t_{A_{2g+1}}\cdot t_{A_1}\cdots t_{A_{2g}}t_{a_g}t_{a'_g}t_{A_{2g}}\cdots t_{A_1}.$$

Proof. The proof is by induction on the genus.

Suppose that $g=1$. The *four-holed torus relation*,

$$t_{a_0}t_{a'_0}t_{a_2}t_{a'_2}=(t_{A_1}t_{A_3}t_{A_2}t_{a_1}t_{a'_1}t_{A_2})^2,$$

was constructed by Korkmaz and Ozbagci [2008, Section 3.4]. Since a_0, a'_0, a_2, a'_2 are disjoint from A_1 and A_1 is disjoint from A_3, by conjugation by t_{A_1}, we have

$$\begin{aligned}t_{a_0}t_{a'_0}t_{a_2}t_{a'_2}&=t_{A_3}t_{A_2}t_{a_1}t_{a'_1}t_{A_2}t_{A_1}\cdot t_{A_3}t_{A_2}t_{a_1}t_{a'_1}t_{A_2}t_{A_1}\\&=t_{A_3}t_{A_2}t_{a_1}t_{a'_1}t_{A_2}t_{A_3}\cdot t_{A_1}t_{A_2}t_{a_1}t_{a'_1}t_{A_2}t_{A_1}.\end{aligned}$$

Hence, the conclusion of the proposition holds for $g=1$.

Next we assume, inductively, that the relation holds in Mod_{g-1}^4. Since then a_0, a'_0, a_g, a'_g are disjoint from $A_1,\ldots,A_{2g-1}$, we have the relation

$$t_{a_0}t_{a'_0}t_{a_g}t_{a'_g}=t_{A_{2g-2}}\cdots t_{A_1}\cdot t_{A_{2g-1}}\cdots t_{A_2}t_{a_1}t_{a'_1}t_{A_2}\cdots t_{A_{2g-1}}\cdot t_{A_1}\cdots t_{A_{2g-2}}t_{a_{g-1}}t'_{a_{g-1}}$$

in Mod_g^4 by conjugation by $t_{A_{2g-2}}\cdots t_{A_1}$. Since $a_{g-1}, a'_{g-1}, a_{g+1}, a'_{g+1}$ are disjoint from $A_{2g-1}, A_{2g}, A_{2g+1}, a_g, a'_g$, by the four-holed torus relation

$$t_{a_{g-1}}t_{a'_{g-1}}t_{a_{g+1}}t_{a'_{g+1}}=(t_{A_{2g-1}}t_{A_{2g+1}}t_{A_{2g}}t_{a_g}t_{a'_g}t_{A_{2g}})^2$$

and conjugation by $t_{A_{2g-1}}t_{A_{2g+1}}t_{A_{2g}}$, we have the relation

$$t_{a_g}^{-1}t_{a'_g}^{-1}t_{a_{g+1}}t_{a'_{g+1}} = t_{a'_{g-1}}^{-1}t_{a_{g-1}}^{-1}t_{A_{2g}}t_{A_{2g-1}}t_{A_{2g+1}}t_{A_{2g}}t_{a_g}t_{a'_g}t_{A_{2g}}t_{A_{2g-1}}t_{A_{2g+1}}t_{A_{2g}}.$$

By combining these relations, we have

$$\begin{aligned} t_{a_0}t_{a'_0}t_{a_{g+1}}t_{a'_{g+1}} = t_{A_{2g-2}}\cdots t_{A_1}\cdot t_{A_{2g-1}}\cdots t_{A_2}t_{a_1}t_{a'_1}t_{A_2}\cdots t_{A_{2g-1}}\cdot t_{A_1}\cdots t_{A_{2g-2}} \\ \cdot t_{A_{2g}}t_{A_{2g-1}}t_{A_{2g+1}}t_{A_{2g}}\cdot t_{a_g}t_{a'_g}t_{A_{2g}}t_{A_{2g-1}}t_{A_{2g+1}}t_{A_{2g}}. \end{aligned}$$

Note that $A_1, \ldots, A_{2g+1}$ are disjoint from $a_0, a'_0, a_{g+1}, a'_{g+1}$. Moreover, A_{2g} and A_{2g+1} are disjoint from $A_1, \ldots, A_{2g-2}$ and $A_1, \ldots, A_{2g-1}$, respectively. Therefore, by conjugation by $t_{A_{2g-2}}\cdots t_{A_1}$ and $t_{A_{2g+1}}t_{A_{2g}}$, we have

$$\begin{aligned} t_{a_0}t_{a'_0}t_{a_{g+1}}t_{a'_{g+1}} &= t_{A_{2g-2}}\cdots t_{A_1}\cdot t_{A_{2g-1}}\cdots t_{A_2}t_{a_1}t_{a'_1}t_{A_2}\cdots t_{A_{2g-1}}\cdot t_{A_1}\cdots t_{A_{2g-2}} \\ &\qquad \cdot t_{A_{2g}}t_{A_{2g-1}}t_{A_{2g+1}}t_{A_{2g}}\cdot t_{a_g}t_{a'_g}t_{A_{2g}}t_{A_{2g-1}}t_{A_{2g+1}}t_{A_{2g}} \\ &= t_{A_{2g+1}}t_{A_{2g}}\cdot t_{A_{2g-1}}\cdots t_{A_2}t_{a_1}t_{a'_1}t_{A_2}\cdots t_{A_{2g-1}}\cdot t_{A_{2g}}t_{A_{2g+1}}\cdot t_{A_1}\cdots t_{A_{2g-2}} \\ &\qquad \cdot t_{A_{2g-1}}t_{A_{2g}}\cdot t_{a_g}t_{a'_g}t_{A_{2g}}t_{A_{2g-1}}t_{A_{2g-2}}\cdots t_{A_1}. \end{aligned}$$

This completes the proof of Proposition 6.3. □

We now prove Proposition 6.1.

Proof of Proposition 6.1. Let c'_{h_1} be the separating simple closed curve as shown in Figure 7. By Lemma 6.2 and Proposition 6.3, we have

$$\begin{aligned} t_{a_{h_1+1}}t_{a'_{h_1+1}} &= t_{c_{h_1}}t_{c'_{h_1}}(t_{D_0}t_{D_1}\cdots t_{D_{2h_1}})^2, \\ t_{c_{h_1}}t_{c'_{h_1}}t_{a_{g+1}}t_{a'_{g+1}} &= t_{a_g}t_{A_{2g}}\cdots t_{A_{4h_1+2}}t_{E_{h_1}}t_{E_{h_1}}t_{A_{4h_1+2}}\cdots t_{A_{2g}}t_{a'_g} \\ &\qquad \cdot t_{A_{2g+1}}\cdots t_{A_{4h_1+2}}t_{a_{h_1+1}}t_{a'_{h_1+1}}t_{A_{4h_1+2}}\cdots t_{A_{2g+1}}. \end{aligned}$$

Since c_{h_1} and c'_{h_1} are disjoint from $A_{2h_1+2}, \ldots, A_{2g}, E_{h_1}, a_{h_1+1}, a'_{h_1+1}$, it follows that

$$\begin{aligned} t_{c'_{h_1}}^{-1}t_{c_{h_1}}^{-1}\cdot t_{a_{h_1+1}}t_{a'_{h_1+1}} &= (t_{D_0}t_{D_1}\cdots t_{D_{2h_1}})^2, \\ t_{a_{g+1}}t_{a'_{g+1}} &= t_{a_g}t_{A_{2g}}\cdots t_{A_{4h_1+2}}t_{E_{h_1}}t_{E_{h_1}}t_{A_{4h_1+2}}\cdots t_{A_{2g}}t_{a'_g} \\ &\qquad \cdot t_{A_{2g+1}}\cdots t_{A_{4h_1+2}}\cdot t_{c'_{h_1}}^{-1}t_{c_{h_1}}^{-1}\cdot t_{a_{h_1+1}}t_{a'_{h_1+1}}\cdot t_{A_{4h_1+2}}\cdots t_{A_{2g+1}}. \end{aligned}$$

Combining these relations gives the relation (21) in Proposition 6.1.

In Σ_g^1, A_{2g+1} is homotopic to a'_g, and (22) follows, completing the proof. □

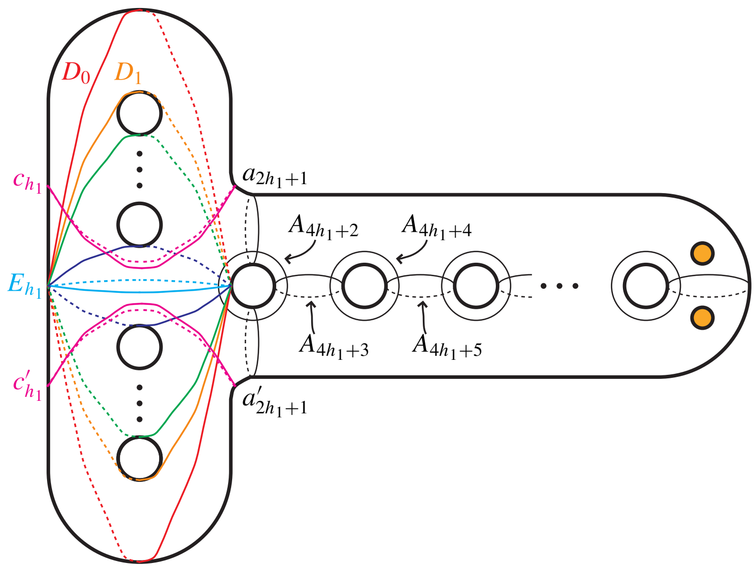

Figure 7. The curve c'_{h_1} on Σ_g^2.

6C. ***A lift of Gurtas' positive relator.*** Since a_{g+1} and a'_{g+1} are null-homotopic in Σ_g, we have $t_{a_{g+1}} = t_{a'_{g+1}} = 1$ in Mod_g, so the relation in Proposition 6.1 is a positive relator in Mod_g. Then we note that A_{2g+1} and a'_g are homotopic to a_g. On the other hand, Gurtas [2004] gave the positive relator

$$(t_{A_{4h_1+2}} \cdots t_{A_{2g}} t_{a_g} t_{a_g} t_{A_{2g}} \cdots t_{A_{4h_1+2}} t_{D_0} t_{D_1} \cdots t_{D_{2h_1}} t_{E_{h_1}})^2 = 1.$$

in Mod_g. Using the following theorem of Kas [1980] and Matsumoto [1996], we show that the relation in Proposition 6.1 gives a lift of Gurtas' positive relator in Mod_g to Mod_g^2.

Theorem 6.4 [Kas 1980; Matsumoto 1996]. *If $g \geq 2$, then the isomorphism class of a Lefschetz fibration is determined by a positive relator modulo simultaneous conjugations*

$$t_{v_n} \cdots t_{v_2} t_{v_1} \sim t_{\phi(v_n)} \cdots t_{\phi(v_2)} t_{\phi(v_1)} \quad \textit{for any } \phi \in \Gamma_g$$

and elementary transformations

$$t_{v_n} \cdots t_{v_{i+2}} t_{v_{i+1}} t_{v_i} t_{v_{i-1}} t_{v_{i-2}} \cdots t_{v_1} \sim t_{v_n} \cdots t_{v_{i+2}} t_{v_i} t_{t_{v_i}^{-1}(v_{i+1})} t_{v_{i-1}} t_{v_{i-2}} \cdots t_{v_1},$$
$$t_{v_n} \cdots t_{v_{i+2}} t_{v_{i+1}} t_{v_i} t_{v_{i-1}} t_{v_{i-2}} \cdots t_{v_1} \sim t_{v_n} \cdots t_{v_{i+2}} t_{v_{i+1}} t_{t_{v_i}(v_{i-1})} t_{v_i} t_{v_{i-2}} \cdots t_{v_1}.$$

The aim of this section is to prove the following proposition. This proposition applied to Proposition 6.1 gives the above mentioned lift.

Proposition 6.5. *In* Mod_g, *the following relation holds*:

$$t_{E_{h_1}} t_{A_{4h_1+2}} \cdots t_{A_{2g}} t_{a_g} t_{a_g} t_{A_{2g}} \cdots t_{A_{4h_1+2}} t_{E_{h_1}} \cdot t_{a_g} t_{A_{2g}} \cdots t_{A_{4h_1+2}} \cdot (t_{D_0} t_{D_1} \cdots t_{D_{2h_1}})^2 \cdot t_{A_{4h_1+2}} \cdots t_{A_{2g}} t_{a_g} \sim (t_{A_{4h_1+2}} \cdots t_{A_{2g}} t_{a_g} t_{a_g} t_{A_{2g}} \cdots t_{A_{4h_1+2}} t_{D_0} t_{D_1} \cdots t_{D_{2h_1}} t_{E_{h_1}})^2.$$

In order to prove this, we need a lemma.

Lemma 6.6. *We deform* Σ_g^2 *as shown in Figure 8(a) and (b). Let* E *and* E' *be the simple closed curves in* Σ_g^2 *as in Figure 8(a) and (b), and let* a *be the arc connecting the boundary components of* Σ_g^2 *as in the figure. Then*

$$t_{B_0} t_{B_1} \cdots t_{B_g}(E) = E', \tag{23}$$

$$t_{B_0} t_{B_1} \cdots t_{B_g} t_E(a) = t_{a_{g+1}} t_{a'_{g+1}}(a). \tag{24}$$

Proof. From the equality (20), we see that

$$t_{B_0} t_{B_1} \cdots t_{B_g} = \Delta_{2g+1} \cdots \Delta_2 \Delta_1 t_{c_r}^{-1}.$$

By drawing corresponding curves and applying the corresponding Dehn twist, we find that

$$\Delta_{2g+1} \cdots \Delta_2 \Delta_1 t_{c_r}^{-1}(E) = E' \quad \text{and} \quad \Delta_{2g+1} \cdots \Delta_2 \Delta_1 t_{c_r}^{-1} t_E(a) = t_{a_{g+1}} t_{a'_{g+1}}(a).$$

This proves the lemma. □

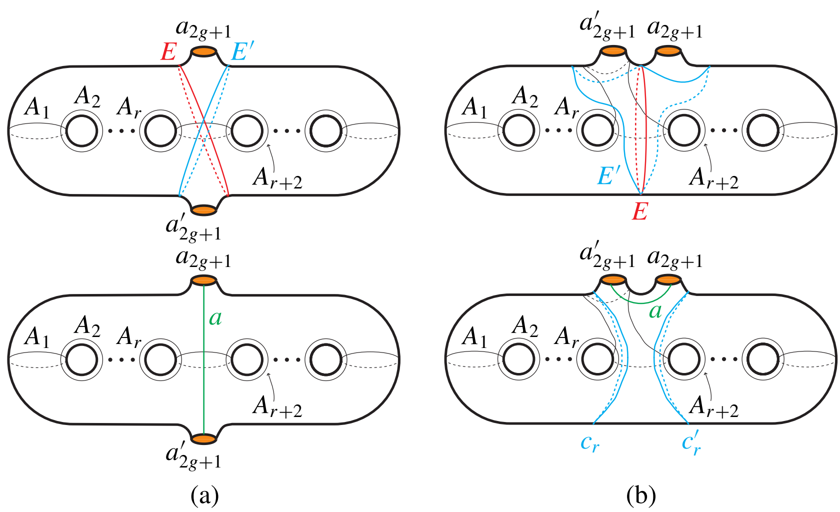

Figure 8. The curves E, E' and the arc a.

Proof of Proposition 6.5. For simplicity of notation, we write

$$\tau := t_{A_{4h_1+2}} \cdots t_{A_{2g}} t_{a_g} \quad \text{and} \quad \overline{\tau} := t_{a_g} t_{A_{2g}} \cdots t_{A_{4h_1+2}}.$$

Note that for each $i = 2h_1+2, \dots, 2g$, we find that

$$t_{E_{h_1}} \tau \overline{\tau} t_{E_{h_1}}(A_i) = A_i \quad \text{and} \quad t_{E_{h_1}} \tau \overline{\tau} t_{E_{h_1}}(a_g) = a_g.$$

This gives

$$t_{E_{h_1}} \tau \overline{\tau} t_{E_{h_1}} \cdot t_{A_i} \sim t_{A_i} \cdot t_{E_{h_1}} \tau \overline{\tau} t_{E_{h_1}} \quad \text{and} \quad t_{E_{h_1}} \tau \overline{\tau} t_{E_{h_1}} \cdot t_{a_g} \sim t_{a_g} \cdot t_{E_{h_1}} \tau \overline{\tau} t_{E_{h_1}},$$

so we obtain the relation

$$t_{E_{h_1}} \tau \overline{\tau} t_{E_{h_1}} \cdot \tau \sim \tau \cdot t_{E_{h_1}} \tau \overline{\tau} t_{E_{h_1}}.$$

Therefore, applying elementary transformations (including cyclic permutations) gives

$$t_{E_{h_1}} \tau \overline{\tau} t_{E_{h_1}} \cdot \overline{\tau} (t_{D_0} t_{D_1} \cdots t_{D_{2h_1}})^2 \cdot \tau \sim t_{E_{h_1}} \tau \overline{\tau} t_{E_{h_1}} \cdot \tau \cdot \overline{\tau} (t_{D_0} t_{D_1} \cdots t_{D_{2h_1}})^2. \tag{25}$$

Since by drawing corresponding curves, applying the corresponding Dehn twist and (24) in Lemma 6.6, we have

$$(\tau \overline{\tau})^{-1}(E_{h_1}) = t_{a_{2h_1+1}} t_{a'_{2h_1+1}}(E_{h_1}) = t_{D_0} t_{D_1} \cdots t_{D_{2h_1}}(E_{h_1}),$$

we thus obtain

$$\tau \overline{\tau} \cdot t_{D_0} t_{D_1} \cdots t_{D_{2h_1}} \cdot t_{E_{h_1}} \sim t_{E_{h_1}} \cdot \tau \overline{\tau} \cdot t_{D_0} t_{D_1} \cdots t_{D_{2h_1}}.$$

Therefore, by using this relation, we have

$$\begin{aligned} t_{E_{h_1}} \tau \overline{\tau} t_{E_{h_1}} \cdot \tau \overline{\tau} \cdot (t_{D_0} t_{D_1} \cdots t_{D_{2h_1}})^2 \\ \sim t_{E_{h_1}} \tau \overline{\tau} \cdot \tau \overline{\tau} \cdot t_{D_0} t_{D_1} \cdots t_{D_{2h_1}} \cdot t_{E_{h_1}} \cdot t_{D_0} t_{D_1} \cdots t_{D_{2h_1}}. \end{aligned} \tag{26}$$

By drawing corresponding curves, applying the corresponding Dehn twist and (23) in Lemma 6.6, we obtain

$$(\tau \overline{\tau})^{-1}(A_{4h_1+2}) = t_{D_0} t_{D_1} \cdots t_{D_{2h_1}} t_{E_{h_1}}(A_{4h_1+2}).$$

Therefore, we have

$$\tau \overline{\tau} \cdot t_{D_0} t_{D_1} \cdots t_{D_{2h_1}} t_{E_{h_1}} \cdot t_{A_{4h_1+2}} \sim t_{A_{4h_1+2}} \cdot \tau \overline{\tau} \cdot t_{D_0} t_{D_1} \cdots t_{D_{2h_1}} t_{E_{h_1}}.$$

Note that for each $i = 4h_1+3, \dots, 2g$, we find that

$$\tau \overline{\tau}(A_i) = A_i \quad \text{and} \quad \tau \overline{\tau}(a_g) = a_g.$$

Moreover, since $A_{4h_1+3}, \ldots, A_{2g}$ and a_g are disjoint from $D_0, \ldots, D_{2h_1}, E_{h_1}$, we therefore obtain, for each $i = 2h_1 + 3, \ldots, 2g$,

$$\tau\bar{\tau}t_{D_0}t_{D_1}\cdots t_{D_{2h_1}}t_{E_{h_1}}\cdot t_{A_i} \sim t_{A_i}\cdot t\tau\bar{\tau}t_{D_0}t_{D_1}\cdots t_{D_{2h_1}}t_{E_{h_1}},$$
$$\tau\bar{\tau}t_{D_0}t_{D_1}\cdots t_{D_{2h_1}}t_{E_{h_1}}\cdot t_{a_g} \sim t_{a_g}\cdot \tau\bar{\tau}t_{D_0}t_{D_1}\cdots t_{D_{2h_1}}t_{E_{h_1}}.$$

This gives

$$\tau\bar{\tau}\cdot\tau\bar{\tau}t_{D_0}t_{D_1}\cdots t_{D_{2h_1}}t_{E_{h_1}} \sim \tau\bar{\tau}t_{D_0}t_{D_1}\cdots t_{D_{2h_1}}t_{E_{h_1}}\cdot\tau\bar{\tau}.$$

From this relation, applying elementary transformations (including cyclic permutations) gives

$$\begin{aligned}(27)\quad & t_{E_{h_1}}\tau\bar{\tau}\cdot\tau\bar{\tau}\cdot t_{D_0}t_{D_1}\cdots t_{D_{2h_1}}t_{E_{h_1}}\cdot t_{D_0}t_{D_1}\cdots t_{D_{2h_1}} \\ & \qquad\sim \tau\bar{\tau}\cdot t_{D_0}t_{D_1}\cdots t_{D_{2h_1}}t_{E_{h_1}}\cdot\tau\bar{\tau}\cdot t_{D_0}t_{D_1}\cdots t_{D_{2h_1}}\cdot t_{E_{h_1}}.\end{aligned}$$

Proposition 6.5 follows from the relations (25)–(27). □

7. Construction of simple loops $R_1, \ldots, R_k$

In this section, we prove Proposition 5.2. This was based on Korkmaz's work [2009] and the argument in [Akhmedov and Ozbagci 2013]. In Proposition 4.3 of [Korkmaz 2009], he defined l as $l = l(r_1) + \cdots + l(r_k)$. However, in this paper, it is sufficient to consider l as $l = \max_{1\le i\le k}\{l(r_i)\}$. Before providing the simple loops in Σ_g in Proposition 5.2, we need the following proposition about simple loops $R_1, \ldots, R_k$ in Σ_{n+l-1}.

Proposition 7.1. *Let F_n be the subgroup of $\pi_1(\Sigma_n)$ generated by $a_1, \ldots, a_n$, i.e., F_n is a free group of rank n. Let $r_1, \ldots, r_k$ be k arbitrary elements in F_n represented as words in $a_1, \ldots, a_n$. Let $l = \max_{1\le i\le k}\{l(r_i)\}$, where $l(r_i)$ is the syllable length of r_i. Then there are simple loops $R_1, \ldots, R_k$ in Σ_{n+l-1} with the property that for each $1 \le i \le k$:*

(a) *R_i is freely homotopic to a simple closed curve which intersects a_{n+l-1} transversely at only one point.*

(b) *$\Phi([R_i]) = r_i$, where $[R_i] \in \pi_1(\Sigma_{n+l-1})$ is the homotopy class of R_i, and $\Phi : \pi_1(\Sigma_{n+l-1}) \to \pi_1(\Sigma_n)$ is the map defined by $\Phi(a_j) = a_j$ for $1 \le j \le n$ and $\Phi(\alpha) = 1$ for $\alpha \in \{a_{n+1}, \ldots, a_{n+l-1}, b_1, \ldots, b_{n+l-1}\}$.*

Proof. Let us consider the surface Σ_n embedded in $\mathbb{R}^3$ as shown in Figure 1 such that for each $1 \le j \le n$, a simple closed curve b'_j in Σ_n which is isotopic to b_j lies on the plane $x = 0$. Write $r_i = a_{i_1}^{m_1}\cdots a_{i_d}^{m_d}$, where $d = l(r_i)$ is the syllable length of r_i. We denote by ξ a constant such that the base point lies in the plane $z = \xi$. Let L be an arc in Σ_n which lies in the half plane $\{z = \xi\} \cap \{x \ge 0\}$.

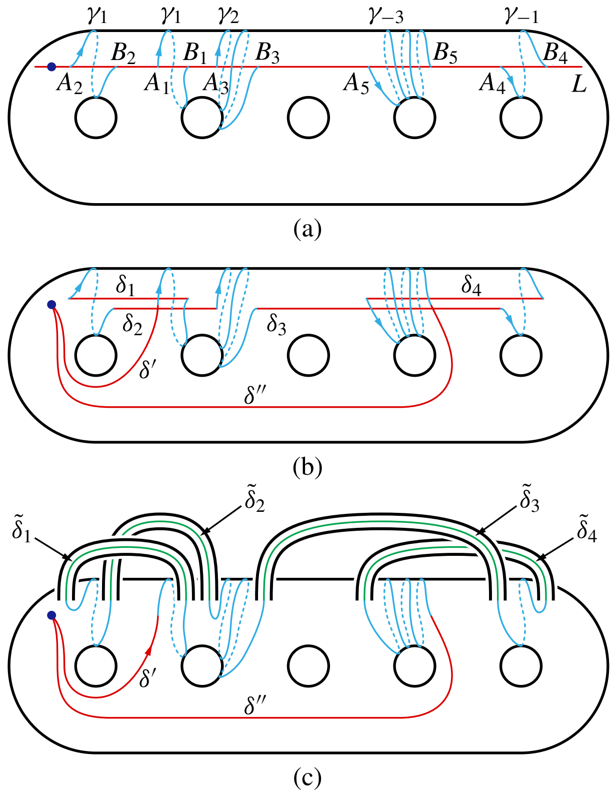

Figure 9. Construction of R_i on Σ_{n+d-1} for $r_i = a_2 a_1 a_2^2 a_5^{-1} a_4^{-3}$ and for $n = 5$.

For $1 \le t \le d$, let α_t be a loop in Σ_n which is isotopic to a_{i_t}. If $j_s = j_{s'}$ for some $s < s'$, then we assume that $\alpha_{s'}$ is to the right of α_s and that $\alpha_{s'}$ is disjoint from α_s. Here, right means the positive direction of the y-axis. Let A_t (resp. B_t) be points on L lying to the left (resp. right) of α_t such that there are no A_s (resp. B_s) between α_t and A_t (resp. B_t).

Let $\gamma_{m_t} = t_{\alpha_t}^{-m_t}(\zeta_t)$, where ζ_t is the subarc of L from the point A_j to the point B_j. For each $1 \le j \le d-1$, let δ_j denote the subarc of L from the point B_j to the point A_{j+1}. Then we can define an arc β in Σ_n connecting A_1 to B_d to be

$$\beta = \gamma_{m_1} \star \delta_1 \star \gamma_{m_2} \star \delta_2 \star \cdots \star \delta_{d-1} \star \gamma_{m_d},$$

where $\gamma \star \delta$ denotes an arc γ followed by an arc δ. Let δ_0 be the subarc of L from the base point to A_1, and δ_d the subarc from B_d to the base point. Then $\delta_0 \star \beta \star \delta_d$ represents r_i (cf. Figure 9(a)). After perturbing β slightly, we assume that $\delta_1, \ldots, \delta_{d-1}$ are pairwise disjoint and lie parallel to the plane $x = 0$. Note that all self-intersection points of $\delta_0 \star \beta \star \delta_d$ lie on $\delta_0 \cup \delta_1 \cup \cdots \cup \delta_d$.

Let δ' and δ'' be arcs from the base point to A_1 and from B_d to the base point, respectively, which are disjoint from $\alpha_1, \alpha_2, \ldots, \alpha_d$ and $b'_1, b'_2, \ldots, b'_n$ and lie in the space $\{z \leq \xi\}$. Suppose that the interiors of δ', δ'' and β are pairwise disjoint. Then the loop $\delta' \star \beta \star \delta''$ represents

$$b_1 b_2 \cdots b_{i_1-1} r_i b_{i_d}^{-1} \cdots b_2^{-1} b_1^{-1}$$

in $\pi_1(\Sigma_n)$ (cf. Figure 9(b)).

Let $D_1, D'_1, \ldots, D_{d-1}, D'_{d-1}$ be pairwise disjoint disks in Σ_n such that for each $1 \leq t \leq d-1$, $\mathrm{Int}(D_t)$ and $\mathrm{Int}(D'_t)$ are disjoint from δ', β and δ'', and $A_t \in \partial D_t$ and $B_t \in \partial D'_t$. We remove $2d-2$ open disks $\mathrm{Int}(D_t)$ and $\mathrm{Int}(D'_t)$ from Σ_n. Then for each $1 \leq t \leq d-1$, by attaching an annulus, denote by $\mathscr{A}_t$, to the surface

$$\Sigma_n \setminus \bigcup_{t=1}^{d-1} \bigl(\mathrm{Int}(D_t) \cup \mathrm{Int}(D'_t)\bigr)$$

along ∂D_t and $\partial D'_t$, we obtain the closed oriented surface

$$\left(\Sigma_n \setminus \bigcup_{t=1}^{d-1} \bigl(\mathrm{Int}(D_t) \cup \mathrm{Int}(D'_t)\bigr)\right) \cap \left(\bigcup_{t=1}^{d-1} \mathscr{A}_t\right)$$

of genus $n+d-1$, denoted by Σ_{n+d-1}. An orientation on Σ_{n+d-1} is given by the orientation on Σ_n.

We define a loop R_i in Σ_{n+d-1} as follows. For each $1 \leq t \leq d-1$, let $\tilde{\delta}_t$ be a simple arc in $\mathscr{A}_t$ from the point B_t to the point A_{t+1} such that $\tilde{\delta}_t$ lies parallel to the plane $x = 0$. Then by "replacing" δ_t in $\delta' \star \beta \star \delta''$ by $\tilde{\delta}_t$, we obtain the loop

$$R = \delta' \star \gamma_{m_1} \star \tilde{\delta}_1 \star \gamma_{m_2} \star \tilde{\delta}_2 \star \cdots \star \tilde{\delta}_{d-1} \star \gamma_{m_d} \star \delta''.$$

In particular, R_i is simple in Σ_{n+d-1} (cf. Figure 9(c)).

Note that from construction, $\tilde{\delta}_t \star \delta_t$ is a simple closed curve in Σ_{n+d-1}. If we collapse each $\mathscr{A}_t$ onto the arc δ_t, then we obtain a map $\Sigma_{n+d-1} \to \Sigma_n$. The induced map $\pi_1(\Sigma_{n+d-1}) \to \pi_1(\Sigma_n)$ takes $[R]$ to

$$b_1 b_2 \cdots b_{i_1-1} r_i b_{i_d}^{-1} \cdots b_2^{-1} b_1^{-1},$$

which in turn is mapped to r_i under the map $\pi_1(\Sigma_n) \to \pi_1(\Sigma_n)$ sending a_j to a_j and b_j to 1 for all j.

Let $h = n + l - 1$, where $l = \max_{1 \leq i \leq k}\{l(r_i)\}$. For each $1 \leq i \leq k$, we now construct a loop R_i in Σ_h as follows. First, by sliding $\mathscr{A}_1, \ldots, \mathscr{A}_{l(r_i)-1}$, we deform the surface $\Sigma_{n+l(r_i)-1}$ into the standard position as shown in Figure 1 in such a way that the simple loop $\tilde{\delta}_t \star \delta_t$ becomes isotopic to b_{n+t} and the boundary curves of $\mathscr{A}_t$ become isotopic to a_{n+t} (cf. Figure 10(a), (b) and (c)). If $l(r_j) = l$ for some j, then we see that the simple closed curve a_h intersects R_j transversely at one point.

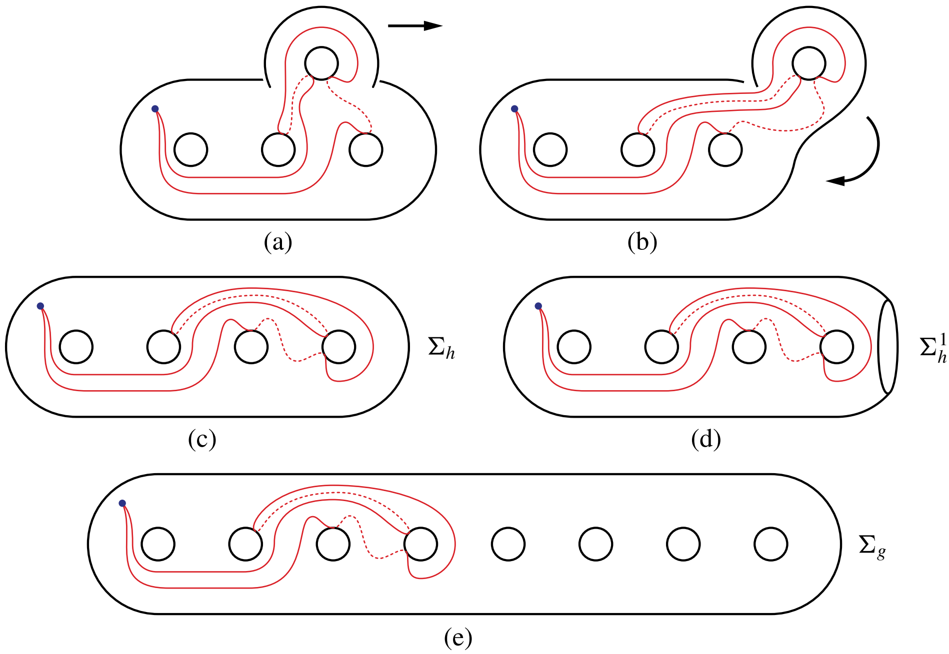

Figure 10. Construction of R_i for $r_i = a_3^{-1}a_2^{-1}$ in the case $n = 3$ and $g = 8$.

Therefore, we assume that $l(r_i) < l$. Next, we remove a small open disk from the deformed surface near $a_{n+l(r_i)-1}$ and disjoint from R_i (cf. Figure 10(d)). Thus, we obtain a surface of genus $n + l(r_i) - 1$ with one boundary component, denoted by $\Sigma^1_{n+l(r_i)-1}$. We embed $\Sigma^1_{n+l(r_i)-1}$ into the standard surface Σ_h in such a way that for each $1 \le t \le n + l(r_i) - 1$, simple loops a_t, b_t in $\Sigma^1_{n+l(r_i)-1}$ correspond to the simple loops a_t, b_t in Σ_h (cf. Figure 10(e)). Finally, we replace R_i with a simple representative of $[R_i]\big((b_1b_2\cdots b_{h-1})(b_1b_2\cdots b_h)^{-1}\big)^{\epsilon}$, where $\epsilon = \pm 1$ (cf. Figure 10(d)). Then we see that the resulting simple loop R_i intersects a_h transversely at one point.

From the above construction, $\Phi : \pi_1(\Sigma_h) \to \pi_1(\Sigma_n)$ maps $[R_i]$ to r_i for each $i = 1, \dots, k$. This gives the required simple loops $R_1, \dots, R_k$. □

Proof of Proposition 5.2. Consider a surface Σ_{n+l-1} and the loops $R_1, \dots, R_k$ constructed in Proposition 7.1. We remove a small open disk from Σ_{n+l-1} near a_{n+l-1} and disjoint from all R_i (cf. Figure 11(a)). Denote by Σ^1_{n+l-1} the resulting surface of genus $n+l-1$ with one boundary component. We embed Σ^1_{n+l-1} into the standard surface Σ_g in such a way that for each $1 \le t \le n+l-1$, simple loops a_t, b_t in Σ^1_{n+l-1} correspond to the simple loops a_t, b_t in Σ_g (cf. Figure 11(b)). Then we can modify $R_1, \dots, R_k$ so that each R_i $(i = 1, \dots, k)$ satisfies the property of Proposition 5.2

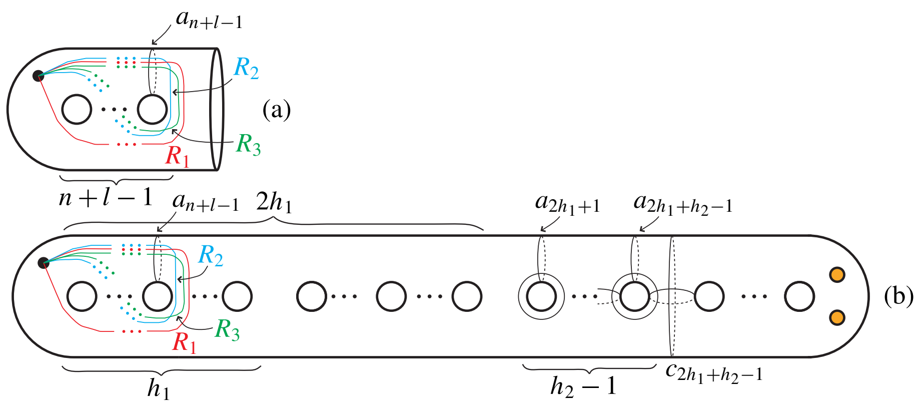

Figure 11. Modified curves $R_1, \dots, R_k$ in Σ_g.

by replacing R_i with a simple representative of $[R_i](b_{2h_1+1}b_{2h_2+2}\cdots b_{2h_1+h_2-i})^\epsilon$ if i is odd, and $[R_i]a^\epsilon_{2h_1+h_2-i}$ if i is even, where $\epsilon = \pm 1$ (cf. Figure 5). Therefore, we obtain the required simple loops $R_1, \dots, R_k$. $\square$

8. Remarks

The results of [Gompf 1995; Donaldson 1999; Gompf and Stipsicz 1999] mentioned in the introduction naturally raise the following two basic questions, which remain open.

Question 8.1 (cf. [Korkmaz and Stipsicz 2009]). Given a symplectic 4-manifold, what is the minimal genus g for which it has a genus-g Lefschetz pencil?

Question 8.2. Given a finitely presented group Γ, what is the minimal genus, denoted by $g_P(\Gamma)$, for which there exists a genus-g Lefschetz pencil on a symplectic 4-manifold with fundamental group Γ?

Although these two questions remain open, for Question 8.2, we can give an upper bound for $g_P(\Gamma)$ as a corollary of Theorem 1.2.

Corollary 8.3. *We have* $g_P(\Gamma) \le 4(n+l-1)+k$ *for* $k \ge 1$, *and* $g_P(F_n) \le 4n+2$.

However, this upper bound for $g_P(\Gamma)$ may not be sharp. In fact, since $\mathbb{CP}^2$ admits a genus-0 Lefschetz pencil, $g_P(\Gamma) = 0$ if Γ is the trivial group. When we replace the relations in Proposition 5.1 and the map ψ_k in Section 5A by another relation and map, we can improve the upper bound of $g_P(\Gamma)$. For example, for every positive integer n, the article [Hamada et al. $\ge$ 2016] gave a genus-g Lefschetz pencil on a 4-manifold X_n such that $\pi_1(X_n) \cong \mathbb{Z} \oplus \mathbb{Z}_n$ for every $g \ge 4$ using a similar construction to this paper. Therefore, $g_P(\mathbb{Z} \oplus \mathbb{Z}_m) \le 4$.

We expect that by a combination of substitution techniques and partial conjugation techniques, we could obtain results for Lefschetz fibrations with (-1)-sections analogous to those obtained by fiber sum operations. The articles [Ozbagci and Stipsicz 2000; Korkmaz 2001; Monden 2014] gave examples of nonholomorphic Lefschetz fibrations by fiber sum operations (and lantern substitutions). By a similar technique to this paper (and a lantern substitution), two kinds of nonholomorphic Lefschetz fibrations with (-1)-sections were constructed in [Hamada et al. $\geq$ 2016]. One is a Lefschetz fibration with noncomplex total space, and the other is a Lefschetz fibration violating the "slope inequality".

Finally, we explain why the Lefschetz fibrations constructed in [Korkmaz 2009; Akhmedov and Ozbagci 2013; Kobayashi 2015] do not have (-1)-sections. In [Korkmaz 2009; Kobayashi 2015], twisted fiber sum operations were adopted, and the fibrations in [Akhmedov and Ozbagci 2013] were obtained by performing Luttinger surgeries and knot surgeries on the symplectic sum of certain symplectic 4-manifolds. The fiber sum of Lefschetz fibrations has no (-1)-sections (see [Stipsicz 2001a], and also [Smith 2001]). In particular, the symplectic sum of symplectic 4-manifolds is minimal, that is, it does not contain any (-1)-spheres (see [Usher 2006], and also [Sato 2006; Baykur 2015]), and Luttinger surgery and knot surgery preserve minimality of symplectic 4-manifolds from the result of [Usher 2006]. Therefore, we see that the fibrations in [Korkmaz 2009; Akhmedov and Ozbagci 2013; Kobayashi 2015] do not have any (-1)-sections.

Acknowledgments

The authors would like to thank Susumu Hirose for his comments on this paper. The second author was supported by Grant-in-Aid for Young Scientists (B) (No. 13276356), Japan Society for the Promotion of Science.

References

[Akbulut and Ozbagci 2002] S. Akbulut and B. Ozbagci, "On the topology of compact Stein surfaces", *Int. Math. Res. Not.* **2002**:15 (2002), 769–782. MR 2003a:57049 Zbl 1007.57023

[Akhmedov and Ozbagci 2013] A. Akhmedov and B. Ozbagci, "Exotic Stein fillings with arbitrary fundamental group", preprint, 2013. arXiv 1212.1743

[Amorós et al. 2000] J. Amorós, F. Bogomolov, L. Katzarkov, and T. Pantev, "Symplectic Lefschetz fibrations with arbitrary fundamental groups", *J. Differential Geom.* **54**:3 (2000), 489–545. MR 2002g:57051 Zbl 1031.57021

[Auroux 2006a] D. Auroux, "The canonical pencils on Horikawa surfaces", *Geom. Topol.* **10** (2006), 2173–2217. MR 2007m:14065 Zbl 1129.57030

[Auroux 2006b] D. Auroux, "Mapping class group factorizations and symplectic 4-manifolds: some open problems", pp. 123–132 in *Problems on mapping class groups and related topics*, edited by B. Farb, Proceedings of Symposia in Pure Mathematics **74**, American Mathematical Society, Providence, RI, 2006. MR 2007h:53134 Zbl 1304.57027

[Baykur 2015] R. I. Baykur, "Minimality and fiber sum decompositions of Lefschetz fibrations", *Proc. Amer. Math. Soc.* (online publication December 2015).

[Cadavid 1998] C. A. Cadavid, *On a remarkable set of words in the mapping class group*, thesis, University of Texas, Austin, TX, 1998, Available at http://search.proquest.com/docview/304458143. MR 2699379

[Donaldson 1999] S. K. Donaldson, "Lefschetz pencils on symplectic manifolds", *J. Differential Geom.* **53**:2 (1999), 205–236. MR 2002g:53154 Zbl 1040.53094

[Etnyre and Honda 2002] J. B. Etnyre and K. Honda, "On symplectic cobordisms", *Math. Ann.* **323**:1 (2002), 31–39. MR 2003c:57026 Zbl 1022.53059

[Gompf 1995] R. E. Gompf, "A new construction of symplectic manifolds", *Ann. of Math.* (2) **142**:3 (1995), 527–595. MR 96j:57025 Zbl 0849.53027

[Gompf and Stipsicz 1999] R. E. Gompf and A. I. Stipsicz, *4-manifolds and Kirby calculus*, Graduate Studies in Mathematics **20**, American Mathematical Society, Providence, RI, 1999. MR 2000h:57038 Zbl 0933.57020

[Gurtas 2004] Y. Z. Gurtas, "Positive Dehn twist expressions for some new involutions in mapping class group", preprint, 2004. arXiv math/0404310

[Hamada ≥ 2016] N. Hamada, "On a combinatorial decomposition for relations among Dehn twists", In preparation.

[Hamada et al. ≥ 2016] M. Hamada, R. Kobayashi, and N. Monden, "Non-holomorphic Lefschetz fibrations with (-1)-sections", In preparation.

[Kas 1980] A. Kas, "On the handlebody decomposition associated to a Lefschetz fibration", *Pacific J. Math.* **89**:1 (1980), 89–104. MR 82f:57012 Zbl 0457.14011

[Kobayashi 2015] R. Kobayashi, "On genera of Lefschetz fibrations and finitely presented groups", *Osaka J. Math.* (online publication January 2015).

[Korkmaz 2001] M. Korkmaz, "Noncomplex smooth 4-manifolds with Lefschetz fibrations", *Int. Math. Res. Not.* **2001**:3 (2001), 115–128. MR 2001m:57036 Zbl 0977.57020

[Korkmaz 2009] M. Korkmaz, "Lefschetz fibrations and an invariant of finitely presented groups", *Int. Math. Res. Not.* **2009**:9 (2009), 1547–1572. MR 2010c:57037 Zbl 1173.57013

[Korkmaz and Ozbagci 2008] M. Korkmaz and B. Ozbagci, "On sections of elliptic fibrations", *Michigan Math. J.* **56**:1 (2008), 77–87. MR 2009f:57043 Zbl 1158.57033

[Korkmaz and Stipsicz 2009] M. Korkmaz and A. I. Stipsicz, "Lefschetz fibrations on 4-manifolds", pp. 271–296 in *Handbook of Teichmüller theory*, vol. 2, edited by A. Papadopoulos, IRMA Lectures in Mathematics and Theoretical Physics **13**, European Mathematical Society, Zürich, 2009. MR 2010k:57048 Zbl 1177.57001

[Loi and Piergallini 2001] A. Loi and R. Piergallini, "Compact Stein surfaces with boundary as branched covers of B^4", *Invent. Math.* **143**:2 (2001), 325–348. MR 2002c:53139 Zbl 0983.32027

[Matsumoto 1996] Y. Matsumoto, "Lefschetz fibrations of genus two: a topological approach", pp. 123–148 in *Topology and Teichmüller spaces* (Katinkulta, 1995), edited by S. Kojima et al., World Scientific, River Edge, NJ, 1996. MR 2000h:14038 Zbl 0921.57006

[Monden 2014] N. Monden, "Lefschetz fibrations with small slope", *Pacific J. Math.* **267**:1 (2014), 243–256. MR 3163482 Zbl 1306.57019

[Ozbagci and Stipsicz 2000] B. Ozbagci and A. I. Stipsicz, "Noncomplex smooth 4-manifolds with genus-2 Lefschetz fibrations", *Proc. Amer. Math. Soc.* **128**:10 (2000), 3125–3128. MR 2000m:57036 Zbl 0951.57015

[Sato 2006] Y. Sato, "The Stipsicz's conjecture for genus-2 Lefschetz fibrations", preprint, 2006.

[Smith 2001] I. Smith, "Geometric monodromy and the hyperbolic disc", *Q. J. Math.* **52**:2 (2001), 217–228. MR 2002c:57046 Zbl 0981.57013

[Stipsicz 2001a] A. I. Stipsicz, "Indecomposability of certain Lefschetz fibrations", *Proc. Amer. Math. Soc.* **129**:5 (2001), 1499–1502. MR 2001h:57029 Zbl 0978.57022

[Stipsicz 2001b] A. I. Stipsicz, "Spin structures on Lefschetz fibrations", *Bull. London Math. Soc.* **33**:4 (2001), 466–472. MR 2002a:53062 Zbl 1037.57019

[Tanaka 2012] S. Tanaka, "On sections of hyperelliptic Lefschetz fibrations", *Algebr. Geom. Topol.* **12**:4 (2012), 2259–2286. MR 3020206 Zbl 1268.57010

[Usher 2006] M. Usher, "Minimality and symplectic sums", *Int. Math. Res. Not.* **2006** (2006), Art. ID #49857. MR 2007h:53139 Zbl 1110.57017

Received March 27, 2015. Revised September 22, 2015.

RYOMA KOBAYASHI
DEPARTMENT OF GENERAL EDUCATION
ISHIKAWA NATIONAL COLLEGE OF TECHNOLOGY
TSUBATA, ISHIKAWA 929-0392
JAPAN

kobayashi_ryoma@ishikawa-nct.ac.jp

NAOYUKI MONDEN
DEPARTMENT OF ENGINEERING SCIENCE
OSAKA ELECTRO-COMMUNICATION UNIVERSITY
HATSU-CHO 18-8
NEYAGAWA 572-8530
JAPAN

monden@isc.osakac.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.389

KNOT HOMOTOPY IN SUBSPACES OF THE 3-SPHERE

YUYA KODA AND MAKOTO OZAWA

We discuss an extrinsic property of knots in a 3-subspace of the 3-sphere S^3 to characterize how the subspace is embedded in S^3. Specifically, we show that every knot in a subspace of the 3-sphere is transient if and only if the exterior of the subspace is a disjoint union of handlebodies, i.e., regular neighborhoods of embedded graphs, where a knot in a 3-subspace of S^3 is said to be transient if it can be moved by a homotopy within the subspace to the trivial knot in S^3. To show this, we discuss the relation between certain group-theoretic and homotopic properties of knots in a compact 3-manifold, which can be of independent interest. Further, using the notion of transient knots, we define an integer-valued invariant of knots in S^3 that we call the transient number. We then show that the union of the sets of knots of unknotting number one and tunnel number one is a proper subset of the set of knots of transient number one.

Introduction

In the list [Eilenberg 1949] of problems edited by Eilenberg, Fox proposed a program to distinguish 3-manifolds by the differences in their "knot theories". Following the program, Brody [1960] reobtained the topological classification of the 3-dimensional lens spaces using knot-theoretic invariants, which are the Alexander polynomials of knots suitably factored out so that it depends only on the homology classes of the knots. Bing's recognition theorem [1958] can be regarded as another example of works that follow Fox's program. The theorem asserts that a closed, connected 3-manifold M is homeomorphic to the 3-sphere if and only if every knot in M can be moved by an isotopy to lie within a 3-ball. We note here that if we replace *isotopy* in this statement by *homotopy*, the assertion implies the Poincaré conjecture, which was proved by Perelman [2002; 2003a; 2003b]. Bing's recognition theorem was generalized by Hass and Thompson [1989] and Kobayashi and Nishi [1994]

Koda is supported by a JSPS Postdoctoral Fellowship for Research Abroad, and by the Grant-in-Aid for Young Scientists (B), JSPS KAKENHI grant number 26800028.
Ozawa is partially supported by the Grant-in-Aid for Scientific Research (C), JSPS KAKENHI grant numbers 23540105 and 26400097.
MSC2010: primary 57M25; secondary 57M15, 57N10, 57Q35.
Keywords: knots, homotopies, transient, persistent, submanifolds of the 3-sphere.

proving that a closed, connected 3-manifold M admits a genus-g Heegaard splitting if and only if there exists a genus-g handlebody V embedded in M such that every knot in M can be moved by an isotopy to lie within V. We note that, as mentioned in [Nakamura 2015], the *homotopy* version of this statement holds when $g = 1$, again by the Poincaré conjecture, whereas the higher genus case fails in general. A result of Brin, Johannson, and Scott [Brin et al. 1985] can also be regarded as a work following Fox's program. This result asserts that if every knot in M can be moved by a homotopy to lie within a collar neighborhood of the boundary ∂M, then there exists a component F of ∂M such that the natural map $\pi_1(F) \to \pi_1(M)$ induced by the inclusion is surjective. In particular, for a compact, connected, orientable, irreducible, boundary-irreducible 3-manifold M, they proved that if every knot in M can be moved by a homotopy to lie within a collar neighborhood of ∂M, then M is homeomorphic to the 3-ball or the product $\Sigma \times [0, 1]$, where Σ is a closed, orientable surface of genus at least one. In the present paper, we will consider a relative version of Fox's program. Namely, we discuss "(extrinsic) knot theories" in 3-subspaces of the 3-sphere S^3 in order to characterize how the 3-subspaces are embedded in S^3.

Let M be a compact, connected, proper 3-submanifold of S^3. We say that M is *unknotted* if its exterior is a disjoint union of handlebodies. A famous theorem of Fox [1948] says that each M can be reembedded in S^3 so that its image is unknotted. A reembedding satisfying this property is called a *Fox reembedding*. Intuitively speaking, unknottedness of $M \subset S^3$ implies that M is embedded in S^3 in one of the "simplest" ways. We note that if M is a handlebody, an unknotted M in S^3 is actually unique up to isotopy [Waldhausen 1968]. The uniqueness up to isotopy and a reflection holds for each knot exterior by a celebrated result of Gordon and Luecke [1989]. However, in other cases M usually admits many mutually nonisotopic Fox reembeddings into S^3.

The unknottedness of a 3-submanifold, and so the existence of a Fox reembedding, can be considered for an arbitrary closed, connected 3-manifold. Scharlemann and Thompson [2005] generalized the above theorem of Fox by proving that any compact, connected, proper 3-submanifold of an irreducible non-Haken 3-manifold N admits a Fox reembedding into N or S^3. Another generalization is given by Nakamura [2015] who proved that a compact, connected, proper 3-submanifold M of a closed, connected 3-manifold N admits a Fox reembedding into N if every knot in N can be moved by an isotopy to lie within M. Here we remark that the property that *every knot in N can be moved by an isotopy to lie within M* does *not* imply that M itself is unknotted in N. This can be seen for example by considering the case where $N = S^3$ and M is not unknotted. In this paper, we will show that the property of a compact, connected, proper 3-submanifold M of S^3 that *every knot in M can be moved by a homotopy in M to be the trivial knot in S^3* implies that M is unknotted in S^3. Following [Letscher 2012], we say that a knot K in M is *transient in M* if

K can be deformed by a homotopy in M to be the trivial knot in S^3; K is said to be *persistent in M* otherwise. Using this terminology, we can state our main theorem:

Theorem 3.2. *Let M be a compact, connected, proper 3-submanifold of S^3. Then every knot in M is transient in M if and only if M is unknotted.*

Roughly speaking, the above theorem implies that a (homotopic) property of knots in M deduces an isotopic property of M inside S^3. We remark that the property that a given knot $K \subset M$ is transient is *extrinsic* with respect to the embedding $M \hookrightarrow S^3$, in the sense that it depends not only on the pair (M, K) but also on the way M is embedded in S^3. Indeed, we can find a persistent knot in a certain genus-two handlebody V embedded in S^3 in such a way that there exists another embedding of V into S^3 such that the reembedded knots in the reembedded V is transient. See Section 3. Now, we can say a little more precisely what is the relative version of Fox's program; we expect that extrinsic properties for knots in a compact, connected, proper 3-submanifold of S^3 distinguish the isotopy class of M inside S^3. Our main theorem is a first step for the program. To obtain the theorem, we discuss the relation between certain group-theoretic and homotopic properties of knots in a compact 3-manifold, which can be of independent interest. See Section 1.

Given a knot K in a compact, connected, proper 3-submanifold M of S^3, it is actually difficult in general to detect if K is persistent in M. One method provided by Letscher [2012] uses what he calls the *persistent Alexander polynomial*. In Section 4, we provide examples of persistent knots in a 3-subspace of S^3 whose persistence are shown by using the notion of *persistent lamination* and *accidental surface*.

In Section 5, we will introduce an integer-valued invariant, the *transient number* of knots in S^3, whose definition is related to Theorem 3.2 as follows. Given a knot K in S^3, we may consider a system of simple arcs in S^3 with their endpoints in K such that K is transient in a regular neighborhood of the union of K and the arcs. The transient number $\mathrm{tr}(K)$ is then defined to be the minimal number of simple arcs in such a system. By an easy observation, we see that the transient number is bounded from above by both the unknotting number and the tunnel number. Further, we will give a knot K that attains $\mathrm{tr}(K) = 1$ while $u(K) = t(K) = 2$, where $u(K)$ and $t(K)$ are the unknotting number and the tunnel number of K, respectively (see Proposition 5.2). In other words, the union of the sets of knots of unknotting number one and tunnel number one is actually a proper subset of the set of knots of transient number one. Section 6 contains some concluding remarks and open questions.

Throughout this paper, we will work in the piecewise linear category.

Notation. Let X be a subset of a given polyhedral space Y. We will denote the interior of X by $\mathrm{Int}\, X$. We will use $\mathrm{Nbd}(X; Y)$ to denote a closed regular neighborhood of X in Y. If the ambient space Y is clear from the context, we denote it briefly by $\mathrm{Nbd}(X)$. Let M be a 3-manifold. Let $L \subset M$ be a submanifold with or without boundary. When L is 1- or 2-dimensional, we write $E(L)=M\setminus\operatorname{Int}\operatorname{Nbd}(L)$. When L is 3-dimensional, we write $E(L)=M\setminus\operatorname{Int}L$. We shall often say "surfaces", "compression bodies", etc., in an ambient manifold to mean their isotopy classes.

1. Knots filling up a handlebody

Let F_g be a free group of rank g with a basis $X_g=\{x_1,x_2,\dots,x_g\}$. We set

$$X_g^{\pm}=X_g\cup\{{x_1}^{-1},{x_2}^{-1},\dots,{x_g}^{-1}\}.$$

A *word* on X_g is a finite sequence of letters of $X_g^{\pm}$. For an element x of a group G, we denote by $c_G(x)$ (or simply by $c(x)$) its conjugacy class in G.

Let G be a group with a decomposition $G=G_1*G_2$. Then G_1 and G_2 are called *free factors* of G. In particular, if $G_2\neq 1$, then G_1 is called a *proper* free factor of G. Following [Lyon 1980], we say that an element x of G *binds* G if x is not contained in any proper free factor of G. Thus, for example, an element of $\mathbb{Z}$ binds $\mathbb{Z}$ if and only if it is nontrivial. We can also see that an element of a rank-2 free group $F_2=\langle x_1,x_2\rangle$ binds F_2 if and only if it is not a power of a primitive element, where an element of a free group is said to be *primitive* if it is a member of some free basis of the free group. For example $x_1x_2x_1x_2$ does not bind F_2, while $x_1x_2x_1{x_2}^3$ binds F. See, e.g., [Osborne and Zieschang 1981] and [Cho and Koda 2015]. Primitive elements of the rank-2 free group have been well understood by, e.g., Osborne and Zieschang [1981] and Cohen, Metzler, and Zimmermann [Cohen et al. 1981], whereas their classification in a free group of higher rank is known to be a hard problem. See [Puder and Wu 2014] (and also [Shpilrain 2005]) and [Puder and Parzanchevski 2015] for some of the deepest results on this problem. On the contrary, an algorithm to detect if a given element x of a free group F_g binds F_g is given by Stallings [1999] using the combinatorics of its Whitehead graph. See (2) in Section 6. It follows immediately from the definition that if x binds G, then any element of its conjugacy class $c(x)$ binds G. In fact, if x lies in G_1 for a decomposition $G=G_1*G_2$, then $a^{-1}xa$ lies in $a^{-1}G_1a$ and $F=(a^{-1}G_1a)*(a^{-1}G_2a)$ is also a decomposition of G for any $a\in G$.

Let K be an oriented knot in a 3-manifold M. We denote by $c_{\pi_1(M)}(K)$ (or simply by $c(K)$) the conjugacy class in $\pi_1(M)$ defined by the homotopy class of K. Here we recall that two oriented knots K and K' in M are homotopic in M if and only if

$$c_{\pi_1(M)}(K)=c_{\pi_1(M)}(K').$$

We say that K *binds* $\pi_1(M)$ if an element (and so every element) of $c(K)$ binds $\pi_1(M)$. It is clear by definition that, if $\bar{K}$ is the knot K with the reversed orientation, K binds $\pi_1(M)$ if and only if $\bar{K}$ also does. For this reason, we can say whether or not a knot K binds $\pi_1(M)$, while ignoring the orientation of K.

Let M be a compact 3-manifold and F a subsurface of ∂M, or a surface properly embedded in M. Here we note that F is possibly disconnected. Recall that F is said to be *compressible* if

(1) there exists a component of F that bounds a 3-ball in M, or

(2) there exists an embedded disk D in M, called a *compression disk* for F, such that $D \cap F = \partial D$ and such that ∂D is an essential simple closed curve on F.

Otherwise, F is said to be *incompressible*. A 3-manifold is said to be *irreducible* if it contains no incompressible 2-spheres and *boundary-irreducible* if its boundary is incompressible. The following lemma is a generalization of [Lyon 1980, Corollary 1].

Lemma 1.1. *Let M be a compact, connected, orientable, irreducible 3-manifold with nonempty boundary. Let K be an oriented simple closed curve in the boundary of M. Then $\partial M \setminus K$ is incompressible in M if and only if K binds $\pi_1(M)$.*

Proof. We fix an orientation and a base point v of K.

Suppose first that K does not bind $\pi_1(M, v)$. Then there exists a decomposition $\pi_1(M, v) = G_1 * G_2$ with $G_2 \neq 1$ and $[K] \in G_1$. Let X_i be a $K(G_i, 1)$-space, and let p be a point not in $X_1 \cup X_2$. We define $\hat{X}_1$ and $\hat{X}_2$ to be the mapping cylinders of maps from p into X_1 and X_2, respectively. Let X denote the space obtained by identifying the copy of p in $\hat{X}_1$ with that of p in $\hat{X}_2$. By the construction, we have $\pi_1(X) = G_1 * G_2$ and $\pi_2(X_1) = \pi_2(X_2) = 0$. Thus there exists a continuous map $f : M \to X$ satisfying the following properties:

(1) $f(v) = p$,

(2) the induced map $f_* : \pi_1(M) \to \pi_1(X)$ is an isomorphism with $f_*(G_i) = \pi_1(X_i)$ for $i \in \{1, 2\}$, and

(3) $f^{-1}(p)$ consists of a finite number of compression disks for ∂M.

Here we use the assumption that M is irreducible. We may assume that $|f^{-1}(p) \cap K|$ is minimal among all continuous maps $M \to X$ satisfying (1)–(3). Suppose that $f^{-1}(p) \cap K$ is nonempty. Then $f(K)$ is a loop in X with base point p that can be decomposed as

$$f(K) = \alpha_1 * \alpha_2 * \cdots * \alpha_r,$$

where each α_i lies in $\hat{X}_1$ or $\hat{X}_2$, and α_i, α_{i+1} do not lie in one of $\hat{X}_1$ and $\hat{X}_2$ at the same time. We note that $r > 1$. Suppose that no $[\alpha_i]$ is trivial in G_1 or G_2. Then $[\alpha_1], [\alpha_2], \ldots, [\alpha_r]$ is a *reduced sequence*, that is, $[\alpha_i]$ is in G_1 or G_2, and $[\alpha_i]$, $[\alpha_{i+1}]$ do not lie in one of G_1 and G_2 at the same time. On the other hand, $[f(K)]$ lies in G_1 by the assumption. This contradicts the uniqueness of reduced sequences; see Theorem 4.1 of Magnus, Karrass, and Solitar's book [Magnus et al. 1976]. Thus at least one of $[\alpha_1], [\alpha_2], \ldots, [\alpha_r]$ is trivial. Consequently, there exists a subarc α of K such that

- $\alpha \cap f^{-1}(p) = \partial\alpha$,
- $f(\alpha) \subset X$ is a contractible loop, and
- α is essential in ∂M cut off by $\partial f^{-1}(p)$.

Then using a standard technique as in [Lyon 1980, Theorem 2], f can be deformed by a homotopy to be a continuous map $f' : M \to X$ satisfying the above (1)–(3) and $|f'^{-1}(p) \cap K| < |f^{-1}(p) \cap K|$. This contradicts the minimality of $|f^{-1}(p) \cap K|$. Thus we have $f^{-1}(p) \cap K = \varnothing$. This implies that $\partial M \setminus K$ is compressible in M.

Next suppose that there exists a compression disk D for $\partial M \setminus K$ in M. Suppose that D separates M into two components M_1 and M_2, where K lies in M_1. Then $\pi_1(M)$ can be decomposed as $\pi_1(M) = \pi_1(M_1) * \pi_1(M_2)$, where $[K] \in \pi_1(M_1)$. If $\pi_1(M_2) = 1$, then $M_2 \cong B^3$ by the Poincaré conjecture proved by Perelman [2002; 2003a; 2003b]. This is a contradiction. Hence $\pi_1(M_2) \neq 1$, which implies that K does not bind $\pi_1(M)$. Suppose that D does not separate M. Let M' be M cut off by D. Then we have $\pi_1(M) = \pi_1(M') * \mathbb{Z}$ and $[K] \in \pi_1(M')$. Hence, again, K does not bind $\pi_1(M)$. $\square$

Let M be a compact, connected 3-manifold. Let K and K' be knots in M. We write $K \overset{M}{\simeq} K'$ if K and K' are homotopic in M. Let K be a knot in the interior of M. We say that K *fills* up M if, for any knot K' in the interior of M such that $K \overset{M}{\simeq} K'$, the exterior $E(K')$ is irreducible and boundary-irreducible.

Example. The knot K_1 shown on the left-hand side in Figure 1 does not fill up the handlebody V (because there exists a compression disk D for ∂V in $V \setminus K_1$ as shown), while the knot K_2 shown on the right-hand side fills up V (see Lemma 1.5).

By a *graph*, we mean the underlying space of a (possibly disconnected) finite 1-dimensional simplicial complex. A handlebody is a 3-manifold homeomorphic to a closed regular neighborhood of a connected graph embedded in the 3-sphere. The *genus* of a handlebody is defined to be the genus of its boundary surface. For a handlebody V, a *spine* is defined to be a graph Γ embedded in V so that V collapses onto Γ. By a 1-*vertex spine* we mean a spine with a single vertex. In other words,

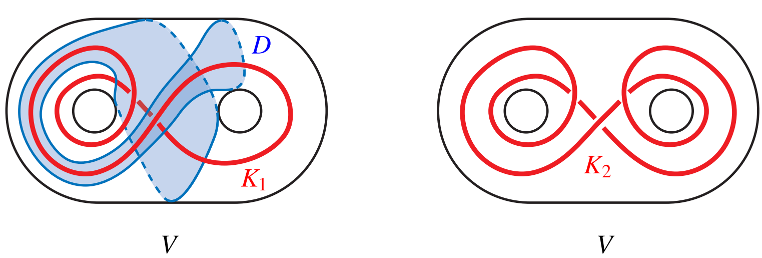

Figure 1. The knot K_1 does not fill up V, while K_2 fills up V.

a 1-vertex spine is a spine of a handlebody that is homeomorphic to a *rose*, i.e., a wedge of circles.

In the remainder of the section we fix the following:

- A handlebody V of genus g at least 1 with a base point v_0.
- A 1-vertex spine Γ_0 of V having the vertex at v_0.
- A standard basis $X = \{x_1, x_2, \ldots, x_g\}$ of $\pi_1(\Gamma_0, v_0) \cong \pi_1(V, v_0)$; that is, we can assign names $e_1^0, e_2^0, \ldots, e_g^0$ and orientations to the edges of Γ_0 so that x_i corresponds to the oriented edge e_i^0 for each $i \in \{1, 2, \ldots, g\}$.

In this setting, we identify $\pi_1(V) = \pi_1(V, v_0)$ with the free group F with basis X.

Let $\{y_1, y_2, \ldots, y_g\}$ be a basis of F, where each y_i is a word on the standard basis X. We say that a 1-vertex spine Γ of V having the vertex at v_0 is *compatible with* the basis $\{y_1, y_2, \ldots, y_g\}$ if we can assign names $e_1, e_2, \ldots, e_g$ and orientations to the edges of Γ so that a word on X corresponding to the oriented edge e_i is y_i for each $i \in \{1, 2, \ldots, g\}$.

Lemma 1.2. *For each basis $Y = \{y_1, y_2, \ldots, y_g\}$ of F, there exists a* 1*-vertex spine of V with the vertex at v_0 that is compatible with Y.*

Proof. Let φ be the automorphism of F that maps x_i to y_i for each $i \in \{1, 2, \ldots, g\}$. By [Nielsen 1924], the map φ can be factored into a composition $\varphi_n \circ \cdots \circ \varphi_2 \circ \varphi_1$, where each φ_j is an *elementary Nielsen transformation*. Here we recall that an elementary Nielsen transformation is one of the four automorphisms $\nu_1, \nu_2, \nu_3, \nu_4$ of F, where

- ν_1 switches x_1 and x_2,
- ν_2 cyclically permutes $x_1, x_2, \ldots, x_g$ to $x_2, \ldots, x_g, x_1$,
- ν_3 replaces x_1 with ${x_1}^{-1}$, and
- ν_4 replaces x_1 with $x_1 x_2$.

We refer the reader to [Magnus et al. 1976] for details. For each φ_i ($i \in \{1, 2, 3, 4\}$), it is easy to see that there exists a homeomorphism g_i of V such that g_i fixes v_0 and $g_i(\Gamma_0)$ is compatible with the basis $\{\nu_i(x_1), \nu_i(x_2), \ldots, \nu_i(x_g)\}$. Let g_j be one of $f_1,\ f_2,\ f_3,\ f_4$ corresponding to φ_j. Then it is clear from the definition that $g_n \circ \cdots \circ g_2 \circ g_1(\Gamma_0)$ is a required 1-vertex spine of V. □

Let M be a compact, connected, orientable, irreducible 3-manifold with nonempty boundary and base point v. We say that M satisfies the *strong bounded Kneser conjecture* (*SBKC*) if, whenever we have subgroups $G_1,\ G_2$ of $\pi_1(M, v)$ with $G_1 \cap G_2 = 1$, $\pi_1(M, v) = G_1 * G_2$ and $G_i \ncong 1$ ($i = 1, 2$), there exists a properly embedded disk D in M containing v such that D separates M into two components M_1 and M_2 with $\iota_{i*}(\pi_1(M_i, v)) = G_i$ ($i = 1, 2$), where $\iota_i : M_i \hookrightarrow M$ is the natural embedding. As we will see in the remark after the proof of Lemma 1.4, there exists a 3-manifold that does not satisfy the SBKC. It follows directly from Lemma 1.2 that a genus-g handlebody V satisfies the SBKC. In fact, for each decomposition $\pi_1(V, v_0) = G_1 * G_2$, we have a 1-vertex spine Γ of V having the vertex at v_0 that is compatible with the basis $\{y_1, y_2, \dots, y_g\}$, where $\{y_1, y_2, \dots, y_{g_1}\}$ is a basis of G_1 and $\{y_{g_1+1}, y_{g_1+2}, \dots, y_g\}$ is a basis of G_2. Using the spine Γ, we have the required disk D. We note that a sufficient condition for a manifold to satisfy the SBKC was given by Jaco as follows.

Lemma 1.3 [Jaco 1969]. *Let M be a compact, connected, orientable, irreducible 3-manifold with nonempty, connected boundary. Suppose that $\pi_1(M)$ is freely reduced, that is, if we have a decomposition $G = G_1 * G_2$ then neither of G_1 and G_2 is a free group. Then M satisfies the SBKC.*

Lemma 1.4. *Let M be a compact, connected, orientable, irreducible 3-manifold with nonempty boundary. Let K be an oriented knot in the interior of M. If K binds $\pi_1(M)$, then K fills up M. Moreover, the converse is true when M satisfies the SBKC.*

Proof. Suppose that K does not fill up M. Then there exists an incompressible sphere or a compression disk D for ∂M in $M \setminus K'$, where K' is a knot with $K \overset{M}{\simeq} K'$. By the same argument as in the second half of the proof of Lemma 1.1, using K' instead of K in the proof, we can show that K does not bind $\pi_1(M)$.

Next, suppose that M satisfies the SBKC and that K does not bind $\pi_1(M)$. We fix an orientation and a base point v of K. There exist subgroups G_1, G_2 of $\pi_1(M, v)$ with $G_1 \cap G_2 = 1$, $\pi_1(M, v) = G_1 * G_2$, $G_2 \not\cong 1$, and $[K] \in G_1$. If $G_1 = 1$, then K is contractible and thus we are done. Suppose that $G_1 \not\cong 1$. Then by the SBKC, there exists a properly embedded disk D in M containing v such that D separates M into two components M_1 and M_2 with $\iota_{i*}(\pi_1(M_i, v)) = G_i$ $(i \in \{1, 2\})$, where $\iota_i : M_i \hookrightarrow M$ is the natural embedding. We may assume that K is moved by a homotopy fixing v so that $|K \cap D|$ is minimal. If $|K \cap D| = 0$, we are done. Suppose that $|K \cap D| > 0$. Then $[K]$ can be decomposed into a product $x_1 x_2 \cdots x_r$, where x_i is in G_1 or G_2, and x_i, x_{i+1} do not lie in one of G_1 and G_2 at the same time. We note that $r > 1$. Since $[K] \subset G_1$, at least one, say x_{i_0}, of $x_1, x_2, \dots, x_r$ is trivial. Then moving a neighborhood of the subarc of K corresponding to x_{i_0} by a homotopy, we can reduce $|K \cap D|$. This contradicts the minimality of $|K \cap D|$. $\square$

We remark that the converse of Lemma 1.4 is not true. This can be seen as follows. Let Σ be a closed orientable surface of genus at least one. Let M be a 3-manifold obtained by attaching a 1-handle H to $\Sigma \times [0, 1]$ so as to connect $D \times \{0\}$ and $D \times \{1\}$ and so that the resulting manifold M is orientable, where D is a disk in Σ. See Figure 2. Clearly, M is compact, connected, orientable and irreducible. Let $K \subset M$ be the knot obtained by extending the core of H

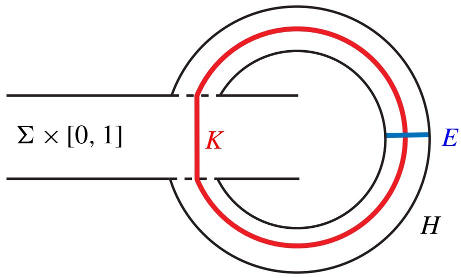

Figure 2. The manifold M.

along a vertical arc $\{*\} \times [0, 1]$ in $\Sigma \times [0, 1]$. We fix a base point v in K and an orientation of K. Then the fundamental group $\pi_1(M, v)$ can be naturally identified with $\pi_1(\Sigma) * \mathbb{Z}$, and under this identification $[K]$ is contained in the factor $\mathbb{Z}$. This implies that K does not bind $\pi_1(M)$. On the contrary, it is easy to see that the cocore E of the 1-handle H is the unique compression disk for ∂M up to isotopy. The algebraic intersection number of K and E is ± 1 after giving an orientation of E. This implies that after deforming K by any homotopy in M, K intersects E, whence K fills up M. We note that M does not satisfy the SBKC.

Lemma 1.5. *Let V be a handlebody. Then there exists a knot in the interior of V that fills up V.*

Proof. Let K be a simple closed curve in ∂V such that $\partial V \setminus K$ is incompressible in V. Such a simple closed curve does exist. In fact, the simple closed curve shown in Figure 3 satisfies this condition (see for instance [Wu 1996, Section 1]). Then by Lemma 1.1 K binds $\pi_1(V)$. It follows from Lemma 1.4 that a knot obtained by moving K by an isotopy to lie in the interior of V fills up V. □

2. Knots filling up a 3-subspace of the 3-sphere

Let V be a handlebody. A (possibly disconnected) subgraph of a spine of V is called a *subspine* if it does not contain a contractible component. A *compression body* W is the complement of an open regular neighborhood of a (possibly empty) subspine Γ

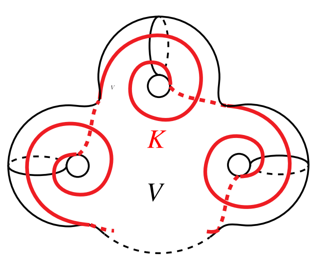

Figure 3. The surface $\partial V \setminus K$ is incompressible in V.

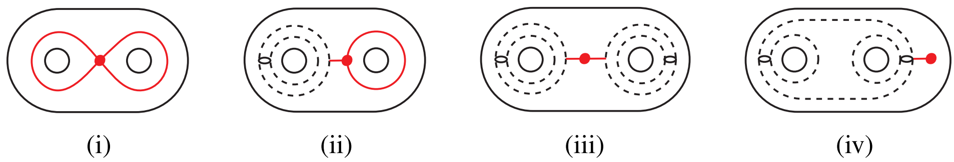

Figure 4

of a handlebody V. The component $\partial_+ W = \partial V$ is called the *exterior boundary* of W, and $\partial_- W = \partial W \setminus \partial_+ W = \partial \operatorname{Nbd}(\Gamma)$ is called the *interior boundary* of W. We remark that the interior boundary is incompressible in W; see [Bonahon 1983].

For a compression body W, a *spine* is defined to be a graph Γ embedded in W such that

(1) $\Gamma \cap \partial W = \Gamma \cap \partial_- W$ consists only of vertices of valence one, and

(2) W collapses onto $\Gamma \cup \partial_- W$.

We note that this is a generalization of a spine of a handlebody. We also note that if V is a handlebody and Γ is a subspine of $\hat{\Gamma}$ of V such that $W \cong V \setminus \operatorname{Int} \operatorname{Nbd}(\Gamma; V)$, then $\hat{\Gamma} \setminus \operatorname{Int} \operatorname{Nbd}(\Gamma; V)$ is a spine of W. As a generalization of the case of handlebodies, a 1-*vertex spine* of a compression body W is defined to be a (possibly empty) connected spine Γ such that

(1) Γ is homeomorphic to the empty set, an interval, a circle, or a graph with a single vertex of valence at least 3,

(2) Γ intersects each component of $\partial_- W$ in a single univalent vertex, and

(3) Γ has no univalent vertices in the interior of W.

If Γ is an interval or a circle, we regard it as a graph containing a unique vertex of valence 2. The spines shown in Figure 4(i)–(iii) are 1-vertex spines while the one shown in Figure 4(iv) is not so because it has a univalent vertex in the interior of the illustrated compression body. We call a vertex of valence at least 2 the *interior vertex*. We note that every 1-vertex spine has a unique interior vertex. This is the reason why it is named so.

Let W be a compression body. Suppose that $\partial_- W$ consists of n closed surfaces $\Sigma_1, \Sigma_2, \dots, \Sigma_n$. A (possibly empty) set $\mathcal{D} = \{D_1, D_2, \dots, D_m, E_{\Sigma_1}, E_{\Sigma_2}, \dots, E_{\Sigma_n}\}$ of pairwise disjoint compression disks for $\partial_+ W$ is called a *cut-system* for W if

(1) each E_{Σ_i} separates from W a component that is homeomorphic to $\Sigma_i \times [0, 1]$ and contains Σ_i,

(2) W cut off by $E_{\Sigma_1} \cup E_{\Sigma_2} \cup \cdots \cup E_{\Sigma_n}$ has at most one handlebody component V, and

(3) $D_1 \cup D_2 \cup \cdots \cup D_m$ cuts off V into a single 3-ball.

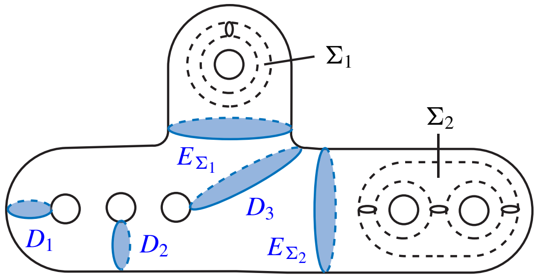

Figure 5. A cut system.

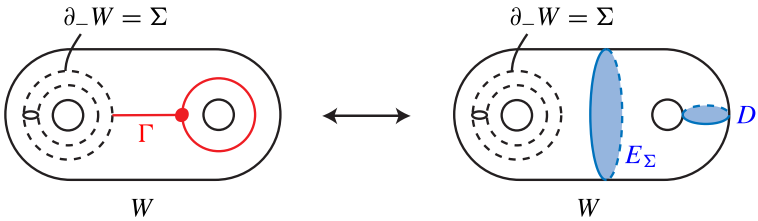

Figure 6. Poincaré–Lefschez duality.

See Figure 5. We note that if $W = \Sigma \times [0, 1]$, where Σ is a closed orientable surface, then $m = n = 0$. If W is a handlebody, then n is 0 and m is its genus.

By virtue of Poincaré–Lefschez duality, we have a one-to-one correspondence between the 1-vertex spines and cut-systems of a compression body W modulo isotopy (see Figure 6). The correspondence can be described as follows. The 1-vertex spine Γ *dual to* a given cut-system $\mathcal{D}$ for a compression body W is obtained by regarding a regular neighborhood of each disk D in $\mathcal{D}$ as a 1-handle with D as the cocore, and then extending the core arcs of the 1-handles in each component W_0 of the exterior of the union of the disks in $\mathcal{D}$ in such a way that

(1) if W_0 is a 3-ball, then the extension is given by radial arcs, and

(2) if W_0 is the product of a closed surface with an interval, then the extension is given by a vertical arc.

By conversing the construction, we get the cut-system *dual to* a 1-vertex spine of W.

Let V be a handlebody of genus g and Γ a subspine of V. Assume that each component of Γ is a rose. A *cut-system* for the pair (V, Γ) is a cut-system for V dual to a spine $\hat{\Gamma}$, where $\hat{\Gamma}$ is obtained by contracting a maximal subtree of a spine Γ' of V that contains Γ as a subgraph. See Figure 7.

Lemma 2.1. *Let W be a compression body. Let D be a compression disk for $\partial_+ W$. Then there exists a cut-system for W disjoint from D.*

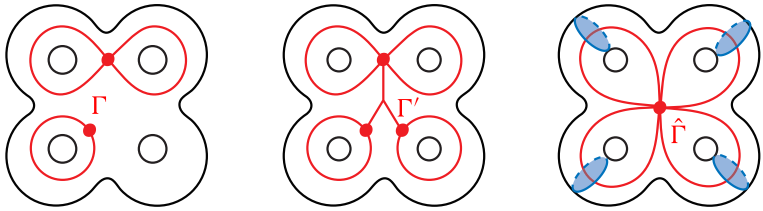

Figure 7. A cut-system for (V, Γ) is a cut-system for V dual to a spine $\hat{\Gamma}$.

Proof. We may identify W with a genus-g handlebody V with an open regular neighborhood of a subspine Γ removed. Further, we may assume that each component of Γ is a rose. Let $\Gamma_1, \Gamma_2, \dots, \Gamma_n$ be the components of Γ. Choose a cut-system $\{D_1, D_2, \dots, D_g\}$ for the pair (V, Γ) so that $|D \cap (D_1 \cup D_2 \cup \cdots \cup D_g)|$ is minimal among all cut-systems for (V, Γ). We note here that each component of the intersection $D \cap (D_1 \cup D_2 \cup \cdots \cup D_g)$ is an arc, for simple closed curves of the intersection can be eliminated by a standard argument.

Suppose for a contradiction that $D \cap (D_1 \cup D_2 \cup \cdots \cup D_g) \neq \varnothing$. Choose an outermost subdisk δ of D cut off by $D_1 \cup D_2 \cup \cdots \cup D_g$. We may assume that $\delta \cap D_1 \neq \varnothing$. Let D_1' and D_1'' be the disks obtained from D_1 by surgery along δ. Then exactly one of $\{D_1', D_2, \dots, D_g\}$ and $\{D_1'', D_2, \dots, D_g\}$, say $\{D_1', D_2, \dots, D_g\}$, is a cut-system for the handlebody V. We note that D_1'' separates the handlebody V cut off by $D_2 \cup D_3 \cup \cdots \cup D_g$. Recall that D_1 intersects Γ in at most one point. If D_1 does not intersect Γ, then it follows that $\{D_1', D_2, \dots, D_g\}$ is a cut-system for the pair (V, Γ) with $|D \cap (D_1' \cup D_2 \cup \cdots \cup D_g)| < |D \cap (D_1 \cup D_2 \cup \cdots \cup D_g)|$. This contradicts the minimality of $|D \cap (D_1 \cup D_2 \cup \cdots \cup D_g)|$. Suppose that D_1 intersects Γ. If D_1'' intersects Γ, then D_1'' cannot separate the handlebody V cut off by $D_2 \cup D_3 \cup \cdots \cup D_g$. This is a contradiction. Thus D_1' intersects Γ. This implies that $\{D_1', D_2, \dots, D_g\}$ is a cut-system for the pair (V, Γ). This contradicts, again, the minimality of $|D \cap (D_1 \cup D_2 \cup \cdots \cup D_g)|$. Thus, we have $D \cap (D_1 \cup D_2 \cup \cdots \cup D_g) = \varnothing$ and $D \cap \Gamma = \varnothing$.

From now on, we assume that each of $D_1, D_2, \dots, D_m$ does not intersect Γ, while each of $D_{m+1}, D_{m+2}, \dots, D_g$ does so. Let B be the 3-ball obtained by cutting V along $D_1 \cup D_2 \cup \cdots \cup D_g$. Then $B \cap \Gamma_i$ is a cone on an even number of points. We note that D is a separating disk in B disjoint from the cones $B \cap \Gamma$. For each $i \in \{1, 2, \dots, m\}$ let D_i^+ and D_i^- be the disks on the boundary of B coming from D_i. Then there exists a set $\{E_{\Sigma_1}, E_{\Sigma_2}, \dots, E_{\Sigma_n}\}$ of mutually disjoint disks properly embedded in B such that

(1) $E_{\Sigma_1} \cup E_{\Sigma_2} \cup \cdots \cup E_{\Sigma_n}$ is disjoint from $\Gamma \cup D \cup D_1^{\pm} \cup D_2^{\pm} \cup \cdots \cup D_g^{\pm}$, and

(2) E_{Σ_i} separates from B a 3-ball B_i such that $B_i \cap (D_1^{\pm} \cup D_2^{\pm} \cup \cdots \cup D_m^{\pm}) = \varnothing$ and $B_i \cap \Gamma = B \cap \Gamma_i$.

Now $\{D_1, D_2, \dots, D_m, E_{\Sigma_1}, E_{\Sigma_2}, \dots, E_{\Sigma_n}\}$ is a required cut-system for W. $\square$

Let M be a compact, connected, orientable, irreducible 3-manifold with connected boundary. Following [Bonahon 1983], a *characteristic compression body* W of M is defined to be a compression body embedded in M such that

(1) $\partial_+ W = \partial M$, and

(2) the closure of $M \setminus W$ is boundary-irreducible.

We remark that, for a given characteristic compression body W of M, by the irreducibility of M, every compression disk for ∂M can be moved by an isotopy to lie in W.

Theorem 2.2 [Bonahon 1983]. *A compact, connected, orientable, irreducible 3-manifold with connected boundary has a unique (up to isotopy) characteristic compression body.*

Lemma 2.3. *Let M be a compact, connected, orientable 3-manifold with connected boundary. Let W be a compression body in M such that $\partial M = \partial_+ W$. Let K be a knot in the interior of W. If K fills up M, then K fills up W. Further, when M is irreducible and W is the characteristic compression body, then K fills up M if and only if K fills up W.*

Proof. Since any knot K' in the interior of W with $K \overset{W}{\sim} K'$ satisfies $K \overset{M}{\sim} K'$, it follows immediately from the definition that if K fills up M, then K fills up W.

Suppose M is irreducible, W is the characteristic compression body, and K is a knot in W that fills up W. We will show that K fills up M. If M is a handlebody, then we have $M = W$ and there is nothing to prove. Suppose that M is not a handlebody. Then M can be decomposed as $M = W \cup X$, where $W \cap X = \partial_- W = \partial X$ and X is the union of boundary-irreducible 3-manifolds. The interior boundary $\partial_- W$ consists of a finite number of closed surfaces $\Sigma_1, \Sigma_2, \dots, \Sigma_n$ of genus at least 1. Let g_i be the genus of Σ_i ($i \in \{1, 2, \dots, n\}$). We recall that each Σ_i is incompressible in M. Suppose for a contradiction that there exists a knot K' in the interior of M with $K \overset{M}{\sim} K'$ such that ∂M is compressible in $M \setminus K'$. Let D be a compression disk for ∂M in $M \setminus K'$. We may assume that D is contained in W.

Suppose first that D does not separate W. By Lemma 2.1, there exists a cut-system for W disjoint from D. By replacing a suitable disk in the system with D, we obtain a cut-system $\mathcal{D} = \{D_1, D_2, \dots, D_m, E_{\Sigma_1}, E_{\Sigma_2}, \dots, E_{\Sigma_n}\}$ where $D = D_1$. Let Γ be the 1-vertex spine of W dual to $\mathcal{D}$. Fix a presentation of the fundamental group of each surface Σ_i as $\pi_1(\Sigma_i) = \langle a_{i,j}, b_{i,j}\ (j \in \{1, 2, \dots, g_i\}) \mid \prod_{j=1}^{g_i} [a_{i,j}, b_{i,j}] \rangle$, where we take the base point at $\Gamma \cap \Sigma_i$.

Let v_0 be the interior vertex of Γ. Let V be the unique component of W cut off by the union of disks in $\mathcal{D}$ that is homeomorphic to a handlebody. We fix a generating set $\{x_1, x_2, \dots, x_m\}$ of $\pi_1(V, v_0)$ so that an element x_i is defined by the loop in Γ dual to D_i. Then by the Seifert–van Kampen theorem, $\pi_1(W, v_0)$ is generated by the x_i, $a_{i,j}$ and $b_{i,j}$. Set

$$G=\{{x_i}^{\pm 1} \mid i \in\{1,2,\dots,m\}\}\cup\{{a_{i,j}}^{\pm 1}, {b_{i,j}}^{\pm 1}\ (j\in\{1,2,\dots,g_i\}) \mid i\in\{1,2,\dots,n\}\}.$$

Let $H_1, H_2, \dots, H_l$ be 1-handles in X attached to $\partial_- W$ so that the closure of $M \setminus (W \cup H_1 \cup H_2 \cup \cdots \cup H_l)$ is the union of handlebodies. Let $h_1, h_2, \dots, h_l$ be the element of $\pi_1(M, v_0)$ corresponding to the core of the 1-handles $H_1, H_2, \dots, H_l$, respectively. We set

$$\hat{G} = G \cup \{{h_i}^{\pm 1} \mid i \in \{1, 2, \dots, l\}\}.$$

We note that the elements of $\hat{G}$ generate the group $\pi_1(M, v_0)$. In other words, any element of $\pi_1(M, v_0)$ can be represented by a word on $\hat{G}$.

Since each Σ_i is incompressible in M, $\pi_1(W, v_0)$ is a subgroup of $\pi_1(M, v_0)$. Consider the conjugation class $c_{\pi_1(W,v_0)}(K)$. Since K fills up W, every word w on G representing an element of $c_{\pi_1(W,v_0)}(K)$ contains $x_1^{\pm 1}$.

By the existence of K', there exists a word w' on $\hat{G} \setminus \{{x_1}^{\pm 1}\}$ representing an element of $c_{\pi_1(M,v_0)}(K)$. Let u be a word on $\hat{G}$ such that $u^{-1}wu$ represents the same element as w' in $\pi_1(M, v_0)$. Let $\varphi : \pi_1(M, v_0) \to \pi_1(W, v_0)$ be the epimorphism obtained by adding the relations $h_i = 1$ for each $i \in \{1, 2, \dots, l\}$. For a word v, we denote by $\varphi(v)$ the word on G obtained from v by replacing each ${h_i}^{\pm}$ in the word with $\varnothing$. Then $\varphi(u^{-1}wu) = \varphi(u)^{-1} w \varphi(u)$ represents an element contained in $c_{\pi_1(W,v_0)}(K)$. It follows that $\varphi(w')$ is a word on $G \setminus \{{x_1}^{\pm}\}$ representing an element of $c_{\pi_1(W,v_0)}(K)$. This is a contradiction.

Next, suppose D separates W into two components W_1 and W_2. By Lemma 2.1, there exists a cut-system $\mathcal{D} = \{D_1, D_2, \dots, D_m, E_{\Sigma_1}, E_{\Sigma_2}, \dots, E_{\Sigma_n}\}$ for W disjoint from D. Without loss of generality, we can assume that the set of disks of $\mathcal{D}$ contained in W_1 is $\{D_1, D_2, \dots, D_{m_1}, E_{\Sigma_1}, E_{\Sigma_2}, \dots, E_{\Sigma_{n_1}}\}$, where $m_1 \in \{1, 2, \dots, m\}$ and $n_1 \in \{0, 1, \dots, n\}$. Here we set $n_1 = 0$ if none of $\{E_{\Sigma_1}, E_{\Sigma_2}, \dots, E_{\Sigma_n}\}$ is contained in W_1.

Let Γ be the 1-vertex spine of W dual to $\mathcal{D}$. Using the spine Γ, fix generating sets

$$G=\{{x_i}^{\pm 1} \mid i\in\{1,2,\dots,m\}\}\cup\{{a_{i,j}}^{\pm 1}, {b_{i,j}}^{\pm 1} \mid i\in\{1,2,\dots,n\},\ j\in\{1,2,\dots,g_i\}\}$$

of $\pi_1(W, v_0)$ and

$$\hat{G} = G \cup \{{h_i}^{\pm 1} \mid i \in \{1, 2, \dots, l\}\}.$$

of $\pi_1(M, v_0)$ and an epimorphism $\varphi : \pi_1(M, v_0) \to \pi_1(W, v_0)$ as above.

If $m_1 \neq m$, then, by the existence of K', there exists a word w' on $\hat{G} \setminus \{{x_1}^{\pm 1}\}$ or $\hat{G} \setminus \{{x_m}^{\pm 1}\}$ representing an element of $c_{\pi_1(M,v_0)}(K)$. By the same argument as in the case where D is nonseparating, this is a contradiction. If $m_1 = m$, then $n_1 \neq n$. Hence, by the existence of K', there exists a word w' on $\hat{G} \setminus \{{x_1}^{\pm 1}\}$ or

$\hat{G} \setminus \{{a_{n,j}}^{\pm 1}, {b_{n,j}}^{\pm 1} \mid j \in \{1, 2, \ldots, g_n\}\}$ representing an element of $c_{\pi_1(M,v_0)}(K)$. It follows that $\varphi(w')$ is a word on $G \setminus \{{x_1}^{\pm 1}\}$ or $G \setminus \{{a_{n,j}}^{\pm 1}, {b_{n,j}}^{\pm 1} \mid j \in \{1, 2, \ldots, g_n\}\}$ representing an element of $c_{\pi_1(W,v_0)}(K)$. However, this is again a contradiction because the fact that K fills up W implies that every word on G representing an element of $c_{\pi_1(W,v_0)}(K)$ contains both one of $\{{a_{n,j}}^{\pm 1}, {b_{n,j}}^{\pm 1} \mid j \in \{1, 2, \ldots, g_n\}\}$ and one of $x_1^{\pm 1}$. This completes the proof. $\square$

Theorem 2.4. *Let M be a compact, connected, orientable, irreducible 3-manifold with connected boundary. Then there exists a knot K in the interior of M that fills up M. Moreover, such a knot K can be taken to lie in* $\mathrm{Nbd}(\partial M; M)$.

Proof. If M is a handlebody, the assertion follows from Lemma 1.5. Suppose that M is not a handlebody. Let W be the characteristic compression body of M. We may identify W with the complement of an open regular neighborhood of a subspine Γ of a handlebody V. Let K be a knot in the interior of V that fills up V. Since K can be taken not to intersect a spine of V containing Γ as a subgraph, we may assume that K lies in a collar neighborhood of $\partial_+ W = \partial M$. By Lemma 2.3, K fills up W. Thus, again by Lemma 2.3, K fills up M. $\square$

3. Transient knots in a subspace of the 3-sphere

Let M be a compact, connected, proper 3-submanifold of S^3. A knot K in $M \subset S^3$ is said to be *transient in M* if K can be deformed by a homotopy in M to be the trivial knot in S^3. Otherwise, K is said to be *persistent in M*.

Example. The knot K_1 described on the left-hand side in Figure 8 is transient in the handlebody V_1 in S^3, while the knot K_2 described on the right-hand side is persistent in V_2.

The next lemma follows straightforwardly from the definition.

Lemma 3.1. *Let M be a compact, connected, proper 3-submanifold of S^3 and let N be a compact, connected 3-submanifold of M. If a knot K in N is persistent in M, then it is also persistent in N.*

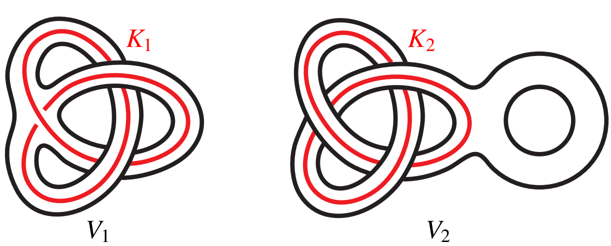

Figure 8. The knot K_1 is transient in V_1, while K_2 is persistent in V_2.

A compact, connected, proper 3-submanifold M of S^3 is said to be *unknotted* if the exterior $E(M)$ is a disjoint union of handlebodies. Otherwise M is said to be *knotted*. We recall that a theorem of Fox [1948] says that any compact, connected, proper 3-submanifold of S^3 can be reembedded in S^3 in such a way that its image is unknotted. See [Scharlemann and Thompson 2005] and [Ozawa and Shimokawa 2015] for certain generalizations and refinements of Fox's theorem.

Remark. As mentioned in the introduction, M usually admits many nonisotopic embeddings into S^3 with the unknotted image. The uniqueness holds for a handlebody by [Waldhausen 1968]. Here the uniqueness is up to isotopy for subsets of S^3, where we recall that two subsets M_1 and M_2 of S^3 are isotopic if and only if there exists an orientation-preserving homeomorphism f of S^3 carrying M_1 onto M_2. If we consider isotopies not between the embedded subsets but between embeddings, it is far from being unique even for a handlebody. This can be explained under a general setting as follows. Let M be a compact, connected 3-submanifold M that can be embedded in S^3. Then its mapping class group $\mathcal{MCG}_+(M)$ is defined to be the group of isotopy classes of orientation-preserving homeomorphisms of M. We fix an embedding $\iota_0 : M \to S^3$. Let $\mathcal{G}_{\iota_0(M)} = \mathcal{MCG}_+(S^3, \iota_0(M))$ be the mapping class group of the pair $(S^3, \iota_0(M))$, that is, the group of isotopy classes of orientation-preserving homeomorphisms of S^3 that preserve $\iota_0(M)$. See [Koda 2015] for details of this group when M is a knotted handlebody. We can define an injective homomorphism $\iota_0^* : \mathcal{G}_{\iota_0(M)} \hookrightarrow \mathcal{MCG}_+(M)$ by assigning to each homeomorphism $\varphi \in \mathcal{G}_{\iota_0(M)}$ a unique element f of $\mathcal{MCG}_+(M)$ satisfying $\varphi \circ \iota_0 = \iota_0 \circ f$. Then the set of embeddings of M into S^3 with the same image up to isotopy can be identified with the right cosets $\iota_0^*(\mathcal{G}_{\iota_0(M)}) \backslash \mathcal{MCG}_+(M)$, where the identification is given by assigning to $f \in \mathcal{MCG}_+(M)$ the embedding $\iota_0 \circ f : M \to S^3$. When M is a handlebody of genus at least two, it is clear that this is an infinite set. We note that, when $\iota_0(M)$ is an unknotted handlebody of genus two, the group $\mathcal{G}_{\iota_0(M)}$ is called the genus-two Goeritz group of S^3 and studied in [Goeritz 1933; Scharlemann 2004; Akbas 2008; Cho 2008].

Let K be a knot in M. Let f be contained in the coset $\iota_0^*(\mathcal{G}_{\iota_0(M)})\,\mathrm{id}_M$. By the observation above and the definition of the persistence of knots in $M \subset S^3$, it follows immediately that $\iota_0 \circ f(K)$ is persistent in M if and only if K is. We note that if f is not contained in the coset $\iota_0^*(\mathcal{G}_{\iota_0(M)})\,\mathrm{id}_M$, then the knot $\iota_0 \circ f(K)$ is not necessarily persistent in M even if K is persistent in M. See Figure 9. Be that as it may be, we discuss in this paper extrinsic properties of knots embedded in submanifolds of S^3, not intrinsic ones.

Theorem 3.2. *Let M be a compact, connected, proper* 3*-submanifold of S^3. Then every knot in M is transient if and only if M is unknotted.*

Proof. Suppose first that M is unknotted, i.e., $M = S^3 \setminus \operatorname{Int} \operatorname{Nbd}(\Gamma)$, where Γ is a graph embedded in M. Let K be a knot in M. Considering a diagram of the spatial

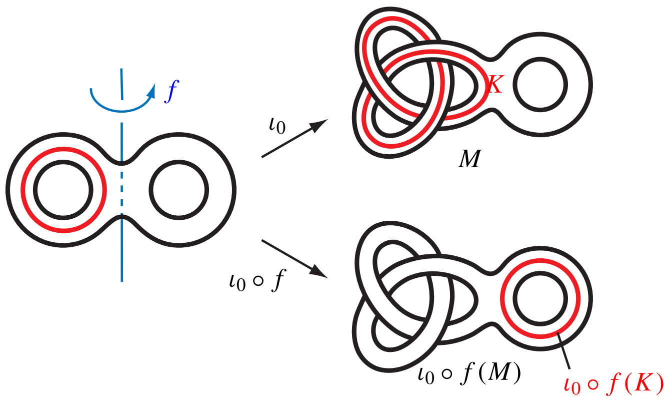

Figure 9. Persistence is an extrinsic property.

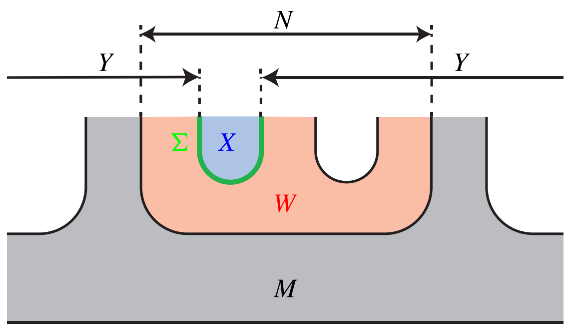

Figure 10. The configurations of M, N, W, Σ, X and Y.

graph $K \cup \Gamma$, we easily see that K can be converted into the trivial knot in S^3 by a finite number of crossing changes of K itself. This implies that K is transient in M.

Next suppose that M is knotted. Then there exists a component N of the exterior of M that is not a handlebody. Let W be the characteristic compression body of N. We note that if N is boundary-irreducible, then W is a collar neighborhood of ∂N in N. Since W is not a handlebody, we can take a nonempty component Σ of $\partial_- W$. Then Σ separates S^3 into two components X and Y so that X is boundary-irreducible and Y contains $M \cup W$. See Figure 10.

By Theorem 2.4, there exists a knot K lying in $\mathrm{Nbd}(\partial Y; Y)$ that fills up Y. In particular K lies in W. Thus by an isotopy we can move K to lie within M. Let $K' \subset M$ be an arbitrary knot with $K \overset{M}{\sim} K'$. Since K fills up Y, Σ is incompressible in $Y \setminus K'$. Thus Σ is incompressible in $S^3 \setminus K'$. This implies that K' is not the trivial knot in S^3. Therefore K is persistent in M. $\square$

Remark. Let M be a compact, connected, knotted, proper 3-submanifold of S^3. In the proof of Theorem 3.2, we explained how to obtain a knot in M that is persistent

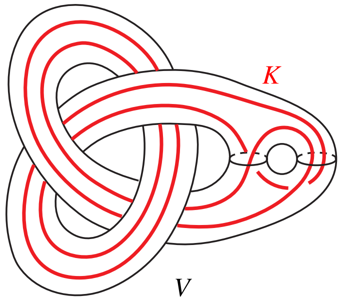

Figure 11. The knot K fills up V, whereas K is transient in V.

in M. In the process, some readers may have guessed that if a knot $K \subset M$ filled up M, then K would already be persistent. If so, the process to consider the characteristic compression body of a nonhandlebody component of the exterior in the proof would not be necessary. However, the guess is not true in fact. Let K be the knot in the genus-two knotted handlebody $V \subset S^3$ as shown in Figure 11. Then we see that K fills up V by the same reason as in the proof of Lemma 1.5 (see also (2) in Section 6, whereas K is apparently transient in V.

4. Construction of persistent knots

Persistent laminations and persistent knots. Let M be a compact, connected, proper 3-submanifold of S^3 whose exterior consists of boundary-irreducible 3-manifolds. It is easy to see that every knot filling up M is persistent in M. Indeed, if a knot K in M fills up M, then each component of ∂M will be an incompressible surface in the exterior of any knot K' homotopic to K in V, hence K' is not the trivial knot in S^3. However, the converse is false in general as we see now:

Proposition 4.1. *There exists a genus-two handlebody V embedded in S^3 with the boundary-irreducible exterior such that there exists a knot $K \subset V$ which is persistent in V, and which does not fill up V.*

Proof. Let V be the genus-two handlebody in S^3 and K the knot in V as shown in Figure 12. We note that the handlebody V is the exterior of Brittenham's branched surface [1999] constructed from a disk spanning the trivial knot in S^3. In particular, the exterior of V is boundary-irreducible. We note that K does not fill up V since there exists a compression disk D for ∂V in $V \setminus K$ as shown in the figure.

We will show that K is persistent in V. As illustrated in the figure, there are meridian disks D_1, D_2 of V each of which intersects K once and transversely. Let K' be any knot homotopic to K in V. Then K' intersects each of D_1 and D_2 at least once. By [Hirasawa and Kobayashi 2001] or [Lee and Oh 2002], which generalizes the result of [Brittenham 1999], in the exterior of V there exists a

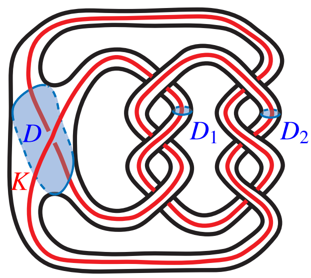

Figure 12. A handlebody V in S^3 with the boundary-irreducible exterior such that there exists a knot $K \subset V$ which is persistent in V, and which does not fill up V.

persistent lamination, that is, an essential lamination that remains essential after performing any nontrivial Dehn surgeries along K'. This implies that K' is not the trivial knot. Thus K is persistent in V. □

Accidental surfaces and persistent knots. A closed essential surface Σ in the exterior of a knot K in the 3-sphere is called an *accidental surface* if there exists an annulus A, called an *accidental annulus*, embedded in the exterior $E(K)$ such that

- the interior of A does not intersect $\Sigma \cup \partial E(K)$,
- $A \cap \Sigma \neq \varnothing$ and $A \cap \partial E(K) \neq \varnothing$, and
- $A \cap \Sigma$ and $A \cap \partial E(K)$ are essential simple closed curves in Σ and $\partial E(K)$, respectively.

In [Ichihara and Ozawa 2000] it is shown that, for each accidental surface in the exterior of a knot in S^3, the boundary curves of accidental annuli determine a unique slope on the boundary of a regular neighborhood of the knot. This slope is called an *accidental slope* for Σ. By the work of Culler, Gordon, Luecke, and Shalen [Culler et al. 1987], an accidental slope is either meridional or integral.

Proposition 4.2. *Let M be a compact, connected, proper 3-submanifold of S^3 with connected boundary such that the exterior of M is boundary-irreducible. Let K be a knot in M such that ∂M is incompressible in $M \setminus K$. If ∂M is an accidental surface with integral accidental slope in the exterior of K, then K is persistent in the submanifold M of S^3 bounded by Σ and containing K.*

Proof. Let $A \subset M$ be an accidental annulus connecting K and a simple closed curve in ∂M. Using this annulus, we move K to a knot K^* lying in ∂M by an isotopy. Since ∂M is incompressible in $E(K)$, $\partial M \setminus K^*$ is incompressible in M. Thus by Lemma 1.1 K^* binds $\pi_1(M)$, and so does K. By Lemma 1.4, K fills up M. Let

$K' \subset M$ be an arbitrary knot lying in the interior of M with $K \overset{M}{\sim} K'$. Since K fills up M, ∂M is incompressible in $M \setminus K'$. Thus ∂M is incompressible in $S^3 \setminus K'$. This implies that K' is not the trivial knot in S^3. Therefore, K is persistent in M. □

5. Transient number of knots

Let K be a knot in S^3. A *crossing move* on a knot K is the operation of passing one strand of K through another. The *unknotting number* $u(K)$ of K, which was first defined by Wendt [1937], is then the minimal number of crossing moves required to convert the knot into the trivial knot. We note that to each crossing move we can associate a simple arc α in S^3 such that $\alpha \cap K = \partial\alpha$ and such that the crossing move is performed in $\mathrm{Nbd}(\alpha)$.

An *unknotting tunnel system* for K is a set $\{\gamma_1, \gamma_2, \dots, \gamma_n\}$ of mutually disjoint simple arcs in S^3 such that $\gamma_i \cap K = \partial\gamma_i$ for each $i \in \{1, 2, \dots, n\}$ and such that the exterior of the union $K \cup \gamma_1 \cup \gamma_2 \cup \cdots \cup \gamma_n$ is a handlebody. The *tunnel number* $t(K)$ of K, first defined in [Clark 1980], is the minimal number of arcs in any of the unknotting tunnel systems for K.

We introduce a new invariant for a knot in the 3-sphere that is strongly related to the above two classical invariants. We define a *transient system* for K to be a set $\{\tau_1, \tau_2, \dots, \tau_n\}$ of mutually disjoint simple arcs in S^3 such that $\tau_i \cap K = \partial\tau_i$ for each $i \in \{1, 2, \dots, n\}$ and such that K is transient in $\mathrm{Nbd}(K \cup \tau_1 \cup \tau_2 \cup \cdots \cup \tau_n)$. The *transient number* $\mathrm{tr}(K)$ of K is defined to be the minimal number of arcs in any of the transient systems for K.

Proposition 5.1. *Let K be a knot in S^3. Then* $\mathrm{tr}(K) \leqslant u(K)$ *and* $\mathrm{tr}(K) \leqslant t(K)$.

Proof. Suppose that $u(K) = m$. Let $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ be a set of mutually disjoint simple arcs associated to m crossing moves that convert K into the trivial knot. Then K is transient in the handlebody $\mathrm{Nbd}(K \cup \alpha_1 \cup \alpha_2 \cup \cdots \cup \alpha_m)$. In other words, $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ is a transient tunnel system for K. This implies that $\mathrm{tr}(K) \leqslant m$.

Suppose that $t(K) = n$. Let $\{\gamma_1, \gamma_2, \dots, \gamma_n\}$ be an unknotting tunnel system for K. Since the handlebody $\mathrm{Nbd}(K \cup \gamma_1 \cup \gamma_2 \cup \cdots \cup \gamma_n)$ is unknotted, K is transient in $\mathrm{Nbd}(K \cup \gamma_1 \cup \gamma_2 \cup \cdots \cup \gamma_n)$ by Theorem 3.2. This implies that $\mathrm{tr}(K) \leqslant n$. □

Proposition 5.2. *There exists a knot K in S^3 with* $\mathrm{tr}(K) = 1$ *and* $u(K) = t(K) = 2$.

Proof. Let K be the satellite knot of the figure-eight knot shown in Figure 13. Clearly, the genus of K is one. The transient number of K is one because K admits a transient tunnel as shown in the figure. In [Kobayashi 1989] and [Scharlemann and Thompson 1989], it is proved that the only knots of genus one and unknotting number one are the doubled knots. It follows that the unknotting number of K is at least two. It is then straightforward to see that the unknotting number is exactly two.

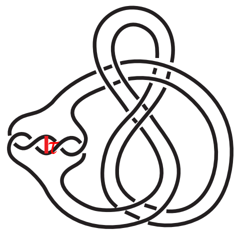

Figure 13. A knot K with $\mathrm{tr}(K) = 1$ and $u(K) = t(K) = 2$.

It is proved in [Morimoto and Sakuma 1991] that the only nonsimple knots having unknotting tunnels are certain satellites of torus knots. It follows that the tunnel number of K is at least two. It is then straightforward to see that the tunnel number is exactly two. □

6. Concluding remarks

(1) Let M be a compact, connected, proper 3-submanifold of S^3. Let K be a knot in the interior of M. In the earlier sections, we have introduced various homotopic properties of knots in M. We summarize their relations. We say that K is *accidental* in M if K can be moved to a knot K' in ∂M by a homotopy in M so that $\partial M \setminus K'$ is incompressible in M. Then we have the following:

(a) If K is accidental, then K binds $\pi_1(M)$ (see Lemma 1.1).

(b) If K binds $\pi_1(M)$, then K fills up M (see Lemma 1.4).

(c) By (a) and (b), if K is accidental, then K fills up M.

The converse of each of these is false. To see this, suppose that M is the exterior of a nontrivial knot in S^3. We note that $\pi_1(M)$ is freely indecomposable by the Kneser conjecture. Let K be a knot in M that cannot be moved by any homotopy in M to lie in ∂M. Such a knot K always exists by, for instance, the work of Brin, Johannson, and Scott [Brin et al. 1985]. This implies that K binds $\pi_1(M)$, whereas K is not accidental in M. A somewhat more subtle example is shown on the left in Figure 14. In the figure, the knot K lies in a genus-two handlebody V, and thus K can be moved by homotopy to lie within a collar neighborhood of ∂V. If K is accidental, then by attaching a 2-handle to V we obtain a 3-manifold M with toroidal boundary whose fundamental group has the presentation $\langle x, y \mid xyx^{-2}y^{-1}\rangle$. This group is called the Baumslag–Solitar group, BS(1), and is known not to be a 3-manifold group; see the work of Aschenbrenner, Friedl, and Wilton [Aschenbrenner et al. 2015]. This implies that K is not accidental in V. On the other hand, it follows straightforwardly

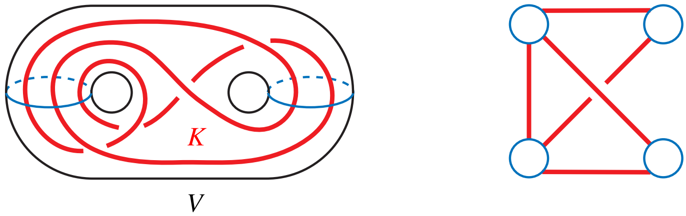

Figure 14. The knot K binds V and is not accidental in V.

from Theorem 6.1 that K binds V since the corresponding Whitehead graph, shown on the right in Figure 14, is connected and contains no cut vertex.

The remark after the proof of Lemma 1.4 shows that the converse of Lemma 1.4 is false. However, the 3-manifold M introduced in the example is not embeddable in S^3. To have a counterexample of the converse of (b), let Σ be a closed orientable surface of genus at least one. Let M be an orientable 3-manifold obtained by attaching a 1-handle to each component of $\partial(\Sigma \times [0, 1])$. We note that M can be embedded in S^3. Let D_0 and D_1 be the cocore of the 1-handles. Then we can easily show as in the remark that there exists a knot K in M, intersecting each of D_0 and D_1 once and transversely, that fills up M, whereas K does not bind $\pi_1(M)$. The relations of these three intrinsic properties are shown on the left-hand side in Figure 15. It is worth noting that, to show that a given knot K in $M \subset S^3$ is persistent, we have used an intrinsic property of K in a subset of S^3 containing M. See Theorem 3.2 and Propositions 4.1 and 4.2.

(2) Let F_g be a rank-g free group. As mentioned in Section 1, an algorithm to detect whether a given element x of a free group F_g binds F_g is described by Stallings using the combinatorics of its Whitehead graph. In fact, the following is proved:

Theorem 6.1 [Stallings 1999]. *Let x be a cyclically reduced word on the set $X_g = \{x_1, x_2, \dots, x_g\}$. If the Whitehead graph of x is connected and contains no cut vertex, then x binds F_g.*

For a simple closed curve in the boundary of a handlebody, this can be seen clearly as follows. Let x be an element of the rank-g free group F_g. We identify

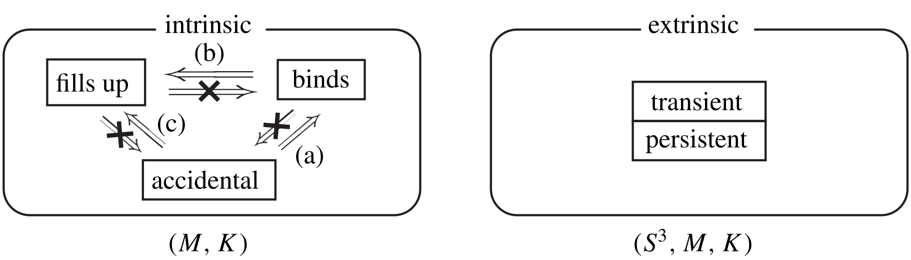

Figure 15. Correlation diagrams of extrinsic and intrinsic properties.

F_g with the fundamental group of a genus-g handlebody. In the case of $M = V_g$ in Lemma 1.1, which is actually [Lyon 1980, Corollary 1], we have seen that if x can be represented by an oriented simple closed curve K in ∂V_g, then x binds F_g if and only if $\partial V_g \setminus K$ is incompressible. On the other hand, Starr [1992] (see also [Wu 1996, Theorem 1.2]) showed that $\partial V_g \setminus K$ is incompressible if and only if there is a complete meridian disk system $D_1, D_2, \dots, D_g$ of V_g such that the planar graph with "fat" vertices obtained by cutting ∂V_g along $\bigcup_{i=1}^{g} D_i$ is connected and contains no cut vertex. This graph is actually nothing else but the Whitehead graph of x. (As explained in [Stallings 1999], we can obtain a geometric interpretation of this for an arbitrary element of F_g if we consider the connected sum of g copies of $S^2 \times S^1$ instead of V_g.)

(3) Let M be a compact, connected, proper 3-submanifold of S^3. In the proofs of Theorem 3.2 and Propositions 4.1 and 4.2, we provided a way to show that a given knot $K \subset M$ is persistent in M. The key idea is to find an essential surface (or lamination) in the exterior of M that is also essential in the exterior of any knot K' homotopic to K in M. As mentioned in the introduction, another way to show persistence was provided by Letscher [2012] and uses what he calls the *persistent Alexander polynomial*.

Problem 1. Provide more methods for detecting whether a knot $K \subset M$ is persistent.

(4) As we have summarized in Figure 15, the only extrinsic property of knots in a 3-subspace of S^3 we have considered in the present paper is transience (or persistence). Using this property, we have actually gotten an "if and only if" condition for a 3-subspace of S^3 being unknotted in Theorem 3.2. This is a first step for a relative version of Fox's program and further progress will be expected.

Problem 2. Consider other extrinsic properties of knots in $M \subset S^3$ in order to characterize how M is embedded in S^3.

We note that the case where M is a handlebody is already a very interesting problem. See, e.g., [Ishii 2008; Koda 2015; Koda and Ozawa 2015].

(5) As mentioned in the introduction, the unknottedness of a 3-submanifold can be considered for an arbitrary closed, connected 3-manifold. Thus it is natural to ask:

Question 1. Can Theorem 3.2 be generalized for M in an arbitrary 3-manifold N?

(6) Finally, in Section 5, we defined an integer-valued invariant $\mathrm{tr}(K)$, the transient number, for a knot K in S^3. This invariant is nice in the sense that it shows the knots of unknotting number 1 and those of tunnel number 1 from the same perspective as we have seen in Proposition 5.1. However, it remains unknown whether there exists a knot whose transient number is more than 1.

Question 2. Can the transient number $\mathrm{tr}(K)$ be arbitrarily large?

Acknowledgements

The authors wish to express their gratitude to Makoto Sakuma and Masakazu Teragaito for helpful suggestions and comments. Part of this work was carried out while Koda was visiting Università di Pisa as a JSPS Postdoctoral Fellow for Research Abroad. He is grateful to the university and its staff for the warm hospitality. Finally, the authors are grateful to the anonymous referee for his or her comments which improved the exposition.

References

[Akbas 2008] E. Akbas, "A presentation for the automorphisms of the 3-sphere that preserve a genus two Heegaard splitting", *Pacific J. Math.* **236**:2 (2008), 201–222. MR 2009d:57029 Zbl 1157.57002

[Aschenbrenner et al. 2015] M. Aschenbrenner, S. Friedl, and H. Wilton, *3-manifold groups*, European Mathematical Society, Zürich, 2015. Zbl 06472948

[Bing 1958] R. H. Bing, "Necessary and sufficient conditions that a 3-manifold be S^3", *Ann. of Math.* (2) **68** (1958), 17–37. MR 20 #1973 Zbl 0081.39202

[Bonahon 1983] F. Bonahon, "Cobordism of automorphisms of surfaces", *Ann. Sci. École Norm. Sup.* (4) **16**:2 (1983), 237–270. MR 85j:57011 Zbl 0535.57016

[Brin et al. 1985] M. Brin, K. Johannson, and P. Scott, "Totally peripheral 3-manifolds", *Pacific J. Math.* **118**:1 (1985), 37–51. MR 86i:57012 Zbl 0525.57010

[Brittenham 1999] M. Brittenham, "Persistently laminar tangles", *J. Knot Theory Ramifications* **8**:4 (1999), 415–428. MR 2000f:57005 Zbl 0941.57013

[Brody 1960] E. J. Brody, "The topological classification of the lens spaces", *Ann. of Math.* (2) **71** (1960), 163–184. MR 22 #7125 Zbl 0119.18901

[Cho 2008] S. Cho, "Homeomorphisms of the 3-sphere that preserve a Heegaard splitting of genus two", *Proc. Amer. Math. Soc.* **136**:3 (2008), 1113–1123. MR 2009c:57029 Zbl 1149.57025

[Cho and Koda 2015] S. Cho and Y. Koda, "Disk complexes and genus two Heegaard splittings for nonprime 3-manifolds", *Int. Math. Res. Not.* **2015**:12 (2015), 4344–4371. MR 3356757 Zbl 1322.57017

[Clark 1980] B. Clark, "The Heegaard genus of manifolds obtained by surgery on links and knots", *Int. J. Math. Math. Sci.* **3**:3 (1980), 583–589. MR 81j:57004 Zbl 0447.57008

[Cohen et al. 1981] M. Cohen, W. Metzler, and A. Zimmermann, "What does a basis of $F(a, b)$ look like?", *Math. Ann.* **257**:4 (1981), 435–445. MR 82m:20028 Zbl 0458.20028

[Culler et al. 1987] M. Culler, C. M. Gordon, J. Luecke, and P. B. Shalen, "Dehn surgery on knots", *Ann. of Math.* (2) **125**:2 (1987), 237–300. MR 88a:57026 Zbl 0633.57006

[Eilenberg 1949] S. Eilenberg, "On the problems of topology", *Ann. of Math.* (2) **50** (1949), 247–260. MR 10,726b Zbl 0034.25304

[Fox 1948] R. H. Fox, "On the imbedding of polyhedra in 3-space", *Ann. of Math.* (2) **49** (1948), 462–470. MR 10,138c Zbl 0032.12502

[Goeritz 1933] L. Goeritz, "Die Abbildungen der Brezelfläche und der Vollbrezel vom Geschlecht 2", *Abh. Math. Sem. Univ. Hamburg* **9**:1 (1933), 244–259. MR 3069602 Zbl 0007.08102

[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 90a:57006a Zbl 0678.57005

[Hass and Thompson 1989] J. Hass and A. Thompson, "A necessary and sufficient condition for a 3-manifold to have Heegaard genus one", *Proc. Amer. Math. Soc.* **107**:4 (1989), 1107–1110. MR 90c:57010 Zbl 0694.57006

[Hirasawa and Kobayashi 2001] M. Hirasawa and T. Kobayashi, "Pre-taut sutured manifolds and essential laminations", *Osaka J. Math.* **38**:4 (2001), 905–922. MR 2002j:57035 Zbl 1025.57027

[Ichihara and Ozawa 2000] K. Ichihara and M. Ozawa, "Accidental surfaces in knot complements", *J. Knot Theory Ramifications* **9**:6 (2000), 725–733. MR 2001f:57007 Zbl 0999.57007

[Ishii 2008] A. Ishii, "Moves and invariants for knotted handlebodies", *Algebr. Geom. Topol.* **8**:3 (2008), 1403–1418. MR 2009g:57022 Zbl 1151.57007

[Jaco 1969] W. Jaco, "Three-manifolds with fundamental group a free product", *Bull. Amer. Math. Soc.* **75** (1969), 972–977. MR 39 #4852 Zbl 0184.48804

[Kobayashi 1989] T. Kobayashi, "Minimal genus Seifert surfaces for unknotting number 1 knots", *Kobe J. Math.* **6**:1 (1989), 53–62. MR 90k:57008 Zbl 0688.57007

[Kobayashi and Nishi 1994] T. Kobayashi and H. Nishi, "A necessary and sufficient condition for a 3-manifold to have genus g Heegaard splitting (a proof of Hass–Thompson conjecture)", *Osaka J. Math.* **31**:1 (1994), 109–136. MR 96b:57020 Zbl 0822.57014

[Koda 2015] Y. Koda, "Automorphisms of the 3-sphere that preserve spatial graphs and handlebody-knots", *Math. Proc. Cambridge Philos. Soc.* **159**:1 (2015), 1–22. MR 3349329

[Koda and Ozawa 2015] Y. Koda and M. Ozawa, "Essential surfaces of non-negative Euler characteristic in genus two handlebody exteriors", *Trans. Amer. Math. Soc.* **367**:4 (2015), 2875–2904. MR 3301885 Zbl 1310.57015

[Lee and Oh 2002] S. Lee and S. Oh, "Constructing persistently laminar knots", *Topology Appl.* **124**:1 (2002), 139–143. MR 2003f:57017 Zbl 1008.57012

[Letscher 2012] D. Letscher, "On persistent homotopy, knotted complexes and the Alexander module", pp. 428–441 in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* (Cambridge, MA, 2012), edited by S. Goldwasser, Association for Computing Machinery, New York, 2012. MR 3388406

[Lyon 1980] H. C. Lyon, "Incompressible surfaces in the boundary of a handlebody: an algorithm", *Canad. J. Math.* **32**:3 (1980), 590–595. MR 82a:57010 Zbl 0443.57008

[Magnus et al. 1976] W. Magnus, A. Karrass, and D. Solitar, *Combinatorial group theory: presentations of groups in terms of generators and relations*, 2nd ed., Dover, New York, 1976. MR 54 #10423 Zbl 0362.20023

[Morimoto and Sakuma 1991] K. Morimoto and M. Sakuma, "On unknotting tunnels for knots", *Math. Ann.* **289**:1 (1991), 143–167. MR 92e:57015 Zbl 0697.57002

[Nakamura 2015] K. Nakamura, "Fox reimbedding and Bing submanifolds", *Trans. Amer. Math. Soc.* **367**:12 (2015), 8325–8346. MR 3403057 Zbl 06503651

[Nielsen 1924] J. Nielsen, "Die Isomorphismengruppe der freien Gruppen", *Math. Ann.* **91**:3-4 (1924), 169–209. MR 1512188 JFM 50.0078.04

[Osborne and Zieschang 1981] R. P. Osborne and H. Zieschang, "Primitives in the free group on two generators", *Invent. Math.* **63**:1 (1981), 17–24. MR 82i:20042 Zbl 0438.20017

[Ozawa and Shimokawa 2015] M. Ozawa and K. Shimokawa, "Dehn surgery and Seifert surface system", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) (online publication November 2015).

[Perelman 2002] G. Perelman, "The entropy formula for the Ricci flow and its geometric applications", preprint, 2002. Zbl 1130.53001 arXiv math/0211159

[Perelman 2003a] G. Perelman, "Finite extinction time for the solutions to the Ricci flow on certain three-manifolds", preprint, 2003. Zbl 1130.53003 arXiv math/0307245

[Perelman 2003b] G. Perelman, “Ricci flow with surgery on three-manifolds”, preprint, 2003. Zbl 1130.53002 arXiv math/0303109

[Puder and Parzanchevski 2015] D. Puder and O. Parzanchevski, “Measure preserving words are primitive”, *J. Amer. Math. Soc.* **28**:1 (2015), 63–97. MR 3264763 Zbl 06394341

[Puder and Wu 2014] D. Puder and C. Wu, “Growth of primitive elements in free groups”, *J. Lond. Math. Soc.* (2) **90**:1 (2014), 89–104. MR 3245137 Zbl 1308.20022

[Scharlemann 2004] M. Scharlemann, “Automorphisms of the 3-sphere that preserve a genus two Heegaard splitting”, *Bol. Soc. Mat. Mexicana* (3) **10**:Special Issue (2004), 503–514. MR 2007c:57020 Zbl 1095.57017

[Scharlemann and Thompson 1989] M. Scharlemann and A. Thompson, “Link genus and the Conway moves”, *Comment. Math. Helv.* **64**:4 (1989), 527–535. MR 91b:57006 Zbl 0693.57004

[Scharlemann and Thompson 2005] M. Scharlemann and A. Thompson, “Surfaces, submanifolds, and aligned Fox reimbedding in non-Haken 3-manifolds”, *Proc. Amer. Math. Soc.* **133**:6 (2005), 1573–1580. MR 2005j:57029 Zbl 1071.57015

[Shpilrain 2005] V. Shpilrain, “Counting primitive elements of a free group”, pp. 91–97 in *Geometric methods in group theory*, edited by J. Burillo et al., Contemporary Mathematics **372**, American Mathematical Society, Providence, RI, 2005. MR 2006b:20032 Zbl 1077.20046

[Stallings 1999] J. R. Stallings, “Whitehead graphs on handlebodies”, pp. 317–330 in *Geometric group theory down under* (Canberra, 1996), edited by J. Cossey et al., de Gruyter, Berlin, 1999. MR 2001i:57028 Zbl 1127.57300

[Starr 1992] E. N. Starr, *Curves in handlebodies*, thesis, University of California, Berkeley, CA, 1992, available at http://search.proquest.com/docview/303991902. MR 2688492

[Waldhausen 1968] F. Waldhausen, “Heegaard-Zerlegungen der 3-Sphäre”, *Topology* **7** (1968), 195–203. MR 37 #3576 Zbl 0157.54501

[Wendt 1937] H. Wendt, “Die gordische Auflösung von Knoten”, *Math. Z.* **42**:1 (1937), 680–696. MR 1545700 Zbl 0016.42005

[Wu 1996] Y.-Q. Wu, “Incompressible surfaces and Dehn surgery on 1-bridge knots in handlebodies”, *Math. Proc. Cambridge Philos. Soc.* **120**:4 (1996), 687–696. MR 97i:57021 Zbl 0888.57015

Received March 30, 2015. Revised September 11, 2015.

YUYA KODA
DEPARTMENT OF MATHEMATICS
HIROSHIMA UNIVERSITY
1-3-1 KAGAMIYAMA
HIGASHI-HIROSHIMA 739-8526
JAPAN

ykoda@hiroshima-u.ac.jp

MAKOTO OZAWA
DEPARTMENT OF NATURAL SCIENCES
KOMAZAWA UNIVERSITY
1-23-1 KOMAZAWA, SETAGAYA-KU
TOKYO 154-8525
JAPAN

w3c@komazawa-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.415

ON THE RELATIONSHIP OF CONTINUITY AND BOUNDARY REGULARITY IN PRESCRIBED MEAN CURVATURE DIRICHLET PROBLEMS

KIRK E. LANCASTER AND JARON MELIN

In 1976, Leon Simon showed that if a compact subset of the boundary of a domain is smooth and has negative mean curvature, then the nonparametric least area problem with Lipschitz continuous Dirichlet boundary data has a generalized solution which is continuous on the union of the domain and this compact subset of the boundary, even if the generalized solution does not take on the prescribed boundary data. Simon's result has been extended to boundary value problems for prescribed mean curvature equations by other authors. In this note, we construct Dirichlet problems in domains with corners and demonstrate that the variational solutions of these Dirichlet problems are discontinuous at the corner, showing that Simon's assumption of regularity of the boundary of the domain is essential.

1. Introduction

For $n \in \mathbb{N}$ with $n \geq 2$, suppose Ω is a bounded, open set in $\mathbb{R}^n$ with locally Lipschitz boundary $\partial\Omega$. Fix $H \in C^2(\mathbb{R}^n \times \mathbb{R})$ such that H is bounded and $H(x, t)$ is nondecreasing in t for $x \in \Omega$. Consider the prescribed mean curvature Dirichlet problem of finding a function $f \in C^2(\Omega) \cap C^0(\overline{\Omega})$ which satisfies

$$\operatorname{div}(Tf) = H(x, f) \quad \text{in } \Omega, \tag{1}$$

$$f = \phi \quad \text{on } \partial\Omega, \tag{2}$$

where $\phi \in C^0(\partial\Omega)$ is a prescribed function and

$$Tf = \frac{\nabla f}{\sqrt{1+|\nabla f|^2}};$$

such a function f, if it exists, is a classical solution of the Dirichlet problem. It has long been known (e.g., Bernstein in 1912) that some type of boundary curvature condition (which depends on H) must be satisfied in order to guarantee that a classical solution exists for each $\phi \in C^0(\partial\Omega)$ (e.g., [Jenkins and Serrin 1968; Serrin

MSC2010: primary 76B45, 35J93; secondary 35J62, 53A10.
Keywords: prescribed mean curvature, nonconvex corner, Dirichlet problem.

1969]). When $H \equiv 0$ and $\partial\Omega$ is smooth, this curvature condition is that $\partial\Omega$ must have nonnegative mean curvature (with respect to the interior normal direction of Ω) at each point [Jenkins and Serrin 1968]. However, Leon Simon [1976] has shown that if $\Gamma_0 \subset \partial\Omega$ is smooth (i.e., C^4), the mean curvature Λ of $\partial\Omega$ is negative on Γ_0, and Γ is a compact subset of Γ_0, then the minimal hypersurface $z = f(x)$, $x \in \Omega$, extends to $\Omega \cup \Gamma$ as a continuous function, even though f may not equal ϕ on Γ. Since [Simon 1976] appeared, the requirement that $H \equiv 0$ has been eliminated and the conclusion remains similar to that which Simon reached (see, for example, [Bourni 2011; Lau and Lin 1985; Lin 1987]).

How important is the role of boundary smoothness in the conclusions reached in [Simon 1976]? We shall show, by constructing suitable domains Ω and Dirichlet data ϕ, that the existence of a "nonconvex corner" P in Γ can cause the unique generalized (e.g., variational) solution to be discontinuous at P even if $\Gamma \setminus \{P\}$ is smooth and the generalized mean curvature Λ^* (i.e., [Serrin 1969]) of Γ at P is $-\infty$; this shows that some degree of smoothness of Γ is required to obtain the conclusions in [Simon 1976]. We shall prove the following.

Theorem 1. *Let $n \in \mathbb{N}$, $n \geq 2$, and assume there exists $\lambda > 0$ such that $|H(x,t)| \leq \lambda$ for $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$. Then there exist a domain $\Omega \subset \mathbb{R}^n$ and a point $P \in \partial\Omega$ such that*

(i) $\partial\Omega \setminus \{P\}$ *is smooth* (C^∞),

(ii) *there is a neighborhood $\mathcal{N}$ of P such that $\Lambda(x) < 0$ for $x \in \mathcal{N} \cap \partial\Omega \setminus \{P\}$, where Λ is the mean curvature of* $\partial\Omega$, *and*

(iii) $\Lambda^*(P) = -\infty$, *where Λ^* is the generalized mean curvature of* $\partial\Omega$,

and there exists Dirichlet boundary data $\phi \in C^\infty(\mathbb{R}^n)$ such that the minimizer $f \in \mathrm{BV}(\Omega)$ of

$$J(u) = \int_\Omega |Du| + \int_\Omega \int_0^u H(x,t)\,dt\,dx + \int_{\partial\Omega} |u - \phi|\,d\mathcal{H}^{n-1}, \quad u \in \mathrm{BV}(\Omega), \tag{3}$$

exists and satisfies (1), $f \in C^2(\Omega) \cap C^0(\overline{\Omega} \setminus \{P\}) \cap L^\infty(\Omega)$, $f \notin C^0(\overline{\Omega})$, *and $f \neq \phi$ in a neighborhood of P in $\partial\Omega$.*

Since there are certainly many examples of Dirichlet problems which have continuous solutions even though their domains fail to satisfy appropriate smoothness or boundary curvature conditions (e.g., by restricting to a smaller domain a classical solution of a Dirichlet problem on a larger domain), the question of necessary or sufficient conditions for the continuity at P of a generalized solution of a particular Dirichlet problem is of interest and the examples here suggest (to us) that a "Concus–Finn" type condition might yield necessary conditions for the continuity at P of solutions (see Section 5).

We view this note as analogous to other articles (e.g., [Shi and Finn 2004; Huff and McCuan 2006; 2009; Korevaar 1980]) which enhance our knowledge of the behavior of solutions of boundary value problems for prescribed mean curvature equations by constructing and analyzing specific examples. One might also compare Theorem 1 with the behavior of generalized solutions of (1)–(2) when $\partial\Omega \setminus \{P\}$ is smooth and $|H(x, \phi(x))| \leq (n-1)\Lambda(x)$ for $x \in \partial\Omega \setminus \{P\}$ (e.g., [Elcrat and Lancaster 1986; Lancaster 1985; 1988]) and with capillary surfaces (e.g., [Lancaster and Siegel 1996]).

2. Nonparametric minimal surfaces in $\mathbb{R}^3$

In this section, we will assume $n = 2$ and $H \equiv 0$; this allows us to use explicit comparison functions and illustrate our general procedure. Let Ω be a bounded, open set in $\mathbb{R}^2$ with locally Lipschitz boundary $\partial\Omega$ such that a point P lies on $\partial\Omega$ and there exist distinct rays $l^{\pm}$ starting at P such that $\partial\Omega$ is tangent to $l^+ \cup l^-$ at P. By rotating and translating the domain, we may assume $P = (0, 1)$ and there exists a $\sigma \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ such that

$$l^- = \{(r\cos(\sigma), 1 + r\sin(\sigma)) : r \geq 0\},$$
$$l^+ = \{(r\cos(\pi - \sigma), 1 + r\sin(\pi - \sigma)) : r \geq 0\},$$
$$\Omega \cap B(P, \delta) = \{(r\cos(\theta), 1 + r\sin(\theta)) : 0 < r < \delta, \theta^-(r) < \theta < \theta^+(r)\} \tag{4}$$

for some $\delta > 0$ and functions $\theta^{\pm} \in C^0([0, \delta))$ which satisfy $\theta^- < \theta^+$, $\theta^-(0) = \sigma$ and $\theta^+(0) = \pi - \sigma$; here $B(P, \delta)$ is the open ball in $\mathbb{R}^2$ centered at P of radius δ. If we set $\alpha = \frac{\pi}{2} - \sigma$, then $\alpha \in (0, \pi)$ and the angle at P in Ω of $\partial\Omega$ has size 2α. As $\sigma < 0$ goes to zero, $2\alpha > \pi$ goes to π and the (upper) region between l^- and l^+ becomes "less nonconvex" and approaches a half-plane through P. We will show that for each choice of $\sigma \in \left(-\frac{\pi}{2}, 0\right)$, there is a domain Ω as above and a choice of Dirichlet data $\phi \in C^\infty(\partial\Omega)$ such that the solution of (1)–(2) for Ω and ϕ is discontinuous at P.

Fix $\sigma \in \left(-\frac{\pi}{2}, -\frac{\pi}{4}\right)$. Let ϵ be a small, fixed parameter, say $\epsilon \in (0, 0.5)$, and let $a = a(\sigma) \in (1, 2)$ be a parameter to be determined. Set $\tau = (1 + \epsilon)\cot(-\sigma)$ and $r_1 = \sqrt{\tau^2 + (1 + \epsilon)^2}$. Define $h_{2/\pi} \in C^2\big((0, 2) \times (-1, 1)\big)$ by

$$h_{2/\pi}(x_1, x_2) = \frac{2}{\pi}\ln\left(\frac{\cos\left(\frac{\pi x_2}{2}\right)}{\sin\left(\frac{\pi x_1}{2}\right)}\right).$$

Notice that the graph of $h_{2/\pi}$ is part of Scherk's first surface, so $\operatorname{div}(Th_{2/\pi}) = 0$ on $(0, 2) \times (-1, 1)$, and $h_{2/\pi}(t, t - 1) = 0$ for each $t \in (0, 2)$. A computation using L'Hospital's Rule shows

$$\lim_{t \to 0^+} h_{2/\pi}\big((t\cos(\theta), 1 + t\sin(\theta))\big) = \tfrac{2}{\pi}\ln(-\tan(\theta)), \quad \theta \in \left(-\tfrac{\pi}{2}, 0\right). \tag{5}$$

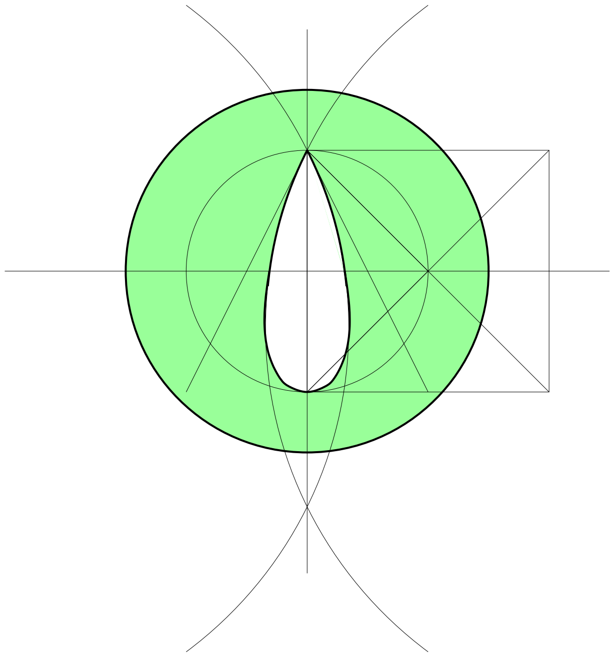

Figure 1. Ω.

Let $D = B(\mathbb{O}, 1) \cap B((\tau, -\epsilon), r_1) \cap B((-\tau, -\epsilon), r_1)$ be the intersection of three open disks and let $E \subset D$ be a strictly convex domain such that $\{x \in \partial E : x_2 < 1\}$ is a C^∞ curve, $E \cap \{x_2 \geq 0\} = D \cap \{x_2 \geq 0\}$, E is symmetric with respect to the x_2-axis, and $(0, -1) \in \partial E$; here $\mathbb{O}$ denotes $(0, 0)$. Define

$$\Omega = B(\mathbb{O}, a) \setminus \overline{E}$$

(see Figure 1); notice that $P \in \partial\Omega$ and (4) holds with the choice of σ above. If we set $C = \{(x_1, x_2) \in \mathbb{R}^2 : 0 < x_1 < 1, x_1 - 1 < x_2 < 1 - x_1\}$, then (5) implies $\sup_{x \in C \cap \partial E} h_{2/\pi}(x) < \infty$.

Let

$$m > \max\left\{ r_1 \cosh^{-1}\left(\frac{2+\sqrt{\tau^2+\epsilon^2}}{r_1}\right), \sup_{x \in C \cap \partial E} h_{2/\pi}(x) \right\}.$$

Notice that m is independent of the parameter a. Define $\phi \in C^\infty(\partial\Omega)$ by $\phi = 0$ on $\partial B(\mathbb{O}, a)$ and $\phi = m$ on ∂E. Let f be the variational solution of (1)–(2) with ϕ as given here (e.g., [Gerhardt 1974; Giusti 1978]). Since $\phi \geq 0$ on $\partial\Omega$ and $\phi > 0$ on ∂E, $f \geq 0$ in Ω (e.g., Lemma 2 (with $h \equiv 0$)) and so $f > 0$ in Ω (e.g., the Hopf boundary point lemma). Notice that $h_{2/\pi} = 0 < f$ on $\Omega \cap \partial C$ and $h_{2/\pi} < \phi$ on $C \cap \partial E = C \cap \partial\Omega$, and therefore $h_{2/\pi} < f$ on $\Omega \cap C$ (see Figure 2). Together with

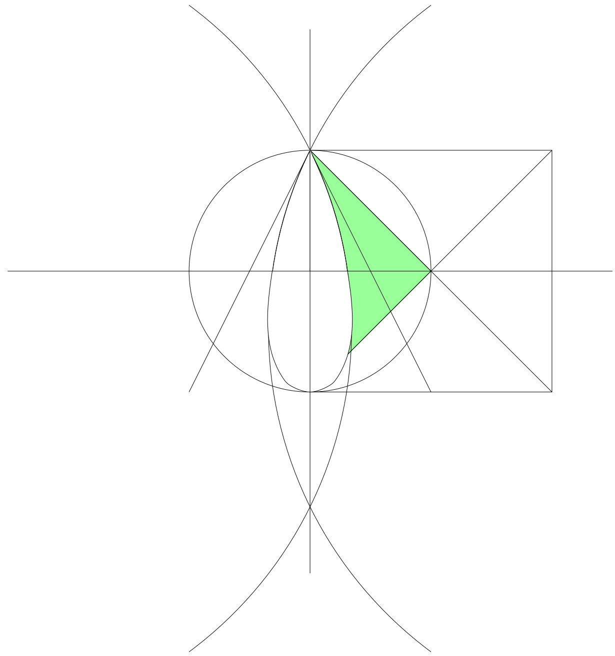

Figure 2. $\Omega \cap C$, the domain of the comparison function for (6).

(5), this implies

$$\liminf_{\Omega\cap C\ni x\to P} f(x) \geq \tfrac{2}{\pi}\ln(\tan(-\sigma)) > 0. \tag{6}$$

Set $W = B(\mathbb{O}, a) \setminus \overline{B(\mathbb{O}, 1)}$ (see Figure 3); then $W \subset \Omega$. Define the function $g \in C^\infty(W) \cap C^0(\overline{W})$ by $g(x) = \cosh^{-1}(a) - \cosh^{-1}(|x|)$ and notice that the graph of g is part of a catenoid, where $g = 0$ on $\partial B(\mathbb{O}, a)$ and $g = \cosh^{-1}(a)$ on $\partial B(\mathbb{O}, 1)$. It follows from the general comparison principle (e.g., [Finn 1986, Theorem 5.1]) that $f \leq g$ on W and therefore

$$f \leq \cosh^{-1}(a) \quad \text{on } W. \tag{7}$$

If we select $a > 1$ so that $\cosh^{-1}(a) < \frac{2}{\pi}\ln(\tan(-\sigma))$, then (6) and (7) imply that f cannot be continuous at P. Notice that [Simon 1976] implies $f \in C^0(\overline{\Omega} \setminus \{P\})$.

This example illustrates the procedure we shall use in Section 4; a somewhat similar approach was used in [Shi and Finn 2004; Korevaar 1980; Lancaster and Siegel 1996; Serrin 1969]. The case $\sigma \in \left[-\frac{\pi}{4}, 0\right)$ has a similar proof with the changes that D is the intersection of the open disk $B(\mathbb{O}, 1)$ with the interiors of two ellipses, and a Scherk surface with rhomboidal domain [Nitsche 1989, pp. 70–71] is used as a comparison surface to obtain the analog of (6); the details can be found in [Melin 2013].

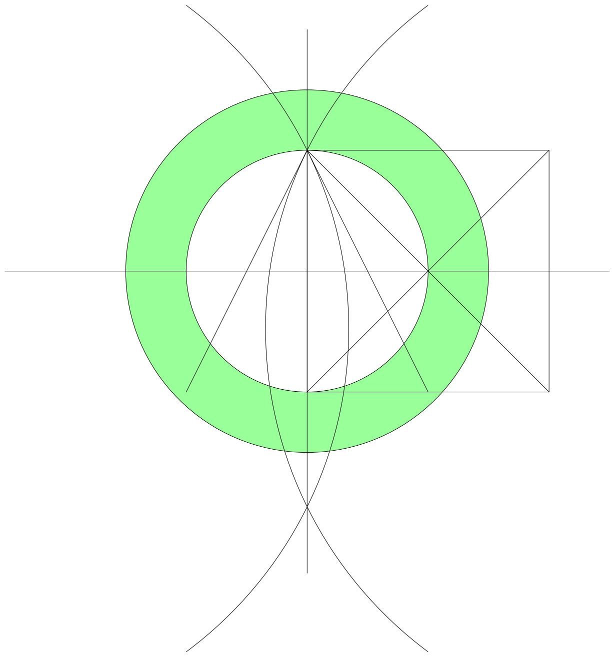

Figure 3. W, the domain of the comparison function for (7).

3. Lemmata

Lemma 1. *Let Ω be a bounded open set in $\mathbb{R}^n$, $n \geq 2$, with locally Lipschitz boundary and let Γ be an open C^2 subset of $\partial\Omega$. Let $\phi \in L^\infty(\partial\Omega) \cap C^{1,\beta}(\Gamma)$. Suppose $g \in C^2(\Omega) \cap L^\infty(\Omega)$ is the variational solution of* (1)–(2) *and $g < \phi$ on Γ. Then*

$$\nu \equiv \frac{(\nabla g, -1)}{\sqrt{1+|\nabla g|^2}} \in C^0(\Omega \cup \Gamma)$$

and $\nu \cdot \eta = 1$ on Γ, where $\eta(x) \in S^{n-1}$ is the exterior unit normal to Γ at x.

Proof. Since g minimizes the functional J in (3) over $\mathrm{BV}(\Omega)$, g also minimizes the functional $K(u) = J(u) - \int_\Gamma \phi \, d\mathcal{H}^{n-1}$. Notice

$$K(u) = \int_\Omega |Du| + \int_\Omega \int_0^u H(x,t)\,dt\,dx + \int_{\partial\Omega\setminus\Gamma} |u-\phi|\,d\mathcal{H}^{n-1} - \int_\Gamma u\,d\mathcal{H}^{n-1}$$

for each $u \in \mathrm{BV}(\Omega)$ with $\mathrm{tr}(u) \leq \phi$ on Γ; in particular, this holds when $u = g$. Therefore, for each $x \in \Gamma$, there exists $\rho > 0$ such that $\partial\Omega \cap B_n(x,\rho) \subset \Gamma$, and the lemma follows as in [Korevaar and Simon 1996]. □

Lemma 2. *Let Ω be a bounded open set in $\mathbb{R}^n$, $n \geq 2$, with locally Lipschitz boundary, $\phi, \psi \in L^\infty(\partial\Omega)$ with $\psi \leq \phi$ on $\partial\Omega$, $H_0 \in C^2(\Omega \times \mathbb{R})$ with $H_0(x,t)$*

nondecreasing in t for $x \in \Omega$, and $H_0 \geq H$ on $\Omega \times \mathbb{R}$. Consider the boundary value problem

$$\operatorname{div}(Tf) = H_0(x, f) \quad \text{in } \Omega, \tag{8}$$

$$f = \psi \quad \text{on } \partial\Omega. \tag{9}$$

Suppose $g \in C^2(\Omega) \cap L^\infty(\Omega)$ is the variational solution of (1)–(2) *and either*

(i) *$h \in C^2(\Omega) \cap L^\infty(\Omega)$ is the variational solution of* (8)–(9), *or*

(ii) *$\psi \in C^0(\partial\Omega)$, $h \in C^2(\Omega) \cap C^0(\overline{\Omega})$, and h satisfies* (8)–(9).

Then $h \leq g$ in Ω.

Proof. Let $A = \{x \in \Omega : h(x) > g(x)\}$. In case (i), let $f = hI_{\Omega\setminus A} + gI_A$, where I_B is the characteristic function of a set B; then a simple calculation using $J(g) \leq J(f)$ shows that $J_1(f) \leq J_1(h)$ and therefore $f = h$ and $A = \varnothing$, where

$$J_1(u) = \int_\Omega |Du| + \int_\Omega \int_0^u H_0(x, t)\, dt\, dx + \int_{\partial\Omega} |u - \psi|\, d\mathcal{H}^{n-1}, \quad u \in \mathrm{BV}(\Omega),$$

is the functional which h minimizes. In case (ii), the conclusion follows from Lemma 1 of [Williams 1978]. □

Lemma 3. *Let $\Omega \subset \{x \in \mathbb{R}^2 : x_2 > 0\}$ be a bounded open set, $n \in \mathbb{N}$ with $n \geq 2$, and $g \in C^2(\Omega)$. Set $\widetilde{\Omega} = \{(x_1, x_2\omega) \in \mathbb{R}^n : (x_1, x_2) \in \Omega, \omega \in S^{n-2}\}$ and define $\tilde{g} \in C^2(\widetilde{\Omega})$ by $\tilde{g}(x_1, x_2\omega) = g(x_1, x_2)$ for $(x_1, x_2) \in \Omega$, $\omega \in S^{n-2}$. Then, for*

$$x = (x_1, \dots, x_n) = (x_1, r\omega) \in \widetilde{\Omega}$$

with $r = \sqrt{x_2^2 + \cdots + x_n^2}$, $\omega = \frac{1}{r}(x_2, \dots, x_n)$, and $(x_1, r) \in \Omega$, we have

$$\operatorname{div}\left(\frac{\nabla\tilde{g}}{\sqrt{1+|\nabla\tilde{g}|^2}}\right)(x) = \operatorname{div}\left(\frac{\nabla g}{\sqrt{1+|\nabla g|^2}}\right)(x_1, r) + \frac{n-2}{r}\frac{g_{x_2}(x_1, r)}{\sqrt{1+|\nabla g(x_1, r)|^2}}.$$

In particular, if $H \geq 0$, $R > 0$, $\Omega \subset \{x \in \mathbb{R}^2 : x_2 \geq R\}$, and

$$\operatorname{div}\left(\frac{\nabla g}{\sqrt{1+|\nabla g|^2}}\right) \geq H + \frac{n-2}{R} \quad \text{on } \Omega,$$

then

$$\operatorname{div}\left(\frac{\nabla\tilde{g}}{\sqrt{1+|\nabla\tilde{g}|^2}}\right) \geq H \quad \text{on } \widetilde{\Omega}.$$

Proof. Notice that

$$1+|\nabla\tilde{g}|^2 = 1+|\nabla g|^2,$$

$$\bigl(1+|\nabla\tilde{g}|^2\bigr)\triangle\tilde{g} = \bigl(1+|\nabla g|^2\bigr)\Bigl(\triangle g+\frac{n-2}{r}g_{x_2}\Bigr),$$

$$\sum_{i,j=1}^{n}\frac{\partial\tilde{g}}{\partial x_i}\frac{\partial\tilde{g}}{\partial x_j}\frac{\partial^2\tilde{g}}{\partial x_i\partial x_j} = \Bigl(\frac{\partial g}{\partial x_1}\Bigr)^2\frac{\partial^2 g}{\partial x_1^2}+2\frac{\partial g}{\partial x_1}\frac{\partial g}{\partial x_2}\frac{\partial^2 g}{\partial x_1\partial x_2}+\Bigl(\frac{\partial g}{\partial x_2}\Bigr)^2\frac{\partial^2 g}{\partial x_2^2},$$

and so

$$\bigl(1+|\nabla\tilde{g}|^2\bigr)\triangle\tilde{g}-\sum_{i,j=1}^{n}\frac{\partial\tilde{g}}{\partial x_i}\frac{\partial\tilde{g}}{\partial x_j}\frac{\partial^2\tilde{g}}{\partial x_i\partial x_j}$$
$$=\bigl(1+g_{x_2}^2\bigr)g_{x_1x_1}-2g_{x_1}g_{x_2}g_{x_1x_2}+\bigl(1+g_{x_1}^2\bigr)g_{x_2x_2}+\frac{n-2}{r}\bigl(1+g_{x_1}^2+g_{x_2}^2\bigr)g_{x_2}.$$

The lemma follows from this. □

4. The *n*-dimensional case

Let $B_k(x,r)$ denote the open ball in $\mathbb{R}^k$ centered at $x\in\mathbb{R}^k$ with radius $r>0$ and $\mathbb{O}_k=(0,\dots,0)\in\mathbb{R}^k$, for $k\in\mathbb{N}$. Now consider $n\ge 2$ and set

$$\lambda = \sup_{(x,t)\in\mathbb{R}^n\times\mathbb{R}}|H(x,t)|;$$

if $\lambda=0$, replace it with a positive constant. For each $a\in\bigl(0,\frac{n}{\lambda}\bigr)$ and $Q\in\mathbb{R}^n$, we have

$$\int_{B_n(Q,a)}\lambda^n\,dx < n^n\omega_n. \tag{10}$$

By translating our problem in $\mathbb{R}^n$, we may (and will) assume $Q=\mathbb{O}_n$. By Proposition 1.1 and Theorem 2.1 of [Giusti 1976], we see that if Ω is a bounded, connected, and open set in $\mathbb{R}^n$ with Lipschitz-continuous boundary, $\overline{\Omega}\subset B_n\bigl(\mathbb{O}_n,\frac{n}{\lambda}\bigr)$, and $\phi\in L^1(\partial\Omega)$, then the functional J in (3) has a minimizer $f\in \mathrm{BV}(\Omega)$, $f\in C^2(\Omega)$ satisfying (1).

The proof in Section 4.1 consists of setting some parameters (e.g., p, r_1, r_2, m_0, b, c, τ, σ, a), determining the domain Ω, finding different comparison functions (e.g., g_1, $g^{[u]}$, $k_\pm$, k_2, k_3, k_4), and mimicking (6) and (7) to show that the variational solution f of (1)–(2) is discontinuous at a nonconvex corner. In particular, we use a torus (i.e., j_a) to obtain (21), unduloids (i.e., $k_\pm$, k_2) to obtain (24) (an analog of (7)), and nodoids (i.e., g_1, $g^{[u]}$), unduloids (i.e., $k_\pm$, k_4), and a helicoidal function (i.e., h_2) to obtain (30) (an analog of (6)) and prove that f is discontinuous at $P=(0,p,0,\dots,0)\in\mathbb{R}^n\in\partial\Omega$.

4.1. ***Codimension* 1 *singular set.*** In this section, we will obtain a domain Ω as above and $\phi \in C^{\infty}(\mathbb{R}^n)$ such that $P \in \partial\Omega$, the minimizer f of (3) is discontinuous at P, $\partial\Omega \setminus T$ is smooth (C^{∞}), and $f \in C^2(\Omega) \cap C^0(\overline{\Omega} \setminus T)$, where T is a smooth set of dimension $n-2$ (i.e., T has codimension 1 in $\partial\Omega$). We will use portions of nodoids, unduloids, and helicoidal surfaces with constant mean curvature as comparison functions. For the convenience of the reader, we will denote functions whose graphs are subsets of nodoids with the letter g (e.g., $g_1(x_1, x_2)$), subsets of CMC helicoids with the letter h, and subsets of unduloids (or onduloids) with the letter k.

Let $\mathcal{N}_1 \subset \mathbb{R}^3$ be a nodoid which is symmetric with respect to the x_3-axis and has mean curvature 1 (when $\mathcal{N}_1$ is oriented "inward", so that the unit normal $\vec{N}_{\mathcal{N}_1}$ to $\mathcal{N}_1$ points toward the x_3-axis at the points of $\mathcal{N}_1$ which are furthest from the x_3-axis). Let $s_1 = \inf_{(x,t)\in\mathcal{N}_1} |x|$ be the inner neck size of $\mathcal{N}_1$ and let s_3 satisfy the condition that the unit normal to $\mathcal{N}_1$ is vertical (i.e., parallel to the x_3-axis) at each point $(x, t) \in \mathbb{R}^2 \times \mathbb{R}$ of $\mathcal{N}_1$ at which $|x| = s_3$; then $s_1 < s_3$. Let $s_2 \in (s_1, s_3)$. (Notice that we can assume s_2/s_1 is close to s_3/s_1 if we wish.)

Let us fix $0 < p < \frac{1}{\lambda}$ and set $w = (0, p) \in \mathbb{R}^2$, $P = (0, p, 0, \dots, 0) \in \mathbb{R}^n$. Let $m_0 = \lambda/2 + (n-2)/(p/3)$. We shall assume $r_2 = s_2/m_0 < p/3$; if necessary, we may increase m_0 to accomplish this. Let $r_1 = s_1/m_0$ and $r_3 = s_3/m_0$. Let $\mathcal{N} = \{(m_0)^{-1}X \in \mathbb{R}^3 : X \in \mathcal{N}_1\}$; then $\mathcal{N}$ is a nodoid with mean curvature m_0. Set $\Delta_1 = \{x \in \mathbb{R}^2 : r_1 < |x| < r_2\}$. Fix $b \in \left(0, \frac{1}{4m_0}\left(1 + 2m_0 p - \sqrt{1 + 4m_0^2 p^2}\right)\right)$.

Define $g_1 \in C^{\infty}(\Delta_1) \cap C^0(\overline{\Delta_1})$ to be a function whose graph is a subset of $\mathcal{N}$ on which $\vec{N}_{\mathcal{N}} = (n_1, n_2, n_3)$ satisfies $n_3 \geq 0$; then

$$\operatorname{div}\left(\frac{\nabla g_1}{\sqrt{1+|\nabla g_1|^2}}\right) = m_0 \geq \lambda + \frac{2(n-2)}{p/3}. \tag{11}$$

By moving $\mathcal{N}$ vertically, we may assume $g_1(x) = 0$ when $|x| = r_2$; then $g_1 > 0$ in Δ_1. Notice that $\frac{\partial g_1}{\partial x_1}(r_1, 0) = -\infty$ and $\frac{\partial g_1}{\partial x_1}(r_2, 0) < 0$; then there exists a $\beta_0 > 0$ such that, for each $\theta \in \mathbb{R}$,

$$\frac{\partial}{\partial r}(g_1(r\Theta)) < -\beta_0 \quad \text{for } r_1 < r < r_2, \tag{12}$$

where $\Theta = (\cos(\theta), \sin(\theta))$. Fix $\beta \in (0, \beta_0)$. Let

$$0 < \tau < \min\left\{\frac{pr_1}{\sqrt{r_2^2 - r_1^2}}, \frac{2(1-p\lambda)}{\lambda(2-p\lambda)}, \frac{b(4p-b)}{4(2p-b)}\right\}. \tag{13}$$

Consider $\sigma \in \left(-\frac{\pi}{2}, 0\right)$. Notice that the distance between L and the point $(0, p-r_2)$ is $r_2\cos(\sigma)$, where L is the closed sector given by

$$L = \{(r\cos(\theta), p + r\sin(\theta)) : r \geq 0, \sigma \leq \theta \leq \pi - \sigma\}.$$

Define $r_4 = \sqrt{p^2+\tau^2}$ and

$$M = B_2((\tau, 0), r_4) \cap B_2((-\tau, 0), r_4).$$

Notice that

$$\tau < \frac{b(4p-b)}{4(2p-b)}$$

and therefore $B_2\big(\mathbb{O}_2, \frac{1}{2}(a+p)-b\big) \subset M$ if $p < a < p+b$.

Set $\sigma = -\arctan(\tau/p)$; then $\cos(\sigma) > r_1/r_2$, since

$$\tau < \frac{p\sqrt{r_2^2-r_1^2}}{r_1},$$

and $L \cap \overline{B_2} = \varnothing$, where $B_2 = B_2((0, p-r_2), r_1)$. Therefore there exists a $\delta_1 > 0$ such that if $u = (u_1, u_2) \in \partial B_2(\mathbb{O}_2, p)$ with $|u-w| < \delta_1$, then

$$B_2\Big(\frac{p-r_2}{p}u, r_1\Big) \subset M. \tag{14}$$

Since

$$\tau < \frac{2(1-p\lambda)}{\lambda(2-p\lambda)},$$

we have $\tau - \big(\frac{2}{\lambda}-r_4\big) < -p$ and so $B_2(\mathbb{O}_2, p) \subset B_2\big((\tau, 0), \frac{2}{\lambda}-r_4\big)$ (see Figure 8, right). Notice that

$$M \setminus \{(0, \pm p)\} = \{(r\cos(\theta), p + r\sin(\theta)) : 0 < r < 2p, \theta^-(r) < \theta < \theta^+(r)\} \tag{15}$$

for some $\theta^{\pm} \in C^0([0, \delta))$ satisfying $\theta^- < \theta^+$, $\theta^-(0) = -\pi-\sigma$, and $\theta^+(0) = \sigma$.

Let $a > p$ and set

$$\mathcal{T} = \Big\{\Big(\big(\tfrac{1}{2}(a+p)+b\cos v\big)\cos u,\ \big(\tfrac{1}{2}(a+p)+b\cos v\big)\sin u,\ b\sin v+c\Big) : (u, v) \in R\Big\},$$

where $R = [0, 2\pi]\times[-\pi, 0]$ and $0 < c < b$; since $b < \frac{1}{4m_0}\big(1+2m_0p-\sqrt{1+4m_0^2p^2}\big)$, we see that

$$\frac{\frac{1}{2}(a+p)-2b}{4b\big(\frac{1}{2}(a+p)-b\big)} > m_0$$

for all $a \geq p$. We shall assume

$$a \in \big(p, \min\{p+b, \tfrac{1}{\lambda}\}\big) \tag{16}$$

and $c = \sqrt{b^2-\big(\frac{1}{2}(a-p)\big)^2}$. Notice that $\mathcal{T}$ is the lower half of a torus whose mean curvature (i.e., one half of the trace of the shape operator) at each point is greater than m_0. Let $\mathcal{T}$ be the graph of a function j_a over

$$\Delta_a = \big\{x \in \mathbb{R}^2 : \tfrac{1}{2}(a+p)-b \leq |x| \leq \tfrac{1}{2}(a+p)+b\big\};$$

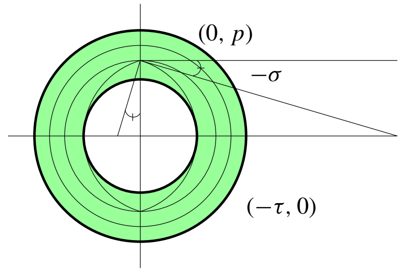

Figure 4. The domain of j_a.

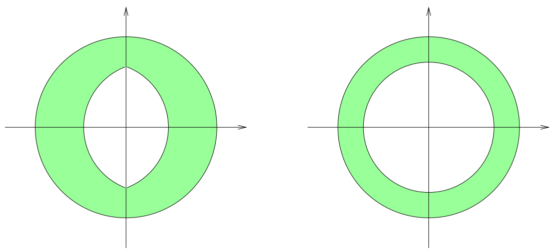

Figure 5. Left: $\Pi_{1,j}(\Omega)$ for $2 \le j \le n$. Right: $\Pi_{i,j}(\Omega)$ for $2 \le i < j \le n$.

then $j_a(x) = 0$ on $|x| = a$ and $|x| = p$, $j_a(x) < 0$ on $p < |x| < a$, and $j_a(x) > 0$ on $\frac{1}{2}(a+p) - b \le |x| < p$ and $a < |x| \le \frac{1}{2}(a+p) + b$ for $x \in \mathbb{R}^2$. Notice that $|j_a(x)| < \frac{1}{2m_0}$ for all $x \in \Delta_a$.

Set

$$\Omega = B_n(\mathbb{O}_n, a) \setminus \overline{\mathcal{M}}, \tag{17}$$

where $\mathcal{M} = \widetilde{M} = \{(x_1, x_2\omega) \in \mathbb{R}^n : (x_1, x_2) \in M, \omega \in S^{n-2}\}$. If we define

$$\Pi_{i,j}(A) = \{(x_i, x_j) : (x_1, \dots, x_n) \in A, x_k = 0 \text{ for } k \ne i, j\}$$

for $A \subset \mathbb{R}^n$ and $1 \le i < j \le n$, then $\Pi_{1,j}(\Omega) = B_2(\mathbb{O}_2, a) \setminus \overline{M}$ for $2 \le j \le n$ and $\Pi_{i,j}(\Omega) = B_2(\mathbb{O}_2, a) \setminus \overline{B_2(\mathbb{O}_2, 1)}$ for $2 \le i < j \le n$ (see Figure 5).

We wish to select a helicoidal surface in $\mathbb{R}^3$ (e.g., [do Carmo and Dajczer 1982]) with constant mean curvature m_0, axis $\{w\} \times \mathbb{R}$, and pitch $-\beta$ (recall $-\beta \in (-\beta_0, 0)$), which we will denote $\mathcal{S}$; then, for each $t \in \mathbb{R}$, $k_t(\mathcal{S}) = \mathcal{S}$, where $k_t : \mathbb{R}^3 \to \mathbb{R}^3$ is the helicoidal motion given by $k_t(x_1, x_2, x_3) = (l_t(x_1, x_2), x_3 - \beta t)$ with $l_t : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$l_t(x_1, x_2) = (x_1 \cos(t) + (x_2 - p)\sin(t),\ p - x_1 \sin(t) + (x_2 - p)\cos(t)).$$

Set $c_0 = \frac{1}{4}\beta\sigma < 0$. By vertically translating $\mathcal{S}$, we may assume that there is an open c_0-level curve $\mathcal{L}_0$ of $\mathcal{S}$ with endpoints $w = (0, p)$ and $b = (b_1, b_2)$ such that

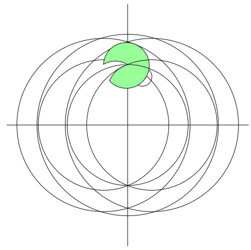

Figure 6. $\mathcal{R}$.

$\mathcal{L}_0 \subset (0, \infty) \times \mathbb{R}$, $\mathcal{L} = \overline{\mathcal{L}_0}$ is tangent to the (horizontal) line $\mathbb{R} \times \{p\}$ at w, and the slope m_v of the tangent line to $\mathcal{L}$ at v satisfies $|m_v| < \tan(-\sigma/5)$ for each $v \in \mathcal{L}_0$; then $\mathcal{L} \times \{c_0\} \subset \mathcal{S}$ and the curves $l_t(\mathcal{L}_0)$, $-\frac{7\pi}{8} < t < \frac{7\pi}{8}$, are mutually disjoint. Notice that the set

$$\mathcal{R} = \left\{ l_t(\mathcal{L}_0) : -\tfrac{7\pi}{8} < t < \tfrac{7\pi}{8} \right\} = \bigcup_{-\frac{7\pi}{8} < t < \frac{7\pi}{8}} l_t(\mathcal{L}_0)$$

is an open subset of $\mathbb{R}^2 \setminus ((-\infty, 0] \times \{p\})$ (see Figure 6), $w \in \overline{\mathcal{R}}$, and $\mathcal{S}$ implicitly defines the smooth function h_2 on $\mathcal{R}$ given by $h_2(x) = \frac{\beta}{4}(\sigma - 4t)$ if $x \in l_t(\mathcal{L}_0)$ for some $t \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Notice that $B_2(w, b_1) \cap \{x_1 > 0\} \subset \mathcal{R}$. Now we have $l_t(\mathcal{L}_0) \cap M = \varnothing$ for $t \in \left(\frac{3\sigma}{4}, \frac{\sigma}{4}\right)$ and, by making $b_1 > 0$ sufficiently small, we may assume that

$$l_t(\mathcal{L}_0) \subset B_2(\mathbb{O}_2, p) \setminus M \quad \text{for each } t \in \left(\tfrac{3\sigma}{4}, \tfrac{\sigma}{4}\right). \tag{18}$$

Notice that $h_2 < \beta(2\sigma^2 - \pi)/(8\sigma)$ on $l_t(\mathcal{L}_0)$ for $-\frac{\pi}{2} < t < \frac{7\pi}{8}$.

Let us fix $u = (u_1, u_2) \in \partial B_2(\mathbb{O}_2, p)$ such that $|u - w| < \min\{\delta_1, b_1\}$ and $u_1 > 0$. Then there exists $\theta_u \in \left(0, \frac{\pi}{2}\right)$ such that $u = (p\cos(\theta_u), p\sin(\theta_u))$. Define

$$g^{[u]}(x) = g_1\left(x + \frac{r_2 - p}{p} u\right)$$

and notice that $g^{[u]}(u) = g_1\left(\frac{r_2}{p}u\right) = 0$, since $\left|\frac{r_2}{p}u\right| = r_2$. Note that the domain

$$\mathcal{D}^{[u]} = \left\{ x + \frac{p - r_2}{p} u : x \in \Delta_1 \right\} = B_2\left(\frac{p - r_2}{p} u, r_2\right) \setminus \overline{B_2\left(\frac{p - r_2}{p} u, r_1\right)}$$

of $g^{[u]}$ is contained in $B_2(\mathbb{O}_2, p)$ since $\partial B_2\left(\frac{p-r_2}{p}u, r_2\right)$ and $\partial B_2(\mathbb{O}_2, p)$ are tangent circles at u and $r_2 < p$ (see Figure 7). Notice that

$$h_2(r\cos(\theta_u), r\sin(\theta_u)) < g^{[u]}(r\cos(\theta_u), r\sin(\theta_u)) \tag{19}$$

when $p - r_2 + r_1 \le r \le p$, because $h_2(u) < 0 = g^{[u]}(u)$, $\beta < \beta_0$, and (12) holds.

Let

$$\mathcal{N}_\pm \subset \{x \in \mathbb{R}^2 : r_4 \le |(x_1 \pm \tau, x_2)| \le \tfrac{2}{\lambda} - r_4\} \times \mathbb{R}$$

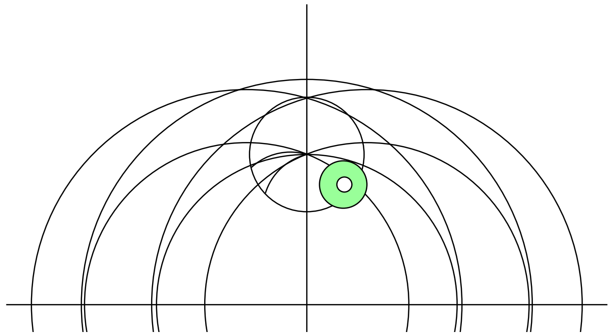

Figure 7. $\mathcal{D}^{[u]}$; $\Omega\cap\widetilde{\mathcal{D}}^{[u]}$ is the domain of the comparison function for (28).

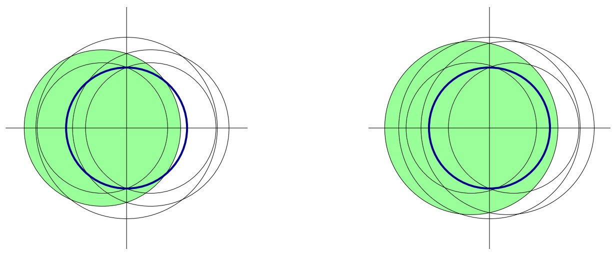

Figure 8. Left: $B_2(\mathbb{O}_2, p) \not\subset B_2\big((-\tau, 0), \frac{2}{\lambda} - r_4\big)$. Right: $B_2(\mathbb{O}_2, p) \subset B_2\big((-\tau, 0), \frac{2}{\lambda} - r_4\big)$.

be unduloids in $\mathbb{R}^3$ with mean curvature $\frac{\lambda}{2}$ such that $\{(\mp\tau, 0)\}\times\mathbb{R}$ are the respective axes of symmetry; the minimum and maximum radii (or "neck" and "waist" sizes) of both unduloids are r_4 and $\frac{2}{\lambda}-r_4$, respectively. Set

$$\Delta_\pm = B_2\big((\mp\tau, 0), \tfrac{2}{\lambda} - r_4\big) \setminus \overline{B_2((\mp\tau, 0), r_4)}$$

and define $k_\pm \in C^\infty(\Delta_\pm)$ so that the graphs of $k_\pm$ are subsets of $\mathcal{N}_\pm$, respectively,

$$\operatorname{div}(Tk_\pm) = -\lambda \quad \text{in } \Delta_\pm,$$

$\frac{\partial}{\partial r}\big(k_\pm((\mp p, 0)+r\Theta)\big)\big|_{r=r_4} = -\infty$ and $\frac{\partial}{\partial r}\big(k_\pm((\mp p, 0)+r\Theta)\big)\big|_{r=2/\lambda-r_4} = -\infty$ for each $\theta\in\mathbb{R}$, where $\Theta = (\cos(\theta), \sin(\theta))$. We may vertically translate $\mathcal{N}_\pm$ so that $k_\pm(x)=0$ for $x\in\mathbb{R}^2$ with $|(x_1\pm\tau, x_2)| = \frac{2}{\lambda}-r_4$. Notice that $k_+(0, p) = k_-(0, p) = \sup_{\Delta_+} k_+ = \sup_{\Delta_-} k_-$.

Let $\mathcal{N}\subset\big\{x\in\mathbb{R}^2 : p\le |x| \le \frac{2}{\lambda}-p\big\}\times\mathbb{R}$ be an unduloid with mean curvature $\frac{\lambda}{2}$ such that the x_3-axis is the axis of symmetry and the minimum and maximum radii (or "neck" and "waist" sizes) are p and $\frac{2}{\lambda}-p$, respectively. Set

$$\Delta_2 = B_2\big(\mathbb{O}_2, \tfrac{2}{\lambda} - p\big) \setminus \overline{B_2(\mathbb{O}_2, p)}$$

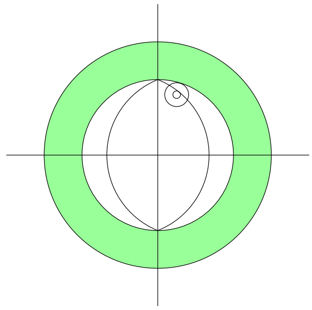

Figure 9. $B_2(\mathbb{O}_2, a) \setminus \overline{B_2(\mathbb{O}_2, p)}$: (22).

and define $k_2 \in C^\infty(\Delta_2)$ so that the graph of k_2 is a subset of $\mathcal{N}$, $\operatorname{div}(Tk_2) = -\lambda$ in Δ_2, and $\frac{\partial}{\partial r}\big(k_2(r\Theta)\big)\big|_{r=p} = \frac{\partial}{\partial r}\big(k_2(r\Theta)\big)\big|_{r=2/\lambda-p} = -\infty$ for each $\theta \in \mathbb{R}$, where $\Theta = (\cos(\theta), \sin(\theta))$.

Define $\phi \in C^\infty(\mathbb{R}^n)$ so that $\phi = 0$ on $\partial B_n(\mathbb{O}_n, a)$ and $\phi = m$ on $\partial \mathcal{M}$, where

$$m > \max\Big\{g_1(0, r_1), \tfrac{1}{2m_0}, k_+(0, r_4 - \tau) + k_2(0, p) - k_2\big(0, \tfrac{2}{\lambda} - p\big)\Big\}; \tag{20}$$

recall then that $m > j_a\big(\frac{1}{2}(a+p) - b\big)$. Let f be the variational solution of (1)–(2) with Ω and ϕ as given here; that is, let f minimize the functional given in (3) and notice that the existence of f follows from (10), (16), §1.D of [Giusti 1976], and [Gerhardt 1974; Giusti 1978]. (Notice that there exists $w : B_2(\mathbb{O}_2, a) \setminus M \to \mathbb{R}$ such that $f = \tilde{w}$.) The comparison principle implies $j_a(x) \le f(x)$ for $x \in \Omega$, and so $f(x) \ge j_a(x) \ge 0$ if $x \in \Omega$ with $|x| \le p$ (recall (16) holds). In particular,

$$f(x) \ge 0 \quad \text{when } x \in \Omega \text{ with } |x| \le p. \tag{21}$$

Set $W = (B_2(\mathbb{O}_2, a) \setminus \overline{B_2(\mathbb{O}_2, p)}) \times \mathbb{R}^{n-2}$. Now

$$\Omega \subset B_2(\mathbb{O}_2, a) \times \mathbb{R}^{n-2} \subset B_2\big(\mathbb{O}_2, \tfrac{2}{\lambda} - p\big) \times \mathbb{R}^{n-2}$$

(see Figure 9). Define $k_3(x) = k_2(x_1, x_2) - k_2(0, a)$ for $x = (x_1, x_2, \dots, x_n) \in W$. Notice that $f = 0 \le k_3$ on $\overline{W} \cap \partial B_n(\mathbb{O}_n, a)$,

$$\operatorname{div}(Tf) = H(x, f(x)) \ge -\lambda = \operatorname{div}(Tk_3) \quad \text{in } \Omega \cap W,$$

and $\frac{\partial}{\partial r}(k_2(r\Theta))|_{r=p} = -\infty$ (so that $\lim_{W \ni y \to x} Tk_3(y) \cdot \xi(x) = 1$, for ξ the unit exterior normal to ∂W and $x \in \partial B_2(\mathbb{O}_2, p) \times \mathbb{R}^{n-2}$). The general comparison principle (e.g., [Finn 1986, Theorem 5.1]) then implies

$$f \le k_3 \quad \text{in } \Omega \cap W \tag{22}$$

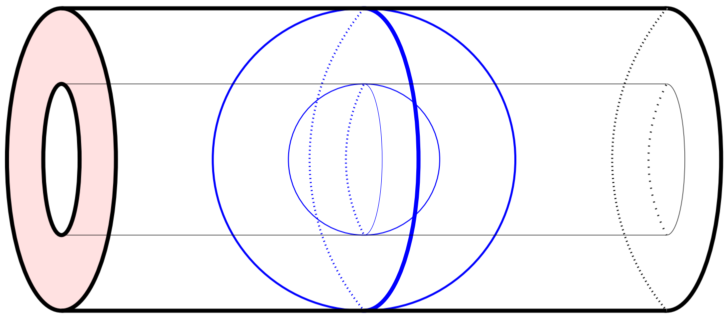

Figure 10. (23): W and $B_n(\mathbb{O}_n, a) \setminus \overline{B_n(\mathbb{O}_n, p)}$ when $n = 3$.

and, in particular,

$$\limsup_{\Omega \cap W \ni y \to x} f(y) \le k_3(x) \quad \text{for } x \in \partial\Omega \cap \overline{W} \tag{23}$$

(see Figure 10). By rotating the axis of symmetry of W through all lines in $\mathbb{R}^n$ containing $\mathbb{O}_n$ (or, equivalently, keeping W fixed and rotating Ω about $\mathbb{O}_n$), we see that

$$\sup\{f(x) : x \in B_n(\mathbb{O}_n, a) \setminus \overline{B_n(\mathbb{O}_n, p)}\} \le k_2(0, p) - k_2(0, a). \tag{24}$$

Now define $k_4 \in C^\infty(\Delta_+ \times \mathbb{R}^{n-2}) \cap C^0(\overline{\Delta_+} \times \mathbb{R}^{n-2})$ by

$$k_4(x) = k_+(x_1, x_2) + k_2(0, p) - k_2(0, a), \qquad x = (x_1, x_2, \dots, x_n) \in \overline{\Delta_+} \times \mathbb{R}^{n-2}.$$

Combining (1) and (24) with the facts that $\operatorname{div}(Tk_4) = -\lambda$ in $\Delta_+ \times \mathbb{R}^{n-2}$ and $\lim_{\Delta_+ \times \mathbb{R}^{n-2} \ni y \to x} Tk_4(y) \cdot \xi_+(x) = 1$ for $x \in \partial B_2((-\tau, 0), r_4) \times \mathbb{R}^{n-2}$, where ξ_+ is the inward unit normal to $\partial B_2((-\tau, 0), r_4) \times \mathbb{R}^{n-2}$, we see that

$$f \le k_4 \quad \text{in } \Omega \cap (\Delta_+ \times \mathbb{R}^{n-2}). \tag{25}$$

(If Figure 8 (left) held, then (25) would not be valid.) Now let $L : \mathbb{R}^n \to \mathbb{R}^n$ be any rotation about $\mathbb{O}_n$ which satisfies $L(\Omega) = \Omega$, notice that $f \circ L$ satisfies (1)–(2), and apply the previous argument to obtain $f \circ L \le k_4$ in $\Omega \cap (\Delta_+ \times \mathbb{R}^{n-2})$ and therefore

$$\sup\{f(x) : x \in \partial\mathcal{M}\} \le k_4(p, 0) < m. \tag{26}$$

From Lemma 1, we see that the downward unit normal N_f to the graph of f satisfies $N_f = (\nu, 0)$ on $\partial\mathcal{M} \setminus \{(0, p\omega) : \omega \in S^{n-2}\}$ and

$$\lim_{\Omega \ni y \to x} Tf(y) \cdot \nu(x) = 1 \quad \text{for } x \in \partial\mathcal{M} \setminus \{(0, p\omega) : \omega \in S^{n-2}\}. \tag{27}$$

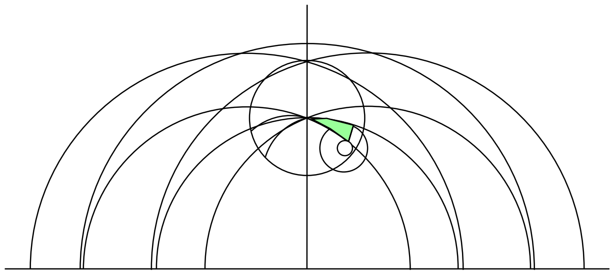

Figure 11. A: (29).

Let us write $B = B_2\left(\frac{p-r_2}{p}u, r_2\right)$; then $\tilde{g}^{[u]} = 0 \le f$ on $\Omega \cap \partial\widetilde{B}$ and $\tilde{g}^{[u]} \le g_1(r_1, 0) < \phi$ on $\widetilde{B} \cap \partial M$. It follows from (1), (11), and Lemma 3 that

$$\tilde{g}^{[u]} < f \quad \text{on } \Omega \cap \widetilde{\mathfrak{D}}^{[u]} = \Omega \cap \widetilde{B}. \tag{28}$$

Set $U = \{r(\cos(\theta), \sin(\theta)\omega) \in \Omega : r \in (0, p), \theta \in (0, \theta_u), \omega \in S^{n-2}\}$. If we write

$$\partial_1 U = \{(p\cos(\theta), p\sin(\theta)\omega) : \theta \in (0, \theta_u], \omega \in S^{n-2}\},$$
$$\partial_2 U = \partial\mathcal{M} \cap \partial U,$$
$$\partial_3 U = \{(r\cos(\theta_u), r\sin(\theta_u)\omega) \in \overline{\Omega} : r \in [0, p], \omega \in S^{n-2}\},$$

then $\partial U = \partial_1 U \cup \partial_2 U \cup \partial_3 U$, $\tilde{h}_2 \le 0 \le f$ on $\partial_1 U \setminus \{P\}$, and $\tilde{h}_2 < \tilde{g}^{[u]} < f$ on $\partial_3 U$ (see (19)); then (27) and the general comparison principle imply

$$\tilde{h}_2 < f \quad \text{in } U = \widetilde{A}, \tag{29}$$

where $A = \{r(\cos(\theta), \sin(\theta)) \in B_2(\mathbb{O}_2, p) \setminus \overline{M} : r \in (0, p), \theta \in (0, \theta_u)\}$ (see Figure 11). Set $\mathcal{R}_2 = \bigcup_{t=3\sigma/4}^{2\sigma/4} l_t(\mathcal{L}_0)$. Now (18) implies $\widetilde{\mathcal{R}}_2 \subset U$ and so

$$f > \tilde{h}_2 \ge -\frac{\beta\sigma}{4} \quad \text{on } \mathcal{R}_2. \tag{30}$$

Using (24) and (30), we see that if $a \in \left(p, \frac{2}{\lambda} - p\right)$ is close enough to p, then $k_2(0, p) - k_2(0, a) < -\frac{\beta\sigma}{4}$ and therefore f cannot be continuous at P or at any point of $T = \{(0, p\omega) \in \mathbb{R}^n : \omega \in S^{n-2}\}$. Note that $f \in C^0(\overline{\Omega} \setminus T)$ (e.g., [Lin 1987]).

4.2. ***One singular point.*** In this section, we obtain a domain Ω and $\phi \in C^\infty(\mathbb{R}^n)$ such that $P \in \partial\Omega$, the minimizer f of (3) is discontinuous at P, $\partial\Omega \setminus \{P\}$ is smooth (C^∞), and $f \in C^0(\overline{\Omega} \setminus \{P\})$. This is accomplished by replacing $\mathcal{M}$ by a convex set $\mathcal{G}$ such that $\partial\mathcal{G} \setminus \{P\}$ is smooth (C^∞) and $\mathcal{G} \subset B_n(\mathbb{O}_n, p)$. We shall use the notation of Section 4.1 throughout this section. We assume $p \in \left(0, \frac{1}{\lambda}\right)$ and set $P = (0, p, 0, \ldots, 0)$. (We will no longer require Figure 8 (right) to hold.)

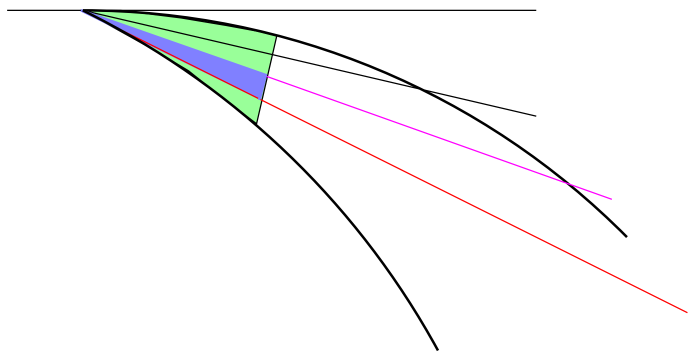

Figure 12. An illustration of $\mathcal{R}_2$ (blue region) and A (green and blue regions).

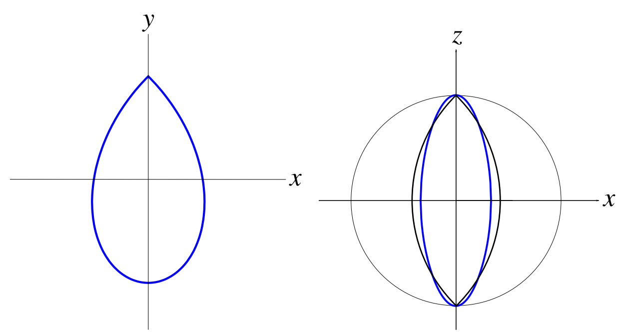

Figure 13. Left: $X\big(\theta, \frac{\pi}{2}, 1\big)$. Right: $X\big(\theta, \frac{1}{2}\arccos(1-\sec(\theta)\sec(2\theta)), 1\big)$.

Let $\alpha > 1$, $n \geq 3$, and $Y : \big[-\frac{\pi}{2\alpha}, \frac{\pi}{2\alpha}\big] \times [0, \pi] \times S^{n-3} \to \mathbb{R}^n$ be defined by

$$Y(\theta, \phi, \omega) = 2\cos(\alpha\theta)\sin(\phi)\big(\cos(\theta)\sin(\phi), \sin(\theta)\sin(\phi), \cos(\phi)\omega\big).$$

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be given by

$$F(x_1, \dots, x_n) = \big(\tfrac{x_2}{p}, \tfrac{1-x_1}{p}, \tfrac{x_3}{p}, \dots, \tfrac{x_n}{p}\big)$$

and define $X(\theta, \phi, \omega) = F(Y(\theta, \phi, \omega))$ for $-\frac{\pi}{2\alpha} \leq \theta \leq \frac{\pi}{2\alpha}$, $0 \leq \phi \leq \pi$, $\omega \in S^{n-3}$ (see Figures 13 and Figure 14 with $n = 3$, $\alpha = 2$; the axes are labeled x, y, z for x_1, x_2, x_3, respectively). Let $\mathcal{G}$ be the open, convex set whose boundary is the image of X; that is,

$$\partial\mathcal{G} = \big\{X(\theta, \phi, \omega) : -\tfrac{\pi}{2\alpha} \leq \theta \leq \tfrac{\pi}{2\alpha}, 0 \leq \phi \leq \pi, \omega \in S^{n-3}\big\}.$$

Notice that $\partial\mathcal{G} \setminus \{P\}$ is a C^∞ hypersurface in $\mathbb{R}^n$ and $\partial\mathcal{G} \subset \overline{B_n(\mathcal{O}_n, p)}$.

Let τ satisfy

$$0 < \tau < \min\left\{\frac{pr_1}{\sqrt{r_2^2-r_1^2}}, \frac{b(4p-b)}{4(2p-b)}\right\}.$$

Set $\sigma = -\arctan(\tau/p)$ and $\alpha = \pi/(\pi+2\sigma)$. Then the tangent cones to $\partial\mathcal{G}$ and $\partial\mathcal{M}$ at P are identical, $\cos(\sigma) > r_1/r_2$, and (14) holds for $u = (u_1, u_2) \in \partial B_2(\mathbb{O}_2, p)$ with $|u - w| < \delta_1$. By making $\tau > 0$ smaller if necessary, we may assume $B_n\big(\mathbb{O}_n, \frac{1}{2}(a+p) - b\big) \subset \mathcal{G}$ if $p < a < p+b$.

Now pick $a \in \big(p, \min\{p+b, \frac{1}{\lambda}\}\big)$ such that $k_2(0, p) - k_2(0, a) < -\frac{\beta\sigma}{4}$, as in (30), and define

$$\Omega = B_n(\mathbb{O}_n, a) \setminus \overline{\mathcal{G}}. \tag{31}$$

Let

$$m > \max\left\{g_1(0, r_1), \frac{1}{2m_0}, \frac{\beta(2\sigma^2-\pi)}{8\sigma}\right\}$$

and define $\phi \in C^\infty(\mathbb{R}^n)$ so that $\phi = 0$ on $\partial B_n(\mathbb{O}_n, a)$ and $\phi = m$ on $\partial\mathcal{G}$, and let f be the variational solution of (1)–(2). Notice that $f \in C^2(\Omega)$ satisfies (1) and $f \in C^0(\overline{\Omega} \setminus \{P\})$ (e.g., [Lin 1987]).

As in (28), let $B = B_2\big(\frac{p-r_2}{p} u, r_2\big)$. Set $U_0 = \{x \in \Omega : x \in \widetilde{B}, x_1 > 0\}$ and $U = \{r(\cos(\theta), \sin(\theta)\omega) \in \Omega : r \in (0, p), \theta \in (0, \theta_u), \omega \in S^{n-2}\}$. Now $\tilde{g}^{[u]} = 0$ on $\partial U_0 \cap \partial\widetilde{B}$ and $\tilde{g}^{[u]} \le g_1(0, r_1) < m$ on $\partial U_0 \cap \partial\mathcal{G}$ and so Lemma 2, Lemma 3, and (1) imply $\tilde{g}^{[u]} \le f$ in U_0 since f minimizes the functional in (3).

As before, set

$$\begin{aligned}
\partial_1 U &= \big\{\big(p\cos(\theta), p\sin(\theta)\omega\big) : \theta \in [0, \theta_u], \omega \in S^{n-2}\big\},\\
\partial_2 U &= \partial\mathcal{G} \cap \partial U,\\
\partial_3 U &= \big\{\big(r\cos(\theta_u), r\sin(\theta_u)\omega\big) \in \overline{\Omega} : r \in [0, p], \omega \in S^{n-2}\big\}.
\end{aligned}$$

Then $f \ge 0$ on $\partial_1 U \setminus \{P\}$, $\partial U = \partial_1 U \cup \partial_2 U \cup \partial_3 U$, $\tilde{h}_2 \le 0 \le f$ on $\partial_1 U$, $\tilde{h}_2 < m = \phi$ on $\partial_2 U$, and $\tilde{h}_2 < \tilde{g}^{[u]} < f$ on $\partial_3 U$; Lemma 2 implies that (30) continues to hold. Then (24) and (30) imply f is discontinuous at P since $k_2(0, p) - k_2(0, a) < -\frac{\beta\sigma}{4}$.

5. The Concus–Finn conjecture

For the moment, assume $n = 2$. Around 1970, Paul Concus and Robert Finn conjectured that if $\kappa \ge 0$, $\Omega \subset \mathbb{R}^2$ has a corner at $P \in \partial\Omega$ of (angular) size 2α, $\alpha \in \big(0, \frac{\pi}{2}\big)$, $\gamma : \partial\Omega \setminus \{P\} \to [0, \pi]$, and $\big|\frac{\pi}{2} - \gamma_0\big| > \alpha$, where

$$\lim_{\partial\Omega \ni x \to P} \gamma(x) = \gamma_0, \tag{32}$$

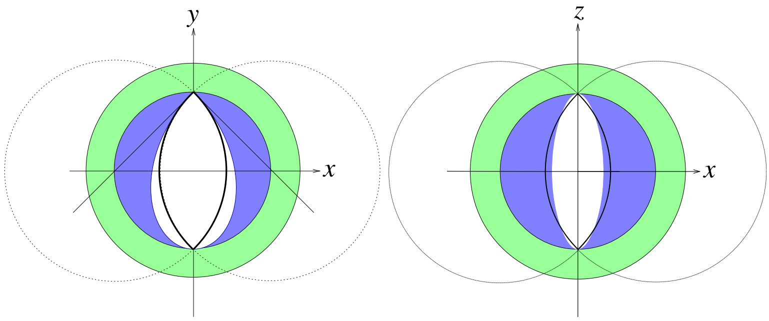

Figure 14. Left: $\Pi_{1,2}(\Omega)$. Right: $\Pi_{1,3}(\Omega)$.

then a function $f \in C^2(\Omega) \cap C^1(\bar{\Omega} \setminus \{P\})$ which satisfies

$$\operatorname{div}(Tf) = \kappa f \qquad \text{in } \Omega, \tag{33}$$

$$Tf \cdot \eta = \cos(\gamma) \quad \text{on } \partial\Omega \setminus \{P\} \tag{34}$$

must be discontinuous at P; here $\eta(x)$ is the exterior unit normal to Ω at $x \in \partial\Omega \setminus \{P\}$.

In the situation above with $\alpha \in \left(\frac{\pi}{2}, \pi\right)$, the "nonconvex Concus–Finn conjecture" states that if $\left|\frac{\pi}{2} - \gamma_0\right| > \pi - \alpha$, then the capillary surface f with contact angle γ must be discontinuous at P. A generalization (including the replacement of (33) by (1)) of this extension of the Concus–Finn conjecture in the case $\gamma_0 \in (0, \pi)$ was proven in [Lancaster 2012]. Both [Lancaster 2010] and [Lancaster 2012] include the possibility of differing limiting contact angles; that is, the limits

$$\lim_{\partial^+\Omega \ni x \to P} \gamma(x) = \gamma_1 \quad \text{and} \quad \lim_{\partial^-\Omega \ni x \to P} \gamma(x) = \gamma_2$$

exist, $\gamma_1, \gamma_2 \in (0, \pi)$, and $\gamma_1 \neq \gamma_2$. Here $\partial^+\Omega$ and $\partial^-\Omega$ are the two components of $\partial\Omega \setminus \{P, Q\}$, where $Q \in \partial\Omega \setminus \{P\}$. When $\gamma_1 \neq \gamma_2$, the necessary and sufficient (when $\alpha \leq \frac{\pi}{2}$) or necessary (when $\alpha > \frac{\pi}{2}$) conditions for the continuity of f at P become slightly more complicated.

The cases where $\gamma_0 = 0$, $\gamma_0 = \pi$, $\min\{\gamma_1, \gamma_2\} = 0$, and $\max\{\gamma_1, \gamma_2\} = \pi$ remain unresolved. If we suppose for a moment that the nonconvex Concus–Finn conjecture with limiting contact angles of 0 or π is proven, then the discontinuity of f at P in Section 2 follows immediately from the fact that $f < \phi$ in a neighborhood in $\partial\Omega \setminus \{P\}$ of P, since then Lemma 1 implies $\gamma_0 = 0$ and therefore $\left|\frac{\pi}{2} - \gamma_0\right| > \pi - \alpha$. In this situation (i.e., the solution f of a Dirichlet problem satisfies a 0 (or π) contact angle boundary condition near P), establishing the discontinuity of f at P would be much easier and a much larger class of domains Ω with a nonconvex corner (i.e., $\alpha > \frac{\pi}{2}$) at P would have this property. For example, if Ω is a bounded locally Lipschitz domain in $\mathbb{R}^2$ for which (4) holds, $f \in C^2(\Omega)$ is a generalized solution of (1)–(2) (and H need not vanish), and ϕ is large enough near P (depending on H and the maximum of ϕ outside some neighborhood of P) that $f < \phi$ on $\partial\Omega \setminus \{P\}$ near P, then the fact that $\gamma_0 = 0$ (Lemma 1) together with the nonconvex Concus–Finn conjecture would imply that f is discontinuous at P.

Now consider $n \in \mathbb{N}$ with $n \geq 3$. Formulating generalizations of the Concus–Finn conjecture in the "convex corner case" (i.e., $\Omega \cap B_n(P, r) \subset \{X \in \mathbb{R}^n : (X-P)\cdot\mu > 0\}$ for some $\mu \in S^{n-1}$, $P \in \partial\Omega$ and $r > 0$) and in other cases where $\partial\Omega$ is not smooth at a point $P \in \partial\Omega$ may be complicated because the geometry of $\partial\Omega \setminus \{P\}$ is much more interesting when $n > 2$. Establishing the validity of a generalization of the Concus–Finn conjecture for solutions of (1) and (34) when $n > 2$ is probably significantly harder than doing so when $n = 2$.

Suppose we knew that a solution f of (1) and (34) is necessarily discontinuous at a "nonconvex corner" $P \in \partial\Omega$ when $\gamma_0 = 0$, where γ_0 is given by (32). In this case, a necessary condition for the continuity of f at P would be that

$$\limsup_{\partial\Omega \ni X \to P} Tf(X)\cdot\eta(X) > 0,$$

$$\liminf_{\partial\Omega \ni X \to P} Tf(X)\cdot\eta(X) < \pi.$$

Then the arguments in Section 4 could be made more easily and the conclusion that f is discontinuous at P would hold in a much larger class of domains Ω; here, of course, we use the ridge point P in Section 4 as an example of a "nonconvex corner" of a domain in $\mathbb{R}^n$. The primary difficulty in proving in Section 4 that f is discontinuous at P is establishing (30); a more "natural" generalization of $\Omega \subset \mathbb{R}^2$ in Section 2 would be

$$\Omega^* = \{(x\omega_1, y, \omega_2, \ldots, \omega_{n-1}) \in \mathbb{R}^n : (x, y) \in B_2(\mathbb{O}_2, a) \setminus \overline{M}, \omega \in S^{n-1}\}.$$

However, the use of Lemma 3 to help establish (30) in Ω^* is highly problematic. On the other hand, an n-dimensional "Concus–Finn theorem" for a nonconvex conical point (e.g., $P \in \partial\Omega^*$) would only require an inequality like (26) to prove that $f < \phi$ on $\partial\Omega \setminus \{P\}$ near P, and hence that f is discontinuous at P; the replacement of (17) by (31) in order to obtain Ω such that $\partial\Omega \setminus \{P\}$ is C^∞ would be unnecessary.

References

[Bourni 2011] T. Bourni, "$C^{1,\alpha}$ theory for the prescribed mean curvature equation with Dirichlet data", *J. Geom. Anal.* **21**:4 (2011), 982–1035. MR 2836589 Zbl 1232.53009

[do Carmo and Dajczer 1982] M. P. do Carmo and M. Dajczer, "Helicoidal surfaces with constant mean curvature", *Tôhoku Math. J.* (2) **34**:3 (1982), 425–435. MR 84f:53003 Zbl 0501.53003

[Elcrat and Lancaster 1986] A. R. Elcrat and K. E. Lancaster, "Boundary behavior of a nonparametric surface of prescribed mean curvature near a reentrant corner", *Trans. Amer. Math. Soc.* **297**:2 (1986), 645–650. MR 87h:35098 Zbl 0602.35042

[Finn 1986] R. Finn, *Equilibrium capillary surfaces*, Grundlehren der Mathematischen Wissenschaften **284**, Springer, New York, 1986. MR 88f:49001 Zbl 0583.35002

[Gerhardt 1974] C. Gerhardt, "Existence, regularity, and boundary behavior of generalized surfaces of prescribed mean curvature", *Math. Z.* **139** (1974), 173–198. MR 55 #10846 Zbl 0316.49005

[Giusti 1976] E. Giusti, "Boundary value problems for non-parametric surfaces of prescribed mean curvature", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **3**:3 (1976), 501–548. MR 58 #2572 Zbl 0344.35036

[Giusti 1978] E. Giusti, "On the equation of surfaces of prescribed mean curvature: Existence and uniqueness without boundary conditions", *Invent. Math.* **46**:2 (1978), 111–137. MR 58 #7337 Zbl 0381.35035

[Huff and McCuan 2006] R. Huff and J. McCuan, "Scherk-type capillary graphs", *J. Math. Fluid Mech.* **8**:1 (2006), 99–119. MR 2006k:76022 Zbl 1125.76012

[Huff and McCuan 2009] R. Huff and J. McCuan, "Minimal graphs with discontinuous boundary values", *J. Aust. Math. Soc.* **86**:1 (2009), 75–95. MR 2010k:35203 Zbl 1161.53010

[Jenkins and Serrin 1968] H. Jenkins and J. Serrin, "The Dirichlet problem for the minimal surface equation in higher dimensions", *J. Reine Angew. Math.* **229** (1968), 170–187. MR 36 #5519 Zbl 0159.40204

[Korevaar 1980] N. J. Korevaar, "On the behavior of a capillary surface at a re-entrant corner", *Pacific J. Math.* **88**:2 (1980), 379–385. MR 82c:49033 Zbl 0475.49034

[Korevaar and Simon 1996] N. Korevaar and L. Simon, "Equations of mean curvature type with contact angle boundary conditions", pp. 175–201 in *Geometric analysis and the calculus of variations*, edited by J. Jost, Int. Press, Cambridge, MA, 1996. MR 98e:35071 Zbl 0932.35091

[Lancaster 1985] K. E. Lancaster, "Boundary behavior of a nonparametric minimal surface in $\mathbf{R}^3$ at a nonconvex point", *Analysis* **5**:1-2 (1985), 61–69. MR 86m:49053 Zbl 0601.35035

[Lancaster 1988] K. E. Lancaster, "Nonparametric minimal surfaces in $\mathbf{R}^3$ whose boundaries have a jump discontinuity", *Internat. J. Math. Math. Sci.* **11**:4 (1988), 651–656. MR 89m:35072 Zbl 0696.35057

[Lancaster 2010] K. E. Lancaster, "A proof of the Concus–Finn conjecture", *Pacific J. Math.* **247**:1 (2010), 75–108. MR 2011m:53011 Zbl 1256.53009

[Lancaster 2012] K. E. Lancaster, "Remarks on the behavior of nonparametric capillary surfaces at corners", *Pacific J. Math.* **258**:2 (2012), 369–392. MR 2981959 Zbl 1253.35069

[Lancaster and Siegel 1996] K. E. Lancaster and D. Siegel, "Existence and behavior of the radial limits of a bounded capillary surface at a corner", *Pacific J. Math.* **176**:1 (1996), 165–194. MR 98g:58030a Zbl 0866.76018

[Lau and Lin 1985] C. P. Lau and F.-H. Lin, "The best Hölder exponent for solutions of the non-parametric least area problem", *Indiana Univ. Math. J.* **34**:4 (1985), 809–813. MR 87d:49058 Zbl 0581.49025

[Lin 1987] F.-H. Lin, "Behaviour of nonparametric solutions and free boundary regularity", pp. 96–116 in *Miniconference on geometry and partial differential equations, 2* (Canberra, 1986), edited by J. E. Hutchinson and L. M. Simon, Proc. Centre Math. Anal. Austral. Nat. Univ. **12**, Austral. Nat. Univ., Canberra, 1987. MR 89c:35047 Zbl 0647.35032

[Melin 2013] J. Melin, *Examples of discontinuity for the variational solution of the minimal surface equation with Dirichlet data on a domain with a nonconvex corner and locally negative mean curvature*, MS Thesis, Wichita State University, 2013, Available at http://hdl.handle.net/10057/10639.

[Nitsche 1989] J. C. C. Nitsche, *Lectures on minimal surfaces, I: Introduction, fundamentals, geometry and basic boundary value problems*, Cambridge Univ. Press, 1989. MR 90m:49031 Zbl 0688.53001

[Serrin 1969] J. Serrin, "The problem of Dirichlet for quasilinear elliptic differential equations with many independent variables", *Philos. Trans. Roy. Soc. London Ser. A* **264** (1969), 413–496. MR 43 #7772 Zbl 0181.38003

[Shi and Finn 2004] D. Shi and R. Finn, "On a theorem of Lancaster and Siegel", *Pacific J. Math.* **213**:1 (2004), 111–119. MR 2004m:76038 Zbl 1156.76361

[Simon 1976] L. Simon, "Boundary regularity for solutions of the non-parametric least area problem", *Ann. of Math.* (2) **103**:3 (1976), 429–455. MR 58 #30681 Zbl 0335.49031

[Williams 1978] G. H. Williams, "Surfaces of prescribed mean curvature with inequalities on the boundary", *Math. Z.* **164**:1 (1978), 31–51. MR 80d:53006 Zbl 0378.49029

Received November 3, 2014. Revised March 21, 2015.

KIRK E. LANCASTER
DEPARTMENT OF MATHEMATICS, STATISTICS, AND PHYSICS
WICHITA STATE UNIVERSITY
WICHITA, KS 67260-0033
UNITED STATES
lancaster@math.wichita.edu

JARON MELIN
DEPARTMENT OF MATHEMATICS, STATISTICS, AND PHYSICS
WICHITA STATE UNIVERSITY
WICHITA, KS 67260-0033
UNITED STATES
melin@math.wichita.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.437

BRIDGE SPHERES FOR THE UNKNOT ARE TOPOLOGICALLY MINIMAL

JUNG HOON LEE

Topologically minimal surfaces were defined by Bachman as topological analogues of geometrically minimal surfaces, and one can associate a topological index to each topologically minimal surface. We show that an $(n+1)$-bridge sphere for the unknot is a topologically minimal surface of index at most n.

1. Introduction

Let S be a closed orientable separating surface embedded in a 3-manifold M. The structure of the set of compressing disks for S, such as how a pair of compressing disks on opposite sides of S intersects, reveals some topological properties of M. For example, if S is a minimal genus Heegaard surface of an irreducible manifold M and S has a pair of disjoint compressing disks on opposite sides, then M contains an incompressible surface [Casson and Gordon 1987].

The *disk complex* $\mathcal{D}(S)$ of S is a simplicial complex defined as follows.

- Vertices of $\mathcal{D}(S)$ are isotopy classes of compressing disks for S.
- A collection of $k+1$ vertices forms a k-simplex if there are representatives for each that are pairwise disjoint.

The disk complex of an incompressible surface is empty. A surface S is *strongly irreducible* if S compresses to both sides and every compressing disk for S on one side intersects every compressing disk on the opposite side. So the disk complex of a strongly irreducible surface is disconnected. Extending these notions, Bachman [2010] defined topologically minimal surfaces, which can be regarded as topological analogues of (geometrically) minimal surfaces.

A surface S is *topologically minimal* if $\mathcal{D}(S)$ is empty or $\pi_i(\mathcal{D}(S))$ is nontrivial for some i. The *topological index* of S is 0 if $\mathcal{D}(S)$ is empty, and the smallest n such that $\pi_{n-1}(\mathcal{D}(S))$ is nontrivial, otherwise.

Topologically minimal surfaces share some useful properties. For example, if an irreducible manifold contains a topologically minimal surface and an incompressible

MSC2010: 57M50.
Keywords: disk complex, topologically minimal surface, bridge splitting, unknot.

surface, then the two surfaces can be isotoped so that any intersection loop is essential in both surfaces. There exist topologically minimal surfaces of arbitrarily high index [Bachman and Johnson 2010], and see also [Lee 2015] for possibly high index surfaces in (closed orientable surface) $\times I$. In this paper we consider bridge splittings of 3-manifolds, and show that the simplest bridge surfaces, bridge spheres for the unknot in S^3, are topologically minimal. The main idea is to construct a retraction from the disk complex of a bridge sphere to S^{n-1} as in [Bachman and Johnson 2010] and [Lee 2015].

Theorem 1.1. *An $(n+1)$-bridge sphere for the unknot is a topologically minimal surface of index at most n.*

In particular, the topological index of a 3-bridge sphere for the unknot is two. We conjecture that the topological index of an $(n+1)$-bridge sphere for the unknot is n. There is another conjecture that the topological index of a genus n Heegaard surface of S^3 is $2n-1$. This correspondence may be due to the fact that a genus n Heegaard splitting of S^3 can be obtained as a 2-fold covering of S^3 branched along an unknot in $(n+1)$-bridge position.

2. Bridge splitting

For a closed 3-manifold M, a *Heegaard splitting* $M = V^+ \cup_S V^-$ is a decomposition of M into two handlebodies V^+ and V^- with $\partial V^+ = \partial V^- = S$. The surface S is called a *Heegaard surface* of the Heegaard splitting.

Let K be a knot in M such that $V^\pm \cap K$ is a collection of n boundary-parallel arcs $\{a_1^\pm, \dots, a_n^\pm\}$ in $V^\pm$. Each $a_i^\pm$ is called a *bridge*. The decomposition

$$(M, K) = (V^+, V^+ \cap K) \cup_S (V^-, V^- \cap K)$$

is called a *bridge splitting* of (M, K), and we say that K is in *n-bridge position* with respect to S. A bridge $a_i^\pm$ cobounds a *bridge disk* $\Delta_i^\pm$ with an arc in S. We can take the bridge disks Δ_i^+ ($i = 1, \dots, n$) to be mutually disjoint, and similarly for Δ_i^- ($i = 1, \dots, n$). By a *bridge surface*, we mean $S - K$. The set of vertices of $\mathcal{D}(S - K)$ consists of compressing disks for $S - K$ in $V^+ - K$ and $V^- - K$.

Two bridge surfaces $S - K$ and $S' - K$ are equivalent if they are isotopic in $M - K$. An n-bridge position of the unknot in S^3 is unique for every n [Otal 1982], so for $n \geq 2$ it is *perturbed*, i.e., there exists a pair of bridge disks Δ_i^+ and Δ_j^- such that the arcs $\Delta_i^+ \cap S$ and $\Delta_j^- \cap S$ intersect at one endpoint. The uniqueness also holds for 2-bridge knots [Scharlemann and Tomova 2008] and torus knots [Ozawa 2011]. However, there are 3-bridge knots that admit multiple 3-bridge spheres [Birman 1976; Montesinos 1976].

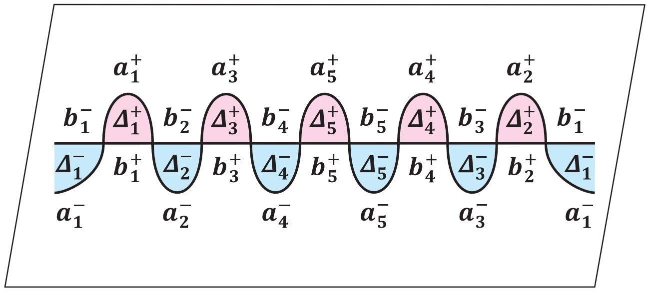

Figure 1. Bridges and bridge disks.

3. Proof of Theorem 1.1

Let S^3 be decomposed into two 3-balls B^+ and B^- with common boundary S. Let K be an unknot in S^3 which is in $(n+1)$-bridge position with respect to S. Then $K \cap B^\pm$ is a collection of $n+1$ bridges $a_i^\pm$ $(i = 1, \dots, n+1)$ in $B^\pm$. We assume that the bridges are arranged with $a_1^\pm$ adjacent to $a_1^\mp$ and $a_2^\mp$, with $a_i^\pm$ adjacent to $a_{i-1}^\mp$ and $a_{i+1}^\mp$ for $2 \le i \le n$, and with $a_{n+1}^\pm$ adjacent to $a_n^\mp$ and $a_{n+1}^\mp$. Let $\{\Delta_i^\pm\}$ be a collection of disjoint bridge disks $\Delta_i^\pm$ for $a_i^\pm$ with $\Delta_i^\pm \cap S = b_i^\pm$. We assume that $\operatorname{int} b_i^+ \cap \operatorname{int} b_j^- = \varnothing$ for any i and j. See Figure 1 for an example.

Let P be the $(2n+2)$-punctured sphere $S - K$. We define compressing disks $D_i^\pm$ $(i = 1, \dots, n)$ for P in $B^\pm - K$ as follows. Let D_1^+ be a disk in $B^+ - K$ such that $\partial D_1^+ = \partial N(b_1^+)$, where $N(b_1^+)$ is a neighborhood of b_1^+ taken in S. Similarly, other disks are defined so as to satisfy the following.

$$\begin{aligned}
\partial D_1^- &= \partial N(b_1^-),\\
\partial D_2^+ &= \partial N(b_1^+ \cup b_1^- \cup b_2^+),\\
\partial D_2^- &= \partial N(b_1^- \cup b_1^+ \cup b_2^-),\\
&\ \vdots\\
\partial D_i^+ &= \partial N(b_1^+ \cup b_1^- \cup \cdots \cup b_{i-1}^+ \cup b_{i-1}^- \cup b_i^+),\\
\partial D_i^- &= \partial N(b_1^- \cup b_1^+ \cup \cdots \cup b_{i-1}^- \cup b_{i-1}^+ \cup b_i^-),\\
&\ \vdots\\
\partial D_n^+ &= \partial N(b_1^+ \cup b_1^- \cup \cdots \cup b_{n-1}^+ \cup b_{n-1}^- \cup b_n^+),\\
\partial D_n^- &= \partial N(b_1^- \cup b_1^+ \cup \cdots \cup b_{n-1}^- \cup b_{n-1}^+ \cup b_n^-).
\end{aligned}$$

The $\partial D_i^\pm$'s in P are depicted in Figure 2.

Now we define subsets $C_i^\pm$ $(i = 1, \dots, n)$ of the set of vertices of $\mathcal{D}(P)$ as

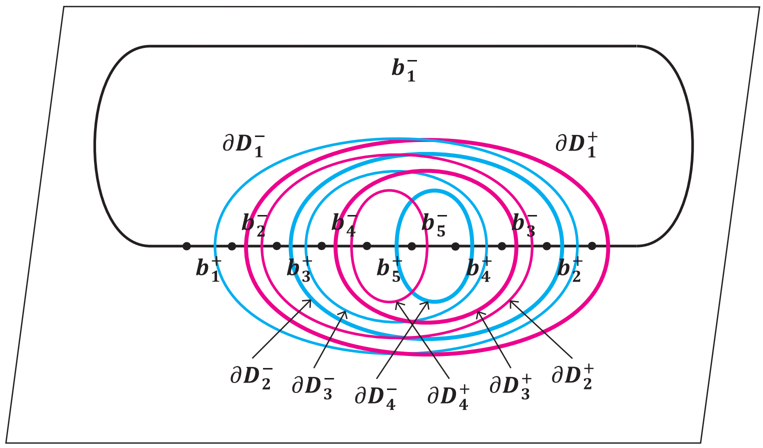

Figure 2. $\partial D_i^{\pm}$ $(i = 1, \ldots, n)$ in P.

follows. For odd i, let

$$C_i^+ = \{D_i^+\},$$
$$C_i^- = \{\text{essential disks in } B^- - K \text{ that intersect } D_i^+ \text{ and are disjoint from } D_1^+, D_3^+, \ldots, D_{i-2}^+\}.$$

For even i, let

$$C_i^+ = \{\text{essential disks in } B^+ - K \text{ that intersect } D_i^- \text{ and are disjoint from } D_2^-, D_4^-, \ldots, D_{i-2}^-\},$$
$$C_i^- = \{D_i^-\}.$$

Note that for all i, $D_i^{\pm}$ belongs to $C_i^{\pm}$.

Lemma 3.1. *The collection* $\{C_i^{\pm}\}$ $(i = 1, \ldots, n)$ *is a partition of the set of essential disks in* $B^{\pm} - K$.

Proof. First we show that $\{C_i^+\}$ $(i = 1, \ldots, n)$ is a partition of the set of essential disks in $B^+ - K$. We show that any essential disk in $B^+ - K$ belongs to one and only one C_i^+.

An essential disk in $B^+ - K$ that intersects D_2^- belongs to C_2^+ by definition. Let $E_2 = N(b_1^- \cup b_1^+ \cup b_2^-)$ be the disk in S such that $\partial E_2 = \partial D_2^-$.

Claim 1. *If an essential disk* D *in* $B^+ - K$ *is disjoint from* D_2^- *and* ∂D *is in* E_2, *then* D *is isotopic to* $D_1^+ \in C_1^+$.

Proof of Claim 1. We assume that D intersects D_1^+ transversely and minimally, so $D \cap D_1^+$ consists of arc components. Let $E_1 = N(b_1^+)$ be the disk in S such that $\partial E_1 = \partial D_1^+$. See Figure 3. Suppose that $D \cap D_1^+ \neq \varnothing$. Consider an outermost

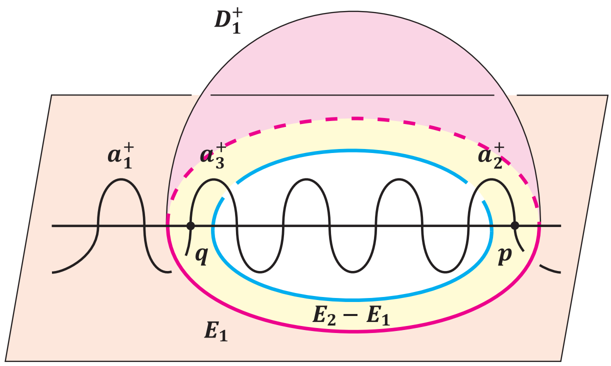

Figure 3. D_1^+ in C_1^+.

disk Δ of D cut off by an outermost arc of $D \cap D_1^+$. By the minimality of $|D \cap D_1^+|$, Δ cannot lie in the 3-ball B bounded by $D_1^+ \cup E_1$ containing a_1^+. So Δ lies outside of B. Let $\overline{D}$ be one of the disks obtained from D_1^+ by surgery along Δ such that $\partial \overline{D}$ bounds a disk $\overline{E}$ in $E_2 - E_1$. Let p be the point $a_2^+ \cap (E_2 - E_1)$ and q be the point $a_3^+ \cap (E_2 - E_1)$.

Suppose $\overline{E}$ contains p. Then the sphere $\overline{D} \cup \overline{E}$ intersects $a_2^+ \cup b_2^+$ in a single point after a slight isotopy of $\operatorname{int} b_2^+$ into B^-, a contradiction. So $\overline{E}$ does not contain p, and by similar reasoning $\overline{E}$ does not contain q. Then $\overline{E}$ is an inessential disk in $E_2 - E_1 - K$, so we can reduce $|D \cap D_1^+|$, a contradiction.

Hence $D \cap D_1^+ = \varnothing$. Let E be the disk in E_2 such that $\partial E = \partial D$. If ∂E is in E_1, then D is isotopic to D_1^+. Suppose ∂E is in $E_2 - E_1$. Then E contains neither p nor q, since otherwise $D \cup E$ intersects $a_2^+ \cup b_2^+$ or $a_3^+ \cup b_3^+$ in a single point as above. So we get the conclusion that D is isotopic to D_1^+. □

Therefore if an essential disk in $B^+ - K$ is disjoint from D_2^- and its boundary is in $S - E_2$, then it belongs to one of $C_3^+, \dots, C_n^+$.

An essential disk in $B^+ - K$ that is disjoint from D_2^- and intersects D_4^- belongs to C_4^+ by definition. Let $E_4 = N(b_1^- \cup b_1^+ \cup \cdots \cup b_3^- \cup b_3^+ \cup b_4^-)$ be the disk in S such that $\partial E_4 = \partial D_4^-$. Let D be an essential disk in $B^+ - K$ that is disjoint from D_2^- and D_4^- and such that $\partial D \subset S - E_2$.

Claim 2. *If ∂D is in E_4 (hence in $E_4 - E_2$), then D is isotopic to $D_3^+ \in C_3^+$.*

Proof of Claim 2. We assume that $|D \cap D_3^+|$ is minimal up to isotopy, so $D \cap D_3^+$ consists of arc components. Let $E_3 = N(b_1^+ \cup b_1^- \cup b_2^+ \cup b_2^- \cup b_3^+)$ be the disk in S such that $\partial E_3 = \partial D_3^+$. See Figure 4.

Suppose that $D \cap D_3^+ \neq \varnothing$. Consider an outermost disk Δ of D cut off by an outermost arc of $D \cap D_3^+$. Without loss of generality, we assume that $\partial \Delta \cap S$ lies in $E_3 - E_2$. Let $\overline{D}$ be one of the disks obtained from D_3^+ by surgery along Δ such

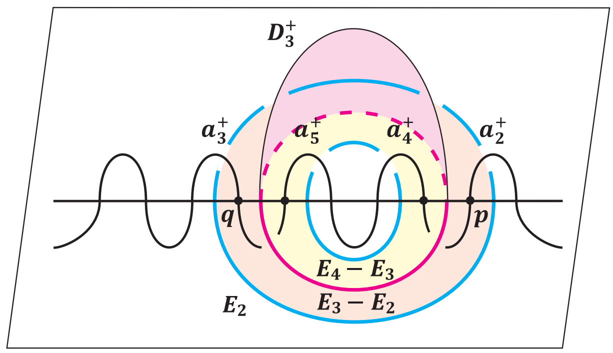

Figure 4. D_3^+ in C_3^+.

that $\partial\bar{D}$ bounds a disk $\bar{E}$ in E_3-E_2. Let p be the point $a_2^+\cap(E_3-E_2)$ and q be the point $a_3^+\cap(E_3-E_2)$.

Suppose $\bar{E}$ contains p. Then the sphere $\bar{D}\cup\bar{E}$ intersects $a_2^+\cup b_2^+$ in a single point after a slight isotopy, a contradiction. So $\bar{E}$ does not contain p, and similarly $\bar{E}$ does not contain q. Then $\bar{E}$ is an inessential disk in E_3-E_2-K, so we can reduce $|D\cap D_3^+|$, a contradiction. Hence $D\cap D_3^+=\varnothing$. Then, reasoning as we did for Claim 1, we see that D is isotopic to D_3^+. □

Therefore if an essential disk in B^+-K is disjoint from D_2^- and D_4^- and its boundary is in $S-E_4$, then it belongs to one of $C_5^+,\dots,C_n^+$.

In general, let $E_{2i}=N(b_1^-\cup b_1^+\cup\cdots\cup b_{2i-1}^-\cup b_{2i-1}^+\cup b_{2i}^-)$ be the disk in S such that $\partial E_{2i}=\partial D_{2i}^-$. Let D be an essential disk in B^+-K that is disjoint from $D_2^-,D_4^-,\dots,D_{2i-2}^-$ and such that $\partial D\subset S-E_{2i-2}$.

- If $\partial D\subset E_{2i}-E_{2i-2}$, then D is isotopic to $D_{2i-1}^+\in C_{2i-1}^+$.
- If D intersects D_{2i}^-, then D belongs to C_{2i}^+ by definition.
- If $\partial D\subset S-E_{2i}$, then D belongs to one of $C_{2i+1}^+,\dots,C_n^+$.

An inductive argument in this way leads to the conclusion that any essential disk in B^+-K belongs to one and only one C_i^+. A similar argument shows that $\{C_i^-\}$ $(i=1,\dots,n)$ is a partition of the set of essential disks in B^--K. □

The collection of disks $\{D_1^+,D_1^-,\dots,D_n^+,D_n^-\}$ spans an $(n-1)$-sphere S^{n-1} in $\mathcal{D}(P)$. There is no edge in $\mathcal{D}(P)$ connecting C_i^+ and C_i^- by definition. There exists an edge in $\mathcal{D}(P)$ connecting $C_i^\pm$ and $C_j^\pm$ for $i\neq j$, e.g., an edge between $D_i^\pm$ and $D_j^\pm$, and there exists an edge in $\mathcal{D}(P)$ connecting C_i^+ and C_j^- for $i\neq j$, e.g., an edge between D_i^+ and D_j^-. Hence if we define a map $\bar{r}$ from the set of vertices of $\mathcal{D}(P)$ to the set of vertices of S^{n-1} by

$$\bar{r}(v)=D_i^\pm \quad \text{if } v\in C_i^\pm,$$

then $\bar{r}$ extends to a continuous map from the 1-skeleton of $\mathcal{D}(P)$ to the 1-skeleton of S^{n-1}. Since higher-dimensional simplices of $\mathcal{D}(P)$ are determined by 1-simplices, the map $\bar{r}$ can be extended to a retraction $r : \mathcal{D}(P) \to S^{n-1}$. Hence $\pi_{n-1}(\mathcal{D}(P)) \neq 1$, and the topological index of P is at most n.

Acknowledgement

The author was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (2015R1D1A1A01056953).

References

[Bachman 2010] D. Bachman, "Topological index theory for surfaces in 3-manifolds", *Geom. Topol.* **14**:1 (2010), 585–609. MR 2011f:57042 Zbl 1206.57020

[Bachman and Johnson 2010] D. Bachman and J. Johnson, "On the existence of high index topologically minimal surfaces", *Math. Res. Lett.* **17**:3 (2010), 389–394. MR 2011e:57025 Zbl 1257.57026

[Birman 1976] J. S. Birman, "On the stable equivalence of plat representations of knots and links", *Canad. J. Math.* **28**:2 (1976), 264–290. MR 53 #6529 Zbl 0339.55005

[Casson and Gordon 1987] A. J. Casson and C. M. Gordon, "Reducing Heegaard splittings", *Topology Appl.* **27**:3 (1987), 275–283. MR 89c:57020 Zbl 0632.57010

[Lee 2015] J. H. Lee, "On topologically minimal surfaces of high genus", *Proc. Amer. Math. Soc.* **143**:6 (2015), 2725–2730. MR 3326050 Zbl 1315.57023

[Montesinos 1976] J. M. Montesinos, "Minimal plat representations of prime knots and links are not unique", *Canad. J. Math.* **28**:1 (1976), 161–167. MR 53 #14470 Zbl 0314.57011

[Otal 1982] J.-P. Otal, "Présentations en ponts du nœud trivial", *C. R. Acad. Sci. Paris Sér. I Math.* **294**:16 (1982), 553–556. MR 84a:57006 Zbl 0498.57001

[Ozawa 2011] M. Ozawa, "Nonminimal bridge positions of torus knots are stabilized", *Math. Proc. Cambridge Philos. Soc.* **151**:2 (2011), 307–317. MR 2012m:57015 Zbl 1226.57012

[Scharlemann and Tomova 2008] M. Scharlemann and M. Tomova, "Uniqueness of bridge surfaces for 2-bridge knots", *Math. Proc. Cambridge Philos. Soc.* **144**:3 (2008), 639–650. MR 2009c:57020 Zbl 1152.57006

Received December 22, 2014. Revised May 29, 2015.

JUNG HOON LEE
DEPARTMENT OF MATHEMATICS AND INSTITUTE OF PURE AND APPLIED MATHEMATICS
CHONBUK NATIONAL UNIVERSITY
JEONJU 54896
SOUTH KOREA

junghoon@jbnu.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.445

ON THE GEOMETRIC CONSTRUCTION OF COHOMOLOGY CLASSES FOR COCOMPACT DISCRETE SUBGROUPS OF $\mathrm{SL}_n(\mathbb{R})$ AND $\mathrm{SL}_n(\mathbb{C})$

SUSANNE SCHIMPF

We construct nontrivial cohomology classes for certain cocompact discrete subgroups of $\mathrm{SL}_n(\mathbb{R})$ and $\mathrm{SL}_n(\mathbb{C})$ using a geometric method. The discrete subgroups are of arithmetic nature, i.e., they arise from arithmetic subgroups of suitably chosen algebraic groups. In certain cases, we show the nonvanishing of automorphic representations as a consequence.

1. Introduction

This paper contributes to the research on cohomology of arithmetic groups by providing a nonvanishing result for the cohomology of certain families of cocompact discrete subgroups of the real Lie groups $\mathrm{SL}_n(\mathbb{R})$ and $\mathrm{SL}_n(\mathbb{C})$. The discrete subgroups are of arithmetic nature, i.e., they arise from arithmetic subgroups of suitably chosen algebraic groups. Our approach is via a geometric argument.

The main result. Let G be a semisimple real Lie group with finite center. Denote by K a maximal compact subgroup and by Γ a torsion-free discrete subgroup of G. The action of Γ on the symmetric space $X := K\backslash G$ is smooth, proper and free, and the quotient X/Γ is a $K(\Gamma, 1)$-space. In particular, one has $H^*(\Gamma, \mathbb{C}) = H^*(X/\Gamma; \mathbb{C})$, i.e., the group cohomology of Γ with respect to the trivial Γ-module $\mathbb{C}$ equals the singular cohomology of X/Γ with complex coefficients.

A particularly interesting case is the situation where the discrete subgroup Γ is cocompact, i.e., the locally symmetric space X/Γ is compact. General results by Borel [1963] and Borel and Harder [1978] imply that such cocompact subgroups can be constructed as arithmetic subgroups of suitable algebraic groups defined over some algebraic number field. One can then use geometric methods to study the cohomology of the compact locally symmetric space X/Γ. Assuming the space

The author was supported by the FWF Austrian Science Fund, grant P 21090-N13.
MSC2010: primary 11F75; secondary 11F70, 11E57.
Keywords: cohomology, arithmetic subgroup, special linear group, geometric cycle, automorphic representation.

X/Γ is orientable, one approach is to construct certain oriented, totally geodesic submanifolds (so-called geometric cycles) and show that their fundamental homology classes contribute nontrivially to the cohomology of X/Γ via Poincaré duality. Such methods have been successfully applied to discrete subgroups of several classical and exceptional Lie groups including $\mathrm{SO}(p,q)$, $\mathrm{SU}(p,q)$, $\mathrm{SU}^*(2n)$ and G_2; see [Millson and Raghunathan 1981; Schwermer and Waldner 2011; Waldner 2010]. In this work, we deal with the special linear group over the real and the complex numbers. We obtain a result of the following form (see Theorems 5.6 and 6.5).

Theorem. *Let $n \in \mathbb{N}$, $n \geq 2$.*

(1) *Let $X := \mathrm{SO}(n)\backslash \mathrm{SL}_n(\mathbb{R})$ denote the symmetric space attached to the real Lie group $\mathrm{SL}_n(\mathbb{R})$. If n is even, there exists a discrete cocompact arithmetically defined subgroup $\Gamma \subset \mathrm{SL}_n(\mathbb{R})$ such that $H^k(X/\Gamma;\mathbb{C})$ contains nontrivial cohomology classes for all k of the form*

$$k = pq \quad \textit{and} \quad k = \tfrac{1}{2}(p^2+q^2+n)-1,$$

where p and q are positive integers with $p+q=n$, and, if $n \neq 2$, for

$$k = \tfrac{1}{4}(n^2+2n) \quad \textit{and} \quad k = \tfrac{1}{4}n^2-1.$$

(2) *Let $X := \mathrm{SU}(n)\backslash \mathrm{SL}_n(\mathbb{C})$ denote the symmetric space attached to the real Lie group $\mathrm{SL}_n(\mathbb{C})$. There exists a discrete cocompact arithmetically defined subgroup $\Gamma \subset \mathrm{SL}_n(\mathbb{C})$ such that $H^k(X/\Gamma;\mathbb{C})$ contains nontrivial cohomology classes for all k of the form*

$$k = 2pq \quad \textit{and} \quad k = p^2+q^2-1,$$

where p and q are positive integers with $p+q=n$, and for

$$k = \tfrac{1}{2}(n^2-n) \quad \textit{and} \quad k = \tfrac{1}{2}(n^2+n)-1.$$

Moreover, if n is even and $n \neq 2$, there are nontrivial cohomology classes in the degrees

$$k = \tfrac{1}{2}(n^2+n), \quad k = \tfrac{1}{2}(n^2-n)-1, \quad k = \tfrac{1}{2}n^2-1 \quad \textit{and} \quad k = \tfrac{1}{2}n^2.$$

When $H^*(X/\Gamma,\mathbb{C})$ is interpreted as the cohomology of the de Rham complex $\Omega^*(X/\Gamma,\mathbb{C})$, the constructed classes are not represented by $\mathrm{SL}_n(\mathbb{R})$- or $\mathrm{SL}_n(\mathbb{C})$-invariant differential forms on X.

A geometric method. The geometric method we are using to obtain our result was developed by Millson and Raghunathan [1981] and is based on an earlier result of Millson [1976] about the nonvanishing of the first Betti number of certain compact hyperbolic manifolds.

Their approach applies to the situation where the Lie group G is the group of real points of a reductive algebraic $\mathbb{Q}$-group and Γ is a cocompact torsion-free arithmetic subgroup of this algebraic group. Under the assumption that the space X/Γ is orientable, they consider so-called geometric cycles, orientable totally geodesic submanifolds of X/Γ. Then the approach of Millson and Raghunathan is based on finding two such geometric cycles of complementary dimension in X/Γ that intersect transversally and with positive multiplicity in all points of intersection. Under this assumption, the fundamental classes of the two submanifolds have nontrivial intersection number, and hence they contribute nontrivially to the cohomology of X/Γ. In 1993, Rohlfs and Schwermer found a way to generalize the method in such a way that it also applies to nontransversal intersections, by using the theory of so-called excess bundles. Their work involves the investigation of deep orientability questions.

As the proof of our result is heavily based on the method of Rohlfs and Schwermer, we give an overview of the relevant notions and their main theorem in Section 3. Then, Section 4 is devoted to introducing the framework of algebraic groups in which the construction of geometric cycles and the associated cohomology classes is carried out: for our groups of interest, $\mathrm{SL}_n(\mathbb{R})$ and $\mathrm{SL}_n(\mathbb{C})$, the algebraic group to start with is the special unitary group

$$\boldsymbol{G} := \mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})$$

defined over an algebraic number field F, where D is a division algebra with involution σ and h denotes a σ-hermitian form on D^m. Under certain assumptions, the associated real Lie group $\boldsymbol{G}_\infty$ is isomorphic to $\mathrm{SL}_n(\mathbb{R})$ or $\mathrm{SL}_n(\mathbb{C})$ up to compact factors and can be used for the construction of cocompact discrete subgroups. In this setting, the construction of geometric cycles and the application of the method of Rohlfs and Schwermer to obtain nontrivial cohomology classes is carried out in Sections 5 and 6, for the real and complex case, respectively. The main results are stated as Theorems 5.6 and 6.5.

Automorphic representations. Nonvanishing results for the cohomology of cocompact discrete subgroups can be applied to the theory of automorphic representations using a well-known result of Matsushima that allows one to interpret the cohomology of X/Γ in terms of the relative Lie algebra cohomology of irreducible unitary representations of G. Thus, we have devoted Section 7 to the study of representations with nontrivial $(\mathfrak{g}, K)$-cohomology occurring in Matsushima's formula for the group $G = \mathrm{SL}_n(\mathbb{C})$. Making explicit general results of Enright [1979] and Delorme [1979] for simply connected complex Lie groups for the case of $\mathrm{SL}_n(\mathbb{C})$, we obtain a complete classification of the equivalence classes of irreducible unitary representations with nontrivial $(\mathfrak{g}, K)$-cohomology. By comparing the occurring degrees in which X/Γ may possibly have nontrivial cohomology with those detected by special cycles, we can identify specific automorphic representations of G with respect to Γ for small values of n.

2. Notation

• For an algebraic number field k, we let $V = V(k)$ and $V_\infty = V_\infty(k)$ denote its set of places and archimedean places, respectively. For a place $v \in V$, we denote by k_v the completion of k at v.

• All algebraic groups are assumed to be linear, i.e., they can be considered as smooth affine algebraic group schemes. We denote algebraic groups by bold letters ($\boldsymbol{G}$, $\boldsymbol{H}$, ...). For an algebraic group $\boldsymbol{G}$ defined over a number field k, we set $\boldsymbol{G}_\infty := \prod_{v \in V_\infty} \boldsymbol{G}(k_v)$.

• Lie groups are denoted by standard Roman letters (G, H, ...). Whenever we speak of a semisimple Lie group, we assume that it has finite center and finitely many connected components.[1] We use the notion of a *reductive* Lie group as in [Knapp 1996, Section VII.2].

• For a semisimple Lie group G, we denote by $\widehat{G}$ the unitary dual of G, that is, the set of unitary equivalence classes of irreducible unitary representations.

• Lie algebras are denoted by small German letters ($\mathfrak{g}$, $\mathfrak{h}$, ...) and can be real or complex depending on context. If $\mathfrak{g}$ is a real Lie algebra, we will denote by $\mathfrak{g}_\mathbb{C}$ its complexification and if $\mathfrak{g}$ is complex we write $\mathfrak{g}_\mathbb{R}$ for the real Lie algebra underlying $\mathfrak{g}$. In general, we denote the Lie algebra of a Lie group G by $\mathfrak{g}$ and consider it as a real or complex Lie algebra depending on whether G is a real or a complex group.

• Let R be a ring and let $n \in \mathbb{N}$. We denote by I_n the $n \times n$ unity matrix in $M_n(R)$ and by $I_{p,q}$ the matrix $\operatorname{diag}(I_p, -I_q) \in M_n(R)$, for $p+q=n$. For even n, we set $J_n := \left(\begin{smallmatrix} 0 & -I_{n/2} \\ I_{n/2} & 0 \end{smallmatrix}\right)$.

3. A geometric method

This section gives a brief summary of the method of Rohlfs and Schwermer [1993] for the geometric construction of nontrivial cohomology classes.

3.1. ***Special cycles.*** Let $\boldsymbol{G}$ be a connected reductive algebraic group defined over $\mathbb{Q}$ and write G for its group of real points $\boldsymbol{G}(\mathbb{R})$. Then G is a real reductive Lie group with a maximal compact subgroup $K \subset G$ and we can form the associated symmetric space $X := K\backslash G$. Let now $\Gamma \subset \boldsymbol{G}(\mathbb{Q})$ be a torsion-free arithmetic subgroup. Then

[1] This is to ensure that the semisimple groups are also reductive in the sense of [Knapp 1996]. Lie groups arising as the groups of real or complex points of semisimple algebraic groups will always have this property.

Γ is a discrete subgroup of G and it acts on the symmetric space X by right translations. This action is smooth, proper and free, and the quotient X/Γ is a Riemannian locally symmetric space.

Let $\boldsymbol{\mu}$ be a $\mathbb{Q}$-rational automorphism of $\boldsymbol{G}$ and assume that K and Γ are invariant under $\boldsymbol{\mu}$. Then the fixed point group $\mathrm{Fix}(\boldsymbol{\mu}, \boldsymbol{G})$ is a reductive subgroup of $\boldsymbol{G}$ and we can associate with it the locally symmetric space $C(\mu, \Gamma) := \mathrm{Fix}(\boldsymbol{\mu}, K)\backslash \mathrm{Fix}(\boldsymbol{\mu}, \boldsymbol{G})(\mathbb{R})/\mathrm{Fix}(\boldsymbol{\mu}, \Gamma)$. This is a connected, totally geodesic submanifold of X/Γ and is called the *special cycle* associated with $\boldsymbol{\mu}$.

Let us now assume that Θ is the group generated by two commuting $\mathbb{Q}$-rational automorphisms $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$ of $\boldsymbol{G}$ of finite order and that K and Γ are Θ-invariant.[2] In general, the locally symmetric space X/Γ and its special cycles need not be orientable. However, it was shown by Rohlfs and Schwermer [1993] that by passing to a suitable subgroup of finite index in Γ, one can always assume that the manifolds X/Γ, $C(\boldsymbol{\tau}_1, \Gamma)$, $C(\boldsymbol{\tau}_2, \Gamma)$ and the (finitely many) connected components of their intersection are orientable.

Let us denote by $[C(\boldsymbol{\tau}_i, \Gamma)]$ the fundamental homology class of $C(\boldsymbol{\tau}_i, \Gamma)$ in $H_*(C(\boldsymbol{\tau}_i, \Gamma))$ and for simplicity also its image in $H_*(X/\Gamma)$, for $i \in \{1, 2\}$. If we assume in addition that the two cycles are of complementary dimension in X/Γ, we can look at their intersection number $[C(\boldsymbol{\tau}_1, \Gamma)][C(\boldsymbol{\tau}_2, \Gamma)]$.

Since these submanifolds need not necessarily intersect transversally, the determination of their intersection number is a complicated issue. It involves the computation of Euler numbers of a certain *excess bundle*. Under certain assumptions connected to deep orientability questions for the involved manifolds, Rohlfs and Schwermer have come up with a nonvanishing result for the intersection number:

Theorem 3.2 [Rohlfs and Schwermer 1993, Theorem 4.11]. *Let $\boldsymbol{G}$ be a reductive algebraic $\mathbb{Q}$-group, let $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ be $\mathbb{Q}$-rational automorphisms of $\boldsymbol{G}$ of finite order, and let Γ be a torsion-free, $\langle\boldsymbol{\tau}_1, \boldsymbol{\tau}_2\rangle$-stable, cocompact arithmetic subgroup of $\boldsymbol{G}$ such that X/Γ, $C(\boldsymbol{\tau}_1, \Gamma)$, $C(\boldsymbol{\tau}_2, \Gamma)$ and all connected components of their intersection are orientable. Suppose that the associated cycles $C(\boldsymbol{\tau}_1)$ and $C(\boldsymbol{\tau}_2)$ are of complementary dimension. Assume that*

(i) *the real Lie groups $\boldsymbol{G}(\mathbb{R})$, $\mathrm{Fix}(\boldsymbol{\tau}_1, \boldsymbol{G})(\mathbb{R})$ and $\mathrm{Fix}(\boldsymbol{\tau}_2, \boldsymbol{G})(\mathbb{R})$ act orientation-preservingly on X, $X(\boldsymbol{\tau}_1)$ and $X(\boldsymbol{\tau}_2)$, respectively, and*

(ii) *the group $\mathrm{Fix}(\langle\boldsymbol{\tau}_1, \boldsymbol{\tau}_2\rangle, \boldsymbol{G})(\mathbb{R})$ is compact.*

Then there exists a $\langle\boldsymbol{\tau}_1, \boldsymbol{\tau}_2\rangle$-stable normal subgroup $\Gamma' \subset \Gamma$ of finite index such that

$$[C(\boldsymbol{\tau}_1, \Gamma')][C(\boldsymbol{\tau}_2, \Gamma')] \neq 0.$$

[2] Such a choice of K and Γ is always possible without loss of generality. For K, this follows from [Helgason 1978, Theorem 13.5]. For Γ, set $\Gamma' := \bigcap_{\theta\in\Theta}\theta(\Gamma)$. Then Γ' is of finite index in Γ and stable under Θ since the elements of Θ are automorphisms of finite order.

Remark. Condition (i) is quite restricting and we will see below natural choices for $\boldsymbol{G}$, $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ where it is not met. Note that the condition is satisfied if $\boldsymbol{G}(\mathbb{R})$, $\mathrm{Fix}(\boldsymbol{\tau}_1, \boldsymbol{G})(\mathbb{R})$ and $\mathrm{Fix}(\boldsymbol{\tau}_2, \boldsymbol{G})(\mathbb{R})$ are connected.

Clearly, the nonvanishing of the intersection number implies the nonvanishing of the homology classes $[C(\boldsymbol{\tau}_i, \Gamma')]$ in $H_*(X/\Gamma', \mathbb{C})$ and of the respective cohomology classes obtained via Poincaré duality, for $i \in \{1, 2\}$.

For a compact quotient X/Γ, it is well-known that there exists an injective homomorphism $\beta_\Gamma^* : H^*(X_u, \mathbb{C}) \to H^*(X/\Gamma, \mathbb{C})$, where X_u denotes the compact dual symmetric space of X. When interpreting $H^*(X/\Gamma, \mathbb{C})$ in terms of de Rham cohomology, the classes in the image of this map can be identified with the G-invariant differential forms on X. It was shown by Millson and Raghunathan [1981] that under certain conditions the classes constructed with Theorem 3.2 are *new* in the sense that they do not lie in the image of β_Γ:

Theorem 3.3. *Let $\boldsymbol{G}$, $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$ and Γ satisfy the assumptions of Theorem 3.2 and suppose moreover that $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are of order two. Then there exists a $\langle \boldsymbol{\tau}_1, \boldsymbol{\tau}_2 \rangle$-stable subgroup Γ'' of Γ' of finite index such that the nontrivial cohomology classes defined by $[C(\boldsymbol{\tau}_1, \Gamma'')]$ and $[C(\boldsymbol{\tau}_2, \Gamma'')]$ via Poincaré duality are not in the image of $\beta_{\Gamma''}^*$.*

Example 3.4. Consider the real Lie group $G = \mathrm{SO}(p, q)$ with maximal compact subgroup $K = S(O(p) \times O(q))$. The group K (and hence also G) is not connected but has two connected components that are distinguished by the determinant of the upper left $(p \times p)$-block. One can show that the action of G on the quotient $X := K \backslash G$ by left translations is orientation-preserving if and only if $n = p + q$ is even. Note that G is the fixed point group of the involution $x \mapsto I_{p,q}(x^t)^{-1} I_{p,q}$ in the connected real Lie group $\mathrm{SL}_n(\mathbb{R})$. Hence, for odd n, this is an example of a fixed point group that does not meet the orientability condition (i) in Theorem 3.2.

A similar argument also applies to the real Lie group $G = S(\mathrm{GL}_p(\mathbb{R}) \times \mathrm{GL}_q(\mathbb{R}))$ with maximal compact subgroup $K = S(O(p) \times O(q))$.

4. The setup: the construction of discrete cocompact subgroups of $\mathrm{SL}_n(\mathbb{R})$ and $\mathrm{SL}_n(\mathbb{C})$

In this section we will see how to construct cocompact discrete subgroups of $\mathrm{SL}_n(\mathbb{R})$ or $\mathrm{SL}_n(\mathbb{C})$ using an arithmetic method based on the compactness criterion by Borel and Harish-Chandra. The starting point is the special unitary group over a division algebra.

Let E be an algebraic number field and D a central division algebra over E of degree d endowed with an involution σ of the first or second kind. Recall that if σ is of the second kind, there exists a subfield F of E of index 2 such that $\sigma|_F = \mathrm{id}$ and $\sigma|_E = \iota$, where ι is the nontrivial Galois automorphism of E over F. For simplicity of notation, we set $F := E$ and $\iota := \mathrm{id}$ in case σ is an involution of the first kind.

Let m be a natural number and let h be a σ-hermitian (or σ-skew-hermitian) form on D^m. Then the *special unitary group of rank m over D* is defined as

$$\mathrm{SU}_m(h, D, \sigma) := \{x \in \mathrm{SL}_m(D) \mid h(xv, xw) = h(v, w) \text{ for all } v, w \in D^m\},$$

where $\mathrm{SL}_m(D)$ denotes the group of matrices in $M_m(D)$ with reduced norm 1.

It is well-known that there exists a simply connected semisimple algebraic group defined over F, whose F-rational points coincide with $\mathrm{SU}_m(h, D, \sigma)$. We denote this group by $\mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})$. Indeed, on the algebraic F-group $\mathrm{Res}_{E/F}\,\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})$ we can define an F-rational morphism $\boldsymbol{\psi}$ that is given on the F-rational points by $\psi(x) = H^{-1}\sigma(x^t)^{-1}H$, where H is the matrix of h with respect to a chosen basis, and we have $\mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma}) = \mathrm{Fix}(\boldsymbol{\psi}, \mathrm{Res}_{E/F}\,\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))$.

Being an F-rational algebraic group, we can look at the real Lie group of F_v-rational points of $\mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})$ for any archimedean place $v \in V_\infty(F)$. The nature of this real Lie group depends on the properties of the place v and the splitting behavior of D at v. Recall that for a quadratic extension E/F a place $v \in V(F)$ is said to be *decomposed* in E if there are exactly two places $w \in V(E)$ such that $w \mid v$, and *nondecomposed* otherwise. We denote by ρ the involution on $M_m(D)$ given by $\rho(x) := H^{-1}\sigma(x)^t H$. The following result can be obtained as an application of results from the theory of algebras with involutions and some easy computations.

Proposition 4.1. (1) *Let σ be an involution of the first kind on D and assume that D splits at all real places of E. Let $w \in V_\infty(E)$ be an archimedean place of E. Then there are the following possibilities*:

- *If w is a complex place*, *we have*

$$\mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})(E_w) \cong \begin{cases} \mathrm{SO}(n, \mathbb{C}) & \text{if } \rho \text{ is of orthogonal type,} \\ \mathrm{Sp}(n, \mathbb{C}) & \text{if } \rho \text{ is of symplectic type.} \end{cases}$$

- *If w is a real place*, *we have*

$$\mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})(E_w) \cong \begin{cases} \mathrm{SO}(p, q) & \text{if } \rho \text{ is of orthogonal type,} \\ \mathrm{Sp}(n, \mathbb{R}) & \text{if } \rho \text{ is of symplectic type,} \end{cases}$$

  for suitable nonnegative integers p and q with $p + q = n$.

(2) *Let σ be an involution of the second kind on D and consider an archimedean place $v \in V_\infty(F)$. Then there are the following possibilities*:

- *If v is a complex place*, *we have* $\mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})(F_v) \cong \mathrm{SL}_n(\mathbb{C})$.
- *If v is a nondecomposed real place*, *we have* $\mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})(F_v) \cong \mathrm{SU}(p, q)$ *for nonnegative integers p, q with $p + q = n$.*

- *If v is a decomposed real place and $w_1 \mid v$ and $w_2 \mid v$ are the real places of E lying above v, we have*

$$\mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})(F_v) \cong \begin{cases} \mathrm{SL}_n(\mathbb{R}) & \text{if } D \text{ splits at } w_1 \text{ and } w_2, \\ \mathrm{SL}_{n/2}(\mathbb{H}) & \text{if } D \text{ ramifies at } w_1 \text{ and } w_2. \end{cases}$$

Using this result, we can now find certain conditions on the number fields E and F, the involution σ and the division algebra D such that arithmetic subgroups of $\mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})$ give rise to cocompact discrete subgroups of either $\mathrm{SL}_n(\mathbb{R})$ or $\mathrm{SL}_n(\mathbb{C})$. An involution σ of the second kind on D is called *definite* if for every real nondecomposed place v of F we have an isomorphism $(D, \sigma)_v \cong (M_d(\mathbb{C}), {}^*)$, where $x \mapsto x^* = \bar{x}^t$ denotes the conjugate-transpose involution on $M_d(\mathbb{C})$. In general, an involution of the second kind need not be definite. However, if there exists an involution of the second kind on D, there is also a definite one (see [Scharlau 1985, Chapter 10, Remark 6.11]).

Theorem 4.2. *Let ℓ be an archimedean local field and let $n \in \mathbb{N}$ be fixed.*

(1) *If $\ell = \mathbb{R}$, let F be a totally real number field with $[F : \mathbb{Q}] \geq 2$, let E/F be a quadratic extension such that there is exactly one place $v \in V_\infty(F)$ that is decomposed in E, and let D be a division algebra of degree $d \mid n$ over E that splits at the places w_1 and w_2 of E lying above v. Moreover, assume that there is a definite involution σ of the second kind on D.*

(2) *If $\ell = \mathbb{C}$, let F be an algebraic number field with $[F : \mathbb{Q}] \geq 3$ that has exactly one complex place v, and let E/F be a quadratic extension such that all real places of F are nondecomposed in E. Let D be a division algebra of degree $d \mid n$ over E that admits a definite involution σ of the second kind.*

Let $m \in \mathbb{N}$ be such that $dm = n$. Then one can choose a σ-hermitian form h on D^m such that any arithmetic subgroup $\Gamma \subset \mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})(F)$ gives rise to a discrete cocompact subgroup of $\mathrm{SL}_n(\ell)$.

Proof. Choose a hermitian form h in such a way that the matrix of h at each nondecomposed place v' of F has only positive eigenvalues (take the trivial hermitian form, for example). Set $\boldsymbol{G}' := \mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})$. Then, using Proposition 4.1 and the fact that σ is definite, it follows that

$$\boldsymbol{G}'(F_{v'}) \cong \mathrm{SU}(dm, 0) = \mathrm{SU}(n)$$

for all nondecomposed real places of F; in particular, these groups are compact. By our choice of E and F there is at least one such nondecomposed place and thus the group $\boldsymbol{G}'$ is anisotropic over F. On the other hand, at the decomposed place v we have $\boldsymbol{G}'(F_v) \cong \mathrm{SL}_n(\ell)$ by Proposition 4.1 and the fact that D is split at w_1 and

w_2 if $\ell = \mathbb{R}$. In particular, we have

$$\boldsymbol{G}'_\infty := \prod_{v \in V_\infty(F)} \boldsymbol{G}'(F_v) \cong \mathrm{SL}_n(\ell) \times \prod_{\substack{v' \in V_\infty(F) \\ \text{real, nondecomposed}}} \mathrm{SU}(n).$$

Let $\Gamma \subset \boldsymbol{G}'(F)$ be an arbitrary arithmetic subgroup of $\boldsymbol{G}'$. The image of Γ in $\boldsymbol{G}'_\infty$ under the diagonal embedding (still denoted by Γ) is a discrete subgroup. Since $\boldsymbol{G}'$ is semisimple and anisotropic over F, it follows from a well-know compactness criterion due to Borel and Harish-Chandra [1962] and Mostow and Tamagawa [1962] that the quotient $\boldsymbol{G}'_\infty / \Gamma$ is compact. Moreover, the image of Γ under the projection onto the noncompact factor of $\boldsymbol{G}'_\infty$ is a discrete cocompact subgroup of $\boldsymbol{G}'(F_v) \cong \mathrm{SL}_n(\ell)$. □

5. Geometric cycles for $\mathrm{SL}_n(\mathbb{R})$

Let F be a totally real number field of degree $r \geq 2$, and let E/F be a quadratic extension such that there is exactly one archimedean place $v \in V_\infty(F)$ that is decomposed in E and all other archimedean places of F are nondecomposed. Let us denote the nontrivial Galois automorphism of E/F by ι. Let D be a central division algebra of degree d over E with a definite involution σ of the second kind.

In this section we will restrict to the cases where D is either the field E itself (with $\sigma = \iota$) or a quaternion division algebra over E that splits at the places w_1 and w_2 of E lying above v and admits a definite involution σ of the second kind. In the latter case, we may assume by a theorem of Albert (see [Knus et al. 1998, Proposition 2.22]) that

$$D = Q(a, b \mid F) \otimes_F E = Q(a, b \mid E)$$

for some $a, b \in F^\times$ and that $\sigma = \tau_{c,0} \otimes \iota$, where $\tau_{c,0}$ denotes the conjugation on the quaternion algebra $Q(a, b \mid F)$.

Let $m \in \mathbb{N}$ be arbitrary and set $n := dm$. Then D, E, F, σ and m satisfy the conditions of Theorem 4.2(1) and we can find a hermitian form h such that any arithmetic subgroup of $\boldsymbol{G}' := \mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})$ gives rise to a discrete cocompact subgroup of $\mathrm{SL}_n(\mathbb{R})$. For technical reasons, we assume that the matrix H of h is a diagonal matrix in $M_m(F)$ that is positive definite under the embedding corresponding to the decomposed place v. If $D = E$ and if m is even, we assume in addition that H is a symplectic matrix, that is, it commutes with the matrix J_m.[3]

In order to construct special cycles for $\mathrm{SL}_n(\mathbb{R})$ we will now define suitable morphisms of finite order. To do this, we need a preparatory lemma. Recall that,

[3]Such a choice of H clearly exists: take the identity matrix, for example.

for each basis element $e \neq 1$ of a quaternion algebra, there exists an orthogonal involution τ_e that sends e to $-e$ and fixes all other basis elements.

Lemma 5.1. *Let E be a number field and $Q := Q(a, b \mid E)$ a quaternion algebra that splits at a real place w of E. Then there exist orthogonal involutions τ and $\tau_{(1,-1)}$ and an isomorphism $Q(a, b \mid E_w) \to M_2(\mathbb{R})$ such that*

$$(Q(a, b \mid E_w), \tau \otimes \mathrm{id}) \cong (M_2(\mathbb{R}),\ x \mapsto x^t),$$

and

$$(Q(a, b \mid E_w), \tau_{(1,-1)} \otimes \mathrm{id}) \cong \left(M_2(\mathbb{R}),\ x \mapsto \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) x^t \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)\right).$$

Proof. Let a_w and b_w denote the images of a and b under the embedding corresponding to w. Since D splits at w, exactly one of the elements a_w, b_w and $-a_w b_w$ is negative, i.e., there exists exactly one basis element $e_0 \in \{i, j, k\}$ such that e_0^2 is negative. Denote the remaining nontrivial basis elements by e_1 and e_2. We set $\tau := \tau_{e_0}$ and $\tau_{(1,-1)} := \tau_{e_1}$.

Now let φ be the $\mathbb{R}$-linear map given on the basis of $Q(a, b \mid E_w)$ by $1 \mapsto I_2$ and

$$e_0 \otimes 1 \mapsto \begin{pmatrix} 0 & \sqrt{-e_0^2} \\ -\sqrt{-e_0^2} & 0 \end{pmatrix}, \quad e_1 \otimes 1 \mapsto \begin{pmatrix} 0 & \sqrt{e_1^2} \\ \sqrt{e_1^2} & 0 \end{pmatrix}, \quad e_2 \otimes 1 \mapsto \begin{pmatrix} \sqrt{e_2^2} & 0 \\ 0 & -\sqrt{e_2^2} \end{pmatrix}.$$

Then $\varphi : Q(a, b \mid E_w) \to M_2(\mathbb{R})$ is a well-defined isomorphism under which $\tau_{e_0} \otimes \mathrm{id}$ goes over to $x \mapsto x^t$ and $\tau_{e_1} \otimes \mathrm{id}$ goes over to $x \mapsto \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) x^t \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. □

Remark. Note that the two orthogonal involutions τ and $\tau_{(1,-1)}$ commute and that we have $\tau \circ \tau_{(1,-1)} = \tau_{(1,-1)} \circ \tau = \mathrm{Int}(e_2)$. Moreover, τ commutes with the conjugation τ_c of Q and we have $\tau \circ \tau_c = \mathrm{Int}(e_0)$.

Let us return to our specific choice of a division algebra $D = Q(a, b \mid F) \otimes_F E$ as described above. Applying Lemma 5.1 to D, we get the existence of orthogonal involutions τ and $\tau_{(1,-1)}$ that are mapped to $x \mapsto x^t$ and $x \mapsto \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) x^t \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ under a suitable splitting isomorphism at the real place w_1.

Using these involutions, we can now define the following automorphisms of order two on $\mathrm{SL}_m(D)$:

$$\theta : \mathrm{SL}_m(D) \to \mathrm{SL}_m(D),$$
$$\theta(x) = \begin{cases} H^{-1}(x^t)^{-1} H & \text{if } D = E, \\ H^{-1}(\tau(x)^t)^{-1} H & \text{if } D = Q(a, b \mid E), \end{cases}$$

$$\mu : \mathrm{SL}_m(D) \to \mathrm{SL}_m(D),$$
$$\mu(x) = \begin{cases} H^{-1} J_m (x^t)^{-1} J_m^{-1} H & \text{if } D = E \text{ and } m > 2 \text{ even}, \\ H^{-1}(\tau_c(x)^t)^{-1} H & \text{if } D = Q(a, b \mid E) \text{ and } m > 1. \end{cases}$$

Moreover, for certain positive integers p and q such that $p+q=n$, we define a family of automorphisms $\nu_{p,q}:\mathrm{SL}_m(D)\to\mathrm{SL}_m(D)$ by

$$\nu_{p,q}(x)=\begin{cases} H^{-1}I_{p,q}((x)^t)^{-1}I_{p,q}H & \text{if } D=E,\\ H^{-1}I_{p/2,q/2}(\tau(x)^t)^{-1}I_{p/2,q/2}H & \text{if } D=Q(a,b\mid E) \text{ and } p,q \text{ even},\\ H^{-1}(\tau_{(1,-1)}(x)^t)^{-1}H & \text{if } D=Q(a,b\mid E)\\ & \qquad\text{and } p=q=n/2 \text{ is odd.}\end{cases}$$

Note that θ commutes with any of the other automorphisms.

5.2. To avoid case distinctions, we put in place the following general assumptions. Whenever we deal with the maps $\nu_{p,q}$ it should be understood that the parameters p and q are nonzero natural numbers satisfying $p+q=n$. Moreover, if D is a quaternion algebra, we assume that both p and q are even or that $p=q=n/2$. Furthermore, statements involving the map μ are only applicable when n is even and $n>2$.

The maps θ, $\nu_{p,q}$ and μ are basically built out of E-linear maps (involutions of the first kind on $\mathrm{SL}_m(D)$) and the group inversion, so they define E-rational morphisms $\boldsymbol{\theta}$, $\boldsymbol{\nu_{p,q}}$ and $\boldsymbol{\mu}$ on the algebraic E-group $\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})$ and F-rational automorphisms $\mathrm{Res}_{E/F}\,\boldsymbol{\theta}$, $\mathrm{Res}_{E/F}\,\boldsymbol{\nu_{p,q}}$ and $\mathrm{Res}_{E/F}\,\boldsymbol{\mu}$ on $\mathrm{Res}_{E/F}\,\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})$ by restriction of scalars. A straightforward computation shows that these maps commute with the morphism $\boldsymbol{\psi}$ whose fixed points in $\mathrm{Res}_{E/F}\,\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})$ define the group $\boldsymbol{G}'$ and can thus be restricted to $\boldsymbol{G}'$.

The fixed points of these morphisms define algebraic subgroups of $\boldsymbol{G}'$ whose F_v-rational points are certain Lie subgroups of $\boldsymbol{G}'(F_v)\cong\mathrm{SL}_n(\mathbb{R})$. We will now determine these subgroups. Recall the definition $\mathrm{GL}_r^{(1)}(\mathbb{C}):=\{g\in\mathrm{GL}_r(\mathbb{C}),\ |\det(g)|=1\}$ for a natural number r.

Proposition 5.3. *Let F be a totally real number field. For $\boldsymbol{G}'$ defined as above, we have $\boldsymbol{G}'(F_v)\cong\mathrm{SL}_n(\mathbb{R})$. The fixed points of the morphisms $\mathrm{Res}_{E/F}\,\boldsymbol{\theta}$, $\mathrm{Res}_{E/F}\,\boldsymbol{\nu_{p,q}}$, $\mathrm{Res}_{E/F}(\boldsymbol{\nu_{p,q}}\circ\boldsymbol{\theta})$, $\mathrm{Res}_{E/F}\,\boldsymbol{\mu}$ and $\mathrm{Res}_{E/F}(\boldsymbol{\mu}\circ\boldsymbol{\theta})$ define the following subgroups of $\mathrm{SL}_n(\mathbb{R})$:*

$$\begin{aligned}\mathrm{Fix}(\mathrm{Res}_{E/F}\,\boldsymbol{\theta},\boldsymbol{G}')(F_v)&\cong\mathrm{SO}(n),\\ \mathrm{Fix}(\mathrm{Res}_{E/F}\,\boldsymbol{\nu_{p,q}},\boldsymbol{G}')(F_v)&\cong\mathrm{SO}(p,q),\\ \mathrm{Fix}(\mathrm{Res}_{E/F}(\boldsymbol{\nu_{p,q}}\circ\boldsymbol{\theta}),\boldsymbol{G}')(F_v)&\cong S(\mathrm{GL}_p(\mathbb{R})\times\mathrm{GL}_q(\mathbb{R})),\\ \mathrm{Fix}(\mathrm{Res}_{E/F}\,\boldsymbol{\mu},\boldsymbol{G}')(F_v)&\cong\mathrm{Sp}(n,\mathbb{R}),\\ \mathrm{Fix}(\mathrm{Res}_{E/F}(\boldsymbol{\mu}\circ\boldsymbol{\theta}),\boldsymbol{G}')(F_v)&\cong\mathrm{GL}^{(1)}_{n/2}(\mathbb{C}).\end{aligned}$$

In particular, $\mathrm{Res}_{E/F}\,\boldsymbol{\theta}$ induces a Cartan involution on $\mathrm{SL}_n(\mathbb{R})$.

Proof. We start with a general observation. Let $\boldsymbol{\varphi}$ denote an E-rational morphism of $\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})$ such that $\operatorname{Res}_{E/F}\boldsymbol{\varphi}$ commutes with $\boldsymbol{\psi}$. Then $\boldsymbol{\varphi}$ can be restricted to $\boldsymbol{G}'$ and we have

$$\operatorname{Fix}(\operatorname{Res}_{E/F}\boldsymbol{\varphi}, \boldsymbol{G}') = \operatorname{Fix}(\operatorname{Res}_{E/F}\boldsymbol{\varphi}, \operatorname{Res}_{E/F}\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})) \cap \boldsymbol{G}'$$

as a subgroup of $\operatorname{Res}_{E/F}\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})$. At the F_v-rational points, there exists an isomorphism

$$\begin{aligned}\operatorname{Res}_{E/F}(\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(F_v) &\cong \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})(E \otimes_F F_v)\\ &\cong \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})(E_{w_1} \oplus E_{w_2}) \cong \mathrm{SL}_n(\mathbb{R}) \times \mathrm{SL}_n(\mathbb{R}).\end{aligned}$$

Moreover,

$$\begin{aligned}\operatorname{Fix}(\operatorname{Res}_{E/F}\boldsymbol{\varphi}, \operatorname{Res}_{E/F}\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(F_v)\\ &= \operatorname{Res}_{E/F}\operatorname{Fix}(\boldsymbol{\varphi}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(F_v)\\ &= \operatorname{Fix}(\boldsymbol{\varphi}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(E_{w_1}) \times \operatorname{Fix}(\boldsymbol{\varphi}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(E_{w_2})\\ &\subset \mathrm{SL}_n(\mathbb{R}) \times \mathrm{SL}_n(\mathbb{R}).\end{aligned}$$

On the other hand, the defining condition of $\boldsymbol{G}'$ identifies the two copies of $\mathrm{SL}_n(\mathbb{R})$ (see Proposition 4.1). Thus we can restrict to one of the components (we choose the first one without loss of generality) and get

$$\begin{aligned}\operatorname{Fix}(\operatorname{Res}_{E/F}\boldsymbol{\varphi}, \boldsymbol{G}')(F_v) &= \operatorname{Fix}(\operatorname{Res}_{E/F}\boldsymbol{\varphi}, \operatorname{Res}_{E/F}\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(F_v) \cap \boldsymbol{G}'(F_v)\\ &\cong \operatorname{Fix}(\boldsymbol{\varphi}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(E_{w_1}) \subset \mathrm{SL}_n(\mathbb{R}).\end{aligned}$$

Let us now specify $\boldsymbol{\varphi}$ to be one of the above morphisms. For $\boldsymbol{\varphi} = \boldsymbol{\theta}$ or $\boldsymbol{\varphi} = \boldsymbol{\nu}_{\boldsymbol{p},\boldsymbol{q}}$, the group $\operatorname{Fix}(\boldsymbol{\varphi}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))$ is a special unitary group with respect to a hermitian form and an orthogonal involution. Therefore, by Proposition 4.1, we have $\operatorname{Fix}(\boldsymbol{\varphi}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(E_{w_1}) \cong \mathrm{SO}(p', q')$ for suitable $p', q' \in \mathbb{N}$. Since H is positive definite at the place v, it does not influence the signature (p', q'). When $D = E$ it is clear from the definitions of θ and $\nu_{p,q}$ that the signature comes from the matrices $I_{p,q}$. When $D = Q(a, b \mid E)$, we can conclude from Lemma 5.1 that, under a suitable splitting at the place w_1, the involution τ^t is mapped to $x \mapsto x^t$ on $M_n(\mathbb{R})$ and $\tau^t_{(1,-1)}$ is mapped to $x \mapsto \operatorname{Int}(I_{n/2,n/2})(x^t)$. Moreover, the matrices $I_{p/2,q/2}$ are mapped to $I_{p,q}$ at the place w_1. Therefore, for both choices of D, we get

$$\operatorname{Fix}(\operatorname{Res}_{E/F}(\boldsymbol{\theta}), \boldsymbol{G}')(F_v) = \operatorname{Fix}(\boldsymbol{\theta}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(E_{w_1}) \cong \mathrm{SO}(n)$$

and

$$\operatorname{Fix}(\operatorname{Res}_{E/F}(\boldsymbol{\nu}_{\boldsymbol{p},\boldsymbol{q}}), \boldsymbol{G}')(F_v) = \operatorname{Fix}(\boldsymbol{\nu}_{\boldsymbol{p},\boldsymbol{q}}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(E_{w_1}) \cong \mathrm{SO}(p, q).$$

In particular, $\mathrm{Res}_{E/F}\,\boldsymbol{\theta}$ induces a Cartan involution on $\mathrm{SL}_n(\mathbb{R})$, as the group of F_v-rational points of its fixed point group is isomorphic to $\mathrm{SO}(n)$, a maximal compact subgroup of $\mathrm{SL}_n(\mathbb{R})$.

For $\boldsymbol{\varphi} = \boldsymbol{\nu_{p,q}} \circ \boldsymbol{\theta}$ one can easily see from the definition of $\nu_{p,q}$ that

$$\boldsymbol{\nu_{p,q}} \circ \boldsymbol{\theta} = \begin{cases} \mathrm{Int}(I_{p,q}) & \text{if } D = E, \\ \mathrm{Int}(I_{p/2,q/2}) & \text{if } D = Q(a,b \mid E) \text{ and } p,q \text{ even}, \\ \mathrm{Int}(\mathrm{diag}(e_2,\dots,e_2)) & \text{if } D = Q(a,b \mid E) \text{ and } p = q = n/2. \end{cases}$$

Here, we use the notation of Lemma 5.1 and its remark for the statement in the last line. Under a suitable splitting isomorphism at the place w_1 of E, these morphisms go over to $\mathrm{Int}(I_{p,q})$ on $\mathrm{SL}_n(\mathbb{R})$ (to see this in the third case, use the isomorphism given in Lemma 5.1). Therefore, we have

$$\begin{aligned} \mathrm{Fix}(\mathrm{Res}_{E/F}(\boldsymbol{\nu_{p,q}} \circ \boldsymbol{\theta}), \boldsymbol{G}')(F_v) &= \mathrm{Fix}(\boldsymbol{\nu_{p,q}} \circ \boldsymbol{\theta}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(E_{w_1}) \\ &\cong \{x \in \mathrm{SL}_n(\mathbb{R}) \mid I_{p,q} x I_{p,q} = x\} \\ &\cong S(\mathrm{GL}_p(\mathbb{R}) \times \mathrm{GL}_q(\mathbb{R})). \end{aligned}$$

Let now $\boldsymbol{\varphi} = \boldsymbol{\mu}$. The group $\mathrm{Fix}(\boldsymbol{\varphi}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))$ is either a special unitary group with respect to a skew-hermitian form and an orthogonal involution (when $D = E$, the matrix $H^{-1} J_m$ occurring in the definition of μ is skew-symmetric and hence it describes a skew-hermitian form over E) or a special unitary group with respect to a hermitian form and a symplectic involution (when D is a quaternion algebra, H is a diagonal matrix with entries in F and thus τ_c-invariant). In both cases, Proposition 4.1 implies

$$\mathrm{Fix}(\mathrm{Res}_{E/F}(\boldsymbol{\mu}), \boldsymbol{G}')(F_v) = \mathrm{Fix}(\boldsymbol{\mu}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(E_{w_1}) \cong \mathrm{Sp}(n,\mathbb{R}).$$

Finally, we consider $\boldsymbol{\varphi} = \boldsymbol{\mu} \circ \boldsymbol{\theta}$. If $D = E$, we have $(\boldsymbol{\mu} \circ \boldsymbol{\theta})(E) = \mathrm{Int}\, J_m$. In the case $D = Q(a,b \mid E)$, we note that $(\boldsymbol{\mu} \circ \boldsymbol{\theta})(E) = \tau_c \circ \tau = \mathrm{Int}\,\mathrm{diag}(e_0,\dots,e_0)$ on $M_m(D)$, by the remark following Lemma 5.1. Under a suitable splitting isomorphism, $\mathrm{diag}(e_0,\dots,e_0)$ is mapped to J_n at the place w_1 of D (see Lemma 5.1). Therefore, for both choices of D, we have

$$\begin{aligned} \mathrm{Fix}(\mathrm{Res}_{E/F}(\boldsymbol{\mu} \circ \boldsymbol{\theta}), \boldsymbol{G}')(F_v) &= \mathrm{Fix}(\boldsymbol{\mu} \circ \boldsymbol{\theta}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(E_{w_1}) \\ &\cong \{x \in \mathrm{SL}_n(\mathbb{R}) \mid J_n x J_n^{-1} = x\} \\ &\cong \mathrm{GL}^{(1)}_{n/2}(\mathbb{C}). \end{aligned}$$

□

5.4. Now that we have defined certain F-rational morphisms on $\boldsymbol{G}'$ and studied their fixed point groups, we are ready to define the corresponding special cycles. To do this, we pass to the algebraic $\mathbb{Q}$-group $\boldsymbol{G} := \operatorname{Res}_{F/\mathbb{Q}} \boldsymbol{G}'$. We have $\boldsymbol{G}(\mathbb{Q}) \cong \boldsymbol{G}'(F)$ and

$$\boldsymbol{G}(\mathbb{R}) \cong \boldsymbol{G}'(\mathbb{R} \otimes_{\mathbb{Q}} F) = \prod_{v' \in V_\infty(F)} \boldsymbol{G}'(F_{v'}) = \mathrm{SL}_n(\mathbb{R}) \times \prod_{\substack{v' \in V_\infty \\ v' \neq v}} \mathrm{SU}(n).$$

Moreover, there exist $\mathbb{Q}$-rational morphisms $\boldsymbol{\theta}$, $\boldsymbol{\nu_{p,q}}$ and $\boldsymbol{\mu}$ of order two on $\boldsymbol{G}$ that are induced from the corresponding morphisms of $\boldsymbol{G}'$. Let $\Gamma \subset \boldsymbol{G}(\mathbb{Q})$ be an arithmetic subgroup of $\boldsymbol{G}$. The image of Γ under the isomorphism $\boldsymbol{G}(\mathbb{Q}) \cong \boldsymbol{G}'(F)$ is an arithmetic subgroup of $\boldsymbol{G}'$ and thus it gives rise to a discrete cocompact subgroup of $\mathrm{SL}_n(\mathbb{R})$ that we will still denote by Γ for simplicity of notation.[4] Let K' denote a maximal compact subgroup of $\boldsymbol{G}(\mathbb{R})$ and $X := K' \backslash \boldsymbol{G}(\mathbb{R})$ the symmetric space attached to $\boldsymbol{G}(\mathbb{R})$. Since $\boldsymbol{G}(\mathbb{R})$ is a product of $\mathrm{SL}_n(\mathbb{R})$ and compact factors, X is isomorphic to $\mathrm{SO}(n) \backslash \mathrm{SL}_n(\mathbb{R})$ and Γ acts on X by right translations. Note that X is a symmetric space of dimension

$$\dim X = \dim(\mathrm{SL}_n(\mathbb{R})) - \dim(\mathrm{SO}(n)) = n^2 - 1 - \tfrac{1}{2}n(n-1) = \tfrac{1}{2}n(n+1) - 1.$$

Let now $\Gamma \subset \boldsymbol{G}(\mathbb{Q})$ be a torsion-free arithmetic subgroup of $\boldsymbol{G}$ and assume that Γ and K' are invariant under the morphisms $\boldsymbol{\theta}$, $\boldsymbol{\nu_{p,q}}$ and $\boldsymbol{\mu}$.[5] Then these morphisms induce certain special geometric cycles in X/Γ, as explained in Section 3.1.

Theorem 5.5. *The morphisms $\boldsymbol{\nu_{p,q}}$ and $\boldsymbol{\nu_{p,q}} \circ \boldsymbol{\theta}$ of $\boldsymbol{G}$ induce a family of pairs of special geometric cycles $C(\boldsymbol{\nu_{p,q}})$, $C(\boldsymbol{\nu_{p,q}} \circ \boldsymbol{\theta})$ in X/Γ, for positive integers p and q with $p+q=n$ if $\boldsymbol{G}$ comes from a special unitary group over an algebraic number field, and for positive integers p and q with $p+q=n$ and p and q even or $p=q=n/2$ if $\boldsymbol{G}$ comes from a special unitary group over a quaternion algebra. If n is even and $n>2$, the morphisms $\boldsymbol{\mu}$ and $\boldsymbol{\mu} \circ \boldsymbol{\theta}$ induce a pair of geometric cycles $C(\boldsymbol{\mu})$, $C(\boldsymbol{\mu} \circ \boldsymbol{\theta})$ in X/Γ. Some properties of these cycles are summarized in Table 1.*

Proof. The existence of the cycles is clear from Section 3.1. The isomorphisms in the second column of Table 1 follow from Proposition 5.3 and the fact that $\boldsymbol{G}(\mathbb{R}) \cong \boldsymbol{G}'(F_v)$ up to compact factors. The dimensions of the cycles can be computed as the dimensions of the associated symmetric spaces, using the dimensions of the occurring real Lie groups and their maximal compact subgroups (for a list of dimensions of classical Lie groups, see, e.g., [Helgason 1978, Table IV, p. 516]). Note that both $\mathrm{SO}(p,q)$ and $S(\mathrm{GL}_p(\mathbb{R}) \times \mathrm{GL}_q(\mathbb{R}))$ have maximal compact

[4] To be precise, the arithmetic subgroup $\Gamma \subset \boldsymbol{G}(\mathbb{Q})$ can be regarded as a subgroup of $\boldsymbol{G}(\mathbb{R}) = \mathrm{SL}_n(\mathbb{R}) \times$ compact factors, and the discrete cocompact subgroup is in fact the projection of Γ to the noncompact factor of $\boldsymbol{G}(\mathbb{R})$.

[5] Such a choice of K' is possible by [Helgason 1978, Theorem 13.5]

$C = C(\boldsymbol{\varphi})$	$\mathrm{Fix}(\boldsymbol{\varphi}, \boldsymbol{G})(\mathbb{R}) \cong$	$\dim C$
$C(\boldsymbol{\nu_{p,q}})$	$\mathrm{SO}(p,q)$	pq
$C(\boldsymbol{\nu_{p,q}} \circ \boldsymbol{\theta})$	$S(\mathrm{GL}_p(\mathbb{R}) \times \mathrm{GL}_q(\mathbb{R}))$	$\frac{1}{2}(p^2+q^2+n)-1$
$C(\boldsymbol{\mu})$	$\mathrm{Sp}(n, \mathbb{R})$	$\frac{1}{4}(n^2+2n)$
$C(\boldsymbol{\mu} \circ \boldsymbol{\theta})$	$\mathrm{GL}^{(1)}_{n/2}(\mathbb{C})$	$\frac{1}{4}n^2-1$

Table 1. Geometric cycles in $\mathrm{SO}(n)\backslash \mathrm{SL}_n(\mathbb{R})/\Gamma$: the isomorphism in the second column is up to compact factors and the lower half of the table is only applicable if n is even and $n > 2$.

subgroup $S(O(p) \times O(q))$ and that both $\mathrm{Sp}(n, \mathbb{R})$ and $\mathrm{GL}^{(1)}_{n/2}(\mathbb{C})$ have maximal compact subgroup isomorphic to $U(n/2)$.[6] □

Finally, we can apply Theorem 3.2 to the constructed cycles to obtain a nonvanishing result for the cohomology of X/Γ. As before, we denote by $X_u \cong \mathrm{SO}(n)\backslash \mathrm{SU}(n)$ the compact dual symmetric space of X.

Theorem 5.6. *Let $n \in \mathbb{N}$ be even.*

(1) *There exists a cocompact discrete subgroup Γ_1 of $\mathrm{SL}_n(\mathbb{R})$ that arises from an arithmetic subgroup of a special unitary group over an algebraic number field, such that $H^k(X/\Gamma_1, \mathbb{C})$ contains nontrivial cohomology classes for*

$$k = pq \quad \text{and} \quad k = \tfrac{1}{2}(p^2+q^2+n)-1,$$

where p and q are positive integers with $p+q=n$, and, if $n \neq 2$, for

$$k = \tfrac{1}{4}(n^2+2n) \quad \text{and} \quad k = \tfrac{1}{4}n^2-1.$$

(2) *There exists a cocompact discrete subgroup Γ_2 of $\mathrm{SL}_n(\mathbb{R})$ that arises from an arithmetic subgroup of a special unitary group over a quaternion algebra, such that $H^k(X/\Gamma_2, \mathbb{C})$ contains nontrivial cohomology classes for*

$$k = pq \quad \text{and} \quad k = \tfrac{1}{2}(p^2+q^2+n)-1,$$

where p and q are positive, even integers with $p+q=n$ or $p=q=n/2$, and, if $n \neq 2$, for

$$k = \tfrac{1}{4}(n^2+2n) \quad \text{and} \quad k = \tfrac{1}{4}n^2-1.$$

In both cases, these classes are not in the image of the respective injective map

$$\beta^*_{\Gamma_i} : H^*(X_u, \mathbb{C}) \to H^*(X/\Gamma_i, \mathbb{C}),$$

i.e., they are not represented by $\mathrm{SL}_n(\mathbb{R})$-invariant forms on X.

[6] Here $U(n/2)$ is considered as a subgroup of $\mathrm{Sp}(n)$ via the embedding $\phi : \mathrm{GL}_{n/2}(\mathbb{C}) \to \mathrm{GL}_n(\mathbb{R})$, $X = A + iB \mapsto \left(\begin{smallmatrix} A & -B \\ B & A \end{smallmatrix}\right)$, for $A, B \in \mathrm{GL}_{n/2}(\mathbb{R})$.

Proof. We give a detailed proof of (1), then (2) follows analogously.

Let $\boldsymbol{G}$ denote the algebraic $\mathbb{Q}$-group whose real points are isomorphic to $\mathrm{SL}_n(\mathbb{R})$ up to compact factors and which is defined via a special unitary group over an algebraic number field. Set $\Psi := \{\boldsymbol{\nu_{p,q}} \mid p+q=n,\ p \neq 0 \neq q\} \cup \{\boldsymbol{\mu}\}$ and let Γ be a torsion-free arithmetic subgroup of $\boldsymbol{G}$ that is stable under the group generated by $\Psi \cup \{\boldsymbol{\theta}\}$. If we choose $\boldsymbol{\tau}_1 \in \Psi$ and set $\boldsymbol{\tau}_2 := \boldsymbol{\tau}_1 \circ \boldsymbol{\theta}$, the pair $(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ is a pair of commuting morphisms of order two and it defines a pair of geometric cycles on X/Γ, whose properties are given in Theorem 5.5. As discussed in Section 3 we may assume that the cycles and the connected components of their intersection are orientable. Moreover, it follows from Table 1 and the fact that $\boldsymbol{\tau}_1 \circ \boldsymbol{\tau}_2 = \boldsymbol{\theta}$ induces a Cartan involution on $\mathrm{SL}_n(\mathbb{R})$ that the cycles $C(\boldsymbol{\tau}_1)$ and $C(\boldsymbol{\tau}_2)$ are of complementary dimension and satisfy condition (ii) of Theorem 3.2.

To apply Theorem 3.2, it remains to check condition (i). It suffices to look at the action of the noncompact factor of $\mathrm{Fix}(\boldsymbol{\tau}_1, \boldsymbol{G})(\mathbb{R})$ and $\mathrm{Fix}(\boldsymbol{\tau}_2, \boldsymbol{G})(\mathbb{R})$ on the respective symmetric space.

For $\boldsymbol{\tau}_1 = \boldsymbol{\mu}$, we have $\mathrm{Fix}(\boldsymbol{\tau}_1, \boldsymbol{G})(\mathbb{R}) \cong \mathrm{Sp}(n, \mathbb{R})$ and $\mathrm{Fix}(\boldsymbol{\tau}_2, \boldsymbol{G})(\mathbb{R}) \cong \mathrm{GL}^{(1)}_{n/2}(\mathbb{C})$ up to compact factors (see Table 1). These are connected Lie subgroups of $\mathrm{SL}_n(\mathbb{R})$ and hence they act orientation-preservingly on the respective symmetric spaces. For $\boldsymbol{\tau}_1 = \boldsymbol{\nu_{p,q}}$, we have $\mathrm{Fix}(\boldsymbol{\tau}_1, \boldsymbol{G})(\mathbb{R}) \cong \mathrm{SO}(p, q)$ and $\mathrm{Fix}(\boldsymbol{\tau}_2, \boldsymbol{G})(\mathbb{R}) \cong S(\mathrm{GL}_p(\mathbb{R}) \times \mathrm{GL}_q(\mathbb{R}))$ up to compact factors, where $p+q=n$. Since n is even, it follows from Example 3.4 that these groups act orientation-preservingly on their associated symmetric spaces.

We conclude that, for any choice of $\boldsymbol{\tau}_1 \in \Psi$, the pair $(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ meets all assumptions of Theorem 3.2. Therefore, for each such $\boldsymbol{\tau}_1$, we can find a normal, $\langle \boldsymbol{\tau}_1, \boldsymbol{\tau}_2 \rangle$-stable subgroup $\Gamma_{\tau_1} \subset \Gamma$ of finite index such that $H^k(X/\Gamma_{\tau_1}, \mathbb{C}) \neq 0$ for $k \in \{\dim C(\boldsymbol{\tau}_1), \dim C(\boldsymbol{\tau}_2)\}$. Moreover, Γ_{τ_1} can be chosen such that the nontrivial cohomology classes detected by $C(\boldsymbol{\tau}_1)$ and $C(\boldsymbol{\tau}_2)$ are not represented by $\mathrm{SL}_n(\mathbb{R})$-invariant differential forms on X, as follows from Theorem 3.3. Set

$$\Gamma' := \bigcap_{\boldsymbol{\tau}_1 \in \Psi} \Gamma_{\tau_1} \quad \text{and} \quad \Gamma_1 := \bigcap_{\boldsymbol{\tau}_1 \in \Psi} \boldsymbol{\tau}_1(\Gamma') \cap \boldsymbol{\tau}_2(\Gamma').$$

Then Γ_1 is a cocompact discrete subgroup of $\mathrm{SL}_n(\mathbb{R})$ that is of finite index in each Γ_{τ_1} and $\langle \boldsymbol{\tau}_1, \boldsymbol{\tau}_2 \rangle$-stable for each $\boldsymbol{\tau}_1 \in \Psi$. The group Γ_1 admits nontrivial cohomology classes in all degrees $k \in \{\dim C(\boldsymbol{\tau}_1), \dim C(\boldsymbol{\tau}_2)\}$ for possible pairs $(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ with $\boldsymbol{\tau}_1 \in \Psi$, and these classes are not represented by $\mathrm{SL}_n(\mathbb{R})$-invariant differential forms on X.[7] The exact dimensions can be read off from Table 1. □

Remark. (1) We do not get any result in the case where n is odd. The morphism $\boldsymbol{\mu}$ is not defined in this case, so we are left with the cycles $C(\boldsymbol{\nu_{p,q}})$ and $C(\boldsymbol{\nu_{p,q}} \circ \boldsymbol{\theta})$.

[7] Here we use the fact that the results of Theorems 3.2 and 3.3 carry over to finite index subgroups of Γ.

	$\dim X/\Gamma$	Cycle	Subgroup of $\mathrm{SL}_n(\mathbb{R})$	Contributing to degree	Occurs for $\Gamma=\Gamma_1$	$\Gamma=\Gamma_2$
$n=2$	2	$C(\boldsymbol{\nu}_{1,1})$	$\mathrm{SO}(1,1)$	1	$\times$	$\times$
		$C(\boldsymbol{\nu}_{1,1}\circ\boldsymbol{\theta})$	$S(\mathrm{GL}_1\times\mathrm{GL}_1)$	1	$\times$	$\times$
$n=4$	9	$C(\boldsymbol{\nu}_{1,3}\circ\boldsymbol{\theta})$	$S(\mathrm{GL}_1\times\mathrm{GL}_3)$	3	$\times$	
		$C(\boldsymbol{\mu})$	$\mathrm{Sp}(4,\mathbb{R})$	3	$\times$	$\times$
		$C(\boldsymbol{\nu}_{2,2}\circ\boldsymbol{\theta})$	$S(\mathrm{GL}_2\times\mathrm{GL}_2)$	4	$\times$	$\times$
		$C(\boldsymbol{\nu}_{2,2})$	$\mathrm{SO}(2,2)$	5	$\times$	$\times$
		$C(\boldsymbol{\mu}\circ\boldsymbol{\theta})$	$\mathrm{GL}_2^{(1)}(\mathbb{C})$	6	$\times$	$\times$
		$C(\boldsymbol{\nu}_{1,3})$	$\mathrm{SO}(1,3)$	6	$\times$	

Table 2. Real case: degrees in $H^*(X/\Gamma)$ in which we have nontrivial cohomology classes coming from special cycles.

These are indeed of complementary dimension and $\mathrm{Fix}(\langle \boldsymbol{\nu}_{p,q}, \boldsymbol{\nu}_{p,q}\circ\boldsymbol{\theta}\rangle, \boldsymbol{G})(\mathbb{R})$ is compact. However, the cycles do not satisfy condition (i) in Theorem 3.2 (see Example 3.4). Therefore, the result of Rohlfs and Schwermer is not applicable and we cannot deduce any statement about the intersection number of $C(\boldsymbol{\nu}_{p,q})$ and $C(\boldsymbol{\nu}_{p,q}\circ\boldsymbol{\theta})$. It is still an open question whether or not this number is nontrivial.

(2) Ash and Ginzburg [1994] show a part of our result in Theorem 5.6(1) to use it in the proof of their Lemma 5.4.2. More precisely, in the case where n is even and $\boldsymbol{G}$ is the algebraic group associated with a special unitary group over a number field, they construct the pair of special cycles $C(\boldsymbol{\nu}_{n/2,n/2})$, $C(\boldsymbol{\nu}_{n/2,n/2}\circ\boldsymbol{\theta})$ (in our notation). Then, using the result of Rohlfs and Schwermer, they show that the intersection number of these cycles is nonzero and deduce the existence of a nonvanishing homology class.

Example 5.7. Let Γ be a cocompact discrete subgroup of $\mathrm{SL}_n(\mathbb{R})$ chosen as in Theorem 5.6(1) or (2). In Table 2 we give an overview of the occurring cycles, the associated subgroups of $\mathrm{SL}_n(\mathbb{R})$ and the degrees in the cohomology of X/Γ to which these cycles contribute,[8] for some choices of n. The last two columns indicate if the respective cycle exists for the choice of Γ as in Theorem 5.6(1) or (2).

Remark. Using the method of crosswise intersection (see [Waldner 2010]), one can show that for $n=2$ the two cohomology classes contributing to degree 1 are linearly independent. Unfortunately, for $n>2$ we do not get a result on the linear independence of the constructed cohomology classes using this technique.

[8]Note that these degrees are not the dimensions of the cycles but their complements, since we are looking at the cohomology classes obtained via Poincaré duality.

6. Geometric cycles for $\mathrm{SL}_n(\mathbb{C})$

We will work in the following general setting. Let F_0 be a totally real number field and let E/F_0 be a totally complex biquadratic extension.[9] Assume that there is a quadratic extension F/F_0 such that F is a subfield of E with exactly one complex place v and denote by v_0 the real place of F_0 with $v \mid v_0$. Moreover, let L and L' denote the other two intermediate fields of the extension E/F_0. Then L is a quadratic extension of F_0 that has two real places v_1, v_2 lying above v_0 and only complex archimedean places otherwise, and L' is a totally complex quadratic extension of F_0. This is to say, we have $F = F_0(\sqrt{D_1})$, $L = F_0(\sqrt{D_2})$ and $L' = F_0(\sqrt{D_1 D_2})$ for $D_1, D_2 \in F_0$ such that none of D_1, D_2 and $D_1 D_2$ is a square in F_0 and such that $(D_1)_{v_0} < 0$, $(D_2)_{v_0} > 0$ and $(D_1)_{v'} > 0 > (D_2)_{v'}$ for $v' \in V_\infty(F_0)$ and $v' \neq v_0$. We write ι and ω for the nontrivial Galois automorphisms of E/F and E/L, respectively. Then ι and ω generate the Galois group of E/F_0 and the third nontrivial element $\iota \circ \omega = \omega \circ \iota$ is the nontrivial Galois automorphism of E over L'. Moreover, we note that $\iota|_L$ is the nontrivial Galois automorphism of the quadratic extension L/F_0, $\omega|_F$ is the one of F/F_0 and $\iota|_{L'} = \omega|_{L'}$ is the one of L'/F_0. The field extension E/F_0, its intermediate subfields and the nontrivial Galois automorphisms corresponding to each extension are illustrated in the following diagram:

$$\begin{tikzcd}
 & E \arrow[dl, dash, "\iota"'] \arrow[d, dash, "\omega"] \arrow[dr, dash, "\iota\circ\omega=\omega\circ\iota"] & \\
F \arrow[dr, dash, "\omega|_F"'] & L \arrow[d, dash, "\iota|_L"] & L' \arrow[dl, dash, "\omega|_{L'}=\iota|_{L'}"] \\
 & F_0 &
\end{tikzcd}$$

Now we let D be a division algebra of degree d over E with an involution σ of the second kind. Again we will restrict to the cases where D is either the field E itself and $\sigma = \iota$ or D is a quaternion division algebra over E constructed in the following way. Let D' over L' be a quaternion division algebra that does not split over E and that admits an involution γ of the second kind (with respect to the subfield $F_0 \subset L'$). Without loss of generality, we may assume that γ is definite. Moreover, by Albert's theorem, we find a quaternion division algebra $D_0 = Q(a, b \mid F_0)$ over F_0 with $a, b \in F_0^\times$ such that $(D', \gamma) \cong (D_0 \otimes_{F_0} L', \tau_{c,0} \otimes \omega|_{L'})$, where $\tau_{c,0}$ denotes the conjugation on D_0. Now set $D := D_0 \otimes_{F_0} E = D' \otimes_{L'} E$. By our choice of D', this is a quaternion division algebra over E that admits the two involutions $\sigma := \tau_{c,0} \otimes \iota$ and $\sigma' := \tau_{c,0} \otimes \omega$, both of the second kind. Note that σ is trivial on the subfield

[9] A *biquadratic extension* of a number field F_0 is an extension of degree 4 with Galois group $\mathrm{Gal}(E/F_0) \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

F of E, and σ' is trivial on the subfield L of E. Moreover, it follows from the choice of γ that both involutions are definite. For simplicity of notation, we will set $D_0 := F_0,\ D' = L',\ \gamma := \omega|_{L'} = \iota|_{L'}$ and $\sigma' := \omega$ when $D = E$.

Let $m \in \mathbb{N}$ be arbitrary and set $n := 2m$. Then D, E, F, σ and m satisfy the conditions of Theorem 4.2(2), and we can choose a hermitian form h on D^m such that each arithmetic subgroup of $\boldsymbol{G}' := \mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}, \boldsymbol{D}, \boldsymbol{\sigma})$ gives rise to a cocompact discrete subgroup of $\mathrm{SL}_n(\mathbb{C})$. As above, h can be chosen such that its diagonal representation H is an element of $M_m(F_0)$ and we will restrict to this situation for technical reasons. Moreover, we suppose that H is positive definite under the embedding corresponding to the real place v_0 of F_0.

We can now define some automorphisms of order two on $\mathrm{SL}_m(D)$:

$$\theta : \mathrm{SL}_m(D) \to \mathrm{SL}_m(D),\qquad \theta(x) = H^{-1}(\sigma'(x)^t)^{-1}H,$$

$$\eta : \mathrm{SL}_m(D) \to \mathrm{SL}_m(D),\qquad \eta(x) = \begin{cases} H^{-1}(x^t)^{-1}H & \text{if } D = E,\\ H^{-1}(\tau_k(x)^t)^{-1}H & \text{if } D = Q(a, b \mid E),\end{cases}$$

$$\mu : \mathrm{SL}_m(D) \to \mathrm{SL}_m(D),\qquad \mu(x) = \begin{cases} H^{-1}J_m(x^t)^{-1}J_m^{-1}H & \text{if } D = E \text{ and } m > 2 \text{ even},\\ H^{-1}(\tau_c(x)^t)^{-1}H & \text{if } D = Q(a, b \mid E) \text{ and } m > 1.\end{cases}$$

Moreover, for certain positive integers p and q such that $p + q = n$, we define a family of automorphisms $\nu_{p,q} : \mathrm{SL}_m(D) \to \mathrm{SL}_m(D)$ by

$$\nu_{p,q}(x) = \begin{cases} H^{-1}I_{p,q}(\sigma'(x)^t)^{-1}I_{p,q}H & \text{if } D = E,\\ H^{-1}I_{p/2,q/2}(\sigma'(x)^t)^{-1}I_{p/2,q/2}H & \text{if } D = Q(a, b \mid E) \text{ and } p, q \text{ even}.\end{cases}$$

Again, θ commutes with each of the other automorphisms.

6.1. As in Section 5.2, we will assume from now on that p and q are positive integers such that $p + q = n$ and that p and q are both even, whenever we deal with the case where D is a quaternion algebra. Statements involving the map μ will again only be applicable if n is even and $n > 2$.

The maps η and μ are built out of E-linear maps (involutions of the first kind on $\mathrm{SL}_m(D)$) and the group inversion, and hence they define E-rational morphisms $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ on the algebraic E-group $\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})$, as expected. However, the maps θ and $\nu_{p,q}$ involve involutions of the second kind with respect to the subfield L of E, and therefore they only define L-rational morphisms $\boldsymbol{\theta}$ and $\boldsymbol{\nu_{p,q}}$ of the algebraic L-group $\mathrm{Res}_{E/L}\,\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})$. On the other hand, the morphism $\boldsymbol{\psi}$ defining the algebraic group $\boldsymbol{G}'$ is an F-rational morphism of the group $\mathrm{Res}_{E/F}\,\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})$. To work with all of these morphisms simultaneously, we need to pass to an algebraic group over the common subfield F_0 of F, L and E.[10] Using restriction of scalars with respect to the field F_0, the morphisms $\boldsymbol{\eta}$, $\boldsymbol{\mu}$, $\boldsymbol{\theta}$ and $\boldsymbol{\nu_{p,q}}$ give rise to F_0-rational morphisms on $\mathrm{Res}_{E/F_0}\,\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})$. An easy computation shows that these morphisms commute with $\mathrm{Res}_{F/F_0}\,\boldsymbol{\psi}$ and can thus be restricted to the group $\boldsymbol{G}'' := \mathrm{Res}_{F/F_0}\,\boldsymbol{G}' = \mathrm{Fix}(\mathrm{Res}_{F/F_0}\,\boldsymbol{\psi}, \mathrm{Res}_{E/F_0}\,\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))$.

Their fixed points define algebraic subgroups of $\boldsymbol{G}''$ whose F_{0,v_0}-rational points are certain Lie subgroups of $\mathrm{SL}_n(\mathbb{C})$. We now determine these subgroups.

Proposition 6.2. *The algebraic F_0-group $\boldsymbol{G}''$ satisfies $\boldsymbol{G}''(F_{0,v_0}) \cong \mathrm{SL}_n(\mathbb{C})$. The fixed points of (certain compositions of) the above-defined F_0-rational morphisms define the following subgroups of $\mathrm{SL}_n(\mathbb{C})$:*

$$\begin{aligned}
\mathrm{Fix}(\mathrm{Res}_{L/F_0}\,\boldsymbol{\theta}, \boldsymbol{G}'')(F_{0,v_0}) &\cong \mathrm{SU}(n),\\
\mathrm{Fix}(\mathrm{Res}_{E/F_0}\,\boldsymbol{\eta}, \boldsymbol{G}'')(F_{0,v_0}) &\cong \mathrm{SO}(n,\mathbb{C}),\\
\mathrm{Fix}(\mathrm{Res}_{E/F_0}\,\boldsymbol{\eta}\circ\mathrm{Res}_{L/F_0}\,\boldsymbol{\theta}, \boldsymbol{G}'')(F_{0,v_0}) &\cong \mathrm{SL}(n,\mathbb{R}),\\
\mathrm{Fix}(\mathrm{Res}_{E/F_0}\,\boldsymbol{\mu}, \boldsymbol{G}'')(F_{0,v_0}) &\cong \mathrm{Sp}(n,\mathbb{C}),\\
\mathrm{Fix}(\mathrm{Res}_{E/F_0}\,\boldsymbol{\mu}\circ\mathrm{Res}_{L/F_0}\,\boldsymbol{\theta}, \boldsymbol{G}'')(F_{0,v_0}) &\cong \mathrm{SU}^*(n),\\
\mathrm{Fix}(\mathrm{Res}_{L/F_0}\,\boldsymbol{\nu_{p,q}}, \boldsymbol{G}'')(F_{0,v_0}) &\cong \mathrm{SU}(p,q),\\
\mathrm{Fix}(\mathrm{Res}_{L/F_0}(\boldsymbol{\nu_{p,q}}\circ\boldsymbol{\theta}), \boldsymbol{G}'')(F_{0,v_0}) &\cong S(\mathrm{GL}_p(\mathbb{C})\times\mathrm{GL}_q(\mathbb{C})),\\
\mathrm{Fix}(\mathrm{Res}_{E/F_0}(\boldsymbol{\eta}\circ\boldsymbol{\mu}), \boldsymbol{G}'')(F_{0,v_0}) &\cong S(\mathrm{GL}_{n/2}(\mathbb{C})\times\mathrm{GL}_{n/2}(\mathbb{C})),\\
\mathrm{Fix}(\mathrm{Res}_{E/F_0}(\boldsymbol{\eta}\circ\boldsymbol{\mu})\circ\mathrm{Res}_{L/F_0}\,\boldsymbol{\theta}, \boldsymbol{G}'')(F_{0,v_0}) &\cong \mathrm{SU}(n/2,n/2).
\end{aligned}$$

In particular, $\mathrm{Res}_{L/F_0}\,\boldsymbol{\theta}$ induces a Cartan involution on $\mathrm{SL}_n(\mathbb{C})$.

Proof. We have $\boldsymbol{G}''(F_{0,v_0}) = \boldsymbol{G}'(F_v) \cong \mathrm{SL}_n(\mathbb{C})$ by construction of the algebraic group $\boldsymbol{G}'$. To determine the fixed points of the morphisms, we need to study each map separately. We start with the morphism $\boldsymbol{\theta}$. The F_0-group $\mathrm{Fix}(\mathrm{Res}_{L/F_0}\,\boldsymbol{\theta}, \boldsymbol{G}'')$ is defined by the equations $\theta(x) = x = \psi(x)$ on $\mathrm{SL}_m(D) = \mathrm{Res}_{E/F_0}\,\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})(F_0)$. We have

$$\begin{aligned}
&\mathrm{Fix}(\mathrm{Res}_{L/F_0}\,\boldsymbol{\theta}, \boldsymbol{G}'')(F_0)\\
&\qquad = \{x \in \mathrm{SL}_m(D) \mid \theta(x) = x = \psi(x)\}\\
&\qquad = \{x \in \mathrm{SL}_m(D) \mid H^{-1}(\sigma'(x)^t)^{-1}H = x = H^{-1}(\sigma(x)^t)^{-1}H\}\\
&\qquad = \{x \in \mathrm{SL}_m(D) \mid (\sigma'\circ\sigma)(x) = x \ \text{ and } \ x = H^{-1}(\sigma(x)^t)^{-1}H\}\\
&\qquad = \{x \in \mathrm{SL}_m(D') \mid x = H^{-1}(\gamma(x)^t)^{-1}H\},
\end{aligned}$$

[10] The reason for this additional complication is that we want the map $\boldsymbol{\theta}$ to define a Cartan involution of $\mathrm{SL}_n(\mathbb{C})$. Unlike in the real case, this involves complex conjugation and can hence not be defined by an E-rational morphism.

which yields $\mathrm{Fix}(\mathrm{Res}_{L/F_0}\boldsymbol{\theta}, \boldsymbol{G}'') = \mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}|_{\boldsymbol{D}'}, \boldsymbol{D}', \boldsymbol{\gamma})$. Now Proposition 4.1 implies

$$\mathrm{Fix}(\mathrm{Res}_{L/F_0}\boldsymbol{\theta}, \boldsymbol{G}'')(F_{0,v_0}) \cong \mathrm{SU}(n),$$

since the real place v_0 of F_0 is nondecomposed in L', the map γ is a definite involution on D' and the matrix $H \in M_m(F_0)$ is chosen positive definite at the place v_0. In particular, this shows that $\mathrm{Res}_{L/F_0}\boldsymbol{\theta}$ induces a Cartan involution on $\mathrm{SL}_n(\mathbb{C})$.

Next, we consider the maps $\mathrm{Res}_{L/F_0}\boldsymbol{\nu_{p,q}}$ and $\mathrm{Res}_{L/F_0}(\boldsymbol{\nu_{p,q}}\circ\boldsymbol{\theta})$. On $\mathrm{SL}_m(D)$ we have $\nu_{p,q} = \mathrm{Int}(I_{p,q})\circ\theta$ and $\nu_{p,q}\circ\theta = \mathrm{Int}(I_{p,q})$ if $D = E$, and $\nu_{p,q} = \mathrm{Int}(I_{p/2,q/2})\circ\theta$ and $\nu_{p,q}\circ\theta = \mathrm{Int}(I_{p/2,q/2})$ if $D = Q(a, b \mid E)$.[11] However, in the latter case, the matrices $I_{p/2,q/2}$ are mapped to $I_{p,q}$ under a suitable splitting of $M_m(D)\otimes\mathbb{C} \to M_n(\mathbb{C})$, and therefore these maps induce the groups

$$\mathrm{Fix}(\mathrm{Res}_{L/F_0}\boldsymbol{\nu_{p,q}}, \boldsymbol{G}'')(F_{0,v_0}) \cong \{x \in \mathrm{SL}_n(\mathbb{C}) \mid I_{p,q}(x^*)^{-1}I_{p,q} = x\} \cong \mathrm{SU}(p,q)$$

and

$$\begin{aligned}\mathrm{Fix}(\mathrm{Res}_{L/F_0}(\boldsymbol{\nu_{p,q}}\circ\boldsymbol{\theta}), \boldsymbol{G}'')(F_{0,v_0}) &\cong \{x \in \mathrm{SL}_n(\mathbb{C}) \mid I_{p,q}xI_{p,q} = x\}\\ &\cong S(\mathrm{GL}_p(\mathbb{C})\times\mathrm{GL}_q(\mathbb{C}))\end{aligned}$$

for both choices of D.

To deal with the maps $\mathrm{Res}_{E/F_0}\boldsymbol{\eta}$, $\mathrm{Res}_{E/F_0}\boldsymbol{\mu}$ and $\mathrm{Res}_{E/F_0}(\boldsymbol{\eta}\circ\boldsymbol{\mu})$ we proceed as in the proof of Proposition 5.3. In fact, for any E-rational morphism $\boldsymbol{\varphi}$ on $\mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D})$ such that $\mathrm{Res}_{E/F}\boldsymbol{\varphi}$ commutes with $\boldsymbol{\psi}$, we have

$$\begin{aligned}\mathrm{Fix}(\mathrm{Res}_{E/F_0}\boldsymbol{\varphi}, \boldsymbol{G}'')(F_{0,v_0}) &= \mathrm{Fix}(\mathrm{Res}_{E/F}\boldsymbol{\varphi}, \boldsymbol{G}')(F_v)\\ &\cong \mathrm{Fix}(\boldsymbol{\varphi}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(E_{w_1}) \subset \mathrm{SL}_n(\mathbb{C}).\end{aligned}$$

Here, the isomorphism is chosen as in the proof of Proposition 5.3. However, since v is now a complex place of F, we obtain a subgroup of $\mathrm{SL}_n(\mathbb{C})$ instead of $\mathrm{SL}_n(\mathbb{R})$. The result then follows from the determination of $\mathrm{Fix}(\boldsymbol{\varphi}, \mathbf{SL}_{\boldsymbol{m}}(\boldsymbol{D}))(E_{w_1})$ for $\boldsymbol{\varphi} \in \{\boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\eta}\circ\boldsymbol{\mu}\}$, where we use in the third case the fact that $\eta\circ\mu$ is mapped to $\mathrm{Int}(I_{n/2,n/2})$ under a suitable splitting isomorphism $M_m(D)\otimes\mathbb{C} \to M_n(\mathbb{C})$.

For the remaining morphisms $\mathrm{Res}_{E/F_0}\boldsymbol{\eta}\circ\mathrm{Res}_{L/F_0}\boldsymbol{\theta}$, $\mathrm{Res}_{E/F_0}\boldsymbol{\mu}\circ\mathrm{Res}_{L/F_0}\boldsymbol{\theta}$ and $\mathrm{Res}_{E/F_0}(\boldsymbol{\eta}\circ\boldsymbol{\mu})\circ\mathrm{Res}_{L/F_0}\boldsymbol{\theta}$, the result follows from straightforward calculations if $D = E$. Thus, we only deal with the more complicated case where $D = Q(a, b \mid E)$.

Recall that $D = D_0\otimes_{F_0} E$. One can show that D_0 ramifies at all archimedean places of F_0 since the involution $\gamma = \tau_{c,0}\otimes\omega|_{L'}$ on $D_0\otimes_{F_0} L'$ is definite and all real places of F_0 are nondecomposed in L'. In particular, we have $a_{v_0} < 0$ and $b_{v_0} < 0$. Moreover, recall that $F = F_0(\sqrt{D_1})$ for some square-free element $D_1 \in F_0$ such that D_1 is negative under the embedding corresponding to the place v_0 of F_0. Therefore, the quaternion algebra $Q_0 := Q(D_1 a, D_1 b \mid F_0)$ is a division algebra

[11] Recall that in the case of a quaternion algebra the maps $\nu_{p,q}$ are only defined for even p and q.

that splits at the place v_0 of F_0. Note that $x \in (Q_0 \otimes_{F_0} L)$ if and only if $x \in D$ and $(\tau_k \circ \tau_c)(x) = (\mathrm{id} \otimes \omega)(x)$. With the help of these observations, we can describe the fixed points of $\eta \circ \theta$ in $\boldsymbol{G}''(F_0)$ with the equation

$$\begin{aligned}
&\operatorname{Fix}(\operatorname{Res}_{E/F_0} \boldsymbol{\eta} \circ \operatorname{Res}_{L/F_0} \boldsymbol{\theta}, \boldsymbol{G}'')(F_0)\\
&\quad = \{x \in \mathrm{SL}_m(D) \mid (\eta \circ \theta)(x) = x = \psi(x)\}\\
&\quad = \{x \in \mathrm{SL}_m(D) \mid (\tau_r \circ (\tau_{c,0} \otimes \omega))(x) = x = H^{-1}(\sigma(x)^t)^{-1} H\}\\
&\quad = \{x \in \mathrm{SL}_m(D) \mid (\tau_r \circ \tau_c)(x) = (\mathrm{id} \otimes \omega)(x) \text{ and } x = H^{-1}((\tau_{c,0} \otimes \iota)(x)^t)^{-1} H\}\\
&\quad = \{x \in \mathrm{SL}_m(Q_0 \otimes_{F_0} L) \mid x = H^{-1}((\tau_{c,Q_0} \otimes \iota|_L)(x)^t)^{-1} H\},
\end{aligned}$$

where τ_{c,Q_0} denotes the canonical symplectic involution of Q_0. This implies

$$\operatorname{Fix}(\operatorname{Res}_{E/F_0} \boldsymbol{\eta} \circ \operatorname{Res}_{F/F_0} \boldsymbol{\theta}, \boldsymbol{G}'') = \mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}|_{\boldsymbol{Q_0} \otimes_{\boldsymbol{F_0}} \boldsymbol{L}}, \boldsymbol{Q_0} \otimes_{\boldsymbol{F_0}} \boldsymbol{L}, \boldsymbol{\tau_{c,Q_0}} \otimes \boldsymbol{\iota}|_{\boldsymbol{L}}),$$

and hence by Proposition 4.1

$$\operatorname{Fix}(\operatorname{Res}_{E/F_0} \boldsymbol{\eta} \circ \operatorname{Res}_{F/F_0} \boldsymbol{\theta}, \boldsymbol{G}'')(F_{0,v_0}) \cong \mathrm{SL}_n(\mathbb{R}),$$

since the real place v_0 of F_0 is decomposed in L and Q_0 splits at v_0.

A similar calculation for the other two morphisms leads to

$$\operatorname{Fix}(\operatorname{Res}_{E/F_0} \boldsymbol{\mu} \circ \operatorname{Res}_{F/F_0} \boldsymbol{\theta}, \boldsymbol{G}'') = \mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}|_{\boldsymbol{D_0} \otimes_{\boldsymbol{F_0}} \boldsymbol{L}}, \boldsymbol{D_0} \otimes_{\boldsymbol{F_0}} \boldsymbol{L}, \boldsymbol{\tau_{c,0}} \otimes \boldsymbol{\iota}|_{\boldsymbol{L}})$$

and

$$\operatorname{Fix}(\operatorname{Res}_{E/F_0}(\boldsymbol{\eta} \circ \boldsymbol{\mu}) \circ \operatorname{Res}_{L/F_0} \boldsymbol{\theta}, \boldsymbol{G}'') = \mathbf{SU}_{\boldsymbol{m}}(\boldsymbol{h}|_{\boldsymbol{Q_0} \otimes_{\boldsymbol{F_0}} \boldsymbol{L}'}, \boldsymbol{Q_0} \otimes_{\boldsymbol{F_0}} \boldsymbol{L}', \boldsymbol{\tau_{c,Q_0}} \otimes \boldsymbol{\iota}|_{\boldsymbol{L}'}).$$

For the first group, Proposition 4.1 implies

$$\operatorname{Fix}(\operatorname{Res}_{E/F_0} \boldsymbol{\mu} \circ \operatorname{Res}_{F/F_0} \boldsymbol{\theta}, \boldsymbol{G}'')(F_{0,v_0}) \cong \mathrm{SL}_{n/2}(\mathbb{H}) \cong \mathrm{SU}^*(n),$$

since the real place v_0 of F_0 is decomposed in L and D_0 ramifies at v_0. For the second group, we note that the involution $\tau_{c,Q_0} \otimes \iota|_{L'}$ of the second kind cannot be definite on $Q_0 \otimes_{F_0} L'$ because Q_0 does not ramify at the real place v_0 of F_0 that is nondecomposed in L'. This means we get a signature of $(n/2, n/2)$ when passing to the F_{0,v_0}-rational points:

$$\operatorname{Fix}(\operatorname{Res}_{E/F_0}(\boldsymbol{\eta} \circ \boldsymbol{\mu}) \circ \operatorname{Res}_{L/F_0} \boldsymbol{\theta})(F_{0,v_0}) \cong \mathrm{SU}(n/2, n/2). \qquad \square$$

6.3. In this section, we study the geometric cycles defined by the various morphisms on $\boldsymbol{G}''$. To do this, we pass to the algebraic $\mathbb{Q}$-group $\boldsymbol{G} := \operatorname{Res}_{F_0/\mathbb{Q}} \boldsymbol{G}''$. This is an algebraic group over $\mathbb{Q}$ with $\boldsymbol{G}(\mathbb{Q}) \cong \boldsymbol{G}''(F_0)$ and

$$\boldsymbol{G}(\mathbb{R}) \cong \boldsymbol{G}''(\mathbb{R} \otimes_{\mathbb{Q}} F_0) = \boldsymbol{G}'(\mathbb{R} \otimes_{\mathbb{Q}} F) = \prod_{v' \in V_\infty(F)} \boldsymbol{G}'(F_{v'}) = \mathrm{SL}_n(\mathbb{C}) \times \prod_{\substack{v' \in V_\infty \\ v' \neq v}} \mathrm{SU}(n).$$

Moreover, we have $\mathbb{Q}$-rational morphisms $\boldsymbol{\theta}$, $\boldsymbol{\nu_{p,q}}$, $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ of order two on $\boldsymbol{G}$ that are induced from the corresponding morphisms of $\boldsymbol{G}''$.

Let $\Gamma \subset \boldsymbol{G}(\mathbb{Q})$ be an arithmetic subgroup of $\boldsymbol{G}$. In analogy to the real case, Γ gives rise to a discrete cocompact subgroup of $\mathrm{SL}_n(\mathbb{C})$ that we still denote by Γ for simplicity of notation. Let K' denote a maximal compact subgroup of $\boldsymbol{G}(\mathbb{R})$ and $X := K' \backslash \boldsymbol{G}(\mathbb{R})$ the symmetric space attached to $\boldsymbol{G}(\mathbb{R})$. Since $\boldsymbol{G}(\mathbb{R})$ is a product of $\mathrm{SL}_n(\mathbb{C})$ and compact factors, X is isomorphic to $\mathrm{SU}(n)\backslash \mathrm{SL}_n(\mathbb{C})$ and Γ acts on X by right translations. Note that X is a symmetric space of real dimension

$$\dim X = \dim(\mathrm{SL}_n(\mathbb{C})) - \dim(\mathrm{SU}(n)) = 2n^2 - 2 - (n^2 - 1) = n^2 - 1.$$

Let now $\Gamma \subset \boldsymbol{G}(\mathbb{Q})$ be a torsion-free arithmetic subgroup of $\boldsymbol{G}$ and assume that Γ and K' are invariant under $\boldsymbol{\theta}$, $\boldsymbol{\nu_{p,q}}$, $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$. Then these morphisms induce certain special geometric cycles in X/Γ, as explained in Section 3:

Theorem 6.4. *The pair of morphisms $(\boldsymbol{\eta}, \boldsymbol{\eta} \circ \boldsymbol{\theta})$ and, if n is even and $n > 2$, the pairs $(\boldsymbol{\mu}, \boldsymbol{\mu} \circ \boldsymbol{\theta})$ and $(\boldsymbol{\eta} \circ \boldsymbol{\mu}, (\boldsymbol{\eta} \circ \boldsymbol{\mu}) \circ \boldsymbol{\theta})$ induce pairs of special geometric cycles $C(\boldsymbol{\eta})$, $C(\boldsymbol{\eta} \circ \boldsymbol{\theta})$, $C(\boldsymbol{\mu})$, $C(\boldsymbol{\mu} \circ \boldsymbol{\theta})$ and $C(\boldsymbol{\eta} \circ \boldsymbol{\mu})$, $C((\boldsymbol{\eta} \circ \boldsymbol{\mu}) \circ \boldsymbol{\theta})$ in X/Γ. Moreover, the morphisms $\boldsymbol{\nu_{p,q}}$ and $\boldsymbol{\nu_{p,q}} \circ \boldsymbol{\theta}$ induce a family of pairs of special geometric cycles $C(\boldsymbol{\nu_{p,q}})$, $C(\boldsymbol{\nu_{p,q}} \circ \boldsymbol{\theta})$ in X/Γ, for positive integers p and q with $p + q = n$ if $\boldsymbol{G}$ is induced from a special unitary group over an algebraic number field, and for positive, even integers p and q with $p + q = n$ if $\boldsymbol{G}$ is induced from a special unitary group over a quaternion algebra. The properties of these cycles are summarized in Table 3.*

Proof. This is proved completely analogously to Theorem 5.5. □

$C = C(\boldsymbol{\varphi})$	$\mathrm{Fix}(\boldsymbol{\varphi}, \boldsymbol{G})(\mathbb{R}) \cong$	$\dim C$
$C(\boldsymbol{\nu_{p,q}})$	$\mathrm{SU}(p, q)$	$2pq$
$C(\boldsymbol{\nu_{p,q}} \circ \boldsymbol{\theta})$	$S(\mathrm{GL}_p(\mathbb{C}) \times \mathrm{GL}_q(\mathbb{C}))$	$p^2 + q^2 - 1$
$C(\boldsymbol{\eta})$	$\mathrm{SO}(n, \mathbb{C})$	$\frac{1}{2}(n^2 - n)$
$C(\boldsymbol{\eta} \circ \boldsymbol{\theta})$	$\mathrm{SL}_n(\mathbb{R})$	$\frac{1}{2}(n^2 + n) - 1$
$C(\boldsymbol{\mu})$	$\mathrm{Sp}(n, \mathbb{C})$	$\frac{1}{2}(n^2 + n)$
$C(\boldsymbol{\mu} \circ \boldsymbol{\theta})$	$\mathrm{SU}^*(n)$	$\frac{1}{2}(n^2 - n) - 1$
$C(\boldsymbol{\eta} \circ \boldsymbol{\mu})$	$S(\mathrm{GL}_{n/2}(\mathbb{C}) \times \mathrm{GL}_{n/2}(\mathbb{C}))$	$\frac{1}{2}n^2 - 1$
$C((\boldsymbol{\eta} \circ \boldsymbol{\mu}) \circ \boldsymbol{\theta})$	$\mathrm{SU}(n/2, n/2)$	$\frac{1}{2}n^2$

Table 3. Geometric cycles in $\mathrm{SU}(n)\backslash \mathrm{SL}_n(\mathbb{C})/\Gamma$: the isomorphism in the second column is up to compact factors and the bottom half of the table is only applicable if n is even and $n > 2$.

Theorem 6.5. *Let $n \in \mathbb{N}$ be arbitrary.*

(1) *There exists a cocompact discrete subgroup Γ_1 of $\mathrm{SL}_n(\mathbb{C})$ that arises from an arithmetic subgroup of a special unitary group over an algebraic number field, such that $H^k(X/\Gamma_1, \mathbb{C})$ contains nontrivial cohomology classes for*

$$k = 2pq \quad \text{and} \quad k = p^2 + q^2 - 1,$$

where p and q are positive integers with $p + q = n$, and for

$$k = \tfrac{1}{2}(n^2 - n) \quad \text{and} \quad k = \tfrac{1}{2}(n^2 + n) - 1.$$

Moreover, if n is even and $n \neq 2$, there are nontrivial cohomology classes in the degrees

$$k = \tfrac{1}{2}(n^2 + n), \quad k = \tfrac{1}{2}(n^2 - n) - 1, \quad k = \tfrac{1}{2}n^2 - 1 \quad \text{and} \quad k = \tfrac{1}{2}n^2.$$

(2) *If n is even, there exists a discrete, cocompact subgroup Γ_2 of $\mathrm{SL}_n(\mathbb{C})$ that arises from an arithmetic subgroup of a special unitary group over a quaternion algebra, such that $H^k(X/\Gamma_2, \mathbb{C})$ contains nontrivial cohomology classes for*

$$k = 2pq \quad \text{and} \quad k = p^2 + q^2 - 1,$$

where p and q are positive, even integers with $p + q = n$, and for

$$k = \tfrac{1}{2}(n^2 - n) \quad \text{and} \quad k = \tfrac{1}{2}(n^2 + n) - 1.$$

Moreover, if $n \neq 2$, there exist nontrivial cohomology classes in the degrees

$$k = \tfrac{1}{2}(n^2 + n), \quad k = \tfrac{1}{2}(n^2 - n) - 1, \quad k = \tfrac{1}{2}n^2 - 1 \quad \text{and} \quad k = \tfrac{1}{2}n^2.$$

In both cases these classes are not in the image of the respective injective map

$$\beta^*_{\Gamma_i} : H^*(X_u, \mathbb{C}) \to H^*(X/\Gamma_i, \mathbb{C}),$$

i.e., they are not represented by $\mathrm{SL}_n(\mathbb{C})$-invariant forms on X.

Proof. The proof is completely analogous to the proof of Theorem 5.6; details are left to the reader. In contrast to the real case, orientability questions are not an issue here, as all occurring fixed point groups are connected Lie subgroups of $\mathrm{SL}_n(\mathbb{C})$. □

Example 6.6. Table 4 summarizes the occurring cycles and the degrees in which they contribute to the cohomology for small values of n. The group Γ denotes a cocompact discrete subgroup of $\mathrm{SL}_n(\mathbb{C})$ chosen as in Theorem 6.5(1) or (2).

Remark. (1) Looking at these examples, the question arises of whether the degrees in which we have constructed nontrivial cohomology classes exhaust all degrees in the cohomology of X/Γ in which there is cohomology that is not coming from the compact dual symmetric space. In general, this is not the case, as we will see in Section 7 using methods from representation theory. For certain choices of n, this

	$\dim X/\Gamma$	Cycle	Subgroup of $\mathrm{SL}_n(\mathbb{C})$	Contributing to degree	Occurs for $\Gamma=\Gamma_1$	$\Gamma=\Gamma_2$
$n=2$	3	$C(\boldsymbol{\nu}_{1,1})$	$\mathrm{SU}(1,1)$	1	$\times$	
		$C(\boldsymbol{\eta}\circ\boldsymbol{\theta})$	$\mathrm{SL}_2(\mathbb{R})$	1	$\times$	$\times$
		$C(\boldsymbol{\nu}_{1,1}\circ\boldsymbol{\theta})$	$S(\mathrm{GL}_1\times\mathrm{GL}_1)$	2	$\times$	
		$C(\boldsymbol{\eta})$	$\mathrm{SO}(2,\mathbb{C})$	2	$\times$	$\times$
$n=3$	8	$C(\boldsymbol{\eta}\circ\boldsymbol{\theta})$	$\mathrm{SL}_3(\mathbb{R})$	3	$\times$	
		$C(\boldsymbol{\nu}_{1,2})$	$\mathrm{SU}(1,2)$	4	$\times$	
		$C(\boldsymbol{\nu}_{1,2}\circ\boldsymbol{\theta})$	$S(\mathrm{GL}_1\times\mathrm{GL}_2)$	4	$\times$	
		$C(\boldsymbol{\eta})$	$\mathrm{SO}(3,\mathbb{C})$	5	$\times$	

Table 4. Complex case: degrees in $H^*(X/\Gamma)$ in which we have nontrivial cohomology classes coming from special cycles.

can also be seen using the Euler characteristic: it is a consequence of the Gauss–Bonnet formula that for compact quotients X/Γ, where $X=\mathrm{SU}(n)\backslash\mathrm{SL}_n(\mathbb{C})$ and $n\geq 2$, the Euler characteristic of X/Γ is always 0. This implies that the sum over the Betti numbers in even degrees equals the sum over the Betti numbers in odd degrees.

Now for certain choices of n (in fact, whenever $n\geq 2$ and $n\equiv 1 \pmod 4$) the cohomology classes constructed in Theorem 6.5 all contribute to even degrees in the cohomology of X/Γ. Therefore, the vanishing of the Euler characteristic implies the existence of at least one nontrivial cohomology class in an odd degree that does not lie in the image of the cohomology of the compact dual symmetric space.

The smallest n to which our argumentation applies is $n=5$. For this case, one can easily read off from Theorem 6.5 that the constructed cycles do indeed only contribute to even degrees.

(2) Again, by using the technique of intersecting crosswise, one can show that when $n=2$ and $\Gamma=\Gamma_1$ the two cohomology classes in each of the degrees 1 and 2 are linearly independent. However, for the case $n=3$ it remains an open question whether or not the two classes in degree 4 are linearly independent.

7. Representation theory and Matsushima's formula

Let G be a connected semisimple Lie group (with finite center), K a maximal compact subgroup, $X:=K\backslash G$ the associated symmetric space and $\Gamma\subset G$ a discrete, cocompact subgroup. By a well-known result of Matsushima [1962], the cohomology of X/Γ decomposes as a finite algebraic sum over the set of equivalence classes of irreducible unitary representations of G,

$$H^*(X/\Gamma, \mathbb{C}) = \bigoplus_{\pi \in \widehat{G}} m(\pi, \Gamma) H^*(\mathfrak{g}, K; H^\infty_{\pi,K}),$$

where the $m(\pi, \Gamma)$ are nonnegative integers and we denote by $H^\infty_{\pi,K}$ the Harish-Chandra module of K-finite, smooth vectors associated with an element $\pi \in \widehat{G}$. Moreover, $m(\mathbb{C}, \Gamma) = 1$, i.e., there is an injection of the $(\mathfrak{g}, K)$-cohomology of the trivial representation into $H^*(X/\Gamma, \mathbb{C})$.

The *unitary representations with nonvanishing* $(\mathfrak{g}, K)$*-cohomology* contributing to the right-hand side of Matsushima's formula are classified by the work of Enright [1979] (for complex groups) and Vogan and Zuckerman [1984] (for real groups). Note that, by a well-known result of Wigner (see [Borel and Wallach 2000, Theorem 5.3(ii)]), the representations π with $H^*(\mathfrak{g}, K; \mathbb{C} \otimes H^\infty_{\pi,K}) \neq 0$ are only those with trivial infinitesimal character. Representations occurring with a nontrivial multiplicity are called *automorphic representations of G with respect to* Γ. In general, given an irreducible unitary representation π of G, it is still an open question whether the corresponding multiplicity $m(\pi, \Gamma)$ is nontrivial or not. For groups admitting discrete series representations, there are nonvanishing results by DeGeorge and Wallach [1978], Wallach [1990], Langlands [1966], and others (see [Schwermer 1990]). We point out that for our cases of interest (i.e., $G = \mathrm{SL}_n(\mathbb{R})$ or $G = \mathrm{SL}_n(\mathbb{C})$), there is no discrete series except for the case $G = \mathrm{SL}_2(\mathbb{R})$.

Against this background, the result from Section 6 can be interpreted as a result in the theory of automorphic representations. To make a precise statement and possibly identify one (or several) automorphic representations explicitly, we will devote this section to the classification of all irreducible unitary representations with nonvanishing $(\mathfrak{g}, K)$-cohomology of the group $\mathrm{SL}_n(\mathbb{C})$ and the determination of their cohomology.

7.1. First we need to fix some notation. Let G be a complex simply connected semisimple Lie group with Lie algebra $\mathfrak{g}$. Considered as a real Lie algebra, $\mathfrak{g}$ has a Cartan involution θ and a corresponding Cartan decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$. Let $\mathfrak{h}$ denote a θ-stable Cartan subalgebra of $\mathfrak{g}$. Then $\mathfrak{h}$ admits the structure of a complex Lie algebra and we denote by $\Phi(\mathfrak{g}, \mathfrak{h})$ and $\Phi^+(\mathfrak{g}, \mathfrak{h})$ the set of roots and a system of positive roots of the pair $(\mathfrak{g}, \mathfrak{h})$, respectively. We denote by $\mathfrak{q}_0$ the minimal parabolic subalgebra of $\mathfrak{g}$ associated with the system of positive roots Φ^+ and by $\mathfrak{q} \supset \mathfrak{q}_0$ a standard parabolic subalgebra of $\mathfrak{g}$. Let $\mathfrak{l}$ denote the Levi factor of $\mathfrak{q}$ and $\mathfrak{s} = [\mathfrak{l}, \mathfrak{l}]$ its derived Lie algebra. Then $\mathfrak{h} \cap \mathfrak{s}$ is a Cartan subalgebra of $\mathfrak{s}$ and we can identify the root system $\Phi_{\mathfrak{s}}$ of $\mathfrak{s}$ with respect to $\mathfrak{h} \cap \mathfrak{s}$ with the set of roots in Φ that are trivial on the center $Z_{\mathfrak{l}}$ of $\mathfrak{l}$. Using this identification, we can set $\Phi^+_{\mathfrak{s}} := \Phi_{\mathfrak{s}} \cap \Phi^+$ and this is a system of positive roots for $\Phi_{\mathfrak{s}}$.

On the other hand, we may consider $\mathfrak{q}$ as a real parabolic subalgebra of $\mathfrak{g}$ and as such it has a Langlands decomposition of the form $\mathfrak{q} = \mathfrak{m} \oplus \mathfrak{a} \oplus \mathfrak{n}$. We denote by Q_0, Q, L, S, M, A and N the connected Lie subgroups of G with Lie algebras $\mathfrak{q}_0$, $\mathfrak{q}$, $\mathfrak{l}$, $\mathfrak{s}$, $\mathfrak{m}$, $\mathfrak{a}$ and $\mathfrak{n}$, respectively.

The irreducible unitary representations of G with trivial infinitesimal character have been completely classified by the work of Delorme and Enright. They have shown that one can associate with each standard parabolic subgroup $Q \supset Q_0$ a principal series representation π_Q that has the desired properties and that these representations exhaust the set of irreducible unitary representations of G with trivial infinitesimal character up to unitary equivalence. Being principal series representations, the $(\mathfrak{g}, K)$-cohomology of the π_Q can be computed with the help of a well-known theorem [Borel and Wallach 2000]. This leads to the following general result.

Theorem 7.2. *Let G be a connected, simply connected complex Lie group. The correspondence $Q \leftrightarrow \pi_Q$ is a bijective correspondence between the standard parabolic subgroups $Q \supset Q_0$ of G and the set of equivalence classes of irreducible unitary representations of G with trivial infinitesimal character.*

The relative Lie algebra cohomology of the representations π_Q is given by

$$H^{k+d_Q}(\mathfrak{g}, K; H^\infty_{\pi_Q,K}) = \bigoplus_{r+s=k} \big(H^r(\mathfrak{m}, K_Q; \mathbb{C}) \otimes \textstyle\bigwedge^s \mathfrak{a}_{\mathbb{C}}\big), \tag{1}$$

where $K_Q := K \cap Q$ and $d_Q := |\Phi^+(\mathfrak{g}, \mathfrak{h})| - |\Phi^+_{\mathfrak{s}}|$.[12]

7.3. Let us apply the above result to the case $G = \mathrm{SL}_n(\mathbb{C})$. On $\mathfrak{sl}_n(\mathbb{C})$ considered as a real Lie algebra, we have a Cartan involution $\theta : X \mapsto -\overline{X}^t$. The subalgebra $\mathfrak{h} := \{X \in \mathfrak{sl}_n(\mathbb{C}) \mid X = \mathrm{diag}(x_1, \dots, x_n)\}$ of diagonal matrices is a θ-stable Cartan subalgebra of $\mathfrak{sl}_n(\mathbb{C})$ and, with the usual choice of positive roots, the algebra $\mathfrak{q}_0$ of upper triangular matrices is a Borel subalgebra in $\mathfrak{sl}_n(\mathbb{C})$. Then the standard parabolic subalgebras of $\mathfrak{g}$ are in bijective correspondence with the set of compositions of n; with a composition $n = \ell_1 + \dots + \ell_m$ we associate the parabolic subalgebra

$$\mathfrak{q} = \mathfrak{q}_{\ell_1,\dots,\ell_m} = \left\{ X \in \mathfrak{g} \;\middle|\; X = \begin{pmatrix} X_1 & * & * \\ & \ddots & * \\ 0 & & X_m \end{pmatrix}, \text{ where } X_j \in \mathrm{GL}_{\ell_j}(\mathbb{C}),\ 1 \le j \le m \right\}.$$

The Levi component of $\mathfrak{q}$ is given by the subalgebra of block diagonal matrices $\mathfrak{l} = \{X \in \mathfrak{q} \mid X = \mathrm{diag}(X_1, \dots, X_m)\}$ and it decomposes into its semisimple part

[12] Note that $K_Q = K \cap Q \subset M$ and that K_Q is a maximal compact subgroup of M by [Borel and Wallach 2000, Section 0.1.6], so taking the relative Lie algebra cohomology of $\mathfrak{m}$ with respect to K_Q is defined.

and its center, given respectively by $\mathfrak{s} = \{X \in \mathfrak{l} \mid \mathrm{tr}(X_j) = 0 \text{ for all } 1 \le j \le m\}$ and $Z_{\mathfrak{l}} = \{X \in \mathfrak{l} \mid X_j = x_j I_{\ell_j} \text{ for some } x_j \in \mathbb{C},\ 1 \le j \le m\}$.

Considering $\mathfrak{q}$ as a real Lie algebra, it also has a Langlands decomposition of the form $\mathfrak{q} = \mathfrak{m} \oplus \mathfrak{a} \oplus \mathfrak{n}$, where

$$\mathfrak{m} = \{X \in \mathfrak{q} \mid X = \mathrm{diag}(X_1, \dots, X_m),\ \mathrm{tr}(X_j) \in i \cdot \mathbb{R} \text{ for all } 1 \le j \le m\},$$
$$\mathfrak{a} = \{X \in \mathfrak{q} \mid X = \mathrm{diag}(X_1, \dots, X_m),\ X_j = x_j I_{\ell_j} \text{ for some } x_j \in \mathbb{R},\ 1 \le j \le m\}$$

and

$$\mathfrak{n} = \left\{ X \in \mathfrak{q} \,\middle|\, X = \begin{pmatrix} 0 & * & * \\ & \ddots & * \\ 0 & & 0 \end{pmatrix}, \text{ where the } j\text{-th diagonal 0-block is of size } \ell_j \times \ell_j \right\}.$$

We use capital letters to denote the connected Lie subgroups of $\mathrm{SL}_n(\mathbb{C})$ corresponding to these Lie algebras.

Theorem 7.4. *The equivalence classes of irreducible unitary representations of* $\mathrm{SL}_n(\mathbb{C})$ *with trivial infinitesimal character are in one-to-one correspondence with the standard parabolic subgroups* $Q \supset Q_0$ *of* $\mathrm{SL}_n(\mathbb{C})$; *a standard parabolic subgroup corresponds to the induced representation* π_Q. *The* $(\mathfrak{g}, K)$*-cohomology of the representation* π_Q *is given by the Poincaré polynomial*

$$P_{H^*(\mathfrak{g},K;H^\infty_{\pi_Q,K})}(t) = t^d \cdot \left(\sum_{k=0}^{m-1} \tbinom{m-1}{k} t^k \right) \cdot \prod_{\substack{j=1 \\ \ell_j \neq 1}}^{m} \prod_{k=2}^{\ell_j} (1 + t^{2k-1}),$$

where m *denotes the number of blocks of* Q, ℓ_j *denotes the length of the* j*-th block and*

$$d := \tfrac{1}{2} n(n-1) - \sum_{j=1}^{m} \tfrac{1}{2} \ell_j (\ell_j - 1).$$

Proof. The first part of the theorem is a direct application of Theorem 7.2 to the connected simply connected complex Lie group $\mathrm{SL}_n(\mathbb{C})$.

To compute the cohomology of the representations π_Q, we use the formula from Theorem 7.2. Note that, in terms of Poincaré polynomials, this formula says

$$P_{H^*(\mathfrak{g},K;H^\infty_{\pi_Q,K})}(t) = t^{d_Q} \cdot P_{H^*(\mathfrak{m},K_Q;\mathbb{C})}(t) \cdot P_{\bigwedge(\mathfrak{a}_{\mathbb{C}})}(t). \tag{2}$$

Therefore, it suffices to determine the number d_Q and the Poincaré polynomials of $H^*(\mathfrak{m}, K_Q; \mathbb{C})$ and $\bigwedge(\mathfrak{a}_{\mathbb{C}})$.

(1) *The Poincaré polynomial of* $\bigwedge(\mathfrak{a}_{\mathbb{C}})$: From the structure of $\mathfrak{a}$ given above we conclude that $\mathfrak{a}_{\mathbb{C}}$ has complex dimension $m-1$, and thus

$$P_{\bigwedge(\mathfrak{a}_{\mathbb{C}})}(t) = \sum_{k=0}^{m-1} \binom{m-1}{k} t^k$$

by the general formula for the Poincaré polynomial of the exterior algebra of a complex vector space.

(2) *The Poincaré polynomial of* $H^*(\mathfrak{m}, K_Q; \mathbb{C})$: First, we consider $Q = Q_0$. The group Q_0 has the Langlands decomposition $Q_0 = M_0 A_0 N_0$, where M_0 is compact. This implies $K_{Q_0} = K \cap M_0 = M_0$, so in fact we consider the relative Lie algebra cohomology $H^*(\mathfrak{m}_0, M_0; \mathbb{C})$. By the definition of relative Lie algebra cohomology, this is one-dimensional in degree 0 and trivial in all higher degrees. In particular,

$$P_{H^*(\mathfrak{m}_0, K_{Q_0}; \mathbb{C})}(t) = 1.$$

Now let $Q \neq Q_0$. The Lie algebra $\mathfrak{m}$ is reductive, has semisimple part $\mathfrak{s}$ and center $Z_{\mathfrak{m}} \subset \mathfrak{k}$. Using the Künneth rule (see [Borel and Wallach 2000, Section I.1.3]) and the fact that K_Q and $K_Q \cap S$ are connected, we obtain $H^*(\mathfrak{m}, K_Q; \mathbb{C}) = H^*(\mathfrak{s}, K_Q \cap S; \mathbb{C})$, so we can restrict to the semisimple part. From the structure of $\mathfrak{s}$ given above we deduce that $S \cong \prod_{j=1}^m \mathrm{SL}_{\ell_j}(\mathbb{C})$, which is clearly the group of real points of a reductive algebraic $\mathbb{R}$-group. Therefore, by [Vogan 1997, Theorem 2.10], $H^*(\mathfrak{s}, K_Q \cap S; \mathbb{C})$ equals the cohomology of the compact symmetric space $\prod_{j=1}^m \mathrm{SU}(\ell_j)$, the compact dual symmetric space of S. For $\ell_j \geq 2$, the Poincaré polynomial of $H^*(\mathrm{SU}(\ell_j); \mathbb{C})$ is given by

$$P_{H^*(\mathrm{SU}(\ell_j); \mathbb{C})}(t) = \prod_{k=2}^{\ell_j} (1 + t^{2k-1})$$

(see [Greub et al. 1976, Theorem VI.X]). For $\ell_j = 1$, we have $\mathrm{SU}(1) = S^1$, so the Poincaré polynomial is given by $P_{H^*(\mathrm{SU}(1), \mathbb{C})}(t) = 1$. Putting everything together, we obtain the formula

$$P_{H^*(\mathfrak{m}, K_Q; \mathbb{C})}(t) = \prod_{j=1}^m P_{H^*(\mathrm{SU}(\ell_j), \mathbb{C})}(t) = \prod_{\substack{j=1 \\ \ell_j \neq 1}}^m \prod_{k=2}^{\ell_j} (1 + t^{2k-1}),$$

where we have used the Künneth rule for singular cohomology in the first step.

(3) *Determination of* d_Q: Recall from Theorem 7.2 that $d_Q = |\Phi^+| - |\Phi_{\mathfrak{s}}^+|$. From the structure of the set of positive roots $\Phi^+(\mathfrak{g}, \mathfrak{h})$ and the definition of $\Phi_{\mathfrak{s}}$ as given above, we conclude that

$$d_Q = \tfrac{1}{2} n(n-1) - \sum_{j=1}^m \tfrac{1}{2} \ell_j (\ell_j - 1). \qquad \square$$

$(\ell_1, \dots, \ell_m)$	$P_{H^*(\mathfrak{g},K;\pi_{Q_{\ell_1,\dots,\ell_m}})}(t)$
$(1, 1, 1)$	$t^3 + 2t^4 + t^5$
$(1, 2)$	$t^2 + t^3 + t^5 + t^6$
$(2, 1)$	$t^2 + t^3 + t^5 + t^6$
(3)	$1 + t^3 + t^5 + t^8$

Table 5. Poincaré polynomials of the irreducible unitary representations of $\mathrm{SL}_3(\mathbb{C})$ with nontrivial $(\mathfrak{g}, K)$-cohomology.

Example 7.5. Let's look at some examples for small values of n.

In the case $n = 2$, $\mathrm{SL}_2(\mathbb{C})$ only has two standard parabolic subgroups, corresponding to the compositions $2 = 1 + 1$ and $2 = 2$ of 2. These are the minimal parabolic subgroup $Q = Q_0$ and the whole group $Q = G$, with associated representations π_{Q_0} and π_G (the latter being the trivial representation). An application of Theorem 7.4 gives us the Poincaré polynomials of the $(\mathfrak{g}, K)$-cohomology of these representations:

$$P_{H^*(\mathfrak{g},K;\pi_{Q_0})}(t) = t + t^2, \quad P_{H^*(\mathfrak{g},K;\pi_G)}(t) = 1 + t^3.$$

For $n = 3$, the situation is more complicated and we will give the results in Table 5. We denote a composition $n = \ell_1 + \cdots + \ell_m$ by the m-tuple $(\ell_1, \dots, \ell_m)$ and the associated parabolic subgroup by $Q_{\ell_1,\dots,\ell_m}$.

7.6. Let us relate our findings to the results of Theorem 6.5. Assume we are given a cocompact discrete subgroup $\Gamma \subset \mathrm{SL}_n(\mathbb{C})$. The irreducible unitary representations with trivial infinitesimal character that we have classified in the previous sections are exactly the representations that can possibly contribute to the cohomology of X/Γ via Matsushima's formula. In general, the question of whether or not a given representation $\pi \in \widehat{G}$ does actually contribute to the cohomology, i.e., $m(\Gamma, \pi) \neq 0$, is still open. However, the nonvanishing results for the cohomology in Theorem 6.5 imply the existence of (at least) one nontrivial automorphic representation for $\mathrm{SL}_n(\mathbb{C})$ with respect to Γ. For small values of n, we can even identify explicit representations with nonvanishing multiplicity. If for one of the degrees for which we have constructed nontrivial cohomology classes in Theorem 6.5 there is exactly one representation π with nontrivial $(\mathfrak{g}, K)$-cohomology that contributes in that degree, we can deduce that the corresponding multiplicity $m(\pi, \Gamma)$ is not zero.

To be able to compare the degrees in which we have cohomology coming from geometric cycles and the degrees to which our representations can possibly contribute, we summarize this information in Tables 6 and 7 for $n = 2, 3$. As above, we denote a representation π_Q by the associated tuple $(\ell_1, \dots, \ell_m)$.

	0	1	2	3
Cycles		×	×	
Trivial representation	×			×
(1, 1)		×	×	

Table 6. Complex case: contribution to the cohomology of X/Γ, $n = 2$.

	0	1	2	3	4	5	6	7	8
Cycles				×	×	×			
Trivial representation	×			×		×			×
(1, 1, 1)				×	×	×			
(2, 1)			×	×		×	×		
(1, 2)			×	×		×	×		

Table 7. Complex case: contribution to the cohomology of X/Γ, $n = 3$.

Corollary 7.7. *Let $n \in \{2, 3\}$ and let Q_0 denote the minimal parabolic subgroup of upper triangular matrices of $\mathrm{SL}_n(\mathbb{C})$. Then there exists a cocompact discrete subgroup $\Gamma \subset \mathrm{SL}_n(\mathbb{C})$ such that the multiplicity $m(\Gamma, \pi_{Q_0})$ is not 0.*

Proof. We choose Γ as in Theorem 6.5 for $n = 2$ or $n = 3$. Then the result can be read off from Tables 6 and 7: when $n = 2$, we have cycles contributing to the cohomology in degrees 1 and 2, and π_{Q_0} is the only unitary representation that has cohomology in these degrees. Therefore, we conclude $m(\Gamma, \pi_{Q_0}) \neq 0$. Similarly, for $n = 3$, we have cycles contributing to degree 4, and π_{Q_0} is the only unitary representation of $\mathrm{SL}_3(\mathbb{C})$ that has cohomology in degree 4. □

Remark. Unfortunately, for bigger n the situation is more complicated and this reasoning is not successful anymore. Already in the case $n = 4$ one can easily see (by looking at a similar table) that there is no degree in the cohomology of X/Γ in which we have a nontrivial class coming from a cycle and to which only one representation can contribute.

Acknowledgements

I would like to thank Professor Joachim Schwermer for supporting and supervising me during my thesis in which the results of this article were obtained.

References

[Ash and Ginzburg 1994] A. Ash and D. Ginzburg, "p-adic L-functions for $\mathrm{GL}(2n)$", *Invent. Math.* **116**:1-3 (1994), 27–73. MR 95f:11038 Zbl 0807.11029

[Borel 1963] A. Borel, "Compact Clifford–Klein forms of symmetric spaces", *Topology* **2** (1963), 111–122. MR 26 #3823 Zbl 0116.38603

[Borel and Harder 1978] A. Borel and G. Harder, "Existence of discrete cocompact subgroups of reductive groups over local fields", *J. Reine Angew. Math.* **298** (1978), 53–64. MR 80b:22022 Zbl 0385.14014

[Borel and Harish-Chandra 1962] A. Borel and Harish-Chandra, "Arithmetic subgroups of algebraic groups", *Ann. of Math.* (2) **75** (1962), 485–535. MR 26 #5081 Zbl 0107.14804

[Borel and Wallach 2000] A. Borel and N. R. Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*, 2nd ed., Mathematical Surveys and Monographs **67**, American Mathematical Society, Providence, RI, 2000. MR 2000j:22015 Zbl 0980.22015

[Delorme 1979] P. Delorme, "Sur la cohomologie continue des représentations unitaires irréductibles des groupes de Lie semi-simples complexes", pp. 179–191 in *Analyse harmonique sur les groupes de Lie, II* (Nancy-Strasbourg, 1976–1978), edited by P. Eymard et al., Lecture Notes in Mathematics **739**, Springer, Berlin, 1979. MR 81f:22018 Zbl 0426.22015

[Enright 1979] T. J. Enright, "Relative Lie algebra cohomology and unitary representations of complex Lie groups", *Duke Math. J.* **46**:3 (1979), 513–525. MR 81i:22007 Zbl 0427.22010

[de George and Wallach 1978] D. L. de George and N. R. Wallach, "Limit formulas for multiplicities in $L^2(\Gamma\backslash G)$", *Ann. of Math.* (2) **107**:1 (1978), 133–150. MR 58 #11231 Zbl 0397.22007

[Greub et al. 1976] W. Greub, S. Halperin, and R. Vanstone, *Connections, curvature, and cohomology, III: Cohomology of principal bundles and homogeneous spaces*, Pure and Applied Mathematics **47**, Academic Press, New York, 1976. MR 53 #4110 Zbl 0372.57001

[Helgason 1978] S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*, Pure and Applied Mathematics **80**, Academic Press, New York, 1978. MR 80k:53081 Zbl 0451.53038

[Knapp 1996] A. W. Knapp, *Lie groups beyond an introduction*, Progress in Mathematics **140**, Birkhäuser, Boston, 1996. MR 98b:22002 Zbl 0862.22006

[Knus et al. 1998] M.-A. Knus, A. Merkurjev, M. Rost, and J.-P. Tignol, *The book of involutions*, American Mathematical Society Colloquium Publications **44**, American Mathematical Society, Providence, RI, 1998. MR 2000a:16031 Zbl 0955.16001

[Langlands 1966] R. P. Langlands, "Dimension of spaces of automorphic forms", pp. 253–257 in *Algebraic groups and discontinuous subgroups* (Boulder, CO, 1965), edited by A. Borel and G. D. Mostow, Proceedings of Symposia in Pure Mathematics **9**, American Mathematical Society, Providence, RI, 1966. MR 35 #3010 Zbl 0215.11802

[Matsushima 1962] Y. Matsushima, "On Betti numbers of compact, locally symmetric Riemannian manifolds", *Osaka Math. J.* **14**:1 (1962), 1–20. MR 25 #4549 Zbl 0118.38401

[Millson 1976] J. J. Millson, "On the first Betti number of a constant negatively curved manifold", *Ann. of Math.* (2) **104**:2 (1976), 235–247. MR 54 #10488 Zbl 0364.53020

[Millson and Raghunathan 1981] J. J. Millson and M. S. Raghunathan, "Geometric construction of cohomology for arithmetic groups, I", *Proc. Indian Acad. Sci. Math. Sci.* **90**:2 (1981), 103–123. MR 83d:22008 Zbl 0514.22007

[Mostow and Tamagawa 1962] G. D. Mostow and T. Tamagawa, "On the compactness of arithmetically defined homogeneous spaces", *Ann. of Math.* (2) **76** (1962), 446–463. MR 25 #5069 Zbl 0196.53201

[Rohlfs and Schwermer 1993] J. Rohlfs and J. Schwermer, "Intersection numbers of special cycles", *J. Amer. Math. Soc.* **6**:3 (1993), 755–778. MR 94a:11075 Zbl 0811.11039

[Scharlau 1985] W. Scharlau, *Quadratic and Hermitian forms*, Grundlehren der Mathematischen Wissenschaften **270**, Springer, Berlin, 1985. MR 86k:11022 Zbl 0584.10010

[Schwermer 1990] J. Schwermer, "Cohomology of arithmetic groups, automorphic forms and L-functions", pp. 1–29 in *Cohomology of arithmetic groups and automorphic forms* (Luminy-Marseille, 1989), edited by J.-P. Labesse and J. Schwermer, Lecture Notes in Mathematics **1447**, Springer, Berlin, 1990. MR 92e:11059 Zbl 0715.11028

[Schwermer and Waldner 2011] J. Schwermer and C. Waldner, "On the cohomology of uniform arithmetically defined subgroups in $SU^*(2n)$", *Math. Proc. Cambridge Philos. Soc.* **151**:3 (2011), 421–440. MR 2012k:11066 Zbl 1228.22013

[Vogan 1997] D. A. Vogan, Jr., "Cohomology and group representations", pp. 219–243 in *Representation theory and automorphic forms* (Edinburgh, 1996), edited by T. N. Bailey and A. W. Knapp, Proceedings of Symposia in Pure Mathematics **61**, American Mathematical Society, Providence, RI, 1997. MR 98k:22064 Zbl 0890.22009

[Vogan and Zuckerman 1984] D. A. Vogan, Jr. and G. J. Zuckerman, "Unitary representations with nonzero cohomology", *Compositio Math.* **53**:1 (1984), 51–90. MR 86k:22040 Zbl 0692.22008

[Waldner 2010] C. Waldner, "Geometric cycles and the cohomology of arithmetic subgroups of the exceptional group G_2", *J. Topol.* **3**:1 (2010), 81–109. MR 2011j:22018 Zbl 1191.55001

[Wallach 1990] N. R. Wallach, "Limit multiplicities in $L^2(\Gamma\backslash G)$", pp. 31–56 in *Cohomology of arithmetic groups and automorphic forms* (Luminy-Marseille, 1989), edited by J.-P. Labesse and J. Schwermer, Lecture Notes in Mathematics **1447**, Springer, Berlin, 1990. MR 92f:22020 Zbl 0749.22006

Received December 1, 2014. Revised October 4, 2015.

SUSANNE SCHIMPF
FACULTY OF MATHEMATICS
UNIVERSITY OF VIENNA
OSKAR-MORGENSTERN-PLATZ 1
1090 VIENNA
AUSTRIA

susanne.schimpf@univie.ac.at

PACIFIC JOURNAL OF MATHEMATICS
Vol. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.479

ON BLASCHKE'S CONJECTURE

XIAOLE SU, HONGWEI SUN AND YUSHENG WANG

Blaschke's conjecture asserts that if a complete Riemannian manifold M satisfies $\mathrm{diam}(M) = \mathrm{Inj}(M) = \frac{\pi}{2}$, then M is isometric to $\mathbb{S}^n\left(\frac{1}{2}\right)$ or to the real, complex, quaternionic or octonionic projective plane with its canonical metric. We prove that the conjecture is true under the assumption that $\sec_M \geq 1$.

Introduction

The projective spaces $\mathbb{KP}^n$ (considered with their canonical metric, induced from the unit sphere) and the sphere $\mathbb{S}^n\left(\frac{1}{2}\right)$ are the only known examples of complete Riemannian manifolds M satisfying

$$\mathrm{diam}(M) = \mathrm{Inj}(M) = \tfrac{\pi}{2}. \tag{0-1}$$

Here $\mathrm{diam}(M)$ and $\mathrm{Inj}(M)$ are the diameter and injective radius of M, and $\mathbb{K}$ is one of the division algebras $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ or $\mathbb{C}a$, with $n \leq 2$ if $\mathbb{K} = \mathbb{C}a$. A longstanding conjecture, whose history is reviewed in [Besse 1978; Berger 2003; Bougas 2013], asserts that these are the only possibilities:

Blaschke's Conjecture. *If a complete Riemannian manifold M satisfies* (0-1), *then M is isometric to $\mathbb{S}^n\left(\frac{1}{2}\right)$ or a $\mathbb{KP}^n$ endowed with the canonical metric.*

(See (1-1) below for the reason why it is called Blaschke's conjecture.) Up to now, the conjecture is still almost open (there are only some partial answers to it) although (0-1) is an extremely strong condition. Note that the conjecture has no restriction on the curvature. The main purpose of the present paper is to give a positive answer to the conjecture under the additional assumption $\sec_M \geq 1$, which is stated as follows.

Main Theorem. *If a complete Riemannian manifold M satisfies* (0-1) *and* $\sec_M \geq 1$, *then M is isometric to $\mathbb{S}^n\left(\frac{1}{2}\right)$ or a $\mathbb{KP}^n$ endowed with the canonical metric.*

If the curvature has an upper bound, we have the following result of Rovenskii and Toponogov [1998] (see also [Shankar et al. 2005]).

Supported by NSFC 11471039. Yusheng Wang is the corresponding author.
MSC2010: 53C20.
Keywords: Blaschke's conjecture, Berger's rigidity theorem, Toponogov's comparison theorem.

Theorem 0.1. *If a complete, simply connected Riemannian manifold M satisfies* (0-1) *and* $\sec_M \le 4$, *then M is isometric to $\mathbb{S}^n\left(\frac{1}{2}\right)$ or a $\mathbb{KP}^n$* ($\mathbb{K} \neq \mathbb{R}$) *endowed with the canonical metric.*

From our Main Theorem and Theorem 0.1, one can see how beautiful the following Berger's rigidity theorem [Cheeger and Ebin 1975] is.

Theorem 0.2. *Let M be a complete, simply connected Riemannian manifold with* $1 \le \sec_M \le 4$. *If* $\mathrm{diam}(M) = \frac{\pi}{2}$, *then M is isometric to $\mathbb{S}^n\left(\frac{1}{2}\right)$ or a $\mathbb{KP}^n$* ($\mathbb{K} \neq \mathbb{R}$) *endowed with the canonical metric.*

In fact, "$1 \le \sec_M \le 4$" and "simply connected" imply that $\mathrm{Inj}(M) \ge \frac{\pi}{2}$ [Cheeger and Gromoll 1980], so "$\mathrm{diam}(M) = \frac{\pi}{2}$" implies that M (in Theorem 0.2) satisfies (0-1) (note that $\mathrm{Inj}(M) \le \mathrm{diam}(M)$). Hence, the Main Theorem implies Theorem 0.2 in the premise of (0-1) (as does Theorem 0.1). (Of course, "$\sec_M \ge 1$" implies that $\mathrm{diam}(M) \le \pi$, and the maximal diameter theorem asserts that *if* $\mathrm{diam}(M) = \pi$, *then M is isometric to* $\mathbb{S}^n(1)$, so Theorem 0.2 is also called the minimal diameter theorem. Moreover, inspired by Theorem 0.2, Grove and Shiohama, Gromoll and Grove, and Wilhelm supply some beautiful (but not purely isometric) classifications under the conditions "$\sec_M \ge 1$ and $\mathrm{diam}(M) \ge \frac{\pi}{2}$ or $\mathrm{Rad}(M) \ge \frac{\pi}{2}$" [Gromoll and Grove 1987; Wilhelm 1996].)

Moreover, from the proof in [Cheeger and Ebin 1975] for Theorem 0.2, it is not hard to see the following.

Theorem 0.3. *Let M be a complete Riemannian manifold satisfying* (0-1) *and* $1 \le \sec_M \le 4$. *Then M is isometric to $\mathbb{S}^n\left(\frac{1}{2}\right)$ or a $\mathbb{KP}^n$ endowed with the canonical metric.*

We end this section with the idea of our proof of the Main Theorem. We first prove that for any $p \in M$, denoting by $|pq|$ the distance between p and q,

$$\{p\}^{=\pi/2} \triangleq \left\{q \in M \mid |pq| = \tfrac{\pi}{2}\right\}$$

is a complete totally geodesic submanifold in M. Then using Theorem 1.3 below and Toponogov's comparison theorem, we derive by induction that $1 \le \sec_M \le 4$, and thus the proof is done by Theorem 0.3. (We would like to point out that, in the premise of Theorem 1.3, we can use the method in [Gromoll and Grove 1987; 1988; Wilhelm 1996] to give the proof (which involves many significant classification results). By comparison, however, our proof is much more direct.)

1. Blaschke manifolds

A closed Riemannian manifold M is called a Blaschke manifold if it is *Blaschke* at each point $p \in M$, i.e., *$\Uparrow_q^p$ is a great sphere in $\Sigma_q M$ for any q in the cut locus of p*

[Besse 1978], where

$$\Sigma_q M \triangleq \{v \in T_q M \mid |v| = 1\},$$
$$\Uparrow_q^p \triangleq \{\text{the unit tangent vector at } q \text{ of a minimal geodesic from } q \text{ to } p\}.$$

On a Blaschke manifold, one can get the following not so obvious fact (p. 137 in [Besse 1978]).

Proposition 1.1. *For a Blaschke manifold M, we have that* $\operatorname{diam}(M) = \operatorname{Inj}(M)$.

A much more difficult observation is the following (p. 138 in [Besse 1978]).

Proposition 1.2. *Given a closed Riemannian manifold M and a point $p \in M$, if $|pq|$ is a constant for all q in the cut locus of p, then M is Blaschke at p.*

Obviously, it follows from Propositions 1.1 and 1.2 that

(1-1) a closed Riemannian manifold M is Blaschke $\Leftrightarrow \operatorname{diam}(M) = \operatorname{Inj}(M)$.

Up to now, Blaschke's conjecture has been solved only for spheres.

Theorem 1.3 [Besse 1978; Berger 2003]. *If a Blaschke manifold is homeomorphic to a sphere, then it is isometric to the unit sphere* (*up to a rescaling*).

2. Proof of the Main Theorem

We first give our main tool of the paper: Toponogov's comparison theorem.

Theorem 2.1 [Petersen 1998; Grove and Markvorsen 1995]. *Let M be a complete Riemannian manifold with $\sec_M \geq \kappa$, and let $\mathbb{S}^2_\kappa$ be the complete, simply connected 2-manifold of curvature κ.*

(i) *To any $p \in M$ and minimal geodesic $[qr] \subset M$, we associate $\tilde p$ and a minimal geodesic $[\tilde q \tilde r]$ in $\mathbb{S}^2_\kappa$ with $|\tilde p \tilde q| = |pq|$, $|\tilde p \tilde r| = |pr|$ and $|\tilde r \tilde q| = |rq|$. Then for any $s \in [qr]$ and $\tilde s \in [\tilde q \tilde r]$ with $|qs| = |\tilde q \tilde s|$, we have that $|ps| \geq |\tilde p \tilde s|$.*

(ii) *To any minimal geodesics $[qp]$ and $[qr]$ in M, we associate minimal geodesics $[\tilde q \tilde p]$ and $[\tilde q \tilde r]$ in $\mathbb{S}^2_\kappa$ with $|\tilde q \tilde p| = |qp|$, $|\tilde q \tilde r| = |qr|$ and $\angle \tilde p \tilde q \tilde r = \angle pqr$. Then we have that $|\tilde p \tilde r| \geq |pr|$.*

(iii) *If equality in (ii)* (*or in (i) for some s in the interior part of* $[qr]$) *holds, then there exists a minimal geodesic $[pr]$ such that the triangle formed by $[qp]$, $[qr]$ and $[pr]$ bounds a surface which is convex*[1] *and can be isometrically embedded into $\mathbb{S}^2_\kappa$.*

[1] We say that a subset A is convex (resp. totally convex) in M if, between any $x \in A$ and $y \in A$, some minimal geodesic $[xy]$ (resp. all minimal geodesics) belongs to A.

In the rest of this paper, M always denotes the manifold in the Main Theorem, and N denotes $\{p\}^{=\pi/2} \triangleq \{q \in M \mid |pq| = \frac{\pi}{2}\}$ for an arbitrary fixed point $p \in M$. We first give an easy observation following from (0-1) (i.e., $\mathrm{Inj}(M) = \mathrm{diam}(M) = \frac{\pi}{2}$), namely that

(2-1) for any $x \in M$,

there is a minimal geodesic $[pq]$ with $q \in N$ such that $x \in [pq]$.

Lemma 2.2. *N is a complete totally geodesic submanifold in M; if $\dim(N) = 0$, then N consists of a single point.*

Remark 2.3. Since $\sec_M \geq 1$, it follows from (i) of Theorem 2.1 that

$$\{p\}^{\geq\pi/2} \triangleq \{q \in M \mid |pq| \geq \tfrac{\pi}{2}\}$$

is totally convex in M. Note that $N = \{p\}^{\geq\pi/2}$ because $\mathrm{diam}(M) = \frac{\pi}{2}$, and that N is closed in M. On the other hand, since M is a Blaschke manifold, we know that N is a submanifold in M [Besse 1978]. It then follows that N is a totally geodesic submanifold in M. This proof is short because we apply the proposition that N is a submanifold in M, which is a significant property of a Blaschke manifold [Besse 1978]. Here, in order to show the importance of "$\sec_M \geq 1$", we will supply a proof only based on the definition of a Blaschke manifold.

Proof of Lemma 2.2. From Remark 2.3, we know that N is totally convex in M, which implies that N consists of a single point if $\dim(N) = 0$. Hence, we can assume that $\dim(N) > 0$; for any geodesic $\gamma(t)|_{t\in[0,\ell]} \subset N$, we need only to show that its prolonged geodesic $\gamma(t)|_{t\in[0,\ell+\varepsilon]}$ in M also belongs to N for some small $\varepsilon > 0$. Note that, without loss of generality, we can assume that there is a unique minimal geodesic between $\gamma(0)$ and $\gamma(\ell+\varepsilon)$. Due to (2-1), we can select $q \in N$ such that $\gamma(\ell+\varepsilon) \in [pq]$. Observe that $q \neq \gamma(0)$ (otherwise, $\gamma(\ell) \in [pq]$ must hold, contradicting $\gamma(\ell) \in N$). Let $[q\gamma(0)]$ be a minimal geodesic in N (note that N is convex in M). By the first variation formula, it is easy to see that

$$|\uparrow_q^{\gamma(0)} \xi| \geq \tfrac{\pi}{2} \quad \text{in } \Sigma_q M, \quad \text{for any } \xi \in \Uparrow_q^p.$$

On the other hand, $\Uparrow_q^p$ is a great sphere in $\Sigma_q M$ because M is Blaschke at p (see Proposition 1.2). It follows that in fact

$$|\uparrow_q^{\gamma(0)} \xi| = \tfrac{\pi}{2} \quad \text{for any } \xi \in \Uparrow_q^p.$$

Then by (iii) of Theorem 2.1, there is a minimal geodesic $[p\gamma(0)]$ such that the triangle formed by $[q\gamma(0)]$, $[pq]$ and $[p\gamma(0)]$ bounds a surface (containing $[\gamma(0)\gamma(\ell+\varepsilon)]$) which is convex and can be isometrically embedded into $\mathbb{S}^2(1)$. It then has to hold that $[\gamma(0)\gamma(\ell+\varepsilon)] = [\gamma(0)q]$ because $[\gamma(0)\gamma(\ell)]$ belongs to N, and so $[\gamma(0)\gamma(\ell+\varepsilon)] \subset N$. $\square$

Since N is a complete totally geodesic submanifold in M, for any $q \in N$, any minimal geodesic $[pq]$ is perpendicular to N at q, i.e.,

$$\Uparrow_q^p \subseteq (\Sigma_q N)^{=\pi/2} \quad \text{in } \Sigma_q M. \tag{2-2}$$

Then from the proof of Lemma 2.2, we have the following corollary.

Corollary 2.4. *For any minimal geodesics* $[pq]$ *and* $[qq'] \subset N$, *there is a minimal geodesic* $[pq']$ *such that the triangle formed by* $[pq]$, $[qq']$ *and* $[pq']$ *bounds a surface which is convex and can be isometrically embedded into* $\mathbb{S}^2(1)$.

Moreover, the "$\subseteq$" in (2-2) can in fact be changed to "$=$".

Lemma 2.5. *For any* $q \in N$, *we have that* $\Uparrow_q^p = (\Sigma_q N)^{=\pi/2}$ *in* $\Sigma_q M$.

Proof. According to (2-2), it suffices to show that for any $\zeta \in (\Sigma_q N)^{=\pi/2}$ there is a minimal geodesic $[qp]$ such that $\uparrow_q^p = \zeta$. Note that there is a minimal geodesic $[qx]$ ($x \in M$) such that $\uparrow_q^x = \zeta$, and we can assume that there is a unique geodesic between q and x. It follows from (2-1) that there is a minimal geodesic $[pq_x]$ with $q_x \in N$ such that $x \in [pq_x]$. Hence, we need only to show that $q_x = q$. If this is not true, then by Corollary 2.4 there are minimal geodesics $[pq]$ and $[qq_x] \subset N$ such that the triangle formed by $[pq]$, $[pq_x]$ and $[qq_x]$ bounds a surface D which is convex and can be isometrically embedded into $\mathbb{S}^2(1)$. Note that $[qx]$ belongs to D. This is impossible because both $[qp]$ (see (2-2)) and $[qx]$ are perpendicular to $[qq_x]$ at q (in D). □

Now we give the proof of our Main Theorem.

Proof of the Main Theorem. Note that, according to Theorem 0.3, we need only to show that

$$1 \le \sec_M \le 4. \tag{2-3}$$

We will apply induction on $\dim(N)$.

• $\dim(N) = 0$: By Lemma 2.2, N consists of a point, so M is homeomorphic to a sphere (because M consists of minimal geodesics between p and N). It follows from Theorem 1.3 that M is isometric to $\mathbb{S}^n\left(\frac{1}{2}\right)$ (which implies (2-3)).

• $\dim(N) = 1$: Note that N is a closed geodesic of length π. Let q_1 and q_2 be two antipodal points of N (i.e., $|q_1q_2| = \frac{\pi}{2}$). It follows that there are only two minimal geodesics between q_1 and q_2 (note that N is totally convex in M). Similarly, we consider $L \triangleq \{q_2\}^{=\pi/2}$ containing p and q_1, which is a totally geodesic submanifold in M of dimension > 0 by Lemma 2.2. Then similar to Lemma 2.5, we have that

$$\Uparrow_p^{q_2} = (\Sigma_p L)^{=\pi/2} = (\Sigma_{q_1} L)^{=\pi/2} = \Uparrow_{q_1}^{q_2}.$$

This implies that there are only two minimal geodesics between p and any $q \in N$ (by Lemma 2.5). It is then easy to see that $\sec_M \equiv 1$ by Corollary 2.4 (in fact, M is isometric to $\mathbb{RP}^2$ with the canonical metric).

• $\dim(N) > 1$: Since N is a complete totally geodesic submanifold in M (see Lemma 2.2), (0-1) implies that

$$\operatorname{diam}(N) = \operatorname{Inj}(N) = \tfrac{\pi}{2}. \tag{2-4}$$

By the inductive assumption on N, we have that

$$1 \le \sec_N \le 4. \tag{2-5}$$

On the other hand, we claim:

Claim. *For any $q \in N$,*

$$S(p,q) \triangleq \{\textit{the point on a minimal geodesic between } p \textit{ and } q\}$$

is totally geodesic in M and is isometric to $\mathbb{S}^m\left(\frac{1}{2}\right)$, where $m = \dim(M) - \dim(N)$.

Note that (2-3) is implied by the claim, (2-5), Lemma 2.5, Corollary 2.4 and Lemma 2.2. Hence, in the rest of the proof, we need only to verify the claim.

By (2-4), we can select $r \in N$ such that $|qr| = \frac{\pi}{2}$. Similarly, we consider $K \triangleq \{r\}^{=\pi/2}$ containing p and q, which is a complete totally geodesic submanifold in M with $\dim(K) > 0$; moreover, we have that

$$\Uparrow_p^r = (\Sigma_p K)^{=\pi/2},$$

and $\Uparrow_p^r$ is isometric to a unit sphere by Lemma 2.5. On the other hand, note that $\Uparrow_r^p$ is isometric to $\mathbb{S}^{m-1}(1)$ by Lemma 2.5, and that $\Uparrow_r^p$ is isometric to $\Uparrow_p^r$. Therefore, it is easy to see (again from Lemma 2.5 on K) that

$$\dim(K) = \dim(N).$$

Hence, by the inductive assumption on K (similar to on N), K is isometric to $\mathbb{S}^l\left(\frac{1}{2}\right)$ or a $\mathbb{KP}^l$ endowed with the canonical metric, which implies the claim above. □

References

[Berger 2003] M. Berger, *A panoramic view of Riemannian geometry*, Springer, Berlin, 2003. MR 2004h:53001 Zbl 1038.53002

[Besse 1978] A. L. Besse, *Manifolds all of whose geodesics are closed*, Ergebnisse der Mathematik und ihrer Grenzgebiete **93**, Springer, Berlin, 1978. MR 80c:53044 Zbl 0387.53010

[Bougas 2013] C. Bougas, *Blaschke's conjecture and Wiedersehen manifolds*, Master's thesis, University of Crete, 2013, http://www.math.uoc.gr:1080/erevna/diplomatikes/Mpougas_MDE.pdf.

[Cheeger and Ebin 1975] J. Cheeger and D. G. Ebin, *Comparison theorems in Riemannian geometry*, North-Holland Mathematical Library **9**, North-Holland, Amsterdam, 1975. MR 56 #16538 Zbl 0309.53035

[Cheeger and Gromoll 1980] J. Cheeger and D. Gromoll, "On the lower bound for the injectivity radius of 1/4-pinched Riemannian manifolds", *J. Differential Geom.* **15**:3 (1980), 437–442. MR 83g:53035 Zbl 0468.53035

[Gromoll and Grove 1987] D. Gromoll and K. Grove, "A generalization of Berger's rigidity theorem for positively curved manifolds", *Ann. Sci. École Norm. Sup.* (4) **20**:2 (1987), 227–239. MR 88k:53062 Zbl 0626.53032

[Gromoll and Grove 1988] D. Gromoll and K. Grove, "The low-dimensional metric foliations of Euclidean spheres", *J. Differential Geom.* **28**:1 (1988), 143–156. MR 89g:53052 Zbl 0683.53030

[Grove and Markvorsen 1995] K. Grove and S. Markvorsen, "New extremal problems for the Riemannian recognition program via Alexandrov geometry", *J. Amer. Math. Soc.* **8**:1 (1995), 1–28. MR 95j:53066 Zbl 0829.53033

[Petersen 1998] P. Petersen, *Riemannian geometry*, Graduate Texts in Mathematics **171**, Springer, New York, NY, 1998. MR 98m:53001 Zbl 0914.53001

[Rovenskii and Toponogov 1998] V. Rovenskii and V. Toponogov, "Great sphere foliations and manifolds with curvature bounded above", Appendix A, pp. 218–234 in *Foliations on Riemannian manifolds and submanifolds*, by V. Rovenskii, Birkhäuser, Boston, 1998.

[Shankar et al. 2005] K. Shankar, R. Spatzier, and B. Wilking, "Spherical rank rigidity and Blaschke manifolds", *Duke Math. J.* **128**:1 (2005), 65–81. MR 2006b:53056 Zbl 1082.53051

[Wilhelm 1996] F. Wilhelm, "The radius rigidity theorem for manifolds of positive curvature", *J. Differential Geom.* **44**:3 (1996), 634–665. MR 97m:53069 Zbl 0870.53032

Received June 11, 2015.

XIAOLE SU
SCHOOL OF MATHEMATICAL SCIENCES
BEIJING NORMAL UNIVERSITY
BEIJING, 100875
CHINA
suxiaole@bnu.edu.cn

HONGWEI SUN
SCHOOL OF MATHEMATICAL SCIENCES
CAPITAL NORMAL UNIVERSITY
BEIJING, 100037
CHINA
5598@cnu.edu.cn

YUSHENG WANG
SCHOOL OF MATHEMATICAL SCIENCES
BEIJING NORMAL UNIVERSITY
BEIJING, 100875
CHINA
wyusheng@bnu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 282, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.282.487

THE ROLE OF THE JACOBI IDENTITY IN SOLVING THE MAURER–CARTAN STRUCTURE EQUATION

ORI YUDILEVICH

We describe a method for solving the Maurer–Cartan structure equation associated with a Lie algebra that isolates the role of the Jacobi identity as an obstruction to integration. We show that the method naturally adapts to two other interesting situations: local symplectic realizations of Poisson structures, in which case our method sheds light on the role of the Poisson condition as an obstruction to realization; and the Maurer–Cartan structure equation associated with a Lie algebroid, in which case we obtain an explicit formula for a solution to the equation which generalizes the well-known formula in the case of Lie algebras.

Introduction

Realization problem for Lie algebras. Any Lie group G carries a canonical 1-form with values in the tangent space to the identity $\mathfrak{g}$,

$$\phi \in \Omega^1(G; \mathfrak{g}),$$

known as the *Maurer–Cartan form* of G. Actually, the Lie group structure is encoded, in some sense, in the 1-form and its properties; this is in fact Cartan's approach to Lie's infinitesimal theory. The two main properties of the Maurer–Cartan form are that it satisfies the so-called *Maurer–Cartan structure equation*[1] and that it is pointwise an isomorphism (the latter is often phrased as the property that the components of the 1-form with respect to some basis form a coframe). The Maurer–Cartan structure equation reveals a Lie algebra structure on $\mathfrak{g}$. Of course, the resulting Lie algebra is the same one obtained in the more common approach of using invariant vector fields.

Conversely, if we begin with an n-dimensional Lie algebra $\mathfrak{g}$, we can formulate the following problem, known as the *realization problem for Lie algebras*: find a

MSC2010: 22A22, 22E60, 53D17.

Keywords: Maurer–Cartan equation, symplectic realization, Maurer–Cartan form, Jacobi identity, structure equations, Lie algebroid, Lie algebra, Poisson structure.

[1]This paper deals with the classical Maurer–Cartan equation. To avoid confusion with the Maurer–Cartan equation that appears in the context of differential graded Lie algebras and other areas, we use the term *Maurer–Cartan structure equation*.

$\mathfrak{g}$-valued 1-form $\phi \in \Omega^1(U; \mathfrak{g})$ defined on some open neighborhood $U \subset \mathfrak{g}$ of the origin such that ϕ is pointwise an isomorphism and satisfies the Maurer–Cartan structure equation

$$d\phi + \tfrac{1}{2}[\phi, \phi] = 0. \tag{1}$$

A solution to the problem induces a local Lie group structure on some open subset of U (see [Greub et al. 1973, pp. 368–369]) and, therefore, we can think of this realization problem as the problem of locally integrating Lie algebras.

A solution to this problem is obtained by supposing that the Lie algebra integrates to a Lie group, and pulling back the canonical Maurer–Cartan form on the Lie group by the exponential map. This produces the following $\mathfrak{g}$-valued 1-form $\phi \in \Omega^1(\mathfrak{g}; \mathfrak{g})$, whose defining formula refers only to data coming from the Lie algebra and not from the Lie group:

$$\phi_x(y) = \int_0^1 e^{-t\,\mathrm{ad}_x} y\, dt, \qquad x \in \mathfrak{g},\ y \in T_x\mathfrak{g}. \tag{2}$$

This formula defines a solution to (1), as can be verified directly, and since it is equal to the identity at the origin, it is pointwise an isomorphism in a neighborhood of the origin. See [Duistermaat and Kolk 2000; Sternberg 2004] for more details.

Observe that neither (1) nor (2) rely on the Jacobi identity; they make perfect sense if we replace the Lie algebra with the weaker notion of a *pre-Lie algebra*, namely a vector space $\mathfrak{g}$ equipped with an antisymmetric bilinear map $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$. However, (2) is a solution of (1) if and only if $\mathfrak{g}$ is a Lie algebra, which is not difficult to show. This leads to the natural question, what is the precise role of the Jacobi identity? Put differently, at what point in the integration process does the Jacobi identity appear?

In Section 1 we present a two-step method for solving the realization problem for Lie algebras which answers this question. The method can be outlined as follows.

- Step 1 (Theorem 1.2): we formulate a weaker version of the realization problem, which admits a unique solution given any pre-Lie algebra.
- Step 2 (Theorem 1.4): we show that the solution of the weak realization problem is a solution of the complete realization problem if and only if the Jacobi identity is satisfied.

Two nice features of the method:

- Step 1 produces an explicit formula for a solution.
- Step 2 gives an explicit relation between the Maurer–Cartan structure equation and the Jacobi identity. Loosely speaking, one is the derivative of the other.

Similar phenomenon: Poisson realizations. There is a striking similarity between the phenomenon we just observed and a phenomenon that occurs in the story of symplectic realizations of Poisson manifolds. Recall that a Poisson manifold (M, π) is a manifold M equipped with a bivector π which satisfies the Poisson equation $[\pi, \pi] = 0$ (of course, the Poisson equation is equivalent to the condition that the induced Poisson bracket satisfy the Jacobi identity). A symplectic realization of a Poisson manifold (M, π) is a symplectic manifold (S, ω) together with a surjective submersion $p : S \to M$ that satisfies the equation

$$dp(\omega^{-1}) = \pi. \tag{3}$$

It was shown in [Crainic and Mărcuţ 2011] that for any Poisson manifold (M, π), a symplectic realization is explicitly given by the cotangent bundle T^*M equipped with the symplectic form

$$\omega = \int_0^1 (\varphi_t)^* \omega_{\text{can}} \, dt \tag{4}$$

together with the projection

$$p : T^*M \to M.$$

Here, ω_{can} is the canonical symplectic form and φ_t is the flow associated with a choice of a contravariant spray on T^*M. See [Crainic and Mărcuţ 2011] for more details.

As in the realization problem of Lie algebras, we observe that neither (3) nor (4) depend on the Poisson equation; they make perfect sense when replacing π with any bivector. And as before, there is the natural question as to the precise role of the Poisson equation in the existence of symplectic realizations, a question which was raised in [Crainic and Mărcuţ 2011] (see the last paragraph of that paper).

An explicit relation between the symplectic realization equation and the Maurer–Cartan structure equation was observed by Alan Weinstein [1983] in his pioneering work on Poisson manifolds. Weinstein showed that, locally, (3) is equivalent to a Maurer–Cartan structure equation associated with an infinite-dimensional Lie algebra, and exploited this to prove the existence of local symplectic realizations by using a heuristic argument to solve this Maurer–Cartan structure equation, producing an explicit local solution of the type (4).

In Section 3, we apply our method to solve the Maurer–Cartan structure equation which Weinstein formulated. As with Lie algebras, we do this by identifying a weaker version of the equation that admits a unique solution given any bivector, not necessarily Poisson, and proceed to show that the solution is a local symplectic realization if and only if the bivector satisfies the Poisson equation. We obtain an explicit relation between the Poisson equation and the symplectic realization condition, thus pinpointing the role of the Poisson equation in the problem of existence of local symplectic realizations.

The Lie algebroid case. In addition to local symplectic realizations, we believe that our method can be adapted to various other situations which generalize or resemble the classical Lie algebra case. One important generalization, which we treat in Section 3, is the realization problem of a Lie algebroid. Although extra difficulties do arise, it is remarkable that the procedure continues to work in this case, despite the fact that the simple-to-handle bilinear bracket of a Lie algebra is replaced by a more cumbersome bidifferential operator. This is largely facilitated by the presence of certain flows, known as infinitesimal flows, associated with time-dependent sections of the Lie algebroid.

As we noted in Step 1 above, our method produces an explicit solution. In the Lie algebra case, this is the well-known formula (2), whereas the formula we obtain in the Lie algebroid case does not appear in the literature to the best of our knowledge (see Theorem 3.3). Having this explicit formula at hand can prove to be useful; in particular, one can attempt to use it to explicitly integrate Lie algebroids locally (as an indication of feasibility, in [Coste et al. 1987] a symplectic realization of a Poisson manifold was used to integrate the associated Lie algebroid to a local symplectic groupoid; see also the discussion starting at the bottom of page 504).

Final remark. We end the introduction with a historical remark and briefly describe our motivation for reopening this classical problem. The Maurer–Cartan structure equation originates in the work of Élie Cartan [1904; 1937] under the name of "structure equations". In his work on Lie pseudogroups, Cartan associates the equation with a Lie pseudogroup, and subsequently extracts out of the equation the Lie pseudogroup's "structure functions", i.e., its infinitesimal data. The reverse direction, the problem of finding and classifying the solutions to the structure equations associated with given infinitesimal data, is known as the realization problem, two special cases of which we discussed above (the Lie algebra case and the Lie algebroid case).

This work arose as part of a larger project aimed at understanding Cartan's original work on Lie pseudogroups in a global, more geometric and coordinate-free fashion, and in particular, the realization problem. Since Cartan's realization problem involves infinitesimal structures that fail to satisfy the Jacobi identity, we first tried to understand the role of the Jacobi identity in the integration process of structures for which the Jacobi identity is satisfied, namely Lie algebras and Lie algebroids. The method and the results that we came across and that we are presenting here seemed to have relevance beyond the realization problem itself, and we, therefore, decided to present it in an independent fashion.

1. The Maurer–Cartan structure equation of a Lie algebra

In this section, we present the two-step method for solving the realization problem for a Lie algebra which was outlined in the Introduction. Let us first recall the necessary definitions.

Definition 1.1. A *pre-Lie algebra* is a vector space $\mathfrak{g}$ equipped with an antisymmetric bilinear map $[\cdot,\cdot]:\mathfrak{g}\times\mathfrak{g}\to\mathfrak{g}$. A *Lie algebra* is a pre-Lie algebra that satisfies the Jacobi identity:

$$[[x,y],z]+[[y,z],x]+[[z,x],y]=0 \quad \forall\, x,y,z\in\mathfrak{g}.$$

Associated with a pre-Lie algebra is the *adjoint map* $\mathrm{ad}:\mathfrak{g}\to\mathrm{End}(\mathfrak{g})$, where $\mathrm{ad}_x(y)=[x,y]$, and the *Jacobiator*

$$\text{(5)}\quad \mathrm{Jac}\in\mathrm{Hom}(\Lambda^3\mathfrak{g},\mathfrak{g}),\qquad \mathrm{Jac}(x,y,z)=[[x,y],z]+[[y,z],x]+[[z,x],y].$$

The space of $\mathfrak{g}$-valued differential forms on $\mathfrak{g}$ is denoted by $\Omega^*(\mathfrak{g};\mathfrak{g})$. This space is equipped with the de Rham differential $d:\Omega^*(\mathfrak{g};\mathfrak{g})\to\Omega^{*+1}(\mathfrak{g};\mathfrak{g})$ and with a bracket, $[\cdot,\cdot]:\Omega^p(\mathfrak{g};\mathfrak{g})\times\Omega^q(\mathfrak{g};\mathfrak{g})\to\Omega^{p+q}(\mathfrak{g};\mathfrak{g})$, that plays the role of the wedge product on $\mathfrak{g}$-valued forms and is defined by the analogous formula:

$$\text{(6)}\quad [\omega,\eta](X_1,\dots,X_{p+q}) = \sum_{\sigma\in S_{p,q}} \mathrm{sgn}(\sigma)[\omega(X_{\sigma(1)},\dots,X_{\sigma(p)}),\eta(X_{\sigma(p+1)},\dots,X_{\sigma(p+q)})],$$

where $S_{p,q}$ is the set of (p,q)-shuffles.

Of course, given any open subset $U\subset\mathfrak{g}$, we also have the space of $\mathfrak{g}$-valued forms $\Omega^*(U;\mathfrak{g})$ on U equipped with a differential and a bracket, defined in the same manner. Given any $\phi\in\Omega^1(U;\mathfrak{g})$, the *Maurer–Cartan 2-form* associated with ϕ is defined by

$$\mathrm{MC}_\phi := d\phi+\tfrac{1}{2}[\phi,\phi]\in\Omega^2(U;\mathfrak{g}),$$

and the *Maurer–Cartan structure equation* is

$$\mathrm{MC}_\phi=0,$$

or more explicitly,

$$(\mathrm{MC}_\phi)_x(y,z)=0 \quad \forall\, x\in U,\ y,z\in\mathfrak{g}.$$

Note that in the last equation, and throughout the paper, we identify the tangent spaces of a vector space with the vector space itself without further mention.

Recall the *realization problem for Lie algebras*: find a 1-form $\phi\in\Omega^1(U;\mathfrak{g})$ on some open neighborhood $U\subset\mathfrak{g}$ of the origin such that ϕ is pointwise an isomorphism and satisfies the Maurer–Cartan structure equation.

We now present our method for solving this realization problem.

Step 1. We show that a weaker version of the realization problem admits a solution given any pre-Lie algebra. We accomplish this by imposing a boundary condition which transforms the equation into a simple ODE that can be easily solved.

Theorem 1.2. *Given any pre-Lie algebra* $\mathfrak{g}$, *the equation*

$$(\mathrm{MC}_\phi)_x(x, y) = 0 \quad \forall\, x, y \in \mathfrak{g} \tag{7}$$

admits a solution in $\Omega^1(\mathfrak{g}; \mathfrak{g})$ *which is pointwise an isomorphism at the origin* (*and thus on some open neighborhood of the origin*). *Moreover*, *if we impose the boundary condition*

$$\phi_x(x) = x \quad \forall\, x \in \mathfrak{g}, \tag{8}$$

then the solution is unique and is given by the formula

$$\phi_x(y) = \int_0^1 e^{-t\,\mathrm{ad}_x}\, y\, dt. \tag{9}$$

Remark 1.3. To get a geometric feel for the equations, note that (7) is the restriction of the Maurer–Cartan structure equation to all two-dimensional subspaces of $\mathfrak{g}$, and (8) is the condition that ϕ restricts to the identity on all one-dimensional subspaces.

Proof. First note that (8) implies that $\phi_0 = \mathrm{id}$, and in particular, ϕ is pointwise an isomorphism at the origin.

Let $\phi \in \Omega^1(\mathfrak{g}; \mathfrak{g})$ be a solution of (7) and (8). We will show that ϕ must be of the form given by (9), which implies uniqueness. Conversely, as we will explain at the end of the proof, reading the steps in the reverse direction will imply that (9) is a solution, thus proving existence.

By linearity, (7) and (8) are equivalent to $(\mathrm{MC}_\phi)_{tx}(x, y) = 0$ and $\phi_{tx}(x) = x$ for all $t \in (0, 1)$ and $x, y \in \mathfrak{g}$. In particular, by continuity, this implies that

$$(\mathrm{MC}_\phi)_0(x, y) = 0 \quad \text{and} \quad \phi_0(x) = x \qquad \forall\, x, y \in \mathfrak{g}. \tag{10}$$

Fix $x, y \in \mathfrak{g}$. The solution ϕ satisfies

$$\big(d\phi + \tfrac{1}{2}[\phi, \phi]\big)_{tx}(x, ty) = 0 \tag{11}$$

for all $t \in (0, 1)$.

To compute $(d\phi)_{tx}(x, ty)$, consider the map $f : (0,1) \times (-\delta, \delta) \to \mathfrak{g}$, where $f(t,\epsilon) = t(x+\epsilon y)$. Then

$$\begin{aligned}
(d\phi)_{tx}(x, ty) &= (f^* d\phi)_{(t,0)}\Big(\frac{\partial}{\partial t}, \frac{\partial}{\partial \epsilon}\Big) \\
&= (d f^*\phi)_{(t,0)}\Big(\frac{\partial}{\partial t}, \frac{\partial}{\partial \epsilon}\Big) \\
&= \frac{\partial}{\partial t}\Big((f^*\phi)\Big(\frac{\partial}{\partial \epsilon}\Big)\Big)\Big|_{(t,0)} - \frac{\partial}{\partial \epsilon}\Big((f^*\phi)\Big(\frac{\partial}{\partial t}\Big)\Big)\Big|_{(t,0)} \\
&= \frac{\partial}{\partial t}(\phi_{t(x+\epsilon y)}(ty))\Big|_{(t,0)} - \frac{\partial}{\partial \epsilon}(\phi_{t(x+\epsilon y)}(x+\epsilon y))\Big|_{(t,0)} \\
&= \frac{\partial}{\partial t}(\phi_{tx}(ty)) - y,
\end{aligned}$$

where, in the last equality, we have used that (8) implies $\phi_{t(x+\epsilon y)}(x+\epsilon y) = x+\epsilon y$.

To compute $\big(\frac{1}{2}[\phi,\phi]\big)_{tx}(x, ty)$, we use (8) again:

$$\big(\tfrac{1}{2}[\phi,\phi]\big)_{tx}(x, ty) = [\phi_{tx}(x), \phi_{tx}(ty)] = [x, \phi_{tx}(ty)] = \mathrm{ad}_x(\phi_{tx}(ty)).$$

Thus for a ϕ that satisfies (8), equation (11) is equivalent to

$$\frac{\partial}{\partial t}(\phi_{tx}(ty)) - y + \mathrm{ad}_x(\phi_{tx}(ty)) = 0,$$

which is equivalent to

$$\frac{\partial}{\partial t}(e^{t\,\mathrm{ad}_x}\phi_{tx}(ty)) = e^{t\,\mathrm{ad}_x} y. \tag{12}$$

Integrating from 0 to t',

$$\phi_{t'x}(t'y) = \int_0^{t'} e^{(t-t')\,\mathrm{ad}_x} y\,dt = \int_0^1 e^{-t\,\mathrm{ad}_{t'x}}(t'y)\,dt. \tag{13}$$

Setting $t' = 1$ proves that ϕ coincides with (9) .

Next, we show that ϕ defined by (9) is a solution. Note that $\phi_x(x) = \int_0^1 e^{-t\,\mathrm{ad}_x} x\,dt = \int_0^1 x\,dt = x$, and thus (8) is satisfied. Equation (9) is equivalent to (13), which is a solution of (12), and since ϕ satisfies (8), it is a solution of (11). In particular, setting $t = 1$ implies that $(\mathrm{MC}_\phi)_x(x, y) = 0$, and thus (7) is satisfied. □

Step 2. By obtaining explicit equations relating the Maurer–Cartan 2-form with the Jacobiator, we show that the solution obtained in the previous step is a solution of the Maurer–Cartan structure equation if and only if the Jacobiator vanishes.

Theorem 1.4. *Let* $\mathfrak{g}$ *be a pre-Lie algebra and* $\phi \in \Omega^1(\mathfrak{g};\mathfrak{g})$ *the solution of* (7) *and* (8). *Then*

$$\mathrm{MC}_\phi = 0 \iff \mathrm{Jac} = 0,$$

or, *more precisely*,

$$\mathrm{Jac}(x, y, z) = -3\,\frac{d}{dt}(\mathrm{MC}_\phi)_{tx}(y, z)\Big|_{t=0}, \tag{14}$$

$$(\mathrm{MC}_\phi)_x(y, z) = -\int_0^1 e^{(t-1)\,\mathrm{ad}_x}\,\mathrm{Jac}(x, \phi_{tx}(ty), \phi_{tx}(tz))\,dt. \tag{15}$$

Proof. Equations (14) and (15) imply that $\mathrm{MC}_\phi = 0$ if and only if $\mathrm{Jac} = 0$. Let us derive these equations. Fix $x, y, z \in \mathfrak{g}$. We will compute

$$d(\mathrm{MC}_\phi)_{tx}(x, ty, tz), \tag{16}$$

with $t \in (0, 1)$, in two different ways.

1. Consider the map $f : (0, 1) \times (-\delta, \delta)^2 \to \mathfrak{g}$, $f(t, \epsilon, \epsilon') = t(x + \epsilon y + \epsilon' z)$. Then

$$\begin{aligned}(d\mathrm{MC}_\phi)_{tx}(x, ty, tz) &= (f^* d\mathrm{MC}_\phi)_{(t,0,0)}\Big(\frac{\partial}{\partial t}, \frac{\partial}{\partial \epsilon}, \frac{\partial}{\partial \epsilon'}\Big)\\ &= (df^*\mathrm{MC}_\phi)_{(t,0,0)}\Big(\frac{\partial}{\partial t}, \frac{\partial}{\partial \epsilon}, \frac{\partial}{\partial \epsilon'}\Big)\\ &= \frac{d}{dt}(\mathrm{MC}_\phi)_{tx}(ty, tz).\end{aligned}$$

In the last equality, terms containing $(\mathrm{MC}_\phi)_{tx}(x, ty)$ and $(\mathrm{MC}_\phi)_{tx}(x, tz)$ vanish by (7).

2. On the other hand,

$$\begin{aligned}(d\mathrm{MC}_\phi)_{tx}(x, ty, tz) &= \big(d\tfrac{1}{2}[\phi, \phi]\big)_{tx}(x, ty, tz)\\ &= ([d\phi, \phi])_{tx}(x, ty, tz)\\ &= [(d\phi)_{tx}(x, ty), \phi_{tx}(tz)] + [(d\phi)_{tx}(tz, x), \phi_{tx}(ty)]\\ &\qquad + [(d\phi)_{tx}(ty, tz), \phi_{tx}(x)]\\ &= -[[x, \phi_{tx}(ty)], \phi_{tx}(tz)] - [[\phi_{tx}(tz), x], \phi_{tx}(ty)]\\ &\qquad + [(\mathrm{MC}_\phi)_{tx}(ty, tz) - [\phi_{tx}(ty), \phi_{tx}(tz)], x]\\ &= -[x, (\mathrm{MC}_\phi)_{tx}(ty, tz)] - \mathrm{Jac}(x, \phi_{tx}(ty), \phi_{tx}(tz)).\end{aligned}$$

In the fourth equality, we have used (7) and (8). In particular, (8) implies that $(d\phi)_{tx}(x, y) + [\phi_{tx}(x), \phi_{tx}(y)] = 0$ and $(d\phi)_{tx}(x, z) + [\phi_{tx}(x), \phi_{tx}(z)] = 0$.

Then (16) becomes

$$\mathrm{Jac}(x, \phi_{tx}(ty), \phi_{tx}(tz)) = -\Big(\frac{d}{dt} + \mathrm{ad}_x\Big)(\mathrm{MC}_\phi)_{tx}(ty, tz),$$

or equivalently,

$$e^{t\,\mathrm{ad}_x}\,\mathrm{Jac}(x, \phi_{tx}(ty), \phi_{tx}(tz)) = -\frac{d}{dt}(e^{t\,\mathrm{ad}_x} n(\mathrm{MC}_\phi)_{tx}(ty, tz)). \tag{17}$$

Integrating from 0 to 1 produces (15), while multiplying both sides of the equation by $1/t^2$, taking the limit as $t \to 0$ and using the fact that $(\mathrm{MC}_\phi)_0(y, z) = 0$ (see (10)) produces (14). $\square$

Remark 1.5. The method we present here was inspired by the method used in [Sternberg 2004, Sections 1.3–1.5] to compute the differential of the exponential map of a Lie group and to derive the Baker–Campbell–Hausdorff formula of a Lie algebra.

2. The Maurer–Cartan structure equation and local symplectic realizations of Poisson structures

In this section, we apply the method from the previous section to the problem of existence of symplectic realizations of Poisson manifolds. The role of the Poisson equation becomes manifest, in the same way that the role of the Jacobi identity was made manifest in the Lie algebra case.

Definition 2.1. A *pre-Poisson manifold* (M, π) is a manifold M together with a choice of a bivector field $\pi \in \mathfrak{X}^2(M)$. A *Poisson manifold* (M, π) is a pre-Poisson manifold with the extra condition that π satisfies the *Poisson equation* $[\pi, \pi] = 0$ (where $[\,\cdot\,,\cdot\,]$ is the Schouten–Nijenhuis bracket).

Equivalently, a pre-Poisson manifold is a manifold M equipped with an $\mathbb{R}$-bilinear antisymmetric operation $\{\,,\,\} : C^\infty(M) \times C^\infty(M) \to C^\infty(M)$ (called the Poisson bracket) that satisfies the Leibniz identity, $\{fg, h\} = f\{g, h\} + \{f, h\}g$ for all $f, g, h \in C^\infty(M)$. A bivector π induces a bracket by $\{f, g\}(m) = \pi_m(df, dg)$ for all $m \in M$, $f, g \in C^\infty(M)$, and vice versa. The Poisson equation is equivalent to the Jacobi identity, i.e., to the condition $\mathrm{Jac} = 0$, where Jac is the Jacobiator associated with $\{\,,\,\}$ (defined as in the previous section).

By the Leibniz identity, a function $f \in C^\infty(M)$ induces a vector field $X_f \in \mathfrak{X}(M)$, the *Hamiltonian vector field* associated with f, by the condition $X_f(g) = \{f, g\}$ for all $g \in C^\infty(M)$, or equivalently, $X_f(g) = \pi(df, dg)$ for all $g \in C^\infty(M)$.

Poisson manifolds can be localized; i.e., if (M, π) is a Poisson manifold and $U \subset M$ is an open subset, then $(U, \pi|_U)$ is a Poisson manifold.

A *symplectic realization* of a Poisson manifold (M, π) is a symplectic manifold (S, ω) together with a surjective submersion $p : S \to M$ such that p is a Poisson map; i.e., the bivector ω^{-1} induced by the symplectic form ω is p-projectable to the bivector π. That is to say,

$$dp(\omega^{-1}) = \pi.$$

A *local symplectic realization* of (M, π) around a point $m \in M$ is a symplectic realization of $(U, \pi|_U)$, where U is some open neighborhood of m.

In the problem of existence of local symplectic realizations it is enough to consider Poisson manifolds of the type $(\mathcal{O}, \pi)$, where

$$\mathcal{O} \subset V$$

is an open subset of a vector space V. The following proposition was proven by Alan Weinstein [1983, Section 9]. To be more precise, Weinstein proved it for the case that $(\mathcal{O}, \pi)$ is a Poisson manifold; however, the arguments do not rely on the Jacobi identity and the proposition also holds for the case that $(\mathcal{O}, \pi)$ is a pre-Poisson manifold.

Proposition 2.2. *Let $(\mathcal{O}, \pi)$ be a pre-Poisson manifold. Let $\phi \in \Omega^1(V^*; C^\infty(\mathcal{O}))$ be defined by*

$$\phi_\xi(\zeta) = \int_0^1 (\varphi_{X_\xi}^{-t})^* \zeta \, dt \quad \forall \xi, \zeta \in V^*. \tag{18}$$

Here ξ and ζ are interpreted as linear functionals on V, X_ξ is the corresponding Hamiltonian vector field and φ_{X_ξ} is its flow.

Let $\tilde{\phi} \in \Omega^1(\mathcal{O} \times V^*)$ be the induced 1-form on $\mathcal{O} \times V^* = T^*\mathcal{O}$ defined by $\tilde{\phi}_{(x,\xi)}(y, \zeta) := \phi_\xi(\zeta)(x)$.

Then, the 2-form $d\tilde{\phi}$ is symplectic on some neighborhood $U \subset \mathcal{O} \times V^*$ of the zero section and, writing $p : \mathcal{O} \times V^* \to \mathcal{O}$ for the projection,

$$p|_U : (U, d\tilde{\phi}) \to (\mathcal{O}, \pi) \text{ is a symplectic realization } \iff d\phi + \tfrac{1}{2}\{\phi, \phi\} = 0.$$

Remark 2.3. The 1-form ϕ defined by (18) and the induced 1-form $\tilde{\phi}$ are only well defined on some open neighborhood of the zero section of $\mathcal{O} \times V^*$, namely on all points (x, ξ) such that $\varphi_{X_\xi}(x)$ is defined up to time 1. This does not pose a problem, since, in the end, we are only interested in the symplectic form $d\tilde{\phi}$ in some neighborhood of the zero section.

Weinstein's remarkable observation was that the symplectic realization condition can be locally rephrased as a Maurer–Cartan structure equation. This equation lives in the space $\Omega^*(V^*; C^\infty(\mathcal{O}))$ consisting of differential forms with values in $C^\infty(\mathcal{O})$, where a 1-form $\phi \in \Omega^1(V^*; C^\infty(\mathcal{O}))$ is smooth if the map $\mathcal{O} \times V^* \to \mathbb{R}$, $(x, \xi) \mapsto \phi_\xi(\zeta)(x)$, is smooth for all $\zeta \in V^*$, and similarly for higher degree forms. This space is equipped with the de Rham differential d defined as usual, and a bracket $\{\,,\}$ defined as in (6) (with the Lie bracket replaced by the Poisson bracket); thus, one can make sense of the Maurer–Cartan 2-form associated with a 1-form $\phi \in \Omega^1(V^*; C^\infty(\mathcal{O}))$:

$$\mathrm{MC}_\phi := d\phi + \tfrac{1}{2}\{\phi, \phi\} \in \Omega^2(V^*; C^\infty(\mathcal{O})).$$

Weinstein proceeded to show that if $(\mathcal{O}, \pi)$ is a Poisson manifold, then the 1-form given by (18) satisfies the Maurer–Cartan structure equation, thus proving the existence of local symplectic realizations. Of course, the fact that the Poisson bracket satisfies the Jacobi identity is used in the proof, but its precise role is somewhat obscure, appearing as a "mere step" in the calculation (see [Weinstein 1983, p. 547]).

The following two theorems shed further light on the role of the Jacobi identity as an obstruction in this problem. The first of the two theorems, an analog of Step 1 of the previous section, demonstrates how close $d\tilde{\phi}$ induced by (18) is from being a symplectic realization, regardless of the Jacobi identity.

Theorem 2.4. *Let* $(\mathcal{O}, \pi)$ *be a pre-Poisson manifold. The* 1*-form*

$$\phi \in \Omega^1(V^*; C^\infty(\mathcal{O}))$$

defined by (18) *satisfies the equation*

$$(\mathrm{MC}_\phi)_\xi(\xi, \zeta) = 0 \quad \forall \xi, \zeta \in V^*. \tag{19}$$

Moreover, it is the unique solution of (19) *together with the boundary condition*

$$\phi_\xi(\xi) = \xi \quad \forall \xi \in V^*. \tag{20}$$

Proof. The proof is essentially the same as the proof of Theorem 1.2. One must only make the following exchanges:

- $\mathfrak{g}$ with V^* (and accordingly x, y with ξ, ζ),
- the Lie bracket $[\ ,\ ]$ with the Poisson bracket $\{\ ,\ \}$,
- $e^{t\,\mathrm{ad}_\xi}$ with $(\varphi^t_{X_\xi})^*$,

and while making the last of the three adjustments, one notes that derivatives of matrix-valued functions of t become derivatives of flows. □

The next theorem, an analog of Step 2 of the previous section, gives an explicit relation between Jac and MC_ϕ which translates into a precise relation between the failure of the Poisson equation and the failure of $d\tilde{\phi}$ to be a symplectic realization. Of course, it follows that if the Poisson equation is satisfied, then $d\tilde{\phi}$ is a symplectic realization.

Theorem 2.5. *Let* $\phi \in \Omega^1(V^*; C^\infty(\mathcal{O}))$ *be a solution to* (19) *and* (20). *Then*

$$\mathrm{Jac} = 0 \iff \mathrm{MC}_\phi = 0,$$

or more precisely,

$$\mathrm{Jac}(\xi, \zeta, \eta) = -3\,\frac{d}{dt}(\mathrm{MC}_\phi)_{t\xi}(\zeta, \eta)\Big|_{t=0},$$

$$(\mathrm{MC}_\phi)_\xi(\zeta, \eta) = -\int_0^1 (\varphi^{t-1}_{X_\xi})^*\,\mathrm{Jac}(\xi, \phi_{t\xi}(t\zeta), \phi_{t\xi}(t\eta))\,dt.$$

Proof. The proof is essentially the same as the proof of Theorem 1.4 after making the necessary adjustments as in the proof of the previous theorem, and using the fact that by the Leibniz identity, the vanishing of the Jacobiator on linear functions implies that it vanishes. □

Remark 2.6. Theorems 1.2 and 1.4 are in fact special cases of Theorems 2.4 and 2.5. Recall that a linear Poisson structure on the vector space $\mathfrak{g}^*$ is a Poisson bracket on $C^\infty(\mathfrak{g}^*)$ satisfying the property that it restricts to a Lie bracket on the linear functions $\mathfrak{g} \subset C^\infty(\mathfrak{g}^*)$. This defines a one-to-one correspondence between linear Poisson structures on $\mathfrak{g}^*$ and Lie algebra structures on $\mathfrak{g}$. In the case of linear Poisson structures, the Hamiltonian vector field on $\mathfrak{g}^*$ associated with an element $x \in \mathfrak{g} = (\mathfrak{g}^*)^*$ is simply the transpose $(\mathrm{ad}_x)^*$ of the linear map $\mathrm{ad}_x : \mathfrak{g} \to \mathfrak{g}$. The flow of $(\mathrm{ad}_x)^*$ is the transpose of the linear map $e^{t\,\mathrm{ad}_x}$, and the pullback by the flow is precisely $e^{t\,\mathrm{ad}_x}$. This implies that the solution (18) takes values in $\mathfrak{g} \subset C^\infty(\mathfrak{g}^*)$, and it follows that Theorems 2.4 and 2.5 for linear Poisson structures coincide with Theorems 1.2 and 1.4.

3. The Maurer–Cartan structure equation of a Lie algebroid

In this section, we generalize our method from the Lie algebra case to the Lie algebroid case. We will begin by recalling the basic definitions and discussing the realization problem for Lie algebroids, after which we will state and prove Theorems 3.3 and 3.5, which generalize Theorems 1.2 and 1.4.

Definition 3.1. A *pre-Lie algebroid* $A \xrightarrow{\pi} M$ is a vector bundle A over M equipped with a vector bundle map (the "anchor") $\rho : A \to TM$ and an antisymmetric bilinear map (the "bracket") $[\,\cdot\,,\cdot\,] : \Gamma(A) \times \Gamma(A) \to \Gamma(A)$ satisfying

$$\begin{aligned} &[\alpha, f\beta] = f[\alpha,\beta] + \mathcal{L}_{\rho(\alpha)}(f)\beta && \forall\, \alpha,\beta \in \Gamma(A),\ f \in C^\infty(M), \\ &\rho([\alpha,\beta]) = [\rho(\alpha),\rho(\beta)] && \forall\, \alpha,\beta \in \Gamma(A). \end{aligned}$$

A pre-Lie algebroid $A \to M$ is called a *Lie algebroid* if it further satisfies the Jacobi identity

$$[[\alpha,\beta],\gamma] + [[\beta,\gamma],\alpha] + [[\gamma,\alpha],\beta] = 0 \quad \forall\, \alpha,\beta,\gamma \in \Gamma(A).$$

Associated with a pre-Lie algebroid is the *Jacobiator* tensor $\mathrm{Jac} \in \mathrm{Hom}(\Lambda^3 A, A)$, defined at the level of sections by

$$\mathrm{Jac}(\alpha,\beta,\gamma) = [[\alpha,\beta],\gamma] + [[\beta,\gamma],\alpha] + [[\gamma,\alpha],\beta] \quad \forall\, \alpha,\beta,\gamma \in \Gamma(A),$$

and easily checked to be $C^\infty(M)$-linear in all slots.

The notions of A-connections, A-paths, geodesics and infinitesimal flows that appear in the context of Lie algebroids remain unchanged when we give up on the Jacobi identity and pass to pre-Lie algebroids. We will assume familiarity with these notions, and otherwise refer the reader to the Appendix (and to [Crainic and Fernandes 2003] for more details).

Let $A \to M$ be a pre-Lie algebroid equipped with an A-connection $\overline{\nabla}$. To every point $a \in A$ we associate the unique maximal geodesic $g_a : I_a \to A$ that satisfies $g_a(0) = a$. We denote its base curve by $\gamma_a : I_a \to M$. Let $A_0 \subset A$ be a neighborhood of the zero section such that g_a is defined up to at least time 1 for all $a \in A_0$. On A_0 we have the exponential map $\exp : A_0 \to A,\ a \mapsto g_a(1)$, and the target map $\tau = \pi \circ \exp : A_0 \to M$. Let $\Omega^*_\pi(A_0; \tau^*A)$ be the space of foliated differential forms (foliated with respect to the foliation by π-fibers) with values in τ^*A. Throughout this section we will use the canonical identification between the vertical bundle of A_0 and the pullback of A to A_0, i.e., $T_a A_0 \cong A_x$ for all $a \in (A_0)_x$. Thus, given a 1-form $\phi \in \Omega^1_\pi(A_0; \tau^*A)$, we will write $\phi_a(b)$ with $a \in (A_0)_x$, $b \in A_x$.

A 1-form $\phi \in \Omega^1_\pi(A_0; \tau^*A)$ is said to be *anchored* if $\rho \circ \phi = d\tau$. Given a vector bundle connection $\nabla : \mathfrak{X}(M) \times \Gamma(A) \to \Gamma(A)$, we define the *Maurer–Cartan 2-form* associated with an anchored 1-form $\phi \in \Omega^1_\pi(A_0; \tau^*A)$ to be

$$\mathrm{MC}_\phi := d_{\tau^*\nabla}\phi + \tfrac{1}{2}[\phi, \phi]_\nabla \in \Omega^2_\pi(A_0; \tau^*A).$$

The differential-like map $d_{\tau^*\nabla}$ and bracket on $\Omega^*_\pi(A_0; \tau^*A)$ are defined in the usual way (see the Appendix). The anchored condition implies that MC_ϕ is independent of the choice of connection (Proposition A.2). The auxiliary connection ∇ should not be confused with the A-connection $\overline{\nabla}$, which is part of the data we fix.

Of course, given any open subset $U \subset A_0$, we have the space of forms $\Omega^*_\pi(U; \tau^*A)$ equipped with a differential-like operator and a bracket in the same manner, and anchored 1-forms have associated Maurer–Cartan 2-forms. The *realization problem for Lie algebroids* can now be stated: find an anchored 1-form $\phi \in \Omega^1_\pi(U; \tau^*A)$ on some open neighborhood of the zero section of A_0 such that ϕ is pointwise an isomorphism and satisfies the *Maurer–Cartan structure equation*:

$$\mathrm{MC}_\phi = 0. \tag{21}$$

Remark 3.2. A solution of the Maurer–Cartan structure equation can also be interpreted as a Lie algebroid map: a 1-form $\phi \in \Omega^1_\pi(A_0; \tau^*A)$ can be viewed as a vector bundle map from the Lie algebroid $T^\pi A_0 \to A_0$ (the vertical bundle, a Lie subalgebroid of $TA_0 \to A_0$) to the Lie algebroid $A \to M$ covering τ, the anchored condition on ϕ is equivalent to the vector bundle map commuting with the anchors, and ϕ satisfies the Maurer–Cartan structure equation if and only if the vector bundle map is a Lie algebroid map (see [Crainic and Fernandes 2011] or [Fernandes and Struchiner 2014] for more details). From this point of view, the Maurer–Cartan structure equation is a special case of the *generalized Maurer–Cartan equation* for vector bundle maps between Lie algebroids which commute with the anchors studied in [Fernandes and Struchiner 2014, Section 3.2].

As in the case of Lie algebras (see the Introduction), one can find a solution to the realization problem by assuming that the Lie algebroid integrates to a Lie groupoid and pulling back the canonical Maurer–Cartan 1-form on the Lie groupoid by the exponential map. The resulting formula will not depend on the Lie groupoid, and one can verify directly that the formula is indeed a solution, and, therefore, not have to require that the Lie algebroid be integrable.

Let us explain this in more detail. Let $\mathcal{G} \rightrightarrows M$ be a Lie groupoid with source/target map s/t. The canonical Maurer–Cartan 1-form $\phi_{\mathrm{MC}} \in \Omega^1_s(\mathcal{G}; t^*A)$ is a foliated differential 1-form on $\mathcal{G}$ (foliated with respect to the foliation by s-fibers) with values in t^*A. It is defined precisely as in the case of Lie groups,

$$(\phi_{\mathrm{MC}})_g = (dR_{g^{-1}})_g \quad \forall\, g \in \mathcal{G}, \tag{22}$$

the difference being that the right multiplication map $R_{g^{-1}}$ is only defined on $s^{-1}(s(g))$. For this reason, the resulting form is foliated. The Maurer–Cartan form satisfies the anchored property $\rho((\phi_{\mathrm{MC}})_g(X)) = (dt)_g(X)$ and the Maurer–Cartan structure equation

$$d_{t^*\nabla}\phi_{\mathrm{MC}} + \tfrac{1}{2}[\phi_{\mathrm{MC}}, \phi_{\mathrm{MC}}]_\nabla = 0$$

(for more details, see [Fernandes and Struchiner 2014, Section 4]).

The exponential map $\mathrm{Exp} := \mathrm{Exp}_{\overline{\nabla}} : A_0 \to \mathcal{G}$ on a Lie groupoid requires a choice of an A-connection $\overline{\nabla}$ on A, where A_0 is as above. Such a choice induces a normal connection on each s-fiber and the exponential map is then defined in the usual way. This choice of an A-connection also gives rise to an exponential on the Lie algebroid, as we saw above, and the two satisfy the relations

$$\exp(a) = (dR_{\mathrm{Exp}(a)^{-1}})_{\mathrm{Exp}(a)} \frac{d}{dt} \mathrm{Exp}(ta)\Big|_{t=1}, \tag{23}$$

$$\pi \circ \exp = t \circ \mathrm{Exp}, \tag{24}$$

$$\pi = s \circ \mathrm{Exp}.$$

If we pull back the Maurer–Cartan form by the exponential map, the resulting form will be an element of $\Omega^1_\pi(A_0; \tau^*A)$. It will be anchored as a result of (24). It is now not difficult to verify that the fact that ϕ_{MC} satisfies the Maurer–Cartan structure equation on the Lie groupoid implies that $\mathrm{Exp}^*\phi_{\mathrm{MC}}$ satisfies the Maurer–Cartan structure equation on the Lie algebroid, i.e., satisfies (21).

In the following two theorems we will obtain a solution by taking a different path, namely by generalizing our method from Section 1. The first theorem is a generalization of Step 1: a weaker version of the realization problem which admits a unique solution for any pre-Lie algebroid. The theorem gives an explicit formula for a solution to the realization problem of Lie algebroids. In Corollary 3.4 we show that our solution coincides with $\mathrm{Exp}^*\phi_{\mathrm{MC}}$.

Theorem 3.3. *Let $A \to M$ be a pre-Lie algebroid equipped with an A-connection $\overline{\overline{\nabla}}$. The equations*

$$(\mathrm{MC}_\phi)_a(a, b) = 0 \qquad \forall\, x \in M,\ a \in (A_0)_x,\ b \in A_x, \tag{25}$$

$$\rho \circ \phi = d\tau \tag{26}$$

admit a solution in $\Omega^1_\pi(A_0; \tau^* A)$ which is pointwise an isomorphism on a small enough neighborhood of the zero section of A_0. Moreover, if we impose the boundary condition

$$\phi_a(a) = \exp(a) \quad \forall\, a \in A_0, \tag{27}$$

then the solution is unique and can be described as follows. Let

$$\xi : [0, 1] \times (-\delta, \delta) \times M \to A$$

be a smooth map such that $\xi^t_\epsilon = \xi(t, \epsilon, \cdot)$ is a section of A and $\xi^t_\epsilon(\gamma_{a+\epsilon b}(t)) = g_{a+\epsilon b}(t)$ for all $(t, \epsilon) \in [0, 1] \times (-\delta, \delta)$, and let ψ_{ξ_0} be the infinitesimal flow associated with the time-dependent section ξ_0 (see the Appendix). The solution is given by

$$\phi_a(b) = \int_0^1 \psi^{1,t}_{\xi_0} \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \xi^t_\epsilon(\gamma_a(t))\, dt. \tag{28}$$

Proof. Equation (27) implies that a solution ϕ is equal to the identity on the zero section of A and thus pointwise an isomorphism on a small enough neighborhood of the zero section.

Let $\phi \in \Omega^1_\pi(A_0; \tau^* A)$ be a solution of (25), (26) and (27). In this proof we show that ϕ must be given by (28). The remaining arguments are precisely as in the proof of Theorem 1.2.

By (27), $\phi_a(a) = \exp(a) = g_a(1)$ for all $a \in A_0$. This implies that $\phi_{ta}(ta) = g_{ta}(1) = t g_a(t)$, by using (38), and by linearity,

$$\phi_{ta}(a) = g_a(t) \tag{29}$$

for all $t \in (0, 1)$. Equation (29) is thus equivalent to (27).

Fix $x \in M$, $a \in (A_0)_x$ and $b \in A_x$. Let ∇ be a vector bundle connection on A. Equation (25) implies that

$$\left(d_{\tau^*\nabla}\phi + \tfrac{1}{2}[\phi, \phi]_\nabla\right)_{ta}(a, tb) = 0 \tag{30}$$

for all $t \in (0, 1)$. We will compute this equation for a fixed $t' \in (0, 1)$.

To compute $(d_{\tau^*\nabla}\phi)_{t'a}(a, t'b)$, consider the map $f : (0, 1) \times (-\delta, \delta) \to (A_0)_x$, $f(t, \epsilon) = t(a + \epsilon b)$. The composition $\tau \circ f$ restricted to $\epsilon = 0$ is the curve $t \mapsto \tau(ta)$, which is precisely γ_a, the base curve of the geodesic g_a, and $\tau \circ f$ restricted to

$t = t'$ is the curve $\gamma_\epsilon : (-\delta, \delta) \to M,\ \epsilon \mapsto \tau(t'(a + \epsilon b))$. Then

$$\begin{aligned}
(d_{\tau^*\nabla}\phi)_{t'a}(a, t'b) &= (f^* d_{\tau^*\nabla}\phi)_{(t',0)}\Big(\frac{\partial}{\partial t}, \frac{\partial}{\partial \epsilon}\Big) \\
&= (d_{f^*\tau^*\nabla} f^*\phi)_{(t',0)}\Big(\frac{\partial}{\partial t}, \frac{\partial}{\partial \epsilon}\Big) \\
&= (f^*\tau^*\nabla)_{\frac{\partial}{\partial t}}(f^*\phi)\Big(\frac{\partial}{\partial \epsilon}\Big)\Big|_{(t',0)} - (f^*\tau^*\nabla)_{\frac{\partial}{\partial \epsilon}}(f^*\phi)\Big(\frac{\partial}{\partial t}\Big)\Big|_{(t',0)} \\
&= (\nabla)_{\dot\gamma_a}\phi_{ta}(tb)|_{t=t'} - (\nabla)_{\dot\gamma_\epsilon} g_{a+\epsilon b}(t')|_{\epsilon=0}.
\end{aligned}$$

In the second equality we have used Lemma A.1 to commute the pullback with $d_{\tau^*\nabla}$ and in the last equality we have used (29), which is equivalent to (27). The two terms in the final expression are covariant derivatives of paths, which make sense because γ_a is the base curve of the curve $t \mapsto \phi_{ta}(tb)$ and γ_ϵ is the base curve of $\epsilon \mapsto g_{a+\epsilon b}(t')$.

To compute $\big(\frac{1}{2}[\phi, \phi]_\nabla\big)_{t'a}(a, t'b)$, let ξ be the map as in the theorem statement and let η be a time-dependent section of A satisfying $\eta^t(\gamma_a(t)) = \phi_{ta}(tb)$. Then

$$\begin{aligned}
\big(\tfrac{1}{2}[\phi, \phi]_\nabla\big)_{t'a}(a, t'b) &= [\xi_0^{t'}, \eta^{t'}]_\nabla(\gamma_a(t')) \\
&= [\xi_0^{t'}, \eta^{t'}](\gamma_a(t')) - \nabla_{\rho(\xi_0^{t'})}\eta^{t'}(\gamma_a(t')) + \nabla_{\rho(\eta^{t'})}\xi_0^{t'}(\gamma_a(t')) \\
&= \frac{d}{dt}\Big|_{t=t'} \psi_{\xi_0}^{t',t}\eta^{t'}(\gamma_a(t)) - \nabla_{\dot\gamma_a}\eta^{t'}(\gamma_a(t')) + \nabla_{\dot\gamma_\epsilon}\xi_0^{t'}(\gamma_a(t')).
\end{aligned}$$

In the last equality, we have used the defining property (37) of the infinitesimal flow for the first term, $\rho(\xi_0^{t'}(\gamma_a(t'))) = \rho(g_a(t')) = \dot\gamma_a(t')$ for the second term, and

$$\begin{aligned}
\rho(\eta^{t'}(\gamma_a(t'))) &= \rho(\phi_{t'a}(t'b)) \\
&= (d\tau)_{t'a}(t'b) \\
&= d(\pi \circ \exp)_{t'a}(t'b) \\
&= \frac{d}{d\epsilon}\Big|_{\epsilon=0}(\pi(\exp(t'a + \epsilon t'b))) \\
&= \frac{d}{d\epsilon}\Big|_{\epsilon=0}(\pi(g_{t'(a+\epsilon b)}(1))) \\
&= \frac{d}{d\epsilon}\Big|_{\epsilon=0}(\pi(t' g_{(a+\epsilon b)}(t'))) \\
&= \dot\gamma_\epsilon(0)
\end{aligned}$$

for the third term, where we have used the anchored property (26) in the second equality.

Thus for ϕ that satisfies (27), equation (30) is equivalent to

$$\frac{d}{dt}\Big|_{t=t'} \psi_{\xi_0}^{t',t}\eta^{t'}(\gamma_a(t)) + \frac{d}{dt}\Big|_{t=t'}\eta^t(\gamma_a(t')) = \frac{d}{d\epsilon}\Big|_{\epsilon=0}\xi_\epsilon^{t'}(\gamma_a(t')),$$

where we have used the characterization (36) of covariant derivatives of curves.

Applying $\psi_{\xi_0}^{1,t'}$ to both sides and using the product rule, the latter equation is equivalent to

$$\frac{d}{dt}\psi_{\xi_0}^{1,t}\eta^t(\gamma_a(t)) = \psi_{\xi_0}^{1,t}\frac{d}{d\epsilon}\Big|_{\epsilon=0}\xi_\epsilon^t(\gamma_a(t)).$$

Integrating t' from 0 to 1, and using the definition of η and the property $\psi_{\xi_0}^{1,1} = \mathrm{id}$, we obtain (28). □

Corollary 3.4. *The pullback of the canonical Maurer–Cartan form of a Lie groupoid by the exponential map* $\mathrm{Exp}^*\phi_{\mathrm{MC}}$ *is equal to the* 1*-form defined by* (28).

Proof. We saw already in the text preceding the last theorem that the 1-form $\mathrm{Exp}^*\phi_{\mathrm{MC}} \in \Omega^1_\pi(A_0; \tau^*A)$ is anchored and satisfies the Maurer–Cartan structure equation, and, in particular, it satisfies (25). Moreover, the initial condition (27) is satisfied since it is precisely the relation (23) when written out explicitly. The corollary now follows from the uniqueness assertion in the theorem. □

The second theorem is a generalization of Step 2 from Section 1. It shows that the solution from the previous theorem is indeed a solution of the realization problem.

Theorem 3.5. *Let A be a pre-Lie algebroid and $\phi \in \Omega^1_\pi(A_0; \tau^*A)$ a solution of* (25), (26) *and* (27). *Choose A_0 to be small enough so that ϕ is pointwise an isomorphism. Then* $\mathrm{MC}_\phi = 0$ *if and only if* $\mathrm{Jac} = 0$, *or more precisely*,

$$\mathrm{Jac}(a,b,c) = -3\frac{d}{dt}(\psi_\xi^{0,t}(\mathrm{MC}_\phi)_{ta}(b,c))\Big|_{t=0}, \tag{31}$$

$$(\mathrm{MC}_\phi)_a(b,c) = -\int_0^1 \psi_\xi^{1,t}\,\mathrm{Jac}\Big(\frac{1}{t}\exp(ta), \phi_{ta}(tb), \phi_{ta}(tc)\Big)\,dt, \tag{32}$$

where ξ is a time-dependent section of A satisfying $\xi^t(\gamma_a(t)) = g_a(t)$ for all $t \in (0,1)$.

Proof. The proof goes along the same lines as the proof of Theorem 1.4. As in Theorem 1.4, we will compute

$$d_{\tau^*\nabla}(\mathrm{MC}_\phi)_{ta}(a, tb, tc) \tag{33}$$

in two different ways, where $t \in (0,1)$, $x \in M$, $a \in (A_0)_x$, $y, z \in A_x$ and ∇ is some vector bundle connection on A.

1. Consider the map $f:(0,1)\times(-\delta,\delta)^2\to\mathfrak{g}$, $f(t,\epsilon,\epsilon')=t(a+\epsilon b+\epsilon' c)$. Recall that γ_a is the base curve of the geodesic g_a that satisfies $\gamma_a(t)=\tau(ta)$. Then

$$\begin{aligned}(d_{\tau^*\nabla}\mathrm{MC}_\phi)_{ta}(a,tb,tc) &= (f^*d_{\tau^*\nabla}\mathrm{MC}_\phi)_{(t,0,0)}\Big(\frac{\partial}{\partial t},\frac{\partial}{\partial\epsilon},\frac{\partial}{\partial\epsilon'}\Big)\\ &= (d_{f^*\tau^*\nabla}f^*\mathrm{MC}_\phi)_{(t,0,0)}\Big(\frac{\partial}{\partial t},\frac{\partial}{\partial\epsilon},\frac{\partial}{\partial\epsilon'}\Big)\\ &= \nabla_{\dot\gamma_a}(\mathrm{MC}_\phi)_{ta}(tb,tc),\end{aligned}$$

where the last expression is the covariant derivative of the curve $t\mapsto(\mathrm{MC}_\phi)_{ta}(tb,tc)$ covering γ_a.

2. Since $\phi\in\Omega^1_\pi(A_0;\tau^*A)$ is a pointwise isomorphism, it induces a linear map $\phi^{-1}:\Gamma(A)\to\mathfrak{X}(A_0)$. Let ξ be as in the statement of the theorem, let η_b and η_c be time-dependent sections of A satisfying $\eta^t_b(\gamma_a(t))=\phi_{ta}(tb)$ and $\eta^t_c(\gamma_a(t))=\phi_{ta}(tc)$ and let σ be a time-dependent section of A satisfying $\sigma^t(\gamma_a(t))=(\mathrm{MC}_\phi)_{ta}(tb,tc)$. Let $\tilde a,\tilde b,\tilde c$ be time-dependent vector fields on A_0 defined by $\tilde a^t=\phi^{-1}(\xi^t)$, $\tilde b^t=\phi^{-1}(\eta^t_b)$, $\tilde c^t=\phi^{-1}(\eta^t_c)$. Then

$$\begin{aligned}&(d_{\tau^*\nabla}\mathrm{MC}_\phi)_{ta}(a,tb,tc)\\ &\quad= (d_{\tau^*\nabla}\mathrm{MC}_\phi)(\tilde a,t\tilde b,t\tilde c)_{ta}\\ &\quad= \nabla_{\dot\gamma_a}\sigma^t(\gamma_a(t))-[[\xi^t,\eta^t_b],\eta^t_c]_{ta}-[[\eta^t_c,\xi^t],\eta^t_b]_{ta}+[\sigma^t-[\eta^t_b,\eta^t_c],\xi^t]_{ta}\\ &\quad= \nabla_{\dot\gamma_a}\sigma^t(\gamma_a(t))-\frac{d}{ds}\Big|_{s=t}\psi^{t,s}_\xi\sigma^t(\gamma_a(s))-\mathrm{Jac}\Big(\frac{1}{t}\exp(ta),\phi_{ta}(tb),\phi_{ta}(tc)\Big).\end{aligned}$$

The second equality is a slightly messy yet straightforward computation. It involves expanding MC_ϕ with respect to the chosen connection, using the choices we made above of time-dependent sections, and using (25), (27) and (26). In particular, it is used that (25) implies that $\phi_{ta}([\tilde a,\tilde b])=[\xi^t,\eta^t_b]_{\gamma_a(t)}$, $\phi_{ta}([\tilde a,\tilde c])=[\xi^t,\eta^t_c]_{\gamma_a(t)}$. Furthermore, $(\mathrm{MC}_\phi)_{ta}(b,c)=-\phi_{ta}([\tilde b,\tilde c])+[\eta^t_b,\eta^t_c]_{\gamma_a(t)}$. In the last equality we express the bracket $[\xi^t,\sigma^t]$ using the infinitesimal flow; see (37).

After equating the two expressions obtained, using characterization (36) of covariant derivatives of curves and applying $\psi^{1,t}_\xi$, (33) becomes

$$\psi^{1,t}_\xi\,\mathrm{Jac}\Big(\frac{1}{t}\exp(ta),\phi_{ta}(tb),\phi_{ta}(tc)\Big)=-\frac{d}{dt}(\psi^{1,t}_\xi(\mathrm{MC}_\phi)_{ta}(tb,tc)).\tag{34}$$

The remaining arguments are identical to those in the proof of Theorem 1.4. □

The Poisson case vs. the Lie algebroid case. Given the well-known relations between Poisson manifolds and Lie algebroids, it is natural to wonder as to the relation between the instances of the Maurer–Cartan structure equation associated with these structures, i.e., as to the relation between Section 2 and Section 3 of this paper. Let us briefly touch upon this.

In one direction, any Lie algebroid $A \to M$ induces a Poisson structure on the total space of the dual vector bundle $A^* \to M$ known as a linear Poisson structure (see [Mackenzie 2005]). This generalizes the construction of a linear Poisson structure on the dual of a Lie algebra. At the level of the associated Maurer–Cartan structure equations, it is not hard to verify that, locally and under obvious identifications, the Maurer–Cartan structure equations as well as the solutions are one and the same on both sides of this correspondence. In particular, trivializing A and computing the 1-form (28) will produce the same result as that obtained by computing the 1-form (18) associated with the induced trivialization of A^*. This is, of course, a generalization of the case of a Lie algebra which was discussed in Remark 2.6.

In the opposite direction, any Poisson manifold (M, π) induces a Lie algebroid structure on the cotangent bundle $T^*M \to M$, as originally shown in [Coste et al. 1987]. In that same paper, the authors proved that the local symplectic realization constructed by Weinstein [1983] (and discussed in Section 2 above) has a canonically induced local symplectic groupoid structure on its total space whose associated Lie algebroid is (the restriction of) $T^*M \to M$. This same phenomenon occurs at the level of the Maurer–Cartan structure equations. Using the notation of Section 2, the local solution of the Maurer–Cartan structure equation associated with the Poisson manifold $(\mathcal{O}, \pi)$, with $\mathcal{O} \subset V$, induces a local solution to the Maurer–Cartan structure equation associated with the Lie algebroid

$$T^*\mathcal{O} = \mathcal{O} \times V^* \xrightarrow{\pi} \mathcal{O}$$

by differentiation of the coefficients, or more precisely, by the map

$$\Omega^1(V^*; C^\infty(\mathcal{O})) \to \Omega^1_\pi((T^*\mathcal{O})_0; \tau^*(T^*\mathcal{O})), \quad \phi \mapsto \hat{\phi}, \tag{35}$$

with $\hat{\phi}_{x,\xi}(\zeta) = d(\phi_\xi(\zeta))_{\tau(x)}$ for all $x \in \mathcal{O}$, $\xi, \zeta \in V^*$.

Note that whereas in the Lie algebroid case we are able to obtain a "wide" solution, i.e., on an open neighborhood of the zero section of $T^*M \to M$, in the Poisson case we only obtain a local one around a point in M. It would be interesting to further investigate the relation given by (35) to see if a wide solution of the Lie algebroid case induces a wide solution of the Poisson case, thus producing yet another proof for the existence of global symplectic realizations.

Appendix: Facts on (pre-)Lie algebroids

In this appendix, various notions are recalled which are needed in Section 3 for the formulation of the Maurer–Cartan structure equation on a Lie algebroid and its solution. For more details, the reader is referred to [Crainic and Fernandes 2003]. Note that all the notions that appear here and that are presented in [Crainic and

Fernandes 2003] do not require the Jacobi identity and are therefore as valid for pre-Lie algebroids as they are for Lie algebroids.

Let $A \to M$ be a pre-Lie algebroid (see Section 3 for the definition). An *A-connection* on a vector bundle $E \to M$ is an $\mathbb{R}$-bilinear map $\overline{\nabla} : \Gamma(A) \times \Gamma(E) \to \Gamma(E)$ satisfying the connection-like properties

$$\overline{\nabla}_{f\alpha} s = f \overline{\nabla}_\alpha s \quad \text{and} \quad \overline{\nabla}_\alpha (fs) = f \overline{\nabla}_\alpha s + \mathcal{L}_{\rho(\alpha)}(f) s$$

for all $\alpha \in \Gamma(A)$, $s \in \Gamma(E)$, $f \in C^\infty(M)$.

For the remainder of the appendix, let $A \to M$ be a pre-Lie algebroid equipped with an A-connection $\overline{\nabla}$. Note that there will be two different connections that will play a role in this appendix (and in Section 3): an A-connection $\overline{\nabla}$ on A that is part of the data, and an auxiliary vector bundle connection ∇ on A that is used to write down the Maurer–Cartan structure equation globally, and which is not part of the data.

Time-dependent sections. A *time-dependent section* ξ of A is a map $\xi : I \times M \to A$, $(t, x) \mapsto \xi^t(x)$ (with I some open interval), such that ξ^t is a section of A for all $t \in I$.

If $\nabla : \mathfrak{X}(M) \times \Gamma(A) \to \Gamma(A)$ is a vector bundle connection, then given a base curve $\gamma : I \to M$ and a curve $u : I \to A$ covering γ, the covariant derivative $(\nabla_{\dot\gamma} u)(t) = ((\gamma^* \nabla)_{\partial/\partial t} u)(t)$ can be characterized using time-dependent sections as follows: choose a time-dependent section ξ of A satisfying $\xi^t(\gamma(t)) = u(t)$ for all $t \in I$; then

$$(\nabla_{\dot\gamma} u)(t) = (\nabla_{\dot\gamma} \xi^t)(x) + \frac{d\xi^t}{dt}(x), \tag{36}$$

where $x = \gamma(t)$.

We will also use time-dependent sections to express the bracket of a pre-Lie algebroid in a Lie derivative-like fashion, as one does for the bracket of vector fields. This involves the notion of an infinitesimal flow. Let ξ be a time-dependent section of A and $\rho(\xi)$ the corresponding time-dependent vector field on M. Let $\varphi^{t,s}_{\rho(\xi)}$ denote the flow of $\rho(\xi)$ from time s to t. The *infinitesimal flow*,

$$\psi^{t,s}_\xi : A_x \to A_{\varphi^{t,s}_{\rho(\xi)}}, \quad x \in M,$$

is the unique linear map satisfying the properties $\psi^{u,t}_\xi \circ \psi^{t,s}_\xi = \psi^{u,s}_\xi$, $\psi^{s,s}_\xi = \mathrm{id}$ and

$$\frac{d}{dt}\Big|_{t=s} \psi^{s,t}_\xi \alpha(\varphi^{t,s}_{\rho(\xi)}(x)) = [\xi^s, \alpha]_x \quad \forall \alpha \in \Gamma(A),\ x \in M.$$

Defining the pullback of sections by the infinitesimal flow as $(\psi^{t,s}_\xi)^*(\alpha)(x) = \psi^{s,t}_\xi \alpha(\varphi^{t,s}_{\rho(\xi)}(x))$ for all $\alpha \in \Gamma(A)$, $x \in M$, the previous equation can be expressed in the more familiar form

$$\frac{d}{dt}\Big|_{t=s} (\psi^{t,s}_\xi)^* \alpha = [\xi^s, \alpha] \quad \forall \alpha \in \Gamma(A). \tag{37}$$

For more on infinitesimal flows and their global counterparts, and flows along invariant time-dependent vector fields on Lie groupoids, see [Crainic and Fernandes 2003].

Geodesics. An *A-path* is a curve $g : I \to A$ with base curve $\gamma : I \to M$, where $\gamma(t) = \pi(g(t))$, such that

$$\rho(g(t)) = \dot{\gamma}(t) \quad \forall\, t \in I.$$

Let g be an A-path with base curve γ, and let $u : I \to A$ be another curve covering γ. The *covariant derivative* of u with respect to g is the curve $\overline{\nabla}_g u : I \to A$, which is defined in analogy to the usual covariant derivative described above: choose a time-dependent section ξ of A satisfying $\xi^t(\gamma(t)) = u(t)$ for all $t \in I$; then

$$(\overline{\nabla}_g u)(t) = (\overline{\nabla}_g \xi^t)(x) + \frac{d\xi^t}{dt}(x),$$

where $x = \gamma(t)$.

A *geodesic* is a curve $g : I \to A$ satisfying the geodesic equation $\overline{\nabla}_g g = 0$. Geodesics are A-paths. Given any point $a \in A$, there is a unique maximal geodesic $g_a : I_a \to A$ satisfying $g_a(0) = a$ with domain I_a. The base curve of g_a will be denoted by γ_a. Geodesics satisfy the basic property

$$g_{sa}(t) = s g_a(st) \quad \forall\, a \in A,\ s, t \in \mathbb{R},\ t \in I_{sa}, \tag{38}$$

which can be easily verified by checking that the curve $t \mapsto s g_a(st)$ satisfies the geodesic equation and then by noting that by uniqueness it must be equal to g_{sa} since at time 0 it takes the value sa.

Let $A_0 \subset A$ be a neighborhood of the zero section such that g_a is defined up to time 1 for all $a \in A_0$. The *exponential map* is defined as $\exp : A_0 \to A,\ a \mapsto g_a(1)$. The point $\pi(\exp(a)) \in M$ will be called the *target* of a and $\tau = \pi \circ \exp : A_0 \to M$ the *target map*.

The Maurer–Cartan 2-form. Let $\Omega^*_\pi(A_0; \tau^* A)$ denote the space of foliated differential forms on A_0 (foliated with respect to the foliation by π-fibers) which take values in $\tau^* A$.

Let $\nabla : \mathfrak{X}(M) \times \Gamma(A) \to \Gamma(A)$ be a vector bundle connection on A. The torsion $[\,\cdot\,,\cdot\,]_\nabla \in \mathrm{Hom}(\Lambda^2 A, A)$ of ∇ is defined at the level of sections by

$$[\alpha, \beta]_\nabla = [\alpha, \beta] - \nabla_{\rho(\alpha)}\beta + \nabla_{\rho(\beta)}\alpha \quad \forall\, \alpha, \beta \in \Gamma(A),$$

and is easily checked to be $C^\infty(M)$-linear in both slots. The torsion induces a bracket $[\,\cdot\,,\cdot\,]_\nabla : \Omega^p_\pi(A_0; \tau^* A) \times \Omega^q_\pi(A_0; \tau^* A) \to \Omega^{p+q}_\pi(A_0; \tau^* A)$ which plays the role of the wedge product on A-valued forms, and similarly to the wedge product, it is defined by the formula

$$[\omega, \eta]_\nabla(X_1, \dots, X_{p+q})_a = \sum_{\sigma \in S_{p,q}} \mathrm{sgn}(\sigma)[\omega(X_{\sigma(1)}, \dots, X_{\sigma(p)})_a, \eta(X_{\sigma(p+1)}, \dots, X_{\sigma(p+q)})_a]_\nabla$$

for all $a \in A_0$, where $S_{p,q}$ is the set of (p, q)-shuffles.

In general, a connection ∇ on a vector bundle $E \to M$ induces a differential-like map $d_\nabla : \Omega^*(M; E) \to \Omega^{*+1}(M; E)$ by the usual Koszul-type formula. For example, if $\phi \in \Omega^1(M; E)$,

$$d_\nabla \phi(X, Y) = \nabla_X \phi(Y) - \nabla_Y \phi(X) - \phi([X, Y]) \quad \forall\, X, Y \in \mathfrak{X}(M).$$

The map d_∇ squares to zero if and only if the connection is flat. If M has a foliation $\mathcal{F}$ and $\Omega^*_{\mathcal{F}}(M; E)$ are the foliated forms, then the map d_∇ descends to a map of foliated forms $d_\nabla : \Omega^*_{\mathcal{F}}(M; E) \to \Omega^{*+1}_{\mathcal{F}}(M; E)$. We will need the following property, whose proof is elementary and will be left out.

Lemma A.1. *Let $E \to M$ be a vector bundle equipped with a connection ∇ and let $f : N \hookrightarrow M$ be a submanifold. Then*

$$f^* d_\nabla \phi = d_{f^*\nabla} f^* \phi$$

for any $\phi \in \Omega^*(M; E)$. If N and M are foliated and f is a foliated map, then the property holds for $\phi \in \Omega^*_{\mathcal{F}}(M; E)$.

In our particular case, the induced pull-back connection $\tau^*\nabla$ on the vector bundle $\tau^*A \to A_0$ induces a differential-like map

$$d_{\tau^*\nabla} : \Omega^*_\pi(A_0; \tau^*A) \to \Omega^{*+1}_\pi(A_0; \tau^*A).$$

A 1-form $\phi \in \Omega^1_\pi(A_0; \tau^*A)$ is said to be *anchored* if $\rho \circ \phi = d\tau$, or more explicitly, if $\rho(\phi_a(b)) = (d\tau)_a(b)$ for all $a \in (A_0)_x$, $b \in A_x$ (where we are using the canonical identification $T_a A_0 \cong A_x$).

Proposition A.2. *Let $\phi \in \Omega^1_\pi(A_0; \tau^*A)$. If ϕ is anchored, then the 2-form*

$$d_{\tau^*\nabla}\phi + \tfrac{1}{2}[\phi, \phi]_\nabla \in \Omega^2_\pi(A_0; \tau^*A) \tag{39}$$

is independent of the choice of connection ∇.

Proof. Let ∇ and ∇' be two connections. Then by the defining properties of a connection, $\omega := \nabla - \nabla' \in \Omega^1(M; \mathrm{Hom}(A, A))$. Let $a \in A_0$ and $X, Y \in T_a A_0$ such that $d\pi(X) = d\pi(Y) = 0$. On the one hand,

$$(d_{\tau^*\nabla}\phi - d_{\tau^*\nabla'}\phi)(X, Y) = \omega_{\tau(a)}((d\tau)_a(X))(\phi_a(Y)) - \omega_{\tau(a)}((d\tau)_a(Y))(\phi_a(X)).$$

On the other hand,

$$\left(\tfrac{1}{2}[\phi,\phi]_\nabla - \tfrac{1}{2}[\phi,\phi]_{\nabla'}\right)(X,Y)$$
$$= -\omega_{\tau(a)}(\rho(\phi_a(X)))(\phi_a(Y)) + \omega_{\tau(a)}(\rho(\phi_a(Y)))(\phi_a(X)).$$

The sum of these two equations vanishes if ϕ is anchored. □

We call the 2-form given by (39) the *Maurer–Cartan* 2*-form* and denote it by MC_ϕ.

Acknowledgements

This research was financially supported by the ERC Starting Grant No. 279729. The author would also like to thank Marius Crainic, Pedro Frejlich, Ionut Mărcuț and Maria Amelia Salazar for helpful discussions and suggestions, and the anonymous referee for excellent remarks and suggestions.

References

[Cartan 1904] É. Cartan, "Sur la structure des groupes infinis de transformations", *Ann. Sci. École Norm. Sup.* (3) **21** (1904), 153–206. MR 1509040 JFM 35.0176.04

[Cartan 1937] É. Cartan, "La structure des groupes infinis", exposé 4-G in *Le Séminaire de mathématiques*, edited by M. Audin, Centre de Diffusion des Revues Académiques de Mathématiques, Grenoble, 1937.

[Coste et al. 1987] A. Coste, P. Dazord, and A. Weinstein, "Groupoïdes symplectiques", *Publ. Dépt. Math. Lyon* **87**:2A (1987), 1–62. MR 90g:58033 Zbl 0668.58017

[Crainic and Fernandes 2003] M. Crainic and R. L. Fernandes, "Integrability of Lie brackets", *Ann. of Math.* (2) **157**:2 (2003), 575–620. MR 2004h:58027 Zbl 1037.22003

[Crainic and Fernandes 2011] M. Crainic and R. L. Fernandes, "Lectures on integrability of Lie brackets", pp. 1–107 in *Lectures on Poisson geometry* (Trieste, 2005), edited by T. Ratiu et al., Geometry and Topology Monographs **17**, Geometry and Topology Publications, Coventry, 2011. MR 2795150 Zbl 1227.22005

[Crainic and Mărcuț 2011] M. Crainic and I. Mărcuț, "On the existence of symplectic realizations", *J. Symplectic Geom.* **9**:4 (2011), 435–444. MR 2900786 Zbl 1244.53091

[Duistermaat and Kolk 2000] J. J. Duistermaat and J. A. C. Kolk, *Lie groups*, Springer, Berlin, 2000. MR 2001j:22008 Zbl 0955.22001

[Fernandes and Struchiner 2014] R. L. Fernandes and I. Struchiner, "The classifying Lie algebroid of a geometric structure, I: Classes of coframes", *Trans. Amer. Math. Soc.* **366**:5 (2014), 2419–2462. MR 3165644 Zbl 1285.53018

[Greub et al. 1973] W. Greub, S. Halperin, and R. Vanstone, *Connections, curvature, and cohomology, II: Lie groups, principal bundles, and characteristic classes*, Pure and Applied Mathematics **47**-II, Academic Press, New York, 1973. MR 49 #1424 Zbl 0335.57001

[Mackenzie 2005] K. C. H. Mackenzie, *General theory of Lie groupoids and Lie algebroids*, London Mathematical Society Lecture Note Series **213**, Cambridge University Press, 2005. MR 2006k:58035 Zbl 1078.58011

[Sternberg 2004] S. Sternberg, "Lie algebras", April 2004, http://www.math.harvard.edu/~shlomo/docs/lie_algebras.pdf. Lecture notes.

[Weinstein 1983] A. Weinstein, "The local structure of Poisson manifolds", *J. Differential Geom.* **18**:3 (1983), 523–557. MR 86i:58059 Zbl 0524.58011

Received June 19, 2015. Revised November 12, 2015.

ORI YUDILEVICH
MATHEMATICAL INSTITUTE
UTRECHT UNIVERSITY
BUDAPESTLAAN 6
3584CD UTRECHT
NETHERLANDS
oriyudilevich@gmail.com

CONTENTS

Volume 282, 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 282 No. 2 June 2016

0030-8730(2016)282:2;1-3