Pacific Journal of Mathematics

Volume 288 No. 2 June 2017

PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm

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

The subscription price for 2017 is US $450/year for the electronic version, and $625/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. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.257

ORDER ON THE HOMOLOGY GROUPS OF SMALE SPACES

MASSOUD AMINI, IAN F. PUTNAM AND SARAH SAEIDI GHOLIKANDI

Smale spaces were defined by D. Ruelle to describe the properties of the basic sets of an Axiom A system for topological dynamics. One motivation for this was that the basic sets of an Axiom A system are merely topological spaces and not submanifolds. One of the most important classes of Smale spaces is shifts of finite type. For such systems, W. Krieger introduced a pair of invariants, the past and future dimension groups. These are abelian groups, but are also with an order which is an important part of their structure. The second author showed that Krieger's invariants could be extended to a homology theory for Smale spaces. In this paper, we show that the homology groups on Smale spaces (in degree zero) have a canonical order structure. This extends that of Krieger's groups for shifts of finite type.

1. Introduction

The original notion of a Smale space is due to David Ruelle, based on the observation that the basic sets of Smale's Axiom A systems do not form submanifolds of the ambient manifold [Ruelle 1978; Smale 1967; Aoki and Hiraide 1994; Fried 1987; Fisher 2013; Bowen 1978]. In fact, Smale spaces are the topological dynamical systems that admit a hyperbolic structure in terms of canonical coordinates of contracting and expanding (or stable and unstable) directions. Hyperbolic toral automorphisms, one-dimensional generalized solenoids as described by R. F. Williams and shifts of finite type are all examples of Smale spaces. In fact, any totally disconnected (irreducible) Smale space is conjugate to a shift of finite type. W. Krieger [1980] defined two abelian groups for shift of finite type, called the past and future dimension groups, in terms of clopen sets of the stable and unstable sets. One of their most important features is a natural order structure.

The second author [Putnam 2014] defined a homology for Smale spaces which extends the dimension groups for shifts of finite type. However, the homology groups as defined in that paper are not given any order structure. In this paper, we

MSC2010: primary 37D20, 55N35; secondary 37B10, 06F15.
Keywords: hyperbolic dynamical systems, dimension groups, homology, ordered groups, Smale spaces.

prove that the homology groups of Smale spaces in degree zero have a canonical order structure.

The paper is organized as follows. In Section 2, we introduce the basic concepts and notations (based on [Putnam 2014]) and state the main results of this paper, which are proved in Sections 4 and 5. The shifts of finite type which play an important role in the homology of Smale spaces are reviewed in Section 3 and their dimension groups are discussed as ordered groups.

2. Preliminaries

2A. ***Smale spaces.*** A pair (X, φ) is called a dynamical system if X is a topological space and φ is a homeomorphism of X. A dynamical system (X, φ) is called irreducible if for every ordered pair of nonempty open sets U, V in X, there is a nonnegative integer n such that $\varphi^n(U) \cap V$ is nonempty. It is called mixing if for every ordered pair of nonempty open sets U, V in X, there is a nonnegative integer N such that $\varphi^n(U) \cap V$ is nonempty for any $n \geq N$ [Aoki and Hiraide 1994; Putnam 2014].

Definition 2.1 [Ruelle 1978; Putnam 2014, Definition 2.1.6]. For a compact metric space X, the dynamical system (X, φ) is called a Smale space if there exist constants ϵ_X and $0 < \lambda < 1$ and a continuous map from

$$\triangle_{\epsilon_X} = \{(x, y) \in X \times X \mid d(x, y) \leq \epsilon_X\}$$

to X (denoted by $[\,\cdot\,,\,\cdot\,]$) such that, for every $x, y, z \in X$,

(B1) $[x, x] = x$,

(B2) $[x, [y, z]] = [x, z]$,

(B3) $[[x, y], z] = [x, z]$,

(B4) $[\varphi(x), \varphi(y)] = \varphi([x, y])$

whenever both sides of the above equations are defined, and

(C1) $d(\varphi(x), \varphi(y)) \leq \lambda\, d(x, y)$ whenever $[x, y] = y$,

(C2) $d(\varphi^{-1}(x), \varphi^{-1}(y) \leq \lambda\, d(x, y)$ whenever $[x, y] = x$.

In a Smale space (X, φ), the local stable and unstable sets are defined, for x in X and $\epsilon_X \geq \epsilon > 0$, by

$$X^s(x, \epsilon) = \{y \in X \mid d(x, y) \leq \epsilon, [x, y] = y\},$$
$$X^u(x, \epsilon) = \{y \in X \mid d(x, y) \leq \epsilon, [y, x] = y\}.$$

It is simple to show that, for any ϵ sufficiently small, $[\,\cdot\,,\,\cdot\,] : X^u(x, \epsilon) \times X^s(x, \epsilon) \to X$ is a homeomorphism to its image, which is a neighbourhood of x in X. The inverse is obtained by mapping y to $([x, y], [y, x])$.

Let (X, φ) be a Smale space. Two points $x, y \in X$ are stably equivalent if

$$\lim_{n\to+\infty} d(\varphi^n(x), \varphi^n(y)) = 0$$

and unstably equivalent if

$$\lim_{n\to+\infty} d(\varphi^{-n}(x), \varphi^{-n}(y)) = 0.$$

We denote the stable and unstable equivalence classes of x by $X^s(x)$ and $X^u(x)$, respectively [Putnam 2014].

Examples of Smale spaces include Anosov diffeomorphisms, the basic sets from Smale's Axiom A systems, various solenoids and certain substitution tiling spaces [Katok and Hasselblatt 1995; Smale 1967; Bowen 1970; 1978; Williams 1967; 1970; 1974; Wieler 2014; Yi 2001]. Key examples are the shifts of finite type, namely the doubly infinite path space of a finite directed graph. We provide a more complete description in the next section. In this case, the underlying space is totally disconnected [Lind and Marcus 1995; Putnam 2014]. Conversely, any irreducible Smale space which is totally disconnected is topologically conjugate to a shift of finite type.

A factor map π between dynamical systems (Y, ψ) and (X, φ) is a continuous, surjective map $\pi : Y \to X$ satisfying $\varphi \circ \pi = \pi \circ \psi$. A factor map π is finite-to-one if there is an upper bound on the cardinality of the sets $\pi^{-1}\{x\}$, as x runs over X [Putnam 2014]. It is almost one-to-one if $\#\pi^{-1}\{x\} = 1$ for each x in some dense G_δ subset of X.

A map $\pi : (Y, \psi) \to (X, \varphi)$ between Smale spaces is called s-bijective (resp. u-bijective) if the restriction of π to $Y^s(y)$ (resp. $Y^u(y)$) is a bijection to $X^s(\pi(y))$ (resp. $X^u(\pi(y))$) for any $y \in Y$. Every s-bijective (or u-bijective) map is finite-to-one [Putnam 2014].

Definition 2.2 [Putnam 2014, Definition 2.6.2]. Let (X, φ) be a Smale space. Then

$$\pi = (Y, \psi, \pi_s, Z, \zeta, \pi_u)$$

is an s/u-bijective pair for (X, φ) if

- (Y, ψ) and (Z, ζ) are Smale spaces,
- $\pi_s : (Y, \psi) \to (X, \varphi)$ is s-bijective and $X^u(y)$ is totally disconnected for every $y \in Y$,
- $\pi_u : (Z, \zeta) \to (X, \varphi)$ is u-bijective and $X^s(y)$ is totally disconnected for every $z \in Z$.

Theorem 2.3 [Putnam 2014, Theorem 2.6.3]. *Every irreducible Smale space (X, φ) admits an s/u-bijective pair.*

This result plays a crucial role in [Putnam 2014]. The homology is defined and computed from such an object. While there may be many such s/u-bijective pairs for a given (X, φ), it is shown in Theorem 5.5.1 of that paper that the homology is independent of the choice.

Our first contribution here is to improve this situation by proving the existence of s/u-bijective pairs with certain advantageous extra features. These will be important for our proofs later, but presumably, will have many other applications.

Theorem 2.4. *Every irreducible Smale space (X, φ) admits an s/u-bijective pair $\pi = (Y, \psi, \pi_s, Z, \zeta, \pi_u)$ such that the Smale spaces (Y, ψ), (Z, ζ) are irreducible and both maps π_s and π_u are almost one-to-one.*

The proof is based on [Putnam 2005] and will be given in Section 4A.

Definition 2.5. For any Smale space (X, φ), we say that an s/u-bijective pair $(Y, \psi, \pi_s, Z, \zeta, \pi_u)$ is *irreducible* if both (Y, ψ) and (Z, ζ) are irreducible and both maps π_s and π_u are almost one-to-one.

For a Smale space (X, φ) and s/u-bijective pair $\pi = (Y, \psi, \pi_s, Z, \zeta, \pi_u)$, for each $L, M \geq 0$ we define

$$\Sigma_{L,M}(\pi) = \big\{(y_0, y_1, \ldots, y_L, z_0, z_1, \ldots, z_M) \in Y^{L+1} \times Z^{M+1} \mid \\ \pi_s(y_l) = \pi_u(z_m) \text{ for all } 0 \leq l \leq L,\ 0 \leq m \leq M\big\}.$$

If we let σ be the obvious map on $\Sigma_{L,M}(\pi)$ induced by ψ and ζ, $(\Sigma_{L,M}(\pi), \sigma)$ is a dynamical system. Indeed, it is also a Smale space with totally disconnected stable and unstable sets, and so is a shift of finite type. In the special case that $L = M = 0$, this is usually called the fibred product of (Y, ψ) and (Z, ζ). On the other hand, $(\Sigma_{L,M}(\pi), \sigma)$ has an obvious action of the group $S_{L+1} \times S_{M+1}$, where S_{N+1} denotes the permutation group of $\{0, 1, \ldots, N\}$ [Putnam 2014].

If the s/u-bijective pair is irreducible in the sense above, then the fibred product is irreducible. By this we mean the shift of finite type $\Sigma_{0,0}(\pi), \sigma)$ is irreducible. The other $\Sigma_{L,M}(\pi), \sigma)$ will not be, in general. The proof of this result is long and will be given in Section 4B.

Theorem 2.6. *Suppose (X, φ), (Y, ψ) and (Z, ζ) are irreducible Smale spaces,*

$$\pi_s : (Y, \psi) \to (X, \varphi)$$

is an almost one-to-one, s-bijective factor map and

$$\pi_u : (Z, \zeta) \to (X, \varphi)$$

is an almost one-to-one, u-bijective factor map. Then the fibred product

$$Y \times_X Z = \{(y, z) \in Y \times Z \mid \pi_s(y) = \pi_u(z)\}$$

of maps π_s and π_u with $\psi \times \zeta$ is an irreducible Smale space. In addition, if (X, φ) is mixing then so is $(Y \times_X Z, \psi \times \zeta)$.

Of course, one application of the theorem is to an irreducible s/u-bijective pair for (X, φ), but the result is more general, since it makes no assumptions that the local stable sets of Z or the local unstable sets of Y are totally disconnected.

For a shift of finite type, (Σ, σ), Krieger introduced the dimension group invariants, denoted here by $D^s(\Sigma, \sigma)$ and $D^u(\Sigma, \sigma)$. These are countable abelian groups and, if the shift of finite type is presented as the edge shift of a finite directed graph G, they may be computed directly as inductive limits from the adjacency matrix of G. We discuss this more thoroughly in the next section.

The second author developed a homology theory for Smale spaces in [Putnam 2014]. Let us briefly review the construction here. First, one considers the dimension groups $D^s(\Sigma_{L,M}(\pi), \sigma)$ of the system, over all $L, M \geq 0$. At each index, a quotient of a certain subgroup is taken, denoted by $D^s_{\mathcal{Q},\mathcal{A}}(\Sigma_{L,M}(\pi), \sigma)$, which takes into account the action of the permutation groups [Putnam 2014, Section 5.1]. These groups are assembled into a double complex, $\mathcal{C}^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M} = D^s_{\mathcal{Q},\mathcal{A}}(\Sigma_{L,M}(\pi), \sigma)$, $L, M \geq 0$, whose homology is denoted by $H^s_*(\pi)$. There is an analogous construction of $H^u_*(\pi)$, using the unstable dimension groups D^u. In [Putnam 2014], it is shown that the result is independent of the choice of π, and so is written as $H^s_*(X, \varphi)$ or $H^u_*(X, \varphi)$ [Putnam 2014, Theorem 5.5.1].

For the remainder of this paper, we will concentrate on $H^s(X, \varphi)$. Analogous results hold for $H^u(X, \varphi)$.

The above construction is analogous to computing the Čech cohomology of a compact manifold by considering a 'nice', finite, open cover and the homology of its nerve. Here, the s/u-bijective pair replaces the open cover. The shifts $(\Sigma_{L,M}(\pi), \sigma)$ evidently play the role of the nerve of the cover, keeping track of the multiplicities of the cover. Finally, Krieger's dimension group invariant replaces the homology of the open balls in the 'nice' cover.

One of the most important features of Krieger's invariant for a shift of finite type is that it also carries a natural order structure. Moreover, this is also easily computed from the corresponding directed graph. The aim of this paper is to define a natural and canonical order structure on the homology groups $H^s_0(X, \varphi)$ and $H^u_0(X, \varphi)$.

Let us begin with the definition of an ordered abelian group.

Definition 2.7 [Blyth 2005]. A pair (G, G^+) is called an ordered abelian group if G is an abelian group with a positive cone G^+, which is a subset of G satisfying

(1) $G^+ + G^+ \subseteq G^+$,

(2) $G^+ - G^+ = G$,

(3) $G^+ \cap -G^+ = \{0\}$.

The elements of G^+ are called positive elements of G, and for g_1, g_2 in G we write $g_1 \geq g_2$ (or $g_2 \leq g_1$) when $g_1 - g_2 \in G^+$.

A homomorphism $\Gamma : G \to H$ of ordered groups is called positive if $\Gamma(G^+) \subseteq H^+$. An isomorphism $\Gamma : G \to H$ of ordered groups is an order isomorphism if both Γ and Γ^{-1} are positive homomorphisms (equivalently, if $\Gamma(G^+) = H^+$). We remark that the inverse of a positive isomorphism is not positive in general. For example, consider the ordered group $\mathbb{Z}^2$ with the positive cone $\{(m,n) \mid m, n \geq 0\}$. The map $\alpha(m,n) = (m+n, n)$ is a positive automorphism of $\mathbb{Z}^2$ whose inverse is not positive.

The groups $D^s(\Sigma_{L,M}(\pi))$, L, $M \geq 0$, all carry canonical orders. Unfortunately, these do not induce orders on the groups $D^s_{\mathcal{Q},\mathcal{A}}(\Sigma_{L,M}(\pi))$ in our double complex, except in the special case when $L = M = 0$, where $D^s_{\mathcal{Q},\mathcal{A}}(\Sigma_{0,0}(\pi))$ and $D^s(\Sigma_{0,0}(\pi))$ are equal. We intend to lift this order to the degree-zero group in our double complex, namely on $\bigoplus_{L-M=0} \mathcal{C}^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}$, by setting the positive cone to be those elements whose entries in the summand $L = M = 0$ are strictly positive, together with the zero element. In particular, the entries in the position $L = M > 0$ do not affect positivity. The positive cone $H^s_0(\pi)^+$ in $H^s_0(\pi)$ is then defined as those elements which are represented by a positive cocycle in $\bigoplus_{L-M=0} \mathcal{C}^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}$. The difficulty is to show that this gives a well-defined and well-behaved order on the homology.

Definition 2.8. Let $\pi = (Y, \psi, \pi_s, Z, \zeta, \pi_u)$ be an s/u-bijective pair for the Smale space (X, φ). Let $(\mathcal{C}^s_{\mathcal{Q},\mathcal{A}}(\pi), d^s_{\mathcal{Q},\mathcal{A}}(\pi))$ be the double complex associated with π and $H^s_*(\pi)$ be the homology of this double complex. We define the corresponding cones as follows:

$$\Big(\bigoplus_{L-M=0} \mathcal{C}^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\Big)^+ = \{0\} \cup \{a \mid 0 \neq a_{0,0} \in \mathcal{C}^s_{\mathcal{Q},\mathcal{A}}(\pi)^+_{0,0}\},$$

and

$$H^s(\pi)^+ = \Big\{a + \operatorname{Im}\Big(\bigoplus_{L-M=1} d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\Big) \;\Big|\; a \in \operatorname{Ker}\Big(\bigoplus_{L-M=0} d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\Big) \cap \Big(\bigoplus_{L-M=0} \mathcal{C}^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\Big)^+\Big\}.$$

Of course, both definitions are the obvious ones. The issue is now to show that this provides good order structures, at least for irreducible Smale spaces. The strategy is simple: we first assume that our s/u-bijective pair is irreducible. We reduce to the case that the shift of finite type, $(\Sigma_{0,0}(\pi), \sigma)$, is mixing. It follows that the order structure on its dimension group is completely determined by the state which arises from its unique measure of maximal entropy, or the Parry measure; see Theorem 3.4.

To take homology, we first pass to a subgroup (the cocycles) and then take a quotient (by the coboundaries). The following rather elementary result summarizes our task.

Theorem 2.9 [Blyth 2005]. *Let (G, G^+) be an ordered abelian group and let $H \subseteq G$ be a subgroup.*

(i) *If* $G^+ \cap H = \{0\}$, *then with*

$$(G/H)^+ = \{a + H \mid a \in G^+\},$$

$(G/H, (G/H)^+)$ *is an ordered abelian group.*

(ii) *If* $G^+ \cap H$ *generates* H, *that is*, $(G^+ \cap H) - (G^+ \cap H) = H$, *then* $(H, G^+ \cap H)$ *is an ordered abelian group.*

The conditions for the subgroup and quotient in the above theorem are complementary and could not hold at the same time (except for trivial cases), but one should note that these conditions are going to be applied to two separate cases with distinct subgroups (the subgroup condition is applied to a "kernel" in the complex, whereas the quotient condition is used for the preceding "image").

Our first task is to show that

$$G = \mathrm{Ker}\big(\textstyle\bigoplus_{L-M=0} d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\big) \quad \text{and} \quad H = \mathrm{Im}\big(\textstyle\bigoplus_{L-M=1} d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\big)$$

satisfy the hypotheses of the first part of Theorem 2.9.

Theorem 2.10. *Let* $\pi = (Y, \psi, \pi_s, Z, \zeta, \pi_u)$ *be an irreducible* s/u*-bijective pair for the irreducible Smale space* (X, φ). *We have*

$$\big(\textstyle\bigoplus_{L-M=0} C^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\big)^+ \cap \mathrm{Im}\big(\textstyle\bigoplus_{L-M=1} d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\big) = \{0\}.$$

Our second task is to show that

$$G = \textstyle\bigoplus_{L-M=0} C^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M} \quad \text{and} \quad H = \mathrm{Ker}\big(\textstyle\bigoplus_{L-M=0} d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\big)$$

satisfy the hypotheses of the second part of Theorem 2.9.

Theorem 2.11. *Let* $\pi = (Y, \psi, \pi_s, Z, \zeta, \pi_u)$ *be an irreducible* s/u*-bijective pair for the irreducible Smale space* (X, φ). *The subgroup generated by*

$$\big(\textstyle\bigoplus_{L-M=0} C^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\big)^+ \cap \mathrm{Ker}\big(\textstyle\bigoplus_{L-M=0} d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\big)$$

is $\mathrm{Ker}\big(\bigoplus_{L-M=0} d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\big)$.

As an immediate consequence of Theorems 2.9, 2.10 and 2.11, we get our main result as follows.

Theorem 2.12. *If* (X, φ) *is an irreducible Smale space and* $\pi = (Y, \psi, \pi_s, Z, \zeta, \pi_u)$ *is an irreducible* s/u*-bijective pair for* (X, φ), *then* $H^s_0(\pi)$ *is an ordered abelian group with the positive cone defined in Definition 2.8.*

The next issue is to see that the resulting order is independent of the choice of π, in a suitable sense.

Theorem 2.13. *Suppose (X, φ) is an irreducible Smale space and*

$$\pi = (Y, \psi, \pi_s, Z, \zeta, \pi_u), \quad \tilde{\pi} = (\tilde{Y}, \tilde{\psi}, \tilde{\pi}_s, \tilde{Z}, \tilde{\zeta}, \tilde{\pi}^u)$$

are s/u-bijective pairs for (X, φ). Assume that π is irreducible. Then

(1) *$H_0^s(\tilde{\pi})$ is an ordered abelian group with the positive cone given in Definition 2.8;*

(2) *there is an order isomorphism $\mathcal{H}$ from $H_0^s(\pi)$ to $H_0^s(\tilde{\pi})$.*

We also want to show that our order structure behaves well as a functor. Already in [Putnam 2014], the functoriality for the groups alone is somewhat subtle; H^s is covariant for s-bijective factor maps and contravariant for u-bijective factor maps. We will show that the maps induced at the level of groups from s-bijective factor maps and u-bijective factor maps between the dynamical systems are positive group homomorphisms.

Theorem 2.14. *Suppose (X, φ) and (X', φ') are irreducible Smale spaces.*

(1) *If $\rho : (X, \varphi) \to (X', \varphi')$ is an s-bijective factor map, then the group homomorphism $\rho_0^s : H_0^s(X, \varphi) \to H_0^s(X', \varphi')$ of [Putnam 2014] is positive; that is,*

$$\rho_0^s(H_0^s(X, \varphi)^+) \subseteq H_0^s(X', \varphi')^+.$$

(2) *If $\rho : (X, \varphi) \to (X', \varphi')$ is a u-bijective factor map, then the group homomorphism $\rho_0^{s*} : H_0^s(X', \varphi') \to H_0^s(X, \varphi)$ of [Putnam 2014] is positive; that is,*

$$\rho_0^{s*}(H_0^s(X', \varphi')^+) \subseteq H_0^s(X, \varphi)^+.$$

A couple of remarks are in order. All of our results are stated for irreducible Smale spaces. They extend easily to Smale spaces in which every point is nonwandering, since any such Smale space is the disjoint union of a finite number of irreducible subsystems.

The ordered groups introduced by Krieger have a number of special features. They are unperforated: if, for any element a, na is positive for some $n \geq 1$, then a itself is positive. They also satisfy the Riesz interpolation property (see [Effros 1981] for details). At this point, it is not clear exactly which nice properties our ordered groups $H_0^s(X, \varphi)$ may have. However, one may observe, using [Amini et al. 2013], that they may have elements of finite order, which means that they are not unperforated in general. It may be reasonable to expect them to be weakly unperforated: if $na > 0$ for some $n \geq 1$, then $a > 0$.

3. Dimension groups and the Perron–Frobenius theorem

3A. ***Shifts of finite type.*** Shifts of finite type are usually defined in terms of the alphabets and (forbidden) words, but here we use an equivalent formulation in terms of graphs, which is more suitable for our purposes.

A graph G consists of finite sets G^0 and G^1, consisting of vertices and edges, respectively, and maps $i, t : G^1 \to G^0$, marking the initial and terminal points. The graph is drawn by depicting each vertex as a dot and each edge e as an arrow from $i(e)$ to $t(e)$.

A path of length k in G is a sequence $(e_1, \dots, e_k)$, with $e_i \in G^1$ for $1 \le i \le k$, such that $t(e_i) = i(e_{i+1})$ for $1 \le i < k$. Let G^k denote the set of all paths of length k. For each k, G^k is a graph with vertices G^{k-1} and edges G^k, and the initial and terminal maps

$$i(e_1, \dots, e_k) = (e_1, \dots, e_{k-1}), \quad t(e_1, \dots, e_k) = (e_2, \dots, e_k)$$

for $(e_1, \dots, e_k)$ in G^k. To any graph G, a pair (Σ_G, σ) is associated, where

$$\Sigma_G = \{(e_n)_{n\in\mathbb{Z}} \mid e_n \in G^1,\ t(e_n) = i(e_{n+1}),\ n \in \mathbb{Z}\},$$

$$\sigma : \Sigma_G \to \Sigma_G, \quad \sigma(e)_n = e_{n+1}.$$

This is a dynamical system with the metric

$$d(e, f) = \inf\{1, 2^{-K-1} \mid K \ge 0,\ e_{[1-K,K]} = f_{[1-K,K]}\}$$

on the Σ_G, where $e_{[K,L]} = (e_K, e_{K+1}, \dots, e_L)$ for $K \le L$, and $e_{[K+1,K]} = t(e_K) = i(e_{K+1})$. It is easy to see that (Σ_G, σ) is a Smale space with constants $\epsilon_X = \lambda = \frac{1}{2}$ and

$$[e, f]_k = \begin{cases} f_k & \text{if } k \le 0, \\ e_k & \text{if } k \ge 1. \end{cases}$$

The system (Σ_G, σ) is called the shift of finite type associated to the graph G.

3B. ***Dimension groups.*** Krieger [1980] defined two ordered groups in terms of the clopen sets for the shift of finite type, called the past and future dimension groups.

Suppose (Σ, σ) is a shift of finite type and $\Sigma^s(e)$ is the stable equivalence class of $e \in \Sigma$. By Proposition 2.1.12 in [Putnam 2014], the set $\Sigma^s(e)$ admits a topology that is second countable and locally compact. This may be different from the relative topology of Σ. Let $\mathrm{CO}(\Sigma, \sigma)$ be the set of nonempty, open and compact subsets of $\Sigma^s(e)$, over all e in Σ, and $\sim$ be the smallest equivalence relation on $\mathrm{CO}(\Sigma, \sigma)$ such that $E \sim F$ if $[E, F] = E$ and $[F, E] = F$ and $E \sim F$ if and only if $\sigma(E) \sim \sigma(F)$, and let $[E]$ denote the equivalence class of E.

Let $\mathcal{D}^s(\Sigma, \sigma)$ be the free abelian group on $\sim$-equivalence classes of $\mathrm{CO}^s(\Sigma, \sigma)$ and H be the subgroup generated by $[E \cup F] - [E] - [F]$, where E, F and $E \cup F$ are in $\mathrm{CO}^s(\Sigma, \sigma)$ and E and F are disjoint. The group $D^s(\Sigma, \sigma)$ is defined to be $\mathcal{D}^s(\Sigma, \sigma)/H$.

The order is obtained by defining

$$\mathcal{D}^s(\Sigma, \sigma)^+ = \{[E] \mid E \in \mathrm{CO}^s(\Sigma, \sigma)\},$$

and then

$$D^s(\Sigma, \sigma)^+ = \{a + H \mid a \in \mathcal{D}^s(\Sigma, \sigma)^+\}.$$

The ordered abelian group $D^u(\Sigma, \sigma)$ is defined in a similar way, by replacing the unstable equivalence classes $\Sigma^u(e)$ by $\Sigma^s(e)$. Krieger showed how this ordered group could be computed from the underlying graph of the shift of finite type.

Before going into more detail, we need some notation. If A is a finite set, then the free abelian group generated by A, $\mathbb{Z}A$, is an ordered abelian group with the positive cone $\{z_1a_1 + \cdots + z_na_n \mid z_1, \ldots, z_n \in \mathbb{Z}^+ \cup \{0\}, a_1, \ldots, a_n \in A, n \in \mathbb{N}\}$. In our notation, A is considered as a subset of $\mathbb{Z}A$. If A, B are finite sets and $\tau : A \to B$ is any function, then there is a unique positive homomorphism $\Gamma : \mathbb{Z}A \to \mathbb{Z}B$ extending τ. For the finite set A, the integer-valued bilinear form $\langle\, , \rangle$ is defined on $\mathbb{Z}A \times \mathbb{Z}A$ which is additive in each variable, and for each $a, b \in A$,

$$\langle a, b\rangle = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{if } a \neq b. \end{cases}$$

For two finite sets A, B and a homomorphism $h : \mathbb{Z}A \to \mathbb{Z}B$, there is a unique homomorphism $h^* : \mathbb{Z}B \to \mathbb{Z}A$ such that

$$\langle h(a), b\rangle = \langle a, h^*(b)\rangle$$

for all a in $\mathbb{Z}A$ and b in $\mathbb{Z}B$.

Let $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_n\}$. We associate a matrix $[h_{ij}]_{n\times m}$ to the homomorphism h such that the entry h_{ij} is equal to the coefficient b_j in $h(a_i)$ when $h(a_i)$ is written in terms of the generating set B. We have

$$\langle h(a), b\rangle = \langle a, h^*(b)\rangle$$

for a in $\mathbb{Z}A$ and b in $\mathbb{Z}B$, that is, $[h^*_{ij}]_{m\times n} = ([h_{ij}]_{n\times m})^T$, where M^T denotes the transpose of a matrix M.

Now we compute the dimension group in terms of the underlying graph of the shift of finite type. Let (G^0, G^1, i, t) be a graph and (Σ_G, σ) be the associated shift of finite type. Suppose $\mathbb{Z}G^0$ is the free abelian group on the generating set G^0, and consider the homomorphism

$$\gamma^s_G : \mathbb{Z}G^0 \to \mathbb{Z}G^0, \qquad \gamma^s_G(v) = \sum_{t(e)=v} i(v) \quad (v \in G^0).$$

The past dimension group $D^s(G)$ is defined as the inductive limit of the system

$$\mathbb{Z}G^0 \xrightarrow{\gamma^s_G} \mathbb{Z}G^0 \xrightarrow{\gamma^s_G} \cdots .$$

Since $\mathbb{Z}G^0$ is an ordered group and γ_G^s is a positive homomorphism, $D^s(G)$ inherits an order structure in a natural way. Let us give a brief and simple description of the elements of $D^s(G)$. Two points (a, m) and (b, n) in $\mathbb{Z}G^0 \times \mathbb{N}$ are equivalent, denoted $(a, m) \sim_s (b, n)$, if there exists $l \in \mathbb{N} \cup \{0\}$ such that $(\gamma_G^s)^{n+l}(a) = (\gamma_G^s)^{m+l}(b)$. The equivalence class of (a, m) is denoted by $[a, m]_s$ and $D^s(G)$ is the set of all equivalence classes. The positive cone in this group consists of those elements $[a, m]_s$ with $(\gamma_G^s)^l(a) \in (\mathbb{Z}G^0)^+$ for some $l \in \mathbb{N}$.

The future dimension group for the graph (G^0, G^1, i, t) is defined in a similar way, by replacing the homomorphism $\gamma_G^s : \mathbb{Z}G^0 \to \mathbb{Z}G^0$ by γ_G^u, where

$$\gamma_G^u(v) = \sum_{i(e)=v} t(v)$$

for all v in G^0. Note that $\gamma_G^u = (\gamma_G^s)^*$.

It is worth noting that in some places in the computation of the homology, it is necessary to use the graph G^k instead of G, which does not affect the answer. This can be viewed as a consequence of the next theorem. The next two results appear as Theorems 3.3.3 and 3.5.5 in [Putnam 2014], but without the order structure.

Theorem 3.1. *Suppose G is a graph, (Σ_G, σ) is the associated shift of finite type and $k \geq 1$. The homomorphism Ψ from $D^s(\Sigma_G, \sigma)$ to $D^s(G^k)$, defined on the generating elements by $\Psi([\Sigma_G^s(e, 2^{-j})]) = [e^{[1-j,k-j-1]}, j-k+1]$, $e \in \Sigma_G$, $j \geq k$, is an order isomorphism.*

We recall some notation from Section 3.1 of [Putnam 2014], that if B is any subset of A, $\mathrm{Sum}(B) = \sum_{b \in B} b \in \mathbb{Z}A$.

Theorem 3.2. *Let G and H be graphs with a graph homomorphism $\pi : H \to G$ and suppose that the associated map $\pi : (\Sigma_H, \sigma) \to (\Sigma_G, \sigma)$ is s-bijective, $k \geq 1$, and K satisfies the conclusion of Lemma 2.7.1 in [Putnam 2014] for π. The induced map $\pi^s[a, j] = [\pi^{s,K}(a), j]$ from $D^s(H^k)$ to $D^s(G^{k+K})$ is a positive homomorphism, where $a \in \mathbb{Z}H^{k-1}$, $j \geq 1$ and $\pi^{s,K}(q) = \mathrm{Sum}\{\pi(q') \mid q' \in H^{k+K}, t^K(q') = q\}$.*

3C. ***The Perron–Frobenius theorem.*** Let G be a finite directed graph. The adjacency matrix, A_G, is $\#G^0 \times \#G^0$ and has entries that are the number of edges between the different vertices of G. The shift of finite type (Σ_G, σ) is irreducible if and only if the graph G is irreducible, in the sense that, for each ordered pair of vertices u and v in G, there exists a path p in G starting at u and terminating at v. This is also equivalent to the adjacency matrix being irreducible, in the sense that for each ordered pair of indices i, j, there is some nonnegative integer n such that $(A_G)_{i,j}^n > 0$.

The shift of finite type (Σ_G, σ) is mixing if and only if there is a positive integer n such that for every ordered pair of vertices u and v in G, there exists a path of length n in G starting at u and terminating at v. This is also equivalent to the adjacency matrix being primitive; that is, there is some positive integer n such that $(A_G)^n_{i,j} > 0$ for all $1 \leq i, j \leq m$. If this holds for some fixed n, it also holds for all higher values of n [Lind and Marcus 1995].

Let us recall the Perron–Frobenius theorem [Lind and Marcus 1995, Theorem 4.2.3]. If A is a nonnegative irreducible square matrix, then it has a positive eigenvalue λ_A and a right positive eigenvector $\boldsymbol{v}_A$ associated to λ_A, called the Perron eigenvalue and the Perron eigenvector, respectively, such that $|\mu| \leq \lambda_A$ for every eigenvalue μ of A, and the corresponding eigenspace of λ_A is both geometrically and algebraically simple.

Given our presentation using homomorphisms rather than matrices, we state this in the following fashion. We apply this to both the adjacency matrix for the graph and its transpose, but these share the same Perron eigenvalue. Assuming that the graph G is irreducible, there are $\lambda_G > 0$ and vectors v^s_G, v^u_G in $\mathbb{R}^+ G^0$ such that

$$\gamma^s_G(v^s_G) = \lambda_G v^s_G, \; \gamma^u_G(v^u_G) = \lambda_G v^u_G.$$

We have extended the definition of γ^s_G, γ^u_G in the obvious way. We remark that if we replace G by G^k, for some $k \geq 1$, we obtain a higher block presentation of the shift (see Definition 1.4.1 of [Lind and Marcus 1995]). The Perron eigenvectors are changed, but not the eigenvalue: $\lambda_{G^k} = \lambda_G$.

The Perron eigenvalue in the above result is related to the notion of entropy as the below result shows. This could be defined for a general dynamical system, but here we only deal with the shifts of finite type. Let G be a graph and (Σ_G, σ) be the corresponding shift of finite type. The entropy of (Σ_G, σ) is defined [Lind and Marcus 1995, Definition 4.1.1] by

$$h(\Sigma_G, \sigma) = \lim_{n \to \infty} \frac{1}{n} \log \# G^n,$$

where $\# G^n$ is the number of paths of length n in G.

Theorem 3.3 [Lind and Marcus 1995, Theorem 4.4.4]. *If G is a graph, then we have $h(\Sigma_G, \sigma) = \log \lambda_G$.*

The Perron–Frobenius theorem also has a nice application for the computation of the order structure of $D^s(\Sigma_G, \sigma)$, particularly in the mixing case. This follows from Corollary 4.2 and Theorem 6.1 of [Effros 1981].

Theorem 3.4. *Let G be a finite directed graph whose associated shift of finite type is mixing. For any $n \geq 1$ and a in $\mathbb{Z}G^0$, the element $[a, n]$ is in $D^s(G)^+ - \{0\}$ if and only if $\langle a, v^u_G \rangle$ is positive.*

We end this section with a result which gives a sufficient condition for the surjectivity of maps between shifts of finite type.

Theorem 3.5 [Lind and Marcus 1995, Corollary 4.4.9]. *Suppose G and H are graphs and $\pi : (\Sigma_G, \sigma) \to (\Sigma_H, \sigma)$ is a finite-to-one map. If the graph H is irreducible and $h(\Sigma_G, \sigma) = h(\Sigma_H, \sigma)$, then π is onto.*

4. Irreducible s/u-bijective pairs and fibred products

4A. ***Irreducible bijective pairs.*** The proof of the existence of s/u-bijective pairs comes from [Putnam 2005]. Our proof of the existence of irreducible ones must go back to the same starting point to see how the results of that paper can be improved.

Suppose (X, φ) and (Y, ψ) are irreducible Smale spaces and $\pi : (X, \varphi) \to (Y, \psi)$ is an almost one-to-one map. In [Putnam 2005], it was shown that there exist irreducible Smale spaces $(\tilde{X}, \tilde{\varphi})$, $(\tilde{Y}, \tilde{\psi})$ and factor maps $\alpha, \beta, \tilde{\pi}$ such that the following diagram is commutative:

(4-1)
$$\begin{array}{ccc} (\tilde{X}, \tilde{\varphi}) & \xrightarrow{\tilde{\pi}} & (\tilde{Y}, \tilde{\psi}) \\ {\scriptstyle \alpha}\big\downarrow & & \big\downarrow{\scriptstyle \beta} \\ (X, \varphi) & \xrightarrow{\pi} & (Y, \psi) \end{array}$$

Moreover, the maps α, β are u-bijective and the map $\tilde{\pi}$ is s-bijective. Regrettably, it was not shown that $\alpha, \beta, \tilde{\pi}$ are almost one-to-one, which is what we undertake now. In fact, it will be enough to consider β. (The space (X, φ) is appearing in a somewhat unfortunate position as the domain, but we follow [Putnam 2005] for the moment.)

The proof involves finding a periodic point y_0 in Y with $\pi^{-1}\{y_0\} = \{x_0\}$, a single point in X. Then W is the unstable set of the orbit of x_0 and it is shown that $\pi(W)$ is the unstable set of the orbit of y_0. Let d_X, d_Y be the metrics on X and Y, respectively. We view X and Y as the completions of the spaces (W, d_X) and $(\pi(W), d_Y)$. The proof of [Putnam 2005] involves introducing new metrics on W and $\pi(W)$, δ_X and δ_Y, respectively, so that $\tilde{X}$ and $\tilde{Y}$ are their completions. As these new metrics are greater than or equal to the old ones, the factor maps α, β appear automatically.

Here, we claim that $\beta^{-1}\{y_0\} = \{x_0\}$. (The references here will all be to [Putnam 2005].) To see this, it suffices to consider a sequence y_n in $\pi(W)$ which is Cauchy in δ_Y and converges to y_0 in d_Y and prove that it converges to y_0 in δ_Y. For n sufficiently large, $[y_0, y_n]$ is defined, and using part 4 of Lemma 2.18, we have

$$\begin{aligned} \delta_Y(y_0, y_n) &\leq \delta_Y(y_0, [y_0, y_n]) + \delta_Y([y_0, y_n], y_n) \\ &\leq \delta_Y(y_0, [y_0, y_n]) + (1 - r\lambda)^{-1} d_Y([y_0, y_n], y_n). \end{aligned}$$

It suffices for us to show that $[y_0, y_n]$ converges to y_0 in δ_Y. By replacing y_n by $[y_0, y_n]$, we may assume that y_n is in $V^s(y_0, \epsilon_Y)$. By part 2 of Proposition 2.12, we may assume that y_n and y_0 are ρ-compatible and then by Lemma 2.10, for all $k \geq 0$ there is $N_k \geq 1$ such that $g^{-k}(y_0)$ and $g^{-k}(y_n)$ are ρ-compatible for $n \geq N_k$.

Let $\epsilon > 0$ be given. From the definition of δ_Y^0 in Definition 2.14, it is bounded by D. We may find $K \geq 1$ such that $\sum_{k>K} r^k D < \epsilon/2$. Find $N \geq \max\{N_k \mid 1 \leq k \leq K\}$ so that for $n \geq N$ and $0 \leq k \leq K$ we have

$$d_Y(g^{-k}(y_0), g^{-k}(y_n) < \frac{\epsilon}{2(K+1)}.$$

It follows from Definition 2.17 and part 4 of Lemma 2.15 that for such n,

$$\begin{aligned}
\delta_Y(y_0, y_n) &= \sum_{k=0}^{\infty} r^k \delta_Y^0(g^{-k}(y_0), g^{-k}(y_n)) \\
&\leq \sum_{k=0}^{K} d_Y(g^{-k}(y_0), g^{-k}(y_n)) + \sum_{k=K+1}^{\infty} r^k D \\
&< \sum_{k=0}^{K} \frac{\epsilon}{2(K+1)} + \frac{\epsilon}{2} \\
&= \epsilon.
\end{aligned}$$

Exactly as in [Putnam 2005], we apply this result as follows. We begin with our irreducible Smale space (X, φ) and find an irreducible shift of finite type (Σ, σ) and an almost one-to-one factor map $\pi : (\Sigma, \sigma) \to (X, \varphi)$. The system which is called $(\tilde{Y}, \tilde{g})$ above, we denote by (Z, ζ) and the map β by π_u. The fact that Z has totally disconnected stable sets follows from the facts that $\tilde{\Sigma}$ is also a shift of finite type and $\tilde{\pi}$ is s-bijective. Now, we also know that there is x_0 in X with $\#\pi_u^{-1}\{x_0\} = 1$.

We next want to show that if there is a single point x with $\#\pi^{-1}\{x\} = 1$, this will also hold for all points with dense forward or backward orbit if we also assume that π is s-bijective or u-bijective. Recall that the forward orbit of a point x is $\{\varphi^n(x) \mid n \geq 0\}$, while the backward orbit is $\{\varphi^n(x) \mid n \leq 0\}$.

Lemma 4.1. *Let (Y, ψ) and (X, φ) be Smale spaces and $\pi : (Y, \psi) \to (X, \varphi)$ be an s-bijective (or u-bijective) factor map. Assume there is x_0 in X such that $\pi^{-1}\{x_0\} = 1$. Then for any point x in X with a dense forward (backward) orbit, we have $\#\pi^{-1}\{x\} = 1$.*

Proof. We prove the result in the case that π is s-bijective. List $\pi^{-1}\{x\} = \{y_1, \ldots, y_I\}$. Since the orbit of x is dense, we may find an increasing sequence of positive integers n_k such that $\varphi^{n_k}(x)$ converges to x_0. Passing to a subsequence, we may assume that for each $1 \leq i \leq I$, the sequence $\psi^{n_k}(y_i)$ converges to some point of Y, and by continuity, these points must all lie in $\pi^{-1}(x_0)$. It remains to show that no two such sequences can have the same limit. If there is $1 \le i \ne j \le I$, then $d(\psi^{n_k}(y_i), \psi^{n_k}(y_j))$ tends to zero as k goes to infinity. Then we have

$$\pi(\psi^{n_k}(y_i)) = \varphi^{n_k}(\pi(y_i)) = \varphi^{n_k}(x) = \varphi^{n_k}(\pi(y_j)) = \pi(\psi^{n_k}(y_i)).$$

Using the fact that π is s-bijective, Proposition 2.5.2 in [Putnam 2014] implies that, for k sufficiently large,

$$\psi^{n_k}(y_i) \in Y^u(\psi^{n_k}(y_j), \epsilon_\pi),$$

which implies that

$$y_i \in Y^u(y_j, \lambda^{n_k}\epsilon_\pi).$$

Since this is true for all k, we conclude $y_i = y_j$, and we are done. □

The set of points with a dense forward orbit is rather large in an irreducible system. The following result is standard; see, for example, Theorem 5.9 of [Walters 1982].

Proposition 4.2. *Let (X, φ) be a dynamical system, with X a compact metric space. If (X, φ) is irreducible, then the set of all points x with dense forward orbit is a dense G_δ subset of X.*

It is probably worth noting that Lemma 4.1 and Proposition 4.2 together prove the following.

Corollary 4.3. *Let (Y, ψ) and (X, φ) be Smale spaces and $\pi : (Y, \psi) \to (X, \varphi)$ be an s-bijective (or u-bijective) factor map. Then π is almost one-to-one if and only if there is a point x_0 in X such that $\#\pi^{-1}\{x_0\} = 1$.*

We have also now proved Theorem 2.3, that every irreducible Smale space has an irreducible s/u-bijective pair.

4B. ***The fibred product of maps.*** Let $\pi_1 : (Y, \psi) \to (X, \varphi)$ and $\pi_2 : (Z, \zeta) \to (X, \varphi)$ be maps between Smale spaces and

$$Y \times_X Z = \{(y, z) \in Y \times Z \mid \pi_1(y) = \pi_2(z)\}$$

be the fibred product of π_1, π_2, with the relative topology of $Y \times Z$. By Theorem 2.4.2 in [Putnam 2014], $Y \times_X Z$ is a Smale space, with $\psi \times \zeta(y, z) = (\psi(y), \zeta(z))$ for $(y, z) \in Y \times_X Z$. We note that there are natural maps $\rho_2 : (Y \times_X Z, \psi \times \zeta) \to (Z, \zeta)$ defined by $\rho_2(y, z) = z$ and $\rho_1 : (Y \times_X Z, \psi \times \zeta) \to (Y, \psi)$ defined by $\rho_1(y, z) = y$. We also note that if π_1 is s-bijective (or u-bijective), then so is ρ_2.

The drawback is that the fibred product of maps on irreducible Smale spaces is not irreducible in general. In this section, we prove the irreducibility of the fibred product $(Y \times_X Z, \psi \times \zeta)$ under certain natural conditions.

Proposition 4.4. *Let $\pi_1 : (Y, \psi) \to (X, \varphi)$ and $\pi_2 : (Z, \zeta) \to (X, \varphi)$ be either s-bijective or u-bijective, almost one-to-one factor maps between irreducible Smale spaces. Then the natural maps ρ_1 and ρ_2 from the fibred product to Y and Z, respectively, are also almost one-to-one.*

Proof. The set of x in X with $\#\pi_1^{-1}\{x\} = 1$ is a dense G_δ, as is the set of x with $\#\pi_2^{-1}\{x\} = 1$. It follows that their intersection is nonempty. If x is in this intersection and $\pi_1(y) = x$, $\pi_2(z) = x$, it is a simple matter to see that

$$\rho_2^{-1}\{z\} = \{(y, z)\} = \rho_1^{-1}\{y\}.$$

We complete the proof by noting that ρ_1 and ρ_2 are also either s-bijective or u-bijective and by recalling Corollary 4.3 □

We will need two technical results for the proof of Theorem 2.6. The first is a characterization of irreducibility.

Lemma 4.5. *Let (X, φ) be a Smale space. If there exists a point x in X whose forward orbit clusters on every periodic point of X, then (X, φ) is irreducible.*

Proof. Let y be an accumulation point of the backward orbit of x. It is clearly nonwandering and so it is in the closure of the periodic points. It follows that y is also a limit point of the forward orbit of x. By patching the forward orbit of x that gets close to y with part of the backward orbit of x that begins close to y, we can form pseudo-orbits from x to itself and conclude that x is in the nonwandering set. The orbit of x will remain in the same irreducible component of the nonwandering set. Hence all periodic points are in the same irreducible component. This implies that there is only one irreducible component. If X contained a wandering point, its forward orbit and backward orbits would limit on two distinct irreducible components. As this is not possible, X has no wandering points. □

Lemma 4.6. *Let $\pi_s : (Y, \psi) \to (X, \varphi)$ be an s-bijective almost one-to-one factor map between irreducible Smale spaces. Let x_0 be a periodic point of X with $\pi^{-1}\{x_0\} = \{y_1, y_2, \dots, y_I\}$. For $\delta, \epsilon > 0$, put $U = [X^u(x_0, \delta), X^s(x_0, \delta)]$ and, for $1 \le i \le I$, let*

$$V_i = \{x \in U \mid \pi^{-1}\{[x, x_0]\} \subseteq Y(y_i, \epsilon)\},$$

where $Y(y_i, \epsilon)$ denotes the open ball at y_i of radius ϵ. Then there exist arbitrarily small positive pairs δ, ϵ such that

(i) *V_i is open,*

(ii) *V_i is nonempty,*

(iii) *$[V_i, U] \subset V_i$.*

Proof. First choose ϵ to be smaller than ϵ_π and also smaller than half of the distance between y_i and y_j, over all $1 \le i \ne j \le I$. Then choose $\delta > 0$ so that Lemma 2.5.11 of [Putnam 2014] holds. It follows easily from the continuity of the bracket and Lemma 2.5.9 of [Putnam 2014] that V_i is open for all i. Let us next fix i and prove that V_i is nonempty. By hypothesis, there exists a point x' with dense forward orbit and $\#\pi^{-1}\{x'\} = 1$. Notice that any point in the orbit of x' also has these properties, as does any point stably equivalent to a point in the orbit of x'. Let y' in Y be the unique point with $\pi(y') = x'$. Since (Y, ψ) is irreducible, the stable equivalence class of the orbit of y' is dense. So there exists y'' stably equivalent to some point in the orbit of y' in $Y^u(y_i, \epsilon)$. Let us check that $\pi(y'')$ is in V_i. As $\epsilon < \epsilon_\pi$, we know $[\pi(y''), x_0] = \pi([y'', y_i]) = \pi(y'')$ and hence $\pi^{-1}\{[\pi(y''), x_0] = \{y''\}\}$ is in $Y(y_i, \epsilon)$. Finally, we verify the last condition. Suppose that x is in V_i and x_1 is in $X^u(x_0, \delta)$ and x_2 in $X^u(x_0, \delta)$. Since $[x, [x_1, x_2]] = [x, x_2]$ is in V_i, $[[x, x_2], x_0] = [x, x_0]$ and the conclusion follows. $\square$

We have most of the ingredients for the proof of Theorem 2.6, but for the last statement, we need some convenient characterizations of mixing.

Lemma 4.7. *Suppose (X, φ) is an irreducible Smale space. The following are equivalent.*

(1) *(X, φ) is mixing.*

(2) *For any periodic point x in X, we have $X^s(x) \cap X^u(\varphi(x)) \ne \varnothing$ and $X^u(x) \cap X^s(\varphi(x)) \ne \varnothing$.*

(3) *For some periodic point x in X, we have $X^s(x) \cap X^u(\varphi(x)) \ne \varnothing$ and $X^u(x) \cap X^s(\varphi(x)) \ne \varnothing$.*

Proof. This is a consequence of Smale's spectral decomposition. Let $\sim$ be the equivalence relation on the periodic points of (X, φ) in Smale's spectral decomposition, that is, for two periodic points $x, y \in X$, $x \sim y$ if and only if $X^s(x) \cap X^u(y) \ne \varnothing$ and $X^u(x) \cap X^s(y) \ne \varnothing$. Then there are pairwise disjoint clopen sets $X_1, \dots, X_N$ whose union is X, $\varphi(X_i) = \varphi(X_{i+1})$ for $1 \le i \le N-1$, $\varphi(X_N) = X_1$ and (X_i, φ^N) is a mixing Smale space, for every $1 \le i \le N$. Moreover each X_i is the closure of an equivalence class of $\sim$ and these sets are unique up to relabeling.

If we assume that (X, φ) is mixing, then N above must equal 1 and the second condition holds. The second part obviously implies the third. Finally, if x is a periodic point, so is $\varphi(x)$. Suppose $x \in X_i$ for some $1 \le i \le N-1$, $X^s(x) \cap X^u(\varphi(x)) \ne \varnothing$ and $X^u(x) \cap X^s(\varphi(x)) \ne \varnothing$. Then $x \sim \varphi(x)$, thus $\varphi(x) \in X_i \cap X_{i+1}$. Since the X_i are pairwise disjoint, $X_i = X_{i+1}$. The same argument shows that $X_i = X_{i+1} = \cdots = X_N$. Similarly, if $x \in X_N$, then $X_N = X_1 = \cdots = X_{N-1}$. Therefore, $N = 1$, hence $X = X_i$ and (X, φ) is a mixing space. $\square$

Proposition 4.8. *If $\pi : (Y, \psi) \to (X, \varphi)$ is an almost one-to-one factor map between Smale spaces, (Y, ψ) is irreducible and (X, φ) is mixing, then (Y, ψ) is mixing also.*

Proof. We will verify the condition of the last lemma. Suppose y is in Y and x is in X such that $\pi^{-1}\{x\} = \{y\}$. Since (X, φ) is mixing, it is irreducible and hence by Proposition 2.3 in [Putnam 2005], we can find a periodic point $x_0 \in X$ with $\#\pi^{-1}\{x_0\} = 1$. Let $y_0 \in Y$ with $\pi(y_0) = x_0$. Since x_0 is periodic and π is finite-to-one, y_0 is a periodic point. By the argument of the proof of Lemma 2.4 in [Putnam 2005], $\pi^{-1}(X^s(x_0)) = Y^s(y_0)$, $\pi^{-1}(X^s(\varphi(x_0))) = Y^s(\psi(y_0))$, $\pi^{-1}(X^u(x_0)) = Y^u(y_0)$ and $\pi^{-1}(X^u(\varphi(x_0))) = Y^u(\psi(y_0))$. Since (X, φ) is a mixing Smale space, we have

$$X^s(x_0) \cap X^u(\varphi(x_0)) \neq \varnothing, \qquad X^u(x_0) \cap X^s(\varphi(x_0)) \neq \varnothing,$$

which implies

$$\begin{aligned} Y^s(y_0) \cap Y^u(\psi(y_0)) &= \pi^{-1}(X^s(x_0)) \cap \pi^{-1}(X^u(\varphi(x_0))) \\ &= \pi^{-1}\big(X^s(x_0) \cap X^u(\varphi(x_0))\big) \neq \varnothing, \\ Y^u(y_0) \cap Y^s(\psi(y_0)) &= \pi^{-1}(X^u(x_0)) \cap \pi^{-1}(X^s(\varphi(x_0))) \\ &= \pi^{-1}\big(X^u(x_0) \cap X^s(\varphi(x_0))\big) \neq \varnothing. \end{aligned}$$

Therefore, by Lemma 4.7, (Y, ψ) is mixing. □

Proof of Theorem 2.6. The sets of points of X with dense forward and backward orbits are both dense G_δ's and so their intersection is nonempty. Let x be a point in X with a dense forward orbit and a dense backward orbit. Let y and z be its unique preimages under π_s and π_u, respectively. By Lemma 4.5, it suffices to prove that the forward orbit of (y, z) clusters on every periodic point. Let (y_1, z_1) be a periodic point in the fibred product. Let $x_1 = \pi_s(y_1) = \pi_u(z_1)$. Enumerate $\pi_s^{-1}\{x_1\} = \{y_1, \dots, y_I\}$ and $\pi_u^{-1}\{x_1\} = \{z_1, \dots, z_J\}$.

For small δ, ϵ, let V_i, $1 \le i \le I$, and W_j, $1 \le j \le J$, be the result of applying Lemma 4.6 to the maps π_s and π_u, respectively. Observe that since π_u is u-bijective, the last condition on W_j is $[U, W_j] \subseteq W_j$. We have

$$V_1 \cap W_1 \supseteq [V_1, U] \cap [U, W_1] \supseteq [V_1, W_1],$$

which is clearly nonempty. Also $V_1 \cap W_1$ is open. It follows that there is $n \ge 1$ with $\varphi^n(x) \in V_1 \cap W_1$. This implies that $\psi^n(y) \in Y(y_1, \epsilon)$ and $\zeta^n(z) \in Z(z_1, \epsilon)$. Since ϵ was arbitrary, this completes the proof of the first part. The mixing case follows from two applications of Proposition 4.8. □

5. Homology

In this section, we prove the main results on the homology of Smale spaces, stated in the first section. If (X, φ) is a Smale space, then so is (X, φ^n), for any positive integer n, and if $\pi = (Y, \psi, \pi_s, Z, \zeta, \pi_u)$ is an s/u-bijective pair for (X, φ), then

$\pi_n = (Y, \psi^n, \pi_s, Z, \zeta^n, \pi_u)$ is an s/u-bijective pair for (X, φ^n). The results in Chapters 4 and 5 of [Putnam 2014] show that $(C_{\mathcal{Q},\mathcal{A}}(\Sigma(\pi)), d_{\mathcal{Q},\mathcal{A}}(\Sigma(\pi)))$ and $(C_{\mathcal{Q},\mathcal{A}}(\Sigma(\pi_n)), d_{\mathcal{Q},\mathcal{A}}(\Sigma(\pi_n)))$ admit the same cocycle and coboundaries. On the other hand, by Smale's spectral decomposition, for every irreducible Smale space (X, φ), X can be written as a union of pairwise disjoint clopen subsets $X_1, \ldots, X_L$ such that $\varphi^L(X_i) = X_i$ for each $1 \leq i \leq L$, and the (X_i, φ^L) are mixing Smale spaces [Smale 1967]. Hence

$$H_N^s(X, \varphi) \cong H_N^s(X, \varphi^L) \cong \bigoplus_{i=1}^{L} H_N^s(X_i, \varphi^L)$$

for any positive integer N, and this along with Theorem 2.13 allows us to replace an irreducible Smale space by a mixing one.

Under the assumption that (X, φ) is mixing, we find π, an irreducible s/u-bijective pair for (X, φ). It follows at once from Propositions 4.8 and 4.4 and from Theorem 2.6 that $(\Sigma_{0,0}(\pi), \sigma) = (Y \times_X Z, \psi \times \zeta)$ is mixing and ρ_s and ρ_u are almost one-to-one.

We start with two lemmas that are simpler versions of Theorems 2.10 and 2.11. Both of these consider the following situation: a shift of finite type (Σ, σ), a Smale space (Y, ψ) and a factor map $\rho : (\Sigma, \sigma) \to (Y, \psi)$ which is either s-bijective or u-bijective. In Chapter 4 of [Putnam 2014], a complex is formed from such a map. It is a simpler object than the double complex associated to an s/u-bijective pair, but its importance lies in the fact that the individual rows and columns of the double complex all arise in this fashion. Applying this to our map $\rho_s : (\Sigma_{0,0}(\pi), \sigma) \to (Y, \psi)$ yields the bottom row of our double complex. Similarly, applying this to our map $\rho_u : (\Sigma_{0,0}(\pi), \sigma) \to (Z, \zeta)$ yields the left column of our double complex.

To a factor map ρ as above, we let

$$\Sigma_N(\rho) = \{(x_0, \ldots, x_N) \in \Sigma^{N+1} \mid \rho(x_0) = \cdots = \rho(x_N)\} \quad \text{for all } N \geq 0.$$

There are obvious maps $\delta_n : \Sigma_N(\rho) \to \Sigma_{N-1}(\rho)$ for $0 \leq n \leq N$ and $N \geq 1$.

Lemma 5.1. *Let (Σ, σ) be a mixing shift of finite type, (Y, ψ) be a mixing Smale space and $\rho_s : (\Sigma, \sigma) \to (Y, \psi)$ be an s-bijective, almost one-to-one factor map. Then* $\mathrm{Im}(\delta_0^s - \delta_1^s) \cap D^s(\Sigma_0(\rho_s))^+ = \{0\}$.

Proof. We begin by finding a graph G whose associated shift (Σ_G, σ) is conjugate to (Σ, σ). (We suppress the conjugacy in our notation.) From Theorem 4.2.8 in [Putnam 2014], this G may be chosen so that the map ρ_s is regular. (The definition of regular is given in Definition 2.3.3 of [Putnam 2014]. We will not really need it here, but we will indicate where it is used shortly.)

If (x_0, x_1) is in $\Sigma_1(\rho_s)$, then x_0 and x_1 are bi-infinite paths in G and if we take their 0-th entries we obtain a pair in $G^1 \times G^1$. We let G_1^1 be the set of all such pairs over all (x_0, x_1) in $\Sigma_1(\rho_s)$ and G_1^0 be the image of this set under $t \times t$. Then G_1 is a graph with obvious i, t maps. The significance of our choice that ρ_s is regular is that

$$\Sigma_{G_1} = \Sigma_1(\rho_s).$$

(The elements of the set on the left are infinite sequences of pairs of edges of G, while those on the right are pairs of infinite sequences of edges of G, but we feel no confusion will arise from equating the two.)

It is clear from letting $x_0 = x_1$ that G_1^i contains all pairs (a, a) where a is in G^i. We denote this subgraph by G_1^Δ. As ρ_s is s-bijective, any edge in G_1 which terminates in G_1^Δ must actually be in G_1^Δ.

Let G_1' consist of those vertices not in G_1^Δ and all edges whose initial vertex is not in G_1^Δ. This is a graph and its infinite path space $\Sigma_{G_1'}$ maps to Σ_G by δ_0. If this map is surjective, then every point of Y has at least two distinct preimages under ρ_s, contrary to our hypothesis. Using Theorem 3.5, we conclude that

$$\log \lambda_{G_1'} = h(\Sigma_{G_1'}, \sigma) = h(\delta_0(\Sigma_{G_1'}, \sigma)) < h(\Sigma_G, \sigma) = \log \lambda_G.$$

It follows that there is a constant C such that $\#(G_1')^j \leq C(\lambda_{G_1'})^j$ for all $j \geq 1$.

Following the discussion prior to Theorem 4.2.13 of [Putnam 2014], for $k \geq 0$ we choose B_1^k to be a subset of G_1^k which contains no paths of the form (p_0, p_1) if $p_0 = p_1$ and for $p_0 \neq p_1$, it contains exactly one of (p_0, p_1) and (p_1, p_0). Following Theorem 4.2.13 of [Putnam 2014], for any $k \geq 0$, $j \geq 1$, p in B_1^k, we let

$$t_{\mathcal{A}}^*(p, j) = \{(q, \alpha) \in G_1^{k+j} \times S_2 \mid t^j(q) = p, i^j(q) \cdot \alpha \in B_1^k\}.$$

The point here is that any path q with $i^j(q) = p \in B_1^k \subseteq (G_1')^k$ must lie entirely in G_1'. It is then clear that $\# t_{\mathcal{A}}^*(p, j) \leq C(\lambda_{G_1'})^{j+k}$. The map $\gamma^s_{B_1^k} : \mathbb{Z}B_1^k \to \mathbb{Z}B_1^k$ is defined just before Theorem 4.2.13 of [Putnam 2014]. We conclude from the first part of Theorem 4.2.13 of [Putnam 2014] that if $\eta : \mathbb{Z}B_1^{k-1} \to \mathbb{R}$ is any group homomorphism and a is in $\mathbb{Z}B_1^{k-1}$, then there is a constant D (depending on a) such that $\eta((\gamma^s_{B_1})^j(a)) < D(\lambda_{G_1'})^j$ for all $j \geq 1$.

Consider the diagram

$$\begin{tikzcd}
\mathbb{Z}B_1^k \arrow[r, "\gamma^s_{B_1^k}"] \arrow[d, "Q"'] & \mathbb{Z}B_1^k \arrow[d, "Q"] \\
\mathcal{Q}(G_1^k, S_2) \arrow[r, "\gamma^s_{G_1^k}"] \arrow[d, "\delta_0^{s,K} - \delta_1^{s,K}"'] & \mathcal{Q}(G_1^k, S_2) \arrow[d, "\delta_0^{s,K} - \delta_1^{s,K}"] \\
\mathbb{Z}G^{k+K} \arrow[r, "\gamma^s_{G^{k+K}}"] & \mathbb{Z}G^{k+K}
\end{tikzcd}$$

The second part of Theorem 4.2.13 of [Putnam 2014] tells us that the top square commutes and that the vertical maps are isomorphisms. The bottom square commutes by Theorem 4.2.3, Definition 4.2.4 and Theorem 4.2.5 of [Putnam 2014].

We consider $\eta(\,\cdot\,) = \langle(\delta_0^{s,K} - \delta_1^{s,K}) \circ Q(\,\cdot\,),\, v^u_{G^{k+K}}\rangle$, where $v^u_{G^{k+K}}$ is the Perron eigenvector for $\gamma^u_{G^{k+K}}$. It follows that for any a in $\mathbb{Z}B_1^k$, there is a D such that

$$\begin{aligned}
D(\lambda_{G_1'})^j &\geq \eta((\gamma^s_{B_1^k})^j(a)) \\
&= \langle(\delta_0^{s,K} - \delta_1^{s,K}) \circ Q((\gamma^s_{B_1^k})^j(a)),\, v^u_{G^{k+K}}\rangle \\
&= \langle(\gamma^s_{G^{k+K}})^j(\delta_0^{s,K} - \delta_1^{s,K}) \circ Q(a),\, v^u_{G^{k+K}}\rangle \\
&= \langle(\delta_0^{s,K} - \delta_1^{s,K}) \circ Q(a),\, (\gamma^u_{G^{k+K}})^j(v^u_{G^{k+K}})\rangle \\
&= \lambda_G^j\langle(\delta_0^{s,K} - \delta_1^{s,K}) \circ Q(a),\, v^u_{G^{k+K}}\rangle.
\end{aligned}$$

As $0 < \lambda_{G_1'} < \lambda_G$, we conclude that $\langle(\delta_0^{s,K} - \delta_1^{s,K}) \circ Q(a),\, v^u_{G^{k+K}}\rangle$ is not positive. This implies that $(\delta_0^{s,K} - \delta_1^{s,K}) \circ Q(a)$ is not in $D^s(G^k)^+ - \{0\}$. This holds for every a in $\mathbb{Z}B_1^k$, but as Q is an isomorphism, we also see that $\mathrm{Im}(\delta_0^{s,K} - \delta_1^{s,K}) \cap D^s(G^k)^+ = \{0\}$. The conclusion follows. □

Lemma 5.2. *Let (Σ, σ) be a mixing shift of finite type, (Z, ζ) be a mixing Smale space and $\rho_u : (\Sigma, \sigma) \to (Z, \zeta)$ be a u-bijective, almost one-to-one factor map. Then the subgroup generated by* $\mathrm{Ker}(\delta_0^{s*} - \delta_1^{s*}) \cap D^s(\Sigma_0(\rho_u))^+$ *is* $\mathrm{Ker}(\delta_0^{s*} - \delta_1^{s*})$.

Proof. First, suppose that we have a strictly positive element a in $(\mathbb{Z}G^{k+K})^+$ such that $(\delta_0^{s*,K} - \delta_1^{s*,K})(a) = 0$. Then $[a, j] \in \mathrm{Ker}(\delta_0^{s*} - \delta_1^{s*})$ for every j in $\mathbb{N}$. It follows that every $[b, j]$ in $\mathrm{Ker}(\delta_0^{s*} - \delta_1^{s*})$ can be written as the difference

$$[b, j] = [b + na, j] - [na, j],$$

in which $n \in \mathbb{N}$. It is a simple consequence of Theorem 3.4 that we may choose n large enough that $b + na \in (\mathbb{Z}G_{0,0}^{k+K})^+$. This means $\mathrm{Ker}(\delta_0^{s*} - \delta_1^{s*}) \cap D^s(\Sigma_0(\rho_u))^+$ generates $\mathrm{Ker}(\delta_0^{s*} - \delta_1^{s*})$.

In order to obtain the element a as above, let us return to the proof of Lemma 5.1, using ρ_u and replacing s with u throughout. We now consider the diagram

$$\begin{tikzcd}
\mathbb{Z}B_1^k \arrow[r, "\gamma^u_{B_1^k}"] \arrow[d, "J"'] & \mathbb{Z}B_1^k \arrow[d, "J"] \\
\mathcal{A}(G_1^k, S_2) \arrow[r, "\gamma^u_{G_1^k}"] \arrow[d, "\delta_0^{u,K} - \delta_1^{u,K}"'] & \mathcal{A}(G_1^k, S_2) \arrow[d, "\delta_0^{u,K} - \delta_1^{u,K}"] \\
\mathbb{Z}G^{k+K} \arrow[r, "\gamma^u_{G^{k+K}}"] & \mathbb{Z}G^{k+K}
\end{tikzcd}$$

The third part of Theorem 4.2.13 of [Putnam 2014] tells us that the top square commutes and that the vertical maps are isomorphisms. The bottom square commutes by Theorem 4.2.3, Definition 4.2.4 and Theorem 4.2.5 of [Putnam 2014].

The same argument as given earlier shows that $\langle (\delta_0^{u,K} - \delta_1^{u,K})(a), v^s_{G^{k+K}} \rangle$ is not positive, for every a in $\mathcal{A}(G_1^k, S_2)$. But this also applies to $-a$ and it follows that

$$0 = \langle (\delta_0^{u,K} - \delta_1^{u,K})(a), v^s_{G^{k+K}} \rangle$$

for every a. Then by Lemma 3.5.6 of [Putnam 2014] (where there is a typo, switching $s*$ and $u*$), we get

$$0 = \langle a, (\delta_0^{s*,K} - \delta_1^{s*,K})(v^s_{G^{k+K}}) \rangle$$

for every a. It follows that $(\delta_0^{s*,K} - \delta_1^{s*,K})(v^s_{G^{k+K}}) = 0$. If $v^s_{G^{k+K}}$ had integer entries, we would be done.

If we view $(\delta_0^{s*,K} - \delta_1^{s*,K})$ as a linear map, the condition above means that it has a nontrivial kernel. That kernel has a basis and since the transformation has matrix with integer entries, we can obtain a basis for the kernel consisting of rational vectors. We know that $v^s_{G^{k+K}}$ is a positive vector and it also must be a linear combination of the rational basis for the kernel. If we carefully choose rational scalars, we may find a rational vector, also in the kernel, and sufficiently close to $v^s_{G^{k+K}}$ that all its entries are positive. If we then multiply by a suitable integer, we find a positive integer vector $a \in \mathbb{Z}G^{k+K}$ in the kernel of $(\delta_0^{s*,K} - \delta_1^{s*,K})$. □

Proof of Theorems 2.11 and 2.10. Consider the fibred product $\Sigma_{0,0}(\pi)$ of maps π_s and π_u, and let G be a presentation of π. Since (X, φ) is mixing, so is $\Sigma_{0,0}(\pi)$, by Theorem 2.6. From Theorem 5.1.4 of [Putnam 2014], the bottom row in our double complex is the same as the complex for the map ρ_s while the first column is the same as the complex for the map ρ_u. Now the two theorems follow from Lemmas 5.1 and 5.2, respectively. □

Suppose π and $\tilde{\pi}$ are the s/u-bijective pairs given in Theorem 2.13. It was shown in [Putnam 2014] that the homology of Smale spaces is independent of the corresponding s/u-bijective pair. This was done in Section 4.5 of that paper, where an isomorphism was found between the homology of the rows of the complexes

$$\Big(\bigoplus_{L-M=N} C^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}, \bigoplus_{L-M=N} d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\Big)$$

and

$$\Big(\bigoplus_{L-M=N} C^s_{\mathcal{Q},\mathcal{A}}(\tilde{\pi})_{L,M}, \bigoplus_{L-M=N} d^s_{\mathcal{Q},\mathcal{A}}(\tilde{\pi})_{L,M}\Big),$$

and then using Theorem 3.9 of [McCleary 2001], it was extended to an isomorphism between the homologies of the complexes

$$\Big(\bigoplus_{L-M=N} C^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}, \bigoplus_{L-M=N} d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{L,M}\Big)$$

and

$$\left(\bigoplus_{L-M=N} C^s_{\mathcal{Q},\mathcal{A}}(\tilde{\pi})_{L,M},\ \bigoplus_{L-M=N} d^s_{\mathcal{Q},\mathcal{A}}(\tilde{\pi})_{L,M}\right).$$

We use these isomorphisms to show that $H^s_0(\tilde{\pi})$ is an ordered group with the positive cone defined in Definition 2.8, and that these are indeed ordered isomorphisms.

Let us first remind the reader that there is a minor mistake in the statement of Theorem 3.5.11 in [Putnam 2014], used to prove the independence and functorial properties of the homology for Smale spaces (see Sections 5.4 and 5.5 in that paper). Deeley and coauthors proved that the surjectivity condition in this theorem must be replaced by the conjugacy condition [Deeley et al. 2016]. It follows that we also need the conjugacy condition in Theorem 5.4.1 in [Putnam 2014]. Here we state the correct versions of these results from [Deeley et al. 2016].

Theorem 5.3. *Suppose that*

$$\begin{array}{ccc} (\Sigma,\sigma) & \xrightarrow{\eta_1} & (\Sigma_1,\sigma) \\ {\scriptstyle \eta_2}\big\downarrow & & \big\downarrow{\scriptstyle \pi_1} \\ (\Sigma_2,\sigma) & \xrightarrow{\pi_2} & (\Sigma_0,\sigma) \end{array}$$

is a commutative diagram of nonwandering shifts of finite type, in which η_1 and π_2 are s-bijective factor maps, and η_2 and π_1 are u-bijective factor maps. If

$$\eta_2 \times \eta_1 : (\Sigma,\sigma) \to (\Sigma_2,\sigma)\ {}_{\pi_2}\!\times_{\pi_1} (\Sigma_1,\sigma)$$

is a conjugacy, then

$$\eta_1^s \circ \eta_2^{s*} = \pi_1^{s*} \circ \pi_2^s : D^s(\Sigma_2,\sigma) \to D^s(\Sigma_1,\sigma). \tag{5-1}$$

Theorem 5.4. *Let $\pi = (Y,\psi,\pi_s,Z,\zeta,\pi_u)$ and $\pi' = (Y',\psi',\pi'_s,Z',\zeta',\pi'_u)$ be s/u-bijective pairs for the Smale spaces (X,φ) and (X',φ'), respectively, and $\eta = (\eta_Y,\eta_X,\eta_Z)$ be a triple of factor maps such that the following diagram commutes:*

$$\begin{array}{ccccc} (Y,\psi) & \xrightarrow{\pi_s} & (X,\varphi) & \xleftarrow{\pi_u} & (Z,\zeta) \\ {\scriptstyle \eta_Y}\big\downarrow & & {\scriptstyle \eta_X}\big\downarrow & & \big\downarrow{\scriptstyle \eta_Z} \\ (Y',\psi') & \xrightarrow{\pi'_s} & (X',\varphi') & \xleftarrow{\pi'_u} & (Z',\zeta') \end{array}$$

(i) *If η is a triple of s-bijective maps and*

$$\pi_u \times \eta_Z : (Z,\zeta) \to (X,\varphi)\ {}_{\eta_X}\!\times_{\pi'_u} (Z',\zeta')$$

is a conjugacy, then for $L \geq 0$, $M \geq 1$,

$$\begin{array}{ccc} (\Sigma_{L,M}(\pi)) & \xrightarrow{\eta_{L,M}} & (\Sigma_{L,M}(\pi')) \\ {\scriptstyle \delta_{,m}}\big\downarrow & & \big\downarrow{\scriptstyle \delta'_{,m}} \\ (\Sigma_{L,M-1}(\pi)) & \xrightarrow{\eta_{L,M-1}} & (\Sigma_{L,M-1}(\pi')) \end{array}$$

and for $L \geq 1$, $M \geq 0$,

$$\begin{array}{ccc} (\Sigma_{L,M}(\pi)) & \xrightarrow{\eta_{L,M}} & (\Sigma_{L,M}(\pi')) \\ {\scriptstyle \delta_{l,}}\big\downarrow & & \big\downarrow{\scriptstyle \delta'_{l,}} \\ (\Sigma_{L-1,M}(\pi)) & \xrightarrow{\eta_{L-1,M}} & (\Sigma_{L-1,M}(\pi')) \end{array}$$

are commutative diagrams and

$$\eta_{L,M} \times \delta_{,m} : (\Sigma_{L,M}(\pi)) \to (\Sigma_{L,M}(\pi'))\ {}_{\delta'_{,m}} \times_{\eta_{L,M-1}} (\Sigma_{L,M-1}(\pi))$$

is a conjugacy. Moreover, η *induces chain maps between the complexes* $C^s_{\mathcal{Q},\mathcal{A}}(\pi)$ *and* $C^s_{\mathcal{Q},\mathcal{A}}(\pi')$, *and hence group homomorphisms* $\eta^{s*} : H^s_N(\pi) \to H^s_N(\pi')$ *for every integer* N.

(ii) *If* η *is a triple of u-bijective maps and*

$$\pi_s \times \eta_Y : (Y, \psi) \to (X, \varphi)\ {}_{\eta_X} \times_{\pi'_s} (Y', \psi')$$

is a conjugacy, then for $L \geq 0$, $M \geq 1$,

$$\begin{array}{ccc} (\Sigma_{L,M}(\pi)) & \xrightarrow{\eta_{L,M}} & (\Sigma_{L,M}(\pi')) \\ {\scriptstyle \delta_{,m}}\big\downarrow & & \big\downarrow{\scriptstyle \delta'_{,m}} \\ (\Sigma_{L,M-1}(\pi)) & \xrightarrow{\eta_{L,M-1}} & (\Sigma_{L,M-1}(\pi')) \end{array}$$

and for $L \geq 1$, $M \geq 0$,

$$\begin{array}{ccc} (\Sigma_{L,M}(\pi)) & \xrightarrow{\eta_{L,M}} & (\Sigma_{L,M}(\pi')) \\ {\scriptstyle \delta_{l,}}\big\downarrow & & \big\downarrow{\scriptstyle \delta'_{l,}} \\ (\Sigma_{L-1,M}(\pi)) & \xrightarrow{\eta_{L-1,M}} & (\Sigma_{L-1,M}(\pi')) \end{array}$$

are commutative diagrams and

$$\eta_{L,M} \times \delta_{l,} : (\Sigma_{L,M}(\pi)) \to (\Sigma_{L,M}(\pi'))\ {}_{\delta'_{l,}} \times_{\eta_{L-1,M}} (\Sigma_{L-1,M}(\pi))$$

is a conjugacy. Moreover, η *induces chain maps between the complexes* $C^s_{\mathcal{Q},\mathcal{A}}(\pi')$ *and* $C^s_{\mathcal{Q},\mathcal{A}}(\pi)$, *and hence group homomorphisms* $\eta^{s*} : H^s_N(\pi') \to H^s_N(\pi)$ *for every integer* N.

We remark that the results obtained in [Putnam 2014] (the independence and functorial properties) are all correct, because the diagrams constructed there satisfy the conjugacy condition.

By Theorem 3.2, both maps $\eta^s_1 \circ \eta^{s*}_2$ and $\pi^{s*}_1 \circ \pi^s_2$ in (5-1) are positive homomorphisms.

Theorem 5.5. *For graphs* G, H, *suppose* $\theta : H \to G$ *is a left-covering graph homomorphism,* (X, φ) *is a Smale space and* $\rho : (\Sigma_G, \sigma) \to (X, \varphi)$ *is a regular* s*-bijective factor map. The map* θ *induces an isomorphism between the homologies of the chain complexes* $(D^s(\Sigma_*(\rho \circ \theta)), d^s(\rho \circ \theta))$ *and* $(D^s(\Sigma_*(\rho), d^s(\rho))$.

In fact, the map θ induces homomorphisms θ^s_* at all levels of the complexes with the commutative diagram

$$\begin{tikzcd} D^s(H^k_N) \arrow[r, "d^s(\rho\circ\theta)_N"] \arrow[d, "\theta^s_N"'] & D^s(H^{k+K}_{N-1}) \arrow[d, "\theta^s_{N-1}"] \\ D^s(G^k_N) \arrow[r, "d^s(\rho)_N"] & D^s(G^{k+K}_{N-1}) \end{tikzcd}$$

for each $N \geq 1$ and $k \geq 0$, where $K \geq 1$ satisfies the conclusion of Lemma 2.7.2 in [Putnam 2014] for the map ρ.

To show that the induced homomorphism on the homology of the above complexes by θ^s_* is an isomorphism, one could choose a lifting map $\lambda : G^0 \to H^0$ with $\theta \circ \lambda = \mathrm{Id}_{G_0}$. Then Lemma 4.5.4 in [Putnam 2014] shows that, for each $N \geq 0$,

$$d^{s,K}(\rho \circ \theta)_N \circ \lambda = \lambda \circ d^{s,K}(\rho)_N. \tag{5-2}$$

We claim that

$$\theta^s\big(\mathrm{Ker}(d^s(\theta \circ \rho)_N) \cap (D^s(H^k_N))^+\big) = \mathrm{Ker}(d^s(\rho)_N) \cap (D^s(G^k_N))^+,$$
$$\theta^s\big(\mathrm{Im}(d^s(\theta \circ \rho)_{N+1}) \cap (D^s(H^{k+K}_N))^+\big) = \mathrm{Im}(d^s(\rho)_{N+1}) \cap (D^s(G^{k+K}_N))^+.$$

By Theorem 3.2,

$$\theta^s\big(\mathrm{Ker}(d^s(\theta \circ \rho)_N) \cap (D^s(H^k_N))^+\big) \subseteq \mathrm{Ker}(d^s(\rho)_N) \cap (D^s(G^k_N))^+$$

and

$$\theta^s\big(\mathrm{Im}(d^s(\theta \circ \rho)_{N+1}) \cap (D^s(H^{k+K}_N))^+\big) \subseteq \mathrm{Im}(d^s(\rho)_{N+1}) \cap (D^s(G^{k+K}_N))^+.$$

Suppose that $b \in \mathbb{Z}G^k_{N+1}$ and $j \geq 0$, with $d^s(\rho)_{N+1}([b, j])$ in $D^s(G^{k+K}_N)^+$. By Theorem 4.2.3 in [Putnam 2014],

$$d^s(\rho)_{N+1}([b, j]) = [d^{s,K}(\rho)_{N+1}(b), j] \in D^s(G^{k+K}_N)^+,$$

which implies that, for some $j' \geq 0$,

$$(\gamma^s_{G_N})^{j'}(d^{s,K}(\rho)_{N+1}(b)) \in (\mathbb{Z}G^{k+K}_N)^+.$$

By Theorem 4.2.3 in [Putnam 2014],

$$(\gamma^s_{G_N})^{j'}(d^{s,K}(\rho)_{N+1}(b)) = d^{s,K}(\rho)_{N+1}((\gamma^s_{G_N})^{j'}(b)) \in (\mathbb{Z}G^{k+K}_N)^+.$$

Let $b_1 = (\gamma^s_{G_N})^{j'}(b)$ and $j_1 = j' + j$. Then

$$[d^{s,K}(\rho)_{N+1}(b), j] = [d^{s,K}(\rho)_{N+1}(b_1), j_1],$$

and since $\lambda((\mathbb{Z}G^k_{N+1})^+) \subseteq (\mathbb{Z}H^k_{N+1})^+$, it follows from (5-2) that

$$d^{s,K}(\rho \circ \theta)_{N+1} \circ \lambda(b_1) = \lambda \circ d^{s,K}(\rho)_{N+1}(b_1) \in (\mathbb{Z}H^{k+K}_N)^+.$$

Let $a_1 = \lambda(b_1)$. Applying $\theta^{s,0} = \theta$ to both sides of the above equality,

$$\theta^{s,0}(d^{s,K}(\rho \circ \theta)_{N+1}(a_1)) = \theta^{s,0}(\lambda \circ d^{s,K}(\rho)_{N+1}(b_1)) = d^{s,K}(\rho)_{N+1}(b_1),$$

hence

$$[\theta^{s,0}(d^{s,K}(\rho \circ \theta)_{N+1}(a_1)), j_1] = [d^{s,K}(\rho)_{N+1}(b_1), j_1] = [d^{s,K}(\rho)_{N+1}(b), j],$$

and so

$$\theta^s(d^s(\rho)_{N+1}[a_1, j_1]) = d^s(\rho)_{N+1}([b, j]).$$

Since b is an arbitrary element in $\mathbb{Z}G^k_{N+1}$ with

$$d^s(\rho)_{N+1}([b, j]) \in D^s(G^{k+K}_N)^+,$$

the last equality implies

$$\mathrm{Im}(d^s(\rho)_{N+1}) \cap (D^s(G^{k+K}))^+ \subseteq \theta^s\big(\mathrm{Im}(d^s(\theta \circ \rho)_{N+1}) \cap (D^s(H^{k+K}))^+\big).$$

A similar argument shows that

$$\mathrm{Ker}(d^s(\rho)_N) \cap (D^s(G^k))^+ \subseteq \theta^s\big(\mathrm{Ker}(d^s(\theta) \circ \rho)_N \cap (D^s(H^k))^+\big).$$

Combining Theorems 3.1 and 5.5 with Theorem 4.5.3 in [Putnam 2014], we get the following result.

Theorem 5.6. *Suppose (X, φ) is a Smale space and (Σ, σ), (Σ', σ) are shifts of finite type with s-bijective maps $\rho : (\Sigma, \sigma) \to (X, \varphi)$ and $\rho' : (\Sigma', \sigma) \to (X, \varphi)$. Let (Y'', ψ'') be the fibred product of maps $\rho : (\Sigma, \sigma) \to (X, \varphi)$ and $\rho' : (\Sigma', \sigma) \to (X, \varphi)$, and η, η' be the natural s-bijective maps from (Y'', ψ'') to (Y, ψ) and (Y', ψ'), respectively. Then*

(i) *a chain map η^s from $\big(D^s(\Sigma_N(\rho \circ \eta)), d^s(\rho \circ \eta)_N\big)$ to $\big(D^s(\Sigma_N(\rho)), d^s(\rho)_N\big)$ exists such that*

$$\eta^s\big(\mathrm{Ker}(d^s(\rho \circ \eta)_N) \cap D^s(\Sigma_N(\rho \circ \eta))^+\big) = \mathrm{Ker}\big(d^s(\rho)_N \cap D^s(\Sigma_N(\rho))^+\big),$$
$$\eta^s\big(\mathrm{Im}(d^s(\rho \circ \eta)_N) \cap D^s(\Sigma_N(\rho \circ \eta))^+\big) = \mathrm{Im}\big(d^s(\rho)_N \cap D^s(\Sigma_N(\rho))^+\big);$$

(ii) *a chain map $\boldsymbol{C}'$ from $\big(D^s(\Sigma_N(\rho' \circ \eta')), d^s(\rho' \circ \eta')_N\big)$ to $\big(D^s(\Sigma_N(\rho')), d^s(\rho')_N\big)$ exists such that*

$$\eta'^s\big(\mathrm{Ker}(d^s(\rho' \circ \eta')_N) \cap D^s(\Sigma_N(\rho' \circ \eta'))^+\big) = \mathrm{Ker}\big(d^s(\rho')_N \cap D^s(\Sigma_N(\rho'))^+\big),$$
$$\eta'^s\big(\mathrm{Im}(d^s(\rho' \circ \eta')_N) \cap D^s(\Sigma_N(\rho' \circ \eta'))^+\big) = \mathrm{Im}\big(d^s(\rho')_N \cap D^s(\Sigma_N(\rho'))^+\big);$$

(iii) *η^s and η'^s induce isomorphisms at the level of the associated homologies of the chain complexes.*

As in Section 5.5 of [Putnam 2014], we prove Theorem 2.13 in the case $Z = \tilde{Z}$, $\zeta = \tilde{\zeta}$ and $\pi_u = \tilde{\pi}_u$. The case $Y = \tilde{Y}$, $\psi = \tilde{\psi}$ and $\pi_s = \tilde{\pi}_s$ is proved in a similar way, and the general result follows from these two special cases.

Let (Y', ψ') denote the fibred product of the maps $\pi_s : (Y, \psi) \to (X, \varphi)$ and $\tilde{\pi}_s : (\tilde{Y}, \tilde{\psi}) \to (X, \varphi)$, and $\eta', \tilde{\eta}'$ denote the natural s-bijective maps from (Y', ψ') to (Y, ψ) and $(\tilde{Y}, \tilde{\psi})$, respectively. Then $\pi' = (Y', \psi', \pi_s \circ \eta', Z, \zeta, \pi_u)$ is an s/u-bijective pair for the Smale space (X, φ), and the following diagram is commutative:

$$\begin{array}{ccccc} (Y', \psi') & \xrightarrow{\pi_s \circ \eta} & (X, \varphi) & \xleftarrow{\pi_u} & (Z, \zeta) \\ {\scriptstyle \eta'}\downarrow & & \downarrow{\scriptstyle \mathrm{Id}_X} & & \downarrow{\scriptstyle \mathrm{Id}_Z} \\ (Y, \psi) & \xrightarrow{\pi_s} & (X, \varphi) & \xleftarrow{\pi_u} & (Z, \zeta) \end{array} \tag{5-3}$$

This diagram satisfies the conditions of the first part of Theorem 5.4, and the triple of s-bijective $\eta = (\eta', \mathrm{Id}_X, \mathrm{Id}_Z)$ induces a chain map on the double complexes used to define $H^s_N(\pi')$ and $H^s_N(\pi)$, $N \in \mathbb{Z}$. Since $D^s_{\mathcal{Q},\mathcal{A}}(\Sigma(\pi), \sigma) = D^s(\Sigma(\pi), \sigma)$, $D^s_{\mathcal{Q},\mathcal{A}}(\Sigma(\pi'), \sigma) = D^s(\Sigma(\pi'), \sigma)$, and $\eta^s_{L,M}$, $\eta^{u*}_{L,M}$ are positive homomorphisms, by Theorem 3.2, we have

$$\begin{aligned} \eta^s_{0,0}\big(\mathrm{Ker}(d^s(\pi)_{0,0}) \cap D^s(\Sigma(\pi), \sigma)^+\big) &\subseteq \mathrm{Ker}(d^s(\pi')_{0,0}) \cap D^s(\Sigma(\pi'), \sigma)^+ \\ \eta^s_{0,0}\big(\mathrm{Im}(d^s(\pi)_{1,0}) \cap D^s(\Sigma(\pi))^+\big) &\subseteq \mathrm{Im}(d^s(\pi')_{1,0}) \cap D^s(\Sigma(\pi'), \sigma)^+ \end{aligned} \tag{5-4}$$

Let $H_N(\eta)$ be the induced homomorphism by the chain map $\eta^s_{*,*}$ at the level of homologies from $H_N(\pi')$ to $H^s_N(\pi)$. This is known to be an isomorphism. We claim that this is an ordered isomorphism after proving that $H_N(\pi')$ is an ordered group. To prove that $H_N(\pi')$ is an ordered group, it suffices to show that the inclusions in (5-4) are indeed equalities.

To prove that $H_N(\eta)$ is an isomorphism, for $N \in \mathbb{Z}$, one needs to consider the filtrations $F^pC^s_{\mathcal{Q},}(\pi')$ and $F^pC^s_{\mathcal{Q},}(\pi)$ for the differential graded abelian groups $(H^s(\pi'), d^s_{\mathcal{Q},\mathcal{A}}(\pi'))$ and $(H^s(\pi), d^s_{\mathcal{Q},\mathcal{A}}(\pi))$, respectively, as in Section 5.5 of [Putnam 2014]. These filtrations satisfy the conditions of Theorem 3.9 in [McCleary 2001]. According to this theorem, every isomorphism Φ_1 between $E_1^{*,*}$ terms of the associated spectral sequences (of these filtration differential graded modules) induces an isomorphism Φ_∞ between E_∞ terms of the associated spectral sequences (roughly, $\Phi_\infty(a) = \Phi_1(a)$, when we regard a as an element of the associated $E_1^{*,*}$ term). The isomorphism $H_N(\eta)$ is then constructed using the isomorphisms between the E_∞ terms, for $N \in \mathbb{N}$. The $E_1^{*,*}$ terms for each of these filtrations are the homologies of the rows of the corresponding complexes, that is,

$$E_1^{p,q}(\pi) = \operatorname{Ker}(\tilde{d}^s(\rho_{,M})_L) / \operatorname{Im}(\hat{d}^s(\rho_{,M})_{L+1}),$$

and the same for π', where

$$\tilde{d}^s(\rho_{,M})_L = d^s(\rho_{,M})_L|_{\bigoplus_{L\geq 2p+q, M=p} C^s_{\mathcal{Q},}(\pi)}$$

and

$$\hat{d}^s(\rho_{,M})_{L+1} = d^s(\rho_{,M})_{L+1}|_{\bigoplus_{L\geq 2p+q+1, M=p} C^s_{\mathcal{Q},}(\pi)}.$$

Since θ and ρ_u in $\Sigma(\pi') \xrightarrow{\theta} \Sigma(\pi) \xrightarrow{\rho_u} (Z, \zeta)$ are s-bijective maps, where $\theta((y, \tilde{y}), z) = (y, z)$, by Theorem 5.6, we have a chain map θ^s from $(C^s(\pi')_{*,M}, d^s((\rho_u \circ \theta)_{,M})_*)$ to $(C^s(\pi)_{*,M}, d^s((\rho_u)_{,M})_*)$ that induces an isomorphism H_θ at the level of homologies of the complexes for fixed $M \geq 0$, so that

$$\theta^s\big(\operatorname{Ker}(d^s(\rho_{,M})_L) \cap (C^s(\pi')_{L,M})^+\big) = \operatorname{Ker}(d^s(\rho_{,M})_L) \cap (C^s(\pi)_{L,M})^+$$

and

$$\theta^s\big(\operatorname{Im}(d^s(\rho_{,M})_{L+1}) \cap (C^s(\pi')_{L,M})^+\big) = \operatorname{Im}(d^s(\rho_{,M})_{L+1}) \cap (C^s(\pi)_{L,M})^+$$

for each $L \geq 0$ and fixed $M \geq 0$.

If one lifts H_θ at the level of homologies of the complexes

$$(C^s_{\mathcal{Q},}(\pi')_{*,M}, d^s_{\mathcal{Q},}((\rho_u \circ \theta)_{,M})_*), \quad (C^s_{\mathcal{Q},}(\pi)_{*,M}, d^s_{\mathcal{Q},}((\rho_u)_{,M})_*),$$

by the first part of Theorem 4.3.1 in [Putnam 2014], for fixed $M \geq 0$, since

$$C^s_{\mathcal{Q},}(\pi')_{0,0} = C^s(\pi')_{0,0} = D^s(\Sigma_{0,0}(\pi'), \sigma)$$

and

$$C^s_{\mathcal{Q},}(\pi_0)_{0,0} = C^s(\pi)_{0,0} = D^s(\Sigma_{0,0}(\pi), \sigma),$$

for

$$K^s(\pi) := \mathrm{Ker}(d^s(\rho_{,0})_0) \cap (D^s(\Sigma_{0,0}(\pi'), \sigma))^+$$

and

$$I^s(\pi) := \mathrm{Im}(d^s(\rho_{,0})_0) \cap (D^s(\Sigma_{0,0}(\pi'), \sigma))^+$$

we have

$$C_\theta(K^s(\pi')) = K^s(\pi), \quad C_\theta(I^s(\pi')) = I^s(\pi').$$

In fact, H_θ is an isomorphism between the terms $E_1^{*,*}(\pi')$ and $E_1^{*,*}(\pi)$. Therefore, Theorem 3.9 in [McCleary 2001] implies that there is an isomorphism $\mathcal{H}_\theta$ at the level of homologies of the complexes $(C^s_{\mathcal{Q},}(\pi'), d^s_{\mathcal{Q},}(\pi'))$ and $(C^s_{\mathcal{Q},}(\pi), d^s_{\mathcal{Q},}(\pi))$, which is constructed by the induced isomorphism H_∞ on $E_\infty^{*,*}$ terms with H_θ (roughly, $H_\infty(a) = H_\theta(a)$ when we regard $a \in E_\infty^{*,*}$ as an element of $E_1^{*,*}$). Since the isomorphism $\mathcal{H}_\theta$ is directly defined by H_∞ (or H_θ), it is the same as the induced homomorphism by the chain map $\eta^s_{\mathcal{Q},}$, where $\eta = (\eta', \mathrm{Id}_X, \mathrm{Id}_Z)$ is the triple s-bijective map in the diagram (5-3) and $\eta^s_{\mathcal{Q},}$ exactly behaves like θ^s, when θ^s is considered as a map on the domain of $\eta^s_{\mathcal{Q},}$. On the other hand, since the maps u and $\bar{u}$ in the proof of Theorem 3.9 in [McCleary 2001] are natural and $D^s_{\mathcal{Q},}(\Sigma_{0,0}, \sigma) = D^s(\Sigma_{0,0}, \sigma)$, for

$$K^s_{\mathcal{Q}},(\pi) := \mathrm{Ker}(d^s_{\mathcal{Q},}(\pi')_{0,0}) \cap (D^s(\Sigma_{0,0}(\pi'), \sigma))^+$$

and

$$I^s_{\mathcal{Q}},(\pi) := \mathrm{Im}(d^s_{\mathcal{Q},}(\pi')_{1,0}) \cap (D^s(\Sigma_{0,0}(\pi'), \sigma))^+$$

we have

$$\eta^s_{\mathcal{Q},}(K^s_{\mathcal{Q}},(\pi')) = K^s_{\mathcal{Q}},(\pi), \quad \eta^s_{\mathcal{Q},}(I^s_{\mathcal{Q}},(\pi')) = I^s_{\mathcal{Q}},(\pi). \tag{5-5}$$

Let $\mathcal{J}(\pi')$ and $\mathcal{J}(\pi)$ be the isomorphisms induced by the chain maps $J_{\mathcal{Q}}(\pi')$ and $J_{\mathcal{Q}}(\pi)$, as in Theorem 5.3.2 in [Putnam 2014], respectively. Then $H_0(\eta) = \mathcal{J}(\pi) \circ \mathcal{H}_\theta \circ \mathcal{J}(\pi')^{-1}$ and it is an isomorphism from $H^s_N(\pi')$ to $H^s_N(\pi)$, and since $D^s_{\mathcal{Q},\mathcal{A}}(\Sigma_{0,0}, \sigma) = D^s(\Sigma_{0,0}, \sigma)$, for

$$K^s_{\mathcal{Q},\mathcal{A}}(\pi) := \mathrm{Ker}(d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{0,0}) \cap (D^s(\Sigma_{0,0}(\pi), \sigma))^+$$

and

$$I^s_{\mathcal{Q},\mathcal{A}}(\pi) := \mathrm{Im}(d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{1,0}) \cap (D^s(\Sigma_{0,0}(\pi), \sigma))^+$$

we have

$$J_{\mathcal{Q}}(\pi)(K^s_{\mathcal{Q},\mathcal{A}}(\pi')) = K^s_{\mathcal{Q},}(\pi), \quad J_{\mathcal{Q}}(\pi)(I^s_{\mathcal{Q},\mathcal{A}}(\pi')) = I^s_{\mathcal{Q},}(\pi). \tag{5-6}$$

The equalities (5-5) and (5-6) show that $\mathrm{Ker}(d^s_{\mathcal{Q},\mathcal{A}}(\pi')_{0,0})$ contains positive elements if and only if $\mathrm{Ker}(d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{0,0})$ does so (and the same holds for $\mathrm{Im}(d^s_{\mathcal{Q},\mathcal{A}}(\pi')_{1,0})$ and $\mathrm{Im}(d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{1,0})$). Since $\mathrm{Im}(d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{1,0})$ does not contain any positive element, and

$\mathrm{Ker}(d^s_{\mathcal{Q},\mathcal{A}}(\pi)_{0,0})$ contains at least one positive element, $\mathrm{Im}(d^s_{\mathcal{Q},\mathcal{A}}(\pi')_{1,0})$ could not contain any positive element and $\mathrm{Ker}(d^s_{\mathcal{Q},\mathcal{A}}(\pi')_{0,0})$ contains at least one positive element, and these imply that $H^s_0(\pi')$ is an ordered group with the positive cone defined as above. Also by (5-5) and (5-6), $H(\eta)$ is an order isomorphism. Replacing $(\tilde{Y}, \tilde{\psi})$ by (Y, ψ) in (5-3), we get that $H^s_0(\tilde{\pi})$ is an ordered group with the positive cone defined as in Definition 2.8 and $H(\tilde{\eta}) = H^s_N(\pi') \to H^s_N(\tilde{\pi})$ is an order isomorphism. Finally, $H_0(\tilde{\eta}) \circ H_0(\eta)^{-1}$ is an order isomorphism from $H^s_0(\pi)$ to $H^s_0(\tilde{\pi})$.

Proof of Theorem 2.14. We only prove the first part. The other is proved in a similar way. By Theorem 4.2 in [Deeley et al. 2016], we can find s/u-bijective pairs $\pi = (Y, \psi, \pi_s, Z, \zeta, \pi_u)$ and $\pi' = (Y', \psi', \pi'_s, Z', \zeta', \pi'_u)$ for Smale spaces (X, φ) and (X', φ'), respectively, and s-bijective maps η_Y and η_Z, such that the diagram

$$\begin{tikzcd}
(Y, \psi) \arrow[r, "\pi_s"] \arrow[d, "\eta_Y"'] & (X, \varphi) \arrow[d, "\rho"'] & (Z, \zeta) \arrow[l, "\pi_u"'] \arrow[d, "\eta_Z"] \\
(Y', \psi') \arrow[r, "\pi'_s"] & (X', \varphi') & (Z', \zeta') \arrow[l, "\pi'_u"']
\end{tikzcd}$$

commutes and $\pi_u \times \eta_Z : (Z, \zeta) \to (X, \varphi)\ {}_{\rho}{\times}_{\pi'_u}\ (Z', \zeta')$ is a conjugacy. Therefore, ρ induces a positive homomorphism $\rho^s_0 : H^s_0(X, \varphi) \to H^s_0(X', \varphi')$, by Theorems 5.4, 3.2 and 2.13, and the order structure is independent of the s/u-bijective pair. □

Acknowledgments

This work was started when Saeidi Gholikandi was at University of Victoria on sabbatical. She would like to express her thanks to Ian F. Putnam for his support and guidance during her visit. The authors would also like to thank the referee for many helpful suggestions and comments.

References

[Amini et al. 2013] M. Amini, I. F. Putnam, and S. Saeidi Gholikandi, “Homology for one-dimensional solenoids”, 2013. To appear in *Math. Scand.*

[Aoki and Hiraide 1994] N. Aoki and K. Hiraide, *Topological theory of dynamical systems*, North-Holland Mathematical Library **52**, North-Holland, Amsterdam, 1994. MR Zbl

[Blyth 2005] T. S. Blyth, *Lattices and ordered algebraic structures*, Springer, Berlin, 2005. MR Zbl

[Bowen 1970] R. Bowen, “Markov partitions for Axiom A diffeomorphisms”, *Amer. J. Math.* **92** (1970), 725–747. MR Zbl

[Bowen 1978] R. Bowen, *On Axiom A diffeomorphisms*, Regional Conference Series in Mathematics **35**, American Mathematical Society, Providence, RI, 1978. MR Zbl

[Deeley et al. 2016] R. J. Deeley, D. B. Killough, and M. F. Whittaker, “Functorial properties of Putnam’s homology theory for Smale spaces”, *Ergodic Theory Dynam. Systems* **36**:5 (2016), 1411–1440. MR Zbl

[Effros 1981] E. G. Effros, *Dimensions and C^*-algebras*, CBMS Regional Conference Series in Mathematics **46**, American Mathematical Society, Providence, RI, 1981. MR Zbl

[Fisher 2013] T. Fisher, "Resolving extensions of finitely presented systems", *Acta Appl. Math.* **126** (2013), 131–163. MR Zbl

[Fried 1987] D. Fried, "Finitely presented dynamical systems", *Ergodic Theory Dynam. Systems* **7**:4 (1987), 489–507. MR Zbl

[Katok and Hasselblatt 1995] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and Its Applications **54**, Cambridge Univ. Press, 1995. MR Zbl

[Krieger 1980] W. Krieger, "On dimension functions and topological Markov chains", *Invent. Math.* **56**:3 (1980), 239–250. MR Zbl

[Lind and Marcus 1995] D. Lind and B. Marcus, *An introduction to symbolic dynamics and coding*, Cambridge Univ. Press, 1995. MR Zbl

[McCleary 2001] J. McCleary, *A user's guide to spectral sequences*, 2nd ed., Cambridge Studies in Advanced Mathematics **58**, Cambridge Univ. Press, 2001. MR Zbl

[Putnam 2005] I. F. Putnam, "Lifting factor maps to resolving maps", *Israel J. Math.* **146** (2005), 253–280. MR Zbl

[Putnam 2014] I. F. Putnam, *A homology theory for Smale spaces*, Memoirs of the American Mathematical Society **1094**, American Mathematical Society, Providence, RI, 2014. MR Zbl

[Ruelle 1978] D. Ruelle, *Thermodynamic formalism*, Encyclopedia of Mathematics and Its Applications **5**, Addison-Wesley, Reading, MA, 1978. MR Zbl

[Smale 1967] S. Smale, "Differentiable dynamical systems", *Bull. Amer. Math. Soc.* **73** (1967), 747–817. MR Zbl

[Walters 1982] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics **79**, Springer, Berlin, 1982. MR Zbl

[Wieler 2014] S. Wieler, "Smale spaces via inverse limits", *Ergodic Theory Dynam. Systems* **34**:6 (2014), 2066–2092. MR Zbl

[Williams 1967] R. F. Williams, "One-dimensional non-wandering sets", *Topology* **6** (1967), 473–487. MR Zbl

[Williams 1970] R. F. Williams, "Classification of one dimensional attractors", pp. 341–361 in *Global Analysis* (Berkeley, CA, 1968), edited by S.-S. Chern and S. Smale, Proc. Symp. Pure Math. **14**, American Mathematical Society, Providence, RI, 1970. MR Zbl

[Williams 1974] R. F. Williams, "Expanding attractors", *Inst. Hautes Études Sci. Publ. Math.* 43 (1974), 169–203. MR

[Yi 2001] I. Yi, "Canonical symbolic dynamics for one-dimensional generalized solenoids", *Trans. Amer. Math. Soc.* **353**:9 (2001), 3741–3767. MR Zbl

Received January 18, 2016. Revised December 7, 2016.

MASSOUD AMINI
DEPARTMENT OF MATHEMATICS
FACULTY OF MATHEMATICAL SCIENCES
TARBIAT MODARES UNIVERSITY
TEHRAN 14115-134
IRAN
mamini@modares.ac.ir

IAN F. PUTNAM
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF VICTORIA
VICTORIA BC V8W 3R4
CANADA
ifputnam@uvic.ca

SARAH SAEIDI GHOLIKANDI
DEPARTMENT OF MATHEMATICS
FACULTY OF MATHEMATICAL SCIENCES
TARBIAT MODARES UNIVERSITY
TEHRAN 14115-134
IRAN
s.saeadi@modares.ac.ir

PACIFIC JOURNAL OF MATHEMATICS
Vol. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.289

CHARACTERIZATIONS OF IMMERSED GRADIENT ALMOST RICCI SOLITONS

CÍCERO P. AQUINO, HENRIQUE F. DE LIMA AND JOSÉ N. V. GOMES

Our purpose is to study the geometry of gradient almost Ricci solitons isometrically immersed either in the hyperbolic space $\mathbb{H}^{n+1}$, in the de Sitter space $\mathbb{S}_1^{n+1}$, or in the anti-de Sitter space $\mathbb{H}_1^{n+1}$. In each one of these ambient spaces we obtain extensions of a classical theorem due to Nomizu and Smith. More precisely, we show that the totally umbilical hypersurfaces are the only immersed hypersurfaces of such ambient spaces which admit a structure of gradient almost Ricci soliton via the tangential component of a certain fixed vector, and whose image of the Gauss mapping is also totally umbilical. Furthermore, in the case that the structure of gradient almost Ricci soliton is nontrivial, we conclude that such a hypersurface must be isometric either to $\mathbb{H}^n$, when the ambient space is $\mathbb{H}^{n+1}$ or $\mathbb{H}_1^{n+1}$, or to $\mathbb{S}^n$, when the ambient space is $\mathbb{S}_1^{n+1}$.

1. Introduction

The concept of a Ricci soliton, introduced in the seminal paper [Hamilton 1982], corresponds to a natural generalization of Einstein metrics. We recall that a Riemannian manifold (M^n, g) is called a Ricci soliton if there exist a complete vector field X and a constant λ satisfying the equation

$$\mathrm{Ric} + \tfrac{1}{2}\mathscr{L}_X g = \lambda g, \tag{1-1}$$

where Ric and $\mathscr{L}$ stand for the Ricci tensor and the Lie derivative on M^n.

Ricci solitons also correspond to selfsimilar solutions of Hamilton's Ricci flow [ibid.] and often arise as limits of dilations of singularities in the Ricci flow. They can be viewed as fixed points of the Ricci flow, as a dynamical system, on the space of Riemannian metrics modulo diffeomorphisms and scalings. For more details on this subject, we recommend the survey [Cao 2010].

Pigola et al. [2011] extended the definition of Ricci solitons by adding the condition that the parameter λ in (1-1) be a smooth real function on M^n; this

MSC2010: primary 53C42; secondary 53B30, 53C50, 53Z05, 83C99.
Keywords: almost Ricci solitons, hyperbolic space, de Sitter space, anti-de Sitter space, mean curvature, Gauss mapping.

attracted much attention in the mathematical community. Such solitons arise from the Ricci–Bourguignon flow as shown recently in [Catino et al. 2016]. In this more general setting, we refer to (1-1) as being the fundamental equation of an almost Ricci soliton (M^n, g, X, λ). Following the terminology of Ricci solitons, an almost Ricci soliton is called *expanding* or *shrinking* if $\lambda < 0$ or $\lambda > 0$, respectively. When $\lambda = 0$ we have a *steady* Ricci soliton. Otherwise, it will be called *indefinite*.

When the vector field X is a gradient of a smooth function $f : M^n \to \mathbb{R}$, the manifold will be called a gradient almost Ricci soliton. In this case, (1-2) becomes

$$\operatorname{Ric} + \nabla^2 f = \lambda g, \tag{1-2}$$

where $\nabla^2 f$ stands for the Hessian of the potential function f. When either the vector field X is trivial, or the potential f is constant, the almost Ricci soliton will be called *trivial*, otherwise it will be a *nontrivial* almost Ricci soliton.

We notice that when $n \geq 3$ and X is a Killing vector field, an almost Ricci soliton is a genuine a Ricci soliton. Indeed, in this case, (M^n, g) is an Einstein manifold and we can apply Schur's lemma to deduce that λ is constant. Conditions under which a nontrivial almost Ricci soliton structure exists were first investigated in [Pigola et al. 2011]. Subsequently, Barros and Ribeiro [2012] obtained some structural equations and deduced corresponding rigidity theorems; jointly with Batista, they also showed in [Barros et al. 2014b] that any compact nontrivial almost Ricci soliton (M^n, g, X, λ) with constant scalar curvature is isometric to a Euclidean sphere $\mathbb{S}^n$. As a consequence, they concluded that every compact nontrivial almost Ricci soliton with constant scalar curvature must be gradient.

Almost Ricci solitons that are realized as Einstein warped products, with a one-dimensional base and Einstein fibers, were constructed in [Pigola et al. 2011]. By using Lemma 1.1 of that paper, we can prove that the warped product $M = \mathbb{R} \times_\psi \mathbb{H}^m$ with metric $g = dt^2 + \psi^2 g_0$, has a structure of almost Ricci soliton $(M, g, \nabla\tilde{f}, \tilde{\lambda})$, where g_0 is the standard metric of $\mathbb{H}^m$ and the functions involved are the respective lifts of $f(t) = \sinh t$ and $\lambda(t) = \sinh t - m$, whereas the warping function is $\psi(t) = \cosh t$. More generally, a necessary and sufficient condition for a warped product Einstein manifold to support a gradient almost Ricci soliton structure was shown in [Feitosa et al. 2015].

Recall also that there exist manifolds that do not admit an almost Ricci soliton structure. For instance, Pigola et al. [2011] proved that $\mathbb{H}^2 \times \mathbb{H}^2$ has this property. For a locally conformally flat gradient almost Ricci soliton, Catino [2012] proved that, around any regular point of the potential f, such a manifold $(M^n, g, \nabla f, \lambda)$ is locally a warped product with fibers of constant sectional curvature.

Jointly with Barros and Ribeiro, the third author studied in [Barros et al. 2011] isometric immersions of an almost Ricci soliton (M^n, g, X, λ) into a Riemannian manifold $\widetilde{M}^{n+p}$. In this context, they presented some obstruction results in order to obtain a minimal immersion under conditions on the sectional curvature of $\widetilde{M}^{n+p}$. In particular, when $\widetilde{M}^{n+p}$ has nonpositive sectional curvature, they proved that if (M^n, g, X, λ) is a traditional Ricci soliton and X has integrable norm on M^n, then M^n cannot be minimal. Moreover, they showed that if (M^n, g, X, λ) is a shrinking Ricci soliton and X has bounded norm on M^n, then M^n must be compact. Hence, when $\widetilde{M}^{n+p}$ is a space-form of nonpositive sectional curvature, such an immersion cannot be minimal. We refer to [Mastrolia et al. 2013] for further discussions.

On the other hand, it is well known that the study of the behavior of the Gauss mapping gives deep information on the geometry of an isometric immersion. For instance, Nomizu and Smyth [1969] showed that a compact connected orientable manifold M^n immersed in the sphere $\mathbb{S}^{n+1}$ with constant mean curvature is a hypersphere if the Gauss image of M^n lies in a closed hemisphere of $\mathbb{S}^{n+1}$. More recently, the first and second authors jointly with Barros [Barros et al. 2014a] showed that a constant mean curvature complete hypersurface of the hyperbolic space $\mathbb{H}^{n+1}$, whose image of the Gauss mapping lies in a totally umbilical spacelike hypersurface of the de Sitter space $\mathbb{S}_1^{n+1}$, must be totally umbilical.

In the Lorentzian setting, Xin [1991] and Aiyama [1992], working independently, characterized spacelike hyperplanes as the only complete constant mean curvature spacelike hypersurfaces in the Lorentz–Minkowski space $\mathbb{R}_1^{n+1}$ whose image of the Gauss mapping is contained in a geodesic ball of $\mathbb{H}^n$; see also [Palmer 1990] for a weaker first version of this result. When the ambient space is $\mathbb{S}_1^{n+1}$, Aledo and Alías [2002] showed that the spacelike geodesic round spheres are the only complete constant mean curvature hypersurfaces in $\mathbb{S}_1^{n+1}$ having the image of its Gauss mapping contained in a geodesic ball of $\mathbb{H}^{n+1}$. The first and second authors [Aquino and de Lima 2014] established another rigidity results showing that a complete spacelike hypersurface immersed with constant mean curvature either in the de Sitter space $\mathbb{S}_1^{n+1}$ or in the anti-de Sitter space $\mathbb{H}_1^{n+1}$ must be totally umbilical, provided that its Gauss mapping has some suitable behavior.

Here, motivated by the works previously described, we apply suitable formulas for the covariant and Lie derivatives of the scalar curvature (see Lemmas 1 and 2, respectively) in order to study the geometry of gradient almost Ricci solitons isometrically immersed either in the hyperbolic space $\mathbb{H}^{n+1}$ or in the de Sitter space $\mathbb{S}_1^{n+1}$ or in the anti-de Sitter space $\mathbb{H}_1^{n+1}$. In this setting, we show that the totally umbilical hypersurfaces of such ambient spaces are the only immersed hypersurfaces which admit a structure of gradient almost Ricci soliton via the tangential component of a certain fixed vector, and whose image of the corresponding Gauss mapping is also totally umbilical (see Theorems 4, 6, and 8). Furthermore, if in addition we impose that the structure of gradient almost Ricci soliton must be nontrivial, then we conclude that such a hypersurface is isometric either to $\mathbb{H}^n$, when the ambient space is $\mathbb{H}^{n+1}$ or $\mathbb{H}_1^{n+1}$, or to $\mathbb{S}^n$, when the ambient space is $\mathbb{S}_1^{n+1}$ (see Corollaries 5, 7, and 9).

To close this introductory section, we also observe that the existence of a Ricci soliton structure on hypersurfaces of the Euclidean space whose potential vector is given by the tangential component of the position vector was recently investigated by Chen and Deshmukhin [2014].

2. Preliminaries

Let $\mathbb{R}^{n+2}_\nu$ denote the $(n+2)$-dimensional semi-Euclidean space of index $\nu \geq 1$, that is, the real vector space $\mathbb{R}^{n+2}$ endowed with the semi-Riemannian metric

$$\langle\cdot,\cdot\rangle = -\sum_{i=1}^{\nu} dx_i^2 + \sum_{j=\nu+1}^{n+2} dx_j^2,$$

where $x = (x_1, \dots, x_{n+2})$ denote the usual coordinates in $\mathbb{R}^{n+2}$. When $\nu = 1$, $\mathbb{R}^{n+2}_1$ is the so-called Lorentz–Minkowski space.

For a vector field X in $\mathbb{R}^{n+2}_\nu$, let $\varepsilon_X = \langle X, X\rangle$. We say that X is a *unit* vector field if $\varepsilon_X = \pm 1$, and *timelike* if $\varepsilon_X = -1$.

The $(n+1)$-dimensional hyperbolic space is the following hyperquadric of $\mathbb{R}^{n+2}_1$

$$\mathbb{H}^{n+1} = \{x \in \mathbb{R}^{n+2}_1;\ \langle x, x\rangle = -1,\ x_{n+2} \geq 1\}.$$

Let us consider a connected and oriented isometrically immersed hypersurface $\psi : \Sigma^n \to \mathbb{H}^{n+1} \subset \mathbb{R}^{n+2}_1$ and let us denote by $A : \mathfrak{X}(\Sigma) \to \mathfrak{X}(\Sigma)$ the Weingarten operator associated to the vector field N as well as $H = \frac{1}{n}\operatorname{tr}(A)$ stands for mean curvature of Σ^n.

Associated to A we have its traceless operator $\Phi : \mathfrak{X}(\Sigma) \to \mathfrak{X}(\Sigma)$ given by

$$\Phi(X) = AX - HX,$$

for every $X \in \mathfrak{X}(\Sigma)$. It is easily checked that the Hilbert–Schmidt norm of Φ (that is, $|\Phi|^2 = \operatorname{tr}(\Phi^*\Phi)$, where Φ^* stands for the adjoint of Φ) satisfies

$$|\Phi|^2 = |A|^2 - nH^2.$$

Consequently, we have that $|\Phi|^2 = 0$ if, and only, if Σ^n is a totally umbilical hypersurface.

Recall that, if ∇^0, $\overline{\nabla}$, and ∇ stands for the Levi–Civita connections in $\mathbb{R}^{n+2}_1$, $\mathbb{H}^{n+1}$, and Σ^n, respectively, then the Gauss and Weingarten formulas for a hypersurface $\psi : \Sigma^n \to \mathbb{H}^{n+1} \subset \mathbb{R}^{n+2}_1$ are given by

$$\overline{\nabla}_X Y = \nabla_X Y + \langle AX, Y\rangle N \tag{2-1}$$

and

$$AX = -\overline{\nabla}_X N = -\nabla^0{}_X N, \tag{2-2}$$

for every tangent vector fields $X, Y \in \mathfrak{X}(\Sigma)$. Consequently, from Gauss equations we have that the Ricci curvature of Σ^n is given by

$$\mathrm{Ric}_\Sigma(X, Y) = (1-n)\langle X, Y\rangle + nH\langle AX, Y\rangle - \langle AX, AY\rangle. \tag{2-3}$$

In addition, for a fixed arbitrary vector $a \in \mathbb{R}^{n+2}_1$, let us consider the *height* and the *angle* functions, defined respectively by, $l_a = \langle \psi, a\rangle$ and $f_a = \langle N, a\rangle$. A direct computation allows us to conclude that the gradient of such functions are given by $\nabla l_a = a^\top$ and $\nabla f_a = -A(a^\top)$, where $a^\top$ is the orthogonal projection of a onto the tangent bundle $T\Sigma$, that this

$$a^\top = a - f_a N + l_a \psi. \tag{2-4}$$

Taking into account that $\nabla^0 a = 0$ and using the Gauss and Weingarten formulas concerning a vector field X tangent to Σ^n,

$$\nabla_X a^\top = f_a AX + l_a X. \tag{2-5}$$

We use (2-5) and the Codazzi equation to deduce

$$\nabla_X A(a^\top) = f_a A^2 X + l_a AX + (\nabla_{a^\top} A)(X). \tag{2-6}$$

Thus, it follows from (2-5) and (2-6) that

$$\Delta l_a = nHf_a + nl_a \tag{2-7}$$

and

$$\Delta f_a = -|A|^2 f_a - nHl_a - n\langle \nabla H, a^\top\rangle. \tag{2-8}$$

See also [Rosenberg 1993].

Now, we deal with hypersurfaces isometrically immersed into two classes of simply connected Lorentzian space-forms. The first one is the $(n+1)$-dimensional de Sitter space

$$\mathbb{S}^{n+1}_1 = \{x \in \mathbb{R}^{n+2}_1;\ \langle x, x\rangle = 1\},$$

a hyperquadric of $\mathbb{R}^{n+2}_1$ with sectional curvature equal to 1. The second one is the $(n+1)$-dimensional anti-de Sitter space

$$\mathbb{H}^{n+1}_1 = \{x \in \mathbb{R}^{n+2}_2;\ \langle x, x\rangle = -1\},$$

a hyperquadric of $\mathbb{R}^{n+2}_2$ with sectional curvature equal to -1. Topologically, $\mathbb{H}^{n+1}_1$ is $\mathbb{S}^1 \times \mathbb{R}^n$ and the semi-Euclidean metric on $\mathbb{R}^{n+2}_2$ induces a Lorentzian metric of constant sectional curvature -1 on $\mathbb{H}^{n+1}_1$. Moreover, the universal covering manifold $\widetilde{\mathbb{H}}^{n+1}_1$ of $\mathbb{H}^{n+1}_1$ is topologically $\mathbb{R}^{n+1}$ (that is, $\widetilde{\mathbb{H}}^{n+1}_1$ is simply connected) and is thus a Lorentzian analogue of the usual Riemannian hyperbolic space of negative curvature -1, which is called the *universal anti-de Sitter spacetime*; see, for instance, [Beem et al. 1996, Section 5.3] or [O'Neill 1983, Section 8.6].

In order to simplify our notation, we will denote by $\mathbb{M}_c^{n+1}$ either the de Sitter space or the anti-de Sitter space, according to whether $c = 1$ or $c = -1$, respectively. In this setting, let $\psi : \Sigma^n \to \mathbb{M}_c^{n+1} \subset \mathbb{R}_\nu^{n+2}$ be a connected spacelike hypersurface immersed into $\mathbb{M}_c^{n+1}$ (that is, the induced metric via ψ is a Riemannian metric on Σ^n). Let us consider $A : \mathfrak{X}(\Sigma) \to \mathfrak{X}(\Sigma)$ the Weingarten operator of Σ^n with respect to a choice of timelike orientation N for Σ^n. We will denote by ∇^0, $\overline{\nabla}$, and ∇ the Levi–Civita connections of $\mathbb{R}_\nu^{n+2}$, $\mathbb{M}_c^{n+1}$, and Σ^n, respectively. Then, the Gauss and Weingarten formulas corresponding to Σ^n are given, respectively, by

$$\nabla^0{}_X Y = \nabla_X Y - \langle AX, Y\rangle N - c\langle X, Y\rangle \psi \quad \text{and} \quad AX = -\overline{\nabla}_X N = -\nabla^0{}_X N,$$

for all tangent vector fields $X, Y \in \mathfrak{X}(\Sigma)$. Thus, from Gauss equation we have that the Ricci curvature of Σ^n is given by

$$\mathrm{Ric}_\Sigma(X, Y) = c(n-1)\langle X, Y\rangle + nH\langle AX, Y\rangle + \langle AX, AY\rangle, \tag{2-9}$$

where $H = -\frac{1}{n}\,\mathrm{tr}(A)$ is the mean curvature of Σ^n.

At this point, we observe that the choice of the sign in our definition of H is motivated by the fact that in that case the mean curvature vector is given by $\overrightarrow{H} = HN$. Hence, $H(p) > 0$ at a point $p \in \Sigma^n$ if and only if $\overrightarrow{H}(p)$ is in the same time-orientation as $N(p)$ (in the sense that $\langle \overrightarrow{H}, N\rangle_p < 0$).

As before, it is also convenient to consider the traceless operator associated to the second fundamental form of Σ^n, $\Phi : \mathfrak{X}(\Sigma) \to \mathfrak{X}(\Sigma)$, which, in the Lorentzian setting, is given by $\Phi(X) = AX + HX$, for all $X \in \mathfrak{X}(\Sigma)$. It is easy to verify that Σ^n is a totally umbilical hypersurface if and only if Φ vanishes identically on Σ^n.

As in the case of immersions in the hyperbolic space, associated to a fixed arbitrary vector $a \in \mathbb{R}_\nu^{n+2}$, let us consider the height function $l_a = \langle \psi, a\rangle$ and the angle function $f_a = \langle N, a\rangle$. A direct computation allows us to conclude that the gradients of such functions are given by $\nabla l_a = a^\top$ and $\nabla f_a = -A(a^\top)$, where $a^\top$ is the orthogonal projection of a onto the tangent bundle $T\Sigma$, that is,

$$a^\top = a + f_a N - c\, l_a \psi. \tag{2-10}$$

Taking into account that $\nabla^0 a = 0$ and using the Gauss and Weingarten formulas concerning a vector field X tangent to Σ^n, it is not difficult to verify that

$$\nabla_X \nabla l_a = -f_a AX - c l_a X. \tag{2-11}$$

Now, we use (2-11) jointly with the Codazzi equation to deduce

$$\nabla_X \nabla f_a = f_a A^2 X + c l_a AX - (\nabla_{a^\top} A)(X). \tag{2-12}$$

Then, it follows from (2-11) and (2-12) that

$$\Delta l_a = nHf_a - cnl_a \tag{2-13}$$

and

$$\Delta f_a = |A|^2 f_a - cnHl_a + n\langle \nabla H, a^\top \rangle. \tag{2-14}$$

To close this section, we will quote three key lemmas, which will be essential in the proofs of our results. The first one corresponds to item (2) of Proposition 1 in [Barros and Ribeiro 2012].

Lemma 1. If Σ^n is a gradient almost Ricci soliton with potential function f, then

$$\nabla R = 2\,\mathrm{Ric}_\Sigma(\nabla f) + 2(n-1)\nabla\lambda, \tag{2-15}$$

where R stands for the scalar curvature of Σ^n.

The second auxiliary lemma is a well known formula of the theory of conformal vector fields in Riemannian geometry; see, for instance, Yano [1970].

Lemma 2. If X is a conformal vector field on a Riemannian manifold Σ^n with metric g such that $\mathscr{L}_X g = 2\sigma g$, then

$$\mathscr{L}_X R = -2(n-1)\Delta\sigma - 2R\sigma, \tag{2-16}$$

where R stands for the scalar curvature of Σ^n.

The third key lemma gives a suitable characterization of totally umbilical hypersurfaces in a semi-Riemannian space-form due to Kim et al. [2002], which can be regarded as a converse of a theorem due to Sharma and Duggal [1985].

Lemma 3. Let Σ^n be a connected semi-Riemannian hypersurface immersion in a semi-Riemannian space-form $\mathbb{M}_c^{n+1}$. Suppose that $\mathbb{M}_c^{n+1}$ carries a conformal vector field V whose tangential component $V^\top$ on Σ^n becomes a conformal vector field. Then, one of the following holds:

(a) Σ^n is a totally umbilical hypersurface;

(b) the restriction of V to Σ^n reduces to a tangent vector field on Σ^n.

3. Characterizing gradient almost Ricci solitons in $\mathbb{H}^{n+1}$

We recall that the totally umbilical hypersurfaces of L_σ of $\mathbb{H}^{n+1}$ can be realized in the Lorentzian model as

$$L_\sigma = \{x \in \mathbb{H}^{n+1};\ \langle x, a\rangle = \sigma\},$$

where $a \in \mathbb{R}_1^{n+2}$ is a fixed vector, and $\sigma^2 + \langle a, a\rangle > 0$; see [López and Montiel 1999]. Furthermore, from a straightforward computation, we see that the Gauss mapping of such hypersurfaces is given by

$$N(x) = \frac{1}{\sqrt{\sigma^2 + \langle a, a\rangle}}(a + \sigma x) \in \mathbb{S}_1^{n+1}.$$

Consequently, from previous expression we obtain that the angle function f_a of a totally umbilical hypersurface of $\mathbb{H}^{n+1}$ satisfies

$$f_a = \langle N, a\rangle = \sqrt{\sigma^2 + \langle a, a\rangle} = \tau = \text{constant}\,.$$

Hence, it follows from [Montiel 1988, Example 1] that $N(L_\sigma)$ is a totally umbilical spacelike hypersurface of $\mathbb{S}_1^{n+1}$ which is isometric to:

(1) an n-dimensional hyperbolic space of constant sectional curvature $-\dfrac{1}{\tau^2 - 1}$, when a is a unit spacelike vector;

(2) the n-dimensional Euclidean space, when a is a nonzero null vector; or

(3) an n-dimensional sphere of constant sectional curvature $\dfrac{1}{\tau^2+1}$, when a is a unit timelike vector.

On the other hand, given $\psi : \Sigma^n \to \mathbb{H}^{n+1} \subset \mathbb{R}_1^{n+2}$, a totally umbilical hypersurface and $a \in \mathbb{R}_1^{n+2}$ a nonzero fixed vector, a straightforward computation yields that $a^\top$ is a conformal vector field on Σ^n. Indeed, after a choice of an orientation on Σ^n by the unit vector field N, we use (2-5) to deduce that

$$\nabla^2 l_a(X, Y) = (Hf_a + l_a)\langle X, Y\rangle. \tag{3-1}$$

Thus, from (2-7) and (3-1) we conclude that the Lie derivative of the Riemannian metric g of Σ^n in the direction of $a^\top$ satisfies

$$\mathcal{L}_{a^\top} g = \frac{2}{n}(\Delta l_a) g. \tag{3-2}$$

On the other hand, since Σ^n is totally umbilical, we obtain from (2-3), that the Ricci curvature of Σ^n satisfies

$$\mathrm{Ric}_\Sigma(X, Y) = (1-n)(H^2 - 1)\langle X, Y\rangle. \tag{3-3}$$

Hence, from (3-2) and (3-3) we arrive at

$$\mathrm{Ric}_\Sigma + \tfrac{1}{2}\mathcal{L}_{a^\top} g = \big((1-n)(H^2-1) + \tfrac{1}{n}\Delta l_a\big) g. \tag{3-4}$$

Therefore, from (3-4) we conclude that, with an appropriate choice of a nonzero vector $a \in \mathbb{R}_1^{n+2}$, the vector field $a^\top$ provides on Σ^n a nontrivial structure of a gradient almost Ricci soliton.

Motivated by the previous digression, we establish the following characterization concerning gradient almost Ricci solitons immersed in the hyperbolic space, which can also be regarded as a version of the rigidity theorem for hyperbolic hypersurfaces in [Barros et al. 2014a].

Theorem 4. Let $\psi : \Sigma^n \to \mathbb{H}^{n+1}$ be a hypersurface immersed in $\mathbb{H}^{n+1}$. Suppose that for some nonzero vector $a \in \mathbb{R}^{n+2}_1$ the vector field $a^\top$ provides the structure of a gradient almost Ricci soliton for Σ^n. If the image of the Gauss mapping of Σ^n lies in a totally umbilical spacelike hypersurface of $\mathbb{S}^{n+1}_1$ determined by a, then Σ^n is a totally umbilical hypersurface of $\mathbb{H}^{n+1}$.

Proof. Initially, we note that our hypothesis under the image of the Gauss mapping N of Σ^n amounts to the fact that the angle function f_a of Σ^n satisfies $f_a = \langle N, a\rangle = \tau$ on Σ^n, for some constant τ satisfying $\tau^2 > \langle a, a\rangle$. Now, since $a^\top = \nabla l_a$ provides the structure of a gradient almost Ricci soliton for Σ^n, from (1-2)

$$\mathrm{Ric}_\Sigma(\nabla l_a) + \nabla^2 l_a(\nabla l_a) = \lambda \nabla l_a, \tag{3-5}$$

for some smooth function $\lambda : \Sigma^n \to \mathbb{R}$.

On the other hand, from (2-3) we have that

$$\mathrm{Ric}_\Sigma(X) = (1-n)X + nHAX - A^2X.$$

Hence, since f_a is a constant function, we conclude from the above expression that

$$\mathrm{Ric}_\Sigma(\nabla l_a) = (1-n)\nabla l_a. \tag{3-6}$$

Now, we use the (2-5), (3-5), and (3-6) to conclude that

$$(1-n+l_a-\lambda)\nabla l_a = 0, \tag{3-7}$$

on Σ^n. Observe that, from (2-4), we arrive at

$$|\nabla l_a|^2 + \tau^2 - l_a^2 = \langle a, a\rangle. \tag{3-8}$$

Since $\tau^2 > \langle a, a\rangle$, we obtain from (3-8) that the height function l_a has strict sign on Σ^n. Moreover, from (3-7) we have that $l_a - \lambda$ is constant on the open set where $\nabla l_a \neq 0$ and, consequently,

$$|\nabla l_a|^2 \nabla(l_a - \lambda) = 0, \tag{3-9}$$

on Σ^n. Hence, from (3-9) we deduce that

$$\langle \nabla l_a, \nabla\lambda\rangle = |\nabla l_a|^2. \tag{3-10}$$

From Lemma 1, we can use equations (3-6) and (3-10) to deduce that

$$\langle \nabla R, \nabla l_a\rangle = 0. \tag{3-11}$$

Contracting (1-2), we have $\Delta l_a = n\lambda - R$; so,

$$\begin{aligned}\langle \nabla\Delta l_a, \nabla l_a\rangle &= n\langle \nabla\lambda, \nabla l_a\rangle - \langle \nabla R, \nabla l_a\rangle \\ &= n|\nabla l_a|^2. \qquad (3\text{-}12)\end{aligned}$$

On the other hand, from (2-7) we have that

$$\langle \nabla\Delta l_a, \nabla l_a\rangle = n\tau\langle \nabla H, \nabla l_a\rangle + n|\nabla l_a|^2. \qquad (3\text{-}13)$$

Thus, from (3-12) and (3-13) it follows immediately that $\langle \nabla H, \nabla l_a\rangle = 0$, when $\tau \neq 0$. In this case, observing that $a^\top = \nabla l_a$, we have from formula (2-8) that

$$|A|^2 = -\frac{nH}{\tau} l_a. \qquad (3\text{-}14)$$

Since the scalar curvature R of Σ^n is given by

$$R = n(1-n) + n^2H^2 - |A|^2 \qquad (3\text{-}15)$$

and we have $\langle \nabla H, \nabla l_a\rangle = \langle \nabla R, \nabla l_a\rangle = 0$ on Σ^n, we obtain from (3-15) after a simple computation that $a^\top(|A|^2) = 0$ on Σ^n. Thus, from (3-14) we deduce

$$H\nabla l_a = 0. \qquad (3\text{-}16)$$

Hence, taking into account that $\langle \nabla H, \nabla l_a\rangle = 0$, we can use once more formula (2-7) jointly with (3-16) to obtain that

$$nH^2 = -\frac{nH}{\tau} l_a. \qquad (3\text{-}17)$$

Therefore, the equations (3-14) and (3-17) allows us to conclude that $|A|^2 = nH^2$ and this means that Σ^n is a totally umbilical hypersurface of $\mathbb{H}^{n+1}$.

When $\tau = 0$, it follows from (2-5) that the Hessian of the height function l_a satisfies $\nabla^2 l_a = l_a g$, where g stands for the Riemannian metric of Σ^n. Consequently, we conclude that $\nabla l_a = a^\top$ is a conformal vector field on Σ^n. Thus, from Lemma 3 we have that either Σ^n is a totally umbilical hypersurface or $a = a^\top$ on Σ^n. But, since l_a has strict sign on Σ^n, from (2-4) we see that this last situation cannot occur. Hence, we conclude that Σ^n is a totally umbilical hypersurface of $\mathbb{H}^{n+1}$.

Moreover, from Lemma 2 we obtain that

$$\mathscr{L}_{\nabla l_a} R = -2(n-1)\Delta l_a - 2Rl_a,$$

Now, combining this latter with formula (2-7) and (3-11) we deduce that

$$(n(n-1) + R)l_a = 0.$$

By using once more that the height function l_a has strict sign, we conclude from the previous equality that the scalar curvature of Σ^n satisfies $R = n(1-n)$. Consequently, since the umbilicity of Σ^n implies that $|A|^2 = nH^2$, from (3-15) we get $H = 0$ on Σ^n. Therefore, Σ^n must be, in fact, a totally geodesic hypersurface of $\mathbb{H}^{n+1}$. □

From the proof of Theorem 4 we also get the following:

Corollary 5. If Σ^n is a complete hypersurface of $\mathbb{H}^{n+1}$ such that, for some nonzero vector $a \in \mathbb{R}_1^{n+2}$, the vector field $a^\top$ provides on it the nontrivial structure of a gradient almost Ricci soliton and the image of its Gauss mapping lies in a totally umbilical spacelike hypersurface of $\mathbb{S}_1^{n+1}$ determined by a, then Σ^n is isometric to $\mathbb{H}^n$.

4. Characterizing gradient almost Ricci solitons in $\mathbb{S}_1^{n+1}$ and $\mathbb{H}_1^{n+1}$

We start by recalling the description of the totally umbilical hypersurfaces of $\mathbb{M}_c^{n+1} \subset \mathbb{R}_\nu^{n+2}$; see [Abe et al. 1987, Section 4], and also [Montiel 1988, Example 1; Lucas and Ramírez-Ospina 2013, Example 2]. Let $a \in \mathbb{R}_\nu^{n+2}$ be a fixed nonzero vector with $\langle a, a\rangle \in \{-1, 0, 1\}$ and consider the smooth function $h_a : \mathbb{M}_c^{n+1} \to \mathbb{R}$ defined by $h_a(x) = \langle x, a\rangle$. A straightforward computation allows us to conclude that for every real number ϱ, with $\langle a, a\rangle - c\varrho^2 \neq 0$, the level set

$$L_\varrho = h_a^{-1}(\varrho) = \{x \in \mathbb{M}_c^{n+1} : \langle x, a\rangle = \varrho\},$$

is a totally umbilical hypersurface in $\mathbb{M}_c^{n+1}$, with Gauss mapping

$$N(x) = \frac{1}{\sqrt{|\langle a, a\rangle - c\varrho^2|}}(a - c\varrho x). \tag{4-1}$$

Consequently, the corresponding angle function f_a of Σ^n satisfies

$$|f_a| = |\langle N, a\rangle| = \sqrt{|\langle a, a\rangle - c\varrho^2|}. \tag{4-2}$$

It follows from (4-1) that if Σ^n is a totally umbilical hypersurface in $\mathbb{M}_c^{n+1}$, then the image of its Gauss mapping lies in a totally umbilical hypersurface of the hyperbolic space, in the case $c = 1$, and in a totally umbilical spacelike hypersurface of the anti-de Sitter space $\mathbb{H}_1^{n+1}$, in the case $c = -1$. Furthermore, from (4-2) we conclude that f_a must be a constant function on Σ^n and, consequently, we have the following possibilities.

When $c = 1$:

(I.1) if a is a unit spacelike vector, then either $|\varrho| > 1$ and L_ϱ is isometric to an n-dimensional hyperbolic space of constant sectional curvature $-1/(\varrho^2 - 1)$, or $|\varrho| < 1$ and L_ϱ is isometric to an n-dimensional de Sitter space of constant sectional curvature $1/(1 - \varrho^2)$;

(I.2) if a is a nonzero null vector, then $\varrho \neq 0$ and L_ϱ is isometric to an n-dimensional Euclidean space;

(I.3) if a is a unit timelike vector, then L_ϱ is isometric to an n-dimensional Euclidean sphere of constant sectional curvature $1/(1+\varrho^2)$.

When $c=-1$:

(II.1) if a is a unit spacelike vector, then L_ϱ is isometric to the n-dimensional anti-de Sitter space of constant sectional curvature $-1/(\varrho^2+1)$;

(II.2) if a is a nonzero null vector, then $\varrho \neq 0$ and L_ϱ is isometric to the n-dimensional Lorentz–Minkowski space;

(II.3) if a is a unit timelike vector, then either $|\varrho|>1$ and L_ϱ is isometric to an n-dimensional de Sitter space of constant sectional curvature $1/(\varrho^2-1)$, or $|\varrho|<1$ and L_ϱ is isometric to an n-dimensional hyperbolic space of constant sectional curvature $-1/(1-\varrho^2)$.

On the other hand, reasoning as in Section 3, we can verify that if Σ^n is a totally umbilical spacelike hypersurface of $\mathbb{M}_c^{n+1}$, then for an arbitrary fixed vector $a \in \mathbb{R}_\nu^{n+2}$ we have

$$\mathrm{Ric}_\Sigma + \tfrac{1}{2}\mathscr{L}_{a^\top} g = \big((1-n)(H^2+c)+\tfrac{1}{n}\Delta l_a\big)g, \tag{4-3}$$

where g stands for the Riemannian metric of Σ^n. Now (4-3) allows us to conclude that, for a suitable choice of a fixed vector $a \in \mathbb{R}_\nu^{n+2}$, the vector field $a^\top$ provides on Σ^n the nontrivial structure of a gradient almost Ricci soliton.

In a similar way to that of Theorem 4, the previous discussion allows us to establish the following characterization concerning gradient almost Ricci solitons immersed in the de Sitter space:

Theorem 6. Let $\psi : \Sigma^n \to \mathbb{S}_1^{n+1}$ be a spacelike hypersurface immersed in $\mathbb{S}_1^{n+1}$. Suppose that for some nonzero vector $a \in \mathbb{R}_1^{n+2}$ the vector field $a^\top$ provides the structure of a gradient almost Ricci soliton for Σ^n. If the image of the Gauss mapping of Σ^n lies in a totally umbilical hypersurface of $\mathbb{H}^{n+1}$ determined by a, then Σ^n is a totally umbilical hypersurface of $\mathbb{S}_1^{n+1}$.

Proof. Observe that the hypothesis on the image of the Gauss mapping N of Σ^n implies that the angle function f_a of Σ^n satisfies $f_a = \langle N, a\rangle = \tau$ on Σ^n, for some constant τ, with $\tau^2 + \langle a, a\rangle > 0$. Since $a^\top = \nabla l_a$ provides the structure of a gradient almost Ricci soliton for Σ^n, from (1-2), the Ricci curvature of Σ^n satisfies

$$\mathrm{Ric}_\Sigma(\nabla l_a) = \lambda \nabla l_a - \nabla^2 l_a(\nabla l_a), \tag{4-4}$$

for some smooth function $\lambda : \Sigma^n \to \mathbb{R}$, where $\nabla^2 l_a$ stands for the Hessian of the height function $l_a = \langle \psi, a\rangle$.

On the other hand, if we denote by A the Weingarten operator of Σ^n with respect to the normal vector field N and taking into account that f_a is a constant function on Σ^n, we have from Gauss equation that

$$\mathrm{Ric}_\Sigma(\nabla l_a) = (n-1)\nabla l_a. \tag{4-5}$$

Now, from the expression of the Hessian of the height function $l_a = \langle \psi, a\rangle$ and using once more that f_a is constant, we conclude from (4-4) and (4-5) that

$$(n-1-l_a-\lambda)\nabla l_a = 0. \tag{4-6}$$

From (4-6), $l_a - \lambda$ is constant on the open set where $\nabla l_a \neq 0$ and, consequently,

$$|\nabla l_a|^2 \nabla(l_a+\lambda) = 0.$$

This equality allows us to conclude that

$$l_a \langle \nabla l_a, \nabla(l_a+\lambda)\rangle = 0. \tag{4-7}$$

We observe from (2-10) that the height function l_a can be sign changing on Σ^n, since $\tau^2 + \langle a, a\rangle > 0$. However, (4-7) provides us the following identity:

$$l_a \langle \nabla l_a, \nabla\lambda\rangle = -l_a |\nabla l_a|^2. \tag{4-8}$$

Now, from (4-4),

$$l_a \langle \nabla \Delta l_a, \nabla l_a\rangle = n l_a \langle \nabla\lambda, \nabla l_a\rangle - l_a\langle \nabla R, \nabla l_a\rangle.$$

From Lemma 1 and (4-7) we conclude that $l_a\langle \nabla R, \nabla l_a\rangle = 0$ on Σ^n. Furthermore, we use the (4-8) to rewrite the above expression as

$$l_a \langle \nabla \Delta l_a, \nabla l_a\rangle = -n l_a |\nabla l_a|^2. \tag{4-9}$$

On the other hand, since $\Delta l_a = nH\tau - nl_a$, we deduce

$$l_a \langle \nabla \Delta l_a, \nabla l_a\rangle = n l_a \tau\langle \nabla H, \nabla l_a\rangle - n l_a|\nabla l_a|^2. \tag{4-10}$$

From (4-9) and (4-10) it follows that $l_a\langle \nabla H, \nabla l_a\rangle = 0$, when $\tau \neq 0$. Thus, in this case, we obtain from formula (2-14) that

$$|A|^2 = \frac{nH}{\tau} l_a. \tag{4-11}$$

Now, we recall that $a^\top = \nabla l_a$ and that the scalar curvature R of Σ^n is given by $R = n(n-1) - n^2H^2 + |A|^2$. Furthermore, since we have $l_a\langle \nabla H, \nabla l_a\rangle = 0$ and $l_a\langle \nabla R, \nabla l_a\rangle = 0$ on Σ^n, it follows from (4-11) that

$$l_a a^\top(|A|^2) = l_a\langle \nabla |A|^2, \nabla l_a\rangle = l_a\langle \nabla R, \nabla l_a\rangle + 2n^2 H l_a\langle \nabla H, \nabla l_a\rangle = 0$$

on Σ^n. On the other hand, by using once more the Equation (4-11) we arrive at

$$l_a a^\top(|A|^2) = \frac{nHl_a}{\tau}|\nabla l_a|^2.$$

Hence, these previous identities allow us to conclude that

$$Hl_a \nabla l_a = 0.$$

Therefore, since $\Delta l_a = nH\tau - nl_a$, we use the above equality to obtain

$$nH^2 l_a^2 = \frac{nH}{\tau} l_a^3. \tag{4-12}$$

From (4-11) and (4-12) it follows that $|A|^2 l_a^2 = nH^2 l_a^2$ and hence,

$$|\Phi|^2 l_a^2 = 0. \tag{4-13}$$

Now, after a simple algebraic argument, we can write

$$R = n(n-1)(1-H^2) + |\Phi|^2. \tag{4-14}$$

Thus, by using once more that $l_a \langle \nabla H, \nabla l_a \rangle = 0$ and $l_a \langle \nabla R, \nabla l_a \rangle = 0$ on Σ^n it follows from (4-14) that

$$l_a a^\top (|\Phi|^2) = 0. \tag{4-15}$$

On the other hand, from (4-13) we obtain that $|\Phi|^2 l_a = 0$ and, consequently, we deduce from (4-15) that

$$|\Phi|^2 |\nabla l_a|^2 = l_a a^\top (|\Phi|^2) + |\Phi|^2 a^\top (l_a) = a^\top (|\Phi|^2 l_a) = 0. \tag{4-16}$$

Therefore, we obtain that $|\Phi|^2 |\nabla l_a|^2 = 0$ on Σ^n. Thus, since we also have $|\Phi|^2 l_a^2 = 0$ on Σ^n it follows from (2-10) that $|\Phi|^2 = 0$ on Σ^n, because $\tau^2 + \langle a, a\rangle > 0$. This means that Σ^n is a totally umbilical hypersurface of $\mathbb{S}_1^{n+1}$.

When $\tau = 0$, it follows from (2-11) that the Hessian of the height function l_a satisfies $\nabla^2 l_a = -l_a g$, where g is the Riemannian metric of Σ^n. Consequently, we conclude that $\nabla\, l_a = a^\top$ is a conformal vector field on Σ^n. Hence, from Lemma 3, either Σ^n is a totally umbilical hypersurface or $a = a^\top$ on Σ^n. From (2-10), we see that this last situation implies that $l_a = 0$ on Σ^n. On the other hand, from (2-10),

$$|\nabla\, l_a|^2 + l_a^2 = \langle a\,, a\rangle. \tag{4-17}$$

Thus, taking into account that $a = a^\top$ implies that $\langle a, a\rangle > 0$, from (4-17) we reach a contradiction. Hence, Σ^n is a totally umbilical hypersurface.

Now, from Lemma 2,

$$\mathscr{L}_{\nabla l_a} R = -2(n-1)\Delta l_a - 2Rl_a.$$

From (2-13) and using that $l_a \langle \nabla R, \nabla l_a \rangle = 0$ on Σ^n, we deduce from above that

$$(n(n-1) - R)l_a^2 = 0. \tag{4-18}$$

We claim that we must have $R = n(n-1)$. Indeed, otherwise (4-18) implies that $l_a = 0$ on Σ^n. But, as before, this cannot occur. Therefore, reasoning as in the last part of the proof of Theorem 4, we conclude that Σ^n must be, in fact, a totally geodesic hypersurface of $\mathbb{S}_1^{n+1}$. □

From the proof of Theorem 6, we also obtain the following result:

Corollary 7. If Σ^n is a complete spacelike hypersurface of $\mathbb{S}_1^{n+1}$ such that, for some nonzero vector $a \in \mathbb{R}_1^{n+2}$, the vector field $a^\top$ provides on it a nontrivial structure of a gradient almost Ricci soliton and the image of its Gauss mapping lies in a totally umbilical hypersurface of $\mathbb{H}^{n+1}$ determined by a, then Σ^n is isometric to $\mathbb{S}^n$.

Finally, we can reason in an analogous way to the proof of Theorem 6 in order to establish corresponding versions of Theorem 6 and Corollary 7 for the case that the ambient space is the anti-de Sitter space $\mathbb{H}_1^{n+1}$. More precisely:

Theorem 8. Let $\psi : \Sigma^n \to \mathbb{H}_1^{n+1}$ be a spacelike hypersurface immersed in $\mathbb{H}_1^{n+1}$. Suppose that for some nonzero vector $a \in \mathbb{R}_2^{n+2}$ the vector field $a^\top$ provides the structure of a gradient almost Ricci soliton for Σ^n. If the image of the Gauss mapping of Σ^n lies in a totally umbilical hypersurface of $\mathbb{H}_1^{n+1}$ determined by a, then Σ^n is a totally umbilical hypersurface of $\mathbb{H}_1^{n+1}$.

Corollary 9. If Σ^n is a complete spacelike hypersurface of $\mathbb{H}_1^{n+1}$ such that, for some nonzero vector $a \in \mathbb{R}_2^{n+2}$, the vector field $a^\top$ provides on it a nontrivial structure of a gradient almost Ricci soliton and the image of its Gauss mapping lies in a totally umbilical hypersurface of $\mathbb{H}_1^{n+1}$ determined by a, then Σ^n is isometric to $\mathbb{H}^n$.

Acknowledgements

Aquino is partially supported by CNPq, Brazil, grant 302738/2014-2, de Lima is partially supported by CNPq, Brazil, grant 303977/2015-9, and Gomes is partially supported by CNPq, Brazil. The authors would like to thank the referee for reading the manuscript in great detail and for his/her valuable suggestions and useful comments which improved the paper.

References

[Abe et al. 1987] N. Abe, N. Koike, and S. Yamaguchi, "Congruence theorems for proper semi-Riemannian hypersurfaces in a real space form", *Yokohama Math. J.* **35**:1-2 (1987), 123–136. MR Zbl

[Aiyama 1992] R. Aiyama, "On the Gauss map of complete space-like hypersurfaces of constant mean curvature in Minkowski space", *Tsukuba J. Math.* **16**:2 (1992), 353–361. MR Zbl

[Aledo and Alías 2002] J. A. Aledo and L. J. Alías, "On the volume and the Gauss map image of spacelike hypersurfaces in de Sitter space", *Proc. Amer. Math. Soc.* **130**:4 (2002), 1145–1151. MR Zbl

[Aquino and de Lima 2014] C. P. Aquino and H. F. de Lima, "On the umbilicity of complete constant mean curvature spacelike hypersurfaces", *Math. Ann.* **360**:3 (2014), 555–569. MR Zbl

[Barros and Ribeiro 2012] A. Barros and E. Ribeiro, Jr., "Some characterizations for compact almost Ricci solitons", *Proc. Amer. Math. Soc.* **140**:3 (2012), 1033–1040. MR Zbl

[Barros et al. 2011] A. Barros, J. N. Gomes, and E. Ribeiro, Jr., "Immersion of almost Ricci solitons into a Riemannian manifold", *Mat. Contemp.* **40** (2011), 91–102. MR Zbl

[Barros et al. 2014a] A. Barros, C. Aquino, and H. de Lima, "Complete CMC hypersurfaces in the hyperbolic space with prescribed Gauss mapping", *Proc. Amer. Math. Soc.* **142**:10 (2014), 3597–3604. MR Zbl

[Barros et al. 2014b] A. Barros, R. Batista, and E. Ribeiro, Jr., "Compact almost Ricci solitons with constant scalar curvature are gradient", *Monatsh. Math.* **174**:1 (2014), 29–39. MR Zbl

[Beem et al. 1996] J. K. Beem, P. E. Ehrlich, and K. L. Easley, *Global Lorentzian geometry*, 2nd ed., Monographs and Textbooks in Pure and Applied Mathematics **202**, Marcel Dekker, New York, 1996. MR Zbl

[Cao 2010] H.-D. Cao, "Recent progress on Ricci solitons", pp. 1–38 in *Recent advances in geometric analysis*, edited by Y.-I. Lee et al., Advanced Lectures in Mathematics **11**, Int. Press, Somerville, MA, 2010. MR Zbl

[Catino 2012] G. Catino, "Generalized quasi-Einstein manifolds with harmonic Weyl tensor", *Math. Z.* **271**:3 (2012), 751–756. MR Zbl

[Catino et al. 2016] G. Catino, L. Cremaschi, Z. Djadli, C. Mantegazza, and L. Mazzieri, "The Ricci–Bourguignon flow", preprint, 2016. arXiv

[Chen and Deshmukh 2014] B.-Y. Chen and S. Deshmukh, "Classification of Ricci solitons on Euclidean hypersurfaces", *Internat. J. Math.* **25**:11 (2014), art. id. 1450104, 22 pp. MR

[Feitosa et al. 2015] F. E. S. Feitosa, A. A. Freitas, J. N. V. Gomes, and R. S. Pina, "On the construction of gradient almost Ricci soliton warped product", preprint, 2015. arXiv

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

[Kim et al. 2002] D.-S. Kim, S.-B. Kim, Y. H. Kim, and S.-H. Park, "Conformal vector fields and totally umbilic hypersurfaces", *Bull. Korean Math. Soc.* **39**:4 (2002), 671–680. MR Zbl

[López and Montiel 1999] R. López and S. Montiel, "Existence of constant mean curvature graphs in hyperbolic space", *Calc. Var. Part. Diff. Eq.* **8**:2 (1999), 177–190. MR Zbl

[Lucas and Ramírez-Ospina 2013] P. Lucas and H. F. Ramírez-Ospina, "Hypersurfaces in pseudo-Euclidean spaces satisfying a linear condition on the linearized operator of a higher order mean curvature", *Differential Geom. Appl.* **31**:2 (2013), 175–189. MR Zbl

[Mastrolia et al. 2013] P. Mastrolia, M. Rigoli, and M. Rimoldi, "Some geometric analysis on generic Ricci solitons", *Commun. Contemp. Math.* **15**:3 (2013), art. id. 1250058, 25 pp. MR

[Montiel 1988] S. Montiel, "An integral inequality for compact spacelike hypersurfaces in de Sitter space and applications to the case of constant mean curvature", *Indiana Univ. Math. J.* **37**:4 (1988), 909–917. MR Zbl

[Nomizu and Smyth 1969] K. Nomizu and B. Smyth, "On the Gauss mapping for hypersurfaces of constant mean curvature in the sphere", *Comment. Math. Helv.* **44** (1969), 484–490. MR Zbl

[O'Neill 1983] B. O'Neill, *Semi-Riemannian geometry: with applications to relativity*, Pure and Applied Mathematics **103**, Academic Press, New York, 1983. MR Zbl

[Palmer 1990] B. Palmer, "The Gauss map of a spacelike constant mean curvature hypersurface of Minkowski space", *Comment. Math. Helv.* **65**:1 (1990), 52–57. MR Zbl

[Pigola et al. 2011] S. Pigola, M. Rigoli, M. Rimoldi, and A. G. Setti, "Ricci almost solitons", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **10**:4 (2011), 757–799. MR Zbl

[Rosenberg 1993] H. Rosenberg, "Hypersurfaces of constant curvature in space forms", *Bull. Sci. Math.* **117**:2 (1993), 211–239. MR Zbl

[Sharma and Duggal 1985] R. Sharma and K. L. Duggal, "Characterization of an affine conformal vector field", *C. R. Math. Rep. Acad. Sci. Canada* **7**:3 (1985), 201–205. MR Zbl

[Xin 1991] Y. L. Xin, "On the Gauss image of a spacelike hypersurface with constant mean curvature in Minkowski space", *Comment. Math. Helv.* **66**:4 (1991), 590–598. MR Zbl

[Yano 1970] K. Yano, *Integral formulas in Riemannian geometry*, Pure and Applied Mathematics **1**, Marcel Dekker, New York, 1970. MR Zbl

Received October 13, 2015. Revised November 18, 2016.

CÍCERO P. AQUINO
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DO PIAUÍ
64049-550 TERESINA
BRAZIL
cicero.aquino@ufpi.edu.br

HENRIQUE F. DE LIMA
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
AV. APRIGIO VELOSO 882
BLOCO CX
BODOCONGO
58429-900 CAMPINA GRANDE
BRAZIL
henrique@mat.ufcg.edu.br

JOSÉ N. V. GOMES
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DO AMAZONAS
69077-070 MANAUS
BRAZIL
jnvgomes@pq.cnpq.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.307

WEIGHTED SOBOLEV REGULARITY OF THE BERGMAN PROJECTION ON THE HARTOGS TRIANGLE

LIWEI CHEN

We prove a weighted Sobolev estimate of the Bergman projection on the Hartogs triangle, where the weight is some power of the distance to the singularity at the boundary. This method also applies to the *n*-dimensional generalization of the Hartogs triangle.

1. Introduction

Setup and background. Let Ω be a domain in $\mathbb{C}^n$. The set of square integrable holomorphic functions on Ω, denoted by $A^2(\Omega)$, forms a closed subspace of the Hilbert space $L^2(\Omega)$. The Bergman projection associated to Ω is the orthogonal projection

$$\mathcal{B} : L^2(\Omega) \to A^2(\Omega),$$

which has an integral representation

$$\mathcal{B}(f)(z) = \int_\Omega B(z,\zeta) f(\zeta)\, d(\zeta), \tag{1-1}$$

for all $f \in L^2(\Omega)$ and $z \in \Omega$. Here the function $B(z,\zeta)$ defined on $\Omega \times \Omega$ is the Bergman kernel, and $d(\zeta) = dV(\zeta)$ is the usual Euclidean volume form.

The regularity of the Bergman projection $\mathcal{B}$ associated to Ω in $L^p(\Omega)$, $W^{k,p}(\Omega)$, and Hölder spaces are of particular interest. When Ω is bounded, smooth, and pseudoconvex with additional geometric condition on the boundary (e.g., strongly pseudoconvex), the regularity of $\mathcal{B}$ in these spaces has been intensively studied in the literature. See, for example, [Lanzani and Stein 2012] and references therein for details.

When Ω is nonsmooth, there are relatively few results in regard to the regularity of the Bergman projection. Even in $L^p(\Omega)$, we cannot expect the regularity to hold for all $p \in (1,\infty)$. If Ω is a simply connected planar domain, then the interval of p for $\mathcal{B}$ to be L^p-bounded highly depends on the geometry of the boundary; see [Lanzani and Stein 2004]. If Ω is a nonsmooth worm domain, then the interval of p depends on the winding of the domain; see [Krantz and Peloso 2008]. If Ω is an

MSC2010: 32A07, 32A25, 32A50.
Keywords: Hartogs triangle, Bergman projection, Sobolev regularity.

inflation of the unit disc by the norm square of a nonvanishing holomorphic function, then the interval of p depends on the boundary behavior of the holomorphic function on the unit disc; see [Zeytuncu 2013].

Results. In this article, we consider the Sobolev regularity of the Bergman projection $\mathcal{B}$ on the Hartogs triangle $\mathbb{H}$, where the Hartogs triangle is defined as

$$\mathbb{H} = \{(z_1, z_2) \in \mathbb{C}^2 \mid |z_1| < |z_2| < 1\}.$$

The Hartogs triangle is a classical nonsmooth domain in $\mathbb{C}^2$. It is well known that the boundary at $(0, 0)$ is not even Lipschitz, and the topological closure of $\mathbb{H}$ does not possess a Stein neighborhood basis. In [Chen 2017a], the L^p regularity of $\mathcal{B}$ on $\mathbb{H}$ was studied: the Bergman projection $\mathcal{B}$ is L^p-bounded if and only if $p \in \left(\frac{4}{3}, 4\right)$. On the other hand, we have $\bar{z}_2 \in W^{k,p}(\mathbb{H})$ for all nonnegative integers k and all $p \in [1, \infty]$, but $\mathcal{B}(\bar{z}_2) = c/z_2 \notin W^{1,p}(\mathbb{H})$ for $p \geq 2$, where c is some nonzero constant. So we cannot expect to obtain regularity in the ordinary Sobolev spaces, nor for all $p \in (1, \infty)$.

A natural way to control the boundary behavior of singularities is the use of weights which measure the distance from the points near the boundary to the singularity at the boundary. Since on the Hartogs triangle we have $|z_2| < |z| < \sqrt{2}|z_2|$, where $z = (z_1, z_2) \in \mathbb{H}$, it is reasonable to consider a weight of the form $|z_2|^s$, for some $s \in \mathbb{R}$. On the other hand, based on the L^p mapping property of the Bergman projection on $\mathbb{H}$ (see [Chakrabarti and Zeytuncu 2016]) and the Sobolev regularity of the weighted canonical solution operator of the $\bar{\partial}$-equation on $\mathbb{H}$ (see [Chakrabarti and Shaw 2013]), it is also reasonable to put a weight of the form $|z_2|^s$ on the target space. Therefore, we consider the following weighted Sobolev spaces:

Definition 1.1. On the Hartogs triangle $\mathbb{H}$, for each $k \in \mathbb{Z}^+ \cup \{0\}$, $s \in \mathbb{R}$, and $p \in (1, \infty)$, we define the *weighted Sobolev space* by

$$W^{k,p}(\mathbb{H}, \delta^s) = \{f \in L^1_{\mathrm{loc}}(\mathbb{H}) \mid \|f\|_{k,p,s} < \infty\},$$

where $\delta(z) = |z_2| \approx |z|$, and the norm is defined as

$$\|f\|_{k,p,s} = \left(\int_{\mathbb{H}} \sum_{|\alpha| \leq k} |D^{\alpha}_{z,\bar{z}}(f)(z)|^p |z_2|^s \, dz \right)^{1/p}.$$

Here $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ is the multi-index running over all $|\alpha| \leq k$, and

$$D^{\alpha}_{z,\bar{z}} = \frac{\partial^{|\alpha|}}{\partial z_1^{\alpha_1} \partial z_2^{\alpha_2} \partial \bar{z}_1^{\alpha_3} \partial \bar{z}_2^{\alpha_4}}.$$

We also denote the usual norm in the (unweighted) Sobolev space $W^{k,p}(\mathbb{H})$ by

$$\|f\|_{k,p} = \left(\int_{\mathbb{H}} \sum_{|\alpha| \le k} |D^{\alpha}_{z,\bar{z}}(f)(z)|^p \, dz \right)^{1/p}.$$

With the definition above, we can state our main result:

Theorem 1.2. *The Bergman projection $\mathcal{B}$ on the Hartogs triangle $\mathbb{H}$ maps continuously from $W^{k,p}(\mathbb{H})$ to $W^{k,p}(\mathbb{H}, \delta^{kp})$ for $p \in \left(\frac{4}{3}, 4\right)$.*

That is, for each $k \in \mathbb{Z}^+ \cup \{0\}$ and $p \in \left(\frac{4}{3}, 4\right)$, there exists a constant $C_{k,p} > 0$, such that

$$\|\mathcal{B}(f)\|_{k,p,kp} \le C_{k,p} \|f\|_{k,p} \quad \textit{for any} \quad f \in W^{k,p}(\mathbb{H}).$$

Remark 1.3. It is clear that $\mathcal{B}$ doesn't lose any derivatives away from the singular point of the Hartogs triangle. If we put a suitable power of the weight δ around the singularity on the target space, then there is no loss of differentiability of $\mathcal{B}(f)$ around the singular point (see also the result in [Chakrabarti and Shaw 2013]).

Remark 1.4. Note that we have $\mathcal{B}(\bar{z}_2) = c/z_2 \notin W^{k,p}(\mathbb{H}, \delta^{kp})$ for $p \ge 4$, where c is some nonzero constant. So we cannot obtain regularity for $p \ge 4$, unless we use more weights on the target space. Conversely, we can only obtain regularity for fewer values of p, if we use less weights on the target space.

Organization and outline. The idea of the proof of the main result is the following. In Section 2, we start with an idea from [Chakrabarti and Shaw 2013] to transfer $\mathbb{H}$ to the product model $\mathbb{D} \times \mathbb{D}^*$, as well as to transfer the differential operators D^{α} to the ones in new variables. From this, we focus on the integration over the punctured disc $\mathbb{D}^*$ in Section 3. We then use an idea from [Straube 1986] to convert D^{α} acting on the Bergman kernel in the holomorphic component to the ones acting on the kernel in the antiholomorphic part. The resulting differential operators can be written as a combination of tangential operators, and therefore, integration by parts applies to the smooth functions. Finally, in Section 4, we apply the weighted L^p estimates in [Chen 2017b] to our integral, and the resulting integral is majorized by the weighted L^p norm of $D^{\alpha}(f)$. To complete the proof, we approximate the weighted Sobolev functions by smooth functions and transfer the product model back to $\mathbb{H}$.

2. Transfer to the product model

Transfer $\mathbb{H}$ to $\mathbb{D} \times \mathbb{D}^*$. In view of Definition 1.1, we adopt the following notation.

Definition 2.1. Let $\beta = (\beta_1, \beta_2)$ be a multi-index, we use the notations below to denote the differential operators

$$D^{\beta}_{z} = \frac{\partial^{|\beta|}}{\partial z_1^{\beta_1} \partial z_2^{\beta_2}} \quad \text{and} \quad D^{\beta}_{z_j,\bar{z}_j} = \frac{\partial^{|\beta|}}{\partial z_j^{\beta_1} \partial \bar{z}_j^{\beta_2}} \quad \text{for} \quad j = 1, 2.$$

From the result in [Chen 2017a], we see that $\mathcal{B}(f) \in A^p(\mathbb{H})$ (the set of L^p functions that are holomorphic), whenever $p \in \left(\frac{4}{3}, 4\right)$ and $f \in L^p(\mathbb{H})$. So we can rewrite the weighted L^p Sobolev norm of $\mathcal{B}(f)$ as

$$\|\mathcal{B}(f)\|^p_{k,p,kp} = \sum_{|\beta|\le k} \int_{\mathbb{H}} \left|D^\beta_z(\mathcal{B}(f))(z)\right|^p |z_2|^{kp}\,dz, \tag{2-1}$$

where β and D^β_z are as in Definition 2.1.

In order to transfer $\mathbb{H}$ to the product model, we first recall the transformation formula for the Bergman kernels.

Proposition 2.2. *Let Ω_j be a domain in $\mathbb{C}^n$ and B_j be its Bergman kernel on $\Omega_j \times \Omega_j$, $j = 1, 2$. Suppose $\Psi : \Omega_1 \to \Omega_2$ is a biholomorphism, then for $(w, \eta) \in \Omega_1 \times \Omega_1$ we have*

$$\det J_{\mathbb{C}}\Psi(w) B_2(\Psi(w), \Psi(\eta)) \det \overline{J_{\mathbb{C}}\Psi(\eta)} = B_1(w, \eta).$$

Proof. See, for example, [Krantz 1992, Proposition 1.4.12]. □

Now let us consider the biholomorphism $\Phi : \mathbb{H} \to \mathbb{D} \times \mathbb{D}^*$ with its inverse $\Psi : \mathbb{D} \times \mathbb{D}^* \to \mathbb{H}$, where

$$\Phi(z_1, z_2) = \left(\frac{z_1}{z_2}, z_2\right) \quad \text{and} \quad \Psi(w_1, w_2) = (w_1 w_2, w_2).$$

A simple computation shows $\det J_{\mathbb{C}}\Psi(w) = w_2$, for $w = (w_1, w_2) \in \mathbb{D} \times \mathbb{D}^*$. Therefore, by the proposition above, we have

$$B(\Psi(w), \Psi(\eta)) = \frac{1}{w_2\bar{\eta}_2} \cdot \frac{1}{(1 - w_1\bar{\eta}_1)^2} \cdot \frac{1}{(1 - w_2\bar{\eta}_2)^2}, \tag{2-2}$$

where B is the Bergman kernel on $\mathbb{H} \times \mathbb{H}$ as in (1-1) and $(w, \eta) \in \mathbb{D} \times \mathbb{D}^* \times \mathbb{D} \times \mathbb{D}^*$.

Transfer the differential operators. We next need to transfer the differential operators D^β_z to the ones in the new variable w. We need a lemma.

Lemma 2.3. *Under the biholomorphism $\Phi(z) = w$, for each β let $m = |\beta|$. Then*

$$D^\beta_z = \sum_{a+b\le m} \frac{p_{a,b,\beta}(w_1)}{w_2^{m-b}} \cdot \frac{\partial^{a+b}}{\partial w_1^a \partial w_2^b}, \tag{2-3}$$

where $p_{a,b,\beta}(w_1)$ is a polynomial of degree at most m in variable w_1. In addition, if $|\beta| \le k$ for some $k \in \mathbb{Z}^+ \cup \{0\}$, then $|p_{a,b,\beta}(w_1)| \le C_k$ on $\mathbb{D}$ uniformly in β, a, and b, for some constant $C_k > 0$ depending only on k.

Proof. We prove (2-3) by induction on $m = |\beta|$. The case $m = 0$ is trivial. When $m = 1$, a direct computation shows

$$\frac{\partial}{\partial z_1} = \frac{1}{w_2} \cdot \frac{\partial}{\partial w_1} \quad \text{and} \quad \frac{\partial}{\partial z_2} = -\frac{w_1}{w_2} \cdot \frac{\partial}{\partial w_1} + \frac{\partial}{\partial w_2}.$$

So both $\partial/\partial z_1$ and $\partial/\partial z_2$ are of the form in (2-3).

Suppose for all β with $|\beta| = m$, the D_z^β are of the form in (2-3). We now check the case $|\beta'| = m+1$. Note that $D_z^{\beta'} = (\partial/\partial z_1) \circ D_z^\beta$ or $D_z^{\beta'} = (\partial/\partial z_2) \circ D_z^\beta$ for some β. By the inductive assumption, we have

$$\begin{aligned}\frac{\partial}{\partial z_1} \circ D_z^\beta &= \frac{1}{w_2} \cdot \frac{\partial}{\partial w_1} \circ \sum_{a+b \le m} \frac{p_{a,b,\beta}(w_1)}{w_2^{m-b}} \cdot \frac{\partial^{a+b}}{\partial w_1^a \partial w_2^b} \\ &= \sum_{a+b \le m} \frac{p'_{a,b,\beta}(w_1)}{w_2^{m+1-b}} \cdot \frac{\partial^{a+b}}{\partial w_1^a \partial w_2^b} + \frac{p_{a,b,\beta}(w_1)}{w_2^{m+1-b}} \cdot \frac{\partial^{a+b+1}}{\partial w_1^{a+1} \partial w_2^b} \\ &= \sum_{a+b \le m+1} \frac{p_{a,b,\beta'}(w_1)}{w_2^{m+1-b}} \cdot \frac{\partial^{a+b}}{\partial w_1^a \partial w_2^b},\end{aligned}$$

and

$$\begin{aligned}\frac{\partial}{\partial z_2} \circ D_z^\beta &= \Big(-\frac{w_1}{w_2} \cdot \frac{\partial}{\partial w_1} + \frac{\partial}{\partial w_2}\Big) \circ \sum_{a+b \le m} \frac{p_{a,b,\beta}(w_1)}{w_2^{m-b}} \cdot \frac{\partial^{a+b}}{\partial w_1^a \partial w_2^b} \\ &= \sum_{a+b \le m} \frac{-w_1 p'_{a,b,\beta}(w_1)}{w_2^{m+1-b}} \cdot \frac{\partial^{a+b}}{\partial w_1^a \partial w_2^b} + \frac{-w_1 p_{a,b,\beta}(w_1)}{w_2^{m+1-b}} \cdot \frac{\partial^{a+b+1}}{\partial w_1^{a+1} \partial w_2^b} \\ &\qquad + \frac{(b-m) p_{a,b,\beta}(w_1)}{w_2^{m+1-b}} \cdot \frac{\partial^{a+b}}{\partial w_1^a \partial w_2^b} + \frac{p_{a,b,\beta}(w_1)}{w_2^{m-b}} \cdot \frac{\partial^{a+b+1}}{\partial w_1^a \partial w_2^{b+1}} \\ &= \sum_{a+b \le m+1} \frac{p_{a,b,\beta'}(w_1)}{w_2^{m+1-b}} \cdot \frac{\partial^{a+b}}{\partial w_1^a \partial w_2^b}.\end{aligned}$$

We see that $p_{a,b,\beta'}(w_1)$ is a polynomial of degree at most $m+1$ and $D_z^{\beta'}$ has the form in (2-3).

When $|\beta| \le k$, all the possible combinations of derivatives in D_z^β are finite. So there are finitely many different coefficients in all of the $p_{a,b,\beta}(w_1)$. Since $|w_1| \le 1$ on $\mathbb{D}$ and $a, b \le m \le k$, we obtain $|p_{a,b,\beta}(w_1)| \le C_k$ on $\mathbb{D}$ as desired. □

Now we can transfer $\mathbb{H}$ to the product model $\mathbb{D} \times \mathbb{D}^*$ by the biholomorphism Φ. Combining (2-2) and (2-3), we see that the right hand side of (2-1) becomes

$$\sum_{|\beta| \le k} \int_{\mathbb{D} \times \mathbb{D}^*} \Big| \sum_{a+b \le |\beta|} \int_{\mathbb{D} \times \mathbb{D}^*} K_{a,b,\beta}(w,\eta) f(\Psi(\eta)) |\eta_2|^2 \, d\eta \Big|^p |w_2|^{kp+2} \, dw, \tag{2-4}$$

where

$$K_{a,b,\beta}(w,\eta) = \frac{p_{a,b,\beta}(w_1)}{w_2^{|\beta|-b}} \cdot \frac{\partial^a}{\partial w_1^a} \Big(\frac{1}{(1-w_1\bar{\eta}_1)^2} \Big) \cdot \frac{\partial^b}{\partial w_2^b} \Big(\frac{1}{w_2 \bar{\eta}_2} \cdot \frac{1}{(1-w_2\bar{\eta}_2)^2} \Big).$$

3. Convert the differential operators on $\mathbb{D}^*$

Convert to the antiholomorphic part. Since $\mathbb{D}^*$ is a Reinhardt domain, by using the idea in [Straube 1986], we can convert the differential operators as follows.

Lemma 3.1. *As in (2-4), for the last factor in $K_{a,b,\beta}(w,\eta)$, we have*

$$\frac{\partial^b}{\partial w_2^b}\left(\frac{1}{w_2\bar{\eta}_2}\cdot\frac{1}{(1-w_2\bar{\eta}_2)^2}\right)=\frac{\bar{\eta}_2^b}{w_2^b}\cdot\frac{\partial^b}{\partial\bar{\eta}_2^b}\left(\frac{1}{w_2\bar{\eta}_2}\cdot\frac{1}{(1-w_2\bar{\eta}_2)^2}\right). \tag{3-1}$$

Proof. The kernel in (3-1) is the weighted Bergman kernel associated to $\mathbb{D}^*$ with the weight $|z|^2$; see [Chen 2017b]. It has the expansion

$$\frac{1}{w_2\bar{\eta}_2}\cdot\frac{1}{(1-w_2\bar{\eta}_2)^2}=\sum_{j=0}^{\infty}(j+1)(w_2\bar{\eta}_2)^{j-1},$$

which converges uniformly on every compact subset $K\times K\subset\mathbb{D}^*\times\mathbb{D}^*$. Differentiate the series term by term, and we see that

$$\begin{aligned}w_2^b\cdot\frac{\partial^b}{\partial w_2^b}\left(\frac{1}{w_2\bar{\eta}_2}\cdot\frac{1}{(1-w_2\bar{\eta}_2)^2}\right)&=\sum_{j=0}^{\infty}(j+1)w_2^b\cdot\frac{\partial^b}{\partial w_2^b}(w_2\bar{\eta}_2)^{j-1}\\&=\sum_{j=0}^{\infty}(j+1)\bar{\eta}_2^b\cdot\frac{\partial^b}{\partial\bar{\eta}_2^b}(w_2\bar{\eta}_2)^{j-1}\\&=\bar{\eta}_2^b\cdot\frac{\partial^b}{\partial\bar{\eta}_2^b}\left(\frac{1}{w_2\bar{\eta}_2}\cdot\frac{1}{(1-w_2\bar{\eta}_2)^2}\right).\end{aligned}$$

□

Integration by parts. Now we focus on the integration over $\mathbb{D}^*$ in (2-4). We first define a "tangential" operator.

Definition 3.2. Let $S_w=w(\partial/\partial w)$ be the *complex normal differential operator* on a neighborhood of $\partial\mathbb{D}$. We define the *tangential operator* by

$$T_w=\Im(S_w)=\frac{1}{2i}\left(w\frac{\partial}{\partial w}-\bar{w}\frac{\partial}{\partial\bar{w}}\right).$$

Remark 3.3. Indeed, T_w is well defined on a neighborhood of $\overline{\mathbb{D}}$. Moreover, for any disc $\mathbb{D}_\rho=\{|w|<\rho\}$ of radius $\rho<1$ with defining function $r_\rho(w)=|w|^2-\rho^2$, we have

$$T_w(r_\rho)=0 \tag{3-2}$$

on $\partial\mathbb{D}_\rho$. That is, T_w is tangential on $\partial\mathbb{D}_\rho$ for all $\rho<1$.

In order to make use of integration by parts, we need the following lemma:

Lemma 3.4. *Let T_w be as above. For $b \in \mathbb{Z}^+ \cup \{0\}$, we have*

$$T_w^b \equiv \sum_{j=0}^{b} c_j \overline{w}^j \frac{\partial^j}{\partial \overline{w}^j} \quad \Big(\mathrm{mod}\ \frac{\partial}{\partial w}\Big), \tag{3-3}$$

where the c_j are constants, $c_b \neq 0$, and T_w^b is the composition of b copies of T_w.

Proof. We prove (3-3) by induction on b. The case $b = 0$ is trivial. When $b = 1$, it is easy to see that

$$T_w \equiv -\frac{1}{2i} \overline{w} \frac{\partial}{\partial \overline{w}} \quad \Big(\mathrm{mod}\ \frac{\partial}{\partial w}\Big).$$

Suppose (3-3) holds for some b. Then we see that

$$T_w^b = \sum_{j=0}^{b} c_j \overline{w}^j \frac{\partial^j}{\partial \overline{w}^j} + A \circ \frac{\partial}{\partial w},$$

for some operator A. So for the case $b+1$, we have

$$\begin{aligned} T_w \circ T_w^b &= \frac{1}{2i}\Big(w\frac{\partial}{\partial w} - \overline{w}\frac{\partial}{\partial \overline{w}}\Big) \circ \Big(\sum_{j=0}^{b} c_j \overline{w}^j \frac{\partial^j}{\partial \overline{w}^j} + A \circ \frac{\partial}{\partial w}\Big) \\ &= \frac{1}{2i}\Big(\sum_{j=0}^{b} c_j w \overline{w}^j \frac{\partial^j}{\partial \overline{w}^j}\frac{\partial}{\partial w} - j c_j \overline{w}^j \frac{\partial^j}{\partial \overline{w}^j} - c_j \overline{w}^{j+1} \frac{\partial^{j+1}}{\partial \overline{w}^{j+1}}\Big) + T_w \circ A \circ \frac{\partial}{\partial w} \\ &= \sum_{j=0}^{b+1} c'_j \overline{w}^j \frac{\partial^j}{\partial \overline{w}^j} + A' \circ \frac{\partial}{\partial w}, \end{aligned}$$

for some constants c'_j with $c'_{b+1} = -(1/2i)c_b \neq 0$ and some operator A'. Therefore, (3-3) holds for T_w^{b+1}. □

Combine (3-1) and (3-3). Since the kernel in (3-1) is antiholomorphic in η_2, the inside integration over $\mathbb{D}^*$ with regard to variable η_2 in (2-4) denoted by I becomes

$$\begin{aligned} I &= \int_{\mathbb{D}^*} \frac{\partial^b}{\partial w_2^b}\Big(\frac{1}{w_2\overline{\eta}_2} \cdot \frac{1}{(1-w_2\overline{\eta}_2)^2}\Big) f(\Psi(\eta))|\eta_2|^2\, d\eta_2 \\ &= \int_{\mathbb{D}^*} \frac{\overline{\eta}_2^b}{w_2^b} \cdot \frac{\partial^b}{\partial \overline{\eta}_2^b}\Big(\frac{1}{w_2\overline{\eta}_2} \cdot \frac{1}{(1-w_2\overline{\eta}_2)^2}\Big) f(\Psi(\eta))|\eta_2|^2\, d\eta_2 \\ &= \frac{1}{w_2^b} \int_{\mathbb{D}^*} \sum_{j=0}^{b} c_j T_{\eta_2}^j \Big(\frac{1}{w_2\overline{\eta}_2} \cdot \frac{1}{(1-w_2\overline{\eta}_2)^2}\Big) f(\Psi(\eta))|\eta_2|^2\, d\eta_2 \\ &= \frac{1}{w_2^b} \sum_{j=0}^{b} c_j \lim_{\epsilon \to 0^+} \int_{\mathbb{D}-\mathbb{D}_\epsilon} T_{\eta_2}^j \Big(\frac{1}{w_2\overline{\eta}_2} \cdot \frac{1}{(1-w_2\overline{\eta}_2)^2}\Big) f(\Psi(\eta))|\eta_2|^2\, d\eta_2. \end{aligned}$$

Let us assume in addition for a moment that $f(\Psi(\eta))$ belongs to $C^\infty(\bar{\mathbb{D}}-\{0\})$ in variable η_2. Then by (3-2) we obtain

$$
\begin{aligned}
(3\text{-}4)\quad I &= \frac{1}{w_2^b}\sum_{j=0}^{b} c_j \lim_{\epsilon\to 0^+}\int_{\mathbb{D}-\mathbb{D}_\epsilon} T^j_{\eta_2}\Big(\frac{1}{w_2\bar\eta_2}\cdot\frac{1}{(1-w_2\bar\eta_2)^2}\Big) f(\Psi(\eta))|\eta_2|^2\,d\eta_2\\
&= \frac{1}{w_2^b}\sum_{j=0}^{b} c_j(-1)^j \lim_{\epsilon\to 0^+}\int_{\mathbb{D}-\mathbb{D}_\epsilon} \frac{1}{w_2\bar\eta_2}\cdot\frac{1}{(1-w_2\bar\eta_2)^2} T^j_{\eta_2}\big(f(\Psi(\eta))|\eta_2|^2\big)\,d\eta_2\\
&= \frac{1}{w_2^b}\sum_{j=0}^{b} (-1)^j c_j \int_{\mathbb{D}^*} \frac{1}{w_2\bar\eta_2}\cdot\frac{1}{(1-w_2\bar\eta_2)^2} T^j_{\eta_2}\big(f(\Psi(\eta))\big)|\eta_2|^2\,d\eta_2,
\end{aligned}
$$

where the last line follows from the fact that $T_{\eta_2}(|\eta_2|^2)=0$.

Definition 3.5. We use the following notation:

$$
F_j(\eta)=T^j_{\eta_2}\big(f(\Psi(\eta))\big)\cdot\eta_2 \quad\text{and}\quad \mathcal{B}_{1,a}(g)(w_1)=\int_{\mathbb{D}}\frac{\partial^a}{\partial w_1^a}\Big(\frac{1}{(1-w_1\bar\eta_1)^2}\Big)g(\eta_1)\,d\eta_1,
$$

for any g whenever the integral is well defined, and

$$
\mathcal{B}_2(h)(w_2)=\int_{\mathbb{D}^*}\frac{h(\eta_2)}{(1-w_2\bar\eta_2)^2}\,d\eta_2,
$$

for any h whenever the integral is well defined.

By (3-4) and the notation above (Definition 3.5), we see that (2-4) becomes

$$
(3\text{-}5)\quad \sum_{|\beta|\le k}\int_{\mathbb{D}\times\mathbb{D}^*}\Big|\sum_{a+b\le|\beta|}\frac{p_{a,b,\beta}(w_1)}{w_2^{|\beta|+1}}\sum_{j=0}^{b}(-1)^j c_j\mathcal{B}_{1,a}\big(\mathcal{B}_2(F_j)\big)(w)\Big|^p |w_2|^{kp+2}\,dw.
$$

4. Proof of the main theorem

L^p boundedness. To finish the proof, we first need two lemmas.

Lemma 4.1. *The operator $\mathcal{B}_{1,a}$ defined as in Definition 3.5 is bounded from $W^{a,p}(\mathbb{D})$ to $L^p(\mathbb{D})$ for $p\in(1,\infty)$.*

Proof. This follows from the well-known result that the Bergman projection on $\mathbb{D}$ is bounded from $W^{k,p}(\mathbb{D})$ to itself for $p\in(1,\infty)$ and all $k\in\mathbb{Z}^+\cup\{0\}$. □

Lemma 4.2. *The integral operator $\mathcal{B}_2$ defined as in Definition 3.5 is bounded from $L^p\big(\mathbb{D}^*,|w|^{2-p}\big)$ to itself for $p\in\big(\frac{4}{3},4\big)$, where $L^p\big(\mathbb{D}^*,|w|^{2-p}\big)$ is the weighted L^p space with $w\in\mathbb{D}^*$.*

Proof. This is equivalent to the statement that the weighted Bergman projection associated to $\mathbb{D}^*$ with the weight $|w|^2$ is bounded from $L^p\big(\mathbb{D}^*,|w|^2\big)$ to itself for $p\in\big(\frac{4}{3},4\big)$. For a proof, see [Chen 2017b]. □

Proof under the additional assumption. With Lemma 4.1 and Lemma 4.2, we can prove Theorem 1.2 under the additional assumption $f(\Psi(\eta)) \in C^\infty(\bar{\mathbb{D}} - \{0\})$ in variable η_2.

Proof of Theorem 1.2 under the additional assumption. By (2-1), (2-4), (3-5) and Lemma 2.3, we obtain

$$\begin{aligned}\|\mathcal{B}(f)\|^p_{k,p,kp} &\le \sum_{|\beta|\le k}\sum_{a+b\le|\beta|}\sum_{j=0}^{b} C_{k,p}\int_{\mathbb{D}\times\mathbb{D}^*}|\mathcal{B}_{1,a}(\mathcal{B}_2(F_j))(w)|^p|w_2|^{kp+2-p(|\beta|+1)}\,dw\\ &\le C_{k,p}\sum_{a+b\le k}\int_{\mathbb{D}\times\mathbb{D}^*}|\mathcal{B}_{1,a}(\mathcal{B}_2(F_b))(w)|^p|w_2|^{2-p}\,dw.\end{aligned}$$

By Lemma 4.1, for $p \in (1,\infty)$ we have

$$\begin{aligned}\|\mathcal{B}(f)\|^p_{k,p,kp} &\le C_{k,p}\sum_{a+b\le k}\int_{\mathbb{D}^*}\biggl(\int_{\mathbb{D}}\sum_{|\beta|\le a}|D^\beta_{w_1,\bar{w}_1}(\mathcal{B}_2(F_b))(w)|^p\,dw_1\biggr)|w_2|^{2-p}\,dw_2\\ &\le C_{k,p}\sum_{|\beta|+b\le k}\int_{\mathbb{D}}\biggl(\int_{\mathbb{D}^*}|\mathcal{B}_2(D^\beta_{w_1,\bar{w}_1}(F_b))(w)|^p|w_2|^{2-p}\,dw_2\biggr)dw_1.\end{aligned}$$

Similarly, by Lemma 4.2, for $p \in \left(\frac{4}{3}, 4\right)$ we have

$$\begin{aligned}\|\mathcal{B}(f)\|^p_{k,p,kp} &\le C_{k,p}\sum_{|\beta|+b\le k}\int_{\mathbb{D}}\biggl(\int_{\mathbb{D}^*}\bigl|D^\beta_{w_1,\bar{w}_1}(F_b)(w)\bigr|^p|w_2|^{2-p}\,dw_2\biggr)dw_1\\ &= C_{k,p}\sum_{|\beta|+b\le k}\int_{\mathbb{D}\times\mathbb{D}^*}\bigl|D^\beta_{w_1,\bar{w}_1}T^b_{w_2}\bigl(f(\Psi(w))\bigr)\cdot w_2\bigr|^p|w_2|^{2-p}\,dw\\ &= C_{k,p}\sum_{|\beta|+b\le k}\int_{\mathbb{D}\times\mathbb{D}^*}\bigl|D^\beta_{w_1,\bar{w}_1}T^b_{w_2}\bigl(f(\Psi(w))\bigr)\bigr|^p|w_2|^{2}\,dw\\ &\le C_{k,p}\sum_{|\beta|+|\beta'|\le k}\int_{\mathbb{D}\times\mathbb{D}^*}\bigl|D^\beta_{w_1,\bar{w}_1}D^{\beta'}_{w_2,\bar{w}_2}\bigl(f(\Psi(w))\bigr)\bigr|^p|w_2|^{2}\,dw,\end{aligned}\tag{4-1}$$

where the last line follows from $T_{w_2} = (1/2i)(w_2(\partial/\partial w_2) - \bar{w}_2(\partial/\partial\bar{w}_2))$, $|w_2| < 1$ for $w_2 \in \mathbb{D}^*$, and a similar equation as (3-3).

By the biholomorphism $\Psi(w) = z$ defined in Section 2, we have

$$\frac{\partial}{\partial w_1} = w_2\frac{\partial}{\partial z_1} \quad\text{and}\quad \frac{\partial}{\partial\bar{w}_1} = \bar{w}_2\frac{\partial}{\partial\bar{z}_1},$$

and also

$$\frac{\partial}{\partial w_2} = w_1\frac{\partial}{\partial z_1} + \frac{\partial}{\partial z_2} \quad\text{and}\quad \frac{\partial}{\partial\bar{w}_2} = \bar{w}_1\frac{\partial}{\partial\bar{z}_1} + \frac{\partial}{\partial\bar{z}_2}.$$

Again, since $(w_1, w_2) \in \mathbb{D} \times \mathbb{D}^*$, we have $|w_1|, |w_2| < 1$. Therefore, by (4-1) and transferring $\mathbb{D} \times \mathbb{D}^*$ back to $\mathbb{H}$, we finally arrive at

$$\|\mathcal{B}(f)\|_{k,p,kp}^{p} \le C_{k,p} \sum_{|\alpha| \le k} \int_{\mathbb{H}} |D_{z,\bar{z}}^{\alpha}(f)(z)|^p \, dz. \qquad \square$$

Remove the additional assumption. To remove the additional assumption that $f(\Psi(\eta)) \in C^\infty(\bar{\mathbb{D}} - \{0\})$ in variable η_2, we need the following lemma.

Lemma 4.3. *The subspace $C^\infty(\bar{\mathbb{D}} - \{0\}) \bigcap W^{k,p}(\mathbb{D}^*, |w|^2)$ is dense in $W^{k,p}(\mathbb{D}^*, |w|^2)$ with regard to the weighted norm in $W^{k,p}(\mathbb{D}^*, |w|^2)$.*

Proof. The argument is based on [Evans 1998, §5.3 Theorem 2 and Theorem 3].

Given any $g \in W^{k,p}(\mathbb{D}^*, |w|^2)$, fix $\varepsilon > 0$. On $V_0 = \mathbb{D} - \overline{\mathbb{D}_{1/2}}$, the weighted norm $W^{k,p}(V_0, |w|^2)$ is equivalent to the unweighted norm $W^{k,p}(V_0)$. Arguing as in the proof of [Evans 1998, §5.3 Theorem 3], we see that there is a $g_0 \in C^\infty(\overline{V_0})$ such that

$$\|g_0 - g\|_{W^{k,p}(V_0, |w|^2)} < \varepsilon.$$

Define $U_j = \mathbb{D}_{\rho - 1/j} - \overline{\mathbb{D}_{1/j}}$ for some $1 > \rho > \frac{1}{2}$ and for $j \in \mathbb{Z}^+$ $(U_1 = \varnothing)$. Let $V_j = U_{j+3} - \overline{U_{j+1}}$, then we see $\bigcup_{j=1}^{\infty} V_j = \mathbb{D}_\rho - \{0\}$. Arguing as in the proof of [Evans 1998, §5.3 Theorem 2], we can find a smooth partition of unity $\{\psi_j\}_{j=1}^{\infty}$ subordinate to $\{V_j\}_{j=1}^{\infty}$, so that $\sum_{j=1}^{\infty} \psi_j = 1$ on $\mathbb{D}_\rho - \{0\}$. Moreover, for each j, the support of $\psi_j g$ lies in V_j (so $|w| > 1/(j+3)$), and hence $\psi_j g \in W^{k,p}(\mathbb{D}_\rho - \{0\})$. Therefore, we can find a smooth function g_j with support in $U_{j+4} - \overline{U_j}$ such that

$$\|g_j - \psi_j g\|_{W^{k,p}(\mathbb{D}_\rho - \{0\})} \le \frac{\varepsilon}{2^j};$$

see [Evans 1998, §5.3 Theorem 2] for details. Write $\tilde{g}_0 = \sum_{j=1}^{\infty} g_j$. It is easy to see that $\tilde{g}_0 \in C^\infty(\mathbb{D}_\rho - \{0\})$ and

$$\|\tilde{g}_0 - g\|_{W^{k,p}(\mathbb{D}_\rho - \{0\}, |w|^2)} \le \|\tilde{g}_0 - g\|_{W^{k,p}(\mathbb{D}_\rho - \{0\})} \le \varepsilon,$$

since $|w| < 1$ on $\mathbb{D}_\rho - \{0\}$.

Let V_0' be an open set such that $\partial\mathbb{D} \subset V_0'$ and $V_0' \bigcap \mathbb{D} = V_0$, then $V_0' \bigcup \mathbb{D}_\rho$ cover $\bar{\mathbb{D}}$. Take a smooth partition of unity $\{\tilde{\psi}_1, \tilde{\psi}_2\}$ on $\bar{\mathbb{D}}$ subordinate to $\{V_0', \mathbb{D}_\rho\}$. Then $h = \tilde{\psi}_1 g_0 + \tilde{\psi}_2 \tilde{g}_0$ belongs to $C^\infty(\bar{\mathbb{D}} - \{0\})$, and

$$\|h - g\|_{W^{k,p}(\mathbb{D}^*, |w|^2)} \le C\big(\|g_0 - g\|_{W^{k,p}(V_0, |w|^2)} + \|\tilde{g}_0 - g\|_{W^{k,p}(\mathbb{D}_\rho - \{0\}, |w|^2)}\big) < 2C\varepsilon$$

as desired $\square$

Now we are ready to remove the extra assumption and prove our main result.

Proof of Theorem 1.2.

For any $f \in W^{k,p}(\mathbb{H})$, we have $f(\Psi(w)) \in W^{k,p}(\mathbb{D}^*, |w_2|^2)$ in variable w_2. Then by Lemma 4.3, we can find a sequence $\{h_j(w)\} \subset C^\infty(\bar{\mathbb{D}} - \{0\})$ tending to $f(\Psi(w))$

in variable w_2 with regard to the norm in $W^{k,p}(\mathbb{D}^*, |w_2|^2)$. We have already seen that (4-1) holds for each $h_j(w)$ replacing $f(\Psi(w))$. Indeed, if we focus on the integration over $\mathbb{D}^*$, by comparing with (2-4), we see that (4-1) is just the following: for each $b = 0, 1, \dots, k$

$$\int_{\mathbb{D}^*} \left| w_2^b \frac{\partial^b}{\partial w_2^b} (\mathcal{B}_3(h_j)) \right|^p |w_2|^2 \, dw_2 \le C_{k,p} \|h_j\|_{W^{k,p}(\mathbb{D}^*, |w_2|^2)}, \tag{4-2}$$

where $\mathcal{B}_3$ is the weighted Bergman projection associated to $\mathbb{D}^*$ with the weight $|w_2|^2$.

Now letting $j \to \infty$, in view of the boundedness of $\mathcal{B}_3$ (Lemma 4.2), we see that $w_2^b(\partial^b/\partial w_2^b)(\mathcal{B}_3(h_j))$ indeed tends to $w_2^b(\partial^b/\partial w_2^b)(\mathcal{B}_3(f(\Psi)))$ in $L^p(\mathbb{D}^*, |w_2|^2)$ for each $b = 0, 1, \dots, k$. Therefore, (4-2) is valid for general $f(\Psi(w)) \in W^{k,p}(\mathbb{D}^*, |w_2|^2)$, which completes the proof for any general $f \in W^{k,p}(\mathbb{H})$. □

Remark 4.4. The method also applies to the n-dimensional generalization of the Hartogs triangle, see [Chen 2017a]. To be precise, for $j = 1, \dots, l$, let Ω_j be a bounded smooth domain in $\mathbb{C}^{m_j}$ with a biholomorphic mapping $\phi_j : \Omega_j \to \mathbb{B}^{m_j}$ between Ω_j and the unit ball $\mathbb{B}^{m_j}$ in $\mathbb{C}^{m_j}$. We use the notation $\tilde{z}_j$ to denote the j-th m_j-tuple in $z \in \mathbb{C}^{m_1+\cdots+m_l}$, that is, $z = (\tilde{z}_1, \dots, \tilde{z}_l)$. Let $n = m_1 + \cdots + m_l + n'$, $n - n' \ge 1$, and $n' \ge 1$, we define the *n-dimensional Hartogs triangle* by

$$\mathbb{H}^n_{\phi_j} = \left\{ (z, z') \in \mathbb{C}^{m_1+\cdots+m_l+n'} \;\middle|\; \max_{1\le j\le l} |\phi_j(\tilde{z}_j)| < |z'_1| < |z'_2| < \cdots < |z'_{n'}| < 1 \right\}.$$

Following the same idea, we see that the Bergman projection $\mathcal{B}$ on $\mathbb{H}^n_{\phi_j}$ is bounded from $W^{k,p}(\mathbb{H}^n_{\phi_j})$ to $W^{k,p}(\mathbb{H}^n_{\phi_j}, |z'_1|^{kp})$ for $p \in (2n/(n+1), 2n/(n-1))$. However, the weight $|z'_1|$ is no longer comparable to $|(z, z')|$, the distance from points near the boundary to the singularity at the boundary.

Acknowledgements

The content of this paper is a part of the author's Ph.D. thesis at Washington University in St. Louis; see [Chen 2015]. The author would like to thank his thesis advisor Prof. S. G. Krantz for giving him a very interesting problem to work on and lots of guidance on his research. The author also wants to thank Prof. E. J. Straube for very helpful comments and suggestions on his work.

References

[Chakrabarti and Shaw 2013] D. Chakrabarti and M.-C. Shaw, "Sobolev regularity of the $\bar{\partial}$-equation on the Hartogs triangle", *Math. Ann.* **356**:1 (2013), 241–258. MR Zbl

[Chakrabarti and Zeytuncu 2016] D. Chakrabarti and Y. E. Zeytuncu, "L^p mapping properties of the Bergman projection on the Hartogs triangle", *Proc. Amer. Math. Soc.* **144**:4 (2016), 1643–1653. MR Zbl

[Chen 2015] L. Chen, *Regularity of the Bergman projection on variants of the Hartogs triangle*, Ph.D. thesis, Washington University in St. Louis, 2015, available at http://search.proquest.com/docview/1677171005. MR

[Chen 2017a] L. Chen, "The L^p boundedness of the Bergman projection for a class of bounded Hartogs domains", *J. Math. Anal. Appl.* **448**:1 (2017), 598–610. MR

[Chen 2017b] L. Chen, "Weighted Bergman projections on the Hartogs triangle", *J. Math. Anal. Appl.* **446**:1 (2017), 546–567. MR Zbl

[Evans 1998] L. C. Evans, *Partial differential equations*, Graduate Studies in Mathematics **19**, American Mathematical Society, Providence, RI, 1998. MR Zbl

[Krantz 1992] S. G. Krantz, *Function theory of several complex variables*, 2nd ed., Wiley, New York, 1992. Reprinted Chelsea/AMS, Providence, 2001. MR Zbl

[Krantz and Peloso 2008] S. G. Krantz and M. M. Peloso, "The Bergman kernel and projection on non-smooth worm domains", *Houston J. Math.* **34**:3 (2008), 873–950. MR Zbl

[Lanzani and Stein 2004] L. Lanzani and E. M. Stein, "Szegö and Bergman projections on non-smooth planar domains", *J. Geom. Anal.* **14**:1 (2004), 63–86. MR Zbl

[Lanzani and Stein 2012] L. Lanzani and E. M. Stein, "The Bergman projection in L^p for domains with minimal smoothness", *Illinois J. Math.* **56**:1 (2012), 127–154. MR Zbl

[Straube 1986] E. J. Straube, "Exact regularity of Bergman, Szegő and Sobolev space projections in nonpseudoconvex domains", *Math. Z.* **192**:1 (1986), 117–128. MR Zbl

[Zeytuncu 2013] Y. E. Zeytuncu, "L^p regularity of weighted Bergman projections", *Trans. Amer. Math. Soc.* **365**:6 (2013), 2959–2976. MR Zbl

Received November 10, 2015. Revised October 9, 2016.

LIWEI CHEN
DEPARTMENT OF MATHEMATICS
THE OHIO STATE UNIVERSITY
COLUMBUS, OH 43210
UNITED STATES

chen.1690@osu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.319

KNOTS OF TUNNEL NUMBER ONE AND MERIDIONAL TORI

MARIO EUDAVE-MUÑOZ AND GRISSEL SANTIAGO-GONZÁLEZ

We give a characterization of knots of tunnel number 1 that admit an essential meridional torus with two boundary components. Let K be a knot in S^3, S an essential meridional torus in the exterior of K with two boundary components, and τ an unknotting tunnel for K. We consider the intersections between S and τ. If the intersection is empty, we conclude that the knot K is an iterate of a satellite knot of tunnel number 1 and one of its unknotting tunnels, and then S is knotted as a nontrivial torus knot. If the intersection is nonempty, we simplify it as much as possible, and conclude that the knot K is a $(1,1)$-knot; it follows from known results that in some cases the torus S is knotted as a nontrivial torus knot, while in others cases the torus S is unknotted.

1. Introduction

An important topic in knot theory is that of studying incompressible surfaces in the exterior of knots. We first make a summary of known results for incompressible surfaces for knots of tunnel number 1. There is a classification of satellite knots of tunnel number 1 in S^3, that is, knots that admit in their exterior an incompressible non-∂-parallel torus; this was given by K. Morimoto and M. Sakuma [1991]. Another proof of this classification was given by M. Eudave-Muñoz [1994]. All these knots are $(1,1)$-knots, that is, knots of 1 bridge with respect to a standard torus in S^3; this is a special class of knots of tunnel number 1. Gordon and Reid [1995] proved that knots of tunnel number 1 do not admit any essential planar meridional surface. Regarding surfaces of higher genus, Eudave-Muñoz [1999; 2006] showed that for any $g \geq 2$, there are infinitely many knots of tunnel number 1 whose exterior contains a closed meridionally incompressible surface of genus g, and gave a characterization of $(1,1)$-knots that admit surfaces of this kind. In [Eudave-Muñoz 2000], he showed that for each pair of integers $g \geq 1$ and $n \geq 1$, there are knots k of tunnel number 1 such that there is an essential meridional surface S in the exterior of k, of genus g, and with $2n$ boundary components. Eudave-Muñoz and

MSC2010: 57M25.
Keywords: knot of tunnel number one, $(1,1)$-knot, meridional torus, iterate knot.

E. Ramírez-Losada [2009] have given a general construction and characterization of $(1,1)$-knots that admit essential meridional surfaces.

In this paper we give a characterization of knots of tunnel number 1 that admit an essential meridional torus with two boundary components. Such knots are either $(1,1)$-knots, and then come from the construction of Eudave-Muñoz and Ramírez-Losada, or are iterates of a satellite knot of tunnel number 1 and one of its unknotting tunnels, i.e., they come from the construction of [Eudave-Muñoz 2000].

In Section 2 we give definitions and statements of the main results. In Section 3 we prove some general lemmas about unknotting tunnels, and in Section 4 we give a proof of the main results.

2. Preliminaries

Let k be a knot in S^3, and denote by $E(k)$ the exterior of k, that is, $E(k) = S^3 - \operatorname{int} N(k)$, where $N(k)$ is a tubular neighborhood of k.

Definition. Let k be a knot in S^3. A surface S properly embedded in $E(k)$ is said to be meridional if ∂S consists of a nonempty collection of meridian curves in $\partial N(k)$.

Definition. Let k be a knot in S^3 and S a surface properly embedded in $E(k)$, which is meridional or disjoint from $\partial N(k)$. We say that S is meridionally compressible in (S^3, k) if there is a disc $D \subset S^3$ such that $D \cap S = \partial D$, D intersects k transversely in one point, and ∂D is essential in S, that is, ∂D does not bound a disc in S and it is not parallel in S to a component of ∂S. The disc D is called a meridional compression disc for S. We say that S is meridionally incompressible in (S^3, k) if S is incompressible and not meridionally compressible in (S^3, k). We say that a meridional surface S is essential if it is meridionally incompressible and not ∂-parallel in $E(k)$.

A meridional surface can be seen as a closed surface $\overline{S}$ in S^3 which a knot intersects transversely in finitely many points. When we say that $\overline{S}$ is a meridional essential surface that intersects a knot k in n points, this means that the surface $S = \overline{S} \cap E(k)$ is a meridional essential surface in $E(k)$ as detailed in the two preceding definitions.

Definition. A knot k in S^3 has tunnel number 1 if there exists an arc τ embedded in S^3 with $\tau \cap k = \partial\tau$, such that $E(k \cup \tau) = S^3 - \operatorname{int} N(k \cup \tau)$ is a genus 2 handlebody. We call τ an unknotting tunnel for k.

Sometimes it is convenient to express a tunnel τ for a knot k as $\tau = \tau_1 \cup \tau_2$, where τ_1 is a simple closed curve in $E(k)$ and τ_2 is an arc in $E(k)$ connecting τ_1 and $\partial N(k)$; by sliding the tunnel we can pass from one expression to the other.

Definition. A knot k in S^3 is a $(1,1)$-knot if there is a standard torus T in S^3 such that k is a 1-bridge knot with respect to T, that is, k intersects T transversely in two points which divide k into two arcs which are parallel to arcs lying on T.

It is not difficult to see that a $(1,1)$-knot k is a knot of tunnel number 1. An unknotting tunnel for k can be seen as $\tau = \tau_1 \cup \tau_2$, where τ_1 is the core of one of the solid tori bounded by T, and τ_2 is a straight arc in that solid torus connecting k and τ_1. Conversely, we have the following result. Though it is well known, we include it for completeness.

Lemma 2.1. *If k is a knot with an unknotting tunnel $\tau = \tau_1 \cup \tau_2$, where τ_1 is a trivial knot in S^3, then k is a $(1,1)$-knot.*

Proof. Note that $E(\tau_1)$ is a solid torus. Slide k over τ_2, until it is an arc k' properly embedded in $E(\tau_1)$. The manifold $E(\tau_1) - \operatorname{int} N(k') \cong E(k \cup \tau)$ has compressible boundary, for it is a handlebody. If every compression disc for $E(\tau_1) - \operatorname{int} N(k')$ intersects a meridian of k', then the manifold obtained by adding a 2-handle along a meridian of k' would have incompressible boundary, by Jaco's addition lemma [1984]. But this is not possible, for the manifold obtained is $E(\tau_1)$, which is a solid torus. Then there is a compression disc disjoint from k'. By compressing along this disc, we get that k' is inside a 3-ball, and then it must be parallel to an arc contained in $\partial E(\tau_1)$. It follows that k can be expressed as $k = k' \cup k''$, where k' is an arc properly embedded in $E(\tau_1)$, and parallel to an arc lying on $\partial E(\tau_1)$, and k'' is an arc contained in $\partial E(\tau_1)$. It follows that k is a 1-bridge knot with respect to the torus $\partial E(\tau_1)$. $\square$

Morimoto and Sakuma's construction [1991] of satellite tunnel number 1 knots is as follows: Let $T(p,q)$ be a torus knot of type (p,q) in S^3, with $|p| \geq 2$, $q \geq 2$, and let $S(\alpha,\beta)$ be a 2-bridge link in S^3 of type (α,β), with $\alpha \geq 4$; that is, $S(\alpha,\beta)$ is neither a trivial link nor a Hopf link. Identify $\partial E(T(p,q))$ and a component of $\partial E(S(\alpha,\beta))$, in such a way that a meridian of $E(S(\alpha,\beta))$ is glued to a fiber of the Seifert fibration of $E(T(p,q))$. The result is the exterior of a satellite knot $K(\alpha,\beta\,;p,q)$ with companion a torus knot, which has tunnel number 1.

The knots $K(\alpha,\beta\,;p,q)$ can also be described in the following way; see [Eudave-Muñoz 1994]. Let T be a standard torus in S^3, and let $A_{p,q} \subset T$ be an annulus so that a component of ∂A is a curve of slope (p,q) on T, $|p| \geq 2$, $q \geq 2$. We say that a knot k belongs to the class of knots $\mathcal{T}$ if k has a 1-bridge presentation with respect to some annulus $A_{p,q}$, that is, k is a 1-bridge knot with respect to T, such that the intersection points of k with T lie in $A_{p,q}$, and the arcs of k are parallel to arcs on $A_{p,q}$. If k is in $\mathcal{T}$ then k can be isotoped to lie in $N(A_{p,q})$, for some $A_{p,q}$. Let $S_{p,q} = \partial N(A_{p,q})$. For any such k that is neither trivial nor the (p,q)-torus knot, the torus S will be essential in the exterior of k. It can be seen that k belongs to $\mathcal{T}$ if and only if it is one of the knots $K(\alpha,\beta\,;p,q)$.

Now we describe the unknotting tunnels for a knot in $\mathcal{T}$. The torus T divides S^3 into two solid tori R_1 and R_2. Let k be a knot in $\mathcal{T}$, such that k has a 1-bridge presentation with respect to an annulus $A_{p,q}$; so $k \subset N(A_{p,q})$. Then k is divided into two arcs k_1 and k_2, which are trivial arcs in R_1 and R_2, respectively. We can consider R_1 as foliated by concentric tori around the core of R_1, and then k_1 as an arc that intersects each of the tori in two or zero points, except for one torus which is tangent to k_1, defining a maximum point in k_1. Similarly we define a minimum of k_2 in R_2. By a straight arc in R_1 or R_2, we mean an arc that intersects each torus in the foliation in at most one point. Take a straight arc ρ_1 which goes from the maximum of k_1 to a point x on $S_{p,q}$. Similarly, take a straight arc ρ_2 which goes from the minimum of k_2 to a point y on $S_{p,q}$. Let ρ_3 be an arc in $S_{p,q}$ joining x and y, which crosses T in one point, and which is disjoint from a meridian of $N(A_{p,q})$. Let τ_x be the union of the core of the solid torus R_1 and a straight arc joining the point x and the core of R_1. Similarly, let τ_y be the union of the core of the solid torus R_2 and a straight arc joining the point y and the core of R_2. Note that τ_x and τ_y are unknotting tunnels for the exterior of $N(A_{p,q})$, that is, for the torus knot $T(p,q)$.

Now define $\tau(1,x) = \tau_x \cup \rho_1$, $\tau(2,x) = \tau_x \cup \rho_3 \cup \rho_2$, $\tau(2,y) = \tau_y \cup \rho_2$, $\tau(1,y) = \tau_y \cup \rho_3 \cup \rho_1$. It is not difficult to see that each of these 1-complexes is an unknotting tunnel for k. Furthermore, it follows from [Morimoto and Sakuma 1991], that if τ is an unknotting tunnel for k, then k is one of the tunnels $\tau(1,x)$, $\tau(2,x)$, $\tau(2,y)$, $\tau(1,y)$, up to homeomorphism of $E(k)$. In the same paper, all unknotting tunnels for k up to ambient isotopy of $E(k)$ are also classified. Here we only need the classification up to homeomorphism of $E(k)$, because if two tunnels are homeomorphic, though not isotopic, they will produce the same family of knots when taking iterates of the knot and the tunnels.

Definition. Let k be a knot of tunnel number 1, and τ an unknotting tunnel for k which is an embedded arc with endpoints lying on $\partial N(k)$. Let k^* be a knot formed by the union of two arcs, $k^* = \tau \cup \gamma$, such that γ is contained in $\partial N(k)$. We say that k^* is an iterate of k and τ.

The knot k^* is also a knot with tunnel number 1, where the tunnel is given by the union of k and a straight arc in $N(k)$ connecting k^* and k.

Eudave-Muñoz [2000] showed that there are knots k of tunnel number 1 for which there is an essential meridional torus S in the exterior of k, with two boundary components. These are constructed by taking iterates of satellite knots of tunnel number 1. Here, we recall this construction.

Let k be a satellite knot of tunnel number 1 in S^3. Let $\overline{S}$ be the essential torus lying in the exterior of k as defined above, so $\overline{S}$ is knotted as a torus knot. Let τ be any of the unknotting tunnels $\tau(1,x)$, $\tau(2,x)$, $\tau(2,y)$, $\tau(1,y)$ for k defined above. Note that τ can be expressed as $\tau = \tau_1 \cup \tau_2$, where τ_1 is a simple closed curve, and τ_2 is an arc with endpoints in $\partial N(k)$ and τ_1, such that τ_1 is disjoint from $\overline{S}$ and τ_2 intersects $\overline{S}$ transversely in one point. The torus $\overline{S}$ divides S^3 into two parts, denoted by M_1 and M_2, where, say, k lies in M_2.

Note that $M_2 \cap N(\tau_2)$ is a cylinder $R \cong D^2 \times I$, such that $R \cap \overline{S}$ is a disc $D_1 = D^2 \times \{1\}$, and $R \cap N(k)$ is a disc $D_0 = D^2 \times \{0\}$. Slide τ_1 over τ_2, to get an arc τ with both endpoints in $D_0 \subset \partial N(k)$, such that $\tau \cap M_2$ consists of two straight arcs contained in R, i.e., arcs which intersect each disc $D_2 \times \{x\}$ transversely in one point. The surface $\overline{S}$ and the arc τ then intersect in two points.

Let k^* be an iterate of k and τ as in the previous definition. So $k^* = \tau \cup \lambda$, where λ is contained in $\partial N(k)$. The torus $\overline{S}$ and the knot k^* then intersect in two points. Push the interior of λ into the interior of $N(k)$, such that now λ is a properly embedded arc in $N(k)$ whose endpoints lie in D_0. Recall that the wrapping number of a knot in a solid torus is defined as the minimal number of times that the knot intersects any meridional disc of such a solid torus. We define the wrapping number of the arc λ in $N(k)$ as the wrapping number of the knot obtained by joining the endpoints of λ with an arc in D_0, and then pushing it into the interior of $N(k)$. This is well defined.

Let $\mathcal{D}$ be the family of knots constructed as above and such that any one of the following conditions is satisfied:

(1) k is not a cable knot, and the wrapping number of λ in $N(k)$ is ≥ 2.

(2) Suppose k is a cable knot. Let A be the annulus spanned by k and $\overline{S}$; that is, $A \subset M_2$, then one boundary component of A is in $\partial N(k)$ and the other is a curve on $\overline{S}$. We can assume that the part of τ lying in M_2 is contained in A. Let $B = \partial N(k) \cap N(A)$; this is an annulus in $\partial N(k)$. Assume that $D_0 \subset B$. In this case we assume that the wrapping number of λ in $N(k)$ is ≥ 2, and that the arc λ cannot be isotoped, relative to D_0, to an arc lying in B.

(3) The wrapping number of λ in $N(k)$ is 1. Embed the solid torus $N(k)$ in S^3 in a standard manner. Let $\widehat{\lambda}$ be the knot obtained by joining the endpoints of λ with an arc lying in D_0. The image of this knot in S^3 is a $(1,1)$-knot, in fact a 2-bridge knot. In the present case assume that $\widehat{\lambda}$ is a nontrivial 2-bridge knot.

Note that if none of the above conditions is satisfied then the torus $\overline{S}$ will be compressible in $E(k^*)$.

Theorem 2.2. *Let k^* be a knot in the family $\mathcal{D}$. Then k^* is a knot of tunnel number 1 and $\overline{S}$ is an essential meridional torus which intersects k^* in two points.*

Proof. The knot k^* has tunnel number 1 because it is an iterate of k and τ. By construction k^* intersects $\overline{S}$ in two points. If conditions (1) or (2) are satisfied, then $\overline{S}$ is essential by [Eudave-Muñoz 2000, Theorem 2.1]. If condition (3) is satisfied, then note that k^* is also a $(1,1)$-knot, and then $\overline{S}$ is essential by [Eudave-Muñoz and Ramírez-Losada 2009]. □

Eudave-Muñoz and Ramírez-Losada [2009] have given a general construction of $(1,1)$-knots that admit essential meridional surfaces. In particular, there are three families of knots, $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, that consist of $(1,1)$-knots that admit an essential meridional torus intersecting the knot in two points. For two of these families, $\mathcal{A}$ and $\mathcal{B}$, the essential torus is knotted as a torus knot, while for the family $\mathcal{C}$, the essential torus is unknotted. Furthermore, if a $(1,1)$-knot k admits an essential torus intersecting it in two points, then k belongs to one of these families.

In this paper we prove the following:

Theorem 2.3. *Let k be a knot of tunnel number* 1, *$\overline{S}$ a meridional essential torus which intersects the knot in two points, and $\tau = \tau_1 \cup \tau_2$ an unknotting tunnel for k. Then one of the following happens*:

(1) *k is a* $(1,1)$*-knot*; *or*

(2) $\overline{S} \cap \tau = \varnothing$, *and*

  (a) *$\overline{S}$ is knotted as a nontrivial torus knot*,

  (b) *the knot τ_1 is a satellite knot of tunnel number* 1, *and*

  (c) *k is an iterate of τ_1 and of an unknotting tunnel for τ_1.*

From Theorem 2.3 and the results of [Eudave-Muñoz and Ramírez-Losada 2009] we get:

Corollary 2.4. *Let $k \subset S^3$ be a knot of tunnel number* 1, *$\overline{S}$ a meridional essential torus which intersects the knot in two points. Then k belongs to one of the families $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ or $\mathcal{D}$ defined above.*

3. Some unknotting lemmas

Let M be a compact, orientable, irreducible 3-manifold whose boundary is a torus T. Suppose τ is an unknotting tunnel for M, that is, τ is an arc properly embedded in M such that $H = M - \operatorname{int} N(\tau)$ is a genus 2 handlebody.

Proposition 3.1. *Suppose that τ has been slid* (*over T and over itself*), *in such a way that $\tau = \tau_1 \cup \tau_2$, where τ_1 is a simple closed curve in the interior of M and τ_2 is an arc joining T and τ_1. Suppose that there is no compression disc for T disjoint from τ. Then τ_1 cannot be contained in a* 3*-ball $B \subset M$.*

Proof. Suppose that τ_1 is contained in a 3-ball $B \subset M$. Let β be a curve on $\partial N(\tau)$ which is a cocore of the arc τ_2, i.e., β bounds a disc in $N(\tau)$ which intersects τ_2 in one point. There are two cases, either there is a compression disc for ∂H disjoint from β, or any compression disc intersects β.

Suppose first that D is a compression disc for ∂H disjoint from β. By isotoping D we can assume that ∂D lies in T or in $\partial N(\tau_1)$. If ∂D lies in T, then either T is compressible and there is a compression disc disjoint from τ, or there exists a disc $D' \subset T$, with $\partial D' = \partial D$, such that $D \cup D'$ bounds a 3-ball in which τ lies. If this happens then by cutting H along D we should get two solid tori, as H is a handlebody. But then M is a solid torus and there is a compression disc for T disjoint from τ. In both cases this establishes the proposition.

Hence, we may assume that ∂D lies on $\partial N(\tau_1)$. Let F be a copy of $\partial N(\tau)$ slightly pushed into the interior of H; this is a once-punctured torus properly embedded in H, whose boundary bounds a disc $D' \subset T$, which is a neighborhood of $\tau_2 \cap T$. We can assume that ∂D lies on F, and then by cutting F along D, we get a disc D'' with $\partial D'' = \partial F = \partial D'$. Note that $D'' \cup D'$ must bound a 3-ball in which τ lies. As before, this shows that there is a compression disc for T disjoint from τ.

Suppose now that any compression disc for ∂H must intersect β. By [Jaco 1984] or [Casson and Gordon 1987], it follows that by adding a 2-handle along β, we get an irreducible manifold with incompressible boundary. This is a contradiction, for what we get is $M - \operatorname{int} N(\tau_1)$, which is reducible, for we are assuming that τ_1 lies inside a 3-ball. This completes the proof. □

The next proposition is somehow natural, but it is not so easy to prove because of certain phenomena. If t_1 is a properly embedded arc in a product $T \times I$, where T is a torus, and $T \times I - \operatorname{int} N(t_1)$ is a handlebody, then by a result of Frohman [1989], t_1 is isotopic to a straight arc in $T \times I$. But if t_1 and t_2 are a pair of arcs properly embedded in $T \times I$ such that $T \times I - \operatorname{int} N(t_1 \cup t_2)$ is a handlebody, then t_1 and t_2 may not be straight arcs simultaneously in $T \times I$. Now, let t_1 be an arc properly embedded in $A \times I$, with endpoints in $A \times \{0\}$ and $A \times \{1\}$, where A is an annulus, such that $A \times I - \operatorname{int} N(t_1)$ is a handlebody. Then by Jaco's addition lemma [1984], t_1 is parallel to an arc lying in $\partial(A \times I)$, but it may not be a straight arc in $A \times I$.

Proposition 3.2. *Let M, T and τ be as above, and assume that T is incompressible. Let T' be a torus embedded in M which is parallel to T; that is, T and T' cobound a region homeomorphic to $T \times I$. Suppose that τ intersects T' in two points. Then $(T \times I) \cap \tau$ consists of two straight arcs in $T \times I$, that is, τ can be isotoped, without intersecting T' in more points, such that $(T \times I, (T \times I) \cap \tau) = (T \times I, \{x, y\} \times I)$, where $x, y \in T$.*

Proof. Let $M' = M - \operatorname{int} T \times I$. Then $M = M' \cup (T \times I)$, where $\partial M' = M' \cap (T \times I) = T'$. We have that $H = M - \operatorname{int} N(\tau)$ is a genus 2 handlebody. Note that the arc τ cannot be isotoped to be disjoint from T', for otherwise T' would be an incompressible torus in the handlebody H, which is not possible. Suppose that τ is divided into 3 arcs $\tau = k_1 \cup k_m \cup k_2$, such that $k_1, k_2 \subset T \times I$, and $k_m \subset M'$. Let $\widetilde{T}' = T' \cap H = T' - \operatorname{int} N(k_1) \cup N(k_2)$; this is a twice punctured torus properly embedded in H. Note that $\widetilde{T}'$ is incompressible in H, for otherwise T' would be compressible in M, or the arc τ could be isotoped to be disjoint from T'.

Let D be a compression disc for H. Assume that D and $\widetilde{T}'$ intersect transversely and that this intersection is minimal. Label the endpoints of the arcs of intersection between D and $\widetilde{T}'$ with 1 or 2, depending on whether the endpoint lies in $\partial N(k_1) \cap \widetilde{T}'$ or in $\partial N(k_2) \cap \widetilde{T}'$. Let γ be an outermost arc of intersection in D, then it bounds a disc $D' \subset D$, with $\partial D' = \alpha \cup \gamma$, where α is an arc on ∂H, and the interior of D' is disjoint from $\widetilde{T}'$.

There are several possibilities for the endpoints of the arc γ.

(1) The endpoints of γ are labeled 1 and 2, and α lies in $\partial N(k_m)$. This implies that the arc k_m is isotopic to an arc over T', and then τ can be isotoped to be disjoint from T', which is not possible.

(2) The endpoints of γ are labeled 1 and 2 and D' lies in $T \times I$. Then α is an arc that goes over $N(k_1)$, then over T, and then over $N(k_2)$. This shows that k_1 and k_2 is a pair of parallel arcs in $T \times I$. As H is a handlebody, by cutting H along the incompressible surface $\widetilde{T}'$ we get a pair of handlebodies; one of these is just $T \times I - \operatorname{int} N(k_1 \cup k_2)$. Note that the disc D' is properly embedded in $T \times I - \operatorname{int} N(k_1 \cup k_2)$. Then by cutting this handlebody with D' we get another handlebody, which is homeomorphic to $T \times I - \operatorname{int} N(k_1)$, for k_1 and k_2 are parallel in $T \times I$. This shows that $T \times I - \operatorname{int} N(k_1)$ is a handlebody, and then by a result of Frohman [1989], k_1 is isotopic to a straight arc in $T \times I$. As k_1 and k_2 are parallel, it follows that both are simultaneously straight in $T \times I$.

(3) The endpoints of γ are labeled 1 and 1 (or 2 and 2), and D' lies in $T \times I$. Then α is an arc that goes over $N(k_1)$, then over T, and then again over $N(k_1)$. The arc α cuts $\partial N(k_1)$ into two discs, let F be either of them. Then $D' \cup F$ is an annulus, in $T \times I$, with one boundary component in T and the other in T', and we can assume that k_1 is a spanning arc of the annulus. If γ is a trivial arc in $\widetilde{T}'$, then it bounds a disc $D'' \subset \widetilde{T}'$, such that k_2' intersects D''. But this would imply that k_2 is an arc parallel to k_1, and then, as in the previous case, k_1 and k_2 are simultaneously straight in $T \times I$. Therefore we can assume that the annulus $D' \cup F$ has to be isotopic to an annulus of the form $\delta \times I$, where δ is an essential simple closed curve in T. This shows that k_1 is a straight arc in $T \times I$.

If there is another outermost arc in D with endpoints labeled 2 and 2, then k_2 would also be a straight arc in $T \times I$, and because there would be two disjoint annuli containing k_1 and k_2 respectively, it would follow that both arcs are simultaneously straight arcs in $T \times I$. We can assume then that all outermost arcs in D have endpoints labeled 1 and 1, for otherwise we have finished.

Note that there is a pair of parallel arcs in D, one outermost with endpoints labeled 1 and 1, and one next to it with endpoints labeled 2 and 2. This is because

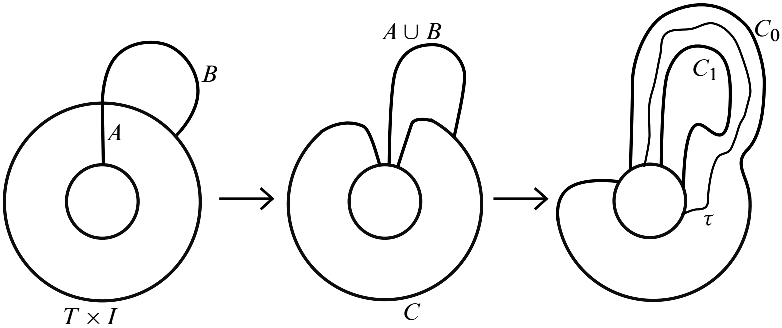


Figure 1. Constructing annuli disjoint from τ.

any outermost arc has endpoints labeled 1 and 1, and next to any label 1 there is a label 2. So assume that there are two arcs γ_1 and γ_2 in D, where γ_1 determines a disc D', with $\partial D' = \alpha \cup \gamma_1$, where α is an arc that goes over $N(k_1)$, then over T, and then again over $N(k_1)$. Then γ_1 is an arc on $\widetilde{T}'$ which goes from $N(k_1)$ to $N(k_1)$, and γ_2 is an arc on $\widetilde{T}'$ which goes from $N(k_2)$ to $N(k_2)$. The arcs γ_1 and γ_2 determine a disc $D'' \subset D$, such that $\partial D'' = \gamma_1 \cup \beta_1 \cup \gamma_2 \cup \beta_2$, where $\beta_1,\ \beta_2$ are arcs on $\partial N(k_m)$. The arc α cuts $\partial N(k_1)$ into two discs, let F be either of them. Then $D' \cup F$ is an annulus A, in $T \times I$, with one boundary component in T and the other in T'. Isotope A in $T \times I$ such that the arc k_1 is a spanning arc of A. The arcs $\beta_1,\ \beta_2$ cut $\partial N(k_m)$ into two discs, let F' be either of them. Then $D'' \cup F'$ is an annulus B, properly embedded in M'. Isotope B in M' such the arc k_m is a spanning arc of B. The annulus B is then incompressible and ∂-incompressible, for otherwise T' would be compressible or the arc k_m would be isotopic to an arc on T'. We can assume that A and B have a boundary component in common; then $A \cup B$ is an annulus, one of its boundaries components lies in T and the other in T'.

Take a product neighborhood $A \times I$ of A, where A is identified with $A \times \{\frac{1}{2}\}$. Consider the annulus $C = (T' - \partial A \times I) \cup (A \times \{0\}) \cup (A \times \{1\})$; note that C is properly embedded in M, it is ∂-parallel in M and it intersects τ in one point. Note that $A \cup B$ and C intersect in a simple closed curve, namely, the boundary component of $A \cup B$ lying in T'. Now take a product neighborhood $(A \cup B) \times I$ of $A \cup B$, where $A \cup B$ is identified with $(A \cup B) \times \{\frac{1}{2}\}$, which intersects C only in a neighborhood of the curve $(A \cup B) \cap C$. Consider the pair of annuli $C_0 \cup C_1 = (C - \partial(A \cup B) \times I) \cup ((A \cup B) \times \{0\}) \cup ((A \cup B) \times \{1\})$. Note that C_0 and C_1 are in fact a pair of annuli properly embedded in M, which are parallel in M, i.e., they cobound a product region $C_0 \times I$, where $C_0 = C_0 \times \{0\}$ and $C_1 = C_0 \times \{1\}$, such that τ is disjoint from C_0 and C_1, but it lies inside the product region $C_0 \times I$ (see Figure 1). Note that C_0 and C_1 are incompressible and ∂-incompressible in M, for these are just extensions of B via $T \times I$ to M. Then C_0 and C_1 are incompressible annuli in H, but they are ∂-compressible in H, for H is a handlebody. Then there is a disc E in H, such that $\partial E = \rho_0 \cup \rho_1$, where ρ_0 is a spanning arc of C_0, say, and ρ_1 lies on ∂H, and furthermore, $E \cap C_0 = \rho_0$, $E \cap C_1 = \varnothing$. The disc E must lie in $H' = C_0 \times I - \operatorname{int} N(\tau)$, for otherwise C_0 would be ∂-compressible in M. Note that H' is a handlebody, for it is one of the components obtained by cutting H along $C_0 \cup C_1$.

Take two parallel copies of E, and join them by the disc $\overline{C_0 - N(E)}$, and push the interior of the resulting disc into the interior of H'. We get a disc E', properly embedded in H', whose boundary is disjoint from $C_0 \cup C_1$. Note that $\partial E'$ is a nontrivial curve in $\partial H'$. Let ξ_0 and ξ_1 be the cores of the annuli C_0 and C_1 respectively. Note that E' is a compression disc for $\partial H' - \xi_0 \cup \xi_1$. Let J be a cocore of τ, that is, a curve in $\partial N(\tau)$ which bounds a disc in $N(\tau)$ intersecting τ in one point. Note that $C_0 \times I$ is the manifold obtained by attaching a 2-handle to H' along J. Then there is a compression disc E'' in $C_0 \times I$ which intersects $\xi_0 \cup \xi_1$ in two points, i.e., E'' is a 2-compression disc for $\partial(C_0 \times I)$ with respect to $\xi_0 \cup \xi_1$, as defined in [Wu 1992]. Then by Theorem 1 of that paper, there is a compression disc G for H' disjoint from J, which intersects $\xi_0 \cup \xi_1$ in at most two points. As ∂G is disjoint from J, we can assume that ∂G lies in $\partial(C_0 \times I)$. There are two possibilities for G, either ∂G is a meridian of $C_0 \times I$ intersecting each ξ_i once, or ∂G is a trivial curve in $\partial(C_0 \times I)$ intersecting ξ_0 or ξ_1 twice. In the latter case ∂G bounds a disc G' in $\partial(C_0 \times I)$ such that $G \cup G'$ is a sphere bounding a 3-ball which must contain τ. Then there is another 2-compression disc for $C_0 \times I$ which is a meridian of $C_0 \times I$. In any case, it follows that there is a meridian disc G of $C_0 \times I$, disjoint from $N(\tau)$. By cutting H' along this disc we get a solid torus. But by cutting $C_0 \times I$ along G, we get a 3-ball containing τ; it follows that τ is an unknotted arc in the 3-ball. This shows that τ is an arc parallel to an arc on C_0, and then that k_1 and k_2 are parallel straight arcs in $T \times I$. □

4. Main proofs

In this section we give a proof of Theorem 2.3.

Proposition 4.1. *Let k be a knot of tunnel number* 1, *$\overline{S}$ a meridional essential torus which intersects the knot in two points, and $\tau = \tau_1 \cup \tau_2$ an unknotting tunnel for k, where τ_1 is a simple closed curve and τ_2 is an arc connecting τ_1 and $\partial N(k)$. Suppose that $\overline{S}$ and τ cannot be made disjoint. Then one of the following happens*:

(1) *τ_1 is a trivial knot*; *or*

(2) *there is a meridional essential torus $\overline{S}'$ which intersects the knot in two points, such that $\overline{S}' \cap \tau = \varnothing$, and such that $\overline{S}'$ bounds a solid torus with τ_1 as its core.*

Let k be a knot of tunnel number 1 and $\overline{S}$ a meridional essential torus which intersects k in two points. So $S = \overline{S} \cap E(k)$ is a meridional essential surface in

$E(k)$ whose boundary consists of two meridians of k. Let τ be an unknotting tunnel for k, so that τ may have been slid over itself so that it can be expressed as $\tau = \tau_1 \cup \tau_2$, where τ_1 is a simple closed curve and τ_2 is an arc connecting τ_1 and $\partial N(k)$. Let ν be the intersection point between τ_1 and τ_2.

Assume that S has been isotoped such that it intersects τ transversely in a finite number of points, say, S meets τ_1 in n points and τ_2 in m points, $n+m>0$, and that this intersection is minimal.

Denote by $\alpha_1, \alpha_2, \dots, \alpha_n$ the discs of intersection between S and $N(\tau_1)$, numbered in order along τ_1 as they intersect S, starting at ν with an arbitrary choice of direction. Denote by $\beta_1, \beta_2, \dots, \beta_m$ the discs of intersection between S and $N(\tau_2)$, numbered in order along τ_2 as they intersect S, starting at ν, going from ν to $\partial N(k)$. Denote by s_1 and s_2 the boundary components of S (or rather, the discs of intersection of $N(k)$ with $\overline{S}$).

Let $M = S^3 - \operatorname{int} N(k \cup \tau)$, so M is a genus 2 handlebody. Let $\widetilde{S} = S \cap M$.

Lemma 4.2. *$\widetilde{S}$ is incompressible in M.*

Proof. Suppose that $\widetilde{S}$ is compressible. Then there exists a compression disc E for $\widetilde{S}$, so that ∂E bounds a disc E' in S, because S is incompressible in $E(k)$, and E' must intersect τ. If we exchange E' by E we obtain a surface S' isotopic to S but with fewer intersections with τ, which cannot happen because the intersection of S and τ is minimal. $\square$

Let D be a compression disc of M. Assume D has been isotoped to intersect $\widetilde{S}$ transversely and that it has minimal intersection with $\widetilde{S}$ among all compression discs for M. If $D \cap \widetilde{S}$ contains a simple closed curve, an innermost disc argument can eliminate it, for $\widetilde{S}$ is incompressible. So we may assume that $D \cap \widetilde{S}$ consists of a collection of arcs. Note that any such arc of intersection is not ∂-parallel in $\widetilde{S}$, for otherwise, if an arc δ in $\widetilde{S}$ is ∂-parallel, then by cutting D with the disc determined by δ in $\widetilde{S}$, we get a compression disc of M with fewer intersections with $\widetilde{S}$, a contradiction.

Label the endpoints of the arcs of intersection in D with the labels of the discs of $S \cap N(\tau)$ or the component of ∂S in which the points lie. Parts of the proof of the following lemma are similar to that of Proposition 2.3 in [Eudave-Muñoz 1994]; we include here a proof for completeness.

Lemma 4.3. *The number n is* 0. *Further, if δ is any arc of intersection between D and $\widetilde{S}$, which is outermost in D, then both ends of δ have labels β_1, and the arc γ of ∂D determined by such an outermost arc wraps at least once around $N(\tau_1)$.*

Proof. Let δ be an outermost arc on D. Then δ cuts a disc $D' \subset D$ with $D' \cap \widetilde{S} = \delta$ and $\partial D' = \delta \cup \gamma$, where γ is an arc on $\partial N(k \cup \tau)$.

There are several possible cases for δ:

Case 1. One endpoint of δ has label α_i, and the other $\alpha_{i+1},\ 1 \leq i < n$ (or β_j and $\beta_{j+1},\ 1 \leq j < m$), and γ is disjoint from $N(\nu)$ and from $\partial N(k)$.

In this case the surface S can be pushed along D' to eliminate α_i and α_{i+1}.

Case 2. One endpoint of δ has label α_1, and the other $\alpha_n,\ n \neq 1$.

Suppose that γ meets either $\partial N(k)$ or $N(\nu)$, for otherwise this would be a special case of Case 1, when $n = 2$. If $m \neq 0$, push S along D'. With this move α_1 and α_n convert into a new β_1, reducing $m + n$.

If $m = 0$ and γ does not meet $\partial N(k)$, then push S as before, creating a new β_1. If γ meets $\partial N(k)$, slide τ_1 over τ_2, then slide τ_1 over $\partial N(k)$ and then again slide over τ_2 following γ, without introducing new intersections with S. So D' is transformed into a disc as in the previous case, where $m = 0$ and $\partial D' \cap \partial N(k) = \varnothing$.

Case 3. Both endpoints of δ are labeled α_1 (or both are labeled α_n).

Note that both endpoints of γ are in the same side of α_1, since $\widetilde{S}$ is a 2-sided surface. Suppose first that $m \neq 0$. We can isotope γ to be completely contained in $N(\tau_1)$. If γ does not meet $N(\nu)$, then the intersection between ∂D and $\widetilde{S}$ is not minimal.

If γ meets $N(\nu)$, then we find a disc E in $N(\tau_1 \cup \tau_2)$ such that E meets τ_1 once and does not intersect τ_2, and $\partial E = \gamma \cup \alpha_1'$, where α_1' is a subarc of $\partial\alpha_1$. Let $E' = E \cup D'$, then $E' \cap S = \partial E' = \delta \cup \alpha_1'$. As E' is contained in $E(k)$ and S is incompressible, $\partial E'$ bounds a disc E'' in S. There are two cases, depending whether α_1 is contained in E'' or it is not. In any case, there must be at least one intersection of τ with E'', other than α_1, for otherwise the arc δ in $\widetilde{S}$ would be ∂-parallel. By exchanging E' by E'' we obtain a surface S' isotopic to S. Suppose first that the disc α_1 is not contained in E''. As E' intersects τ once, and E'' intersects τ at least once, the new surface has at most as many intersections with τ as S. Note that $S' \cap N(\tau)$ contains the disc $E \cup \alpha_1$, which intersects τ in two points. Then by isotoping S', the disc $E \cup \alpha_1$ becomes a new β_1, intersecting τ just once. Then S' has fewer intersections with τ than S, which is a contradiction. Suppose now that the disc α_1 is contained in E''. In this case, E'' intersects τ in at least two points, and E' intersects τ just once. So, S' has fewer intersections with τ than S. Note that in this case the intersection of S' with $N(\tau)$ contains the disc E. So, we are eliminating α_1 and some other α_i or β_j, and getting a new α_n.

Suppose now that $m = 0$. If we can isotope γ such that it is contained in $\partial N(\tau)$, then the proof is identical to the previous case. In the other case, a subarc of γ is contained in $\partial N(k)$ and does not intersect ∂S. Slide τ_1 over τ_2, such that τ is a properly embedded arc in $E(k)$. This can be done following γ such that no new intersections between $\widetilde{S}$ and D are created. There is a disc E contained in $N(k \cup \tau)$, $\partial E = \gamma \cup \alpha_1$, where α_1' is a subarc of $\partial\alpha_1$, and such that E meets k

once. Let $E' = E \cup D'$, then $E' \cap S = \partial E' = \delta \cup \alpha_1$. Since E' meets k once and S is meridionally incompressible, $\partial E'$ bounds a disc F in $\overline{S}$ which intersects k in one point, say it intersects $N(k)$ in s_1. Then $E' \cup F$ is a sphere which bounds a ball that intersects k in an unknotted spanning arc, for k is a prime knot. Let $\overline{S}' = (\overline{S} - F) \cup E'$; this is a surface intersecting k in two points, so that the corresponding meridional surface $S' = \overline{S}' \cap E(k)$ is isotopic to S. By slightly isotoping the tunnel τ, we see that S' has fewer intersections with τ than S, since at least we eliminated α_1, which is a contradiction.

Case 4. Both endpoints of δ are labeled β_m (and if $m = 1$, suppose that γ is on the side of β_1 closest to $\partial N(k)$).

If γ can be isotoped on ∂M such that it is contained in $\partial N(\tau)$, then the intersection between ∂D and $\widetilde{S}$ is not minimal. Otherwise, a subarc of γ is contained in $\partial N(k)$ and does not meet ∂S. Now the proof is identical to that of Case 3 when $m = 0$, with β_m in place of α_1.

Case 5. One endpoint of δ is labeled α_1, α_n or β_m, and the other s_i, $i = 1, 2$.

Suppose first that one endpoint of δ is labeled α_1 (or α_n); note that in this case $m = 0$. Slide τ_1 over τ_2, following γ, without introducing new intersections between S and τ, until τ is an arc properly embedded in $E(k)$. Now pushing τ along D', the disc of intersection α_1 is eliminated. If one endpoint of δ is labeled β_m, then push τ as before to eliminate β_m.

Case 6. One endpoint of δ is labeled s_1, and the other s_2.

As $m + n \neq 0$, γ can be made disjoint from $\partial N(\tau)$, by sliding τ if necessary. This implies that S is ∂-compressible, a contradiction.

Case 7. Both endpoints of δ are labeled s_i, $i = 1, 2$.

Again, we can assume that γ does not intersect $\partial N(\tau)$. As S is ∂-incompressible, δ cuts a disc E from S, which may contain some α_i or β_j. Note that $\partial E = \delta \cup s_i'$, where s_i' is a subarc of s_i. Then $D' \cup E$, glued along δ, is a disc whose boundary is in $N(k)$, and because $\partial N(k)$ is incompressible in $E(k)$, it bounds a disc E' in $\partial N(k)$. Note that E' must intersect τ, for otherwise D can be isotoped along E', to reduce the number of intersections between ∂D and S, which is not possible. So, $D' \cup E \cup E'$ bounds a 3-ball in $E(k)$. As τ intersects E', it must also intersect E in at least one point. Now exchange E with D', to get an essential surface S' isotopic to S in $E(k)$, with fewer intersections with τ. Note that one boundary component of S' is $\gamma \cup s_i''$, where s_i'' is the other subarc of s_i, and that, in fact, $\gamma \cup s_i''$ is a meridian of $N(k)$.

Case 8. One endpoint of δ is labeled β_1, and the other α_1 (or α_n).

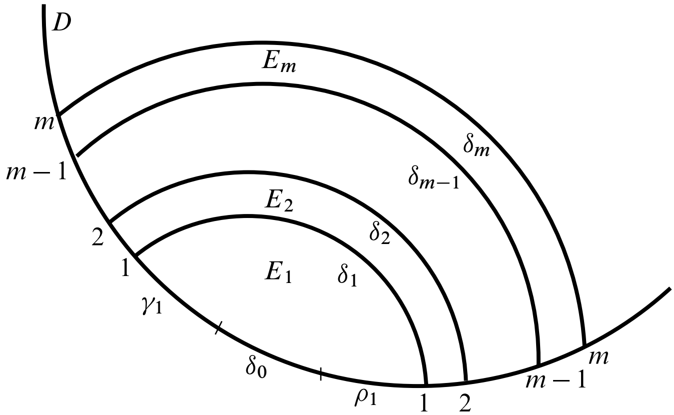


Figure 2. Outermost arcs in D.

Pushing S along D', α_1 and β_1 convert into a curve parallel to α_n, and this reduces $n+m$.

Case 9. Both endpoints of δ are labeled β_1, and the arc δ can be isotoped into $N(\tau_2) \cup N(\nu)$.

If γ is disjoint from $N(\nu)$, then the intersection between ∂D and $\widetilde{S}$ is not minimal. If γ meets $N(\nu)$, then it can be arranged such that γ intersects $N(\tau_2)$ in two arcs. There exists a disc E contained in $N(\tau)$ such that $\partial E = \gamma \cup \beta_1'$, where β_1' is a subarc of β_1. Let $E' = D' \cup E$, then $\partial E' = \delta \cup \beta_1'$ is contained in S, and because of the incompressibility of S, it bounds a disc D'' in S. We can choose the discs E and D'' such that τ_2 meets D'' in a point corresponding to β_1, and τ intersects E' once. The disc D'' necessarily intersects τ in more points, for otherwise the arc δ would be ∂-parallel in $\widetilde{S}$. Exchanging D'' with E' we get a surface S' isotopic to S, with $m' + n' < m + n$.

With this we have already considered all the possible cases for the arc δ, except if the ends of δ are in β_1 and the arc γ cannot be isotoped to $N(\tau_2) \cup N(\nu)$, i.e., γ is wrapped one or more times around $N(\tau_1)$, but this is possible only if $n = 0$, that is, S intersects τ only in the arc τ_2. □

Lemma 4.4. *There is a collection of m arcs, say $\delta_1, \delta_2, \ldots, \delta_m$ in $D \cap \widetilde{S}$, which are parallel in D and δ_1 is an outermost arc in D.*

Proof. $D \cap \widetilde{S}$ consists of a collection of arcs in D. We construct a tree in D as follows: assign a vertex for each region of $D - \widetilde{S}$, then connect two vertices if their respective regions are adjacent, that is, they have an arc of $D \cap \widetilde{S}$ common. The resultant graph G is a tree, because D is a disc. The ends of the tree, (that is, the vertices of degree 1), correspond to the outermost regions of D.

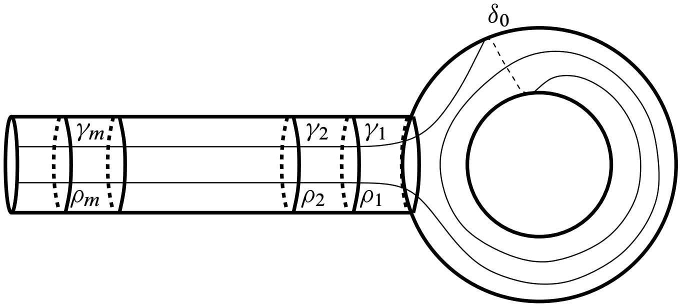


Figure 3. Curves in $\partial N(\tau)$.

A branch of G is a trajectory that begins at an end of G and finishes in a vertex of degree > 2, such that the intermediate vertices of the branch are all of degree 2. If all the vertices of G are of degree 1 or 2, then all the arcs are parallel, and there are at least $2m$ such arcs. Otherwise, let G' be the graph obtained by eliminating the branches, that is, by clearing the vertices of degree 1 and 2 of branches and the corresponding edges. Let V be a vertex of degree 1 of G' (if vertices of degree 1 do not exist, let V be the unique vertex of G'). Then at least two branches arrive at V, say r_1 and r_2 are two adjacent branches that arrive at V. Let η_1 and η_2 be the outermost arcs corresponding to r_1 and r_2, respectively. The endpoints of η_1 and η_2 are labeled β_1 and β_1, by Lemma 4.3. Let ϕ be an arc of ∂D that goes from one endpoint of η_1 to one endpoint of η_2. Then ϕ must cross labels $\beta_1, \beta_2, \dots, \beta_m, \beta_m, \beta_{m-1}, \dots, \beta_2, \beta_1$, and perhaps more labels between β_m and β_m. Any arc of intersection that leaves these labels corresponds to an edge of r_1 or r_2, by the selection of the branches. This implies that $r_1 \cup r_2$ has at least $2m$ edges, and then at least one of the branches has m or more edges corresponding to m parallel arcs. □

Label with i the endpoints of δ_i for $1 \le i \le m$. Call $E_1 \subset D$ the disc determined by δ_1. Let β_0 be a disc in $N(\tau)$ which intersects τ just in the point ν, such that $\partial\beta_0$ is a curve on $\partial N(\tau_2)$ parallel to $\partial\beta_1$. $E_1 \cap \partial N(\tau)$ can be isotoped so that it intersects β_0 in two points which divide $E_1 \cap \partial N(\tau)$ into three arcs, say γ_1, ρ_1 and δ_0, where γ_1, ρ_1 are in $\partial N(\tau_2)$ and δ_0 is in $\partial N(\tau_1)$ (see Figure 2).

Denote by γ_i and ρ_i the arcs in ∂D with endpoints $i-1$ and i. Call $E_i \subset D$ the disc determined by $\delta_i, \delta_{i-1}, \rho_i$ and γ_i, for $2 \le i \le m$. The arcs γ_i and ρ_i are contained in $\partial N(\tau_2)$ and decompose $\partial\beta_i$ into two arcs, call them β_i^1 and β_i^2, for $0 \le i \le m$. Note that $\beta_i^1, \beta_{i-1}^1, \gamma_i$ and ρ_i, for $1 \le i \le m$, determine a disc in $\partial N(\tau_2)$, call it C_i, and $\beta_i^2, \beta_{i-1}^2, \gamma_i$ and ρ_i also determine a disc, call it C_i' (see Figure 3).

Lemma 4.5. *There is an annulus A with interior disjoint from S, such that one of the boundary components is $\delta_1 \cup \beta_1^1 \subset S$, and the other is $\delta_0 \cup \beta_0^1 \subset \partial N(\tau_1)$ with some slope p/q, where $q \ge 2$.*

Proof. Note that $E_1 \cup C_1$ is an annulus A, where one of its boundary components is $\delta_1 \cup \beta_1^1 \subset S$, and the other boundary component is $\delta_0 \cup \beta_0^1$, which is contained in $\partial N(\tau_1)$, with some slope p/q. If $q = 1$, that is, $\delta_0 \cup \beta_0^1$ only turns once around $N(\tau_1)$, then τ_1 is isotopic to $\delta_1 \cup \beta_1^1$ on S, so we can push the tunnel through S, using the annulus A, eliminating one intersection with S corresponding to β_1. Thus $q \geq 2$. □

Since $\overline{S}$ is a torus in S^3, it is boundary of a solid torus R. We have two cases, depending whether τ_1 is contained in R or not.

Case 1. Suppose that τ_1 is not contained in R.

In this case the interior of the annulus A is disjoint from R. One boundary component of A lies in $\partial N(\tau)$, and the other in $\partial R = \overline{S}$.

Lemma 4.6. *The core of R is a cable around τ_1 and ∂A is a longitude of R, or the core of R and τ_1 form a Hopf link.*

Proof. The component of ∂A in $N(\tau_1)$ is a curve with slope p/q and $q \geq 2$ by Lemma 4.5. If the component of ∂A in R is a curve with slope r/s and $s \geq 2$, then the unique possibility is that τ_1 and the core of R form a Hopf link, by [Eudave-Muñoz and Uchida 1996, Theorem 1(iv)]. Otherwise, the slope of ∂A in R is longitudinal, in which case the core of R is a cable around τ_1. □

If the core of R and τ_1 form a Hopf link, then τ_1 is a trivial knot and we are done. So, we suppose now that the core of R is a cable around τ_1 and ∂A is a longitude of R.

Lemma 4.7. *The number of points of intersection, m, is* 1.

Proof. Suppose that $m \geq 2$, and consider the annulus $F = E_2 \cup C_2$, where E_2 and C_2 are glued along γ_2 and ρ_2, with its boundary lying on S. We have that $F \subset R$, and ∂F consists of two longitudes of R, so one of these boundary components is $\delta_1 \cup \beta_1$, which is contained in ∂A. The annulus F divides R into two solid tori, only one of which intersects the knot, and we can push the arc τ_2 along the other solid torus to eliminate at least two intersections with it, which is a contradiction. □

Suppose then that $S \cap \tau_2$ is one point. Let $N(A)$ be a neighborhood of A such that $z_1 = N(A) \cap R$ is a neighborhood of $\delta_1 \cup \beta_1^1$ in S, and $N(A) \cap N(\tau_1)$ is a neighborhood of $\delta_0 \cup \beta_0^1$ in $\partial N(\tau_1)$. We can assume that $N(A)$ and k are disjoint.

Let $W = R \cup N(A) \cup N(\tau_1)$. Then W is a solid torus and τ_1 is a core of W. Let $T_1 = \partial W$. The surface T_1 is a torus which intersects k in two points.

Lemma 4.8. *Either the punctured surface $T_1 - k$ is incompressible in $S^3 - k$, or τ_1 is a trivial knot.*

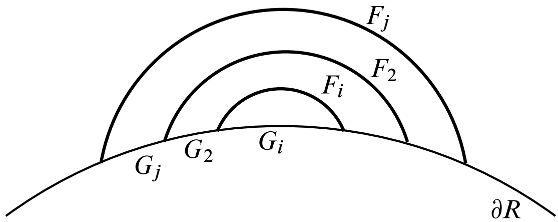


Figure 4. Constructing parallel annuli.

Proof. We prove first that $T_1 - k$ is incompressible in $W - k$. Note that z_1 is an annulus properly embedded in W, with slope p/q, which does not meet k. Suppose Q is a compression disc for $T_1 - k$. Then $Q \cap z_1$ consists of simple closed curves and arcs, and the simple closed curves can be eliminated, because z_1 is essential in W. Now we take an outermost arc η in Q. If η is trivial in z_1, then we can isotope Q to eliminate intersections with this annulus. If η is essential in z_1, then the outermost disc determined by z_1 in Q is contained in R, since $q \geq 2$. This implies that S is compressible in $R - k$, which is not possible.

If $T_1 - k$ is compressible in $S^3 - \operatorname{int} W$, we have two cases, either the boundary of a compression disc Q is essential in the torus T_1, or is trivial in that torus. If the curve ∂Q is essential in T_1, we have that the solid torus W is unknotted and then τ_1 is a trivial knot.

If the curve ∂Q is trivial in T_1, then it bounds a disc $Q' \subset T_1$, which meets k in two points. If W is unknotted, then τ_1 is a trivial knot. Suppose that W is knotted; exchanging Q' for Q, we have a bigger torus T_1', parallel to T_1, which does not touch the knot. The torus T_1' is incompressible in $S^3 - \operatorname{int} N(k \cup \tau)$, since it bounds a knotted solid torus and τ_1 is a core of W, but this cannot happen because $S^3 - \operatorname{int} N(k \cup \tau)$ is a handlebody. □

In this case we concluded that either τ_1 is a trivial knot, or that there is another meridional essential torus which intersects k in two points that is disjoint from τ, and such that τ_1 is a core of the solid torus bounded by T_1.

Case 2. Suppose that τ_1 is contained in R. In this case τ_1 is a core of R.

Lemma 4.9. *Either $m = 1$, or τ_1 is a trivial knot.*

Proof. Suppose that $m \geq 2$. Let F_2 be defined as before, $F_2 = E_2 \cup C_2$, where E_2 and C_2 are glued along γ_2 and ρ_2, with its boundary lying on S. Now F_2 is not contained in R. Note that ∂F_2 consists of two curves in ∂R, with slope p/q and $q \geq 2$. That is, F_2 is an annulus in the exterior of R, and F_2 is parallel to an annulus $G_2 \subset \partial R$, since the slope of its boundary is not integral. If k is not in the region bounded by $F_2 \cup G_2$, we can eliminate two intersections with τ, by pushing τ_2 through the solid torus with boundary $F_2 \cup G_2$. Suppose then that k is in such a

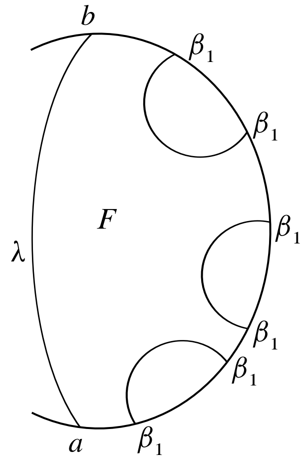


Figure 5. Outermost arcs in D when $m = 1$.

region. Consider any other of the annuli F_i defined as before, $F_i = E_i \cup C_i$, where E_i and C_i are glued along γ_i and ρ_i, with its boundary lying on S. Suppose that F_i is not contained in R. Again, F_i is parallel to an annulus $G_i \subset \partial R$ and k must be contained in the region bounded by $F_i \cup G_i$. This shows that F_2 and F_i must be parallel (see Figure 4).

Let F_j be the annulus not contained in R, bounding a maximal parallel region between F_j and G_j. Let $T = (\partial R - G_j) \cup F_j$. By slightly pushing T, we have that $T \cap \tau = \varnothing$, and $T \cap k = \varnothing$. The torus T bounds a solid torus R' with τ_1 as its core. If τ_1 is not the trivial knot, then T is incompressible in $S^3 - N(k \cup \tau)$, which is not possible, for $S^3 - N(k \cup \tau)$ is a handlebody. Then τ_1 is a trivial knot. $\square$

Suppose now that $m = 1$. Remember that D denotes a meridian disc of $S^3 - \operatorname{int} N(k \cup \tau)$. By Lemma 4.3 we have that $n = 0$, and we can suppose that the intersections of the disc D with S consist of collections of arcs in D, where the outermost arcs have ends in β_1.

We construct a tree in D as in the proof of Lemma 4.4. Consider the graph obtained by cutting the outermost vertices, and choose one of the outermost vertices in the new graph. Now we consider the region F associated with this vertex. This disc is bordered by intersection arcs where all the arcs are outermost arcs except at most one, which we denote by λ.

The outermost arcs have endpoints in β_1, and the endpoints $\{a, b\}$ of the arc λ are one of the pairs from the set $\{\{s_1, s_2\}, \{s_i, s_i\}, \{s_i, \beta_1\}, \{\beta_1, \beta_1\}\}$, with $i = 1$ or 2 (see Figure 5).

Case 1. The arc λ in the region F has its endpoints in $\{s_1, s_2\}$. The arc λ connects s_1 with s_2. Let γ be an arc in $\partial N(k)$, lying in the part of $N(k)$ which is in the solid torus R, so that $\partial\gamma = \partial\lambda$. Let L the link formed by τ_1 and $\gamma \cup \lambda$. Note that $\lambda \subset \partial R$, and that the interior of γ is inside R. We will show that L has an unknotting tunnel.

Let k' be the arc of k lying in the exterior or R. Let k_i be an arc in $\partial N(k')$ that connects s_i and the point $\tau_2 \cap N(k')$, $i = 1, 2$. Assume that $k_1 \cap k_2$ is just the point $\tau_2 \cap N(k')$. Suppose that $N(k') = N(k_1) \cup N(k_2)$. An unknotting tunnel $\hat{\tau}$ for L is formed by the union of τ_2 and k_1. Let F' be the disc in D cut by λ and which contains F. Note that $\partial F' = \lambda \cup \rho$, where ρ is an arc in $N(k') \cup N(\tau)$, and furthermore $\rho = \rho_1 \cup \rho_2 \cup \rho_3$, where $\rho_1 \subset \partial N(k_1)$ and $\rho_3 \subset \partial N(k_2)$. We slide λ along $\hat{\tau}$, following ρ, by first sliding λ over $N(k_1)$, then sliding λ over $N(\tau_2)$, then sliding λ over $N(\tau_1)$, and so on. We do this according to $\partial F'$, until we get to the point $\rho_2 \cap \rho_3$. Now we push the previous arc (equivalent to $\lambda \cup \rho_1 \cup \rho_2$) through F', deforming it into ρ_3. We see that a neighborhood of the complex

$$L \cup \hat{\tau} = \lambda \cup \gamma \cup k_1 \cup \tau_2 \cup \tau_1$$

is deformed into a neighborhood of the complex

$$k_2 \cup k_1 \cup \gamma \cup \tau_2 \cup \tau_1 = k \cup \tau.$$

This proves that $\hat{\tau}$ is a tunnel for L.

We can isotope the link L into R, since $\lambda \subset \partial R$ and the interior of γ is inside R. This link has a tunnel number 1 and does not meet $\overline{S}$. By the classification [Eudave-Muñoz and Uchida 1996] of links which have tunnel number 1 and contain an incompressible torus in their exteriors, this cannot happen unless τ_1 is the trivial knot, and in this case we have the first assertion of Proposition 4.1.

In what follows, suppose that the arc τ_2 is very short, that is, isotope τ_2 until it is almost contained in the boundary of the solid torus R. Let R' be the solid torus $R' = R \cup N(\tau_2)$, and let $S' = \partial R'$. Note that S' intersects k in four points, and then there are two arcs of k in the complement of R', say k^1 and k^2, where k^i is the arc with one endpoint in s_i, $i = 1, 2$.

Case 2. The arc λ in the region F has its ends in $\{s_i, s_i\}$, $i = 1, 2$.

Suppose without loss of generality that the arc λ in S connects s_1 with s_1. In S we have a collection of arcs with ends in β_1, which correspond to the outermost arcs determined by F. These arcs are parallel in S, since each outermost disc determines an annulus with boundary in $\overline{S}$ and $\partial N(\tau_1)$, like in Lemma 4.5. Furthermore the boundary of each of these annuli in $\overline{S}$ is a curve with slope p/q, with $q \geq 2$. Since the arc λ is disjoint from these curves, there are two possibilities for this arc. Either it bounds a disc or punctured disc D' in S, or with a subarc of s_1 is a curve of slope p/q in S.

If λ bounds such a disc D', then there is an intersection arc between S and D, which is trivial and outermost in S. This is clear if s_2 is not contained in D''. If s_2 is contained in D'', then there is a trivial arc with endpoints in s_2, as ∂D intersects s_1 and s_2 in the same number of points. This is not possible.

Then we have that λ with a subarc of s_1 is a curve of slope p/q in S. We can consider F as a disc whose boundary consists of the arc λ, two arcs μ_1 and μ_2 in $N(k^1)$, plus one arc λ' in S'. Note that μ_1 and μ_2 are parallel in $N(k^1)$; that is, there is a disc G in $\partial N(k^1)$, such that $F \cap G = \mu_1 \cup \mu_2$. Let $H = F \cup G$. This is an annulus whose boundary is contained in S', and each of these curves has slope p/q. Then H is an annulus properly embedded in the exterior of R' and its boundary consists of curves with nonintegral slope. Then H is parallel to an annulus H' contained in S', that is, H and H' bound a solid torus. Let

$$T = H \cup (S' - H'),$$

and push this torus slightly such that the arc k^1 is contained in the interior of the solid torus bounded by H and H'.

We have two possibilities: 1) T is disjoint from k and τ. This case is not possible if R is knotted, for T would be an incompressible torus in the handlebody $S^3 - N(k \cup \tau)$, which is not possible. So, τ_1 must be a trivial knot. 2) The torus T intersects k in two points and does not meet τ. We claim that T is incompressible in $E(k)$ or that τ_1 is a trivial knot. Note that T and S cobound a product region, and each of these tori intersects k in two points. So T must be incompressible in the region containing R. Suppose that there is a compression disc E lying in the region not containing R. Let $\gamma = \partial E$. Then we have two cases: γ is essential in T or γ is trivial in T (without considering the intersections with k). If γ is essential in T, then T is not knotted, so τ_1 is a trivial knot. If γ is trivial in T, then it bounds a disc $E' \subset T$. Since γ is essential in $T - N(k)$, E' must contain the intersection points between k and T, then the arc of k is contained in the ball bounded by $E \cup E'$. Now,

$$T' = (T - E') \cup E$$

is a torus which intersects neither k nor τ. If T' is incompressible in $S^3 - N(k \cup \tau)$, then there would be an incompressible torus in $S^3 - N(k \cup \tau)$, which cannot happen. If T' is compressible, then it is not knotted, so τ_1 is a trivial knot.

We conclude that either τ_1 is a trivial knot, or that there is another torus T intersecting k in two points, incompressible in $E(k)$, disjoint from τ, but such that τ_1 is a core of the solid torus bounded by T.

Case 3. The arc λ in the region F has its ends in $\{s_i, \beta_1\}$, $i = 1, 2$. Suppose without loss of generality that the arc λ connects s_1 with β_1. We can suppose that ∂F consist of the arc λ, an arc μ_1 in $N(k^1)$, plus an arc in S'. Push the arc k^1, using the disc F, until it is in a neighborhood of R'. We can take a bigger torus T, which does not intersect the tunnel and meets k twice.

We claim that T is incompressible in $E(k)$ or that τ_1 is a trivial knot. The proof is similar to the proof in the previous case.

Case 4. The arc λ in the region F has the ends in $\{\beta_1, \beta_1\}$. In this case all the arcs have the ends in β_1. We can assume that ∂F lies in the torus S'. If the boundary of the disc F is nontrivial in S', the torus R' cannot be knotted and then τ_1 is a trivial knot. If the boundary of F is trivial in S', then for homological reasons, the arc λ must be parallel in S' to the other arcs with ends in β_1, and there are in total an even number of arcs with ends in β_1. It follows that F bounds a disc E in S', which contains the two points of intersection of k with $\partial N(\tau_2)$. Then both arcs k^1 and k^2 are inside the 3-ball bounded by $F \cup E$, and by exchanging F for E, we find a bigger torus which does not intersect k nor τ. As before, the torus is unknotted, i.e., τ_1 is a trivial knot.

This completes the proof of Proposition 4.1. □

Let k be a knot of tunnel number 1 and $\overline{S}$ a meridional essential torus for (S^3, k), which intersects the knot in two points. As before, let $S = \overline{S} \cap E(k)$. Let $\tau = \tau_1 \cup \tau_2$ be an unknotting tunnel for k such that $S \cap \tau = \varnothing$. The surface $\overline{S}$ divides S^3 in two parts $S^3 = V \cup W$, and one of them is a solid torus. Suppose that τ is contained in V. Let $M = S^3 - \operatorname{int} N(k \cup \tau)$. Then M is a handlebody, and S divides M in two handlebodies, say $M = V' \cup W'$, where $V' = V - \operatorname{int} N(k \cup \tau)$ and $W' = W - \operatorname{int} N(k)$.

Lemma 4.10. *V is a solid torus and W is not a solid torus.*

Proof. Suppose that W is a solid torus. As W' is a handlebody, $\partial W'$ is compressible. Let c be the boundary of a meridian disc of k which is in W. Note that $\partial W' - c$ is incompressible in W', for otherwise S would be compressible. Applying Jaco's addition lemma [1984], we have that $W'[c]$ has incompressible boundary (where $W'[c]$ denotes W' with a 2-handle attached along the curve c). On the other hand $W'[c] = W$ which has compressible boundary, and this is not possible. Therefore W cannot be a solid torus, and then V is a solid torus. □

This implies that V is knotted in S^3. As V is a solid torus, we have 3 cases:

(a) τ_1 is inside a 3-ball in V;

(b) τ_1 is a core of V; or

(c) τ_1 is essential in V (that is, cases (a) and (b) do not happen).

Lemma 4.11. *Case (b) cannot happen, and if case (a) happens, τ_1 is a trivial knot.*

Proof. Suppose that case (a) happens; that is, τ_1 is inside a 3-ball B contained in V. Then $k \cap V$ consists of an arc k' properly embedded in V. Let $k' = k_1 \cup k_2$, where k_1 and k_2 are arcs such that $k_1 \cap k_2 = k \cap \tau_2$. Let D be a compression disc for M. The intersection between S and D consists of simple closed curves and arcs, and the simple closed curves can be deleted as usual, because $S \cap M$ is incompressible in M. Let γ be an outermost arc in D, so γ cuts a disc F. If

F were contained in W', it would be a ∂-compression disc for S, which is not possible. Then $F \subset V'$. Note that $\partial F = \gamma \cup \beta$, $\gamma \subset S$ and $\beta \subset N(k \cup \tau)$. Then $\beta = \beta_1 \cup \beta_2 \cup \beta_3$, where β_1 is contained in $\partial N(k_i)$, β_2 is contained in $\partial N(\tau_2 \cup \tau_1)$ and β_3 is contained in $\partial N(k_j)$. Suppose first that $i \neq j$. Shrink τ_2 into τ_1, such that k_1 and k_2 can be seen as arcs with one endpoint in $\partial N(\tau_1)$. Then the arc β can be seen as $\beta = \beta_1 \cup \beta_2 \cup \beta_3$, where β_1 and β_3 are as before and β_2 is an arc on $\partial N(\tau_1)$. By sliding k_1 along $\partial N(\tau_1)$ following β_2, we see that k_1 and k_2 are parallel arcs; that is, there is a disc F' in V' such that $\partial F' = \gamma \cup \beta_1 \cup \beta_2' \cup \beta_3$, where β_2' is an arc in $N(\tau)$, disjoint from a meridian of τ_1. Cut V' along F, to get a handlebody V'', which is homeomorphic to $V - N(\tau_1 \cup k'')$, where k'' is an arc with endpoints on S and τ_1 (it can be considered as k_1). This is not possible by Proposition 3.1.

Suppose now that β_1 and β_3 are both contained in $\partial N(k_1)$. Shrink τ_2 into τ_1 again, such that k_1 and k_2 can be seen as arcs each with one endpoint in $\partial N(\tau_1)$. There is a disc $C \subset \partial N(k_1)$ such that $C \cup F$ is an annulus with one boundary component, say C_1, lying on S, and the other boundary component, C_2, lying on $\partial N(\tau_1)$. The closed curve C_2 is either trivial in $\partial N(\tau_1)$ or it is essential. Suppose first that C_2 is trivial in $\partial N(\tau_1)$. Then it bounds a disc $E \subset \partial N(\tau_1)$ which contains an endpoint of k_2. If C_1 is trivial on S, then k_2 must be an arc parallel to k_1, and we proceed as in the previous case. If C_1 is nontrivial on S, then it must be a meridian of S, for $C \cup F \cup E$ is a disc in V with boundary C_1. By taking a copy of $C \cup F \cup E$ and pushing it to be disjoint from $k_1 \cup \tau_1$, we get a disc whose boundary is a meridian of S and which intersects k_2 in one point, and then it is a meridian disc that intersects k in one point. This is not possible because S is meridionally incompressible.

Suppose now that C_2 is essential in $\partial N(\tau_1)$. Assume that the annulus $C \cup F$ and the sphere ∂B intersect transversely, and note that $\partial(C \cup F)$ is disjoint from ∂B. Let α be an innermost curve of intersection on ∂B. If α is a trivial curve in $C \cup F$, we can find another 3-ball containing τ_1 whose boundary has fewer intersections with $C \cup F$. If α is essential in $C \cup F$, then by cutting $C \cup F$ with the disc in ∂B bounded by α we get an embedded disc whose boundary is C_1. If C_1 is not a longitudinal curve in $\partial N(\tau_1)$, this implies that there is a punctured lens space embedded in V, which is impossible. So, C_1 must be a longitude of $\partial N(\tau_1)$, and then τ_1 must be a trivial knot.

Suppose now that case (b) happens; that is, τ_1 is a core of V. As above, let k' be the arc $k \cap V$, such that $k' = k_1 \cup k_2$, where k_1 and k_2 are arcs with $k_1 \cap k_2 = k \cap \tau_2$. Slide k_2 over τ_2, getting two arcs, k_1' and k_2', each with one endpoint on S and one in τ_1. By Proposition 3.2, it follows that k_1' and k_2' are a pair of simultaneously straight arcs in the space product $V - N(\tau_1)$. By sliding back k_2' over k_1', we see that k' is an arc in V that is isotopic to an arc contained in ∂V. This implies that T is compressible in $S^3 - k$, a contradiction. □

Proof of Theorem 2.3. Let k be a knot of tunnel number 1, $\overline{S}$ a meridional essential torus which intersects the knot in two points and $\tau = \tau_1 \cup \tau_2$ an unknotting tunnel for k. Suppose first that τ cannot be made disjoint from $\overline{S}$. Then by Proposition 4.1, either τ_1 is a trivial knot, or there is another essential meridional torus $\overline{S}'$, which intersects k in two points, is disjoint from τ, and such that τ_1 is a core of the solid torus bounded by $\overline{S}'$. However, the existence of such a torus contradicts Lemma 4.11, so this case is not possible. Therefore, τ_1 is a trivial knot, and by Lemma 2.1, k is a $(1,1)$-knot.

Suppose now that τ and $\overline{S}$ are disjoint. By Lemma 4.10, $\overline{S}$ bounds a solid torus V in which τ lies. Then by Lemma 4.11, either τ_1 is a trivial knot, and then k is a $(1,1)$-knot, or we have case (c), that is, τ_1 is an essential curve in V. So, suppose that case (c) happens. Then $\overline{S}$ is essential in $S^3 - \tau_1$ and $\tau_2 \cup k$ is a tunnel for τ_1. Then τ_1 is a satellite knot with tunnel number 1, and this implies that $\overline{S}$ is knotted as a torus knot, by the result of Morimoto and Sakuma [1991]. Slide k over τ_2 until it becomes an arc k' with endpoints on τ_1. Then k' has to be one of the unknotting tunnels for τ_1 as classified by Morimoto and Sakuma [1991]; that is, by sliding k' over $\partial N(\tau_1)$ we get an arc ρ which is one of the tunnels $\tau(1,x)$, $\tau(2,x)$, $\tau(2,y)$ or $\tau(1,y)$ for τ_1, as defined in Section 2. To get k from ρ, we have to slide ρ over $\partial N(\tau_1)$ and then over itself, but this is equivalent to taking an arc on $\partial N(\tau_1)$ joining the endpoints of ρ, in fact the arc γ determined by the sliding of ρ over $\partial N(\tau_1)$, and then taking the iterate of ρ and τ_1 using the arc γ.

This completes the proof of Theorem 2.3. □

Acknowledgements

We would like to thank the referee for their careful reading and many helpful suggestions and comments. Eudave-Muñoz was partially supported by project PAPIIT-UNAM IN102814.

References

[Casson and Gordon 1987] A. J. Casson and C. M. Gordon, "Reducing Heegaard splittings", *Topology Appl.* **27**:3 (1987), 275–283. MR Zbl

[Eudave-Muñoz 1994] M. Eudave-Muñoz, "On nonsimple 3-manifolds and 2-handle addition", *Topology Appl.* **55**:2 (1994), 131–152. MR Zbl

[Eudave-Muñoz 1999] M. Eudave-Muñoz, "Incompressible surfaces in tunnel number one knot complements", *Topology Appl.* **98**:1-3 (1999), 167–189. MR Zbl

[Eudave-Muñoz 2000] M. Eudave-Muñoz, "Essential meridional surfaces for tunnel number one knots", *Bol. Soc. Mat. Mexicana* (3) **6**:2 (2000), 263–277. MR Zbl

[Eudave-Muñoz 2006] M. Eudave-Muñoz, "Incompressible surfaces and $(1,1)$-knots", *J. Knot Theory Ramifications* **15**:7 (2006), 935–948. MR Zbl

[Eudave-Muñoz and Ramírez-Losada 2009] M. Eudave-Muñoz and E. Ramírez-Losada, "Meridional surfaces and $(1,1)$-knots", *Trans. Amer. Math. Soc.* **361**:2 (2009), 671–696. MR Zbl

[Eudave-Muñoz and Uchida 1996] M. Eudave-Muñoz and Y. Uchida, "Non-simple links with tunnel number one", *Proc. Amer. Math. Soc.* **124**:5 (1996), 1567–1575. MR Zbl

[Frohman 1989] C. Frohman, "An unknotting lemma for systems of arcs in $F \times I$", *Pacific J. Math.* **139**:1 (1989), 59–66. MR Zbl

[Gordon and Reid 1995] C. M. Gordon and A. W. Reid, "Tangle decompositions of tunnel number one knots and links", *J. Knot Theory Ramifications* **4**:3 (1995), 389–409. MR Zbl

[Jaco 1984] W. Jaco, "Adding a 2-handle to a 3-manifold: an application to property R", *Proc. Amer. Math. Soc.* **92**:2 (1984), 288–292. MR Zbl

[Morimoto and Sakuma 1991] K. Morimoto and M. Sakuma, "On unknotting tunnels for knots", *Math. Ann.* **289**:1 (1991), 143–167. MR Zbl

[Wu 1992] Y. Q. Wu, "A generalization of the handle addition theorem", *Proc. Amer. Math. Soc.* **114**:1 (1992), 237–242. MR Zbl

Received July 1, 2015. Revised August 7, 2016.

MARIO EUDAVE-MUÑOZ
INSTITUTO DE MATEMÁTICAS
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
CIRCUITO EXTERIOR
CIUDAD UNIVERSITARIA
04510 MÉXICO DF
MEXICO

and

CIMAT
CALLEJÓN JALISCO S/N
COL. VALENCIANA 36240
GUANAJUATO
MEXICO

mario@matem.unam.mx

GRISSEL SANTIAGO-GONZÁLEZ
FACULTAD DE CIENCIAS
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
CIRCUITO EXTERIOR
CIUDAD UNIVERSITARIA
04510 MÉXICO, DF
MEXICO

gissel@matem.unam.mx

PACIFIC JOURNAL OF MATHEMATICS
Vol. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.343

ON BISECTIONAL NONPOSITIVELY CURVED COMPACT KÄHLER–EINSTEIN SURFACES

DANIEL GUAN

We prove a conjecture on the pinching of the bisectional curvature of nonpositively curved Kähler–Einstein surfaces. We also prove that any compact Kähler–Einstein surface M is a quotient of the complex two-dimensional unit ball or the complex two-dimensional plane if M has nonpositive Einstein constant and, at each point, the average holomorphic sectional curvature is closer to the minimum than to the maximum.

1. Introduction

In [Siu and Yang 1981] the authors conjectured that any compact Kähler surface with negative bisectional curvature is a quotient of the complex two-dimensional unit ball. They proved that there is a number $a \in \left(\frac{1}{3}, \frac{2}{3}\right)$ such that if, at every point P, $K_{\mathrm{av}} - K_{\min} \le a[K_{\max} - K_{\min}]$ then M is a quotient of the complex ball. Here $K_{\min}$, $K_{\max}$ and K_{av} is the minimal, maximal and average value of the holomorphic sectional curvature, respectively. The number a they obtained was $a < 2/\left(3\left[1+\sqrt{\frac{6}{11}}\,\right]\right) < 0.38$ (see [Polombo 1992, p. 398]). In [Hong et al. 1988], Yi Hong pointed out that this is also true if $a \le 2/\left(3\left[1+\sqrt{\frac{1}{6}}\right]\right) < 0.476$. We observed in [Hong et al. 1988, Theorem 2] that if $a \le \frac{1}{2}$, then there is a ball-like point P. That is $K_{\max} = K_{\min}$ at P. We notice here that $\sqrt{\frac{1}{6}} > \frac{1}{3}$. Therefore, we conjectured in [Hong et al. 1988] that M is a quotient of the complex ball if $a = \frac{1}{2}$. In general, we believe that we might not get a quotient of the complex ball if $a > \frac{1}{2}$. Around 1992 Hong Cang Yang almost proved this conjecture except for some technical difficulties, see the argument of Theorem 1.2 in [Chen et al. 2011]. Polombo [1988; 1992] used a different method and proved that a can be $(3 + (4\sqrt{3})/3)/11 < 0.48$ (according to [Chen et al. 2011, p. 2628 right before Theorem 1.2]), see [Polombo 1988, p. 669] or [Polombo 1992, p. 398]. In [Chen et al. 2011], the authors improved the constant to $a < \frac{1}{2}$ which gave a proof of a weaker version of the conjecture.

We first notice that in the proof of Theorem 2 in [Hong et al. 1988] (for which this author was responsible) we proved that if $K_{\mathrm{av}} - K_{\min} = \frac{1}{2}[K_{\max} - K_{\min}]$ at P,

MSC2010: 32M15, 32Q20, 53C21, 53C55.

Keywords: Kähler–Einstein metrics, compact complex surfaces, bisectional curvature, pinching of the curvatures.

then P must be a ball-like point (for this part, any negativity of the curvature is not needed except to use the result from [Siu and Yang 1981] when $A = B$). See the remark after the Theorem 1 in [Hong et al. 1988]. According to [Siu and Yang 1981, p. 485, Proposition 4] the subset of ball-like points is either the whole manifold or a real codimension two analytic subvariety. Since the function considered in Theorem 1.2 of [Chen et al. 2011] is bounded, it can be extended to all of M, is a constant and must be zero. Notice that we only need that the bisectional curvature is nonpositive. With this in mind, we also have the possibility of the flat case. That is, the manifold could also be a quotient of $\mathbb{C}^2$ if the Einstein constant is zero. This case should also be included in the main theorem of [Siu and Yang 1981, p. 472] and Theorems A and 1 of [Hong et al. 1988].

Since [Hong et al. 1988] was only written in Chinese, we provide a mostly self contained account here. Also, Polombo [1988; 1992] had something more general than stated above. Therefore, we generalized our result to the case of nonpositive Einstein constant.

Theorem. *Let M be a connected compact Kähler–Einstein surface with nonpositive scalar curvature, if we have*

$$K_{\mathrm{av}} - K_{\mathrm{min}} \leq \tfrac{1}{2}[K_{\mathrm{max}} - K_{\mathrm{min}}]$$

at every point, then M is a compact quotient of either the complex two-dimensional unit ball or the two-complex-dimensional plane.

This note is written in such a way that experts who are familiar with [Hong et al. 1988; Chen et al. 2011] will be able to understand the proof of the conjecture stated in those works from the present introduction. For those only familiar with the second of those references, the present Section 2 should be enough to understand the proof of the conjecture. Notice that we do not need the nonpositivity of the bisectional curvature except to apply the result of [Siu and Yang 1981] or [Chen et al. 2011] to the case $A = B$. We shall give a complete proof of the conjecture in Section 3, with a simpler explanation than that of [Chen et al. 2011] for the last step, that also explains away the mystery of the negativity. In Section 4, we apply these methods to prove our theorem.

To the author, the conjecture in [Siu and Yang 1981] is very important to complex geometry. This work is heavily dependent on earlier works in this subject. Although we are able to prove the conjecture from [Hong et al. 1988; Chen et al. 2011] and our main theorem, there is more work which needs to be done in the direction of compact complex surfaces with negative holomorphic bisectional, or even real, sectional curvatures. Therefore, the author thinks that it is proper to write this paper with an emphasis on the nonpositive holomorphic bisectional curvature case instead of the case of our main theorem.

2. Existence of ball-like points

Here, we repeat the argument in the proof of Theorem 2 in [Hong et al. 1988].

Proposition 1 [Hong et al. 1988, p. 597–599]. *Suppose that*

$$K_{\mathrm{av}} - K_{\min} \le \tfrac{1}{2}[K_{\max} - K_{\min}]$$

for every point on the compact Kähler–Einstein surface with nonpositive holomorphic bisectional curvatures. There is at least one ball-like point.

Proof. Throughout this paper, as in [Siu and Yang 1981; Chen et al. 2011], we assume that $\{e_1, e_2\}$ is a unitary basis at a given point P with

$$\begin{aligned} R_{1\bar{1}1\bar{1}} &= R_{2\bar{2}2\bar{2}} = K_{\min}, \\ R_{1\bar{1}1\bar{2}} &= R_{2\bar{2}2\bar{1}} = 0, \\ A &= 2R_{1\bar{1}2\bar{2}} - R_{1\bar{1}1\bar{1}} \ge 0, \\ B &= |R_{1\bar{2}1\bar{2}}|. \end{aligned}$$

As in [Siu and Yang 1981], we always have $A \ge |B|$ and we assume that $B = R_{1\bar{2}1\bar{2}}$ (i.e., the latter is nonnegative).

If P is not a ball-like point, according to [Siu and Yang 1981], we can do as above for a neighborhood $U(P)$ of P whenever $A > B$ (Case 1 in [Siu and Yang 1981], page 475). We should handle the case in which $A = B$ at the end of this proof. We write

$$\begin{aligned} \alpha &= e_1 = \sum a_i \partial_i, \\ \beta &= e_2 = \sum b_i \partial_i, \\ S_{1\bar{1}1\bar{1}} &= R(e_1, \bar{e}_1, e_1, \bar{e}_1) = \sum R_{i\bar{j}k\bar{l}} a_i \bar{a}_j a_k \bar{a}_l, \end{aligned}$$

and so on. In particular, we have

$$S_{1\bar{1}1\bar{1}} = S_{2\bar{2}2\bar{2}} = K_{\min}, \quad S_{1\bar{1}1\bar{2}} = S_{2\bar{2}2\bar{1}} = 0.$$

According to [Siu and Yang 1981], we have

$$\begin{gathered} K_{\max} = K_{\min} + \tfrac{1}{2}(A + B), \\ K_{\mathrm{av}} = K_{\min} + \tfrac{1}{3}A, \\ \tfrac{1}{3}[K_{\max} - K_{\min}] \le K_{\mathrm{av}} - K_{\min} \le \tfrac{2}{3}[K_{\max} - K_{\min}]. \end{gathered}$$

Our condition in Proposition 1 is therefore equivalent to $A \le 3B$. As in [Hong et al. 1988], we let $\Phi_1 = |B|^2/A^2 = \tau^2$.

If there is no ball-like point, since $\frac{1}{3} \le \tau \le 1$, there is a minimal point.

We shall calculate the Laplacian of Φ_1 at a minimal point, which is not ball-like. For example, when $A = 3B$, the minimum $\Phi_1 = \frac{1}{9}$, is achieved. The Laplacian at that point should be nonnegative.

We let

$$x_i = \nabla_i \Phi_1 = 2\frac{\tau}{A}[\operatorname{Re} \nabla_i S_{1\bar{2}1\bar{2}} + 3\tau \nabla_i S_{1\bar{1}1\bar{1}}].$$

As we pointed out earlier, we first assume that A does not equal B, then we can apply the argument in case 1 of [Siu and Yang 1981, p. 475] at the minimal point since $A > B$.

As in [Siu and Yang 1981; Hong et al. 1988; Chen et al. 2011], we have

$$\Delta R_{1\bar{1}1\bar{1}} = -AR_{1\bar{1}2\bar{2}} + B^2, \quad \Delta R_{1\bar{2}1\bar{2}} = 3(R_{1\bar{1}2\bar{2}} - A)B.$$

At P we have $a_1 = b_2 = 1$, $a_2 = b_1 = 0$, $\nabla a_1 = \nabla b_2 = 0$ and $\nabla a_2 + \nabla \bar{b}_1 = 0$. Therefore, we write $y_{i1} = \nabla_i a_2$ and $y_{i2} = \nabla_i \bar{a}_2$. We also have

$$\Delta(a_1 + \bar{a}_1) = -|\nabla a_2|^2, \quad \Delta(a_2 + \bar{b}_2) = 0, \quad \nabla_i R_{1\bar{1}1\bar{2}} = -Ay_{i1} - By_{i2},$$

since

$$0 = \nabla S_{1\bar{1}1\bar{2}} = \nabla R_{1\bar{1}1\bar{2}} + 2R_{2\bar{1}1\bar{2}} \nabla a_2 + B\nabla \bar{a}_2 + R_{1\bar{1}1\bar{1}} \nabla \bar{b}_1,$$

i.e.,

$$\nabla R_{1\bar{1}1\bar{2}} = -A\nabla a_2 - B\nabla \bar{a}_2.$$

This also gives a similar formula for $\nabla_{\bar{i}} R_{1\bar{1}1\bar{2}}$. Similarly,

$$\begin{aligned}
\nabla S_{1\bar{1}1\bar{1}} &= \nabla R_{1\bar{1}1\bar{1}}, \\
\nabla S_{1\bar{2}1\bar{2}} &= \nabla R_{1\bar{2}1\bar{2}}, \\
\Delta S_{1\bar{1}1\bar{1}} &= -2A \sum |y|^2 - 4B \operatorname{Re} \sum y_{i1}\bar{y}_{i2} - AR_{1\bar{1}2\bar{2}} + B^2, \\
\operatorname{Re} \Delta S_{1\bar{2}1\bar{2}} &= 4A \sum \operatorname{Re} y_{i1}\bar{y}_{i2} + 2B \sum |y|^2 + 3(R_{1\bar{1}2\bar{2}} - A)B., \\
\nabla_{\bar{1}} S_{1\bar{2}1\bar{2}} &= -A\bar{y}_{22} - B\bar{y}_{21}, \\
\nabla_2 S_{1\bar{2}1\bar{2}} &= Ay_{11} + By_{12}, \\
\nabla_1 S_{1\bar{2}1\bar{2}} &= -A(6\tau^2 - 1)y_{22} - 5A\tau y_{21} + x_1, \\
\nabla_{\bar{2}} S_{1\bar{2}1\bar{2}} &= 5A\tau \bar{y}_{12} + A(6\tau^2 - 1)\bar{y}_{11} + \bar{x}_2.
\end{aligned}$$

As in [Hong et al. 1988, p. 598] at P we have

$$
\begin{aligned}
\Delta\Phi_1 &= \frac{2\tau\Delta B}{A} + \frac{6\tau^2}{A}\Delta S_{1\bar{1}1\bar{1}} + \frac{1}{A^2}\sum\bigl(|\nabla S_{1\bar{2}1\bar{2}}|^2 + |\overline{\nabla} S_{1\bar{2}1\bar{2}}|^2\bigr) \\
&\quad + \frac{54\tau^2}{A^2}\sum|\nabla S_{1\bar{1}1\bar{1}}|^2 + \frac{12\tau}{A^2}\sum\operatorname{Re}\bigl(\nabla_i S_{1\bar{1}1\bar{1}}(\nabla_{\bar{i}}(S_{1\bar{2}1\bar{2}} + S_{2\bar{1}2\bar{1}}))\bigr) \\
&= 2\tau\Bigl[3A\tau(\tau^2-1) - 4\tau\sum|y|^2 + 4(1-3\tau^2)\sum\operatorname{Re}(y_{i1}\bar{y}_{i2})\Bigr] \\
&\quad + |y_{22}+\tau y_{21}|^2 + |y_{11}+\tau y_{12}|^2 \\
&\quad + \frac{1}{A^2}\Bigl[\bigl|x_1 + A[(1-6\tau^2)y_{22} - 5\tau y_{21}]\bigr|^2 \\
&\qquad\qquad + \bigl|x_2 + A\bigl[(6\tau^2-1)y_{11} + 5\tau y_{12}\bigr]\bigr|^2\Bigr] \\
&\quad - 18\tau^2\bigl[|y_{12}+\tau y_{11}|^2 + |y_{21}+\tau y_{22}|^2\bigr] \\
&\quad + \frac{12\tau}{A}\bigl[\operatorname{Re}[(y_{21}+\tau y_{22})\bar{x}_1] - \operatorname{Re}[(y_{21}+\tau y_{11})\bar{x}_2]\bigr]
\end{aligned}
$$

Here we notice that $\Delta\Phi_1$ has two general terms. The first term is constant with respect to x and y, and is always nonpositive since $\frac{1}{3} \le \tau \le 1$.

The second term can be regarded as a hermitian form h in x and y. We can separate x and y into two groups: x_1, y_{2j} in one group and x_2, y_{1j} in the other. These two groups of variables are orthogonal to each other with respect to this hermitian form. That is, $h = h_1 + h_2$ where h_1 and h_2 depend only on the first and second group of variables, respectively.

We need to check the nonpositivity for each term.

For x_2, y_{11}, y_{12}, the corresponding matrix of h_2 is

$$
\begin{bmatrix}
1/A^2 & -1/A & -\tau/A \\
-1/A & 2(9\tau^2-1)(\tau^2-1) & 0 \\
-\tau/A & 0 & 0
\end{bmatrix},
$$

and the matrix for h_1 of x_1, y_{21}, y_{22} is

$$
\begin{bmatrix}
1/A^2 & \tau/A & 1/A \\
\tau/A & 0 & 0 \\
1/A & 0 & 2(9\tau^2-1)(\tau^2-1)
\end{bmatrix}.
$$

When P is a critical point of Φ_1, then $x_1 = x_2 = 0$. The matrix for y is clearly seminegative. Therefore, if there is no ball-like point, then we have, at the minimal point of Φ_1, that $\tau^2 = 1$ or $A = 0$ since $\tau \ge \frac{1}{3}$.

If $A = 0$, then we have a ball-like point, and we are done.

On the other hand,[1] if $\tau = 1$, we have $A = B$ at P. Since P is a minimal point, this implies that $A = B$ on the whole manifold. According to [Siu and Yang 1981, p. 475, case 2], we have smooth coordinates with $K_{\max} = R_{1\bar{1}1\bar{1}}$. (Fortunately, this works whenever $A = B$. In general, the original argument might not always work since one might not have $A = B$ in a neighborhood. However, as was pointed out in [Siu and Yang 1981, case 1], under our condition the directions for $K_{\max}$ are always isolated. Therefore, it might be better to choose $K_{\max}$ instead of $K_{\min}$ from the very beginning. But this is not in the scope of this paper.) Using this new coordinate, we can define similar functions A_1 and B_1. In general, $B_1 = \frac{1}{2}(A - B)$ and $A_1 = -\frac{1}{2}(A + 3B)$. In our case, $B_1 = 0$ and $A_1 = -2A$. Using this new coordinate, one can do the calculation for any of the functions in [Siu and Yang 1981; Polombo 1988; 1992; Chen et al. 2011] that the set of ball-like points is the whole manifold. If one does not like Polombo's function Φ_α [1992, p. 418] with $\alpha = -\frac{8}{7}$, then one might simply use the function with $\alpha = -1$ (in [Polombo 1988; Polombo 1992], not the vector we mentioned in this paper earlier), i.e., the new function is proportional to $\Phi_2 = (3B - A)A$. In our case, this is just $2A^2$. We can apply $\Phi_2^{1/3}$. This is relatively easy and is left to the reader. We can also use the argument in [Siu and Yang 1981, case 1], in which the minimal vectors are not isolated but they are points in a smooth circle bundle over the manifold so we could choose a smooth section instead.

Also, the preceding paragraph is not needed in Corollary 2 and Lemma 3 since, in those two propositions, we already have $A = 3B$. With $A = B$, one could readily get that $A = B = 0$.

If $A = 0$, $K_{\max} = K_{\min}$ and P is a ball-like point, then we have a contradiction. Therefore, the set of ball-like points is not empty. □

Observe that if $A = 3B$ at P, then Φ_1 achieves the minimal value at P and $A \neq B$ unless P is a ball-like point. That is the first part of the proof of Proposition 1 goes through. That is, P must be a ball-like point.

Corollary 2. *Assume the above, if $K_{\mathrm{av}} - K_{\min} = \frac{1}{2}[K_{\max} - K_{\min}]$ at P, then P is a ball-like point.*

Therefore, we have:

Lemma 3. *If $K_{\mathrm{av}} - K_{\min} \leq \frac{1}{2}[K_{\max} - K_{\min}]$ on M, then $K_{\mathrm{av}} - K_{\min} < \frac{1}{2}[K_{\max} - K_{\min}]$ on $M - N$, where N is the subset of all the ball-like points.*

Therefore, we can apply the argument of [Chen et al. 2011]. To do that one needs the following proposition:

Proposition 4 (see [Siu and Yang 1981] and [Hong et al. 1988, Theorem 3]). *If $N \neq M$, then N is a real analytic subvariety and* $\operatorname{codim} N \geq 2$.

[1]This paragraph is not needed for the proofs of Corollary 2 and Lemma 3. Also, in this special case, the original frame in [Siu and Yang 1981] works. So, one could apply [Siu and Yang 1981].

As in [Siu and Yang 1981], Proposition 4 gives us a way to the conjecture by finding a superharmonic function on M which was obtained by Hong Cang Yang around 1992. In [Siu and Yang 1981; Hong et al. 1988], the authors used $\Phi = 6B^2 - A^2$. Polombo [1992, p. 417, Lemma] used $(11A-3B)(B-A)+16AB$. One might ask why do we need another function, why do we not use Φ_1? The answer is that by a power of Φ_1 we can only correct the Laplacian by $|\nabla\Phi_1|^2$. But that could only change the upper left coefficients of our matrices as it only provides $|x|^2$ terms. In the case of Φ_1, it does not work since $\tau/A \neq 0$ but the coefficients of $|y_{12}|^2$ and $|y_{21}|^2$ are zeros. Therefore, we need another function, which was provided by Hong Cang Yang.

Remark 5. Whenever there is a bounded continuous nonnegative function f on M such that $f(N)=0$, f is real analytic on $M-N$ and $\Delta f \le 0$ on $M-N$, then $f=0$. *Here N could be just a codimension two subset.* This is in general true for extending continuous superharmonic functions over a codimension two subset, see [Siu and Yang 1981; Hong et al. 1988; Chen et al. 2011]. Here, we would like to give our own reasons why this is true in these special cases. If we define $M_s = \{x \in M|_{\mathrm{dist}(x,N)\ge s}\}$ and $h_s = \partial M_s$, then the measure of h_s is smaller than $O(s)$ when s tends to zero. Therefore,

$$0 \ge \ln 2 \int_{M_{2\delta}} \Delta f \omega^n \ge \int_\delta^{2\delta}\left[\int_{M_s} \Delta f \omega^n\right] s^{-1}\,ds = \int_\delta^{2\delta}\left[\int_{h_s} \frac{\partial f}{\partial n}\,d\tau\right] s^{-1}\,ds.$$

But, by applying an integration by parts to the single variable integral, the last term is about $(\delta)^{-1}\int_{h_{2\delta}}(f-g)\,d\tau \to 0$, since f is bounded and $f-g$ tends to 0 near N, where g is the f value of the corresponding point on h_δ. For example, if $f=r^a$ with $a>0$, then

$$\frac{\partial f}{\partial n} = ar^{a-1} = as^{a-1} \quad \text{and} \quad \int_{h_s} \frac{\partial f}{\partial n}\,d\tau = O(s^a) \to 0.$$

Therefore, $\Delta f = 0$ on $M-N$. Hence f extends over N as a harmonic function. This implies that $f=0$ on M.

Now, let $f=(3B-A)^a$, this is natural after the proof of Proposition 1, we will show in the next section that $\Delta f \le 0$ for $a \le \frac{1}{3}$ (see the proof in [Chen et al. 2011]). Therefore, $A=3B$ always. By Corollary 2, we have $A=B=0$. This function is also related to the functions in [Polombo 1992, p. 417] with $a_1=a_3=0$. Polombo had to pick up functions with $a_1=a_2$ to avoid a complication of the singularities. See page 406 and the first paragraph in page 418 in [Polombo 1992] and the last paragraph of page 668 in [Polombo 1988]. We shall completely resolve the difficulty in the next section.

3. Hong Cang Yang's function

Let $\Psi = 3B - A$. Around 1992 Hong Cang Yang considered $f = \Psi^{1/3}$.

Lemma 6 [Chen et al. 2011, p. 2630 (13)]. *We have*

$$\Delta(3B-A)=3\big[\Psi R_{1\bar{1}2\bar{2}}+B(B-3A)\big]+\tfrac{3}{B}|\nabla(\operatorname{Im} R_{1\bar{2}1\bar{2}})|^2+6(B-A)\sum|y_{i1}-y_{i2}|^2.$$

Let $z_i = \nabla_i \Psi$, then

$$z_1 = \nabla_1(3B-A) = \tfrac{3}{2}\nabla_1(R_{1\bar{2}1\bar{2}} + R_{2\bar{1}2\bar{1}} + 2R_{1\bar{1}1\bar{1}}),$$

$$\begin{aligned}\sqrt{-1}\nabla_1(\operatorname{Im} R_{1\bar{2}1\bar{2}}) &= \tfrac{1}{2}\nabla_1(R_{1\bar{2}1\bar{2}} - R_{2\bar{1}2\bar{1}})\\ &= \tfrac{1}{3}z_1 - \nabla_1 R_{2\bar{1}2\bar{1}} - \nabla_1 R_{1\bar{1}1\bar{1}}\\ &= \tfrac{1}{3}z_1 - \nabla_2 \overline{R}_{1\bar{1}1\bar{2}} + \nabla_2 R_{1\bar{1}1\bar{2}}\\ &= \tfrac{1}{3}z_1 + (A-B)y_{22} + (B-A)y_{21},\end{aligned}$$

$$z_2 = \nabla_2(3B-A) = \tfrac{3}{2}\nabla_2(R_{2\bar{1}2\bar{1}} + R_{1\bar{2}1\bar{2}} + 2R_{1\bar{1}1\bar{1}}),$$

$$\begin{aligned}\sqrt{-1}\nabla_2(\operatorname{Im} R_{1\bar{2}1\bar{2}}) &= \tfrac{1}{2}\nabla_2(R_{1\bar{2}1\bar{2}} - R_{2\bar{1}2\bar{1}})\\ &= -\tfrac{1}{3}z_2 + \nabla_2 R_{1\bar{1}1\bar{1}} + \nabla_2 R_{1\bar{2}1\bar{2}}\\ &= -\tfrac{1}{3}z_2 + \nabla_1 R_{2\bar{1}1\bar{1}} - \nabla_1 R_{1\bar{1}1\bar{2}}\\ &= -\tfrac{1}{3}z_2 + (B-A)y_{12} + (A-B)y_{11}.\end{aligned}$$

We can write the formula in the Lemma 6 as

$$\begin{aligned}\Delta\Psi = 3\big[\Psi R_{1\bar{1}2\bar{2}} + B(B-3A)\big] - 3\frac{A-B}{B}\Psi\sum|y_{i1}-y_{i2}|^2\\ + 2\frac{A-B}{B}\operatorname{Re}\big[(y_{12}-y_{11})\bar{z}_2 + (y_{22}-y_{21})\bar{z}_1\big] + \sum\frac{1}{3B}|z|^2.\end{aligned}$$

As in the last section, we have two general terms, the first is negative as is the constant term of z with respect to y. The second is a hermitian form in z and y. We can actually let $w_i = y_{i^*1} - y_{i^*2}$ with $i^* \neq i$. Then the second term is a sum of two hermitian forms. One of them is on w_1, z_1 and the other is on w_2, z_2. We notice that the second term is also nonpositive on y (or nonpositive on w, if we assume that $z = 0$). We can modify the coefficient of $|z|^2$ (only) by taking the power of Ψ. More precisely, if we let $g = \Psi^a$, to make sure that $\Delta g < 0$, after taking out a factor $3(A-B)/B$ we need

$$\begin{vmatrix} -\Psi & \frac{1}{3} \\ \frac{1}{3} & \frac{1-3\Psi^{-1}(1-a)B}{9(A-B)} \end{vmatrix} \geq 0.$$

That is,

$$A - 3B + 3(1-a)B - A + B = (3(1-a)-2)B \geq 0.$$

We have $1-3a \geq 0$. So, $a \leq \frac{1}{3}$.

Therefore, we have:

Lemma 7. *$\Delta g < 0$ for $a \leq \frac{1}{3}$ on $M - N$.*

This is exactly the same as in [Chen et al. 2011]. Actually, the number $\frac{1}{6}$ was already in [Siu and Yang 1981; Hong et al. 1988; Polombo 1988; 1992] for those quadratic functions.

So, finally we have:

Theorem 8. *If $K_{\mathrm{av}} - K_{\min} \leq \frac{1}{2}[K_{\max} - K_{\min}]$, then M has a constant holomorphic sectional curvature.*

Remark 9. The reason we did not get this earlier was that there was a difficulty when $A = B$. In that case, the argument in [Siu and Yang 1981, p. 475, case 2] seems not to work. Polombo resolved the problem by using a function which is symmetric about $\lambda_1 = -A/3$ and $\lambda_2 = A - 3B/6$ (see [Polombo 1992] the first paragraph of page 418 and the end of page 397). However, Hong Cang Yang's function Ψ is only $-6\lambda_2$ and therefore is not symmetric after all. To overcome this difficulty, we let $\Omega = \{x \in M|_{A=B}\}$. Then according to [Siu and Yang 1981], all our calculation are good on $M - \Omega$ since $N \subset \Omega$. In [Chen et al. 2011, p. 2632] there was a suggestion on how to prove that $\operatorname{codim} \Omega \leq 2$, although it was not very well explained. By doing this, everything went through. The relation was that if we use the argument in [Siu and Yang 1981, p. 475, case 2] using the maximum instead of the minimum, and we let $B_1 = |R_{1\bar{2}1\bar{2}}|$ then $2B_1 = A - B$. That is $\Omega = \{x \in M|_{B_1=0}\}$. The argument goes as follows:

Case 1: If Ω is a closed region, we have

$$\begin{aligned} 0 \geq \int_{M-\Omega} \Delta g &= a \int_{-\partial\Omega} \Psi^{a-1} \frac{\partial(-A_1 - 3B_1)}{\partial n} \\ &\geq a \int_{-\partial\Omega} (2A)^{a-1} \frac{\partial(-A_1)}{\partial n} = -\int_{\Omega} \Delta F_1 \geq 0, \end{aligned}$$

where F_1 can be chosen from one of the functions in [Polombo 1992] which satisfy the symmetric condition on M, e.g., a power of Φ_2 from the proof of Proposition 1, or one of our functions with a calculation using the new smooth coordinate in [Siu and Yang 1981, p. 475] with $R_{1\bar{1}1\bar{1}} = K_{\max}$. Actually, A_1 itself is proportional to the λ_2 in [Polombo 1992] and is symmetric in the sense of Polombo. On Ω, F_1 is just our g since $B_1 = 0$. We notice that there is a sign difference for the Laplacian operator in [Polombo 1992]. Again, on Ω, since $A = B$ on a neighborhood, the set of minimum directions is an S^1 bundle over Ω, therefore one can choose a smooth section of it locally such that the calculation of [Siu and Yang 1981] still works in our case. That is, one could simply choose F_1 to be g.

Case 2: If Ω is a hypersurface then the same argument goes through except that

$$\int_{\partial(M-\Omega)} (A)^{a-1}\frac{\partial A}{\partial n} = 0,$$

since $A \neq 0$ outside a codimension one subset and on $\Omega_1 = \{x \in \Omega|_{A\neq 0}\}$, the integral is integrated from both sides.

Therefore, Ω is a subset of codimension two and we can apply Remark 5. By the calculation in Remark 5, we see that g is harmonic on $M-\Omega$. Now, by Lemma 6, that implies that $B(B-3A)=0$ and hence $A=B=0$ by our assumptions.

4. The generalization

Actually, in the first section of [Siu and Yang 1981], the authors did not require any negativity. We also see that in Section 2, we do not really need any negativity except when we apply the formula from Lemma 6 in the Section 3.

In the first section of [Siu and Yang 1981], they also consider the coordinate in which $R_{1\bar{1}1\bar{1}}$ achieves the maximum instead of the minimum. We let $C = R_{1\bar{1}2\bar{2}}$ from the earlier sections and C_1 be the bisectional curvature for the maximal case. Then

$$K_{\min} + C = K_{\max} + C_1$$

is the Einstein constant Q,

$$C_1 - C = K_{\min} - K_{\max} = -\tfrac{1}{2}(A+B)$$

and

$$C_1 = C - \tfrac{1}{2}(A+B) = \tfrac{1}{2}(R_{1\bar{1}1\bar{1}} - B) = \tfrac{1}{2}\big(Q - C_1 - \tfrac{1}{2}(A+B) - B\big).$$

Therefore

$$3C_1 = Q - \tfrac{1}{2}(A+B) - B \le 0,$$

always. Also, $C_1 = 0$ implies that $A = B = Q = 0$.

The constant term in Lemma 6 is

$$\begin{aligned}
3\big[(3B-A)C - B(3A-B)\big] &= 3\big[(3B-A)\big(C_1 + \tfrac{1}{2}(A+B)\big) - B(3A-B)\big] \\
&= \tfrac{3}{2}[2\Psi C_1 - (A-B)(A+5B)] \\
&\le 0,
\end{aligned}$$

always. Therefore, we have the same result only if $Q \le 0$, unless $C_1 = 0$. As above if $C_1 = 0$ we have $A = B = 0$, then $C = 0$ and therefore $K_{\min} = Q = 0$. The manifold is flat.

Thus we conclude the general case. One might conjecture that our theorem is also true in the higher dimensional cases.

Remark 10. Notice that this generalization basically covers the results in [Polombo 1988; Polombo 1992] for the Kähler–Einstein case (see Corollary on page 398 of [Polombo 1992]). See also [Derdziński 1983, p. 415, Proposition 2] for the W^+ for a Kähler surface. One might ask whether our result could be generalized to the Riemannian manifolds with closed half Weyl curvature tensors. This is out of the scope of this paper although a similar result is true, i.e., $\lambda_2 \leq 1$ at every point. To make the relation between this paper and [Polombo 1988; Polombo 1992] clearer to the reader, we mention that any one of the half Weyl tensors is harmonic if and only if it is closed since the tensor is dual to either itself or the negative of itself. Remark (i) in [Polombo 1992, p. 397] notes that if M is Riemannian–Einstein, then the second Bianchi identity says that the half Weyl tensors are closed (see also [Derdziński 1983] page 408 formula (9) and page 411 Remark 1).

Acknowledgements

I thank the referee for useful comments and encouragements, Professors Poon, Wong and the Department of Mathematics, University of California at Riverside for their support. I thank Professor Hong Cang Yang for showing me his work when I was a graduate student in Berkeley. I also thank Professor Paul Yang for telling me of the article [Polombo 1988].

References

[Chen et al. 2011] D. Chen, Y. Hong, and H. Yang, "Kähler–Einstein surface and symmetric space", *Sci. China Math.* **54**:12 (2011), 2627–2634. MR Zbl

[Derdziński 1983] A. Derdziński, "Self-dual Kähler manifolds and Einstein manifolds of dimension four", *Compositio Math.* **49**:3 (1983), 405–433. MR Zbl

[Hong et al. 1988] Y. Hong, Z. D. Guan, and H. C. Yang, "A remark on Kähler–Einstein manifolds", *Acta Math. Sinica* **31**:5 (1988), 595–602. MR Zbl

[Polombo 1988] A. Polombo, "Condition d'Einstein et courbure négative en dimension 4", *C. R. Acad. Sci. Paris Sér. I Math.* **307**:12 (1988), 667–670. MR Zbl

[Polombo 1992] A. Polombo, "De nouvelles formules de Weitzenböck pour des endomorphismes harmoniques, applications géométriques", *Ann. Sci. École Norm. Sup.* (4) **25**:4 (1992), 393–428. MR Zbl

[Siu and Yang 1981] Y. T. Siu and P. Yang, "Compact Kähler–Einstein surfaces of nonpositive bisectional curvature", *Invent. Math.* **64**:3 (1981), 471–487. MR Zbl

Received December 8, 2014. Revised December 1, 2016.

Daniel Guan
Department of Mathematics
The University of California at Riverside
Riverside, CA 92521
United States
zguan@math.ucr.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.355

EFFECTIVE LOWER BOUNDS FOR $L(1, \chi)$ VIA EISENSTEIN SERIES

PETER HUMPHRIES

We give effective lower bounds for $L(1, \chi)$ via Eisenstein series on $\Gamma_0(q) \setminus \mathbb{H}$. The proof uses the Maass–Selberg relation for truncated Eisenstein series and sieve theory in the form of the Brun–Titchmarsh inequality. The method follows closely the work of Sarnak in using Eisenstein series to find effective lower bounds for $\zeta(1+it)$.

1. Introduction

Let q be a positive integer, let χ be a Dirichlet character modulo q, and let

$$L(s, \chi) := \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$$

be the associated Dirichlet L-function, which converges absolutely for $\Re(s) > 1$ and extends holomorphically to the entire complex plane except when χ is principal, in which case there is a simple pole at $s = 1$. It is well known that Dirichlet's theorem on the infinitude of primes in arithmetic progressions is equivalent to showing that $L(1, \chi) \neq 0$ for every Dirichlet character χ modulo q. Of further interest is obtaining lower bounds for $L(1, \chi)$ in terms of q. By complex analytic means [Montgomery and Vaughan 2007, Theorems 11.4 and 11.11], one can show that if χ is complex, then

$$|L(1, \chi)| \gg \frac{1}{\log q},$$

while

$$L(1, \chi) \gg \frac{1}{\sqrt{q}}$$

if χ is quadratic. In both cases, the implicit constants are effective. For quadratic characters, the Landau–Siegel theorem states that

$$L(1, \chi) \gg_{\varepsilon} q^{-\varepsilon}$$

MSC2010: primary 11M20; secondary 11M36.
Keywords: Dirichlet L-function, lower bounds, Eisenstein series.

for all $\varepsilon > 0$ [Montgomery and Vaughan 2007, Theorem 11.14], though this estimate is ineffective due to the possible existence of a Landau–Siegel zero of $L(s, \chi)$.

In this article, we give a novel proof of effective lower bounds for $L(1, \chi)$, albeit in slightly weaker forms.

Theorem 1.1. *Let $q \geq 2$ be a positive integer, and let χ be a primitive character modulo q. If χ is complex, then*

$$|L(1, \chi)| \gg \frac{1}{(\log q)^3},$$

while

$$L(1, \chi) \gg \frac{1}{\sqrt{q}(\log q)^2}$$

if χ is quadratic. In both cases, the implicit constants are effective.

Our proof of Theorem 1.1 makes use of the fact that $L(s, \chi)$ appears in the Fourier expansion of an Eisenstein series associated to χ on $\Gamma_0(q) \backslash \mathbb{H}$, together with sieve theory — specifically the Brun–Titchmarsh inequality — to find these lower bounds. As is well-known, improving the constant in the Brun–Titchmarsh inequality is essentially equivalent the nonexistence of Landau–Siegel zeroes; it is for this same reason that the lower bounds in Theorem 1.1 are weak for quadratic characters, as we discuss in Remark 4.7.

That one can use Eisenstein series to prove nonvanishing of L-functions is well known, first appearing in unpublished work of Selberg, but such methods were not shown to give good effective lower bounds for L-functions on the line $\Re(s) = 1$ until the work of Sarnak [2004]. He showed that

$$|\zeta(1+it)| \gg \frac{1}{(\log |t|)^3}$$

for $|t| > 1$ by exploiting the inhomogeneous form of the Maass–Selberg relation for the Eisenstein series $E(z, s)$ for the group $\mathrm{SL}_2(\mathbb{Z})$.

More precisely, for $t > 1$, Sarnak studied the integral

$$\mathcal{I} := \int_{1/t}^{\infty} \int_0^1 |\zeta(1+2it)|^2 \left|\Lambda^t\left(z, \frac{1}{2}+it\right)\right|^2 \frac{dx\,dy}{y^2},$$

involving a truncated Eisenstein series $\Lambda^T E(z, s)$ and found an upper bound up to a scalar multiple for this integral of the form

$$t(\log t)^2 |\zeta(1+2it)|$$

via the Maass–Selberg relation, and a lower bound up to a scalar multiple of the form

$$\frac{1}{t}\sum_{\frac{t^2}{8}\le m\le\frac{t^2}{4}}|\sigma_{-2it}(m)|^2$$

via Parseval's identity, where

$$\sigma_{-2it}(m):=\sum_{d|m}d^{-2it}.$$

By restricting the summation over m to primes, Sarnak was able to use sieve theory to show that

$$\sum_{\frac{t^2}{8}\le p\le\frac{t^2}{4}}|\sigma_{-2it}(p)|^2\gg\frac{t^2}{\log t},$$

from which the result follows. Indeed, the use of sieve theory to prove lower bounds for $\zeta(1+it)$ (and also $L(1+it,\chi)$) has its roots in work of Balasubramanian and Ramachandra [1976].

The chief novelty of Sarnak's work is to use the Maass–Selberg relation to obtain effective lower bounds for $\zeta(1+it)$; more precisely, it is the inhomogeneous nature of the Fourier expansion of the Eisenstein series $E(z,s)$, whose constant term involves $\zeta(2s-1)/\zeta(2s)$ and whose nonconstant terms involve $1/\zeta(2s)$. This method has been generalized by Gelbart and Lapid [2006] to determine effective lower bounds on the line $\Re(s)=1$ for L-functions associated to automorphic representations on arbitrary reductive groups over number fields, albeit with the lower bound being in the weaker form $C|t|^{-n}$ for some constants C, n depending on the L-function, for Gelbart and Lapid make no use of sieve theory in this generalized setting. More recently, Goldfeld and Li [2016] have succeeded in generalizing Sarnak's method to show that

$$|L(1+it,\pi\times\widetilde{\pi})|\gg_\pi\frac{1}{(\log|t|)^3}$$

for any cuspidal automorphic representation π of $\mathrm{GL}_n(\mathbb{A}_\mathbb{Q})$ that is unramified and tempered at every place, with the implicit constant in the lower bound dependent on π.

All three of these results give lower bounds for L-functions on the line $\Re(s)=1$ in the height aspect, namely in terms of t. In this article, we give the first example of Sarnak's method being used to give lower bounds for L-functions on the line $\Re(s)=1$ in the level aspect, namely in terms of q.

2. Eisenstein series

We introduce Eisenstein series for the group $\Gamma_0(q)$ associated to a primitive Dirichlet character χ modulo q. Standard references for this material are [Deshouillers and Iwaniec 1982], [Duke et al. 2002], and [Iwaniec 2002].

Cusps. Let $\mathbb{H}$ be the upper half plane, upon which $\mathrm{SL}_2(\mathbb{R})$ acts via Möbius transformations $\gamma z = (az+b)/(cz+d)$ for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ and $z \in \mathbb{H}$. Let q be a positive integer, and let $\mathfrak{a}$ be a cusp of $\Gamma_0(q) \backslash \mathbb{H}$, where

$$\Gamma_0(q) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : c \equiv 0 \pmod{q} \right\},$$

and we denote the stabilizer of $\mathfrak{a}$ by $\Gamma_{\mathfrak{a}} := \{\gamma \in \Gamma_0(q) : \gamma\mathfrak{a} = \mathfrak{a}\}$. This subgroup of $\Gamma_0(q)$ is generated by two parabolic elements $\pm\gamma_{\mathfrak{a}}$, where

$$\gamma_{\mathfrak{a}} := \sigma_{\mathfrak{a}} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \sigma_{\mathfrak{a}}^{-1},$$

and the scaling matrix $\sigma_{\mathfrak{a}} \in \mathrm{SL}_2(\mathbb{R})$ is such that

$$\sigma_{\mathfrak{a}}\infty = \mathfrak{a}, \qquad \sigma_{\mathfrak{a}}^{-1}\Gamma_\infty\sigma_{\mathfrak{a}} = \Gamma_\infty,$$

where

$$\Gamma_\infty := \left\{ \pm \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \in \Gamma_0(q) : n \in \mathbb{Z} \right\}$$

is the stabilizer of the cusp at infinity. The scaling matrix is unique up to translation on the right.

Let χ be a primitive character modulo q. A cusp $\mathfrak{a}$ of $\Gamma_0(q) \backslash \mathbb{H}$ is said to be singular with respect to χ if $\chi(\gamma_{\mathfrak{a}}) = 1$, where $\chi(\gamma) := \chi(d)$ for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(q)$. As χ is primitive, any singular cusp is equivalent to $1/v$ for a single unique divisor v of q satisfying $vw = q$ and $(v, w) = 1$, where w is the width of the cusp; when $v = q$, this cusp is equivalent to the cusp at infinity, while when $v = 1$, the cusp is equivalent to the cusp at zero. Note that if $q = 1$, so that χ is the trivial character, there is merely a single equivalence class of cusps, namely the cusp at infinity.

The scaling matrix $\sigma_{\mathfrak{a}} \in \mathrm{SL}_2(\mathbb{R})$ for a singular cusp $\mathfrak{a} \sim 1/v$, $v \neq q$, can be chosen to be

$$\sigma_{\mathfrak{a}} := \begin{pmatrix} \sqrt{w} & 0 \\ v\sqrt{w} & 1/\sqrt{w} \end{pmatrix},$$

while for the cusp at infinity, we simply take σ_∞ to be the identity.

The Bruhat decomposition for $\sigma_{\mathfrak{a}}^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}}$ [Iwaniec 2002, Theorem 2.7] states that

$$\sigma_{\mathfrak{a}}^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}} = \delta_{\mathfrak{a}\mathfrak{b}}\Omega_\infty \sqcup \bigsqcup_{c>0} \bigsqcup_{d \,(\mathrm{mod}\, c)} \Omega_{d/c},$$

where $\delta_{\mathfrak{a}\mathfrak{b}} = 1$ if $\mathfrak{a} \sim \mathfrak{b}$ and 0 otherwise, and

$$\Omega_\infty := \Gamma_\infty \omega_\infty, \qquad \omega_\infty = \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \in \sigma_{\mathfrak{a}}^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}},$$

$$\Omega_{d/c} := \Gamma_\infty \omega_{d/c} \Gamma_\infty, \qquad \omega_{d/c} = \begin{pmatrix} * & * \\ c & d \end{pmatrix} \in \sigma_{\mathfrak{a}}^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}} \quad \text{with } c > 0,$$

and c, d run over all real numbers such that $\sigma_{\mathfrak{a}}^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}}$ contains $\left(\begin{smallmatrix} * & * \\ c & d \end{smallmatrix}\right)$. In particular, for the cusp at infinity we have the Bruhat decomposition

$$\sigma_\infty^{-1}\Gamma_0(q)\sigma_\infty = \Gamma_\infty \sqcup \bigsqcup_{\substack{c=1 \\ c \equiv 0 \ (\mathrm{mod}\ q)}}^{\infty} \ \bigsqcup_{\substack{d \ (\mathrm{mod}\ c) \\ (c,d)=1}} \Gamma_\infty \begin{pmatrix} * & * \\ c & d \end{pmatrix} \Gamma_\infty.$$

For $\mathfrak{a} \sim \infty$ and $\mathfrak{b} \sim 1/v$ a nonequivalent singular cusp with $1 \le v < q$, v dividing q, $vw = q$, and $(v, w) = 1$, and for any $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(q)$, we have that

$$\sigma_\infty^{-1}\gamma\sigma_{\mathfrak{b}} = \begin{pmatrix} (a+bv)\sqrt{w} & b/\sqrt{w} \\ (c+dv)\sqrt{w} & d/\sqrt{w} \end{pmatrix},$$

and so

$$\text{(2.1)} \quad \sigma_\infty^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}} = \left\{ \begin{pmatrix} a\sqrt{w} & b/\sqrt{w} \\ c\sqrt{w} & d/\sqrt{w} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}), \right.$$
$$\left. c \equiv 0 \ (\mathrm{mod}\ v),\ d \equiv c/v \ (\mathrm{mod}\ w),\ (c,d) = 1,\ (c,w) = 1 \right\}.$$

So the Bruhat decomposition in this case can be explicitly written in the form

$$\text{(2.2)} \qquad \sigma_\infty^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}} = \bigsqcup_{\substack{c=1 \\ (c,w)=1 \\ c \equiv 0 \ (\mathrm{mod}\ v)}}^{\infty} \ \bigsqcup_{\substack{d \ (\mathrm{mod}\ cw) \\ (cw,d)=1 \\ d \equiv c/v \ (\mathrm{mod}\ w)}} \Gamma_\infty \begin{pmatrix} * & * \\ c\sqrt{w} & d/\sqrt{w} \end{pmatrix} \Gamma_\infty.$$

Eisenstein series. Given a primitive Dirichlet character χ modulo q and a singular cusp $\mathfrak{a}$ of $\Gamma_0(q) \setminus \mathbb{H}$, we define the Eisenstein series $E_{\mathfrak{a}}(z, s, \chi)$ for $z \in \mathbb{H}$ and $\Re(s) > 1$ by

$$E_{\mathfrak{a}}(z, s, \chi) := \sum_{\gamma \in \Gamma_{\mathfrak{a}} \setminus \Gamma_0(q)} \bar{\chi}(\gamma) j_{\sigma_{\mathfrak{a}}^{-1}\gamma}(z)^{-\kappa} \Im(\sigma_{\mathfrak{a}}^{-1}\gamma z)^s,$$

where $\kappa \in \{0, 1\}$ is such that $\chi(-1) = (-1)^\kappa$, and for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$,

$$j_\gamma(z) := \frac{cz+d}{|cz+d|} = e^{i \arg(cz+d)}.$$

The Eisenstein series associated to a singular cusp $\mathfrak{a}$ is independent of the choice of representative of $\mathfrak{a}$ and of the scaling matrix $\sigma_{\mathfrak{a}}$. For fixed $z \in \mathbb{H}$, the Eisenstein series $E_{\mathfrak{a}}(z, s, \chi)$ converges absolutely for $\Re(s) > 1$ and extends meromorphically to the entire complex plane with no poles on the closed right half-plane $\Re(s) \geq \frac{1}{2}$ except at $s = 1$ when $q = 1$, so that χ is the trivial character.

For any $z \in \mathbb{H}$ and $\gamma_1, \gamma_2 \in \mathrm{SL}_2(\mathbb{R})$, the j-factor satisfies the cocycle relation

$$j_{\gamma_1\gamma_2}(z) = j_{\gamma_2}(z) j_{\gamma_1}(\gamma_2 z), \tag{2.3}$$

while the Eisenstein series satisfies the automorphy condition

$$E_{\mathfrak{a}}(\gamma z, s, \chi) = \chi(\gamma) j_\gamma(z)^\kappa E_{\mathfrak{a}}(z, s, \chi) \tag{2.4}$$

for any $\gamma \in \Gamma_0(q)$.

For any singular cusps $\mathfrak{a}$, $\mathfrak{b}$ of $\Gamma_0(q)$, one can show using the Bruhat decomposition that there exists a function $\varphi_{\mathfrak{a}\mathfrak{b}}(s, \chi)$ such that the constant term in the Fourier expansion for the function $j_{\sigma_{\mathfrak{b}}}(z)^{-\kappa} E_{\mathfrak{a}}(\sigma_{\mathfrak{b}} z, s, \chi)$ is

$$c_{\mathfrak{a}\mathfrak{b}}(z, s, \chi) := \int_0^1 j_{\sigma_{\mathfrak{b}}}(z)^{-\kappa} E_{\mathfrak{a}}(\sigma_{\mathfrak{b}} z, s, \chi)\, dx = \delta_{\mathfrak{a}\mathfrak{b}} y^s + \varphi_{\mathfrak{a}\mathfrak{b}}(s, \chi) y^{1-s}.$$

The functions $\varphi_{\mathfrak{a}\mathfrak{b}}(s, \chi)$ are the entries of the scattering matrix associated to χ. We will calculate $\varphi_{\mathfrak{a}\mathfrak{b}}(s, \chi)$ when $\mathfrak{a} \sim \infty$ for each nonsingular cusp $\mathfrak{b}$ of $\Gamma_0(q)$ with respect to χ, and also find the rest of the Fourier coefficients of $E_\infty(z, s, \chi)$.

Fourier expansion of $E_\infty(z, s, \chi)$.

Lemma 2.5. *Let χ be a primitive character modulo q. For $m \neq 0$ and $c \equiv 0 \pmod{q}$,*

$$\sum_{\substack{d \,(\mathrm{mod}\, c)\\ (c,d)=1}} \chi(d) e\left(\frac{md}{c}\right) = \chi(\operatorname{sgn}(m)) \tau(\chi) \sum_{d \mid \left(|m|, \frac{c}{q}\right)} d \overline{\chi}\left(\frac{|m|}{d}\right) \chi\left(\frac{c}{dq}\right) \mu\left(\frac{c}{dq}\right).$$

Here, as usual, we define $e(x) := e^{2\pi i x}$ for $x \in \mathbb{R}$.

Proof. For m positive, this is [Miyake 1989, Lemma 3.1.3]. The result for m negative follows by replacing m with $|m|$ and χ with $\overline{\chi}$, then taking complex conjugates of both sides and using the fact that $\overline{\tau(\overline{\chi})} = \chi(-1)\tau(\chi)$. ☐

Proposition 2.6 (cf. [Iwaniec 2002, Theorem 3.4]). *The Eisenstein series associated to the cusp at infinity has the Fourier expansion*

$$E_\infty(z, s, \chi) = y^s + \varphi_{\infty\infty}(s, \chi) y^{1-s} + \sum_{\substack{m=-\infty\\ m \neq 0}}^{\infty} \rho_\infty(m, s, \chi) W_{\operatorname{sgn}(m)\frac{\kappa}{2}, s-\frac{1}{2}}(4\pi |m| y) e(mx),$$

where $W_{\alpha,\nu}(y)$ is the Whittaker function,

$$\varphi_{\infty\infty}(s,\chi)=\begin{cases}\sqrt{\pi}\dfrac{\Gamma\left(s-\frac{1}{2}\right)}{\Gamma(s)}\dfrac{\zeta(2s-1)}{\zeta(2s)} & \text{if } q=1,\\ 0 & \text{if } q\geq 2,\end{cases}$$

and for $m\neq 0$,

$$\rho_{\infty}(m,s,\chi)=\frac{\chi(\operatorname{sgn}(m))i^{-\kappa}\tau(\overline{\chi})\pi^{s}|m|^{s-1}}{q^{2s}\Gamma\left(s+\operatorname{sgn}(m)\frac{\kappa}{2}\right)L(2s,\overline{\chi})}\sigma_{1-2s}(|m|,\chi),$$

where $\tau(\chi)$ is the Gauss sum of χ and

$$\sigma_s(m,\chi):=\sum_{d\mid m}d^{s}\chi\left(\frac{m}{d}\right).$$

Note in particular that if $\kappa=0$, so that χ is even, the Whittaker function is simply

$$W_{0,s-\frac{1}{2}}(4\pi|m|y)=\sqrt{4|m|y}\,K_{s-\frac{1}{2}}(2\pi|m|y),$$

where $K_\nu(y)$ is the K-Bessel function. On the other hand, if $\kappa=1$, so that χ is odd, and we set $s=\frac{1}{2}$, then

$$W_{\operatorname{sgn}(m)\frac{\kappa}{2},0}(4\pi|m|y)=\begin{cases}\sqrt{4\pi|m|y}\,e^{-2\pi|m|y} & \text{if } m>0,\\ \sqrt{4\pi|m|y}\,e^{2\pi|m|y}\int_{4\pi|m|y}^{\infty}e^{-u}/u\,du & \text{if } m<0.\end{cases}$$

Proof. Via the Bruhat decomposition (2.2), $E_\infty(z,s,\chi)$ is equal to

$$y^{s}+\sum_{\substack{c=1\\ c\equiv 0\ (\mathrm{mod}\ q)}}^{\infty}\ \sum_{\substack{d\ (\mathrm{mod}\ c)\\ (c,d)=1}}\overline{\chi}(d)\sum_{n=-\infty}^{\infty}\left(\frac{c(z+n)+d}{|c(z+n)+d|}\right)^{-\kappa}\frac{y^{s}}{|c(z+n)+d|^{2s}}.$$

So if $m=0$, the zeroth Fourier coefficient of $E_\infty(z,s,\chi)$ is

$$y^{s}+\sum_{\substack{c=1\\ c\equiv 0\ (\mathrm{mod}\ q)}}^{\infty}\ \sum_{\substack{d\ (\mathrm{mod}\ c)\\ (c,d)=1}}\overline{\chi}(d)\int_{-\infty}^{\infty}\left(\frac{cz+d}{|cz+d|}\right)^{-\kappa}\frac{y^{s}}{|cz+d|^{2s}}\,dx$$
$$=y^{s}+y^{1-s}\int_{-\infty}^{\infty}\left(\frac{t+i}{|t+i|}\right)^{-\kappa}\frac{1}{|t+i|^{2s}}\,dt\sum_{\substack{c=1\\ c\equiv 0\ (\mathrm{mod}\ q)}}^{\infty}\frac{1}{c^{2s}}\sum_{\substack{d\ (\mathrm{mod}\ c)\\ (c,d)=1}}\overline{\chi}(d)$$

by the change of variables $x\mapsto yt-d/c$. From [Gradshteyn and Ryzhik 2007, (8.381.1)], we have that

$$\int_{-\infty}^{\infty}\left(\frac{t+i}{|t+i|}\right)^{-\kappa}\frac{1}{|t+i|^{2s}}\,dt=i^{-\kappa}\sqrt{\pi}\frac{\Gamma\left(\frac{1}{2}(2s-1+\kappa)\right)}{\Gamma\left(\frac{1}{2}(2s+\kappa)\right)},$$

while for $c \equiv 0 \pmod{q}$, the fact that χ is primitive implies that

$$\sum_{\substack{c=1 \\ c\equiv 0 \ (\mathrm{mod}\ q)}}^{\infty} \frac{1}{c^{2s}} \sum_{\substack{d \ (\mathrm{mod}\ c) \\ (c,d)=1}} \overline{\chi}(d) = \begin{cases} \displaystyle\sum_{c=1}^{\infty} \frac{\varphi(c)}{c^{2s}} = \frac{\zeta(2s-1)}{\zeta(2s)} & \text{if } q = 1, \\ 0 & \text{if } q \geq 2. \end{cases}$$

If $m \neq 0$, on the other hand, then the m-th Fourier coefficient is

$$\sum_{\substack{c=1 \\ c\equiv 0 \ (\mathrm{mod}\ q)}}^{\infty} \sum_{\substack{d \ (\mathrm{mod}\ c) \\ (c,d)=1}} \overline{\chi}(d) \int_{-\infty}^{\infty} \left(\frac{cz+d}{|cz+d|}\right)^{-\kappa} \frac{y^s}{|cz+d|^{2s}} e(-mx)\, dx$$

$$= y^{1-s} \int_{-\infty}^{\infty} \left(\frac{t+i}{|t+i|}\right)^{-\kappa} \frac{e(-myt)}{|t+i|^{2s}}\, dt \sum_{\substack{c=1 \\ c\equiv 0 \ (\mathrm{mod}\ q)}}^{\infty} \frac{1}{c^{2s}} \sum_{\substack{d \ (\mathrm{mod}\ c) \\ (c,d)=1}} \overline{\chi}(d) e\left(\frac{md}{c}\right)$$

again by the change of variables $x \mapsto yt - d/c$. Moreover, [Gradshteyn and Ryzhik 2007, (3.384.9)] implies that

$$\int_{-\infty}^{\infty} \left(\frac{t+i}{|t+i|}\right)^{-\kappa} \frac{e(-myt)}{|t+i|^{2s}}\, dt = \frac{i^{-\kappa} \pi^s |m|^{s-1} y^{s-1}}{\Gamma\left(s + \operatorname{sgn}(m)\frac{\kappa}{2}\right)} W_{\operatorname{sgn}(m)\frac{\kappa}{2}, s-\frac{1}{2}}(4\pi|m|y),$$

and via Lemma 2.5,

$$\sum_{\substack{c=1 \\ c\equiv 0 \ (\mathrm{mod}\ q)}}^{\infty} \frac{1}{c^{2s}} \sum_{\substack{d \ (\mathrm{mod}\ c) \\ (c,d)=1}} \overline{\chi}(d) e\left(\frac{md}{c}\right)$$

$$= \chi(\operatorname{sgn}(m)) \tau(\overline{\chi}) \sum_{d \mid |m|} d \chi\left(\frac{|m|}{d}\right) \sum_{\substack{c=1 \\ c\equiv 0 \ (\mathrm{mod}\ dq)}}^{\infty} \frac{\overline{\chi}\left(\frac{c}{dq}\right) \mu\left(\frac{c}{dq}\right)}{c^{2s}}$$

$$= \chi(\operatorname{sgn}(m)) \frac{\tau(\overline{\chi})}{q^{2s}} \sum_{d \mid |m|} d^{1-2s} \chi\left(\frac{|m|}{d}\right) \sum_{n=1}^{\infty} \frac{\overline{\chi}(n)\mu(n)}{n^{2s}}$$

$$= \chi(\operatorname{sgn}(m)) \frac{\tau(\overline{\chi})}{q^{2s} L(2s, \overline{\chi})} \sigma_{1-2s}(|m|, \chi)$$

where we have let $c = dqn$. We thereby obtain the desired identity. □

Proposition 2.7. *Suppose that $q \geq 2$. Then $\varphi_{\infty\mathfrak{b}}(s, \chi)$ vanishes unless $\mathfrak{b} \sim 1$, in which case*

$$\varphi_{\infty 1}(s, \chi) = \frac{\overline{\tau(\chi)}}{q^s} \frac{\Lambda(2-2s, \chi)}{\Lambda(2s, \overline{\chi})}, \tag{2.8}$$

where

$$\Lambda(s,\chi) := \left(\frac{\pi}{q}\right)^{-\frac{s+\kappa}{2}} \Gamma\left(\frac{s+\kappa}{2}\right) L(s,\chi), \tag{2.9}$$

is the completed Dirichlet L-function. In particular,

$$\left|\varphi_{\infty 1}\left(\tfrac{1}{2}+it,\chi\right)\right| = 1. \tag{2.10}$$

Proof. The fact that $\varphi_{\infty\mathfrak{b}}(s,\chi) = 0$ when $\mathfrak{b}$ is the cusp at infinity follows from Proposition 2.6. For the entries of the scattering matrix at other cusps, we use (2.3) to write

$$E_{\mathfrak{a}}(\sigma_{\mathfrak{b}} z, s, \chi) = j_{\sigma_{\mathfrak{b}}}(z)^{\kappa} \sum_{\gamma \in \Gamma_\infty \backslash \sigma_{\mathfrak{a}}^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}}} \overline{\chi}(\sigma_{\mathfrak{a}}\gamma\sigma_{\mathfrak{b}}^{-1}) j_\gamma(z)^{-\kappa} \Im(\gamma z)^s.$$

The singular cusp $\mathfrak{b}$ is equivalent to $1/v$ for some divisor v of q with $v<q,\ vw=q$, and $(v,w)=1$. Given a matrix

$$\gamma = \begin{pmatrix} a\sqrt{w} & b/\sqrt{w} \\ c\sqrt{w} & d/\sqrt{w} \end{pmatrix}$$

in $\sigma_\infty^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}}$ as in (2.1), we have that

$$\sigma_\infty \gamma \sigma_{\mathfrak{b}}^{-1} = \begin{pmatrix} a-bv & b \\ c-dv & d \end{pmatrix},$$

and so as $d \equiv c/v \pmod{w}$,

$$\overline{\chi}(\sigma_\infty\gamma\sigma_{\mathfrak{b}}^{-1}) = \overline{\chi_v}(d)\,\overline{\chi_w}\left(\frac{c}{v}\right),$$

where we have decomposed the primitive character χ modulo q into the product of primitive characters χ_v modulo v and χ_w modulo w. From this and (2.2), we see that $j_{\sigma_{\mathfrak{b}}}(z)^{-\kappa} E_\infty(\sigma_{\mathfrak{b}} z, s, \chi)$ is equal to

$$\sum_{\substack{c=1 \\ (c,w)=1 \\ c\equiv 0 \pmod{v}}}^{\infty} \overline{\chi_w}\left(\frac{c}{v}\right) \sum_{\substack{d \pmod{cw} \\ (cw,d)=1 \\ d\equiv c/v \pmod{w}}} \overline{\chi_v}(d) \times \sum_{n=-\infty}^{\infty} \left(\frac{c(z+n)\sqrt{w}+d/\sqrt{w}}{|c(z+n)\sqrt{w}+d/\sqrt{w}|}\right)^{-\kappa} \frac{y^s}{|c(z+n)\sqrt{w}+d/\sqrt{w}|^{2s}},$$

and so integrating from 0 to 1 with respect to x, making the change of variables $x \mapsto yt - d/(cw)$, and dividing by y^{1-s} yields

$$\varphi_{\infty\mathfrak{b}}(s,\chi) = \frac{1}{w^s}\int_{-\infty}^{\infty}\left(\frac{t+i}{|t+i|}\right)^{-\kappa}\frac{1}{|t+i|^{2s}}\,dt \sum_{\substack{c=1\\(c,w)=1\\c\equiv 0 \pmod{v}}}^{\infty}\frac{\overline{\chi_w}(c/v)}{c^{2s}}\sum_{\substack{d \pmod{cw}\\(cw,d)=1\\d\equiv c/v \pmod{w}}}\overline{\chi_v}(d).$$

From [Gradshteyn and Ryzhik 2007, (8.381.1)], the integral is equal to

$$i^{-\kappa}\sqrt{\pi}\frac{\Gamma\left(\frac{1}{2}(2s-1+\kappa)\right)}{\Gamma\left(\frac{1}{2}(2s+\kappa)\right)}.$$

To evaluate the sum over d, we write $d = \bar{v}c + wd'$, where $\bar{v}v \equiv 1 \pmod{w}$ and $(d', c) = 1$. This allows us to replace the sum over d with a sum over d' modulo c with $(c, d') = 1$, so that

$$\sum_{\substack{d \pmod{cw}\\(cw,d)=1\\d\equiv c/v \pmod{w}}}\overline{\chi_v}(d) = \overline{\chi_v}(w)\sum_{\substack{d' \pmod{c}\\(c,d')=1}}\overline{\chi_v}(d')$$

by the fact that $c \equiv 0 \pmod{v}$.

If $\overline{\chi_v}$ is nonprincipal, this sum vanishes, and as χ is a primitive character, $\overline{\chi_v}$ can only be the principal character if $v = 1$; consequently, $\varphi_{\infty\mathfrak{b}}(s,\chi)$ vanishes if $\mathfrak{b}$ is inequivalent to the cusp at 1.

If $\mathfrak{b} \sim 1$, so that $v = 1$ and $w = q$, then this sum over d' is merely $\varphi(c)$, and so

$$\sum_{\substack{c=1\\(c,w)=1\\c\equiv 0 \pmod{v}}}^{\infty}\frac{\overline{\chi_w}(c/v)}{c^{2s}}\sum_{\substack{d \pmod{cw}\\(cw,d)=1\\d\equiv c/v \pmod{w}}}\overline{\chi_v}(d) = \sum_{c=1}^{\infty}\frac{\varphi(c)\overline{\chi}(c)}{c^{2s}} = \frac{L(2s-1,\overline{\chi})}{L(2s,\overline{\chi})}.$$

Using the definition of the completed Dirichlet L-function together with the fact that it satisfies the functional equation

$$\Lambda(s,\chi) = \frac{\tau(\chi)}{i^{\kappa}\sqrt{q}}\Lambda(1-s,\overline{\chi}),$$

we see that we may write

$$\varphi_{\infty 1}(s,\chi) = \frac{i^{-\kappa}}{q^{s-\frac{1}{2}}}\frac{\Lambda(2s-1,\overline{\chi})}{\Lambda(2s,\overline{\chi})} = \frac{\overline{\tau(\chi)}}{q^s}\frac{\Lambda(2-2s,\chi)}{\Lambda(2s,\overline{\chi})}.$$

As $\overline{\Lambda(s,\chi)} = \Lambda(\bar{s},\overline{\chi})$ and $|\tau(\chi)| = \sqrt{q}$, the result follows. □

3. Maass–Selberg relation

For $z \in \mathbb{H}$ and $T \geq 1$, we define the truncated Eisenstein series

$$\Lambda^T E_{\mathfrak{a}}(z,s,\chi) := E_{\mathfrak{a}}(z,s,\chi) - \sum_{\mathfrak{c}} \sum_{\substack{\gamma \in \Gamma_{\mathfrak{c}} \backslash \Gamma_0(q) \\ \Im(\sigma_{\mathfrak{c}}^{-1}\gamma z) > T}} \overline{\chi}(\gamma) j_{\sigma_{\mathfrak{c}}^{-1}\gamma}(z)^{-\kappa} c_{\mathfrak{a}\mathfrak{c}}(\sigma_{\mathfrak{c}}^{-1}\gamma z, s, \chi), \tag{3.1}$$

where the summation over $\mathfrak{c}$ is over all singular cusps of $\Gamma_0(q) \backslash \mathbb{H}$. It is not difficult to see that $\Lambda^T E_{\mathfrak{a}}(z,s,\chi)$ satisfies the automorphy condition

$$\Lambda^T E_{\mathfrak{a}}(\gamma z, s, \chi) = \chi(\gamma) j_\gamma(z)^\kappa \Lambda^T E_{\mathfrak{a}}(z,s,\chi) \tag{3.2}$$

for any $\gamma \in \Gamma_0(q)$. We will show that, unlike $E_{\mathfrak{a}}(z,s,\chi)$, the function $\Lambda^T E_{\mathfrak{a}}(z,s,\chi)$ is square-integrable on $\Gamma_0(q) \backslash \mathbb{H}$, and give an explicit expression for the resulting integral.

Lemma 3.3. *Let $\mathfrak{b}$ and $\mathfrak{c}$ be singular cusps of $\Gamma_0(q) \backslash \mathbb{H}$, and let $\gamma \in \sigma_{\mathfrak{c}}^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}}$. Then for any $z = x + iy \in \mathbb{H}$, we have that $\Im(z)\Im(\gamma z) \leq 1$ if $\mathfrak{b}$ and $\mathfrak{c}$ are inequivalent or if $\mathfrak{b}$ and $\mathfrak{c}$ are equivalent but $\gamma \notin \Gamma_\infty \omega_\infty$. If $\mathfrak{b}$ and $\mathfrak{c}$ are equivalent and $\gamma \in \Gamma_\infty\omega_\infty$, then $\Im(\gamma z) = \Im(z)$.*

Proof. We deal with the cases where neither $\mathfrak{b}$ nor $\mathfrak{c}$ are equivalent to the cusp at infinity; when $\mathfrak{b} \sim \infty$ or $\mathfrak{c} \sim \infty$, the proof is similar but simpler. Let $\mathfrak{b} \sim 1/v$ and $\mathfrak{c} \sim 1/v'$, $1 \leq v, v' < q$, with w, w' such that $vw = v'w' = q$. For $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(q)$, we have that

$$\sigma_{\mathfrak{c}}^{-1} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \sigma_{\mathfrak{b}} = \begin{pmatrix} (a+bv)\sqrt{w/w'} & b/\sqrt{w'w} \\ (c - av' + dv - bv'v)\sqrt{w'w} & (d - bv')\sqrt{w'/w} \end{pmatrix}.$$

So for

$$\gamma = \begin{pmatrix} * & * \\ C\sqrt{w'w} & D\sqrt{w'/w} \end{pmatrix} \in \sigma_{\mathfrak{c}}^{-1}\Gamma_0(q)\sigma_{\mathfrak{b}},$$

where $C = c - av' + dv - bv'v$ and $D = d - bv'$ are integers, we have that

$$\Im(\gamma z) = \frac{1}{w'w} \frac{y}{(Cx + Dw^{-1})^2 + C^2y^2}.$$

By the Bruhat decomposition, if $\mathfrak{b}$ and $\mathfrak{c}$ are inequivalent, then $C\sqrt{w'w}$ must be nonzero, and so $C^2 \geq 1$. In particular, if $\mathfrak{b}$ and $\mathfrak{c}$ are inequivalent, then

$$\Im(z)\Im(\gamma z) \leq \frac{1}{w'w} \leq 1.$$

If $\mathfrak{b}$ and $\mathfrak{c}$ are equivalent and $\gamma \notin \Gamma_\infty\omega_\infty$, then again $C\sqrt{w'w} \neq 0$, and the same result holds. Finally, if $\mathfrak{b}$ and $\mathfrak{c}$ are equivalent and $\gamma \in \Gamma_\infty\omega_\infty$, then it is clear that $\Im(\gamma z) = \Im(z)$. $\square$

Corollary 3.4. *If $\Im(z) > T \geq 1$, then for any singular cusp $\mathfrak{b}$, we have that*

$$\Lambda^T E_{\mathfrak{a}}(\sigma_{\mathfrak{b}} z, s, \chi) = E_{\mathfrak{a}}(\sigma_{\mathfrak{b}} z, s, \chi) - j_{\sigma_{\mathfrak{b}}}(z)^{\kappa} c_{\mathfrak{a}\mathfrak{b}}(z, s, \chi).$$

Proof. From the definition of $\Lambda^T E_{\mathfrak{a}}(z, s, \chi)$ and (2.3), we must show that for any singular cusp $\mathfrak{c}$ and $\gamma \in \Gamma_{\mathfrak{c}} \backslash \Gamma_0(q)$ that the inequalities $\Im(z) > T$ and $\Im(\sigma_{\mathfrak{c}}^{-1} \gamma \sigma_{\mathfrak{b}} z) > T$ are simultaneously satisfied only when $\mathfrak{c} \sim \mathfrak{b}$ and $\gamma = \omega_\infty$. This is equivalent to showing that if $\gamma \in \Gamma_\infty \setminus \sigma_{\mathfrak{c}}^{-1} \Gamma_0(q) \sigma_{\mathfrak{b}}$ is such that $\Im(z) > T$ and $\Im(\gamma z) > T$, then $\mathfrak{c} \sim \mathfrak{b}$ and $\gamma = \omega_\infty$, which follows immediately from Lemma 3.3. $\square$

With these results in hand, we can prove the following Maass–Selberg relation.

Proposition 3.5. *For any two singular cusps $\mathfrak{a}, \mathfrak{b}$, $T \geq 1$, and $s \neq \bar{r}$, $s + \bar{r} \neq 1$,*

$$\int_{\Gamma_0(q)\backslash\mathbb{H}} \Lambda^T E_{\mathfrak{a}}(z, s, \chi) \overline{\Lambda^T E_{\mathfrak{b}}(z, r, \chi)}\, d\mu(z)$$
$$= \overline{\varphi_{\mathfrak{b}\mathfrak{a}}(r, \chi)} \frac{T^{s-\bar{r}}}{s-\bar{r}} + \varphi_{\mathfrak{a}\mathfrak{b}}(s, \chi) \frac{T^{\bar{r}-s}}{\bar{r}-s} + \delta_{\mathfrak{a}\mathfrak{b}} \frac{T^{s+\bar{r}-1}}{s+\bar{r}-1}$$
$$+ \sum_{\mathfrak{c}} \varphi_{\mathfrak{a}\mathfrak{c}}(s, \chi) \overline{\varphi_{\mathfrak{b}\mathfrak{c}}(r, \chi)} \frac{T^{1-s-\bar{r}}}{1-s-\bar{r}},$$

where the sum is over singular cusps $\mathfrak{c}$. Here $d\mu(z) = dx\, dy/y^2$ is the $\mathrm{SL}_2(\mathbb{R})$-invariant measure on $\mathbb{H}$.

Proof. We initially assume that $\Re(s), \Re(r) > 1$ with $\Re(s) - \Re(r) > 1$; the identity then extends to all $s, r \in \mathbb{C}$ with $s \neq \bar{r}$ and $s + \bar{r} \neq 1$ by analytic continuation.

We first show that

$$\int_{\Gamma_0(q)\backslash\mathbb{H}} \Lambda^T E_{\mathfrak{a}}(z, s, \chi) \bigl(\overline{\Lambda^T E_{\mathfrak{b}}(z, r, \chi)} - \overline{E_{\mathfrak{b}}(z, r, \chi)}\bigr)\, d\mu(z) = 0.$$

Indeed, the left-hand side is equal to

$$\sum_{\mathfrak{c}} \int_{\Gamma_0(q)\backslash\mathbb{H}} \Lambda^T E_{\mathfrak{a}}(z, s, \chi) \sum_{\substack{\gamma \in \Gamma_{\mathfrak{c}} \backslash \Gamma_0(q) \\ \Im(\sigma_{\mathfrak{c}}^{-1} \gamma z) > T}} \overline{\chi(\gamma) j_{\sigma_{\mathfrak{c}}^{-1} \gamma}(z)^{-\kappa} c_{\mathfrak{b}\mathfrak{c}}(\sigma_{\mathfrak{c}}^{-1} \gamma z, r, \chi)}\, d\mu(z),$$

which, by (2.3) and (3.2), is equal to

$$-\sum_{\mathfrak{c}} \int_{\Gamma_0(q)\backslash\mathbb{H}} \sum_{\substack{\gamma \in \Gamma_{\mathfrak{c}} \backslash \Gamma_0(q) \\ \Im(\sigma_{\mathfrak{c}}^{-1} \gamma z) > T}} \overline{c_{\mathfrak{b}\mathfrak{c}}(\sigma_{\mathfrak{c}}^{-1} \gamma z, r, \chi)}\, j_{\sigma_{\mathfrak{c}}}(\sigma_{\mathfrak{c}}^{-1} \gamma z)^{-\kappa} \Lambda^T E_{\mathfrak{a}}(\gamma z, s, \chi)\, d\mu(z),$$

and this integral can be unfolded to yield

$$-\sum_{\mathfrak{c}} \int_T^\infty \int_0^1 \overline{c_{\mathfrak{b}\mathfrak{c}}(z, r, \chi)}\, j_{\sigma_{\mathfrak{c}}}(z)^{-\kappa} \Lambda^T E_{\mathfrak{a}}(\sigma_{\mathfrak{c}} z, s, \chi) \frac{dx\, dy}{y^2}.$$

But $\overline{c_{\mathfrak{bc}}(z,r,\chi)}$ is independent of x, while for $\Im(z) > T \geq 1$, the zeroth Fourier coefficient of the function $j_{\sigma_{\mathfrak{c}}}(z)^{-\kappa}\Lambda^T E_{\mathfrak{a}}(\sigma_{\mathfrak{c}}z,s,\chi)$ vanishes via Corollary 3.4, and so this vanishes. Consequently,

$$\int_{\Gamma_0(q)\backslash\mathbb{H}} \Lambda^T E_{\mathfrak{a}}(z,s,\chi)\overline{\Lambda^T E_{\mathfrak{b}}(z,r,\chi)}\,d\mu(z) = \int_{\Gamma_0(q)\backslash\mathbb{H}} \Lambda^T E_{\mathfrak{a}}(z,s,\chi)\overline{E_{\mathfrak{b}}(z,r,\chi)}\,d\mu(z).$$

The right-hand side can be written as

$$\begin{aligned}
&\int_{\Gamma_0(q)\backslash\mathbb{H}}\Bigg(\sum_{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma_0(q)} \overline{\chi}(\gamma) j_{\sigma_{\mathfrak{a}}^{-1}\gamma}(z)^{-\kappa}\Im(\sigma_{\mathfrak{a}}^{-1}\gamma z)^s\,\overline{E_{\mathfrak{b}}(z,r,\chi)} \\
&-\sum_{\mathfrak{c}}\sum_{\substack{\gamma\in\Gamma_{\mathfrak{c}}\backslash\Gamma_0(q)\\ \Im(\sigma_{\mathfrak{c}}^{-1}\gamma z)>T}} \overline{\chi}(\gamma) j_{\sigma_{\mathfrak{c}}^{-1}\gamma}(z)^{-\kappa} c_{\mathfrak{ac}}(\sigma_{\mathfrak{c}}^{-1}\gamma z,s,\chi)\overline{E_{\mathfrak{b}}(z,r,\chi)}\Bigg)\,d\mu(z) \\
&\quad= \int_{\Gamma_0(q)\backslash\mathbb{H}} \sum_{\substack{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma_0(q)\\ \Im(\sigma_{\mathfrak{a}}^{-1}\gamma z)\leq T}} \overline{\chi}(\gamma) j_{\sigma_{\mathfrak{a}}^{-1}\gamma}(z)^{-\kappa}\Im(\sigma_{\mathfrak{a}}^{-1}\gamma z)^s\,\overline{E_{\mathfrak{b}}(z,r,\chi)}\,d\mu(z) \\
&\qquad+ \int_{\Gamma_0(q)\backslash\mathbb{H}} \sum_{\substack{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma_0(q)\\ \Im(\sigma_{\mathfrak{a}}^{-1}\gamma z)>T}} \overline{\chi}(\gamma) j_{\sigma_{\mathfrak{a}}^{-1}\gamma}(z)^{-\kappa}\varphi_{\mathfrak{aa}}(s,\chi)\Im(\sigma_{\mathfrak{a}}^{-1}\gamma z)^{1-s}\,\overline{E_{\mathfrak{b}}(z,r,\chi)}\,d\mu(z) \\
&\qquad\quad- \sum_{\mathfrak{c}\neq\mathfrak{a}}\int_{\Gamma_0(q)\backslash\mathbb{H}} \sum_{\substack{\gamma\in\Gamma_{\mathfrak{c}}\backslash\Gamma_0(q)\\ \Im(\sigma_{\mathfrak{c}}^{-1}\gamma z)>T}} \overline{\chi}(\gamma) j_{\sigma_{\mathfrak{c}}^{-1}\gamma}(z)^{-\kappa} c_{\mathfrak{ac}}(\sigma_{\mathfrak{c}}^{-1}\gamma z,s,\chi)\overline{E_{\mathfrak{b}}(z,r,\chi)}\,d\mu(z).
\end{aligned}$$

By (2.3) and (2.4), the first term is

$$\int_{\Gamma_0(q)\backslash\mathbb{H}} \sum_{\substack{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma_0(q)\\ \Im(\sigma_{\mathfrak{a}}^{-1}\gamma z)\leq T}} \Im(\sigma_{\mathfrak{a}}^{-1}\gamma z)^s\,\overline{j_{\sigma_{\mathfrak{a}}}(\sigma_{\mathfrak{a}}^{-1}\gamma z)^{-\kappa}E_{\mathfrak{b}}(\gamma z,r,\chi)}\,d\mu(z),$$

and upon unfolding the integral, this becomes

$$\begin{aligned}
\int_0^T\int_0^1 y^s\,\overline{j_{\sigma_{\mathfrak{a}}}(z)^{-\kappa}E_{\mathfrak{b}}(\sigma_{\mathfrak{a}}z,r,\chi)}\,\frac{dx\,dy}{y^2} &= \int_0^T y^s\,\overline{c_{\mathfrak{ba}}(z,r,\chi)}\,\frac{dy}{y^2} \\
&= \delta_{\mathfrak{ab}}\frac{T^{s+\bar{r}-1}}{s+\bar{r}-1} + \overline{\varphi_{\mathfrak{ba}}(r,\chi)}\frac{T^{s-\bar{r}}}{s-\bar{r}}.
\end{aligned}$$

Similarly, the second term is

$$\int_T^\infty \varphi_{\mathfrak{aa}}(s,\chi)y^{1-s}\,\overline{c_{\mathfrak{ba}}(z,s,\chi)}\,\frac{dy}{y^2} = \delta_{\mathfrak{ab}}\varphi_{\mathfrak{ab}}(s,\chi)\frac{T^{\bar{r}-s}}{\bar{r}-s} + \varphi_{\mathfrak{aa}}(s,\chi)\overline{\varphi_{\mathfrak{ba}}(r,\chi)}\frac{T^{1-s-\bar{r}}}{1-s-\bar{r}},$$

and the third term is

$$-\sum_{\mathfrak{c}\neq\mathfrak{a}}\int_T^\infty c_{\mathfrak{a}\mathfrak{c}}(z,s,\chi)\overline{c_{\mathfrak{b}\mathfrak{c}}(z,r,\chi)}\,\frac{dy}{y^2}$$
$$=(1-\delta_{\mathfrak{a}\mathfrak{b}})\varphi_{\mathfrak{a}\mathfrak{b}}(s,\chi)\frac{T^{\bar{r}-s}}{\bar{r}-s}+\sum_{\mathfrak{c}\neq\mathfrak{a}}\varphi_{\mathfrak{a}\mathfrak{c}}(s,\chi)\overline{\varphi_{\mathfrak{b}\mathfrak{c}}(r,\chi)}\frac{T^{1-s-\bar{r}}}{1-s-\bar{r}}.$$

Combining these identities yields the result. □

Corollary 3.6. *For $T\geq 1$ and $t\in\mathbb{R}$, we have that*

$$\int_{\Gamma_0(q)\backslash\mathbb{H}}\left|\Lambda^T E_\infty\left(z,\tfrac{1}{2}+it,\chi\right)\right|^2 d\mu(z)=2\log T-\Re\left(\frac{\varphi'_{\infty 1}}{\varphi_{\infty 1}}\left(\frac{1}{2}+it,\chi\right)\right).$$

Proof. We take $\mathfrak{a}\sim\mathfrak{b}\sim\infty$ and $s=r=\frac{1}{2}+it+\varepsilon$ with $\varepsilon>0$ in the Maass–Selberg relation to obtain

$$\int_{\Gamma_0(q)\backslash\mathbb{H}}\left|\Lambda^T E_\infty\left(z,\frac{1}{2}+it+\varepsilon,\chi\right)\right|^2 d\mu(z)=\frac{T^{2\varepsilon}}{2\varepsilon}-\left|\varphi_{\infty 1}\left(\frac{1}{2}+it+\varepsilon,\chi\right)\right|^2\frac{T^{-2\varepsilon}}{2\varepsilon}.$$

The result then follows by taking the limit as ε tends to zero and using the Taylor expansions

$$T^{2\varepsilon}=1+2\varepsilon\log T+O(\varepsilon^2),$$
$$\varphi_{\infty 1}\left(\tfrac{1}{2}+it+\varepsilon,\chi\right)=\varphi_{\infty 1}\left(\tfrac{1}{2}+it,\chi\right)+\varepsilon\varphi'_{\infty 1}\left(\tfrac{1}{2}+it,\chi\right)+O(\varepsilon^2).$$

together with (2.10). □

Remark 3.7. This proof of the Maass–Selberg relation is via unfolding as in Section 4 of [Arthur 1980], and makes use of the Arthur truncation $\Lambda^T E_{\mathfrak{a}}(z,s,\chi)$ of the Eisenstein series $E_{\mathfrak{a}}(z,s,\chi)$ given by (3.1); compare Section 1 of the same work. One can instead prove the Maass–Selberg relation without recourse to the automorphy of the truncated Eisenstein series by only defining $\Lambda^T E_{\mathfrak{a}}(z,s,\chi)$ within a fundamental domain of $\Gamma_0(q)\setminus\mathbb{H}$. Let

$$\mathcal{F}\supset\{z\in\mathbb{H}:0<\Re(z)<1,\ \Im(z)\geq 1\}$$

be the usual fundamental domain of $\Gamma_0(q)\setminus\mathbb{H}$, and for each singular cusp $\mathfrak{a}$, we define the cuspidal zone

$$\mathcal{F}_{\mathfrak{a}}(T):=\{z\in\mathcal{F}:0<\Re(\sigma_{\mathfrak{a}}^{-1}z)<1,\ \Im(\sigma_{\mathfrak{a}}^{-1}z)\geq T\}$$

for $T\geq 1$; note that any two cuspidal zones will be disjoint provided that T is sufficiently large. Then from Lemma 3.3, we have that for $T \geq 1$,

$$\Lambda^T E_{\mathfrak{a}}(z,s,\chi) = \begin{cases} E_{\mathfrak{a}}(z,s,\chi) & \text{if } z \in \mathcal{F} \setminus \bigcup_{\mathfrak{c}} \mathcal{F}_{\mathfrak{c}}(T), \\ E_{\mathfrak{a}}(z,s,\chi) \\ \quad - \sum_{\mathfrak{c}\in A} j_{\sigma_{\mathfrak{c}}^{-1}}(z)^{-\kappa}(\delta_{\mathfrak{ac}}\Im(\sigma_{\mathfrak{c}}^{-1}z)^s + \varphi_{\mathfrak{ac}}(s,\chi)\Im(\sigma_{\mathfrak{c}}^{-1}z)^{1-s}) & \text{if } z \in \bigcap_{\mathfrak{c}\in A} \mathcal{F}_{\mathfrak{c}}(T), \end{cases}$$

where A is any subset of the set of singular cusps. The Maass–Selberg relation may then be proved using Green's theorem along the same lines as the proof of [Iwaniec 2002, Proposition 6.8].

4. Upper bounds and lower bounds for the integral $\mathcal{I}$

For $\eta \leq 1$, we consider the integral

$$\mathcal{I} = \mathcal{I}(\chi,\eta,T) := \int_\eta^\infty \int_0^1 \left| \Lambda^T E_\infty\left(z, \frac{1}{2}, \chi\right)\right|^2 \frac{dx\,dy}{y^2}.$$

Our goal is to find upper and lower bounds for this integral: upper bounds via the Maass–Selberg relation and lower bounds via Parseval's identity and the Brun–Titchmarsh inequality. Combining these bounds will yield lower bounds for $L(1,\chi)$.

Upper bounds for $\mathcal{I}$.

Proposition 4.1. *For $\eta \ll 1/q$ and $T \geq 1$, we have that*

$$\mathcal{I} \ll \frac{\log q \log qT}{q\eta |L(1,\chi)|}.$$

Proof. By folding the integral, one can write

$$\mathcal{I} = \int_{\Gamma_0(q)\backslash\mathbb{H}} N_q(z,\eta) \left|\Lambda^T E_\infty\left(z, \tfrac{1}{2}, \chi\right)\right|^2 d\mu(z),$$

where for $\eta \leq 1$,

$$N_q(z,\eta) := \#\{\gamma \in \Gamma_\infty \backslash \Gamma_0(q) : \Im(\gamma z) > \eta\}.$$

The Maass–Selberg relation then implies the upper bound

$$\mathcal{I} \leq \sup_{z\in\Gamma_0(q)\backslash\mathbb{H}} N_q(z,\eta)\left(2\log T - \Re\left(\frac{\varphi_{\infty 1}'}{\varphi_{\infty 1}}\left(\frac{1}{2},\chi\right)\right)\right).$$

From [Iwaniec 2002, Lemma 2.10], we have the bound

$$N_q(z,\eta) < 1 + \frac{10}{q\eta}.$$

By taking logarithmic derivatives of (2.8),

$$\frac{\varphi'_{\infty 1}}{\varphi_{\infty 1}}(s,\chi) = -\log q - 2\frac{\Lambda'}{\Lambda}(2-2s,\chi) - 2\frac{\Lambda'}{\Lambda}(2s,\overline{\chi}).$$

Taking logarithmic derivatives of (2.9) and letting $s = \frac{1}{2}$ then shows that

$$\frac{\varphi'_{\infty 1}}{\varphi_{\infty 1}}\Bigl(\frac{1}{2},\chi\Bigr) = -4\Re\Bigl(\frac{L'}{L}(1,\chi)\Bigr) - 2\log q + \log 8\pi + \gamma_0 + (-1)^{\kappa}\frac{\pi}{2},$$

where γ_0 denotes the Euler–Mascheroni constant, and we have used the fact that

$$\frac{\Gamma'}{\Gamma}\Bigl(\frac{1+\kappa}{2}\Bigr) = -\log 8 - \gamma_0 - (-1)^{\kappa}\frac{\pi}{2}.$$

So if $\eta \ll 1/q$,

$$\mathcal{I} \ll \frac{(|L(1,\chi)|\log qT + |L'(1,\chi)|)}{q\eta|L(1,\chi)|}.$$

The desired upper bound then follows from the bounds

$$|L(1,\chi)| \ll \log q, \qquad |L'(1,\chi)| \ll (\log q)^2,$$

which are both easily shown via partial summation. See, for example, [Montgomery and Vaughan 2007, Lemma 10.15] for the former estimate; the latter follows by a similar argument. □

Lower bounds for $\mathcal{I}$.

Proposition 4.2. *If $T \geq 1$ and $\eta = 1/T$, we have the lower bound*

$$\mathcal{I} \gg \frac{1}{q|L(1,\chi)|^2}\sum_{T \leq m \leq 2T} |\sigma_0(m,\chi)|^2.$$

Proof. If $\eta = 1/T$, then Lemma 3.3 implies that

$$\Lambda^T E_\infty(z,s,\chi) = \begin{cases} E_\infty(z,s,\chi) & \text{if } 1/T < \Im(z) \leq T, \\ E_\infty(z,s,\chi) - c_{\infty\infty}(z,s,\chi) & \text{if } \Im(z) > T. \end{cases}$$

It follows that the nonzero Fourier coefficients of $\Lambda^T E_\infty(z,s,\chi)$ coincide with those of $E_\infty(z,s,\chi)$ for $\Im(z) > 1/T$. So by Parseval's identity, using the fact that $|\tau(\chi)| = \sqrt{q}$, and making the change of variables $y \mapsto y/|m|$ in the integral, we have that

$$\mathcal{I} \gg \begin{cases} \dfrac{1}{q|L(1,\chi)|^2}\displaystyle\sum_{m=1}^{\infty} |\sigma_0(m,\chi)|^2 \int_{m/T}^{\infty} |K_0(2\pi y)|^2\,\frac{dy}{y} & \text{if } \kappa = 0, \\ \dfrac{1}{q|L(1,\chi)|^2}\displaystyle\sum_{m=1}^{\infty} |\sigma_0(m,\chi)|^2 \int_{m/T}^{\infty} e^{-4\pi y}\,\frac{dy}{y} & \text{if } \kappa = 1. \end{cases}$$

If we simply consider the contribution of the positive integers m for which $m/T \asymp 1$ — say $T \le m \le 2T$ — then we find that

$$\mathcal{I} \gg \frac{1}{q|L(1, \chi)|^2} \sum_{T \le m \le 2T} |\sigma_0(m, \chi)|^2,$$

as desired. □

Combining the upper and lower bounds for $\mathcal{I}$, we derive the following inequality for $L(1, \chi)$:

Corollary 4.3. *For all $T \ge q$, we have that*

$$|L(1, \chi)| \gg \frac{1}{T(\log T)^2} \sum_{T \le m \le 2T} |\sigma_0(m, \chi)|^2.$$

So to obtain lower bounds for $|L(1, \chi)|$, we must find lower bounds for

$$\sum_{T \le m \le 2T} |\sigma_0(m, \chi)|^2. \tag{4.4}$$

Sieve methods. For quadratic characters, lower bounds for (4.4) follow by restricting the sum to perfect squares.

Lemma 4.5. *If χ is a quadratic character, then*

$$\sum_{T \le m \le 2T} |\sigma_0(m, \chi)|^2 \ge (\sqrt{2} - 1)\sqrt{T}.$$

Proof. We restrict the sum over m to perfect squares and use the fact that $\sigma_0(m, \chi) \ge 1$ whenever m is a perfect square in order to find that

$$\sum_{T \le m \le 2T} |\sigma_0(m, \chi)|^2 \ge \sum_{T \le m^2 \le 2T} |\sigma_0(m^2, \chi)|^2 \ge (\sqrt{2} - 1)\sqrt{T}. \qquad \square$$

For complex characters, we instead restrict the sum in (4.4) to primes and use the Brun–Titchmarsh inequality to show that there are sufficiently many primes for which $\bar{\chi}(p)$ is not close to -1, so that $|\sigma_0(p, \chi)|^2$ is not small. This is a result of Balasubramanian and Ramachandra [1976, Lemma 4], who combine it with an identity of Ramanujan together with a complex analytic argument to obtain lower bounds for $L(1 + it, \chi)$, and consequently derive zero-free regions for $L(s, \chi)$. We reproduce a proof of this result here for the sake of completeness.

Lemma 4.6 [Balasubramanian and Ramachandra 1976, Lemma 4]. *There exists a large constant $K \ge 2$ such that for all complex characters χ modulo q with $q \ge 2$ and for $T = q^K$,*

$$\sum_{T \le m \le 2T} |\sigma_0(m, \chi)|^2 \gg_K \frac{T}{\log T}.$$

Proof. We restrict the sum over m to primes p in order to find that

$$\sum_{T \le m \le 2T} |\sigma_0(m,\chi)|^2 \ge \sum_{T \le p \le 2T} |1+\chi(p)|^2$$
$$= 2 \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} (1+\Re(\chi(a)))(\pi(2T;q,a)-\pi(T;q,a)),$$

where $\pi(x;q,a) := \#\{p \le x : p \equiv a \pmod q\}$.

Let Q be the order of the Dirichlet character χ; this divides $\varphi(q)$, and as χ is complex, $Q \ge 3$. For any integer M between 0 and $\lfloor Q/2 \rfloor$, we have that

$$\sum_{T \le m \le 2T} |\sigma_0(m,\chi)|^2 \ge 2\Bigl(1+\cos\frac{2\pi M}{Q}\Bigr)(\pi(2T)-\pi(T))$$
$$-2\Bigl(1+\cos\frac{2\pi M}{Q}\Bigr) \sum_{\substack{a \in (\mathbb{Z}/q\mathbb{Z})^\times \\ \Re(\chi(a)) < \cos\frac{2\pi M}{Q}}} (\pi(2T;q,a)-\pi(T;q,a)).$$

For the former sum, we have that for fixed $\delta > 0$ to be chosen,

$$\pi(2T)-\pi(T) \ge (1-\delta)\frac{T}{\log T}$$

for all sufficiently large T dependent on δ. See, for example, [Diamond and Erdős 1980]; in particular, this does not require the full strength of the prime number theorem.

For the latter sum, we first observe that there are $\varphi(q)/Q$ reduced residue classes a modulo Q for which $\chi(a) = e^{2\pi i m/Q}$ for each integer m between 0 and $Q-1$, and so the number of reduced residue classes modulo q for which $\Re(\chi(a)) < \cos(2\pi M/Q)$ is

$$\frac{\varphi(q)}{Q}\#\{M < m < Q-M\} = \varphi(q)\frac{Q-2M-1}{Q}.$$

To find an upper bound for $\pi(2T;q,a)-\pi(T;q,a)$, we use the Brun–Titchmarsh inequality, which states that for $(q,a)=1$, $x \ge 2$, and $y \ge 2q$,

$$\pi(x+y;q,a)-\pi(x;q,a) \le \frac{2y}{\varphi(q)\log y/q}\Bigl(1+\frac{8}{\log y/q}\Bigr).$$

We take $x = y = T$, assuming that $T \ge 2q$, in order to obtain

$$\sum_{\substack{a \in (\mathbb{Z}/q\mathbb{Z})^\times \\ \Re(\chi(a)) < \cos\frac{2\pi M}{Q}}} (\pi(2T;q,a)-\pi(T;q,a)) \le \frac{2(Q-2M-1)}{Q}\frac{T}{\log T/q}\Bigl(1+\frac{8}{\log T/q}\Bigr).$$

We take $T = q^K$ with $K \geq 2$ sufficiently large and dependent on δ but not on q, such that

$$\frac{1}{\log T/q}\Big(1+\frac{8}{\log T/q}\Big) \leq (1+\delta)\frac{1}{\log T}.$$

Combined, these estimates imply that for $T = q^K$ with $K \geq 2$ a sufficiently large constant,

$$\sum_{T\leq m\leq 2T} |\sigma_0(m,\chi)|^2 \geq 2(1-\cos\pi X)\Big(1-\delta-2(1+\delta)X+\frac{2(1+\delta)}{Q}\Big)\frac{T}{\log T}$$

for $X = (Q-2M)/Q$.

For $Q \geq 3$, we may choose

$$\delta = \frac{1}{10}, \qquad M = \Big\lfloor \frac{1+4\delta}{2(1+\delta)}\frac{Q}{2}+\frac{1}{2}\Big\rfloor,$$

so that

$$X = \frac{1-2\delta}{2(1+\delta)} - \frac{1}{Q} + \frac{2}{Q}\Big\{\frac{1+4\delta}{2(1+\delta)}\frac{Q}{2}+\frac{1}{2}\Big\},$$

and hence

$$1-\delta-2(1+\delta)X+\frac{2(1+\delta)}{Q} = \delta + \frac{4(1+\delta)}{Q}\Big(1-\Big\{\frac{1+4\delta}{2(1+\delta)}\frac{Q}{2}+\frac{1}{2}\Big\}\Big) \geq \delta.$$

Moreover, the fact that $\delta = \frac{1}{10}$ and $Q \geq 3$ implies that $1 \leq M \leq \lfloor Q/2\rfloor$ and $\frac{1}{33} \leq X \leq \frac{23}{33}$. So

$$\sum_{T\leq m\leq 2T} |\sigma_0(m,\chi)|^2 \gg_K \frac{T}{\log T}. \qquad \square$$

Remark 4.7. If χ is quadratic, so that the order of χ is $Q = 2$, then

$$\sum_{T\leq m\leq 2T} |\sigma_0(m,\chi)|^2 \geq 2(\pi(2T)-\pi(T)) - 2\sum_{\substack{a\in(\mathbb{Z}/q\mathbb{Z})^\times\\ \chi(a)=-1}} (\pi(2T;q,a)-\pi(T;q,a)).$$

The Brun–Titchmarsh inequality is insufficient to show that the first term on the right-hand side dominates the second term; in its place, we would require a strengthening of the Brun–Titchmarsh inequality of the form

$$\pi(x+y;q,a)-\pi(x;q,a) \leq \frac{(2-\delta)y}{\varphi(q)\log y/q}(1+o(1)) \tag{4.8}$$

for some $\delta > 0$. With this in hand, we would then be able to show that

$$\sum_{T\leq m\leq 2T} |\sigma_0(m,\chi)|^2 \gg \frac{T}{\log T},$$

so that

$$L(1,\chi) \gg \frac{1}{(\log q)^3},$$

which would imply the nonexistence of a Landau–Siegel zero for $L(1, \chi)$. Of course, the fact that the strengthened Brun–Titchmarsh inequality (4.8) implies (and is in fact equivalent to) the nonexistence of Landau–Siegel zeroes is well known.

5. Proof of Theorem 1.1

With these upper and lower bounds established, we are in a position to prove Theorem 1.1.

Proof of Theorem 1.1. If χ is quadratic, we have via Corollary 4.3 and Lemma 4.5 that for $T \geq q$,

$$L(1, \chi) \gg \frac{1}{\sqrt{T}(\log T)^2},$$

and so taking $T = q$ yields the desired lower bound.

If χ is complex, we have via Corollary 4.3 and Lemma 4.6 that for $T = q^K$,

$$|L(1, \chi)| \gg_K \frac{1}{(\log T)^3} \gg_K \frac{1}{(\log q)^3}. \qquad \square$$

References

[Arthur 1980] J. Arthur, "A trace formula for reductive groups, II: Applications of a truncation operator", *Compositio Math.* **40**:1 (1980), 87–121. MR Zbl

[Balasubramanian and Ramachandra 1976] R. Balasubramanian and K. Ramachandra, "The place of an identity of Ramanujan in prime number theory", *Proc. Indian Acad. Sci. Sect. A* **83**:4 (1976), 156–165. MR Zbl

[Deshouillers and Iwaniec 1982] J.-M. Deshouillers and H. Iwaniec, "Kloosterman sums and Fourier coefficients of cusp forms", *Invent. Math.* **70**:2 (1982), 219–288. MR Zbl

[Diamond and Erdős 1980] H. G. Diamond and P. Erdős, "On sharp elementary prime number estimates", *Enseign. Math.* (2) **26**:3-4 (1980), 313–321. MR Zbl

[Duke et al. 2002] W. Duke, J. B. Friedlander, and H. Iwaniec, "The subconvexity problem for Artin L-functions", *Invent. Math.* **149**:3 (2002), 489–577. MR Zbl

[Gelbart and Lapid 2006] S. S. Gelbart and E. M. Lapid, "Lower bounds for L-functions at the edge of the critical strip", *Amer. J. Math.* **128**:3 (2006), 619–638. MR Zbl

[Goldfeld and Li 2016] D. Goldfeld and X. Li, "A standard zero free region for Rankin–Selberg L-functions", preprint, 2016. arXiv

[Gradshteyn and Ryzhik 2007] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed., edited by A. Jeffrey and D. Zwillinger, Elsevier, Amsterdam, 2007. MR Zbl

[Iwaniec 2002] H. Iwaniec, *Spectral methods of automorphic forms*, 2nd ed., Graduate Studies in Mathematics **53**, American Mathematical Society, Providence, RI, 2002. MR Zbl

[Miyake 1989] T. Miyake, *Modular forms*, Springer, Berlin, 1989. MR Zbl

[Montgomery and Vaughan 2007] H. L. Montgomery and R. C. Vaughan, *Multiplicative number theory, I: Classical theory*, Cambridge Studies in Advanced Mathematics **97**, Cambridge Univ. Press, 2007. MR Zbl

[Sarnak 2004] P. Sarnak, "Nonvanishing of L-functions on $\Re(s) = 1$", pp. 719–732 in *Contributions to automorphic forms, geometry, and number theory*, edited by H. Hida et al., Johns Hopkins Univ. Press, Baltimore, 2004. MR Zbl

Received June 14, 2016.

PETER HUMPHRIES
DEPARTMENT OF MATHEMATICS
PRINCETON UNIVERSITY
PRINCETON, NJ 08544
UNITED STATES
peterch@math.princeton.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.377

ASYMPTOTIC ORDER-OF-VANISHING FUNCTIONS ON THE PSEUDOEFFECTIVE CONE

SHIN-YAO JOW

Let v be a discrete valuation on the function field of a normal projective variety X. Ein, Lazarsfeld, Mustaţă, Nakamaye, and Popa showed that v induces a nonnegative real-valued continuous function on the big cone of X, which they called the asymptotic order of vanishing along v. The case where v is given by the order of vanishing along a prime divisor was studied earlier by Nakayama, who extended the domain of the function to the pseudoeffective cone and investigated the continuity of the extended function.

Here we generalize Nakayama's results to any discrete valuation v, using an approach inspired by Lazarsfeld and Mustaţă's construction of the global Okounkov body, which has a quite different flavor from the arguments employed by Nakayama.

A corollary is that the asymptotic order-of-vanishing function can be extended continuously to the pseudoeffective cone PE(X) of X if PE(X) is polyhedral (note that we do *not* require PE(X) to be *rational* polyhedral).

Let X be a normal projective variety over an algebraically closed field k, and let $K(X)$ be the function field of X. Let v be a discrete valuation of $K(X)$ over k, and let Z be the center of v on X. Ein, Lazarsfeld, Mustaţă, Nakamaye, and Popa gave the following definitions:

Definition 1 [Ein et al. 2006]. Let D be an effective big Cartier divisor on X. We establish the following notation:

(i) $v(D) = v(f)$, where f is a local equation of D at the generic point of Z.

(ii) $v(|D|) = \min\{v(D') : D' \in |D|\} = v(D')$ for general $D' \in |D|$.

(iii) $v(\|D\|) = \lim_{m\to\infty} v(|mD|)/m$. This is called the asymptotic order of vanishing of D along v.

MSC2010: 14C20.

Keywords: asymptotic order of vanishing, pseudoeffective cone, global Okounkov body, Nakayama's σ-decomposition.

By taking m to be sufficiently divisible, the definition of $v(\|D\|)$ can also be extended to big $\mathbb{Q}$-divisors D. It is proved in [Ein et al. 2006, Theorem A] that $v(\|D\|)$ depends only on the numerical equivalence class of D, so it induces a function on the set of numerical equivalence classes of big $\mathbb{Q}$-divisors. Moreover, this function extends uniquely to a continuous function on the cone $\mathrm{Big}(X)$ of numerical equivalence classes of big $\mathbb{R}$-divisors. In view of this result, it is natural to ask whether this function can be extended continuously to the pseudoeffective cone $\mathrm{PE}(X)$, the closure of the big cone $\mathrm{Big}(X)$ in the Néron–Severi space $N^1(X)_{\mathbb{R}}$.

The case where v is a divisorial valuation was investigated earlier by Nakayama [2004] during his study of Zariski decomposition in higher dimensions. More precisely, let Γ be a prime divisor on a smooth projective variety X, and let v be the discrete valuation of $K(X)$ given by the order of vanishing at the generic point of Γ. Nakayama used the notation $\sigma_\Gamma(D)$ to denote the asymptotic order of vanishing $v(\|D\|)$ of a big divisor class $D \in \mathrm{Big}(X)$. If $D \in \mathrm{PE}(X)$ is a pseudoeffective class, he defined $\sigma_\Gamma(D)$ by picking an arbitrary ample class $A \in N^1(X)_{\mathbb{R}}$ and setting $\sigma_\Gamma(D)$ to be the limit

$$\sigma_\Gamma(D) = \lim_{\epsilon \to 0^+} \sigma_\Gamma(D + \epsilon A),$$

after establishing that this limit does not depend on the choice of A, in [Nakayama 2004, III.1.5]. In III.1.7 of the same work, Nakayama showed that the function $\sigma_\Gamma : \mathrm{PE}(X) \to \mathbb{R}_{\geq 0}$ is lower semicontinuous, and he gave an example where it is not continuous in IV.2.8. It is interesting to note that in his example $\mathrm{PE}(X)$ is not polyhedral. The goal of this short note is to generalize Nakayama's results to any discrete valuation v of $K(X)/k$, using an approach inspired by Lazarsfeld and Mustață's construction [2009] of the global Okounkov body, which has a quite different flavor from the arguments employed by Nakayama. In addition, we will see that the function $v(\|\cdot\|) : \mathrm{Big}(X) \to \mathbb{R}_{\geq 0}$ can be extended continuously to $\mathrm{PE}(X)$ if $\mathrm{PE}(X)$ is polyhedral.

Theorem 2. *Let X be a normal projective variety over an algebraically closed field k, and let v be a discrete valuation of $K(X)$ over k. If $D \in \mathrm{PE}(X)$ is a pseudoeffective class, then for any ample class $A \in N^1(X)_{\mathbb{R}}$, $\lim_{\epsilon \to 0^+} v(\|D + \epsilon A\|)$ does not depend on the choice of A. Moreover, if we denote this limit by $\sigma_v(D)$, then the function*

$$\sigma_v : \mathrm{PE}(X) \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$$

is lower semicontinuous, and is continuous at every point where $\mathrm{PE}(X)$ *is locally polyhedral.*

A subset S of $\mathbb{R}^n$ is said to be *locally polyhedral* at a point $x \in S$ if there exist a polytope $P \subset \mathbb{R}^n$ and an open subset U of $\mathbb{R}^n$ containing x such that $U \cap S = U \cap P$. It follows from Theorem 2 that the function $v(\|\cdot\|) : \mathrm{Big}(X) \to \mathbb{R}_{\geq 0}$ can be extended continuously to $\mathrm{PE}(X)$ if $\mathrm{PE}(X)$ is polyhedral, which is the case, for example, when the Picard number of X is 2. Note that we do *not* require $\mathrm{PE}(X)$ to be *rational* polyhedral (cf. [Ein et al. 2006, Theorem D]).

Remark 3. If v is divisorial, the limit $\lim_{\epsilon\to 0^+} v(\|D+\epsilon A\|)$ in Theorem 2 is finite [Nakayama 2004, III.1.5]. We do not know if this is true for all discrete valuations v, which is why we include $+\infty$ in the target of σ_v. In case the value of σ_v is $+\infty$ at a point of $\mathrm{PE}(X)$, the (semi)continuity of σ_v should be interpreted with respect to the usual order topology on $\mathbb{R}_{\geq 0}\cup\{+\infty\}$.

Let us introduce some notions from convex analysis which will be useful in the proof of Theorem 2. Let $f : S \to \mathbb{R}\cup\{+\infty\}$ be a function on a convex subset S of $\mathbb{R}^n$. We say that f is *convex* if

$$f\big(\lambda x_1 + (1-\lambda)x_2\big) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

for all $x_1, x_2 \in S$ and all $0 \leq \lambda \leq 1$. The *epigraph* of f is the set

$$\{(x, y) \in S\times\mathbb{R} : y \geq f(x)\}.$$

A convex function f is said to be *closed* if its epigraph is a closed subset of $\mathbb{R}^{n+1}$. It is not difficult to show that if f is a closed convex function, then f is lower semicontinuous.

Proof of Theorem 2. As mentioned earlier, our approach is inspired by the construction of the global Okounkov body due to [Lazarsfeld and Mustață 2009]. The strategy is to construct the epigraph of the asymptotic order-of-vanishing function as the closed convex cone spanned by a certain lattice semigroup. To see how this works for one big divisor D, let $\mathbb{N}$ denote the set of nonnegative integers, and let

$$S(D) = \{(m, y) \in \mathbb{N}^2 : y \geq v(|mD|)\},$$

which is a subsemigroup of $\mathbb{N}^2$. Let $C(D) = \overline{\mathrm{cone}}(S(D))$ be the closed convex cone spanned by $S(D)$ in $\mathbb{R}^2$. Then $C(D)$ is the epigraph of the function $x \mapsto v(\|xD\|)$. In order to get the epigraph of the function $v(\|\cdot\|) : \mathrm{Big}(X) \to \mathbb{R}_{\geq 0}$, pick a $\mathbb{Z}$-basis $D_1, \ldots, D_n$ for $N^1(X)$ such that, after identifying $N^1(X)_{\mathbb{R}}$ with $\mathbb{R}^n$ by this basis, we have $\mathrm{PE}(X) \subseteq \mathbb{R}^n_{\geq 0}$. Let

$$S(X) = \{(m_1, \ldots, m_n, y) \in \mathbb{N}^n\times\mathbb{N} : y \geq v(|m_1D_1 + \cdots + m_nD_n|)\},$$

and let

$$C(X) = \overline{\mathrm{cone}}(S(X)) \subseteq \mathbb{R}^n_{\geq 0}\times\mathbb{R}_{\geq 0}$$

be the closed convex cone spanned by $S(X)$. Let

$$f : \mathrm{PE}(X) \to \mathbb{R}_{\geq 0}\cup\{+\infty\}$$

be the function whose epigraph is $C(X)$. Then f is a closed convex function since $C(X)$ is a closed convex cone. Moreover, on the big cone $\mathrm{Big}(X)$, f coincides with the function $v(\|\cdot\|)$ by [Lazarsfeld and Mustaţă 2009, Proposition 4.9].

To see what $f(D)$ is if D is on the boundary of $\mathrm{PE}(X)$, we invoke a theorem of Gale, Klee, and Rockafellar, which states that a closed convex function is continuous at every point where its domain is locally polyhedral ([Gale et al. 1968, Theorem 2]; see also the introduction of [Ernst 2013]). It follows that for any ample $A \in N^1(X)_{\mathbb{R}}$, the restriction of f to the half-line $D + \mathbb{R}_{\geq 0} A$ is continuous. Hence

$$f(D) = \lim_{\epsilon \to 0^+} f(D + \epsilon A) = \lim_{\epsilon \to 0^+} v(\|D + \epsilon A\|).$$

This shows that the limit on the right does not depend on the choice of A, and that in fact $f = \sigma_v$. Since σ_v is a closed convex function, it is lower semicontinuous, and is continuous at every point where $\mathrm{PE}(X)$ is locally polyhedral by the theorem of Gale, Klee and Rockafellar. $\square$

Acknowledgments

The author gratefully acknowledges the support of MOST (Ministry of Science and Technology, Taiwan), NCTS (National Center for Theoretical Sciences), and the Kenda Foundation.

References

[Ein et al. 2006] L. Ein, R. Lazarsfeld, M. Mustaţă, M. Nakamaye, and M. Popa, "Asymptotic invariants of base loci", *Ann. Inst. Fourier* (*Grenoble*) **56**:6 (2006), 1701–1734. MR Zbl

[Ernst 2013] E. Ernst, "A converse of the Gale–Klee–Rockafellar theorem: continuity of convex functions at the boundary of their domains", *Proc. Amer. Math. Soc.* **141**:10 (2013), 3665–3672. MR Zbl

[Gale et al. 1968] D. Gale, V. Klee, and R. T. Rockafellar, "Convex functions on convex polytopes", *Proc. Amer. Math. Soc.* **19** (1968), 867–873. MR Zbl

[Lazarsfeld and Mustaţă 2009] R. Lazarsfeld and M. Mustaţă, "Convex bodies associated to linear series", *Ann. Sci. Éc. Norm. Supér.* (4) **42**:5 (2009), 783–835. MR Zbl

[Nakayama 2004] N. Nakayama, *Zariski-decomposition and abundance*, MSJ Memoirs **14**, Mathematical Society of Japan, Tokyo, 2004. MR Zbl

Received May 13, 2016.

SHIN-YAO JOW
DEPARTMENT OF MATHEMATICS
NATIONAL TSING HUA UNIVERSITY
NO. 101, SEC. 2, GUANGFU RD
HSINCHU 300
TAIWAN
syjow@math.nthu.edu.tw

PACIFIC JOURNAL OF MATHEMATICS
Vol. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.381

AUGMENTATIONS AND RULINGS OF LEGENDRIAN LINKS IN $\#^k(S^1 \times S^2)$

CAITLIN LEVERSON

Given a Legendrian link in $\#^k(S^1 \times S^2)$, we extend the definition of a normal ruling from $J^1(S^1)$ given by Lavrov and Rutherford and show that the existence of an augmentation to any field of the Chekanov–Eliashberg differential graded algebra over $\mathbb{Z}[t, t^{-1}]$ is equivalent to the existence of a normal ruling of the front diagram. For Legendrian knots, we also show that any even graded augmentation must send t to -1. We use the correspondence to give nonvanishing results for the symplectic homology of certain Weinstein 4-manifolds. We show a similar correspondence for the related case of Legendrian links in $J^1(S^1)$, the solid torus.

1. Introduction

Augmentations and normal rulings are important tools in the study of Legendrian knot theory, especially in the study of Legendrian knots in $\mathbb{R}^3$. Here, augmentations are augmentations of the Chekanov–Eliashberg differential graded algebra introduced by Chekanov [2002] and Eliashberg [1998]. Chekanov describes the noncommutative differential graded algebra (DGA) over $\mathbb{Z}/2$ associated to a Lagrangian diagram of a Legendrian link in $(\mathbb{R}^3, \xi_{\text{std}})$ combinatorially: The DGA is generated by crossings of the link; the differential is determined by a count of immersed polygons whose corners lie at crossings of the link and whose edges lie on the link. This is called the Chekanov–Eliashberg DGA and Chekanov showed that the homology of this DGA is invariant under Legendrian isotopy. Etnyre, Ng, and Sabloff [Etnyre et al. 2002] defined a lift of the Chekanov–Eliashberg DGA to a DGA over $\mathbb{Z}[t, t^{-1}]$ in. Following ideas introduced by Eliashberg [1987] and motivated by generating families (functions whose critical values generate front diagrams of Legendrian knots), Fuchs [2003] and Chekanov and Pushkar [2005] gave invariants of Legendrian knots in $\mathbb{R}^3$. Fuchs looked at decompositions of these generating families, generally called "normal rulings."

These two invariants are very closely related; Fuchs [2003], Fuchs and Ishkhanov [2004], and Sabloff [2005] showed that the existence of a normal ruling is equivalent

MSC2010: primary 57R17; secondary 53D42, 57M27.
Keywords: Legendrian knot, Legendrian submanifold, contact manifold, normal ruling.

to the existence of an augmentation to $\mathbb{Z}/2$ of the Chekanov–Eliashberg DGA $\mathcal{A}$ for Legendrian knots in $\mathbb{R}^3$. Here, given a unital ring S, an augmentation of $\mathcal{A}$ is a ring map $\epsilon : \mathcal{A} \to S$ such that $\epsilon \circ \partial = 0$ and $\epsilon(1) = 1$. One of the main results of [Leverson 2016] is that the equivalence remains true when one looks at augmentations to a field of the lift of the Chekanov–Eliashberg DGA from [Etnyre et al. 2002] to the DGA over $\mathbb{Z}[t, t^{-1}]$ for Legendrian knots in $\mathbb{R}^3$. We extend the result to Legendrian *links* in $\mathbb{R}^3$ to prove the main result of this paper.

Theorem 1.1. *Let Λ be an s-component Legendrian link in $\mathbb{R}^3$. Given a field F, the Chekanov–Eliashberg DGA $(\mathcal{A}, \partial)$ over $\mathbb{Z}[t_1^{\pm 1}, \dots, t_s^{\pm 1}]$ has a ρ-graded augmentation $\epsilon : \mathcal{A} \to F$ if and only if a front diagram of Λ has a ρ-graded normal ruling. Furthermore, if ρ is even, then $\epsilon(t_1 \cdots t_s) = (-1)^s$.*

The final statement tells us that for all even graded augmentations $\epsilon : \mathcal{A} \to F$, $\epsilon(t_1 \cdots t_s) = (-1)^s$. In particular, if Λ is a knot, then any even graded augmentation sends t to -1.

For $k \geq 0$, an analogous correspondence can be shown for Legendrian links in $\#^k(S^1 \times S^2)$. A Legendrian link in $\#^k(S^1 \times S^2)$ with the standard contact structure is an embedding $\Lambda : \coprod_s S^1 \to \#^k(S^1 \times S^2)$ which is everywhere tangent to the contact planes. We will think of them as Gompf [1998] does. For an example, see Figure 2. In this paper, we extend the definition of normal ruling of a Legendrian link in $\mathbb{R}^3$ to a Legendrian link in $\#^k(S^1 \times S^2)$. We then define the ruling polynomial for a Legendrian link in $\#^k(S^1 \times S^2)$ and show that the ruling polynomial is invariant under Legendrian isotopy. Note that Lavrov and Rutherford [2013] did this previously in the case where $k = 1$.

Theorem 1.2. *The ρ-graded ruling polynomial $R^{\rho}_{(\Lambda,m)}$ with respect to the Maslov potential m (which changes under Legendrian isotopy) is a Legendrian isotopy invariant.*

Ekholm and Ng [2015] extend the definition of the Chekanov–Eliashberg DGA over $\mathbb{Z}[t, t^{-1}]$ to Legendrian links in $\#^k(S^1 \times S^2)$. The main result of this paper uses Theorem 1.1 to extend the correspondence between normal rulings and augmentations to a correspondence for Legendrian links in $\#^k(S^1 \times S^2)$.

Theorem 1.3. *Let Λ be an s-component Legendrian link in $\#^k(S^1 \times S^2)$ for some $k \geq 0$. Given a field F, the Chekanov–Eliashberg DGA $(\mathcal{A}(\Lambda), \partial)$ over $\mathbb{Z}[t_1^{\pm 1}, \dots, t_s^{\pm 1}]$ has a ρ-graded augmentation $\epsilon : \mathcal{A}(\Lambda) \to F$ if and only if a front diagram of Λ has a ρ-graded normal ruling. Furthermore, if ρ is even, then $\epsilon(t_1 \cdots t_s) = (-1)^s$.*

Notice that one can consider Legendrian links in $\mathbb{R}^3$ as being Legendrian links in $\#^0(S^1 \times S^2)$. In this way, this result is a generalization of the correspondence in [Leverson 2016] and Theorem 1.1. An immediate corollary is the following:

Corollary 1.4. *If Λ is a Legendrian link in $\#^k(S^1 \times S^2)$ and there exists ℓ such that N_ℓ is odd, then there does not exist a ρ-graded augmentation of the DGA $\mathcal{A}(\Lambda)$ for any ρ.*

In other words, if Λ has a 1-handle with an odd number of strands going through it, then there does not exist a ρ-graded augmentation of the DGA $\mathcal{A}(\Lambda)$ for any ρ. This follows from the fact that every involution of a set with an odd number of elements has a fixed-point.

Along with the work of Bourgeois, Ekholm, and Eliashberg [Bourgeois et al. 2012], Theorem 1.3 gives nonvanishing results for Weinstein (Stein) 4-manifolds. (Note that proofs of the results in [loc. cit.] have not appeared yet.) In particular:

Corollary 1.5. *If X is the Weinstein 4-manifold obtained from attaching 2-handles along a Legendrian link Λ to $\#^k(S^1 \times S^2)$ and Λ has a graded normal ruling, then the full symplectic homology $S\mathbb{H}(X)$ is nonzero.*

This follows from Theorem 1.3 as the existence of a normal ruling implies the existence of an augmentation to $\mathbb{Q}$, which, by [Bourgeois et al. 2012], is a sufficient condition for the full symplectic homology to be nonzero.

We show a correspondence for Legendrian links in the 1-jet space of the circle $J^1(S^1)$. Ng and Traynor [2004] extend the definition of the Chekanov–Eliashberg DGA to Legendrian links in $J^1(S^1)$. Lavrov and Rutherford [2012] extend the definition of normal ruling to a "generalized normal ruling" of Legendrian links in $J^1(S^1)$ and show that the existence of a generalized normal ruling is equivalent to the existence of an augmentation to $\mathbb{Z}/2$ of the Chekanov–Eliashberg DGA over $\mathbb{Z}/2$ of a Legendrian link in $J^1(S^1)$. In Section 6, we show that this correspondence holds for augmentations to any field of the Chekanov–Eliashberg DGA over $\mathbb{Z}[t_1^{\pm 1}, \dots, t_s^{\pm 1}]$.

Theorem 1.6. *Suppose that Λ is a Legendrian link in $J^1(S^1)$. Given a field F, the Chekanov–Eliashberg DGA $(\mathcal{A}, \partial)$ over $\mathbb{Z}[t_1^{\pm 1}, \dots, t_s^{\pm 1}]$ has a ρ-graded augmentation $\epsilon : \mathcal{A} \to F$ if and only if a front diagram of Λ has a ρ-graded generalized normal ruling.*

1A. ***Outline of the article.*** In Section 2, we recall background on Legendrian links in $\#^k(S^1 \times S^2)$ and $\mathbb{R}^3$. We give definitions of the Chekanov–Eliashberg DGA over $\mathbb{Z}[t, t^{-1}]$, with sign conventions, and augmentations of the DGA in both $\#^k(S^1 \times S^2)$ and $\mathbb{R}^3$. We also define normal rulings for links in $\#^k(S^1 \times S^2)$ and show that the ruling polynomial is invariant under Legendrian isotopy, proving Theorem 1.2. In Section 3, we prove Theorem 1.1. In Section 4, given an augmentation, we construct a normal ruling proving one direction of Theorem 1.3. In Section 5, given a normal ruling, we construct an augmentation, finishing the proof of Theorem 1.3. In Section 6, we prove Theorem 1.6. In the Appendix, we give the nonvanishing symplectic homology result.

2. Background material

2A. *Legendrian links in $\#^k(S^1 \times S^2)$*. In this section we will briefly discuss necessary concepts of Legendrian links in $\#^k(S^1 \times S^2)$. We will follow the notation in [Ekholm and Ng 2015].

Definition 2.1. Let $A, M > 0$ be fixed. A tangle in $[0, A]\times[-M, M]\times[-M, M]$ is *Legendrian* if it is everywhere tangent to the standard contact structure $dz - y dx$. Informally, a Legendrian tangle T in $[0, A]\times[-M, M]\times[-M, M]$ is in *normal form* if

- T meets $x = 0$ and $x = A$ in k groups of strands, where the groups are of size $N_1, \dots, N_k$, from top to bottom in both the xy- and xz-projections,
- and within the ℓ-th group, we label the strands by $1, \dots, N_\ell$ from top to bottom at $x = 0$ in both the xy- and xz-projections and $x = A$ in the xz-projection, and from bottom to top at $x = A$ in the xy-projection.

Every Legendrian tangle in normal form gives a Legendrian link in $\#^k(S^1 \times S^2)$ by attaching k 1-handles which join parts of the xz projection of the tangle at $x = 0$ to the parts at $x = A$. In particular, the ℓ-th 1-handle joins the ℓ-th group at $x = 0$ to the ℓ-th group at $x = A$ and connects the strands in this group with the same label at $x = 0$ and $x = A$ through the 1-handle. See Figure 2.

Every Legendrian link in $\#^k(S^1 \times S^2)$ has an xz-diagram of the form given by Gompf [1998], which we will call *Gompf standard form*. The left diagram of Figure 2 is an example of a link in Gompf standard form. Any link in Gompf standard form can be isotoped to a link whose xy-projection is obtained from the xz-diagram by *resolution*. The resolution of an xz-diagram of a link is obtained by the replacements given in Figure 1. For an example, see Figure 2. By [Ekholm and Ng 2015], an xy-diagram obtained by the resolution of an xz-diagram of a link in Gompf standard form is in normal form. Thus, we can assume that the xy-diagram of any Legendrian link is in normal form.

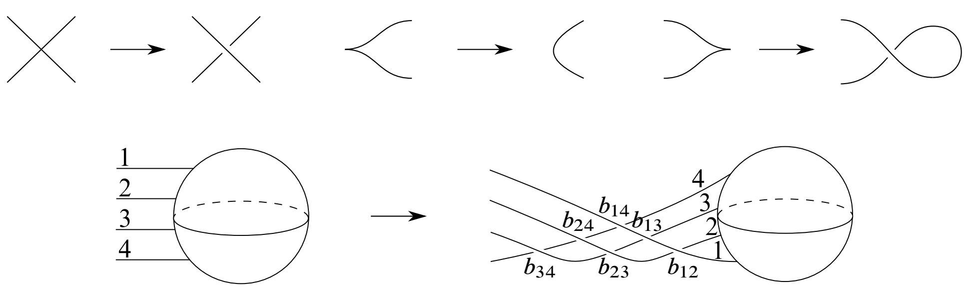


Figure 1. Resolutions of an xz-diagram in Gompf standard form.

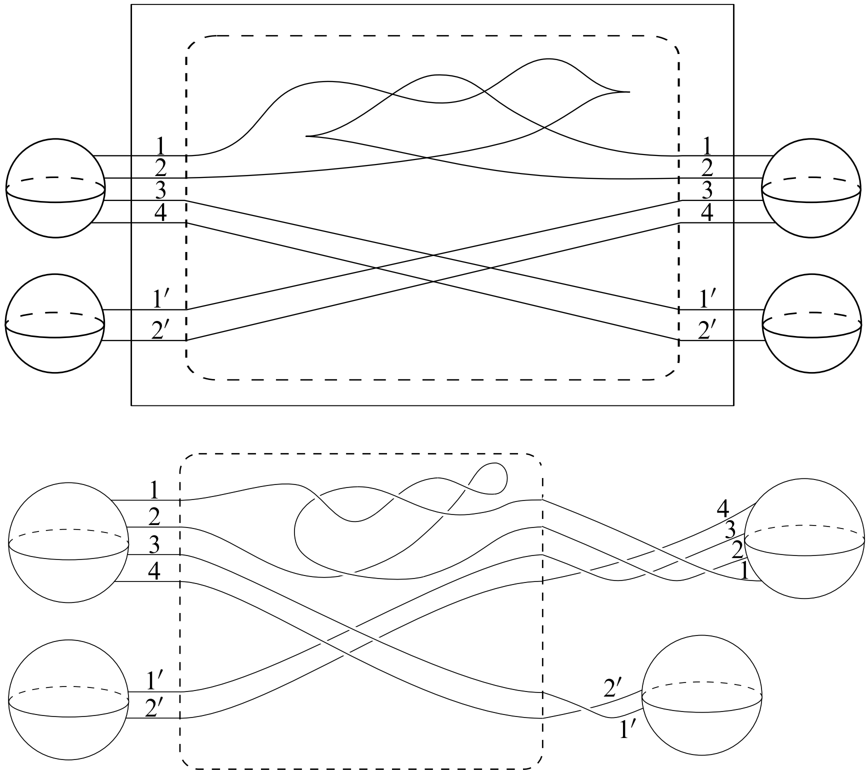


Figure 2. A Legendrian xz-diagram of a link in $\#^2(S^1 \times S^2)$ in Gompf standard form (top), and the resolution of the Legendrian link to an xy-diagram of a Legendrian isotopic link (bottom).

2B. ***Definition of the DGA and augmentations in $\#^k(S^1 \times S^2)$.*** This section contains an overview of the differential graded algebra over the ring $\mathbb{Z}[t_1^{\pm 1}, \dots, t_s^{\pm 1}]$ presented by Ekholm and Ng [2015]. Let $\Lambda = \Lambda_1 \sqcup \cdots \sqcup \Lambda_n$ be a Legendrian link in $\#^k(S^1 \times S^2)$ in normal form, where the Λ_i denote the components of Λ and $n \leq s$. On each link component Λ_i, label a point by $*_i$ (corresponding to t_i) within the tangle (away from crossings). We will discuss the case where there is more than one basepoint on a given component in Section 2K. Let $N_i \geq 1$ be the number of strands of Λ which go through the i-th 1-handle with $N = \sum N_i$ the total number of strands at $x = 0$.

2C. ***Internal DGA.*** We will define the internal DGA for a Legendrian link in $S^1 \times S^2$, but one can easily extend the definition to the internal DGA for a Legendrian link in $\#^k(S^1 \times S^2)$ by defining the internal DGA as follows for each 1-handle separately.

Let $(r_1, \dots, r_n) \in \mathbb{Z}^n$ be the n-tuple where r_i is the rotation number of the i-th component Λ_i, let $r = \gcd(r_1, \dots, r_n)$, and let $(m(1), \dots, m(N)) \in (\mathbb{Z}/2r)^N$ be the N-tuple of a choice of Maslov potential for each strand passing through the 1-handle (see Section 2E).

Let $(\mathcal{A}_N, \partial_N)$ denote the DGA defined as follows. Let $\mathcal{A}$ be the tensor algebra over $R = \mathbb{Z}[t_1^{\pm 1}, \dots, t_s^{\pm 1}]$ generated by c_{ij}^0 for $1 \le i < j \le N$ and c_{ij}^p for $1 \le i, j \le N$ and $p \ge 1$. Set $|t_i| = -2r_i$, $|t_i^{-1}| = 2r_i$, and

$$|c_{ij}^p| = 2p - 1 + m(i) - m(j)$$

for all i, j, p. Define the differential ∂_N on the generators by

$$\begin{aligned}
\partial_N(c_{ij}^0) &= \sum_{\ell=i+1}^{j-1} (-1)^{|c_{i\ell}^0|+1} c_{i\ell}^0 c_{\ell j}^0, \\
\partial_N(c_{ij}^1) &= \delta_{ij} + \sum_{\ell=i+1}^{N} (-1)^{|c_{i\ell}^0|+1} c_{i\ell}^0 c_{\ell j}^1 + \sum_{\ell=1}^{j-1} (-1)^{|c_{i\ell}^1|+1} c_{i\ell}^1 c_{\ell j}^0, \\
\partial_N(c_{ij}^p) &= \sum_{\ell=0}^{p} \sum_{m=1}^{N} (-1)^{|c_{im}^\ell|+1} c_{im}^\ell c_{mj}^{p-\ell},
\end{aligned}$$

where $p \ge 2$, δ_{ij} is the Kronecker delta function, and we set $c_{ij}^0 = 0$ for $i \ge j$. Extend ∂_N to $\mathcal{A}_N$ by the Leibniz rule

$$\partial_N(xy) = (\partial_N x)y + (-1)^{|x|} x(\partial_N y).$$

From [Ekholm and Ng 2015], we know ∂_N has degree -1, $\partial_N^2 = 0$, and $(\mathcal{A}_N, \partial_N)$ is infinitely generated as an algebra, but is a filtered DGA, where c_{ij}^p is a generator of the ℓ-th component of the filtration if $p \le \ell$.

Given a Legendrian link $\Lambda \subset \#^k(S^1 \times S^2)$, we can associate a DGA $(\mathcal{A}_{N_i}, \partial_{N_i})$ to each of the 1-handles. We then call the DGA generated by the collection of generators of $\mathcal{A}_i$ for $1 \le i \le k$ with differential induced by ∂_{N_i}, the *internal DGA* of Λ.

2D. *Algebra.* Suppose we have a Legendrian link $\Lambda = \Lambda_1 \sqcup \cdots \sqcup \Lambda_n \subset \#^k(S^1 \times S^2)$ in normal form with exactly one point labeled $*_i$ within the tangle (away from crossings) on each link component Λ_i of Λ (corresponding to t_i). We will discuss the case where there is more than one basepoint on a given component in Section 2K.

Notation 2.2. Let $\tilde{a}_1, \dots, \tilde{a}_m$ denote the crossings of the xy tangle diagram in normal form. Label the k 1-handles in the diagram by $1, \dots, k$ from top to bottom. Recall that N_i denotes the number of strands of the tangle going through the i-th 1-handle. For each i, label the strands going through the i-th 1-handle on the left side of the diagram $1, \dots, N_i$ from top to bottom and from bottom to top on the right side, as in Figure 2.

Let $\mathcal{A}(\Lambda)$ be the tensor algebra over $R = \mathbb{Z}[t_1^{\pm 1}, \dots, t_s^{\pm 1}]$ generated by

- $\tilde{a}_1, \dots, \tilde{a}_m$;
- $c^0_{ij;\ell}$ for $1 \le \ell \le k$ and $1 \le i < j \le N_\ell$;
- $c^p_{ij;\ell}$ for $1 \le \ell \le k$, $p \ge 1$, and $1 \le i, j \le N_\ell$.

(In general, we will drop the index ℓ when the 1-handle is clear.)

2E. *Grading.* The following are a few preliminary definitions which will allow us to define the grading on the generators of $\mathcal{A}(\Lambda)$.

Definition 2.3. A *path* in $\pi_{xy}(\Lambda)$ is a path that traverses part (or all) of $\pi_{xy}(\Lambda)$ which is connected except for where it enters a 1-handle, meaning, where it approaches $x = 0$ (respectively $x = A$) along a labeled strand and exits the 1-handle along the strand with the same label from $x = A$ (respectively $x = 0$). Note that the tangent vector in $\mathbb{R}^2$ to the path varies continuously as we traverse a path as the strands entering and exiting 1-handles are horizontal.

The *rotation number* $r(\gamma)$ of a path γ is the number of counterclockwise revolutions around S^1 made by the tangent vector $\gamma'(t)/|\gamma'(t)|$ to γ as we traverse γ. Generally this will be a real number, but will be an integer if and only if γ is smooth and closed.

Thus, the rotation number $r_i = r(\Lambda_i)$ is the rotation number of the path in $\pi_{xy}(\Lambda)$ which begins at the basepoint $*_i$ on the link component Λ_i and traverses the link component, following the orientation of the component. In the case where Λ is a link with components $\Lambda_1, \dots, \Lambda_n$, we define

$$r(\Lambda) = \gcd(r_1, \dots, r_n).$$

Define

$$|t_i| = -2r(\Lambda_i).$$

If $\pi_{xy}(\Lambda)$ is the resolution of an xz-diagram of an n-component link in Gompf standard form, then the method assigning gradings follows: Choose a *Maslov potential* m that associates an integer modulo $2r(\Lambda)$ to each strand in the tangle T associated to Λ, minus cusps and basepoints, such that the following conditions hold:

(1) For all $1 \le \ell \le k$ and all $1 \le i \le N_\ell$, the strand labeled i going through the ℓ-th 1-handle at $x = 0$ and the $x = A$ must have the same Maslov potential.

(2) If a strand is oriented to the right, meaning it enters the 1-handle at $x = A$ and exits at $x = 0$, then the Maslov potential of the strand must be even. Otherwise the Maslov potential of the strand must be odd.

(3) At a cusp, the upper strand (strand with higher z-coordinate) has Maslov potential one more than the lower strand.

The Maslov potential is well-defined up to an overall shift by an even integer for knots. (Ekholm and Ng [2015] give another method for defining the gradings using the rotation numbers of specified paths.)

Set $|t_i| = -2r(\Lambda_i)$ and $|c^p_{ij;\ell}| = 2p-1+m(i)-m(j)$, where $m(i)$ means the Maslov potential of the strand with label i going through the ℓ-th 1-handle. It remains to define the grading on crossings in the tangle, crossings resulting from resolving right cusps, and crossings from the half-twists in the resolution. If a is a crossing in the tangle T, then define

$$|a| = m(S_o) - m(S_u),$$

where S_o is the strand which crosses over the strand S_u at a in the xy-projection of Λ. If a is a right cusp, define $|a| = 1$ (assuming there is not a basepoint in the loop). If a is a crossing in one of the half-twists in the resolution where strand i crosses over strand j $(i < j)$, then

$$|a| = m(i) - m(j).$$

2F. *Differential.* It suffices to define the differential ∂ on generators and extend by the Leibniz rule. Define $\partial(\mathbb{Z}[t_1^{\pm 1}, \dots, t_s^{\pm 1}]) = 0$. Set $\partial = \partial_{N_\ell}$ on $\mathcal{A}_{N_\ell}$ as in Section 2C.

In [Ekholm and Ng 2015], the DGA on crossings a_i is defined by looking for immersed disks in the xy-diagrams of Legendrian links, (see the left diagram in Figure 3). However, Ekholm and Ng note that it is equivalent to look for immersed disks in dip versions of the diagram, (see the right diagram in Figure 3). See Figure 4 for the labeling of the crossings in Figure 3.

Definition 2.4. Let $a, b_1, \dots, b_\ell$ be generators. Define $\Delta(a; b_1, \dots, b_\ell)$ to be the set of orientation-preserving maps

$$f : D^2 \to \mathbb{R}^2$$

(up to smooth reparametrization) that map ∂D^2 to the dip version of Λ such that

(1) f is a smooth immersion except at $a, b_1, \dots, b_\ell$,

(2) $a, b_1, \dots, b_\ell$ are encountered as one traverses $f(\partial D^2)$ counterclockwise,

(3) near $a, b_1, \dots, b_\ell$, $f(D^2)$ covers exactly one quadrant, specifically, a quadrant with positive Reeb sign near a and a quadrant with negative Reeb sign near

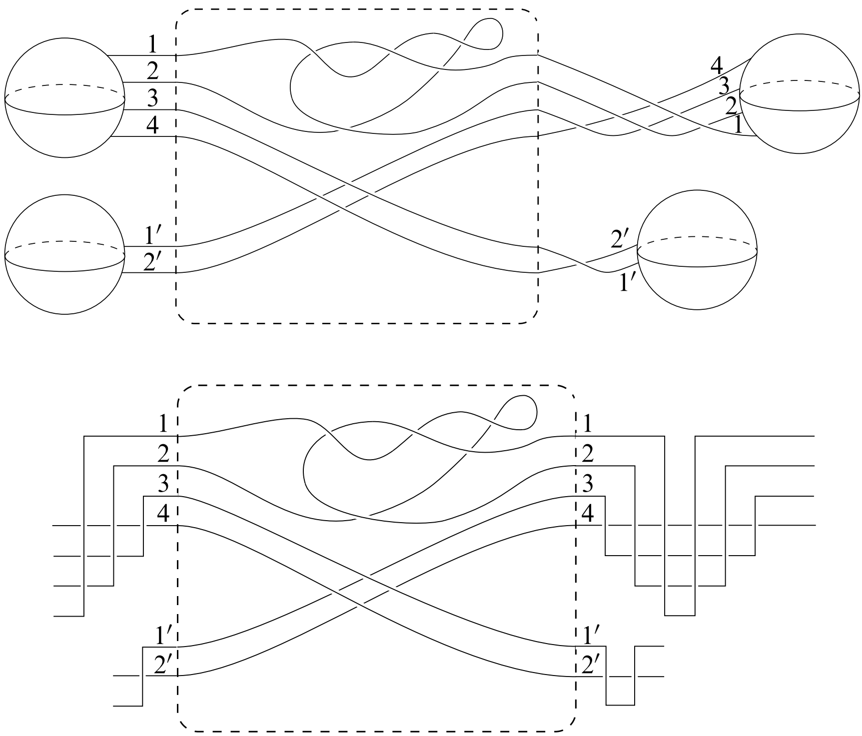


Figure 3. A Legendrian xy-diagram of a link in $\#^2(S^1 \times S^2)$ which has resulted from the resolution of a link in Gompf standard form (top) and the dipped version of the link where the half of a dip on the left side of the dipped version is identified with the right half of the dip on the right side. See Figure 4 for the labeling of the crossings in the dips (bottom).

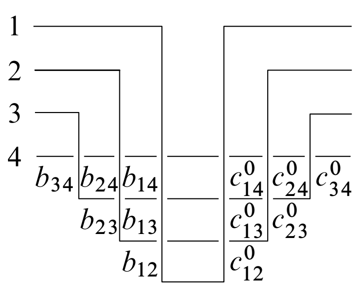


Figure 4. This is the dip at the right of the bottom figure in Figure 3 with strands and crossings labeled. The labels of the partial dip at the left of the bottom figure in Figure 3 are the same as the right half of the dip depicted.

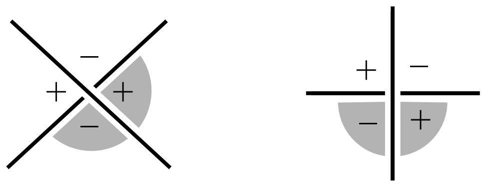


Figure 5. The signs in the figure give the Reeb signs of the quadrants around the crossings. The orientation signs are $+1$ for all quadrants of crossings of odd degree. For crossings of even degree, we use the convention indicated in the left figure if the crossing comes from the xz-projection and the convention in the right figure if the crossing is in a dip, which will be discussed in Section 2J, where the shaded quadrants have orientation sign -1 and the other quadrants have orientation sign $+1$.

$b_1, \dots, b_\ell$, where the Reeb sign of a quadrant near a crossing is defined as in Figure 5.

To each immersed disk, we can assign a word in $\mathcal{A}(\Lambda)$ by starting with the first corner where the quadrant covered has negative Reeb sign, b_1, and listing the crossing labels of all negative corners as encountered while following the boundary of the immersed polygon counterclockwise, $b_1 \cdots b_\ell$. We associate an *orientation sign* $\delta_{Q,a}$ to each quadrant Q in the neighborhood of a crossing a, defined in Figure 5, and use these to define the sign of a disk $f(D^2)$ to be the product of the orientation signs over all the corners of the disk. We denote this sign by $\delta(f)$. In many cases there is a unique disk with positive corner at a (with respect to Reeb sign) and negative corners at $b_1, \dots, b_\ell$ and in these we define $\delta(a; b_1, \dots, b_\ell)$ to be the sign of the unique disk. (In exceptional cases there may be more than one disk with positive corner at a and negative corners at $b_1, \dots, b_\ell$.)

Define $n_{*_i}(f)$ or $n_{*_i}(a; b_1, \dots, b_\ell)$ to be the signed count of the number of times one encounters the basepoint $*_i$ while following $f(\partial D^2)$ counterclockwise, where the sign is positive if we encounter the basepoint while following the orientation of the link component and negative if we encounter the basepoint while going against the orientation.

We define

$$\partial(a_i) = \sum_{\ell \geq 0} \sum_{(b_1,\dots,b_\ell)} \sum_{f \in \Delta(a_i; b_1,\dots,b_\ell)} \delta(f)\, t_1^{n_{*_1}(f)} \cdots t_s^{n_{*_s}(f)} b_1 \cdots b_\ell$$

and extend to $\mathcal{A}(\Lambda)$ by the Leibniz rule.

Ekholm and Ng [2015] prove that the map ∂ has degree -1 and is a differential, i.e., $\partial^2 = 0$.

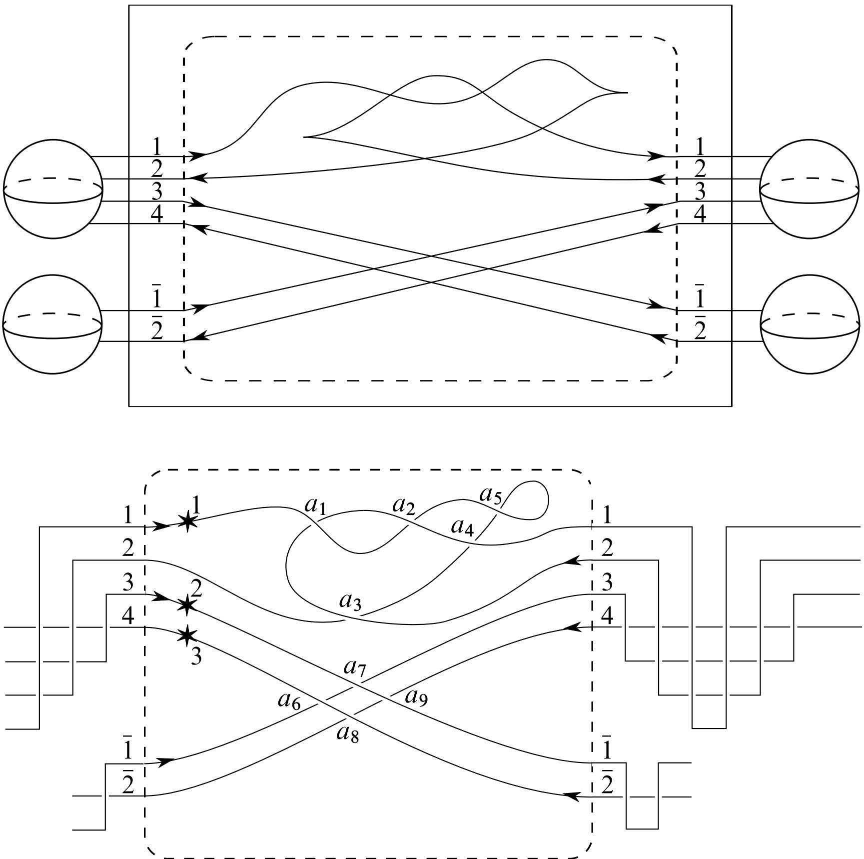


Figure 6. A Legendrian xz-diagram in $\#^2(S^1 \times S^2)$ in Gompf standard form (top) and the dip form of the normal form (bottom). Recall the labels on the crossings in the dips from Figure 4 for the top 1-handle and label the left crossing $\bar{b}_{12}$ and the right $\bar{c}_{12}$ in the dip of the bottom 1-handle.

Example 2.5. The following is the definition of the DGA $(\mathcal{A}(\Lambda), \partial)$ for the Legendrian link Λ in Figure 6. Here $\mathcal{A}(\Lambda)$ is generated by $a_1, \dots, a_9, b_{ij}, c_{ij}^p$ over $\mathbb{Z}[t_1^{\pm 1}, t_2^{\pm 1}, t_3^{\pm 1}]$. We set $|t_i| = 2r(\Lambda_i) = 0$ for $i = 1, 2, 3$. Define a Maslov potential m on the strands near $x = 0$ by

i	1	2	3	4	$\bar{1}$	$\bar{2}$
$m(i)$	2	1	0	-1	0	-1

Then we have the following gradings:

$$|a_1| = |a_2| = |a_3| = |a_7| = |a_8| = 0, \quad |a_4| = |a_5| = |a_9| = 1, \quad |a_6| = -1,$$

ij	12	13	14	23	24	34	$\overline{12}$
$\lvert b_{ij}\rvert$	1	2	3	2	2	1	1
$\lvert c^0_{ij}\rvert$	0	1	2	0	1	0	0

	$\lvert c^1_{ij}\rvert$	j = 1	2	3	4
i	1	1	2	3	4
	2	0	1	2	3
	3	−1	0	1	2
	4	−2	−1	0	1

	$\lvert c^2_{ij}\rvert$	j = 1	2	3	4
i	1	3	4	5	6
	2	2	3	4	5
	3	1	2	3	4
	4	0	1	2	3

where $\overline{12}$ is the crossing of the strands in the bottom 1-handle. Since $|c^p_{ij}| = 2p-1+m(i)-m(j)$, we know $|c^p_{ij}|>0$ for $p>2$.

For ease of notation, we will use $\bar{c}^p_{12}$ to denote $c^p_{\overline{12}}$. We then have the following differentials:

$$\begin{aligned}
\partial a_1 &= \partial a_2 = \partial a_3 = \partial a_6 = 0\\
\partial a_4 &= (1+a_2a_1)a_3 - t_1^{-1}a_2c^0_{12}\\
\partial a_5 &= 1-a_1a_3+t_1^{-1}c^0_{12}\\
\partial a_7 &= t_2^{-1}t_3^{-1}c^0_{34}a_6\\
\partial a_8 &= a_6\bar{c}^0_{12}\\
\partial a_9 &= t_2^{-1}t_3^{-1}c^0_{34}a_8 - a_7\bar{c}^0_{12}
\end{aligned}$$

$$\begin{aligned}
\partial b_{12} &= 1+a_2a_1-c^0_{12}\\
\partial b_{13} &= (1+a_2a_1)b_{23}+a_4(t_2c^0_{23}a_7+t_3^{-1}c^0_{24}a_6)\\
&\qquad -t_1^{-1}a_2(t_2c^0_{13}a_7+t_3^{-1}c^0_{14}a_6)-c^0_{13}+b_{12}c^0_{23}\\
\partial b_{14} &= (1+a_2a_1)b_{24}\\
&\qquad -\big[a_4(t_2c^0_{23}a_7+t_3^{-1}c^0_{24}a_6)-t_1^{-1}a_2(t_2c^0_{13}a_7+t_3^{-1}c^0_{14}a_6)\big]b_{34}\\
&\qquad +(a_4c^0_{23}-t_1^{-1}a_2c^0_{13})t_2a_9+(a_4c^0_{24}-t_1^{-1}a_2c^0_{14})t_3^{-1}a_8\\
&\qquad -c^0_{14}+b_{12}c^0_{24}-b_{13}c^0_{34}\\
\partial b_{23} &= -a_3(t_2c^0_{23}a_7+t_3^{-1}c^0_{24}a_6)-c^0_{23}\\
\partial b_{24} &= -a_3(t_2c^0_{23}a_7+t_3^{-1}c^0_{24}a_6)b_{34}-t_3^{-1}a_3c^0_{24}a_8\\
&\qquad -c^0_{24}+b_{23}c^0_{34}-t_2a_3c^0_{23}a_9\\
\partial b_{34} &= \bar{c}^0_{12}-c^0_{34}
\end{aligned}$$

$$\partial \bar{b}_{12} = t_2^{-1} t_3^{-1} c_{34}^0 - \bar{c}_{12}^0$$

$$\partial c_{ij}^p = \delta_{ij}\delta_{1p} + \sum_{\ell=0}^{p}\sum_{m=1}^{4}(-1)^{|c_{im}^\ell|+1} c_{im}^\ell c_{mj}^{p-\ell}$$

$$\partial \bar{c}_{ij}^p = \delta_{ij}\delta_{1p} + \sum_{\ell=0}^{p}\sum_{m=1}^{2}(-1)^{|\bar{c}_{im}^\ell|+1} \bar{c}_{im}^\ell \bar{c}_{mj}^{p-\ell}$$

Definition 2.6. Let $(\mathcal{A}, \partial)$ be a semifree DGA over R generated by $\{a_i | i \in I\}$. Let J be a countable (possibly finite) index set. A *stabilization* of $(\mathcal{A}, \partial)$ is the semifree DGA $(S(\mathcal{A}), \partial)$, where $S(\mathcal{A})$ is the tensor algebra over R generated by $\{a_i \mid i \in I\} \cup \{\alpha_j \mid j \in J\} \cup \{\beta_j | j \in J\}$ and the grading on a_i is inherited from $\mathcal{A}$ and $|\alpha_j| = |\beta_j| + 1$ for all $j \in J$. Let the differential on $S(\mathcal{A})$ agree with the differential on $\mathcal{A} \subset S(\mathcal{A})$, define

$$\partial(\alpha_j) = \beta_j \text{ and } \partial(\beta_j) = 0$$

for all $j \in J$, and extend by the Leibniz rule.

Definition 2.7 [Ekholm and Ng 2015]. Two semifree DGAs $(\mathcal{A}, \partial)$ and $(\mathcal{A}', \partial')$ are *stable tame isomorphic* if some stabilization of $(\mathcal{A}, \partial)$ is tamely isomorphic to some stabilization of $(\mathcal{A}', \partial')$.

Theorem 2.8 [op. cit., Theorem 2.18]. *Let Λ and Λ' be Legendrian isotopic Legendrian links in $\#^k(S^1 \times S^2)$ in normal form. Let $(\mathcal{A}(\Lambda), \partial)$ and $(\mathcal{A}(\Lambda'), \partial')$ be the semifree DGAs over $R = \mathbb{Z}[t_1^{\pm 1}, \dots, t_s^{\pm 1}]$ associated to the diagrams $\pi_{xy}(\Lambda)$ and $\pi_{xy}(\Lambda')$, which are in normal form. Then $(\mathcal{A}(\Lambda), \partial)$ and $(\mathcal{A}(\Lambda'), \partial')$ are stable tame isomorphic.*

Definition 2.9. Let F be a field. An *augmentation* of $(\mathcal{A}(\Lambda), \partial)$ to F is a ring map $\epsilon : \mathcal{A}(\Lambda) \to F$ such that $\epsilon \circ \partial = 0$ and $\epsilon(1) = 1$. If $\rho \mid 2r(\Lambda)$ and ϵ is supported on generators of degree divisible by ρ, then ϵ is ρ-graded. In particular, if $\rho = 0$, we say it is *graded* and if $\rho = 1$, we say if is *ungraded*. We call a generator a *augmented* if $\epsilon(a) \neq 0$.

Example 2.10. Recalling the DGA associated with the Legendrian link in Figure 6 of Example 2.5, given a field F, one can check that any graded augmentation ϵ to F satisfies: $\epsilon(t_1) = -1$, $\epsilon(t_3) = \epsilon(t_2)^{-1}$ where $\epsilon(t_2) \neq 0$, $\epsilon(b_{ij}) = \epsilon(\bar{b}_{12}) = 0$, and for $a, b, c, d, e, f \in F$ such that $1 + ab, d, e \neq 0$,

i	1	2	3	4	5	6	7	8	9
$\epsilon(a_i)$	a	b	$-b$	0	0	0	c	c	0

ij	12	13	14	23	24	34	$\overline{12}$
$\epsilon(c_{ij}^0)$	$1+ab$	0	0	0	0	d	d

$\lvert c^1_{ij}\rvert$	$j=1$	2	3	4
$i=1$	0	0	0	0
2	e	0	0	0
3	0	f	0	0
4	0	0	$(1+ab)d^{-1}e$	0

$\lvert c^2_{ij}\rvert$	$j=1$	2	3	4
$i=1$	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	$-(1+ab)d^{-1}f$	0	0	0

Note that any augmentation of a stabilization $S(\mathcal{A})$ restricts to an augmentation of the smaller algebra $\mathcal{A}$ and any augmentation of the algebra $\mathcal{A}$ extends to an augmentation of the stabilization $S(\mathcal{A})$ where the augmentation sends β_j to 0 and α_j to an arbitrary element of F if ρ divides $|\alpha_j|$ and 0 otherwise for all $j \in J$.

2G. ***Normal rulings in*** $\#^k(S^1\times S^2)$**.** In this section, we extend the definition of a normal ruling from Legendrian links in $\mathbb{R}^3$ to Legendrian links in $\#^k(S^1\times S^2)$. We formulate the definition similarly to how Lavrov and Rutherford [2012] define normal rulings in the case of Legendrian links in the solid torus.

Consider the tangle portion of the $\pi_{xz}(\Lambda)$ diagram in normal form of a Legendrian link $\Lambda \subset \#^k(S^1\times S^2)$. A normal ruling can be viewed locally as a decomposition of $\pi_{xz}(\Lambda)$ into pairs of paths.

Let $C\subset S^1$ be the set of x-coordinates of crossings and cusps of $\pi_{xz}(\Lambda)$ where $S^1=[0,A]/\{0=A\}$. We can write

$$S^1\backslash C=\coprod_{\ell=1}^{M} I_\ell$$

where I_ℓ is an open interval (or all of S^1) for each ℓ. We use the convention that $I_0=I_M$ and the I_ℓ are ordered $I_0,\dots,I_M$ from $x=0$ to $x=A$ (from left to right in the xz-diagram) so that $I_{\ell-1}$ appears to the left of (has lower x-coordinates than) I_ℓ. Note that $(I_\ell\times[-M,M])\cap\pi_{xz}(\Lambda)$ consists of some number of nonintersecting components which project homeomorphically onto I_ℓ. We call these components *strands* of $\pi_{xz}(\Lambda)$ and number them from top to bottom by $1,\dots,N(\ell)$. For each ℓ, choose a point $x_\ell\in I_\ell$.

Definition 2.11. A *normal ruling* of $\pi_{xz}(\Lambda)$ is a sequence of involutions $\sigma=(\sigma_1,\dots,\sigma_M)$,

$$\sigma_m:\{1,\dots,N(m)\}\to\{1,\dots,N(m)\},\qquad (\sigma_m)^2=\mathrm{id},$$

satisfying:

(1) Each σ_m is fixed-point-free.

(2) If the strands above I_m labeled ℓ and $\ell+1$ meet at a left cusp in the interval (x_{m-1}, x_m), then

$$\sigma_m(i) = \begin{cases} \ell+1 & \text{if } i = \ell, \\ J(\sigma_{m-1}(i)) & \text{if } i < \ell, \\ J(\sigma_{m-1}(i-2)) & \text{if } i > \ell+1, \end{cases}$$

where

$$J(i) = \begin{cases} i, & i < \ell, \\ i+2, & i \geq \ell, \end{cases}$$

and a similar condition at right cusps.

(3) If strands above I_m labeled ℓ and $\ell+1$ meet at a crossing on the interval (x_{m-1}, x_m), then $\sigma_{m-1}(\ell) \neq \ell+1$ and either
- $\sigma_m = (\ell\ \ell+1) \circ \sigma_{m-1} \circ (\ell\ \ell+1)$, where $(\ell\ \ell+1)$ denotes transposition or
- $\sigma_m = \sigma_{m-1}$.

When the second case occurs, we call the crossing *switched*.

(4) (Normality condition) If there is a switched crossing on the interval (x_{m-1}, x_m), then one of the following holds:
- $\sigma_m(\ell+1) < \sigma_m(\ell) < \ell < \ell+1$,
- $\sigma_m(\ell) < \ell < \ell+1 < \sigma_m(\ell)$,
- $\ell < \ell+1 < \sigma_m(\ell+1) < \sigma_m(\ell)$.

(5) Near $x = 0$ and $x = A$, both the strand with label ℓ and the strand with label $\sigma_0(\ell)$ must go through the same 1-handle; in other words, there exists p such that $\sum_{i=1}^{p-1} N_i < \ell, \sigma_0(\ell) \leq \sum_{i=1}^{p} N_i$.

The final condition is the only condition which is different from how normal rulings are defined in [Lavrov and Rutherford 2012] for the case of solid torus knots. This condition ensures the ruling "behaves well" with the 1-handles.

Remark 2.12. As in [loc. cit.], one can equivalently see normal rulings as pairings of strands in the xz-diagram with certain conditions. Here we think of strands i and j being paired for $x_{m-1} \leq x \leq x_m$ if $\sigma_m(i) = j$. In this way, we can cover the xz-diagram with pairs of paths which have monotonically increasing x-coordinate. Note that if a path goes all the way from $x = 0$ to $x = A$, it may end up on a different strand than it started, but strand i is paired with strand j at $x = 0$ if and only if they are paired at $x = A$. Condition (5) also specifies that the paired strands must go through the same 1-handle. The conditions mentioned above are as follows: Paired paths can only meet at a cusp. This also means that at a crossing, the crossings strands must be paired with other strands. These *companion strands* can either lie above or below the crossing. Conditions (3) and (4) specify that near a crossing the pairings must be one of those depicted in Figure 7.

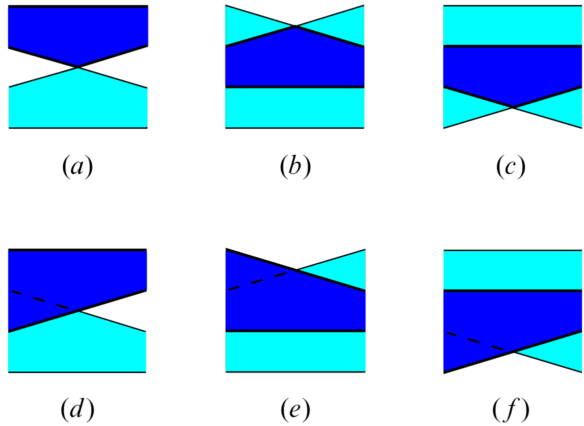


Figure 7. These configurations, along with vertical reflections of (d), (e), and (f), are all possible configurations of a normal ruling near a crossing. The top row contains all possible configurations for switched crossings in a normal ruling. (This figure is taken from [Leverson 2016].)

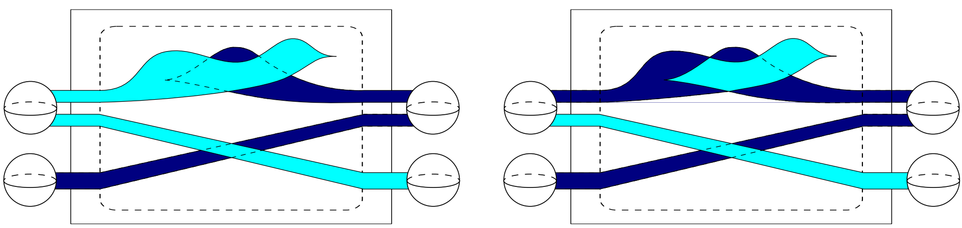


Figure 8. These are the two normal rulings of the Legendrian link of Example 2.5 seen in Figure 6.

Example 2.13. Figure 8 gives the normal rulings of the Legendrian link from Example 2.5.

Definition 2.14. Given ρ such that $\rho \mid 2r(\Lambda)$ and an $\mathbb{Z}/\rho$-valued Maslov potential on Λ, a normal ruling is *ρ-graded* with respect to the $\mathbb{Z}/\rho$-valued Maslov potential if whenever two strands are paired by one of the σ_m, the upper strand (strand with lower label) has Maslov potential one higher than the lower strand (strand with higher label).

Remark 2.15. Note that the condition for being a ρ-graded normal ruling of a Legendrian link in $\#^k(S^1 \times S^2)$ implies that $\rho \,\big|\, |c|$ if the normal ruling is switched at a crossing c. Further, any Legendrian link in $\mathbb{R}^3$ is also a Legendrian link in

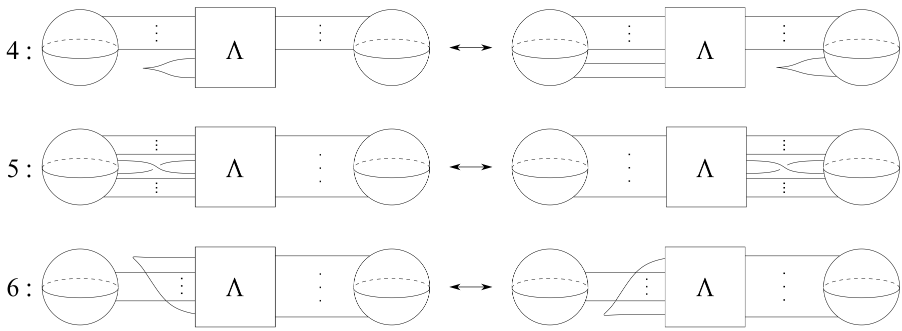


Figure 9. Gompf moves 4, 5, and 6.

$\#^k(S^1 \times S^2)$ for any k (no strands of this link go through any of the 1-handles). We then see that the definition of a ρ-graded normal ruling for the Legendrian link in $\#^k(S^1 \times S^2)$ is equivalent to the definition of a ρ-graded normal ruling for the Legendrian link in $\mathbb{R}^3$.

Similarly to $\mathbb{R}^3$, we can define a ρ-graded ruling polynomial.

Definition 2.16. If m is a $\mathbb{Z}/\rho$-valued Maslov potential for a Legendrian link Λ, then the *ρ-graded ruling polynomial* of Λ with respect to m is

$$R^{\rho}_{(\Lambda,m)} = \sum_{\sigma} z^{j(\sigma)},$$

where the sum is over all ρ-graded normal rulings of Λ and

$$j(\sigma) = \#\,\text{switches} - \#\,\text{right cusps}.$$

Note that in the case where Λ is a knot, the ruling polynomial does not depend on the Maslov potential. Restated from the introduction:

Theorem 1.2. *The ρ-graded ruling polynomial $R^{\rho}_{(\Lambda,m)}$ with respect to the Maslov potential m (which changes under Legendrian isotopy) is a Legendrian isotopy invariant.*

Proof. By Gompf [1998], any Legendrian link in $\#^k(S^1 \times S^2)$ can be represented by an xz-diagram in Gompf standard form and two such xz-diagrams represent links that are Legendrian isotopic if and only if they are related by a sequence of Legendrian Reidemeister moves of the xz-diagram of the tangle inside the rectangle $[0, A] \times [-M, M]$ and three additional moves, which we will, following the nomenclature of [Ekholm and Ng 2015], call Gompf moves 4, 5, and 6 (see Figure 9). By [Pushkar and Chekanov 2005], we know the ruling polynomial is invariant under Legendrian isotopy of the tangle, so we need only show it is invariant under Gompf moves 4, 5, and 6.

Gompf moves 4 and 5 clearly do not change the ruling polynomial. For Gompf move 6, note that any normal ruling cannot pair a strand going through the 1-handle with one of the strands incident to the cusp. Instead, the ruling must pair the two strands incident to the left cusp and not have any switches in the portion of the diagram depicted in Figure 9, thus the ruling polynomial does not change. □

Example 2.17. The normal rulings for the Legendrian link from Example 2.5 are given in Figure 8. Thus the ruling polynomial is

$$R_\Lambda = z^{-1} + z.$$

2H. ***Legendrian links in*** $\mathbb{R}^3$**.** The classical invariants for Legendrian isotopy classes of knots in $\mathbb{R}^3$ are: topological knot type, Thurston–Bennequin number, and rotation number; see [Etnyre 2005]. The *Thurston–Bennequin number* of a knot measures the self-linking of a Legendrian knot Λ. Given a push off Λ' of Λ in a direction tangent to the contact structure, then $tb(\Lambda)$ is the linking number of Λ and Λ'. Given the xz-projection of Λ,

$$tb(\Lambda) = \mathrm{writhe}(\Lambda) - \tfrac{1}{2}(\#\,\mathrm{cusps}).$$

The *rotation number* $r(\Lambda)$ of an oriented Legendrian knot Λ is the rotation of its tangent vector field with respect to any global trivialization. (This definition agrees with the definition of the rotation number of a path given earlier.) Given the xz-projection of Λ,

$$r(\Lambda) = \tfrac{1}{2}(\#\,\mathrm{down\ cusps} - \#\,\mathrm{up\ cusps}).$$

Given a Legendrian link $\Lambda = \Lambda_1 \sqcup \cdots \sqcup \Lambda_n$, we define $tb_i = tb(\Lambda_i)$ and $r_i = r(\Lambda_i)$ for $1 \le i \le n$, and define

$$r(\Lambda) = \gcd(r_1, \ldots, r_n).$$

2I. ***Satellites, the DGA, and augmentations in*** $\mathbb{R}^3$**.** This section gives the results and notation for Legendrian links in $\mathbb{R}^3$ necessary to prove Theorem 1.3.

We will first extend the idea of satelliting a knot in $J^1(S^1)$ to an unknot (see [Ng and Rutherford 2013]) to satelliting each 1-handle of a knot in $\#^k(S^1 \times S^2)$ around a twice stabilized unknot.

Definition 2.18. Given the xy- or xz-diagram for a Legendrian link Λ in $\#^k(S^1 \times S^2)$, *satellited* Λ is denoted $S(\Lambda)$, the xy-diagram of which is depicted in Figure 10 and the xz-diagram of a Legendrian isotopic link of which is depicted in Figure 12 for the Legendrian link from Figure 6. Label the crossings as indicated, where $i \le j$ and label the basepoints in $S(\Lambda)$ as they are labeled in Λ. Note that the xy- or xz-diagram of Λ defines $S(\Lambda)$ up to Legendrian isotopy.

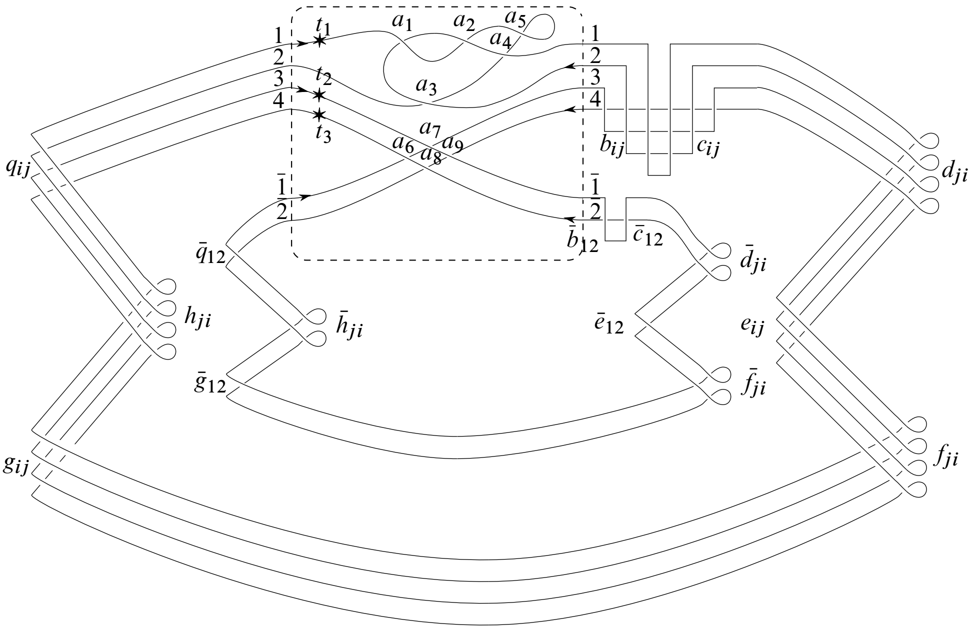


Figure 10. The xy-projection of the satellited link $S(\Lambda)$. The crossings in the c_{ij}-, b_{ij}-, $\bar{c}_{ij}$, and $\bar{b}_{ij}$-lattices are labeled as in Figure 4. The crossings in the d, e, f, g, h, q-lattices are labeled according to Figure 11.

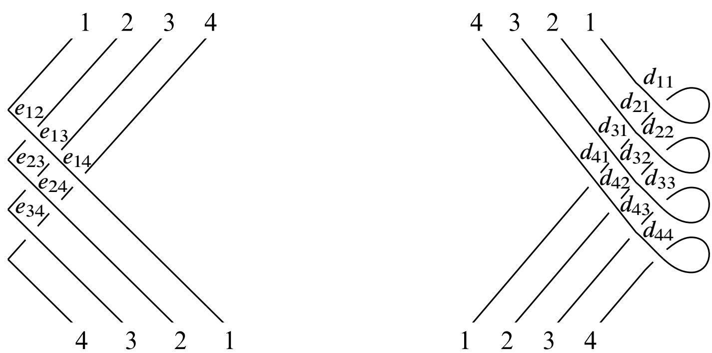


Figure 11. The labels for the crossings in the e- and d-lattices of the satellited link $S(\Lambda)$ as seen in Figure 10. The f- and h-lattices are analogous to the d-lattice. The g- and q-lattices are analogous to the e-lattice.

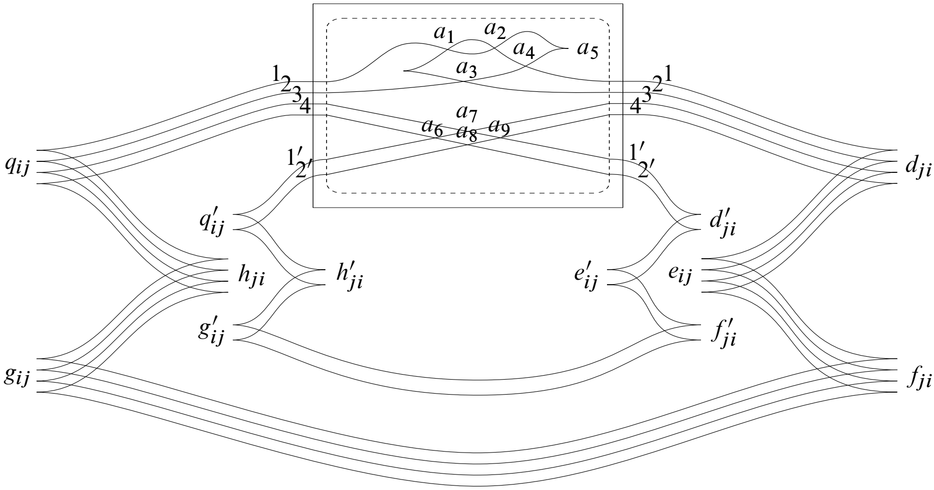


Figure 12. The xz-projection of a link which is Legendrian isotopic to the satellited link $S(\Lambda)$.

Remark 2.19. The Chekanov–Eliashberg DGA was originally defined on Legendrian links in $(\mathbb{R}^3, dz - y\,dx)$; see [Chekanov 2002; Sabloff 2005]. Note that the same DGA results from defining the DGA as we did in $\#^k(S^1 \times S^2)$ where $k = 0$.

2J. ***Dips.*** Dips will be defined analogously to those defined in [Leverson 2016].

Given a diagram $\pi_{xy}(\Lambda)$ in normal form which is the result of resolution, we construct a *dip* in the vertical slice of the diagram between two crossings, a crossing and a cusp, or two cusps, by a sequence of Reidemeister II moves, as seen in Figure 13 in the xz-projection and xy-projection. From the xz-projection, it is clear that the diagram with the dip is Legendrian isotopic to the original diagram. To construct a dip, number the N strands from top to bottom. Using a type II Reidemeister move, push strand $N-1$ over strand N, then strand $N-2$ over strand $N-1$, then strand $N-2$ over strand N, and so on. In this way, strand i is pushed over strand j in antilexicographic order.

Given an xy-diagram for a link $\Lambda \subset \mathbb{R}^3$ in normal form, where all crossings and resolutions of left cusps having distinct x-coordinates, the *dipped diagram* $D(\Lambda)$ is the result of adding a dip between each pair of crossings or resolution of a cusp and crossing. For each Reidemeister II move, we have two new generators. Call the left crossing b_{ij} and the right crossing c_{ij} if strands $i < j$ cross. One can check that $|b_{ij}| = m(j) - m(i)$ and since ∂ lowers degree by 1, we know $|c_{ij}| = |b_{ij}| - 1$.

While dipped diagrams have many more crossings than the original link diagram, the differential ∂ on $\mathcal{A}(D(\Lambda))$ is generally much simpler. In fact, a *totally augmented disk* (a disk from the definition of the differential of the DGA where all crossings at corners are augmented), cannot "go through" or "span" more than one dip.

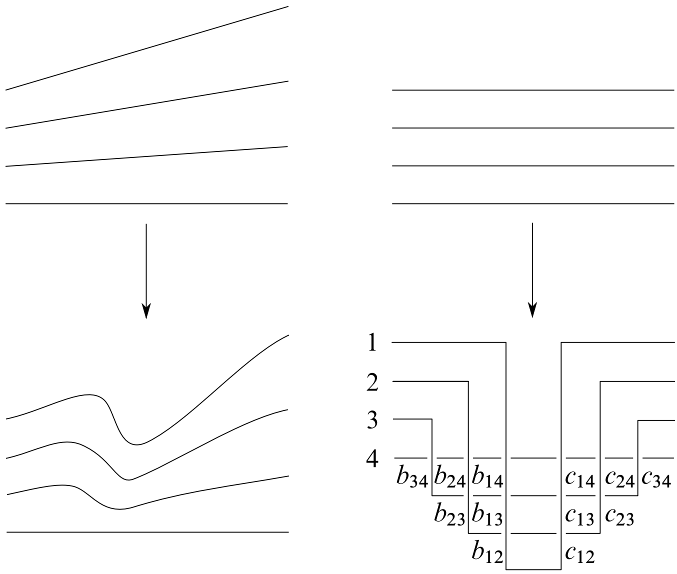


Figure 13. The modification of the xz-diagram when creating a dip (left) and the modification of the xy-diagram (right). (This figure is taken from [Leverson 2016].)

2K. ***Augmentations before and after basepoints and type II moves.*** In certain cases, we will find that adding basepoints will simplify the signs. For Legendrian links in $\mathbb{R}^3$, Ng and Rutherford [2013] give the DGA homomorphisms induced by adding a basepoint to a diagram and by moving a basepoint around a link. One can easily extend their results to $\#^k(S^1 \times S^2)$.

The following theorem is the analog of [op. cit, Theorem 2.21]:

Theorem 2.20. *Let $*_1, \dots, *_k$ and $*'_1, \dots, *'_k$ denote two collections of basepoints on the Lagrangian resolution of the front diagram of a Legendrian knot Λ, each of which is cyclically ordered along Λ, and let $(\mathcal{A}(\Lambda, *_1, \dots, *_k), \partial)$ and $\mathcal{A}(\Lambda, *'_1, \dots, *'_k), \partial')$ denote the corresponding multipointed DGAs. Then there is a DGA isomorphism $\Psi : (\mathcal{A}(\Lambda, *_1, \dots, *_k), \partial) \to (\mathcal{A}(\Lambda, *'_1, \dots, *'_k), \partial')$ such that $\Psi(t_i) = t_i$ for all i.*

In the proof of the theorem, Ψ is defined so that $\Psi(c) = c$ if no basepoint is moved over or under the crossing c. However, if the basepoint $*_i$ is moved over the crossing c, then $\Psi(c) = t_i^{\pm 1} c$, where the sign depends on whether the basepoint is moved along the knot following the orientation of the knot or against the orientation of the knot. If, instead, the basepoint is moved under the crossing c, then $\Psi(c) = c t_i^{\pm 1}$, where the sign, again, depends on the orientation of the knot.

Thus, If ϵ' is an augmentation of the DGA of the diagram after moving the basepoint $*_i$ over the crossing c, then $\epsilon = \epsilon' \circ \Psi$ is an augmentation of the DGA of the diagram before moving the basepoint.

The following theorem is the analog of [Ng and Rutherford 2013, Theorem 2.22]:

Theorem 2.21. *Let $*_1, \dots, *_k$ be a cyclically ordered collection of basepoints along Λ, and let $*$ be a single basepoint on Λ. Then there is a DGA homomorphism $\phi : (\mathcal{A}(\Lambda, *), \partial) \to (\mathcal{A}(\Lambda, *_1, \dots, *_k), \partial)$ such that $\phi \circ \partial = \partial \circ \phi$ and $\phi(t) = t_1 \cdots t_k$.*

Remark 2.22. In summary, if we have an augmentation $\epsilon : \mathcal{A} \to F$ with $\epsilon(t_i) = -1$, then moving the basepoint $*_i$ through a crossing c only changes the augmentation by changing the sign of the augmentation on the crossing c. Suppose we have a diagram with a basepoint $*$ corresponding to t and the same diagram with basepoints $*_1, \dots, *_s$ associated to $t_1, \dots, t_s$ on the same component of the link and we move all of the basepoints $*_1, \dots, *_s$ to the location of $*$. By the above results, if ϵ is an augmentation to F of the multiple basepoint diagram, there exists an augmentation ϵ' to F of the single basepoint diagram such that for all crossings c there exists $x_c \in F$ such that $\epsilon'(c) = x_c \epsilon(c)$ and

$$\epsilon'(t) = \epsilon(t_1 \cdots t_s) = \prod_{i=1}^{s} \epsilon(t_i).$$

Etnyre, Ng, and Sabloff [Etnyre et al. 2002] give a DGA isomorphism relating the DGA of a diagram of a Legendrian knot in $\mathbb{R}^3$ before and after a Reidemeister II move. One can easily extend this to a similar result for $\#^k(S^1 \times S^2)$, which gives a way to extend an augmentation of the diagram before a Reidemeister II move to an augmentation of the diagram after the move; see [Leverson 2016] for the analogous result in $\mathbb{R}^3$.

3. Correspondence between augmentations and normal rulings for links in $\mathbb{R}^3$

We have the following result for *knots* in $\mathbb{R}^3$:

Theorem 3.1 [Leverson 2016, Theorem 1.1]. *Let Λ be a Legendrian knot in $\mathbb{R}^3$. Given a field F, $(\mathcal{A}, \partial)$ has a ρ-graded augmentation $\epsilon : \mathcal{A} \to F$ if and only if any front diagram of Λ has a ρ-graded normal ruling. Furthermore, if ρ is even, then $\epsilon(t) = -1$.*

This result is proven by construction. Using the same method we can prove an analogous result for *links* in $\mathbb{R}^3$. Restating from the introduction:

Theorem 1.1. *Let Λ be an n-component Legendrian link in $\mathbb{R}^3$ with s basepoints (at least one basepoint on each component). Given a field F, the Chekanov–Eliashberg DGA $(\mathcal{A}, \partial)$ over $\mathbb{Z}[t_1^{\pm 1}, \dots, t_s^{\pm 1}]$ has a ρ-graded augmentation $\epsilon : \mathcal{A} \to F$ if and* *only if a front diagram of Λ has a ρ-graded normal ruling. Furthermore, if ρ is even, then $\epsilon(t_1 \cdots t_s) = (-1)^s$.*

The following result will be necessary for the proof of Theorem 1.1. Analogous to the knot case in $\mathbb{R}^3$, we have the following extension of [Leverson 2016, Lemma 3.2]:

Lemma 3.2. *If c gives the number of right cusps, sw is the number of switches in the ruling, a_- is the number of $-(a)$ crossings, and n is the number of components, then*

$$c + sw + a_- \equiv n \mod 2.$$

Proof. As in the knot case, one can easily show each of the following statements:

$$\sum_{i=1}^{n} tb_i + \sum_{i=1}^{n} r_i \equiv n \mod 2 \tag{1}$$

$$\sum_{i=1}^{n} tb_i \equiv c + cr \mod 2 \tag{2}$$

$$cr \equiv sw \mod 2 \tag{3}$$

$$\sum_{i=1}^{n} r_i \equiv a_- \mod 2 \tag{4}$$

where r_i is the rotation number of Λ_i and cr is the number of crossings. Note that if we add these four equations together, we get that

$$c + sw + a_- \equiv n \mod 2$$

as desired. □

Proof of Theorem 1.1. After a series of Legendrian isotopies, we can assume the front diagram of Λ has the following form where from left to right (lowest x-coordinate to highest x-coordinate) we have: all left cusps have the same x-coordinate, no two crossings of Λ have the same x-coordinate, and all right cusps have the same x-coordinate (in [Leverson 2016], this is called plat position). Label the crossings in the right cusps by $q_1, \ldots, q_m$ from top to bottom and label the other crossings by $c_1, \ldots, c_\ell$ from left to right.

Augmentation to ruling: Beginning with a ρ-graded augmentation of the Chekanov–Eliashberg DGA of the resolution of $\pi_{xz}(\Lambda)$ to a Lagrangian diagram, define a ρ-graded normal ruling of $\pi_{xz}(\Lambda)$ by simultaneously defining a ρ-graded augmentation of the dipped diagram $D(\Lambda)$ as in the knot case, using Figure 14.

Ruling to augmentation: Given a ρ-graded normal ruling of $\pi_{xz}(\Lambda)$, define a ρ-graded augmentation of the dipped diagram $D(\Lambda)$ with basepoints where specified in Figure 14 and at each right cusps as in the knot case, using Figure 14.

Using Lemma 3.2 and the methods in the proof of [Leverson 2016, Theorem 3.1], one can show the final statement of Theorem 1.1. Given a ρ-graded augmentation $\epsilon : \mathcal{A} \to F$, consider the associated ρ-graded normal ruling. If ρ is even, then the ruling is only switched at crossings c_k with $\rho \mid |c_k|$ and so $2 \mid |c_k|$. Thus, any strands paired by the ruling must have opposite orientation. As in the case of knots, this implies that near a crossing where the ruling is switched the crossing must be a positive crossing. Thus each ruling path is an oriented unknot.

If we consider the dipped diagram of the link, by induction we can show that

$$\prod \epsilon(b_{ij}^k)^{\pm 1} = 1,$$

where the product is taken over all paired strands i and j in the ruling between c_k and c_{k+1} and the sign is determined by the orientation of the paired strands as in [op. cit.]. By considering ∂q_k, we see that

$$\begin{aligned}\epsilon(t_1 \cdots t_s) = (-1)^{s-m} \prod_{k=1}^{m} \bigl(-\epsilon(b^\ell_{2k,2k-1})^{\pm 1}\bigr) \\ = (-1)^s \prod_{i<j \text{ paired}} \epsilon(b^\ell_{ij})^{\pm 1} = (-1)^s = (-1)^n\end{aligned}$$

by Lemma 3.2 and the fact that the number of basepoints $s \equiv c + sw + a_-$ mod 2. □

4. Augmentation to ruling

In this section, we will show that a quotient of the DGA of the satellited version of any Legendrian link Λ in $\#^k(S^1 \times S^2)$ is a subalgebra of the DGA of Λ in $\#^k(S^1 \times S^2)$ and use the construction from Theorem 1.1 to construct a ruling of the satellited link in $\mathbb{R}^3$ to then give a normal ruling of Λ in $\#^k(S^1 \times S^2)$. This shows the forward direction of Theorem 1.3.

Given an xy-diagram for the Legendrian link Λ in $\#^k(S^1 \times S^2)$ which results from the resolution of an xz-diagram in normal form with basepoints indicated. We can construct an xy-diagram for $S(\Lambda)$, satellited Λ, (see Figure 10) with basepoints in the same location as they were for Λ.

We will use the notation for Legendrian links in $\#^k(S^1 \times S^2)$ with tildes added for the Legendrian link Λ in $\#^k(S^1 \times S^2)$:

$$\mathcal{A}(\Lambda) = \mathbb{Z}[\tilde{t}_1^{\pm 1}, \dots, \tilde{t}_s^{\pm 1}]\langle \tilde{a}_i, \tilde{b}_{ij;\ell}, \tilde{c}^p_{ij;\ell} \rangle$$

with differential $\tilde{\partial}$, where $1 \le \ell \le k$, $i < j$ for all $\tilde{b}_{ij;\ell}$ and $i < j$ for $\tilde{c}^p_{ij;\ell}$ if $p = 1$. We will use the notation for Legendrian links from Figure 10 for $S(\Lambda)$:

$$\mathcal{A}(S(\Lambda)) = \mathbb{Z}[t_1^{\pm 1}, \dots, t_s^{\pm 1}]\langle a_i, b_{ij;\ell}, c_{ij;\ell}, d_{ji;\ell}, e_{ij;\ell}, f_{ji;\ell}, g_{ij;\ell}, h_{ji;\ell}, q_{ij;\ell} \rangle$$

with differential ∂, where $1\le \ell\le k$, $1\le i\le m$ for a_i, $i<j$ for $b_{ij;\ell}$, $c_{ij;\ell}$, $e_{ij;\ell}$, $g_{ij;\ell}$, and $q_{ij;\ell}$, and $i\le j$ for $d_{ji;\ell}$, $f_{ji;\ell}$, and $h_{ji;\ell}$.

Suppose we have a Legendrian link Λ in $\#^k(S^1\times S^2)$ with associated DGA $(\mathcal{A}(\Lambda),\partial)$. If $(\mathcal{A}(S(\Lambda)),\partial)$ is the DGA associated to satellited Λ, then let $\pi:\mathcal{A}(S(\Lambda))\to\mathcal{A}(S(\Lambda))/B$ be the quotient algebra homomorphism where B is the ideal in $\mathcal{A}(S(\Lambda))$ generated by

$$\{c_{ij;\ell}-g_{ij;\ell},\ c_{ij;\ell}-q_{ij;\ell},\ c_{ij;\ell}-(-1)^{|e_{ij;\ell}|+1}e_{ij;\ell},\\ h_{ji;\ell}-(-1)^{|f_{ji;\ell}|+1}f_{ji;\ell},\ h_{ji;\ell}-(-1)^{|d_{ji;\ell}|+1}d_{ji;\ell}\}.$$

Define $\gamma:\mathcal{A}(S(\Lambda))/B\to\mathcal{A}(\Lambda)$ by

$$\begin{aligned}\gamma\ :\ \mathcal{A}(S(\Lambda))/B &\longrightarrow \mathcal{A}(\Lambda)\\ [a_i] &\longmapsto \tilde{a}_i\\ [b_{ij;\ell}] &\longmapsto \tilde{b}_{ij;\ell}\\ [c_{ij;\ell}] &\longmapsto \tilde{c}^0_{ij;\ell}\\ [h_{ji;\ell}] &\longmapsto \tilde{c}^1_{ji;\ell}\\ [t_i] &\longmapsto \tilde{t}_i\end{aligned}$$

Proposition 4.1. *If $\phi=\gamma\circ\pi$, then ϕ is a graded algebra homomorphism such that $\tilde{\partial}\phi(c)=\pm\phi\partial(c)$ for all $c\in\{a_i,b_{ij;\ell},c_{ij;\ell},d_{ji;\ell},e_{ij;\ell},f_{ji;\ell},g_{ij;\ell},h_{ji;\ell},q_{ij;\ell}\}$.*

Proof. Grading: We will first show that π and γ (and thus ϕ) are graded algebra homomorphisms. First, let m be the Maslov potential used to assign the gradings of the crossings of Λ in $\#^k(S^1\times S^2)$. We will use m to define a Maslov potential μ on $S(\Lambda)$ in $\mathbb{R}^3$ as follows: Define μ on $T\subset S(\Lambda)$ the same as m is defined on $T\subset\Lambda$ and extend μ to the rest of $S(\Lambda)$. Notice that there is only one way to do this which keeps μ of the upper strand (higher z-coordinate) entering a cusp one higher than μ of the lower strand (lower z-coordinate) entering a cusp. The fact that ∂ has degree -1 and properties of the Maslov potential immediately give us that in the p-th 1-handle:

$$\begin{aligned}|\tilde{c}^1_{ji}|=|d_{ji}|=|f_{ji}|&=|h_{ji}|, &&i\le j\\ |\tilde{c}^0_{ij}|=|c_{ij}|=|e_{ij}|=|g_{ij}|&=|q_{ij}|, &&i<j\\ -|d_{ji}|&=|e_{ij}|, &&i<j\\ |b_{ij}|&=|c_{ij}|+1, &&i<j\end{aligned}\tag{5}$$

Thus, π and γ are graded algebra homomorphisms and so ϕ is as well.

$\tilde{\partial}\phi(c)=\pm\phi\partial(c)$: From the definition of their gradings, in the p-th 1-handle:

$$|\tilde{c}^0_{ij}|\equiv|\tilde{c}^0_{i\ell}|+|\tilde{c}^0_{\ell j}|\mod 2\quad\text{and}\quad|\tilde{c}^1_{ji}|\equiv|\tilde{c}^1_{j\ell}|+|\tilde{c}^1_{\ell i}|\mod 2\tag{6}$$

With (5), we have analogous statements for b_{ij}, c_{ij}, d_{ji}, e_{ij}, f_{ji}, g_{ij}, h_{ji}, and q_{ij}.

By considering the disks which contribute terms to ∂a_i and $\tilde{\partial}\tilde{a}_i$ (and analogously ∂b_{ij} and $\tilde{\partial}\tilde{b}_{ij}$ in the p-th 1-handle for $i < j$), it is clear that

$$\tilde{\partial}\phi(a_i)\tilde{\partial}(\tilde{a}_i) = \phi\partial(a_i) \quad \text{and} \quad \tilde{\partial}\phi(b_{ij})\tilde{\partial}(\tilde{b}_{ij}) = \phi\partial(b_{ij}).$$

Given $1 \le p \le k$ and $1 \le i < j \le N_p$. In the p-th 1-handle:

$$\begin{aligned}
\tilde{\partial}\phi c_{ij} &= \tilde{\partial}\tilde{c}^0_{ij} \\
&= \sum_{i<\ell<j} (-1)^{|\tilde{c}^0_{i\ell}|+1}\tilde{c}^0_{i\ell}\tilde{c}^0_{\ell j} \\
&= \phi\left(\sum_{i<\ell<j} (-1)^{|c_{ij}|+1} c_{i\ell}c_{\ell j}\right) && \text{by (5)} \\
&= \phi\partial c_{ij},
\end{aligned}$$

$$\begin{aligned}
\tilde{\partial}\phi d_{ii} &= \partial(-1)^{|d_{ii}|+1}\tilde{c}^1_{ii} \\
&= (-1)^{|d_{ii}|+1}\left(1 + \sum_{i<\ell\le N_p} (-1)^{|\tilde{c}^0_{i\ell}|+1}\tilde{c}^0_{i\ell}\tilde{c}^1_{\ell i} + \sum_{1\le\ell<i} (-1)^{|\tilde{c}^1_{i\ell}|+1}\tilde{c}^1_{i\ell}\tilde{c}^0_{\ell i}\right) \\
&= 1 + \sum_{i<\ell\le N_p} (-1)^{|\tilde{c}^0_{i\ell}|+|d_{\ell i}|}\phi(c_{i\ell}d_{\ell i}) \\
&\qquad + \sum_{1\le\ell<i} (-1)^{|\tilde{c}^1_{i\ell}|+|d_{i\ell}|+|e_{\ell i}|+1}\phi(d_{i\ell}e_{\ell i}) && \text{since } |d_{ii}| = 1 \\
&= 1 + \sum_{i<\ell\le N_p} \phi(c_{i\ell}d_{\ell i}) + \sum_{1\le\ell<i} (-1)^{|d_{i\ell}|+1}\phi(d_{i\ell}e_{\ell i}) && \text{by (5)} \\
&= \phi\partial d_{ii},
\end{aligned}$$

$$\begin{aligned}
\tilde{\partial}\phi d_{ji} &= \tilde{\partial}(-1)^{|d_{ji}|+1}\tilde{c}^1_{ji} \\
&= (-1)^{|d_{ji}|+1}\left(0 + \sum_{j<\ell\le N_p} (-1)^{|\tilde{c}^0_{j\ell}|+1}\tilde{c}^0_{j\ell}\tilde{c}^1_{\ell i} + \sum_{1\le\ell<i} (-1)^{|\tilde{c}^1_{j\ell}|+1}\tilde{c}^1_{j\ell}c^0_{\ell i}\right) \\
&= \phi\Bigg((-1)^{|d_{ji}|+1}\Bigg(0 + \sum_{j<\ell\le N_p} (-1)^{|\tilde{c}^0_{j\ell}|+|d_{\ell i}|}c_{j\ell}d_{\ell i} \\
&\qquad + \sum_{1\le\ell<i} (-1)^{|\tilde{c}^1_{j\ell}|+|d_{j\ell}|+|e_{\ell i}|+1}d_{j\ell}e_{\ell i}\Bigg)\Bigg)
\end{aligned}$$

$$\begin{aligned}
&= \phi\Bigg((-1)^{|d_{ji}|+1}\Bigg(0 + (-1)^{|d_{ji}|}\sum_{j<\ell\le N_p} c_{j\ell}d_{\ell i} \\
&\qquad\qquad + (-1)^{|d_{ji}|}\sum_{1\le\ell<i}(-1)^{|d_{j\ell}|+1}d_{j\ell}e_{\ell i}\Bigg)\Bigg) \qquad \text{by (5) and (6)} \\
&= -\phi\Bigg(0 + \sum_{j<\ell\le N_p} c_{j\ell}d_{\ell i} + \sum_{1\le\ell<i}(-1)^{|d_{j\ell}|+1}d_{j\ell}e_{\ell i}\Bigg) \\
&= -\phi\partial d_{ji}
\end{aligned}$$

One can similarly show that for $i<j$

$$\begin{aligned}
\tilde\partial\phi e_{ij} &= \phi\partial e_{ij}, \\
\tilde\partial\phi f_{ii} &= \phi\partial f_{ii}, \qquad \tilde\partial\phi f_{ji} = -\phi\partial f_{ji}, \\
\tilde\partial\phi g_{ij} &= \phi\partial g_{ij}, \qquad \tilde\partial\phi h_{ii} = \phi\partial h_{ii}, \\
\tilde\partial\phi h_{ji} &= \phi\partial h_{ji}, \\
\tilde\partial\phi q_{ij} &= \phi\partial q_{ij}.
\end{aligned}$$

□

Given a field F and a ρ-graded augmentation $\tilde\epsilon:\mathcal{A}(\Lambda)\to F$, we will construct a ρ-graded augmentation $\epsilon:\mathcal{A}(S(\Lambda))\to F$. Define $\epsilon=\tilde\epsilon\circ\phi$. Thus, on the generators of $\mathcal{A}(S(\Lambda))$ in the p-th 1-handle,

$$\epsilon(c) = \begin{cases}
\tilde\epsilon(\tilde a_i) & \text{if } c = a_i \\
\tilde\epsilon(\tilde b_{ij}) & \text{if } c = b_{ij} \\
\tilde\epsilon(\tilde c^0_{ij}) & \text{if } c \in \{c_{ij}, g_{ij}, q_{ij}\} \\
(-1)^{|\tilde c^0_{ij}|+1}\tilde\epsilon(\tilde c^0_{ij}) & \text{if } c = e_{ij} \\
\tilde\epsilon(\tilde c^1_{ji}) & \text{if } c = h_{ji} \\
(-1)^{|\tilde c^1_{ji}|+1}\tilde\epsilon(\tilde c^1_{ji}) & \text{if } c \in \{d_{ji}, f_{ji}\} \\
\tilde\epsilon(\tilde t_i) & \text{if } c = t_i.
\end{cases}$$

We see that ϵ is an augmentation because on any generator c of $\mathcal{A}(S(\Lambda))$,

$$\begin{aligned}
\epsilon\partial(c) &= \tilde\epsilon\phi\partial(c) \\
&= \pm\tilde\epsilon\tilde\partial\phi(c) \qquad \text{by Proposition 4.1} \\
&= 0,
\end{aligned}$$

since ϵ' is an augmentation. And, since ϵ' is a ρ-graded augmentation and ϕ is a graded algebra homomorphism, ϵ is a ρ-graded augmentation.

Thus an augmentation $\tilde\epsilon:\mathcal{A}(\Lambda)\to F$ of the DGA of Λ in $\#^k(S^1\times S^2)$ gives an augmentation $\epsilon:\mathcal{A}(S(\Lambda))\to F$ of the DGA of $S(\Lambda)$ in $\mathbb{R}^3$. By [Leverson 2016, Theorem 1.1], the augmentation ϵ gives an augmentation of the DGA of $S(\Lambda)$ with dips in $\mathbb{R}^3$, which gives a normal ruling of $S(\Lambda)$ with no dips in $\mathbb{R}^3$. We must check that if two strands are paired in this normal ruling, then they go through the same 1-handle. Clearly this normal ruling must be *thin*, meaning outside of the tangle T associated to Λ the ruling only has switches at crossings where the crossing strands go through the same 1-handle. By restricting the ρ-graded normal ruling of $S(\Lambda)$ in $\mathbb{R}^3$ to a ρ-graded normal ruling of T, we get a ρ-graded normal ruling of Λ in $\#^k(S^1 \times S^2)$.

5. Ruling to augmentation

Let F be a field. We will now prove the existence of a ρ-graded normal ruling implies the existence of a ρ-graded augmentation, the backward direction of Theorem 1.3, by constructing a ρ-graded augmentation $\epsilon : \mathcal{A}(D(\Lambda)) \to F$ given a ρ-graded normal ruling of Λ in $\#^k(S^1 \times S^2)$.

Given an xz-diagram of a Legendrian link Λ in $\#^k(S^1 \times S^2)$ in normal form, we will consider the resolution to an xy-diagram of a Legendrian isotopic link. Using Legendrian isotopy, we can ensure all crossings, left cusps, and right cusps have different x coordinates and all right cusps occur "above" (have higher y or z coordinate than) the remaining strands of the tangle at that x coordinate. Place a basepoint on every strand at $x = 0$ and one in every loop coming from the resolution of a right cusp.

Define the augmentation $\epsilon : \mathcal{A}(D(\Lambda)) \to F$ of the DGA for the dipped diagram $D(\Lambda)$ on generators as follows: If the ruling is switched at a crossing a_ℓ, then set $\epsilon(a_\ell) = 1$. If not, set $\epsilon(a_\ell) = 0$. (Note that we can augment the switched crossings to any nonzero element of F and still get an augmentation. But in the case where Λ is a knot, by augmenting the switched crossing to 1, we will be able to ensure $\epsilon(t) = -1$.) Add basepoints and augment the crossings in the dips following Figure 14. On the remaining generators, set

$$\epsilon(c_{ij}^\ell) = \begin{cases} 1 & \text{if } \ell = 0 \text{ and strands } i < j \text{ are paired in the normal ruling} \\ & \qquad \text{and go through the } p\text{-th 1-handle} \\ (-1)^{|c_{ij}^\ell|} & \text{if } \ell = 1,\ i > j, \text{ and strands } i, j \text{ are paired in the normal} \\ & \qquad \text{ruling and go through the } p\text{-th 1-handle} \\ 0 & \text{otherwise.} \end{cases}$$

Augment all basepoints to -1.

By considering Figure 14, it is involved but straightforward to check that ϵ is an augmentation on the a_ℓ and the crossings in the dips.

Notation 5.1. $c_{\{ij\}}^\ell = c_{\min(i,j),\,\max(i,j)}^\ell$

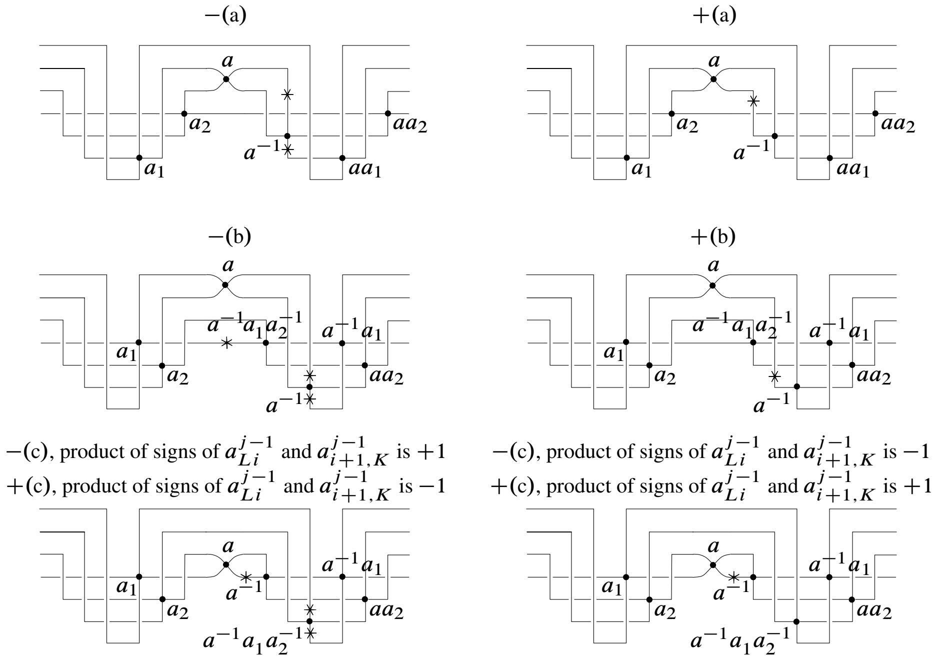


Figure 14. In the diagrams, $*$ denotes a basepoint. A dot denotes the specified crossing is augmented and the augmentation sends the crossing to the labeled value. For example, in the left dip of the $-$(a) configuration, $\epsilon(c_{12}) = a_1$ and $\epsilon(c_{34}) = a_2$. All other crossings are sent to 0 by the augmentation. Here $-/+$(a) denotes a negative/positive crossing where the ruling has configuration (a) and the rest are defined analogously. See Figure 15 for configurations (d), (e), and (f). (This figure is taken from [Leverson 2016].)

We will now check that ϵ is an augmentation on the c_{ij}^{ℓ} generators from the p-th 1-handle.

$\underline{\epsilon\partial c_{ij}^0 = 0}$: For any ruling, at the left end of the diagram, each strand is paired with another strand going through the same 1-handle. So for each strand i going through the p-th 1-handle, there exists a strand $j \neq i$ such that strand i and j are paired and $1 \leq i, j \leq N_p$. So if $i < j$, then $\epsilon(c_{ij}^0) = 1$, $\epsilon(c_{\{i\ell\}}^0) = 0$ for all $\ell \neq j$, and $\epsilon(c_{\{j\ell\}}^0) = 0$ for all $\ell \neq i$. Suppose $i < r < \ell$. Then $\epsilon(c_{ir}^0) = 0$ if $r \neq j$ and $\epsilon(c_{r\ell}^0) = 0$ if $r = j$. Thus $\epsilon(c_{ir}^0 c_{r\ell}^0) = 0$ for all $i < r < \ell$ and so for $i < \ell$,

$$\epsilon\partial c_{i\ell}^0 = \sum_{i<r<\ell} (-1)^{|c_{ir}^0|+1} \epsilon(c_{ir}^0 c_{r\ell}^0) = 0.$$

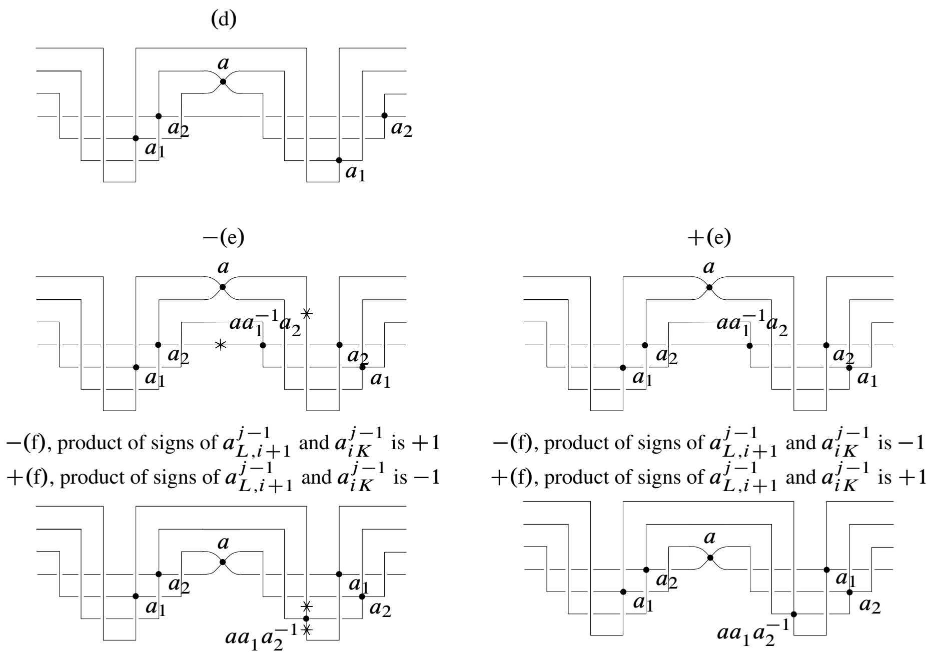


Figure 15. A continuation of Figure 14. (This figure is taken from [Leverson 2016].)

$\underline{\epsilon\partial c^1_{ij} = 0}$: Recall that in the p-th 1-handle

$$\partial c^1_{ij} = \delta_{ij} + \sum_{i<\ell\le N_p} (-1)^{|c^0_{i\ell}|+1} c^0_{i\ell} c^1_{\ell j} + \sum_{1\le \ell<j} (-1)^{|c^1_{i\ell}|+1} c^1_{i\ell} c^0_{\ell j}.$$

If $i \neq j$, then $\epsilon(c^0_{i\ell}c^1_{\ell j}) = 0$ and $\epsilon(c^1_{i\ell}c^0_{\ell j}) = 0$ for all ℓ since it is not possible for strand i to be paired with strand ℓ and for strand ℓ to be paired with strand j when $i \neq j$. Thus

$$\epsilon\partial c^1_{ij} = \sum_{i<\ell\le N_p} (-1)^{|c^0_{i\ell}|+1} \epsilon(c^0_{i\ell}c^1_{\ell j}) + \sum_{1\le\ell<j} (-1)^{|c^1_{i\ell}|+1} \epsilon(c^1_{i\ell}c^0_{\ell j}) = 0.$$

To show $\epsilon\partial c^1_{ii} = 0$, suppose strand i is paired with strand ℓ through the p-th 1-handle. Then by (5),

$$\begin{aligned}
\epsilon\partial c^1_{ii} &= \begin{cases} 1 + (-1)^{|c^0_{i\ell}|+1}\epsilon(c^0_{i\ell}c^1_{\ell i}), & i<\ell, \\ 1 + (-1)^{|c^1_{i\ell}|+1}\epsilon(c^1_{i\ell}c^0_{\ell i}), & i>\ell, \end{cases} \\
&= \begin{cases} 1 + (-1)^{|c^0_{i\ell}|+1}(-1)^{|c^1_{\ell i}|}, & i<\ell, \\ 1 + (-1)^{|c^1_{i\ell}|+1}(-1)^{|c^1_{i\ell}|}, & i>\ell, \end{cases} \\
&= 0.
\end{aligned}$$

$\epsilon\partial c_{ij}^{\ell} = 0$ for $1 < \ell$: Recall

$$\partial c_{ij}^{\ell} = \sum_{r=0}^{\ell} \sum_{s=1}^{N_p} (-1)^{|c_{is}^r|+1} c_{is}^r c_{sj}^{\ell-r}$$

for $1 < \ell$, $1 \le p \le k$, and $1 \le i, j \le N_p$. We will show that

$$\epsilon(c_{is}^r c_{sj}^{\ell-r}) = 0,$$

which implies that $\epsilon\partial c_{ij}^{\ell} = 0$. If $\ell > 2$, then for all $0 \le r \le \ell$, either $r > 1$ or $\ell - r > 1$, so $\epsilon(c_{is}^r c_{sj}^{\ell-r}) = 0$ for all i, j, s. If $\ell = 2$, then $r > 1$, $\ell - r > 1$, or $r = 1 = \ell - r$. The first and second case clearly imply $\epsilon(c_{is}^r c_{sj}^{\ell-r}) = 0$. In the final case, this is also clearly true, unless $i = j$ and strands i and s are paired in the ruling. In this case, either $i < s$ or $s < i = j$, so either $\epsilon(c_{is}^1) = 0$ or $\epsilon(c_{sj}^1) = 0$. So

$$\epsilon\partial c_{ii}^{\ell} = \sum_{r=0}^{\ell} \sum_{s=1}^{N_p} (-1)^{|c_{is}^r|+1} \epsilon(c_{is}^r c_{si}^{\ell-r}) = 0$$

for all $1 \le p \le k$, $1 \le i \le N_p$, and $\ell > 1$. So for $1 < \ell$

$$\epsilon\partial c_{ij}^{\ell} = 0.$$

Grading: From the definition, a_i is augmented only if the ρ-graded normal ruling is switched at a_i and thus $\rho \,\big|\, |a_i|$. Since $|a_i| = |\tilde{a}_i|$, we have $\rho \,\big|\, |a_i|$. By definition, if $c_{ij;p}^{\ell}$ is augmented, then either $\ell = 0$, $i < j$, and strands i and j are paired by the normal ruling and go through the p-th 1-handle or $\ell = 1$, $i > j$, and strands i and j are paired in the normal ruling and go through the p-th 1-handle. In the first case, $\mu(i) \equiv \mu(j) + 1 \mod \rho$ and so

$$|c_{ij;p}^{0}| = 2(0) - 1 + \mu(i) - \mu(j) \equiv 0 \mod \rho.$$

In the second case, $\mu(j) \equiv \mu(i) + 1 \mod \rho$ and so

$$|c_{ij;p}^{1}| = 2(1) - 1 + \mu(i) - \mu(j) \equiv 0 \mod \rho.$$

Following arguments similar to those in [Leverson 2016], one can also check that if a crossing c in a dip is augmented then $\rho \,\big|\, |c|$.

Proposition 5.2. *If $\Lambda \subset \#^k(S^1 \times S^2)$ is an n-component link, $\rho \mid 2r(\Lambda)$ is even, and Λ has a ρ-graded normal ruling, then the ρ-graded augmentation $\epsilon : \mathcal{A}(\Lambda) \to F$ constructed above sends $t_1 \cdots t_s$ to $(-1)^n$.*

Proof. Given a ρ-graded ruling of Λ in $\#^k(S^1 \times S^2)$, there is a unique way to extend it to a normal ruling of $S(\Lambda)$ by switching at $d_{ji}, e_{ij}, f_{ji}, g_{ij}, h_{ji}, q_{ij}$ if and only if strands $i < j$ are paired in the ruling of Λ. Let $\tilde{\epsilon} : \mathcal{A}(\Lambda) \to F$ be the ρ-graded augmentation resulting from the ρ-graded normal ruling and let $\epsilon : \mathcal{A}(S(\Lambda)) \to F$

be the ρ-graded augmentation resulting from the ρ-graded normal ruling of $S(\Lambda)$ as constructed in [Leverson 2016] in $\mathbb{R}^3$. Note that

$$\frac{\epsilon(t_1\cdots t_s)}{\tilde{\epsilon}(\tilde{t}_1\cdots\tilde{t}_r)}=\left(\prod_{1\le p\le k}(-1)^{3N_p}\right)\prod_{i,j\text{ paired}}(-1)^6.$$

If strands $i<j$ are paired near $x=0$ in the ruling of Λ, then the ruling of $S(\Lambda)$ must be switched at d_{ji}, e_{ij}, f_{ji}, g_{ij}, h_{ji}, and q_{ij} with configuration $+$(a) since the ruling is ρ-graded and ρ is even. So there is one additional basepoint augmented to -1 per crossing. Thus, there are six additional basepoints augmented to -1 for each pair of strands. Each right cusp contributes one extra basepoint augmented to -1 and there are three additional right cusps for each strand. However, N_p is even for all $1\le p\le k$ by Corollary 1.4 and $\epsilon(t_1\cdots t_s)=(-1)^n$ by Theorem 1.1, so

$$\frac{(-1)^n}{\tilde{\epsilon}(\tilde{t}_1\cdots\tilde{t}_r)}=1$$

and so $\tilde{\epsilon}(\tilde{t}_1\cdots\tilde{t}_r)=(-1)^n$. □

All that remains to be proven is the final statement of Theorem 1.3, which says:

Proposition 5.3. *Given a field F, if Λ is an n component link in $\#^k(S^1\times S^2)$, $\epsilon(t)=(-1)^n$ for all even-graded augmentations $\epsilon:\mathcal{A}(\Lambda)\to F$.*

Proof. Suppose that $\tilde{\epsilon}:\mathcal{A}(\Lambda)\to F$ is an even-graded augmentation (ρ-graded augmentation where $2\mid\rho$). As in Section 4, we construct a ρ-graded augmentation $\epsilon:\mathcal{A}(S(\lambda))\to F$. By definition, $\epsilon(t_i)=\tilde{\epsilon}(\tilde{t}_i)$ for all $1\le i\le s$ and so

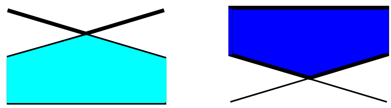
$$\tilde{\epsilon}(\tilde{t}_1\cdots\tilde{t}_s)=\epsilon(t_1\cdots t_s)=(-1)^n,$$

where the final equality follows from Theorem 1.1. □

6. Correspondence for links in $J^1(S^1)$

Recall that the 1-jet space of the circle, $J^1(S^1)$, is diffeomorphic to the solid torus $S^1_x\times\mathbb{R}^2_{y,z}$ with contact structure given by $\xi=\ker(dz-y\,dx)$. As in [Ng and Traynor 2004], by viewing S^1 as a quotient of the unit interval, $S^1=[0,1]/(0\sim 1)$, we can see Legendrian links in $J^1(S^1)$ as quotients of arcs in $I\times\mathbb{R}^2$ with boundary conditions which are everywhere tangent to the contact planes. Given a Legendrian link $\Lambda\subset J^1(S^1)$ we will use the methods of Lavrov and Rutherford [2012] to show the following theorem, restated from the introduction:

Theorem 1.6. *Suppose Λ is a Legendrian link in $J^1(S^1)$. Given a field F, the Chekanov–Eliashberg DGA $(\mathcal{A},\partial)$ over $\mathbb{Z}[t_1^{\pm1},\dots,t_s^{\pm1}]$ has a ρ-graded augmentation $\epsilon:\mathcal{A}\to F$ if and only if a front diagram of Λ has a ρ-graded generalized normal ruling.*



(g) (h)

Figure 16. These figures give the configuration of a generalized normal ruling near a switched crossing involving exactly one self-paired strand. With the top row of configurations in Figure 7, these are all possible configurations of a generalized normal ruling near a switched crossing.

We recall the definition of generalized normal ruling.

Definition 6.1 [Lavrov and Rutherford 2012]. A *generalized normal ruling* is a sequence of involutions $\sigma = (\sigma_1, \dots, \sigma_M)$ as in Definition 2.11 with the following differences:

(1) Remove the requirement that σ_m is fixed-point-free and the condition about 1-handles.

(2) If strands ℓ and $\ell+1$ cross in the interval (x_{m-1}, x_m) above I_{m-1}, where exactly one of the crossing strands is a fixed point of σ_m, then the crossing is a switch if σ_m satisfies the conditions in (3) of Definition 2.11. If crossing is a switch, then we require an additional normality condition:

$$\sigma_m(\ell) = \ell < \ell+1 < \sigma_m(\ell+1) \quad \text{or} \quad \sigma_m(\ell) < \ell < \ell+1 = \sigma_m(\ell+1).$$

A *strictly generalized normal ruling* is a generalized normal ruling which is not a normal ruling, in other words, a generalized normal ruling with at least one fixed point.

Thus, near a crossing, a generalized normal ruling looks like the crossings in Figure 7 or Figure 16.

Remark 6.2. (1) If a crossing involving strands ℓ and $\ell+1$ occurs in the interval (x_{m-1}, x_m) and both crossing strands are fixed by the ruling, self-paired, in other words, $\sigma_{m-1}(\ell) = \ell$ and $\sigma_{m-1}(\ell+1) = \ell+1$, then $\sigma_m = (\ell \quad \ell+1) \circ \sigma_{m-1} \circ (\ell \quad \ell+1)$ and so we will not consider such crossings to be switched.

(2) Note that the number of generalized normal rulings of a Legendrian link is not invariant under Legendrian isotopy.

The definition of the Chekanov–Eliashberg DGA of a Legendrian link in $\mathbb{R}^3$ can be extended to Legendrian links in $J^1(S^1)$. (One can find the full definition of the Chekanov–Eliashberg DGA of a Legendrian link in $J^1(S^1)$ in [Ng and Traynor 2004].) Note that given an augmentation of the Chekanov–Eliashberg DGA over

$\mathbb{Z}[t, t^{-1}]$ of a Legendrian link in $S^1 \times S^2$, one can define an augmentation of the DGA of the analogous link (where if a strand goes through the 1-handle with $y = y_0$ at $x = 0$, then it is paired with the strand going through the 1-handle with $y = y_0$ at $x = A$) in $J^1(S^1)$ and similarly for normal rulings. (The resulting normal ruling of the link in $J^1(S^1)$ will not have any self-paired strands.) However, there is no reason to think the converse is true.

6A. ***Matrix definition of the DGA in $J^1(S^1)$.*** Ng and Traynor [2004] define a version of the Chekanov–Eliashberg DGA $\mathcal{A}$ over $R = \mathbb{Z}[t, t^{-1}]$ in. For ease of definition, note that we can assume all left and right cusps involve the two strands with lowest z-coordinate (and thus highest labels) and that there is one basepoint at $x = 0$ on each strand with the basepoint on strand i corresponding to t_i, and one basepoint in each loop resulting from the resolution of a right cusp. We give the definition of the DGA for the dipped version Λ, $D(\Lambda)$ as in [Lavrov and Rutherford 2012] with an extra dip immediately to the right of the basepoints at $x = 0$. Label the dips as in Figure 13 with b_{ij}^m and c_{ij}^m in the dip at x_m. Place these generators in upper triangular matrices

$$B_m = (b_{ij}^m) \text{ and } C_m = (c_{ij}^m).$$

Note that since the x-coordinate is S^1-valued, we need to add the convention that $B_0 = B_M$ and $C_0 = C_M$. We then see that

$$\begin{aligned}
\partial C_m &= (\Sigma_m C_m)^2, \\
\partial B_1 &= T C_0 T^{-1}(I + B_1) - \Sigma_1 (I + B_1) \Sigma_1 C_1, \\
\partial B_m &= \widetilde{C}_{m-1}(I + B_m) - \Sigma_m (I + B_m) \Sigma_m C_m,
\end{aligned}$$

where Σ_m is the diagonal matrix with $(-1)^{\mu_m(i)}$ the i-th entry on the diagonal for Maslov potential μ_m at $x = x_m$, T is the diagonal matrix with $t_i^{o_1(i)}$ the i-th entry on the diagonal where

$$o_m(i) = \begin{cases} -1 & \text{if strand } i \text{ is oriented to the right at } x = x_m, \\ 1 & \text{otherwise,} \end{cases}$$

and I is the appropriately sized identity matrix. The form of $\widetilde{C}_m$ will depend on the tangle appearing in the interval (x_{m-1}, x_m).

If (x_{m-1}, x_m) contains a crossing a_m of strands k and $k+1$, then

$$\begin{aligned}
\partial a_m &= c_{k,k+1}^{m-1}, \\
\widetilde{C}_{m-1} &= U_{k,k+1} \widehat{C}_{m-1} V_{k,k+1},
\end{aligned}$$

where $U_{k,k+1}$ and $V_{k,k+1}$ are the identity matrix with the 2×2 block in rows k and $k+1$ and columns k and $k+1$ replaced with

$$\begin{pmatrix} 0 & 1 \\ 1 & (-1)^{|a_m|+1} a_m \end{pmatrix}$$

for $U_{k,k+1}$ and

$$\begin{pmatrix} a_m & 1 \\ 1 & 0 \end{pmatrix}$$

for $V_{k,k+1}$, and $\widehat{C}_{m-1}$ is C_{m-1} with 0 replacing the entry $c^{m-1}_{k,k+1}$.

If (x_{m-1}, x_m) contains a left cusp, by assumption strands $N(m)-1$ and $N(m)$ are incident to the cusp. In this case,

$$\widetilde{C}_{m-1} = J C_{m-1} J^T + W,$$

where J is the $N(m-1) \times N(m-1)$ identity matrix with two rows of zeroes added to the bottom and W is $N(m) \times N(m)$ matrix where the $(N(m)-1, N(m))$-entry is 1 and all other entries are zero.

Finally, if (x_{m-1}, x_m) contains a right cusp a_m with basepoint $*_\alpha$ corresponding to t_α in the loop, by assumption strands $N(m)-1$ and $N(m)$ are incident to the cusp. In this case

$$\begin{aligned} \partial a_m &= t_\alpha^{o_{m-1}(N(m-1)-1)} + c^{m-1}_{N(m-1)-1,N(m-1)}, \\ \widetilde{C}_{m-1} &= K C_{m-1} K^T, \end{aligned}$$

where K is the $N(m-1) \times N(m-1)$ identity matrix with two columns of zeroes added to the right.

6B. ***Proof of correspondence.*** We will use the methods of [Lavrov and Rutherford 2012] to prove Theorem 1.3. Given an involution σ of $\{1, \dots, N\}$, $\sigma^2 = \mathrm{id}$, we define $A_\sigma = (a_{ij})$ the $N \times N$ matrix with entries

$$a_{ij} = \begin{cases} 1 & \text{if } i < \sigma(i) = j, \\ 0 & \text{otherwise.} \end{cases}$$

Ruling to augmentation: Given a generalized normal ruling $\sigma = (\sigma_1, \dots, \sigma_M)$, we will define a ρ-graded augmentation $\epsilon : \mathcal{A}(D(\Lambda)) \to F$ satisfying Property (R) (as in [Sabloff 2005]) by defining ϵ on the crossings in the dip involving crossings b^0_{ij} and c^0_{ij} and extending to the right.

Property (R): In any dip, the generator c^m_{rs} is augmented (to 1) if and only if $\sigma_m(r) = s$.

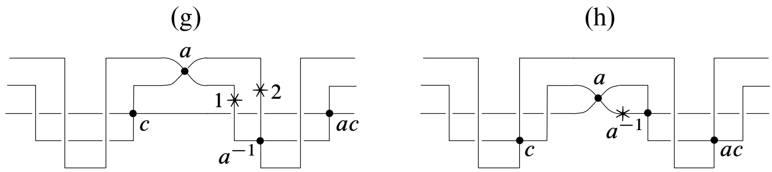


Figure 17. In the diagrams, $*_i$ denotes the basepoint associated to t_i. A dot denotes the specified crossing is augmented and the augmentation sends the crossing to the label. In configuration (g), $\epsilon(t_1) = (-1)^{|a|+1}$ and $\epsilon(t_2) = (-1)^{|c_{i,i+1}|+1}$. In configuration (h), $\epsilon(t) = -1$.

Add a basepoint to the loop in each resolution of a right cusp. Augment all basepoints to -1. Given a crossing a, set

$$\epsilon(a) = \begin{cases} 1 & \text{if the ruling is switched at } a, \\ 0 & \text{otherwise.} \end{cases}$$

Define $\epsilon(B_0) = 0$ and $\epsilon(C_0) = A_{\sigma_0}$. We will now extend ϵ to the right. Suppose ϵ is defined on all crossings in the interval $(0, x_{m-1})$. If (x_{m-1}, x_m) contains a crossing, define ϵ on crossings b_{ij}^m and c_{ij}^m and add basepoints as in Figure 14 and Figure 17. If (x_{m-1}, x_m) contains a left cusp, set

$$\epsilon(B_m) = J\epsilon(B_{m-1})J^T + W.$$

If (x_{m-1}, x_m) contains a right cusp, set

$$\epsilon(B_m) = K\epsilon(B_{m-1})K^T.$$

It is easy to check that by our definition the augmentation satisfies Property (R), which tells us $\epsilon(B_0) = \epsilon(B_M)$ and $\epsilon(C_0) = \epsilon(C_M)$, and our augmentation is a ρ-graded augmentation.

Augmentation to ruling: This direction of the proof follows that of the $\mathbb{Z}/2$ case in [Lavrov and Rutherford 2012] and is based on canonical form results from linear algebra due to Barannikov [1994].

Definition 6.3. An *M-complex* $(V, \mathcal{B}, d)$ is a vector space V over a field F with an ordered basis $\mathcal{B} = \{v_1, \ldots, v_N\}$ and a differential $d : V \to V$ of the form $dv_i = \sum_{j=i+1}^N a_{ij} v_j$ satisfying $d^2 = 0$.

The following two propositions are essentially in [Lavrov and Rutherford 2012, Propositions 5.4 and 5.6] and [Barannikov 1994, Lemmas 2 and 4].

Proposition 6.4. *Suppose that* $(V, \mathcal{B}, d)$ *is an* M*-complex, then there exists a triangular change of basis* $\{\tilde{v}_1, \ldots, \tilde{v}_N\}$ *with* $\tilde{v}_i = \sum_{j=i}^N a_{ij} v_j$ *and an involution*

$\tau : \{1, \dots, N\} \to \{1, \dots, N\}$ *such that*

$$d\tilde{v}_i = \begin{cases} \tilde{v}_j, & \text{if } i < \tau(i) = j, \\ 0, & \text{otherwise}. \end{cases}$$

Moreover, *the involution* τ *is unique.*

Remark 6.5. We have the following properties of the involution:

(1) If the basis elements v_i have been assigned degrees $|v_i| \in \mathbb{Z}/\rho$ such that V is $\mathbb{Z}/\rho$-graded and d has degree -1, then it can be assumed that the change of basis preserves degree. Thus, if $i < \tau(i) = j$, then $|v_i| = |v_j| + 1$.

(2) The set $\{[\tilde{v}_i] : \tau(i) = i\}$ forms a basis for the homology $H(V, d)$.

(3) In matrix formulation, according to Proposition 6.4, there is a unique function $D \mapsto \tau(D)$ which assigns an involution $\tau = \tau(D)$ to each strictly upper triangular matrix D with $D^2 = 0$ and there is an invertible upper triangular matrix P so that $PDP^{-1} = A_\tau$. The uniqueness statement tells us that $\tau(QDQ^{-1}) = \tau(D)$ if Q is a nonsingular upper triangular matrix.

Proposition 6.6. *Suppose* $(V, \mathcal{B}, d)$ *is an* M*-complex and* $k \in \{1, \dots, N\}$ *such that* $dv_k = \sum_{j=k+2}^{N} a_{kj} v_j$ *so the triple* $(V, \mathcal{B}', d)$ *with* $\mathcal{B}' = \{v_1, \dots, v_{k+1}, v_k, \dots, v_N\}$ *is also an* M*-complex. Then the associated involutions* τ *and* τ' *from* Proposition 6.4 *are related as follows*:

(1) *If*

$$\begin{aligned} &\tau(k+1) < \tau(k) < k < k+1, \\ &\tau(k) < k < k+1 < \tau(k+1), \\ &k < k+1 < \tau(k+1) < \tau(k), \\ &\tau(k) < k < k+1 = \tau(k+1), \\ &\tau(k) = k < k+1 < \tau(k+1), \end{aligned}$$

then either $\tau' = \tau$ *or* $\tau' = (k\ \ k+1) \circ \tau \circ (k\ \ k+1)$.

(2) *Otherwise* $\tau' = (k\ \ k+1) \circ \tau \circ (k\ \ k+1)$.

Augmentation to ruling: This part of the proof is the same as the analogous statement in [Lavrov and Rutherford 2012] with $\Sigma_{m-1}\epsilon(C_{m-1})$ replacing $\epsilon(Y_{m-1})$.

Suppose $\epsilon : \mathcal{A}(D(\Lambda)) \to F$ is a ρ-graded augmentation. Then for all m, $\epsilon(C_m)$ is an $N(m) \times N(m)$ strictly upper triangular matrix such that

$$0 = \epsilon\partial C_m = (\Sigma_m \epsilon(C_m))^2.$$

As in Remark 6.5, we can set $\tau_m = \tau(\Sigma_m C_m)$ and obtain the sequence $\tau = \{\tau_0, \dots, \tau_M\}$ of involutions where τ_m is an involution of $\{1, \dots, N(m)\}$. We will show that τ satisfies the requirements of a generalized normal ruling (see Definition 6.1).

We also have $N(m) \times N(m)$ strictly upper triangular matrices $\epsilon(B_m)$ which satisfy

$$
\begin{aligned}
0 = \epsilon\partial B_1 &= T\epsilon(C_0)T^{-1}(I+\epsilon(B_1)) - \Sigma_1(I+\epsilon(B_1))\Sigma_1\epsilon(C_1),\\
0 = \epsilon\partial B_m &= \epsilon(\widetilde{C}_{m-1})(I+\epsilon(B_m)) - \Sigma_m(I+\epsilon(B_m))\Sigma_m\epsilon(C_m).
\end{aligned}
$$

In the case where $m = 1$, this tells us

$$
\begin{aligned}
\Sigma_1\epsilon(C_1) &= (I+\epsilon(B_1))^{-1}\Sigma_1 T\epsilon(C_0)T^{-1}(I+\epsilon(B_1))\\
&= (I+\epsilon(B_1))^{-1}T\Sigma_1\epsilon(C_0)T^{-1}(I+\epsilon(B_1))
\end{aligned}
$$

since T and Σ_1 are diagonal matrices. So Remark 6.5 tells us

$$
\begin{aligned}
\tau_1 = \tau(\Sigma_1\epsilon(C_1)) = \tau((I+\epsilon(B_1))^{-1}T\Sigma_1\epsilon(C_0)T^{-1}(I+\epsilon(B_1)))\\
= \tau(\Sigma_1\epsilon(C_0)) = \tau(\Sigma_0\epsilon(C_0)) = \tau_0
\end{aligned}
$$

since $\Sigma_1 = \Sigma_0$ and $T^{-1}(I+\epsilon(B_1))$ is a nonsingular upper triangular matrix. Thus τ_1 satisfies the definition of generalized normal ruling since τ_0 does.

More generally, for $m > 1$, we have

$$
\Sigma_m\epsilon(C_m) = (I+\epsilon(B_m))^{-1}\Sigma_m\epsilon(\widetilde{C}_{m-1})(I+\epsilon(B_m)).
$$

So Remark 6.5 tells us

$$
\tau_m = \tau(\Sigma_m\epsilon(C_m)) = \tau(\Sigma_m\epsilon(\widetilde{C}_{m-1})).
$$

Recall that $\widetilde{C}_{m-1}$ depends on whether the interval (x_{m-1}, x_m) contains a left cusp, right cusp, or crossing.

Crossing: In the case where the interval (x_{m-1}, x_m) contains a crossing a_m of strands k and $k+1$, recall that $0 = \epsilon\partial(a_m) = \epsilon(c^{m-1}_{k,k+1})$. In this case,

$$
\widetilde{C}_{m-1} = U_{k,k+1}\widehat{C}_{m-1}V_{k,k+1},
$$

where $\widehat{C}_{m-1}$ is C_{m-1} with 0 replacing the entry $c^{m-1}_{k,k+1}$. Thus $\epsilon(\widehat{C}_{m-1}) = \epsilon(C_{m-1})$. So $\epsilon(\widetilde{C}_{m-1}) = \epsilon(U_{k,k+1}C_{m-1}V_{k,k+1})$. Note that $\mu_{m-1}(k) = \mu_m(k+1)$ and $\mu_{m-1}(k+1) = \mu_m(k)$, so $\Sigma_{m-1} = P_{k,k+1}\Sigma_m P_{k,k+1}$. We also see that

$$
\begin{aligned}
\Sigma_m U_{k,k+1} &= \Sigma_m P_{k,k+1}(I+(-1)^{|a_m|+1}\epsilon(a_m)E_{k,k+1})\\
&= P_{k,k+1}(I-\epsilon(a_m)E_{k,k+1})P_{k,k+1}\Sigma_m P_{k,k+1}\\
&= P_{k,k+1}(I-\epsilon(a_m)E_{k,k+1})\Sigma_{m-1},\\
V_{k,k+1} &= (I+\epsilon(a_m)E_{k,k+1})P_{k,k+1},
\end{aligned}
$$

where $E_{k,k+1}$ is a matrix with a single nonzero entry of 1 in the $(k, k+1)$ position. Thus

$$\Sigma_m\epsilon(\widetilde{C}_{m-1}) \\ = P_{k,k+1}(I-\epsilon(a_m)E_{k,k+1})\,\Sigma_{m-1}\epsilon(C_{m-1})(I+\epsilon(a_m)E_{k,k+1})P_{k,k+1}.$$

Since the $(k, k+1)$-entry of $(I-\epsilon(a_m)E_{k,k+1})\,\Sigma_{m-1}\epsilon(C_{m-1})(I+\epsilon(a_m)E_{k,k+1})$ is 0 no matter the value of $\epsilon(a_m)$, the matrix $\Sigma_m\epsilon(\widetilde{C}_{m-1})$ is strictly upper triangular. Therefore

$$\tau_m=\tau(\Sigma_m\epsilon(C_m))=\tau(\Sigma_m\epsilon(\widetilde{C}_{m-1}))$$

and

$$\tau((I-\epsilon(a_m)E_{k,k+1})\,\Sigma_{m-1}\epsilon(C_{m-1})(I+\epsilon(a_m)E_{k,k+1})) \\ =\tau(\Sigma_{m-1}\epsilon(C_{m-1}))=\tau_{m-1}$$

are related as in Proposition 6.6. So, as τ_{m-1} satisfies the conditions of a generalized normal ruling, so does τ_m. The left and right cusp cases follow similarly.

As in Remark 6.5, $\Sigma_m\epsilon(C_m)$ denotes the matrix of an M-complex with basis $v_1,\dots,v_{N(m)}$ corresponding to the strands of Λ at x_m. If ϵ is ρ-graded with respect to μ, then we can assign the gradings $|v_i|=\mu_m(i)$ and the differential will have degree -1. So (1) of Remark 6.5 tells us that the resulting involution $\tau_m=\tau(\Sigma_m\epsilon(C_m))$ is ρ-graded and thus τ is ρ-graded.

6C. *Corollary.* The following proposition uses certain techniques from the proof of Theorem 1.6 to show that

$$\mathrm{Aug}_\rho(\Lambda)=F\backslash 0$$

for any field F and any ρ if Λ has a strictly generalized normal ruling.

Proposition 6.7. *Given a field F and a Legendrian link $\Lambda\subset J^1(S^1)$ with n components and a strictly generalized normal ruling, for all $0\neq x\in F$ there exists an augmentation $\epsilon:\mathcal{A}\to F$ such that*

$$\epsilon(t_1\cdots t_s)=x.$$

Proof. Fix $0\neq x\in F$. Given a generalized normal ruling $\sigma=(\sigma_1,\dots,\sigma_M)$ for Λ with a self-paired strand, we will construct an augmentation $\epsilon:\mathcal{A}(D(\Lambda))\to F$ such that $\epsilon(t_1\cdots t_s)=x$.

Suppose k is the label at $x=0$ of a self-paired strand of the generalized normal ruling σ, in other words, $\sigma_0(k)=k$. We can assume that $D(\Lambda)$ has one basepoint corresponding to t_i on strand i at $x=0$ and one basepoint in the loop in the resolution of each right cusp, and no other basepoints. Define

$$\epsilon(t_i) = \begin{cases} (-1)^{N+c-1}x & \text{if } i = k, \\ -1 & \text{otherwise,} \end{cases}$$

where c is the number of right cusps and N is the number of strands at $x = 0$.

Define ϵ on all crossings as in the proof of ruling to augmentation in Theorem 1.6. Note that t_k does not appear on the boundary of any totally augmented disks and so ϵ is still an augmentation, but now

$$\epsilon(t_1 \cdots t_s) = x$$

as desired. □

Remark 6.8. For any link $\Lambda \subset J^1(S^1)$, one can consider the analogous link $\Lambda' \subset S^1 \times S^2$. Note that $\mathcal{A}(\Lambda) \to \mathcal{A}(\Lambda')$ where the map is inclusion. Thus, any augmentation $\epsilon' : \Lambda' \to F$ gives an augmentation $\epsilon : \Lambda \to F$. As one would expect from Theorems 1.3 and 1.6, it is also clear that any normal ruling of $\Lambda' \subset S^1 \times S^2$ gives a generalized normal ruling of $\Lambda \subset J^1(S^1)$.

Appendix

The appendix will address Corollary 1.5 which follows from

(1) Theorem 1.3 over $\mathbb{Q}$, and

(2) the result that if a graded augmentation to the rationals exists then the full symplectic homology is nonzero.

The second result is known to experts; assumes the results of [Bourgeois et al. 2012]. We will outline the proof here for completeness. Statement (2) is a straight forward consequence of work of Bourgeois, Ekholm, and Eliashberg [Bourgeois et al. 2012] and has previously been observed in [Lidman and Sivek 2016].

Every connected Weinstein (Stein) 4-manifold X can be decomposed into 1- and 2-handle attachments to D^4 along $\partial D^4 = S^3$. Thus, for each such 4-manifold there exists a Legendrian link Λ in $\#^k(S^1 \times S^2)$ so that attaching 2-handles along Λ to $\#^k(S^1 \times S^2)$ results in X.

Results of Bourgeois, Ekholm, and Eliashberg (using their notation) tell us the following:

Proposition A.1 [Bourgeois et al. 2012, Corollary 5.7].

$$S\mathbb{H}(X) = L\mathbb{H}^{\mathrm{Ho}}(\Lambda),$$

where $L\mathbb{H}^{\mathrm{Ho}}(\Lambda)$ is the homology of the Hochschild complex associated to the Chekanov–Eliashberg differential graded algebra over $\mathbb{Q}$.

Therefore, if the DGA for Λ has a graded augmentation to $\mathbb{Q}$, then $S\mathbb{H}(X)$ is nonzero. By Theorem 1.3, we know that the DGA for Λ has a graded augmentation to $\mathbb{Q}$ if and only if Λ has a graded normal ruling. Thus, restated from the introduction:

Corollary 1.5. *If X is the Weinstein 4-manifold that results from attaching 2-handles along a Legendrian link Λ to $\#^k(S^1 \times S^2)$ and Λ has a graded normal ruling, then the full symplectic homology $S\mathbb{H}(X)$ is nonzero.*

For completeness, we give an outline of the proof of statement (2). Recall that full symplectic homology is a symplectic invariant of Weinstein 4-manifolds which coincides with the Floer–Hofer symplectic homology.

We will show that given a graded augmentation ϵ' of the Chekanov–Eliashberg DGA of a Legendrian link Λ over $\mathbb{Z}[t,t^{-1}]$ to $\mathbb{Q}$, one can define a graded augmentation $\epsilon : LH^{\text{Ho}}(\Lambda) \to \mathbb{Q}$, where the homology of $LH^{\text{Ho}}(\Lambda)$ is $L\mathbb{H}^{\text{Ho}}(\Lambda)$. Recall that

$$LH^{\text{Ho}}(\Lambda) = \widecheck{\text{LHO}^+}(\Lambda) \oplus \mathbb{Q}\langle \tau_1, \ldots, \tau_n\rangle \oplus \widehat{\text{LHO}^+}(\Lambda)$$

is generated by elements of the form $\check{w}$, τ_i, and $\hat{v}$, where $w, v \in \text{LHO}(\Lambda) \subset \text{LHA}(\Lambda)$ and n is the number of components of the link. Define

$$\epsilon : LH^{\text{Ho}}(\Lambda) \to \mathbb{Q}$$

by $\check{w} \mapsto \epsilon'(w)$, $\tau_i \mapsto 1$, $\hat{v} \mapsto 0$. Let us check that this gives an augmentation. It suffices to check the generators. Clearly $\epsilon \circ d_{\text{Ho}}(\tau_i) = 0$ for all i. If $d_{\text{LHO}^+}(w) = \sum_{j=1}^r w_j$, then we recall that

$$d_{\text{Ho}}(\check{w}) = d_{\text{Ho}_+}(\check{w}) + \delta_{\text{Ho}}(\check{w}) = \check{d}_{\text{LHO}^+}(\check{w}) + \delta_{\text{Ho}}(\check{w}).$$

Let w be a chord in $\text{LHO}^+(\Lambda)$. Then, there exists i such that $w \in C_i$ and

$$d_{\text{Ho}}(\check{w}) = \sum_{j=1}^r \check{w}_j + \alpha_{wi}\tau_i,$$

where α_{wi} is the algebraic number of components of the 1-dimensional moduli space of holomorphic disks with one positive and no negative boundary punctures. Thus

$$\epsilon \circ d_{\text{Ho}}(\check{w}) = \sum_{j=1}^r \epsilon'(w_j) + \sum_{i=1}^n \alpha_{wi} = \epsilon' \circ d_{\text{LHO}}(w) = 0,$$

since α_{wi} is exactly the constant term of $d_{\text{LHA}}(w)$, ϵ' is an augmentation of $\text{LHA}(\Lambda)$, $\text{LHO}(\Lambda) \subset \text{LHA}(\Lambda)$, and $d_{\text{LHO}} = d_{\text{LHA}}|_{\text{LHO}}$. If $w \in \text{LHO}^+(\Lambda)$ is a linearly composable monomial which is not a chord, then

$$d_{\mathrm{Ho}}(\check{w}) = \sum_{j=1}^{r} \check{w}_j$$

and so

$$\epsilon \circ d_{\mathrm{Ho}}(\check{w}) = \sum_{j=1}^{r} \epsilon'(w_j) = \epsilon' \circ d_{\mathrm{LHO}}(w) = 0$$

since $d_{\mathrm{LHO}}(w)$ does not have a constant term.

If $v = c_1 \cdots c_\ell \in \mathrm{LHO}^+(\Lambda)$, then we recall that

$$\begin{aligned} d_{\mathrm{Ho}}(\hat{v}) &= d_{\mathrm{Ho}_+}(\hat{v}) + \delta_{\mathrm{Ho}}(\hat{v}) = d_{M\ Ho_+}(\hat{v}) + \hat{d}_{\mathrm{LHO}^+}(\hat{v}) + 0 \\ &= \check{c}_1 c_2 \cdots c_\ell - c_1 \cdots c_{\ell-1}\check{c}_\ell + \hat{d}_{\mathrm{LHO}^+}(\hat{v}) \\ &= \check{c}_1 c_2 \cdots c_\ell - (-1)^{|c_\ell|(|c_1|+\cdots+|c_{\ell-1}|)}\check{c}_\ell c_1 \cdots c_{\ell-1} + \hat{d}_{\mathrm{LHO}^+}(\hat{v}). \end{aligned}$$

Thus

$$\begin{aligned} &\epsilon \circ d_{\mathrm{Ho}}(\hat{v}) \\ &\qquad = \epsilon'(c_1 \cdots c_\ell) - (-1)^{|c_\ell|(|c_1|+\cdots+|c_{\ell-1}|)}\epsilon'(c_\ell c_1 \cdots c_{\ell-1}) + \epsilon' \circ \hat{d}_{\mathrm{LHO}^+}(\hat{v}) \\ &\qquad = \epsilon'(c_1 \cdots c_\ell) - (-1)^{|c_\ell|(|c_1|+\cdots+|c_{\ell-1}|)}\epsilon'(c_\ell c_1 \cdots c_{\ell-1}) + 0 \\ &\qquad = 0 \end{aligned}$$

since ϵ' is a graded augmentation of $\mathrm{LHA}(\Lambda)$ so if $\epsilon(c_1 \cdots c_\ell) \neq 0$, then $\epsilon(c_i) \neq 0$ for all i and thus $|c_i| = 0$ for all i.

Acknowledgements

The author thanks Lenhard Ng and Dan Rutherford for many helpful discussions and the referee for many valuable comments. This work was partially supported by NSF grants DMS-0846346 and DMS-1406371.

References

[Barannikov 1994] S. A. Barannikov, "The framed Morse complex and its invariants", pp. 93–115 in *Singularities and bifurcations*, edited by V. I. Arnol'd, Adv. Soviet Math. **21**, American Mathematical Society, Providence, RI, 1994. MR Zbl

[Bourgeois et al. 2012] F. Bourgeois, T. Ekholm, and Y. Eliashberg, "Effect of Legendrian surgery", *Geom. Topol.* **16**:1 (2012), 301–389. MR Zbl

[Chekanov 2002] Y. Chekanov, "Differential algebra of Legendrian links", *Invent. Math.* **150**:3 (2002), 441–483. MR Zbl

[Ekholm and Ng 2015] T. Ekholm and L. Ng, "Legendrian contact homology in the boundary of a subcritical Weinstein 4-manifold", *J. Differential Geom.* **101**:1 (2015), 67–157. MR Zbl

[Eliashberg 1987] Y. M. Eliashberg, "A theorem on the structure of wave fronts and its application in symplectic topology", *Funktsional. Anal. i Prilozhen.* **21**:3 (1987), 65–72. MR

[Eliashberg 1998] Y. Eliashberg, "Invariants in contact topology", pp. 327–338 in *Proceedings of the International Congress of Mathematicians* (Berlin, 1998), vol. 2, extra volume of *Documenta Mathematica*, Deutsche Mathematiker-Vereinigung, Bielefeld, 1998. MR Zbl

[Etnyre 2005] J. B. Etnyre, "Legendrian and transversal knots", pp. 105–185 in *Handbook of knot theory*, edited by W. Menasco and M. Thistlethwaite, Elsevier, Amsterdam, 2005. MR Zbl

[Etnyre et al. 2002] J. B. Etnyre, L. L. Ng, and J. M. Sabloff, "Invariants of Legendrian knots and coherent orientations", *J. Symplectic Geom.* **1**:2 (2002), 321–367. MR Zbl

[Fuchs 2003] D. Fuchs, "Chekanov–Eliashberg invariant of Legendrian knots: existence of augmentations", *J. Geom. Phys.* **47**:1 (2003), 43–65. MR Zbl

[Fuchs and Ishkhanov 2004] D. Fuchs and T. Ishkhanov, "Invariants of Legendrian knots and decompositions of front diagrams", *Mosc. Math. J.* **4**:3 (2004), 707–717. MR Zbl

[Gompf 1998] R. E. Gompf, "Handlebody construction of Stein surfaces", *Ann. of Math.* (2) **148**:2 (1998), 619–693. MR Zbl

[Lavrov and Rutherford 2012] M. Lavrov and D. Rutherford, "Generalized normal rulings and invariants of Legendrian solid torus links", *Pacific J. Math.* **258**:2 (2012), 393–420. MR Zbl

[Lavrov and Rutherford 2013] M. Lavrov and D. Rutherford, "On the $S^1 \times S^2$ HOMFLY-PT invariant and Legendrian links", *J. Knot Theory Ramifications* **22**:8 (2013), art. id. 1350040, 21 pp. MR

[Leverson 2016] C. Leverson, "Augmentations and rulings of Legendrian knots", *J. Symplectic Geom.* **14**:4 (2016), 1089–1143. MR

[Lidman and Sivek 2016] T. Lidman and S. Sivek, "Contact structures and reducible surgeries", *Compos. Math.* **152**:1 (2016), 152–186. MR Zbl

[Ng and Rutherford 2013] L. Ng and D. Rutherford, "Satellites of Legendrian knots and representations of the Chekanov–Eliashberg algebra", *Algebr. Geom. Topol.* **13**:5 (2013), 3047–3097. MR Zbl

[Ng and Traynor 2004] L. Ng and L. Traynor, "Legendrian solid-torus links", *J. Symplectic Geom.* **2**:3 (2004), 411–443. MR Zbl

[Pushkar and Chekanov 2005] P. E. Pushkar' and Y. V. Chekanov, "Combinatorics of fronts of Legendrian links and the Arnol'd 4-conjectures", *Uspekhi Mat. Nauk* **60**:1(361) (2005), 99–154. In Russian; translated in *Russian Math. Surveys* **60**:1 (2005), 95–149. MR Zbl

[Sabloff 2005] J. M. Sabloff, "Augmentations and rulings of Legendrian knots", *Int. Math. Res. Not.* **2005**:19 (2005), 1157–1180. MR Zbl

Received November 10, 2015. Revised September 2, 2016.

CAITLIN LEVERSON
SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
SCHOOL OF MATHEMATICS
686 CHERRY STREET
ATLANTA, GA 30332
UNITED STATES

leverson@math.gatech.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.425

THE FABER–KRAHN INEQUALITY FOR THE FIRST EIGENVALUE OF THE FRACTIONAL DIRICHLET p-LAPLACIAN FOR TRIANGLES AND QUADRILATERALS

FRANCO OLIVARES CONTADOR

We prove the Faber–Krahn inequality for the first eigenvalue of the fractional Dirichlet p-Laplacian for triangles and quadrilaterals of a given area. The proof is based on a nonlocal Pólya–Szegő inequality under Steiner symmetrization and the continuity of the first eigenvalue of the fractional Dirichlet p-Laplacian with respect to the convergence, in the Hausdorff distance, of convex domains.

1. Introduction and main result

The classical isoperimetric problem reads as follows: "among all domains in $\mathbb{R}^n$ of a given volume with rectifiable boundary, the sphere has the minimum perimeter."

In line with this, various isoperimetric problems have been studied (see [Osserman 1978]). For example, the Faber–Krahn inequality, originally conjectured in [Rayleigh 1894, 339–340], can be stated as follows: "among all open sets of a given volume in Euclidean space the ball minimizes the first eigenvalue of the Dirichlet Laplacian."

The Faber–Krahn inequality for variants of the Laplacian or by restriction to special classes of domains have generated interest in recent years. In fact, inspired by the Faber–Krahn inequality, Pólya and Szegő [1951] conjectured that among all polygons with n sides of fixed area, the regular n-polygon of the same area minimizes the first eigenvalue of the Dirichlet Laplacian. This conjecture is known to hold for $n = 3$ and $n = 4$, but for n-gons with $n \geq 5$ it still remains a conjecture. On the other hand, the Faber–Krahn inequality has been generalized, for example, to the case of the Dirichlet p-Laplacian [Bhattacharya 1999; Ly 2005; Chorwadwala et al. 2015; Toledo Oñate 2012].

The author acknowledges the financial support of CONICYT through the grants CONICYT-PCHA/Magíster Nacional/2013-22131441 and FONDECYT Regular 1130595.

MSC2010: primary 49Q20, 52B60; secondary 47A75, 47J10.

Keywords: fractional p-Laplacian, Dirichlet condition, Faber–Krahn inequality, polygonal domains, Steiner symmetrization, Riesz's inequality.

Recently, partial differential equations involving nonlocal versions of the Laplacian and in particular eigenvalue problems involving such operators have generated a lot of interest and have been studied (see [Di Nezza et al. 2012; Lindgren and Lindqvist 2014; Frank et al. 2008; Brasco et al. 2014]).

The first eigenvalue of fractional Dirichlet p-Laplacian is defined as follows:

Definition. Let $n \geq 1,\ 0 < s < 1$ and $1 < p < \infty$. Given an open and bounded set $\Omega \subset \mathbb{R}^n$ we define

$$\lambda_{1,p}^{s}(\Omega) = \inf\left\{ \frac{\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{|u(x)-u(y)|^p}{|x-y|^{n+ps}}\,dx\,dy}{\int_{\mathbb{R}^n}|u(x)|^p\,dx} : u \in \widetilde{W}_0^{s,p}(\Omega) \quad \text{and} \quad u \not\equiv 0 \right\}, \tag{1-1}$$

where $\widetilde{W}_0^{s,p}(\Omega)$ is the closure of $C_0^\infty(\Omega)$ with respect to the norm

$$u \mapsto [u]_{W^{s,p}(\mathbb{R}^n)} + \|u\|_p \tag{1-2}$$

where $[u]_{W^{s,p}(\mathbb{R}^n)}$ is defined in (2-1).

Inspired by the nonlocal Faber–Krahn inequality proved in [Brasco et al. 2014] for the fractional Dirichlet p-Laplacian and the Pólya–Szegő conjecture for the usual Laplacian for polygonal domains, we prove a Faber–Krahn inequality for the fractional Dirichlet p-Laplacian in the class of polygonal domains. This is our main result.

Theorem 1.1. *The equilateral triangle has the least first eigenvalue for the fractional Dirichlet p-Laplacian among all triangles of given area. The square has the least first eigenvalue for the fractional Dirichlet p-Laplacian among all quadrilaterals of given area. Moreover, the equilateral triangle and the square are the unique minimizers in the above problems.*

For proving this result we shall study the effect of Steiner symmetrization in nonlocal functionals and the continuity properties of the first eigenvalue of the fractional Dirichlet p-Laplacian with respect to the Hausdorff convergence of convex domains. In particular we will prove the following two results which will be used in the proof of Theorem 1.1:

Proposition 1.2 (nonlocal Pólya–Szegő inequality). *Let $n \geq 1,\ 0 < s < 1,\ 1 \leq p \leq n/s$ and $u \in \widetilde{W}_0^{s,p}(\Omega)$. Then*

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{|u^\star(x) - u^\star(y)|^p}{|x-y|^{n+ps}}\,dx\,dy \leq \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{|u(x)-u(y)|^p}{|x-y|^{n+ps}}\,dx\,dy, \tag{1-3}$$

where $u^\star$ is the Steiner symmetrization of u with respect to a given hyperplane. If $p > 1$, then equality holds if and only if u is proportional to a translate of a function which is symmetric with respect to the hyperplane.

Proposition 1.3. *Let B be a fixed compact set in $\mathbb{R}^n$ and Ω_n be a family of convex open subsets of B which converges, for the Hausdorff distance, to a set Ω. Furthermore, assume that there exist $r > 0$ such that $B(0, r) \subset \Omega_n$ and $B(0, r) \subset \Omega$. Then $\lambda^s_{1,p}(\Omega) = \lim_{n\to\infty} \lambda^s_{1,p}(\Omega_n)$.*

The basic definitions, notions and results which will be used in this paper are to be given in the next section.

2. Tools

Fractional Sobolev spaces and the first eigenvalue. Let $p \in [1, \infty)$ and $s \in (0, 1)$. Then

$$[u]_{W^{s,p}(\mathbb{R}^n)} = \left(\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{|u(x) - u(y)|^p}{|x - y|^{n+ps}} \, dx \, dy \right)^{1/p} \tag{2-1}$$

denotes the (s, p)-Gagliardo seminorm in $\mathbb{R}^n$ of a measurable function u. The Gagliardo seminorm satisfies the following Poincaré-type inequality:

Proposition 2.1 (Poincaré-type inequality). *Let $1 \le p < \infty$ y $s \in (0, 1)$, $\Omega \subset \mathbb{R}^n$ be an open and bounded set. There then exists a constant $C_{n,s,p}$, depending only on n, s, p and Ω, so that, for every function $u \in C^\infty_0(\Omega)$ we have*

$$\|u\|^p_p \le C_{n,s,p}(\Omega)[u]^p_{W^{s,p}(\mathbb{R}^n)}$$

Proof. See Lemma 2.4, [Brasco et al. 2014]. □

Proposition 2.1 shows that for an open and bounded set $\Omega \subset \mathbb{R}^n$ the space $\widetilde{W}^{s,p}_0(\Omega)$ can be equivalently defined as the closure of $C^\infty_0(\Omega)$ with respect to the Gagliardo seminorm. The space $\widetilde{W}^{s,p}_0(\Omega)$ is a reflexive Banach space for $1 < p < \infty$.

Theorem 2.2 (Rellich–Kondrachov theorem). *Let $p \in [1, \infty)$ and $s \in (0, 1)$, $\Omega \subset \mathbb{R}^n$ be a open and bounded set. Let $\{u_n\}^\infty_{n=1} \subset \widetilde{W}^{s,p}_0(\Omega)$ be a bounded sequence. Then there exists a subsequence $\{u_{n_k}\}^\infty_{k=1}$ of $\{u_n\}^\infty_{n=1}$ which converges strongly in $L^p(\Omega)$ to a function u. Moreover, if $p > 1$ then $u \in \widetilde{W}^{s,p}_0(\Omega)$.*

Proof. See Theorem 2.7, [Brasco et al. 2014]. □

Remark 1. Following Theorem 2.2, it can be shown that the infimum in (1-1) is a minimum and by the homogeneity of the Rayleigh quotient, the expression (1-1) can be written as

$$\lambda^s_{1,p}(\Omega) = \min\left\{ \|u\|^p_{\widetilde{W}^{s,p}_0(\Omega)} : u \in \widetilde{W}^{s,p}_0(\Omega),\ \|u\|_p = 1 \right\}. \tag{2-2}$$

Observe also that $\lambda^s_{1,p}(\Omega)$ equals the inverse of the best constant in the Poincaré inequality (Proposition 2.1).

The minimizer in (1-1) satisfies the following Euler–Lagrange equation

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{|u(x)-u(y)|^{p-2}(u(x)-u(y))(\phi(x)-\phi(y))}{|x-y|^{n+ps}}\,dx\,dy = \lambda_{1,p}^s(\Omega)\int_{\mathbb{R}^n}|u(x)|^{p-2}u(x)\phi(x)\,dx, \tag{2-3}$$

for all $\phi \in \widetilde{W}_0^{s,p}(\Omega)$ (see Theorem 5, [Lindgren and Lindqvist 2014]).

One can easily check that the following properties hold:

Proposition 2.3. *Let $\Omega \subset \mathbb{R}^n$ be an open and bounded set.*

(1) (*Homothety law*) $\lambda_{1,p}^s(t\Omega) = t^{-sp}\lambda_{1,p}^s(\Omega)$ *for* $t > 0$.

(2) (*Translation invariance*) $\lambda_{1,p}^s(\Omega) = \lambda_{1,p}^s(\Omega + x)$ *for all* $x \in \mathbb{R}^n$.

(3) (*Invariance under orthonormal transformations*) $\lambda_{1,p}^s(\Omega) = \lambda_{1,p}^s(T(\Omega))$ *for every orthonormal transformation* T.

(4) (*Domain monotony*) *If* $A \subset B$ *are open sets*, *then* $\lambda_{1,p}^s(B) \le \lambda_{1,p}^s(A)$.

Steiner symmetrizations of sets and functions. Let $n \ge 2$ and $\Omega \subset \mathbb{R}^n$ be a measurable set. We denote by Ω' the projection of Ω in the x_n-direction:

$$\Omega' := \{x' \in \mathbb{R}^{n-1} : \text{ there exists } x_n \text{ such that } (x', x_n) \in \Omega\},$$

and, for $x' \in \mathbb{R}^{n-1}$, we denote by $\Omega(x')$ the section of Ω in x':

$$\Omega(x') := \{x_n \in \mathbb{R} : (x', x_n) \in \Omega\},\ x' \in \Omega'.$$

Definition. Let $\Omega \subset \mathbb{R}^n$ be a measurable set. The set

$$\Omega^\star := \left\{x = (x', x_n) : -\tfrac{1}{2}|\Omega(x')| < x_n < \tfrac{1}{2}|\Omega(x')|,\ x' \in \Omega'\right\} \tag{2-4}$$

is the Steiner symmetrization of Ω with respect to the hyperplane $x_n = 0$. In the above, $|\Omega(x')|$ denotes the one-dimensional Lebesgue measure of $\Omega(x')$.

The Steiner symmetrization of a convex set with respect to a given hyperplane can be similarly defined.

A convex body is a compact convex set. For a convex body A in $\mathbb{R}^n$, the inradius $r(A)$ is the maximum of the radii of balls contained in A and the circumradius $R(A)$ is the minimum of the radii of balls containing A.

The Steiner symmetrization of sets has the following properties:

Proposition 2.4. *Let A, B be convex bodies. Then*

(1) $A^\star \subseteq B^\star$ *for* $A \subseteq B$.

(2) $r(A) \le r(A^\star)$.

(3) $R(A^\star) \le R(A)$.

(4) $V(A) = V(A^\star)$ *where* $V(A)$ *denotes the volume of* A.

Proof. See Proposition 9.1, page 169–171 of [Gruber 2007]. □

Definition. Let f be a nonnegative measurable function defined on Ω, which vanishes on $\partial\Omega$. The Steiner symmetrization of f is the function $f^\star$ defined on $\Omega^\star$ by

$$f^\star(x) = \sup\{c : x \in \{y \in \Omega : f(y) \geq c\}^\star\}. \tag{2-5}$$

The Steiner symmetrization of functions has the following properties.

Proposition 2.5. (1) *The definitions of $A^\star$ and $f^\star$ are consistent, i.e.,*

$$\chi_{A^\star} = (\chi_A)^\star \quad \textit{and} \quad \{x : f(x) \geq t\}^\star = \{x : f^\star(x) \geq t\}.$$

(2) *Let f and g be two nonnegative measurable functions such that $f(x) \leq g(x)$. Then $f^\star(x) \leq g^\star(x)$.*

(3) *Let $\Phi : \mathbb{R}^+ \to \mathbb{R}^+$ be a nondecreasing function. Then $(\Phi \circ f)^\star = \Phi \circ f^\star$.*

(4) *Let f be a nonnegative measurable function defined on Ω vanishing on $\partial\Omega$. Let $F : \mathbb{R}^+ \to \mathbb{R}$ be a measurable function. Then,*

$$\int_\Omega F(f(x))\,dx = \int_{\Omega^\star} F(f^\star(x))\,dx.$$

(5) *Let f, g and h be nonnegative measurable functions on $\mathbb{R}^n$. Then with $I(f, g, h) = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(x)g(x-y)h(y)\,dx\,dy$, we have*

$$I(f, g, h) \leq I(f^\star, g^\star, h^\star). \tag{2-6}$$

Moreover, if g is strictly symmetric decreasing, then there is equality in (2-6) if only if $f(x) = f^\star(x-y)$ and $h(x) = h^\star(x-y)$ almost everywhere for some $y \in \mathbb{R}^n$.

Proof. The proof of (1)–(4) is straightforward. For the proof of (5), we refer to Theorem 3.7, page 87 and Theorem 3.9, page 93 of [Lieb and Loss 2001] and [Brascamp et al. 1974] . □

For J a nonnegative, convex function on $\mathbb{R}$ with $J(0) = 0$ and k a nonnegative measurable function on $\mathbb{R}^n$, we let

$$E[u] = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} J(u(x) - u(y))k(x-y)\,dx\,dy.$$

Following the same ideas given in Lemma A.2. of [Frank and Seiringer 2008], using principally part (5) of Proposition 2.5 for Steiner symmetrization instead of symmetric decreasing rearrangement, we get the following lemma:

Lemma 2.6. *Let J be a nonnegative, convex function on $\mathbb{R}$ with $J(0) = 0$ and let $k \in L_1(R^n)$ be a nonnegative function which is symmetric and decreasing. Then for* *all nonnegative measurable u with $E[u]$ and $|\{u > \tau\}|$ finite for all $\tau > 0$ one has*

$$E[u] \geq E[u^\star],$$

with $u^\star$ the Steiner symmetrization of u with respect a hyperplane. If, in addition, J is strictly convex and k is strictly decreasing, then equality holds if and only if u is a translate of a function which is symmetric with respect to the hyperplane.

Hausdorff distance.

Definition. Let K_1 and K_2 be two nonempty compact sets in $\mathbb{R}^n$. Taking $d(x, K_2) := \inf\{|y - x| : y \in K_2\}$ for $x \in \mathbb{R}^n$, we set

$$\rho(K_1, K_2) := \sup\{d(x, K_2) : x \in K_1\}.$$

Let C^n be the family of compact subsets of $\mathbb{R}^n$. It is a metric space when equipped with the Hausdorff distance

$$d^H(K_1, K_2) := \max(\rho(K_1, K_2), \rho(K_2, K_1)). \tag{2-7}$$

For open sets inside a fixed compact set, we define the Hausdorff distance through their complement.

Definition. Let O_1, O_2 be two open sets of a compact set B. Then their Hausdorff distance is defined by

$$d_H(O_1, O_2) = d^H(B \setminus O_1, B \setminus O_2). \tag{2-8}$$

The Minkowski addition and Minkowski difference.

Definition. The Minkowski addition of two sets $A, B \subset \mathbb{R}^n$ can be defined by

$$A \oplus B := \bigcup_{b \in B} (A + b). \tag{2-9}$$

Definition. The Minkowski difference of two sets $A, B \subset \mathbb{R}^n$ can be defined by

$$A \ominus B := \bigcap_{b \in B} (A - b). \tag{2-10}$$

Clearly, we may also write $A \ominus B := \{x \in \mathbb{R}^n : B + x \subset A\}$. If $B = -B$, then

$$A \ominus B := \bigcap_{b \in B} (A + b).$$

The following proposition can be obtained without much difficulty using the above definition:

Proposition 2.7. *Let A, B and C be subsets of $\mathbb{R}^n$ such that $B = -B$, $A \subset C$ and $B \subset C$. Then*

$$A \ominus B \subseteq C \setminus ((C \setminus A) \oplus B).$$

Recall that, $K \ominus B(0,\epsilon)$ is the inner parallel body of K at distance ϵ. The main tool in the proof of Proposition 1.2 is the following lemma, which states that a suitable contraction of a convex body is contained in the inner parallel body of the convex body.

Lemma 2.8. *Let K be a convex body in $\mathbb{R}^n$, with $B(0,r) \subset K \subset B(0,R)$ for some numbers $r > 0$ and $R > 0$. If $0 < \epsilon < r^2/4R$, then*

$$\Big(1 - 4\frac{R\epsilon}{r^2}\Big)K \subset K \ominus B(0,\epsilon) \subset K. \tag{2-11}$$

Proof. See Lemma 2.3.6, page 93 of [Schneider 2014]. □

3. Proofs

The proof of Proposition 1.2 is given in Theorem A.1 of [Frank and Seiringer 2008] for the symmetric decreasing rearrangement. We sketch the proof of the adaptation to the case of Steiner symmetrization for the sake of completeness.

Proof of Proposition 1.2. Since $u^\star(x)$ is nonnegative and $\big||u(x)| - |u(y)|\big| \le |u(x) - u(y)|$, it suffices to prove the theorem for nonnegative functions. By definition of the Gamma function and following a change of variables we obtain

$$\frac{1}{\Gamma(\frac{n+ps}{2})}\int_0^\infty \alpha^{\frac{n+ps}{2}-1} \mathrm{e}^{-\alpha|x-y|^2}\, d\alpha = \frac{1}{|x-y|^{n+ps}}. \tag{3-1}$$

Using (3-1) and Tonelli's theorem for nonnegative integrands and we have

$$\begin{aligned}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{|u(x)-u(y)|^p}{|x-y|^{n+ps}}\,dx\,dy &= C\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}\int_0^\infty \alpha^{\frac{n+ps}{2}-1}\mathrm{e}^{-\alpha|x-y|^2}|u(x)-u(y)|^p d\alpha\,dx\,dy \\ &= C\int_0^\infty I_\alpha[u]\alpha^{\frac{n+ps}{2}-1}\,d\alpha\end{aligned}$$

with

$$I_\alpha[u] := \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |u(x)-u(y)|^p \mathrm{e}^{-\alpha|x-y|^2}\,dx\,dy \quad \text{and} \quad C = \frac{1}{\Gamma(\frac{n+ps}{2})}.$$

The function $J : \mathbb{R} \to \mathbb{R},\ x \mapsto |x|^p$ is strictly convex and nonnegative with $J(0) = 0$. The function $k : \mathbb{R}^n \to \mathbb{R},\ x \mapsto \mathrm{e}^{-|x|^2}$ is a strictly decreasing symmetric function and $k \in L_1(\mathbb{R}^n)$. Applying Lemma 2.6 to the functional I_α we obtain the desired result. □

Proof of Proposition 1.3. Since by hypothesis the sequence of sets $\{\Omega_n\}_{n=1}^\infty$ converges in the Hausdorff distance to Ω, then for any $\epsilon > 0$ there exist n_ϵ such that

$$B \setminus \Omega_n \subset (B \setminus \Omega) \oplus B(0,\epsilon) \quad \text{for all} \quad n \ge n_\epsilon \tag{3-2}$$

and

$$B \setminus \Omega \subset (B \setminus \Omega_n) \oplus B(0,\epsilon) \quad \text{for all} \quad n \ge n_\epsilon. \tag{3-3}$$

By Proposition 2.7, we have

$$\Omega \ominus B(0,\epsilon) \subseteq B \setminus ((B \setminus \Omega) \oplus B(0,\epsilon)). \tag{3-4}$$

It is clear that

$$\overline{\Omega} \ominus B(0,\epsilon) = \Omega \ominus B(0,\epsilon). \tag{3-5}$$

By using (3-2) (after taking the complement), (3-4), and (3-5) we obtain

$$\overline{\Omega} \ominus B(0,\epsilon) \subseteq B \setminus ((B \setminus \Omega) \oplus B(0,\epsilon)) \subset \Omega_n. \tag{3-6}$$

Using Lemma 2.8 and (3-6) we get

$$\Big(1 - 4\frac{R\epsilon}{r^2}\Big)\Omega \subset \Big(1 - 4\frac{R\epsilon}{r^2}\Big)\overline{\Omega} \subset \overline{\Omega} \ominus B(0,\epsilon) \subset \Omega_n. \tag{3-7}$$

Then applying parts (1) and (4) of Proposition 2.3 to (3-7) we obtain:

$$\Big(1 - 4\frac{R\epsilon}{r^2}\Big)^{sp} \lambda^s_{1,p}(\Omega_n) \le \lambda^s_{1,p}(\Omega). \tag{3-8}$$

Taking the upper limit in (3-8) gives:

$$\Big(1 - 4\frac{R\epsilon}{r^2}\Big)^{sp} \varlimsup_{n\to\infty} \lambda^s_{1,p}(\Omega_n) \le \lambda^s_{1,p}(\Omega). \tag{3-9}$$

Now, taking the limit as ϵ goes to 0 in (3-9) we get

$$\varlimsup_{n\to\infty} \lambda^s_{1,p}(\Omega_n) \le \lambda^s_{1,p}(\Omega). \tag{3-10}$$

Similarly, applying (3-4) and (3-5) in (3-3), and arguing as above, we can get

$$\lambda^s_{1,p}(\Omega) \le \varliminf_{n\to\infty} \lambda^s_{1,p}(\Omega_n). \tag{3-11}$$

The result follows immediately from (3-10) and (3-11). □

Proof of Theorem 1.1. Since $\lambda^s_{1,p}$ is translation and rotation invariant (see parts (2) and (3) of Proposition 2.3), to prove Theorem 1.1 for triangles, it is sufficient to find one equilateral triangle T' such that $\lambda^s_{1,p}(T') \le \lambda^s_{1,p}(T)$.

Let T_1 be an arbitrary triangle. We define recursively T_{n+1} to be the Steiner symmetrization of T_n with respect to the perpendicular bisector of one side (a side with respect to which there is no symmetry). Let u_n be a normalized function for the fractional Dirichlet p-Laplacian on T_n. Then, by Proposition 1.2 we have,

$$\lambda^s_{1,p}(T_n) = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{|u_n(x) - u_n(y)|^p}{|x-y|^{n+ps}}\,dx\,dy \ge \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{|u_n^\star(x) - u_n^\star(y)|^p}{|x-y|^{n+ps}}\,dx\,dy,$$

and by part (4) of Proposition 2.5 we have $\|u_n\|_p = \|u_n^\star\|_p = 1$. Therefore, using the definition on page 426, we obtain

$$\lambda_{1,p}^s(T_{n+1}) \le \lambda_{1,p}^s(T_n) \quad \text{for each} \quad n. \tag{3-12}$$

Now, recall the fact that the sequence of Steiner symmetrizations T_n of the arbitrary initial triangle T_1 converges to an equilateral triangle T with respect to the Hausdorff distance (see page 158 of [Pólya and Szegő 1951]). Then, using part (2) of Proposition 2.4, and if necessary a translation, we can show that there is a fixed ball contained in all the triangles T_n. Using part (3) of Proposition 2.4 we also conclude that all the triangles T_n are contained in a fixed ball. This allows us to apply Proposition 1.3, and we get

$$\lambda_{1,p}^s(T) = \lim_{n\to\infty} \lambda_{1,p}^s(T_n) \le \lambda_{1,p}^s(T_1)\,.$$

In the case of quadrilaterals, a similar argument can be used. In fact, a sequence of Steiner symmetrizations of a given quadrilateral, done alternatingly, with respect to the perpendicular bisector of a side and the diagonal, converges in the Hausdorff distance to a square (see page 158–159 of [Pólya and Szegő 1951]). This fact together with a reasoning as in the case of triangles leads to the Faber–Krahn inequality for quadrilaterals.

We now turn to the question of uniqueness. Suppose that T is any triangle for which the minimum is attained in the Faber–Krahn inequality. We can assume without loss of generality that T is not an equilateral triangle. Then T is not symmetric respect to the perpendicular bisector L to at least one side l. Let $T^\star$ the Steiner symmetrization of T respect to L. Let u be a normalized eigenfunction of $\lambda_{1,p}^s(T)$. Applying Proposition 1.2 and $\|u\|_p = \|u^\star\|_p = 1$, we get

$$\lambda_{1,p}^s(T^\star) \le \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{|u^\star(x)-u^\star(y)|^p}{|x-y|^{n+ps}}\,dx\,dy \le \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{|u(x)-u(y)|^p}{|x-y|^{n+ps}}\,dx\,dy = \lambda_{1,p}^s(T).$$

Since, $\lambda_{1,p}^s(T)$ is minimum, we obtain $\lambda_{1,p}^s(T^\star) = \lambda_{1,p}^s(T)$. This means that there is equality in the nonlocal Pólya–Szegő inequality and so, by the uniqueness part of Proposition 1.2, we get that u is a translate of $u^\star$. This is possible only if the triangles T and $T^\star$ are translates of each other. However, $T^\star$ is symmetric with respect to the L and T and $T^\star$ being translates of each other, T would have to be symmetric with respect to L. This gives a contradiction. So, the only minimizers are equilateral triangles.

The uniqueness in the case of quadrilaterals is completely analogous to case of the triangles. □

References

[Bhattacharya 1999] T. Bhattacharya, "A proof of the Faber–Krahn inequality for the first eigenvalue of the p-Laplacian", *Ann. Mat. Pura Appl.* (4) **177** (1999), 225–240. MR Zbl

[Brascamp et al. 1974] H. J. Brascamp, E. H. Lieb, and J. M. Luttinger, "A general rearrangement inequality for multiple integrals", *J. Functional Analysis* **17** (1974), 227–237. MR Zbl

[Brasco et al. 2014] L. Brasco, E. Lindgren, and E. Parini, "The fractional Cheeger problem", *Interfaces Free Bound.* **16**:3 (2014), 419–458. MR Zbl

[Chorwadwala et al. 2015] A. M. H. Chorwadwala, R. Mahadevan, and F. Toledo, "On the Faber–Krahn inequality for the Dirichlet p-Laplacian", *ESAIM Control Optim. Calc. Var.* **21**:1 (2015), 60–72. MR Zbl

[Di Nezza et al. 2012] E. Di Nezza, G. Palatucci, and E. Valdinoci, "Hitchhiker's guide to the fractional Sobolev spaces", *Bull. Sci. Math.* **136**:5 (2012), 521–573. MR Zbl

[Frank and Seiringer 2008] R. L. Frank and R. Seiringer, "Non-linear ground state representations and sharp Hardy inequalities", *J. Funct. Anal.* **255**:12 (2008), 3407–3430. MR

[Frank et al. 2008] R. L. Frank, E. H. Lieb, and R. Seiringer, "Hardy–Lieb–Thirring inequalities for fractional Schrödinger operators", *J. Amer. Math. Soc.* **21**:4 (2008), 925–950. MR Zbl

[Gruber 2007] P. M. Gruber, *Convex and discrete geometry*, Grundlehren der Mathematischen Wissenschaften **336**, Springer, Berlin, 2007. MR Zbl

[Lieb and Loss 2001] E. H. Lieb and M. Loss, *Analysis*, 2nd ed., Graduate Studies in Mathematics **14**, American Mathematical Society, Providence, RI, 2001. MR Zbl

[Lindgren and Lindqvist 2014] E. Lindgren and P. Lindqvist, "Fractional eigenvalues", *Calc. Var. Partial Differential Equations* **49**:1-2 (2014), 795–826. MR Zbl

[Ly 2005] I. Ly, "The first eigenvalue for the p-Laplacian operator", *JIPAM. J. Inequal. Pure Appl. Math.* **6**:3 (2005), art. id. 91, 12 pp. MR

[Osserman 1978] R. Osserman, "The isoperimetric inequality", *Bull. Amer. Math. Soc.* **84**:6 (1978), 1182–1238. MR Zbl

[Pólya and Szegő 1951] G. Pólya and G. Szegö, *Isoperimetric inequalities in mathematical physics*, Annals of Mathematics Studies **27**, Princeton University Press, 1951. MR Zbl

[Rayleigh 1894] J. W. S. Rayleigh, *The theory of sound*, vol. 1, 2nd ed., McMillan, London, 1894. Reprinted Dover, New York, 1945. MR

[Schneider 2014] R. Schneider, *Convex bodies: the Brunn–Minkowski theory*, 2nd ed., Encyclopedia of Mathematics and its Applications **151**, Cambridge University Press, 2014. MR Zbl

[Toledo Oñate 2012] F. J. Toledo Oñate, *Minimization of the first eigenvalue of the Dirichlet p-Laplacian in certain classes of domains*, Master's thesis, Universidad de Concepción, 2012, available at http://dmat.cfm.cl/colloquium/2012-12-14-13:30.pdf.

Received April 11, 2016. Revised August 11, 2016.

FRANCO OLIVARES CONTADOR
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CONCEPCIÓN
4030000 CONCEPCIÓN
CHILE

folivaresc@udec.cl

PACIFIC JOURNAL OF MATHEMATICS
Vol. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.435

TOPOLOGICAL INVARIANCE OF QUANTUM QUATERNION SPHERES

BIPUL SAURABH

The C^*-algebra of continuous functions on the quantum quaternion sphere H_q^{2n} can be identified with the quotient algebra $C(\mathrm{SP}_q(2n)/\mathrm{SP}_q(2n-2))$. In the commutative case, i.e., for $q=1$, the topological space $\mathrm{SP}(2n)/\mathrm{SP}(2n-2)$ is homeomorphic to the odd-dimensional sphere $\mathcal{S}^{4n-1}$. In this paper, we prove the noncommutative analogue of this result. Using homogeneous C^*-extension theory, we prove that the C^*-algebra $C(H_q^{2n})$ is isomorphic to the C^*-algebra $C(S_q^{4n-1})$. This further implies that for different values of q in $[0,1)$, the C^*-algebras underlying the noncommutative spaces H_q^{2n} are isomorphic.

1. Introduction

Quantization of Lie groups and their homogeneous spaces has played an important role in linking the theory of compact quantum groups with noncommutative geometry. Many authors (see [Vaksman and Soibelman 1990; Podkolzin and Vainerman 1999; Chakraborty and Pal 2008; Pal and Sundar 2010]) have studied different aspects of the theory of quantum homogeneous spaces. However, in these papers, the main examples have been the quotient spaces of the compact quantum group $\mathrm{SU}_q(n)$. Neshveyev and Tuset [2012] studied quantum homogeneous spaces in a more general setup and gave a complete classification of the irreducible representations of the C^*-algebra $C(G_q/H_q)$ where G_q is the q-deformation of a simply connected semisimple compact Lie group and H_q is the q-deformation of a closed Poisson–Lie subgroup H of G. Moreover, Neshveyev and Tuset [2012] proved that $C(G_q/H_q)$ is KK-equivalent to the classical counterpart $C(G/H)$. In [Saurabh 2017], we studied the quantum symplectic group $\mathrm{SP}_q(2n)$ and its homogeneous space $\mathrm{SP}_q(2n)/\,\mathrm{SP}_q(2n-2)$, and obtained K-groups of $C(\mathrm{SP}_q(2n)/\,\mathrm{SP}_q(2n-2))$ with explicit generators.

The C^*-algebra $C(H_q^{2n})$ of continuous functions on the quantum quaternion sphere is defined as the universal C^*-algebra given by a finite set of generators and relations; see [Saurabh 2017]. In the same paper, the isomorphism between

MSC2010: 19K33, 46L80, 58B34.

Keywords: homogeneous extension, quantum double suspension, corona factorization property.

the quotient algebra $C(\mathrm{SP}_q(2n)/\mathrm{SP}_q(2n-2))$ and $C(H_q^{2n})$ was established. Now several questions arise about this noncommutative space H_q^{2n}:

(1) Topologically, is H_q^{2n} the same as S_q^{4n-1}, i.e., are the C^*-algebras $C(H_q^{2n})$ and $C(S_q^{4n-1})$ isomorphic?

(2) Are the C^*-algebras $C(H_q^{2n})$ isomorphic for different values of q?

(3) Does the quantum quaternion sphere admit a good spectral triple equivariant under the $\mathrm{SP}_q(2n)$-group action?

We attempt the first two questions in this paper. In the commutative case, that is, for $q=1$, the quotient space $\mathrm{SP}(2n)/\mathrm{SP}(2n-2)$ can be realized as the quaternion sphere H^{2n}. It can be easily verified that the quaternion sphere H^{2n} is homeomorphic to the odd-dimensional sphere $\mathcal{S}^{4n-1}$. One can now expect the quotient algebra $C(\mathrm{SP}_q(2n)/\mathrm{SP}_q(2n-2))$, or equivalently, the C^*-algebra $C(H_q^{2n})$, to be isomorphic to the C^*-algebra underlying the odd-dimensional quantum sphere S_q^{4n-1}. Using homogeneous C^*-extension theory, we show that this is indeed the case.

The remarkable work done by L. G. Brown, R. G. Douglas and P. A. Fillmore [Brown et al. 1977] on extensions of commutative C^*-algebras by compact operators has led many authors to extend this theory further in order to provide a tool for analyzing the structure of C^*-algebras. For a nuclear, separable C^*-algebra A and a separable C^*-algebra B, G. G. Kasparov [1979] constructed the group $\mathrm{Ext}(A, B)$ consisting of stable equivalence classes of C^*-algebra extensions of the form

$$0 \to B \otimes \mathcal{K} \to E \to A \to 0.$$

Here E will be called the middle C^*-algebra. One of the important features of this construction is that the group $\mathrm{Ext}(A, B)$ coincides with the group $KK^1(A, B)$. Another important aspect is that it does not demand much. It does not require the extensions to be unital or essential. But at the same time, it does not provide much information about the middle C^*-algebras. Since elements of the group $\mathrm{Ext}(A, B)$ are stable equivalence classes and not strongly unitary equivalence classes of extensions, two elements in the same class may have nonisomorphic middle C^*-algebras. For a nuclear C^*-algebra A and a finite-dimensional compact metric space Y (i.e., a closed subset of $\mathcal{S}^n$ for some $n \in \mathbb{N}$), M. Pimsner, S. Popa and D. Voiculescu [Pimsner et al. 1979] constructed another group $\mathrm{Ext}_{\mathrm{PPV}}(Y, A)$ consisting of strongly unitary equivalence classes of unital homogeneous extensions of A by $C(Y) \otimes \mathcal{K}$. For $y_0 \in Y$, the subgroup $\mathrm{Ext}_{\mathrm{PPV}}(Y, y_0, A)$ consists of those elements of $\mathrm{Ext}_{\mathrm{PPV}}(Y, A)$ that split at y_0. For a commutative C^*-algebra A, the group $\mathrm{Ext}_{\mathrm{PPV}}(Y, A)$ was computed by Schochet [1980]. Further, Rosenberg and Schochet [1981] showed that

$$\mathrm{Ext}_{\mathrm{PPV}}(Y, A^+) = \mathrm{Ext}(A, C(Y)) \quad \text{and} \quad \mathrm{Ext}_{\mathrm{PPV}}(Y^+, +, A^+) = \mathrm{Ext}(A, C(Y)),$$

where Y is a finite-dimensional locally compact Hausdorff space, $+$ is the point at infinity and A^+ is the C^*-algebra obtained by adjoining unity to A.

To prove the claim, our idea is to exhibit two short exact sequences of C^*-algebras in the same equivalence class in the group $\mathrm{Ext}_{\mathrm{PPV}}(\mathbb{T}, C(S_0^{4n-3}))$ with $C(S_q^{4n-1})$ and $C(H_q^{2n})$ as middle C^*-algebras and then to compare their middle C^*-algebras. First, we prove an isomorphism between groups $\mathrm{Ext}_{\mathrm{PPV}}(Y, y_0, A)$ and $\mathrm{Ext}_{\mathrm{PPV}}(Y, y_0, \Sigma^2 A)$ under certain assumptions on the topological space Y where $\Sigma^2 A$ is the quantum double suspension of A and $y_0 \in Y$. Using this, we describe all elements of the group $\mathrm{Ext}_{\mathrm{PPV}}(\mathbb{T}, C(S_0^{2\ell+1}))$ explicitly. We compute K-groups of all middle C^*-algebras that occur in all the extensions of the group $\mathrm{Ext}_{\mathrm{PPV}}(\mathbb{T}, C(S_0^{2\ell+1}))$. Then using the ideal structure of $C(H_q^{2n})$, we show that the extension

$$0 \to C(\mathbb{T}) \otimes \mathcal{K} \to C(H_q^{2n}) \to C(S_0^{4n-3}) \to 0$$

is unital and homogeneous. Now by comparing the K-groups of middle C^*-algebras, we prove that the above extension is strongly unitarily equivalent to either the extension

$$0 \to C(\mathbb{T}) \otimes \mathcal{K} \to C(S_0^{4n-1}) \to C(S_0^{4n-3}) \to 0,$$

or its inverse in the group $\mathrm{Ext}_{\mathrm{PPV}}(\mathbb{T}, C(S_0^{2\ell+1}))$, having $C(S_0^{4n-1})$ as a middle C^*-algebra. This proves that the C^*-algebras $C(H_q^{2n})$ and $C(S_0^{4n-1})$ are isomorphic; see [Blackadar 1998, page 147]. For $q = 0$, it follows immediately as the defining relations of $C(H_0^{2n})$ (see [Saurabh 2017]) are exactly the same as those of $C(S_0^{4n-1})$. In [Hong and Szymański 2002], it was proved that for different values of q in $[0, 1)$, the C^*-algebras $C(S_q^{4n-1})$ are isomorphic. As a consequence, the C^*-algebras $C(H_q^{2n})$ and $C(S_q^{4n-1})$ are isomorphic for all q in $[0, 1)$. Also, this establishes the q-invariance of the quantum quaternion spheres, as it shows that the C^*-algebras $C(H_q^{2n})$ are isomorphic for different values of q. Here we must point out that to the best of our knowledge, the group $\mathrm{Ext}_{\mathrm{PPV}}(Y, A)$ has not been used before to show that two C^*-algebras are isomorphic. In that sense, our idea can be considered as the first of its kind.

We now set up some notation. The standard basis of the Hilbert space $L_2(\mathbb{N})$ will be denoted by $\{e_n : n \in \mathbb{N}\}$. We denote the left shift operator on $L_2(\mathbb{N})$ and $L_2(\mathbb{Z})$ by the same notation S. For $m < 0$, $(S^*)^m$ denotes the operator S^{-m}. Let p_i be the rank-1 projection sending e_i to e_i. The operator p_0 will be denoted by p. We write $\mathcal{L}(\mathcal{H})$ and $\mathcal{K}(\mathcal{H})$ for the sets of all bounded linear operators and compact operators on $\mathcal{H}$, respectively. We denote by $\mathcal{K}$ the C^*-algebra of compact operators. For a C^*-algebra A, $\Sigma^2 A$ and $M(A)$ are used to denote the quantum double suspension (see [Hong and Szymański 2002; 2008]) of A and multiplier algebra of A, respectively. The map π will denote the canonical homomorphism from $M(A)$ to $Q(A) := M(A)/A$ and for $a \in M(A)$, $[a]$ stands for the image of a

under the map π. For a locally compact Hausdorff space Y, we write Y^+ to denote one-point compactification of Y. For a C^*-algebra A, A^+ is the C^*-algebra obtained by adjoining unity to A. The symbol $\mathcal{S}^n$ will be reserved for the n-dimensional sphere. However, sometimes we will use $\mathbb{T}$ in place of $\mathcal{S}^1$ to denote the circle. Unless otherwise stated, q will denote a real number in the interval (0, 1).

2. C^*-algebra extensions

In this section, we briefly recall some notions related to C^*-extension theory. For a detailed treatment, we refer the reader to [Blackadar 1998]. Let A be a unital separable nuclear C^*-algebra. Let B be a stable C^*-algebra. An extension of A by B is a short exact sequence

$$0 \to B \xrightarrow{i} E \xrightarrow{j} A \to 0.$$

In such cases there exists a unique homomorphism $\sigma : E \to M(B)$ such that $\sigma(i(b)) = b$ for all $b \in B$. We can now define the Busby invariant for the extension $0 \to B \xrightarrow{i} E \xrightarrow{j} A \to 0$ by the homomorphism $\tau : A \to M(B)/B$ given by $\tau(a) = \pi \circ \sigma(e)$, where e is a preimage of a and π is the quotient map from $M(B)$ to $M(B)/B$. It is easy to see that τ is well-defined. Up to strong isomorphism, an extension can be identified with its Busby invariant. In this paper, we will not distinguish between an extension and its Busby invariant, as all the equivalence relations given here are weaker than the strong isomorphism relation.

An extension $\tau : A \to M(B)/B$ is called essential if τ is injective or, equivalently, the image of B is an essential ideal of E. We call an extension unital if it is a unital homomorphism or, equivalently, E is a unital C^*-algebra. An extension τ is called a trivial (or split) extension if there exists a homomorphism $\lambda : A \to M(B)$ such that $\tau = \pi \circ \lambda$. Extensions τ_1 and τ_2 are said to be unitarily equivalent if there exists a unitary u in $Q(B)$ such that $u\tau_1(a)u^* = \tau_2(a)$ for all $a \in A$. The two extensions are said to be strongly unitarily equivalent if there exists a unitary U in $M(B)$ such that $\pi(U)\tau_1(a)\pi(U^*) = \tau_2(a)$ for all $a \in A$. We denote a strongly unitary equivalence relation by $\sim_{su}$. Let $\mathrm{Ext}_{\sim_{su}}(A, B)$ denote the set of strongly unitary equivalence classes of extensions of A by B. One can put a binary operation $+$ on $\mathrm{Ext}_{\sim_{su}}(A, B)$ as follows. Since $M(B)$ is a stable C^*-algebra, we can get two isometries ν_1 and ν_2 in $M(B)$ such that $\nu_1\nu_1^* + \nu_2\nu_2^* = 1$. Let $[\tau_1]_{su}$ and $[\tau_2]_{su}$ be two elements in $\mathrm{Ext}_{\sim_{su}}(A, B)$. Define the extension $\tau_1 + \tau_2 : A \to Q(B)$ by $(\tau_1 + \tau_2)(a) := \pi(\nu_1)\tau_1(a)\pi(\nu_1^*) + \pi(\nu_2)\tau_2(a)\pi(\nu_2^*)$. The binary operation $+$ on $\mathrm{Ext}_{\sim_{su}}(A, B)$ can now be defined as

$$[\tau_1]_{su} + [\tau_2]_{su} := [\tau_1 + \tau_2]_{su}. \tag{2-1}$$

This makes $\mathrm{Ext}_{\sim_{su}}(A, B)$ a commutative semigroup. Moreover, the set of trivial extensions forms a subsemigroup of $\mathrm{Ext}_{\sim_{su}}(A, B)$. We denote the quotient of

$\mathrm{Ext}_{\sim_{su}}(A, B)$ with the set of trivial extensions by $\mathrm{Ext}(A, B)$. For a separable nuclear C^*-algebra A, the set $\mathrm{Ext}(A, B)$ under the operation $+$ is a group; see [Blackadar 1998]. Two extensions τ_1 and τ_2 represent the same element in $\mathrm{Ext}(A, B)$ if there exist two trivial extensions ϕ_1 and ϕ_2 such that $\tau_1 + \phi_1 \sim_{su} \tau_2 + \phi_2$. We denote an equivalent class in the group $\mathrm{Ext}(A, B)$ of an extension τ by $[\tau]_s$. One can show that for a stable C^*-algebra B, $\mathrm{Ext}(A, B) = \mathrm{Ext}(A, B \otimes \mathcal{K})$. Now for an arbitrary C^*-algebra B, define $\mathrm{Ext}(A, B) := \mathrm{Ext}(A, B \otimes \mathcal{K})$. For $B = \mathbb{C}$, we denote the group $\mathrm{Ext}(A, \mathbb{C})$ by $\mathrm{Ext}(A)$. Note that in this case, two unital essential extensions τ_1 and τ_2 are in the same equivalence class (i.e., $[\tau_1]_s = [\tau_2]_s$) if and only if they are strongly unitarily equivalent. Suppose that Y is a finite-dimensional compact metric space, i.e., a closed subset of $\mathcal{S}^n$ for some $n \in \mathbb{N}$. Let $M(Y)$, $Q(Y)$ and Q be the C^*-algebras $M(C(Y) \otimes \mathcal{K})$, $M(C(Y) \otimes \mathcal{K})/C(Y) \otimes \mathcal{K}$ and $\mathcal{L}(\mathcal{H})/\mathcal{K}(\mathcal{H})$ (Calkin algebra) respectively. One can easily see that $M(Y)$ is the set of all $*$-strong continuous functions from Y to $\mathcal{L}(\mathcal{H})$. We call an extension τ of A by $C(Y) \otimes \mathcal{K}$ homogeneous if for all $y \in Y$, the map $ev_y \circ \tau : A \to Q$ is injective where $ev_y : Q(Y) \to Q$ is the evaluation map at y. Let $\mathrm{Ext}_{\mathrm{PPV}}(Y, A)$ be the set of strongly unitary equivalence classes of unital homogeneous extensions of A by $C(Y) \otimes \mathcal{K}$. For a nuclear C^*-algebra A, Pimsner, Popa and Voiculescu [Pimsner et al. 1979] showed that $\mathrm{Ext}_{\mathrm{PPV}}(Y, A)$ is a group with the additive operation defined as in (2-1). We denote the equivalence class in the group $\mathrm{Ext}_{\mathrm{PPV}}(Y, A)$ of an extension τ by $[\tau]_{su}$. For $y_0 \in Y$, define the set

$$\mathrm{Ext}_{\mathrm{PPV}}(Y, y_0, A) = \big\{[\tau]_{su} \in \mathrm{Ext}_{\mathrm{PPV}}(Y, A) : ev_{y_0} \circ \tau \text{ is split}\big\}.$$

The set $\mathrm{Ext}_{\mathrm{PPV}}(Y, y_0, A)$ is a subgroup of $\mathrm{Ext}_{\mathrm{PPV}}(Y, A)$.

***The groups* $\mathbf{Ext_{PPV}(Y, A)}$ *and* $\mathbf{Ext_{PPV}(Y, \Sigma^2 A)}$.** Here we will show that for a separable nuclear C^*-algebra A and a finite-dimensional compact metric space Y such that K-groups of $C(Y)$ are finitely generated, the groups $\mathrm{Ext}_{\mathrm{PPV}}(Y, A)$ and $\mathrm{Ext}_{\mathrm{PPV}}(Y, \Sigma^2 A)$ are isomorphic. Let us recall some definitions. We say that two elements a and b in $Q(B)$ are strongly unitarily equivalent if there exists a unitary $U \in M(B)$ such that $[U]a[U^*] = b$. Two elements a and b in $Q(B)$ are said to be unitarily equivalent if there exists unitary $u \in Q(B)$ such that $uau^* = b$. We call an element a in a C^*-algebra B norm-full if it is not contained in any proper closed ideal in B. Suppose that A and B are separable C^*-algebras. An extension $\tau : A \to Q(B \otimes \mathcal{K})$ is said to be norm-full if for every nonzero element $a \in A$, $\tau(a)$ is norm-full element of $Q(B \otimes \mathcal{K})$.

Definition 2.1 [Lin 2009]. Let B be a separable σ-unital C^*-algebra. We say $Q(B \otimes \mathcal{K})$ has property (P) if for any norm-full element $b \in Q(B \otimes \mathcal{K})$, there exist $x, y \in Q(B \otimes \mathcal{K})$ such that $xby = 1$.

Definition 2.2 [Kucerovsky and Ng 2006]. Let B be a separable C^*-algebra. Then $B \otimes \mathcal{K}$ is said to have the corona factorization property if every norm-full projection in $M(B \otimes \mathcal{K})$ is Murray–von Neumann equivalent to the unit element of $M(B \otimes \mathcal{K})$.

One can show that in a C^*-algebra $B \otimes \mathcal{K}$ with the corona factorization property, any norm-full projection in $Q(B \otimes \mathcal{K})$ is Murray–von Neumann equivalent to the unit element of $Q(B \otimes \mathcal{K})$; see [Kucerovsky and Ng 2006]. Also note that the fact that $Q(B \otimes \mathcal{K})$ has property (P) implies that $B \otimes \mathcal{K}$ has the corona factorization property. It is proved in [Lin 2007] that for a finite-dimensional compact metric space Y, $Q(C(Y) \otimes \mathcal{K})$ has property (P) and hence $C(Y) \otimes \mathcal{K}$ has the corona factorization property. We will see that these properties play important roles in proving the isomorphism between the groups $\mathrm{Ext}_{\mathrm{PPV}}(Y, A)$ and $\mathrm{Ext}_{\mathrm{PPV}}(Y, \Sigma^2 A)$. But for that, we need the following proposition that says that for a C^*-algebra with certain properties, the group $\mathrm{Ext}_{\mathrm{PPV}}(Y, A)$ can be viewed as a subgroup of the group $KK^1(A, C(Y))$.

Proposition 2.3. *Let A be a unital separable nuclear C^*-algebra which satisfies the universal coefficient theorem. Assume that $K_0(A) = G \oplus \mathbb{Z}$ with $[1_A] = (0, 1)$. Suppose that Y is a finite-dimensional compact metric space. Then the map*

$$i : \mathrm{Ext}_{\mathrm{PPV}}(Y, A) \to KK^1(A, C(Y)), \qquad [\tau]_{su} \mapsto [\tau]_s$$

is an injective homomorphism.

Proof. Since strongly unitary equivalence implies stable equivalence, the map i is well-defined. Any unital homogeneous extension is a purely large extension and hence a norm-full extension; see [Elliott and Kucerovsky 2001, page 19]. Therefore, from [Lin 2009, Theorem 2.4 and Corollary 3.9], it follows that i is injective. $\square$

From now on, without loss of generality, we will assume that the Hilbert space $\mathcal{H}$ is $L_2(\mathbb{N})$. Let τ be a unital homogeneous extension of A by $C(Y) \otimes \mathcal{K}(\mathcal{H})$. Define $\tilde{\tau} : A \to Q\big(C(Y) \otimes \mathcal{K}(\mathcal{H}) \otimes \mathcal{K}(\mathcal{H})\big)$ by $\tilde{\tau}(a) = [\tau_a \otimes p]$ where $[\tau_a] = \tau(a)$. By the universal property of quantum double suspension (see [Hong and Szymański 2008, Proposition 2.2]), there exists a unique homomorphism

$$\Sigma^2\tau : \Sigma^2 A \to Q\big(C(Y) \otimes \mathcal{K}(\mathcal{H}) \otimes \mathcal{K}(\mathcal{H})\big) \tag{2-2}$$

such that $\Sigma^2\tau(a \otimes p) = \tilde{\tau}(a) = [\tau_a \otimes p]$ and $\Sigma^2\tau(1 \otimes S) = [1 \otimes 1 \otimes S]$. Clearly $\Sigma^2\tau$ is a unital extension. Since τ is homogeneous, the map $ev_y \circ \tau$ is injective for all $y \in Y$. Therefore the map $ev_y \circ \Sigma^2\tau$ is injective on the C^*-algebra $A \otimes p$ as $ev_y \circ \Sigma^2\tau(a \otimes p) = [(ev_y \circ \tau)_a \otimes p]$ where $[(ev_y \circ \tau)_a] = ev_y \circ \tau(a)$. Making use of the fact that $(1 \otimes p)A \otimes \mathcal{K}(1 \otimes p) = A \otimes p$, one can prove that the map $ev_y \circ \Sigma^2\tau$ is injective on $A \otimes \mathcal{K}$. Since $A \otimes \mathcal{K}$ is an essential ideal of $\Sigma^2 A$, we conclude that the map $ev_y \circ \Sigma^2\tau$ is injective on $\Sigma^2 A$ and hence $\Sigma^2\tau$ is a homogeneous extension. Moreover, if τ_1 and τ_2 are strongly unitarily equivalent by a unitary $U \in M\big(C(Y) \otimes \mathcal{K}(\mathcal{H})\big)$ then so are $\Sigma^2\tau_1$ and $\Sigma^2\tau_2$ by the unitary $U \otimes 1 \in M(C(Y) \otimes \mathcal{K}(\mathcal{H}) \otimes \mathcal{K}(\mathcal{H}))$. This gives a well-defined map:

$$\beta : \mathrm{Ext}_{\mathrm{PPV}}(Y, A) \to \mathrm{Ext}_{\mathrm{PPV}}(Y, \Sigma^2 A), \qquad [\tau]_{su} \mapsto [\Sigma^2\tau]_{su}. \tag{2-3}$$

Proposition 2.4. *The map* $\beta : \mathrm{Ext}_{\mathrm{PPV}}(Y, A) \to \mathrm{Ext}_{\mathrm{PPV}}(Y, \Sigma^2 A)$ *given above is an injective group homomorphism.*

Proof. Let τ be a unital homogeneous extension of A by $C(Y) \otimes \mathcal{K}$ such that $\Sigma^2\tau$ is a split extension. In this case, there exists a homomorphism $\lambda : \Sigma^2 A \to M(Y)$ such that $\pi \circ \lambda = \Sigma^2\tau$. Define $\alpha : A \to M(Y)$ by $\alpha(a) := \lambda(a \otimes p)$ for $a \in A$. It is easy to check that $\pi \circ \alpha = \tau$ which implies that τ is a split extension. This proves that the map β is injective. □

To get surjectivity of the map β, we need to put certain assumptions on the topological space Y.

Proposition 2.5. *Let Y be a finite-dimensional compact metric space. Assume that $K_0(C(Y))$ and $K_1(C(Y))$ are finitely generated abelian groups. Then, letting $V \in Q\big(C(Y) \otimes \mathcal{K}(\mathcal{H}) \otimes \mathcal{K}(\mathcal{H})\big)$ be an isometry such that VV^* and $1 - VV^*$ both are norm-full projections, V is unitarily equivalent to* $[1 \otimes 1 \otimes S^*]$.

Proof. Let $G_n := \mathrm{Ext}_{\mathrm{PPV}}(Y, \Sigma^{2n} C(\mathbb{T}))$. Since

$$KK^1(\Sigma^{2n} C(\mathbb{T}), C(Y)) \equiv K_0(C(Y)) \oplus K_1(C(Y)),$$

one can consider the groups G_n as subgroups of $K_0(C(Y)) \oplus K_1(C(Y))$ thanks to Proposition 2.3. This implies that the G_n are finitely generated abelian groups. For $n \in \mathbb{N}$, define the map

$$\beta_n : \mathrm{Ext}_{\mathrm{PPV}}(Y, \Sigma^{2n} C(\mathbb{T})) \to \mathrm{Ext}_{\mathrm{PPV}}(Y, \Sigma^{2n+2} C(\mathbb{T})), \qquad [\tau]_{su} \mapsto [\Sigma^2\tau]_{su}, \tag{2-4}$$

where $\Sigma^2\tau$ is as in (2-2). From Proposition 2.4, it follows that the maps β_n are injective homomorphisms. Assume that V is not unitarily equivalent to $[1 \otimes 1 \otimes S^*]$. For each $n \in \mathbb{N}$, the isometry V will induce an isometry $V_n \in Q(C(Y) \otimes \mathcal{K}(\mathcal{H})^{\otimes n+1})$ (where $^{\otimes k}$ means the tensor product of k copies) such that $V_n V_n^*$ and $1 - V_n V_n^*$ both are norm-full projections and V_n is not unitarily equivalent to $[1^{\otimes n+1} \otimes S^*]$. Since $C(Y) \otimes \mathcal{K}$ has the corona factorization property, it follows that $V_n V_n^*$ and $1 - V_n V_n^*$ both are Murray–von Neumann equivalent to $[1]$. Also, one can easily verify that $[1^{\otimes n+1} \otimes p]$ and $[1 - 1^{\otimes n+1} \otimes p] = [1^{\otimes n+1} \otimes (1 - p)]$ are Murray–von Neumann equivalent to $[1]$. This implies that $V_n V_n^*$ is unitarily equivalent to $1 - [1^{\otimes n+1} \otimes p]$. So, without loss of generality, we can assume that V_n has final projection $1 - [1^{\otimes n+1} \otimes p]$. Take a split unital homogeneous extension τ of $C(\mathbb{T})$ by $C(Y) \otimes \mathcal{K}(\mathcal{H})$. Clearly $\Sigma^{2n}\tau$ is a split unital homogeneous extension of

$\Sigma^{2n}C(\mathbb{T})$ by $C(Y)\otimes\mathcal{K}(\mathcal{H})^{\otimes n+1}$. Let $\Sigma_V^2(\Sigma^{2n}\tau)$ be a unital homogeneous extension of $\Sigma^{2n+2}C(\mathbb{T})$ by $C(Y)\otimes\mathcal{K}(\mathcal{H})^{\otimes n+2}$ given by

$$\Sigma_V^2(\Sigma^{2n}\tau)(a\otimes p)=[\Sigma^{2n}\tau_a\otimes p] \quad \text{and} \quad \Sigma_V^2(\Sigma^{2n}\tau)(1\otimes S^*)=V_{n+1},$$

where $[\Sigma^{2n}\tau_a]=\Sigma^{2n}\tau(a)$. From [Lin 2009, Corollary 3.8] and the fact that V_{n+1} is not unitarily equivalent to $[1^{\otimes n+2}\otimes S^*]$, it follows that $[\Sigma_V^2(\Sigma^{2n}\tau)]_{su}$ is not in the image of the map β_n defined as in (2-4). Let $m[\Sigma_V^2(\Sigma^{2n}\tau)]_{su}=\beta_n([\phi]_{su})$ for some $m\in\mathbb{Z}-\{0\}$ and for some unital homogeneous extension ϕ of $\Sigma^{2n}C(\mathbb{T})$ by $C(Y)\otimes\mathcal{K}$. It is easy to see that ϕ must be split and in that case $m[\Sigma_V^2(\Sigma^{2n}\tau)]$ is the class of split extensions. This shows that for all $n\in\mathbb{N}$, the group G_{n+1} has either one more free generator or one more element of finite order than the group G_n. Since $K_0(C(Y))\oplus K_1(C(Y))$ is a finitely generated group for all $n\in\mathbb{N}$, $G_n\subset K_0(C(Y))\oplus K_1(C(Y))$, and we reach a contradiction. This proves that V is unitarily equivalent to $[1\otimes 1\otimes S^*]$. □

Remark 2.6. Here we should point out that the above proposition may hold for a more general finite-dimensional compact metric space Y. But since we could not find any general result along this direction in literature, we prove the proposition under certain assumptions on Y.

Corollary 2.7. *Let Y and V be as in the above proposition. Then V is strongly unitarily equivalent to* $[1\otimes 1\otimes S^*]$.

Proof. Consider the unital extension $\Sigma_V^2\tau$ constructed in Proposition 2.5 where τ is a split unital homogeneous extension of $C(\mathbb{T})$ by $C(Y)\otimes\mathcal{K}(\mathcal{H})$. Using Proposition 2.5, one can show that $\Sigma_V^2\tau$ is unitarily equivalent to $\Sigma^2\tau$ defined in (2-2) with $A=C(\mathbb{T})$. Therefore, by [Lin 2009, Corollary 3.10], it follows that $\Sigma_V^2\tau$ is strongly unitarily equivalent to $\Sigma^2\tau$. Hence V is strongly unitarily equivalent to $[1\otimes 1\otimes S^*]$. □

Lemma 2.8 establishes the isomorphism between the groups $\mathrm{Ext}_{\mathrm{PPV}}(Y, A)$ and $\mathrm{Ext}_{\mathrm{PPV}}(Y, \Sigma^2 A)$ under certain assumptions on the space Y.

Lemma 2.8. *Let Y be a finite-dimensional compact metric space. Assume that the groups $K_0(C(Y))$ and $K_1(C(Y))$ are finitely generated abelian groups. Then the map $\beta:\mathrm{Ext}_{\mathrm{PPV}}(Y, A)\to\mathrm{Ext}_{\mathrm{PPV}}(Y, \Sigma^2 A)$ given above is an isomorphism.*

Proof. We only need to show that β is surjective thanks to Proposition 2.4. Let ϕ be a unital homogeneous extension of $\Sigma^2 A$ by $C(Y)\otimes\mathcal{K}(\mathcal{H})\otimes\mathcal{K}(\mathcal{H})$. Let $\phi(1\otimes S^*)=V$. Since ϕ is a unital homogeneous extension and hence a norm-full extension, it follows that VV^* and $1-VV^*$ are norm-full projections. Therefore, by Corollary 2.7, there exists a unitary $U\in M(C(Y)\otimes\mathcal{K}(\mathcal{H})\otimes\mathcal{K}(\mathcal{H}))$ such that $[U]V[U^*]=[1\otimes 1\otimes S^*]$. So without loss of generality, we can assume that ϕ maps $1\otimes S^*$ to $[1\otimes 1\otimes S^*]$. This implies that $\phi(1\otimes p)=[1\otimes 1\otimes p]$. But then $\phi(A\otimes p)\subset(1\otimes 1\otimes p)\phi(A\otimes p)(1\otimes 1\otimes p)\subset Q(C(Y)\otimes\mathcal{K}(\mathcal{H}))\otimes p$ which induces a map $\tau : A \to Q(C(Y) \otimes \mathcal{K}(\mathcal{H}))$ by omitting the projection p. Therefore, we get a unital homogeneous extension τ of A such that $\beta([\tau]_u) = [\phi]_u$. This proves that the map β is surjective. □

Corollary 2.9. *For $y_0 \in Y$, the map*

$$\beta_{|\,\mathrm{Ext}_{\mathrm{PPV}}(Y, y_0, A)} : \mathrm{Ext}_{\mathrm{PPV}}(Y, y_0, A) \to \mathrm{Ext}_{\mathrm{PPV}}(Y, y_0, \Sigma^2 A)$$

is an isomorphism.

Proof. It is easy to check that if $ev_{y_0} \circ \tau$ is split then so is $ev_{y_0} \circ \Sigma^2 \tau$ and vice versa. Now the claim will follow from Lemma 2.8. □

3. Elements of $\mathrm{Ext}_{\mathrm{PPV}}\big(\mathbb{T}, C\big(S_0^{2\ell+1}\big)\big)$

In this section, we will write down all elements of the groups $\mathrm{Ext}(C(S_0^{2\ell+1}))$ and $\mathrm{Ext}_{\mathrm{PPV}}(\mathbb{T}, C(S_0^{2\ell+1}))$ explicitly in terms of their Busby invariants. We start with the definition of $C(S_0^{2\ell+1})$. The C^*-algebra $C(S_0^{2\ell+1})$ is defined as the C^*-subalgebra of $\mathcal{L}\big(L_2(\mathbb{N})^{\otimes \ell}\big) \otimes C(\mathbb{T})$ generated by the following operators:

$$\begin{gathered}
S^* \otimes 1 \otimes \cdots \otimes 1, \\
p \otimes S^* \otimes 1 \otimes \cdots \otimes 1, \\
\vdots \\
p \otimes p \otimes \cdots \otimes p \otimes S^* \otimes 1, \\
p \otimes p \otimes \cdots \otimes p \otimes p \otimes t.
\end{gathered}$$

Hong and Szymański [2002] showed that for $q \in (0, 1)$, the C^*-algebra $C(S_q^{2\ell+1})$ of continuous functions on the odd-dimensional quantum sphere $S_q^{2\ell+1}$ is isomorphic to the C^*-algebra $C(S_0^{2\ell+1})$. Since for calculation purposes, the generators of $C(S_0^{2\ell+1})$ given above are easier to deal with in comparison to those of $C(S_q^{2\ell+1})$, we will, without loss of generality, take the C^*-algebra $C(S_0^{2\ell+1})$. Define the $*$-homomorphisms φ_m as follows:

$$\begin{aligned}
\varphi_m : C(S_0^{2\ell+1}) &\to Q\big(\mathcal{K}\big(L_2(\mathbb{N})^{\otimes \ell+1}\big)\big), \\
S^* \otimes 1 \otimes \cdots \otimes 1 &\mapsto S^* \otimes 1 \otimes \cdots \otimes 1, \\
p \otimes S^* \otimes 1 \otimes \cdots \otimes 1 &\mapsto p \otimes S^* \otimes 1 \otimes \cdots \otimes 1, \\
&\ \,\vdots \\
p \otimes p \otimes \cdots \otimes p \otimes S^* \otimes 1 &\mapsto p \otimes p \otimes \cdots \otimes S^* \otimes 1, \\
p \otimes p \otimes \cdots \otimes t &\mapsto p \otimes p \otimes \cdots \otimes p \otimes (S^*)^m.
\end{aligned}$$

The following proposition says that for each $m \in \mathbb{Z}$, the homomorphism $[\varphi_m]_s$ is an element of the group $\mathrm{Ext}(C(S_0^{2\ell+1}))$.

Proposition 3.1. *For each $m \in \mathbb{Z}$, the extension φ_m is an essential unital extension of $C(S_0^{2\ell+1})$ by compact operators.*

Proof. Clearly the φ_m are unital extensions of $C(S_0^{2\ell+1})$ by compact operators. We need to show that the φ_m are injective homomorphisms. Let $C_t(\mathbb{T})$ be the set of all continuous functions on $\mathbb{T}$ vanishing at t. Using irreducible representations of $C(S_0^{2\ell+1})$, it is easy to see that

(1) $\{\mathcal{K}(L_2(\mathbb{N}))^{\otimes\ell} \otimes C_t(\mathbb{T})\}_{t\in\mathbb{T}}$ are primitive ideals of $C(S_0^{2\ell+1})$,

(2) all other primitive ideals contain $p\otimes p\otimes\cdots\otimes p\otimes p\otimes t$ and $\mathcal{K}(L_2(\mathbb{N}))^{\otimes\ell}\otimes C_t(\mathbb{T})$ for all $t \in \mathbb{T}$.

Since $\ker\varphi_m$ is the intersection of all primitive ideals that contain $\ker\varphi_m$ and since $p \otimes p \otimes \cdots \otimes t \notin \ker\varphi_m$, we conclude that $\ker\varphi_m = \mathcal{K}(L_2(\mathbb{N}))^{\otimes\ell} \otimes C_F(\mathbb{T})$ for some closed subset F of $\mathbb{T}$ where $C_F(\mathbb{T})$ is the set of all continuous functions on $\mathbb{T}$ vanishing on F. Consider the function $\chi : C(\mathbb{T}) \to Q$ such that $\chi(t) = [(S^*)^m]$. Since $[(S^*)^m]$ is unitary in Q with spectrum equal to $\mathbb{T}$, it follows that the map χ is injective. This shows that for any nonzero continuous complex valued function f on $\mathbb{T}$, $\varphi_m(p \otimes p \otimes \cdots \otimes f(t)) \neq 0$. Hence $F = \mathbb{T}$ and $\ker\varphi_m = \{0\}$. □

We shall show that each element in the group $\operatorname{Ext}(C(S_0^{2\ell+1}))$ is of the form $[\varphi_m]_s$ for some $m \in \mathbb{Z}$. Let $\mathcal{H}_0$ be the Hilbert space $L_2(\mathbb{N})^{\otimes\ell} \otimes L_2(\mathbb{Z})$. For $m \in \mathbb{Z}$, let ϑ_m be the representation of $C(S_0^{2\ell+1})$ given by

$$\begin{aligned}
\vartheta_m : C(S_0^{2\ell+1}) &\to \mathcal{L}(\mathcal{H}_0),\\
S^* \otimes 1 \otimes \cdots \otimes 1 &\mapsto S^* \otimes 1 \otimes \cdots \otimes 1,\\
p \otimes S^* \otimes 1 \otimes \cdots \otimes 1 &\mapsto p \otimes S^* \otimes 1 \otimes \cdots \otimes 1,\\
&\vdots\\
p \otimes p \otimes \cdots \otimes p \otimes S^* \otimes 1 &\mapsto p \otimes p \otimes \cdots \otimes S^* \otimes 1,\\
p \otimes p \otimes \cdots \otimes t &\mapsto p \otimes p \otimes \cdots \otimes p \otimes (S^*)^m.
\end{aligned}$$

Let P be the self-adjoint projection in $\mathcal{L}(\mathcal{H}_0)$ on the subspace spanned by the basis elements $\{e_{n_1} \otimes \cdots \otimes e_{n_{\ell+1}} : n_i \in \mathbb{N} \text{ for all } i \in \{1, 2, \ldots, \ell + 1\}\}$. One can check that $\mathcal{F}_m := \big(C(S_0^{2\ell+1}), \mathcal{H}_0, 2P - 1\big)$ with the underlying representation ϑ_m is a Fredholm module. By [Blackadar 1998, Proposition 17.6.5, page 157], the group $\operatorname{Ext}(C(S_0^{2\ell+1}))$ is isomorphic to the group $K^1(C(S_0^{2\ell+1}))$. Under this identification, one can easily show that the equivalence class of the Fredholm module $\mathcal{F}_m$ corresponds to the equivalence class $[\varphi_m]_s$.

Proposition 3.2. *For $\ell \in \mathbb{N}$, one has*

$$\operatorname{Ext}\big(C(S_0^{2\ell+1})\big) = \big\{[\varphi_m]_s : m \in \mathbb{Z}\big\}.$$

Proof. To prove the claim, we will use the index pairing between the groups $K_1(C(S_0^{2\ell+1}))$ and $K^1(C(S_0^{2\ell+1}))$ which is given by the Kasparov product. The group $K_1(C(S_0^{2\ell+1}))$ is generated by the unitary $u := p^{\otimes\ell} \otimes t + 1 - p^{\otimes\ell} \otimes 1$. For $m \in \mathbb{Z}$, let $R_m : P\mathcal{H}_0 \to P\mathcal{H}_0$ be the operator $P\vartheta_m(u)P$. Hence we get

$$\langle u, \mathcal{F}_m \rangle = \operatorname{Index}(R_m) = m. \qquad \square$$

To describe all elements of $\operatorname{Ext}_{\mathrm{PPV}}(\mathbb{T}, C(S_0^{2\ell+1}))$, define the $*$-homomorphisms ϕ_m:

$$\begin{aligned}
\phi_m : C(S_0^{2\ell+1}) &\to Q\big(\mathcal{K}\big(L_2(\mathbb{N})^{\otimes\ell+1}\big) \otimes C(\mathbb{T})\big),\\
S^* \otimes 1 \otimes \cdots \otimes 1 &\mapsto S^* \otimes 1 \otimes \cdots \otimes 1,\\
p \otimes S^* \otimes 1 \otimes \cdots \otimes 1 &\mapsto p \otimes S^* \otimes 1 \otimes \cdots \otimes 1,\\
&\vdots\\
p \otimes p \otimes \cdots \otimes p \otimes S^* \otimes 1 &\mapsto p \otimes p \otimes \cdots \otimes S^* \otimes 1 \otimes 1,\\
p \otimes p \otimes \cdots \otimes t &\mapsto p \otimes p \otimes \cdots \otimes p \otimes (S^*)^m \otimes 1.
\end{aligned}$$

It follows from Proposition 3.1 that the ϕ_m are essential unital extensions. Since the last component is 1, these extensions are homogeneous. Let A_m be the C^*-subalgebra of $C(S_0^{2\ell+3})$ generated by the operators

$$\begin{gathered}
S^* \otimes 1 \otimes \cdots \otimes 1 \otimes 1,\\
p \otimes S^* \otimes 1 \otimes \cdots \otimes 1 \otimes 1,\\
\vdots\\
p \otimes p \otimes \cdots \otimes S^* \otimes 1 \otimes 1,\\
p \otimes p \otimes \cdots \otimes p \otimes (S^*)^m \otimes 1
\end{gathered}$$

and $\mathcal{K}\big(L_2(\mathbb{N})^{\otimes\ell+1}\big) \otimes C(\mathbb{T})$. Then for each $m \in \mathbb{Z}$, we have the exact sequence

$$0 \to \mathcal{K}\big(L_2(\mathbb{N})^{\otimes\ell+1}\big) \otimes C(\mathbb{T}) \to A_m \to C(S_0^{2\ell+1}) \to 0$$

with the Busby invariant ϕ_m. By using the six-term exact sequence, one can show

$$K_0(A_m) = \mathbb{Z} \oplus \mathbb{Z}/m\mathbb{Z}, \qquad K_1(A_m) = \mathbb{Z}. \tag{3-1}$$

Lemma 3.3. *For $\ell \in \mathbb{N}$ and $t_0 \in \mathbb{T}$, one has*

$$\operatorname{Ext}_{\mathrm{PPV}}\big(\mathbb{T}, t_0, C(S_0^{2\ell+1})\big) = \{0\}, \qquad \operatorname{Ext}_{\mathrm{PPV}}\big(\mathbb{T}, C(S_0^{2\ell+1})\big) = \mathbb{Z}.$$

Proof. It follows from Theorem 1.5 in [Rosenberg and Schochet 1981] that

$$\operatorname{Ext}_{\mathrm{PPV}}\big(\mathbb{T}, t_0, C(\mathbb{T})\big) = \operatorname{Ext}_{\mathrm{PPV}}\big(\mathbb{R}^+, t_0, C_0(\mathbb{R})^+\big) = \operatorname{Ext}\big(C_0(\mathbb{R}), C_0(\mathbb{R})\big) = \{0\}.$$

The C^*-algebra $C(S_0^{2\ell+1})$ can be obtained by applying quantum double suspension on $C(\mathbb{T})$ repeatedly; see [Hong and Szymański 2002]. Therefore, from Corollary 2.9, we have

$$\mathrm{Ext}_{\mathrm{PPV}}\big(\mathbb{T}, t_0, C(S_0^{2\ell+1})\big) = \mathrm{Ext}_{\mathrm{PPV}}\big(\mathbb{T}, t_0, C(\mathbb{T})\big) = \{0\}.$$

Further, from Theorem 1.4 in [Rosenberg and Schochet 1981], we get

$$\mathrm{Ext}_{\mathrm{PPV}}\big(\mathbb{T}, C(\mathbb{T})\big) = \mathrm{Ext}_{\mathrm{PPV}}\big(\mathbb{T}, C_0(\mathbb{R})^+\big) = \mathrm{Ext}\big(C_0(\mathbb{R}), C(\mathbb{T})\big) = \mathbb{Z}.$$

Hence by applying Lemma 2.8, we get the claim. □

The following lemma says that each element of the group $\mathrm{Ext}_{\mathrm{PPV}}\big(\mathbb{T}, C(S_0^{2\ell+1})\big)$ is of the form $[\phi_m]_{su}$ for some $m \in \mathbb{Z}$.

Lemma 3.4. *For $\ell \in \mathbb{N}$, one has*

$$\mathrm{Ext}_{\mathrm{PPV}}\big(\mathbb{T}, C(S_0^{2\ell+1})\big) = \big\{[\phi_m]_{su} : m \in \mathbb{Z}\big\}.$$

Proof. Fix $t_0 \in \mathbb{T}$. Define a homomorphism Ψ as follows:

$$\Psi : \mathrm{Ext}_{\mathrm{PPV}}\big(\mathbb{T}, C(S_0^{2\ell+1})\big) \to \mathrm{Ext}\big(C(S_0^{2\ell+1})\big), \qquad [\tau]_{su} \mapsto [ev_{t_0} \circ \tau]_s.$$

Clearly $\ker \Psi = \mathrm{Ext}_{\mathrm{PPV}}\big(\mathbb{T}, t_0, C(S_0^{2\ell+1})\big) = \{0\}$. Therefore, Ψ is an injective group homomorphism. Since $ev_{t_0} \circ \phi_m = \varphi_m$, for all $m \in \mathbb{Z}$, it follows that the homomorphism Ψ is surjective. This proves the claim. □

4. Quantum quaternion sphere

We first recall the definition and representation theory of the C^*-algebra $C(H_q^{2n})$ of continuous functions on the quantum quaternion sphere. Then we prove our main result that the C^*-algebra $C(H_q^{2n})$ is isomorphic to the C^*-algebra $C(S_q^{4n-1})$.

Definition 4.1. Let $i' = 2n+1-i$. The C^*-algebra $C(H_q^{2n})$ of continuous functions on the quantum quaternion sphere is defined as the universal C^*-algebra generated by elements $z_1, z_2, \ldots, z_{2n}$ satisfying the following relations:

(4-1) $z_i z_j = q z_j z_i$ for $i > j,\ i + j \neq 2n+1$,

(4-2) $z_i z_{i'} = q^2 z_{i'} z_i - (1-q^2) \sum_{k>i} q^{i-k} z_k z_{k'}$ for $i > n$,

(4-3) $z_i^* z_{i'} = q^2 z_{i'} z_i^*$

(4-4) $z_i^* z_j = q z_j z_i^*$ for $i + j > 2n+1,\ i \neq j$,

(4-5) $z_i^* z_j = q z_j z_i^* + (1-q^2)\epsilon_i \epsilon_j q^{\rho_i+\rho_j} z_{i'} z_{j'}^*$ for $i + j < 2n+1,\ i \neq j$,

(4-6) $z_i^* z_i = z_i z_i^* + (1-q^2) \sum_{k>i} z_k z_k^*$ for $i > n$,

(4-7) $z_i^* z_i = z_i z_i^* + (1-q^2)\Big(q^{2\rho_i} z_{i'} z_{i'}^* + \sum_{k>i} z_k z_k^*\Big)$ for $i \leq n$,

(4-8) $\sum_{i=1}^{2n} z_i z_i^* = 1.$

In [Saurabh 2017], we showed that the C^*-algebra $C(H_q^{2n})$ is isomorphic to the quotient algebra $C(\mathrm{SP}_q(2n)/\mathrm{SP}_q(2n-2))$ that can also be described as the C^*-subalgebra of $C(\mathrm{SP}_q(2n))$ generated by $\{u_m^1, u_m^{2n} : m \in \{1, 2, \dots, 2n\}\}$, i.e., elements of the first and last row of the fundamental matrix of the quantum symplectic group $\mathrm{SP}_q(2n)$. Here we briefly describe all irreducible representations of $C(H_q^{2n})$. For a detailed treatment on this, we refer the reader to [Saurabh 2017]. Let N be the number operator given by $N : e_n \mapsto ne_n$ and S be the shift operator given by $S : e_n \mapsto e_{n-1}$ on $L_2(\mathbb{N})$. We denote by $\mathscr{T}$ the Toeplitz algebra. Let $E_{i,j} \in M_n(\mathbb{R})$ be the $n \times n$ matrix with the only nonzero entry at the ij-th place and equal to 1. Define

$$\begin{aligned} s_i &= I - E_{i,i} - E_{i+1,i+1} + E_{i,i+1} + E_{i+1,i}, && \text{for } i = 1, 2, \dots, n-1, \\ s_n &= I - 2E_{n,n}, && \text{for } i = n. \end{aligned}$$

One can prove that the Weyl group W_n of sp_{2n} is isomorphic to a subgroup of $\mathrm{GL}(n, \mathbb{R})$ generated by $s_1, s_2, \dots, s_n$. We refer the reader to [Fulton and Harris 1991] for a proof of this fact. For $i = 1, 2, \dots, n-1$, let π_{s_i} denote the following representation of $C(\mathrm{SP}_q(2n))$,

$$\pi_{s_i}(u_l^k) = \begin{cases} \sqrt{1-q^{2N+2}}\,S & \text{if } (k,l) = (i,i) \text{ or } (2n-i, 2n-i), \\ S^*\sqrt{1-q^{2N+2}} & \text{if } (k,l) = (i+1,i+1) \text{ or } (2n-i+1, 2n-i+1), \\ -q^{N+1} & \text{if } (k,l) = (i, i+1), \\ q^N & \text{if } (k,l) = (i+1, i), \\ q^{N+1} & \text{if } (k,l) = (2n-i, 2n-i+1), \\ -q^N & \text{if } (k,l) = (2n-i+1, 2n-i), \\ \delta_{kl} & \text{otherwise.} \end{cases}$$

For $i = n$,

$$\pi_{s_n}(u_l^k) = \begin{cases} \sqrt{1-q^{4N+4}}\,S & \text{if } (k,l) = (n,n), \\ S^*\sqrt{1-q^{4N+4}} & \text{if } (k,l) = (n+1,n+1), \\ -q^{2N+2} & \text{if } (k,l) = (n, n+1), \\ q^{2N} & \text{if } (k,l) = (n+1, n), \\ \delta_{kl} & \text{otherwise.} \end{cases}$$

Each π_{s_i} is an irreducible representation and is called an elementary representation of $C(\mathrm{SP}_q(2n))$. For any two representations φ and ψ of $C(\mathrm{SP}_q(2n))$, define $\varphi * \psi$ to be $(\varphi \otimes \psi) \circ \Delta$, where Δ is the comultiplication map of $C(\mathrm{SP}_q(2n))$. Let ϑ be an element of W_n such that $s_{i_1} s_{i_2} \cdots s_{i_k}$ is a reduced expression for ϑ. Then $\pi_\vartheta = \pi_{s_{i_1}} * \pi_{s_{i_2}} * \cdots * \pi_{s_{i_k}}$ is an irreducible representation which is independent of the reduced expression. Now for $t = (t_1, t_2, \dots, t_n) \in \mathbb{T}^n$, define the map $\tau_t : C(\mathrm{SP}_q(2n) \to \mathbb{C}$ by

$$\tau_t(u^i_j) = \begin{cases} \overline{t_i}\delta_{ij} & \text{if } i \le n, \\ t_{2n+1-i}\delta_{ij} & \text{if } i > n. \end{cases}$$

Then τ_t is a $*$-algebra homomorphism. For $t \in \mathbb{T}^n$, $\vartheta \in W$, let $\pi_{t,\vartheta} = \tau_t * \pi_\vartheta$. Define the representation $\eta_{t,\vartheta}$ of $C(H^{2n}_q)$ as the representation $\pi_{t,\vartheta}$ restricted to $C(H^{2n}_q)$. Denote by ω_k the following reduced word of Weyl group of sp_{2n},

$$\omega_k = \begin{cases} I & \text{if } k = 1, \\ s_1 s_2 \cdots s_{k-1} & \text{if } 2 \le k \le n, \\ s_1 s_2 \cdots s_{n-1} s_n s_{n-1} \cdots s_{2n-k+1} & \text{if } n < k \le 2n. \end{cases}$$

For $k=1$, define $\eta_{t,I} : C(H^{2n}_q) \to \mathbb{C}$ such that $\eta_{t,I}(z_j) = t\delta_{1j}$. The set $\{\eta_{t,I} : t \in T\}$ gives all one-dimensional irreducible representations of $C(H^{2n}_q)$.

Theorem 4.2 [Saurabh 2017]. *The set $\{\eta_{t,\omega_k} : 1 \le k \le 2n, t \in \mathbb{T}\}$ gives a complete list of irreducible representations of $C(H^{2n}_q)$.*

Define

$$\eta_{\omega_k} : C(H^{2n}_q) \to C(\mathbb{T}) \otimes \mathscr{T}^{\otimes k-1}$$

such that $\eta_{\omega_k}(a)(t) = \eta_{t,\omega_k}(a)$ for all $a \in C(H^{2n}_q)$. Let $C^{2n}_1 = C(\mathbb{T})$ and for $2 \le k \le 2n$, $C^{2n}_k = \eta_{\omega_k}(C(H^{2n}_q))$.

Corollary 4.3. *The set $\{\eta_{t,\omega_l} : 1 \le l \le k, t \in \mathbb{T}\}$ gives a complete list of irreducible representations of C^{2n}_k.*

By Corollary 4.3, one can find all primitive ideals, i.e., kernels of irreducible representations of C^{2n}_k. Define $y^k_l := \eta_{\omega_k}(z_l)$ and $I^k_{t,l} := \ker(\eta_{t,\omega_l})$ for $1 \le l \le k$ and $t \in \mathbb{T}$. Then

$$\{I^k_{t,k} = C_t(\mathbb{T}) \otimes \mathcal{K}(L_2(\mathbb{N}))^{\otimes(k-1)}\}_{t\in\mathbb{T}}, \{I^k_{t,k-1}\}_{t\in\mathbb{T}}, \dots, \{I^k_{t,1}\}_{t\in\mathbb{T}} \tag{4-9}$$

is a complete list of primitive ideals of C^{2n}_k. Moreover for $t, t' \in \mathbb{T}$ and $1 \le l \le k-1$, we have $C_t(\mathbb{T}) \otimes \mathcal{K}(L_2(\mathbb{N}))^{\otimes(k-1)} \subset I^k_{t',l}$ and $y^k_k \in I^k_{t',l}$. In Lemma 5.1 of [Saurabh 2017], we established the exact sequence

$$0 \to C(\mathbb{T}) \otimes \mathcal{K}(L_2(\mathbb{N}))^{\otimes(k)} \to C^{2n}_{k+1} \xrightarrow{\sigma_{k+1}} C^{2n}_k \to 0,$$

where σ_{k+1} is the restriction of $1 \otimes 1^{\otimes(k-1)} \otimes \sigma$ to C^{2n}_{k+1} and the map $\sigma : \mathscr{T} \to \mathbb{C}$ is the homomorphism such that $\sigma(S) = 1$. The following lemma says that this exact sequence is a unital homogeneous extension of C^{2n}_k by $C(\mathbb{T}) \otimes \mathcal{K}$:

Lemma 4.4. *For $1 \le k \le 2n$, the exact sequence*

$$0 \to C(\mathbb{T}) \otimes \mathcal{K}(L_2(\mathbb{N}))^{\otimes(k)} \to C^{2n}_{k+1} \xrightarrow{\sigma_{k+1}} C^{2n}_k \to 0$$

is a unital homogeneous extension of C^{2n}_k by $C(\mathbb{T}) \otimes \mathcal{K}$.

Proof. Since C_{k+1}^{2n} is unital, the given extension is unital. Let $\tau : C_k^{2n} \to Q(\mathbb{T})$ be the Busby invariant corresponding to this extension. For $t_0 \in \mathbb{T}$, let $\tau_{t_0} : C_k^{2n} \to Q$ be the map $ev_{t_0} \circ \tau$ where $ev_{t_0} : Q(\mathbb{T}) \to Q$ is the evaluation map at t_0. Assume that $J_{t_0} = \ker(\tau_{t_0})$. To show that the given short exact sequence is a homogeneous extension, we need to prove that $J_{t_0} = \{0\}$ for all $t_0 \in \mathbb{T}$.

Case 1: $n < k < 2n$. We have

$$\begin{aligned} \tau_{t_0}(y_k^k) &= \tau_{t_0}\big(t \otimes q^{N\otimes(n-1)} \otimes q^{2N} \otimes q^{N\otimes(k-n-1)}\big) \\ &= t_0\big[q^{N\otimes(n-1)} \otimes q^{2N} \otimes q^{N\otimes(k-n-1)} \otimes \sqrt{1-q^{2N}}S^*\big] \neq 0. \end{aligned} \tag{4-10}$$

This shows $y_k^k \notin J_{t_0}$. Since J_{t_0} is the intersection of all primitive ideals that contain J_{t_0}, and $y_k^k \in I_{t',l}^k$ and $C_t(\mathbb{T}) \otimes \mathcal{K}(L_2(\mathbb{N}))^{\otimes(k-1)} \subset I_{t',l}^k$ for $t, t' \in \mathbb{T}$ and $1 \le l \le k-1$, we conclude that J_{t_0} is equal to $C_F(\mathbb{T}) \otimes \mathcal{K}$ for some closed subset F of $\mathbb{T}$ where $C_F(\mathbb{T})$ is the set of all continuous functions on $\mathbb{T}$ vanishing on F. From (4-10), we get

$$\begin{aligned} \tau_{t_0}((y_k^k)(y_k^k)^*) &= \big[q^{2N\otimes(n-1)} \otimes q^{4N} \otimes q^{2N\otimes(k-n-1)} \otimes (1-q^{2N})\big] \\ &= \big[q^{2N\otimes(n-1)} \otimes q^{4N} \otimes q^{2N\otimes(k-n-1)} \otimes 1\big]. \end{aligned}$$

Therefore,

$$\tau_{t_0}\big(1 \otimes p^{\otimes(k-1)}\big) = \big[p^{\otimes(k-1)} \otimes 1\big].$$

Hence,

$$\begin{aligned} \tau_{t_0}(t \otimes p^{\otimes(k-1)}) &= t_0\big[p^{\otimes(k-1)} \otimes \sqrt{1-q^{2N}}S^*\big] \\ &= t_0\big[p^{\otimes(k-1)} \otimes S^*\big]. \end{aligned}$$

Since the function $\chi : C(\mathbb{T}) \to Q$ such that $\chi(t) = [S^*]$ is an injective homomorphism as shown in Proposition 3.1, it follows that for any nonzero continuous function f on $\mathbb{T}$,

$$\tau_{t_0}\big(f(t) \otimes p^{\otimes(k-1)}\big) \neq 0.$$

This proves that $F = \mathbb{T}$ and $J_{t_0} = \{0\}$.

Case 2: $1 \le k \le n$. For $k = n$,

$$\tau_{t_0}(y_n^n) = t_0\big[q^{N\otimes(n-1)} \otimes \sqrt{1-q^{4N}}S^*\big].$$

For $1 \le k < n$,

$$\tau_{t_0}(y_k^k) = t_0\big[q^{N\otimes(k-1)} \otimes \sqrt{1-q^{2N}}S^*\big].$$

Similar calculations to those in Case 1 show that $J_{t_0} = \{0\}$. This proves the claim. □

We now state the main result of this paper.

Theorem 4.5. *For all $n \geq 2$ and $1 \leq k \leq 2n$, the C^*-algebra C_k^{2n} is isomorphic to the C^*-algebra $C(S_0^{2k-1})$ of continuous functions on odd-dimensional quantum spheres. In particular, $C(H_q^{2n})$ is isomorphic to $C(S_0^{4n-1})$ or, equivalently, to $C(S_q^{4n-1})$.*

Proof. Fix n. To prove the theorem, we use induction on k. For $k = 1$, $C_1^{2n} = C(\mathbb{T})$. So the claim is true for $k = 1$. Assume that the claim is true for k, i.e., C_k^{2n} is isomorphic to $C(S_0^{2k-1})$. From Lemma 4.4, it follows that the short exact sequence

$$0 \to C(\mathbb{T}) \otimes \mathcal{K} \to C_{k+1}^{2n} \to C_k^{2n} \to 0 \tag{4-11}$$

is a unital homogeneous extension. Therefore, it can be viewed as an element of the group $\mathrm{Ext}_{\mathrm{PPV}}(\mathbb{T}, C(S_0^{2k-1}))$. It follows from Lemma 3.4 that it is strongly unitarily equivalent to ϕ_m or, equivalently, to the following exact sequence

$$0 \to C(\mathbb{T}) \otimes \mathcal{K} \to A_m \to C(S_0^{2k-1}) \to 0,$$

for some $m \in \mathbb{Z}$. From Theorem 5.3 in [Saurabh 2017] and equation (3-1), we have

$$K_0(C_{k+1}^{2n}) = \mathbb{Z}, \qquad K_0(A_m) = \mathbb{Z} \oplus \mathbb{Z}/m\mathbb{Z}.$$

Since strongly unitary equivalence gives an isomorphism of the middle C^* algebras and hence an isomorphism of the K-groups of middle C^*-algebras, it follows that the exact sequence (4-11) is strongly unitarily equivalent to ϕ_1 or ϕ_{-1}. This implies that C_{k+1}^{2n} is isomorphic to A_1 or A_{-1}. Since $A_1 = A_{-1} = C(S_0^{2k+1})$, it follows that C_{k+1}^{2n} is isomorphic to $C(S_0^{2k+1})$. Hence by induction, it follows that $C(H_q^{2n})$ is isomorphic to $C(S_0^{4n-1})$. From Theorem 4.4 in [Hong and Szymański 2002], it follows that the C^*-algebra $C(S_q^{4n-1})$ is isomorphic to $C(S_0^{4n-1})$, for $q \in (0, 1)$. This proves that $C(H_q^{2n})$ is isomorphic to $C(S_q^{4n-1})$. □

Remark 4.6. In the case where $q = 0$, we need to be slightly careful to get the defining relations of $C(H_0^{2n})$. In the relation (4-2), we first start with $i = 2n$. This gives the relation $z_{2n} z_1 = 0$. Then we take $i = 2n - 1$ and so on and get the relation $z_i z_{i'} = 0$ for $i < n$. Further, in the relation (4-5), it is easy to check that for $i + j < 2n + 1$, $\rho_i + \rho_j > 0$. Now by putting $q = 0$ into the relations (4-3), (4-4) and (4-4), we get $z_i^* z_j = 0$ for $i \neq j$. The other relations are obtained by putting $q = 0$ in the remaining relations. By looking at the relations, one can see that the defining relations of $C(H_0^{2n})$ are exactly the same as those of $C(S_0^{4n-1})$. These facts together with Theorem 4.5 prove that for different values of $q \in [0, 1)$, the C^*-algebras $C(H_q^{2n})$ are isomorphic.

Acknowledgement

I would like to thank Arup Kumar Pal, my supervisor, for his constant support and for his valuable suggestions.

References

[Blackadar 1998] B. Blackadar, *K-theory for operator algebras*, 2nd ed., Mathematical Sciences Research Institute Publications **5**, Cambridge Univ. Press, 1998. MR Zbl

[Brown et al. 1977] L. G. Brown, R. G. Douglas, and P. A. Fillmore, "Extensions of C^*-algebras and K-homology", *Ann. of Math.* (2) **105**:2 (1977), 265–324. MR Zbl

[Chakraborty and Pal 2008] P. S. Chakraborty and A. Pal, "Characterization of $\mathrm{SU}_q(\ell+1)$-equivariant spectral triples for the odd-dimensional quantum spheres", *J. Reine Angew. Math.* **623** (2008), 25–42. MR Zbl

[Elliott and Kucerovsky 2001] G. A. Elliott and D. Kucerovsky, "An abstract Voiculescu–Brown–Douglas–Fillmore absorption theorem", *Pacific J. Math.* **198**:2 (2001), 385–409. MR Zbl

[Fulton and Harris 1991] W. Fulton and J. Harris, *Representation theory: a first course*, Graduate Texts in Mathematics **129**, Springer, New York, 1991. MR Zbl

[Hong and Szymański 2002] J. H. Hong and W. Szymański, "Quantum spheres and projective spaces as graph algebras", *Comm. Math. Phys.* **232**:1 (2002), 157–188. MR Zbl

[Hong and Szymański 2008] J. H. Hong and W. Szymański, "Noncommutative balls and mirror quantum spheres", *J. Lond. Math. Soc.* (2) **77**:3 (2008), 607–626. MR Zbl

[Kasparov 1979] G. G. Kasparov, "The K-functors in the theory of extensions of C^*-algebras", *Funktsional. Anal. i Prilozhen.* **13**:4 (1979), 73–74. In Russian; translated in *Functional Anal. Appl.* **13**:4 (1979), 296–297. MR

[Kucerovsky and Ng 2006] D. Kucerovsky and P. W. Ng, "The corona factorization property and approximate unitary equivalence", *Houston J. Math.* **32**:2 (2006), 531–550. MR Zbl

[Lin 2007] H. Lin, "Full extensions and approximate unitary equivalence", *Pacific J. Math.* **229**:2 (2007), 389–428. MR Zbl

[Lin 2009] H. Lin, "Unitary equivalences for essential extensions of C^*-algebras", *Proc. Amer. Math. Soc.* **137**:10 (2009), 3407–3420. MR Zbl

[Neshveyev and Tuset 2012] S. Neshveyev and L. Tuset, "Quantized algebras of functions on homogeneous spaces with Poisson stabilizers", *Comm. Math. Phys.* **312**:1 (2012), 223–250. MR Zbl

[Pal and Sundar 2010] A. Pal and S. Sundar, "Regularity and dimension spectrum of the equivariant spectral triple for the odd-dimensional quantum spheres", *J. Noncommut. Geom.* **4**:3 (2010), 389–439. MR Zbl

[Pimsner et al. 1979] M. Pimsner, S. Popa, and D. Voiculescu, "Homogeneous C^*-extensions of $C(X) \otimes K(H)$, I", *J. Operator Theory* **1**:1 (1979), 55–108. MR Zbl

[Podkolzin and Vainerman 1999] G. B. Podkolzin and L. I. Vainerman, "Quantum Stiefel manifold and double cosets of quantum unitary group", *Pacific J. Math.* **188**:1 (1999), 179–199. MR Zbl

[Rosenberg and Schochet 1981] J. Rosenberg and C. Schochet, "Comparing functors classifying extensions of C^*-algebras", *J. Operator Theory* **5**:2 (1981), 267–282. MR Zbl

[Saurabh 2017] B. Saurabh, "Quantum quaternion spheres", *Proc. Indian Acad. Sci. Math. Sci.* **127**:1 (2017), 133–164. MR

[Schochet 1980] C. Schochet, "Homogeneous extensions of C^*-algebras and K-theory, I", *Bull. Amer. Math. Soc.* (*N.S.*) **3**:1, part 1 (1980), 715–718. MR Zbl

[Vaksman and Soibelman 1990] L. L. Vaksman and Y. S. Soibelman, "Algebra of functions on the quantum group $\mathrm{SU}(n+1)$, and odd-dimensional quantum spheres", *Algebra i Analiz* **2**:5 (1990), 101–120. In Russian; translated in *Leningrad Math. J.* **2**:5 (1991), 1023–1042. MR

Received June 22, 2016. Revised December 12, 2016.

BIPUL SAURABH
INSTITUTE OF MATHEMATICAL SCIENCES
CIT CAMPUS, TARAMANI
CHENNAI 600113
INDIA
saurabhbipul2@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.453

GAP THEOREMS FOR COMPLETE λ-HYPERSURFACES

HUIJUAN WANG, HONGWEI XU AND ENTAO ZHAO

An n-dimensional λ-hypersurface $X : M \to \mathbb{R}^{n+1}$ is the critical point of the weighted area functional $\int_M e^{-\frac{1}{4}|X|^2}\,\mathrm{d}\mu$ for weighted volume-preserving variations, which is also a generalization of the self-shrinking solution of the mean curvature flow. We first prove that if the L^n-norm of the second fundamental form of the λ-hypersurface $X : M \to \mathbb{R}^{n+1}$ with $n \geq 3$ is less than an explicit positive constant $K(n, \lambda)$, then M is a hyperplane. Secondly, we show that if the L^n-norm of the trace-free second fundamental form of M with $n \geq 3$ is less than an explicit positive constant $D(n, \lambda)$ and the mean curvature is suitably bounded, then M is a hyperplane. We also obtain similar results for λ-surfaces in $\mathbb{R}^3$ under L^4-curvature pinching conditions.

1. Introduction

Let $X : M \to \mathbb{R}^{n+1}$ be an n-dimensional immersed smooth hypersurface in the $(n+1)$-dimensional Euclidean space $\mathbb{R}^{n+1}$. We call the hypersurface a λ-hypersurface if it satisfies

$$H + \tfrac{1}{2}\langle X, N\rangle = \lambda,$$

where λ is a constant, H is the mean curvature and N is the unit inward normal vector of $X : M \to \mathbb{R}^{n+1}$.

McGonagle and Ross [2015] studied λ-hypersurfaces from the viewpoint of variation. Let $A_\mu(M)$ be the functional defined by $A_\mu(M) = \int_M e^{-\frac{1}{4}|X|^2}\,\mathrm{d}\mu$. They showed that the critical points of $\delta A_\mu(u) = 0$ for $u \in C_0^\infty$ satisfying

$$\int_M e^{-\frac{1}{4}|X|^2} u \,\mathrm{d}\mu = 0$$

are λ-hypersurfaces. Cheng and Wei [2014a] also introduced λ-hypersurfaces in a different way by investigating the weighted volume-preserving mean curvature flow. Obviously, when $\lambda = 0$, a λ-hypersurface is a self-shrinker of the mean curvature flow. It is well known that self-shrinkers play an important role in the study of mean

Research supported by the National Natural Science Foundation of China, Grant Nos. 11531012, 11371315, 11201416.
MSC2010: 53C42, 53C44.
Keywords: gap theorem, lambda-hypersurfaces, integral curvature pinching.

curvature flow because they describe the singularity models of the mean curvature flow and they arise as tangent flows of mean curvature flow at singularities; see, for example, [Colding and Minicozzi 2012; Huisken 1990; Ilmanen 1995; White 1997].

The rigidity phenomena of self-shrinkers has been studied extensively [Cheng and Peng 2015; Cheng and Wei 2015; Colding et al. 2015; Colding and Minicozzi 2012; Ding and Xin 2013; 2014; Huisken 1990; Le and Sesum 2011]. For example, Le and Sesum [2011] proved that a smooth self-shrinker with polynomial volume growth and satisfying $|A|^2 < \frac{1}{2}$ is a hyperplane. Here A denotes the second fundamental form of the immersion. Cao and Li [2013] generalized this result to arbitrary codimension and proved that any smooth complete self-shrinker with polynomial volume growth and $|A|^2 \leq \frac{1}{2}$ is a generalized cylinder. On the other hand, Ding and Xin [2014] showed that a smooth complete self-shrinker satisfying $\left(\int_M |A|^n \,\mathrm{d}\mu\right)^{1/n} < C$ for a certain positive constant C is a linear space. For more curvature pinching theorems for self-shrinkers, see [Cao et al. 2014; Li and Wei 2014; Lin 2016].

The geometric properties of λ-hypersurfaces were recently investigated by Cheng, Wei, Ogata, Guang [Cheng and Wei 2014a; Cheng et al. 2016; Guang 2014]. As generalizations of self-shrinkers of the mean curvature flow, complete λ-hypersurfaces with polynomial area growth and $H - \lambda \geq 0$ were classified by Cheng and Wei [2014a]. They also defined an $\mathcal{F}$-functional and studied $\mathcal{F}$-stability of λ-hypersurfaces. Cheng, Ogata and Wei [Cheng et al. 2016] proved some gap and rigidity theorems for complete λ-hypersurfaces. See [Cheng and Wei 2014b; Guang 2014; Ogata 2015] for more results on the rigidity of λ-hypersurfaces.

We study the integral curvature pinching theorems for λ-hypersurfaces. We first prove the following L^n-pinching theorem of the second fundamental form.

Theorem 1. *Let $X : M^n \to \mathbb{R}^{n+1}$ $(n \geq 3)$ be an n-dimensional complete λ-hypersurface in the Euclidean space $\mathbb{R}^{n+1}$. If*

$$\left(\int_M |A|^n \,\mathrm{d}\mu\right)^{1/n} < K(n, \lambda),$$

where $K(n, \lambda)$ is an explicit positive constant depending only on n and λ, then $|A| \equiv 0$ and M is a hyperplane.

Remark. It is easy to see from the expression of $K(n, \lambda)$ that $\lim_{\lambda\to 0} K(n, \lambda) = K_n$ for a positive constant K_n depending only on n. Hence if $\lambda = 0$, Theorem 1 reduces to the L^n-pinching theorem for self-shrinkers due to Ding and Xin [2014].

Let $\mathring{A}$ denote the trace-free second fundamental form, which is defined by $\mathring{A} = A - (H/n)g$ with g denoting the induced metric on M. We prove an L^n-pinching theorem of the trace-free second fundamental form for λ-hypersurfaces provided that the mean curvature is suitably bounded.

Theorem 2. *Let $X : M^n \to \mathbb{R}^{n+1}$ $(n \geq 3)$ be an n-dimensional complete λ-hypersurface in the Euclidean space $\mathbb{R}^{n+1}$. Suppose the mean curvature satisfies*

$$|H| \leq \sqrt{\tfrac{1}{3}n + \lambda^2} - |\lambda|.$$

If

$$\left(\int_M |\mathring{A}|^n \,\mathrm{d}\mu\right)^{1/n} < D(n, \lambda),$$

where $D(n, \lambda)$ is an explicit positive constant depending on n and λ, then M is a hyperplane.

For the case $n = 2$, we obtain the following results.

Theorem 3. *Let $X : M^2 \to \mathbb{R}^3$ be a 2-dimensional complete λ-hypersurface in the Euclidean space $\mathbb{R}^3$. If*

$$\left(\int_M |A|^4 \,\mathrm{d}\mu\right)^{1/2} < K(\lambda),$$

where $K(\lambda)$ is an explicit positive constant depending only on λ, then $|A| \equiv 0$ and M is a hyperplane.

Theorem 4. *Let $X : M^2 \to \mathbb{R}^3$ be a 2-dimensional complete λ-hypersurface in the Euclidean space $\mathbb{R}^3$. Suppose the mean curvature satisfies*

$$|H| \leq \sqrt{\tfrac{2}{3} + \lambda^2} - |\lambda|.$$

If

$$\left(\int_M |\mathring{A}|^4 \,\mathrm{d}\mu\right)^{1/2} < D(\lambda),$$

where $D(\lambda)$ is an explicit positive constant depending on λ, then M is a hyperplane.

The rest of our paper is organized as follows. Some notation and several lemmas are prepared in Section 2. In Section 3, we prove Theorems 1 and 2. Theorems 3 and 4 will be proved in Section 4.

2. Preliminaries

Let $X : M^n \to \mathbb{R}^{n+1}$ be an n-dimensional connected hypersurface. Denote by g and $\mathrm{d}\mu$ the induced metric and the volume form on M, respectively. We shall make use of the following convention on the range of indices:

$$1 \leq A, B, C, \ldots \leq n+1, \quad 1 \leq i, j, k, \ldots \leq n.$$

Choose local orthonormal frame fields $\{e_A\}$ in $\mathbb{R}^{n+1}$ such that, restricted to M, the e_i are tangent to M. Let $\{\omega_A\}$ and $\{\omega_{AB}\}$ be the dual frame fields and the connection 1-forms of $\mathbb{R}^{n+1}$, respectively. Then we have the following structure equations:

$$dX = \sum_i \omega_i e_i, \quad de_i = \sum_j \omega_{ij} e_j + \sum_j h_{ij}\omega_j e_{n+1},$$

and

$$de_{n+1} = -\sum_{i,j} h_{ij}\omega_j e_i.$$

Restricting these forms to M, we have

$$\omega_{n+1i} = \sum_j h_{ij}\omega_j, \quad h_{ij} = h_{ji},$$

where h_{ij} denotes the components of the second fundamental form of M. $H = \sum_i h_{ii}$ is the mean curvature and $A = \sum_{i,j} h_{ij}\omega_i \otimes \omega_j$ is the second fundamental form of $X : M^n \to \mathbb{R}^{n+1}$. The trace-free second fundamental form is defined by $\mathring{A} = A - (H/n)g$.

Let $h_{ijk} = \nabla_k h_{ij}$, $h_{ijkl} = \nabla_l \nabla_k h_{ij}$, where ∇ is the Levi-Civita connection on M. Gauss equations, Codazzi equations and Ricci formulas are given by

$$R_{ijkl} = h_{ik}h_{jl} - h_{il}h_{jk}, \quad h_{ijk} = h_{ikj},$$

$$h_{ijkl} - h_{ijlk} = \sum_{m=1}^{n} h_{im}R_{mjkl} + \sum_{m=1}^{n} h_{mj}R_{mikl}.$$

For λ-hypersurfaces, an elliptic operator $\mathcal{L}$ is given by

$$\mathcal{L} = \Delta - \tfrac{1}{2}\langle X, \nabla(\,\cdot\,)\rangle = e^{\frac{1}{4}|X|^2}\operatorname{div}\big(e^{-\frac{1}{4}|X|^2}\nabla(\,\cdot\,)\big),$$

where Δ and div denote the Laplacian and divergence on the λ-hypersurface, respectively. The $\mathcal{L}$ operator was introduced by Colding and Minicozzi [2012] when they investigated self-shrinkers. They showed that $\mathcal{L}$ is self-adjoint with respect to the measure $e^{-\frac{1}{4}|X|^2}\,d\mu$. We set $\rho = e^{-\frac{1}{4}|X|^2}$ and the volume form $d\mu$ might be omitted in the integrations for notational simplicity.

The following lemma, which was proved in [Cheng and Wei 2014a], is needed in order to prove our results. For convenience, we also include the proof here.

Lemma 5. *Let $X : M \to \mathbb{R}^{n+1}$ be a λ-hypersurface satisfying $H + \frac{1}{2}\langle X, N\rangle = \lambda$. Then*

$$\tfrac{1}{2}\mathcal{L}H^2 = |\nabla H|^2 + \tfrac{1}{2}H^2 + |A|^2(\lambda - H)H, \tag{1}$$

$$\tfrac{1}{2}\mathcal{L}|A|^2 = |\nabla A|^2 + \big(\tfrac{1}{2} - |A|^2\big)|A|^2 + \lambda f_3, \tag{2}$$

where $f_3 = \sum_{i,j,k} h_{ij}h_{jk}h_{ki}$.

Proof. Since $H + \frac{1}{2}\langle X, N\rangle = \lambda$, one has

$$\nabla_i H = \frac{1}{2}\sum_j h_{ij}\langle X, e_j\rangle,$$

and

$$\nabla_k\nabla_i H = \frac{1}{2}\sum_j h_{ijk}\langle X, e_j\rangle + \tfrac{1}{2}h_{ik} + \sum_j h_{ij}h_{jk}(\lambda - H).$$

Hence,

$$\Delta H = \sum_i \nabla_i\nabla_i H = \frac{1}{2}\sum_i \nabla_i H\langle X, e_i\rangle + \tfrac{1}{2}H + |A|^2(\lambda - H),$$

and

$$\mathcal{L}H = \Delta H - \frac{1}{2}\sum_i \nabla_i H\langle X, e_i\rangle = \tfrac{1}{2}H + |A|^2(\lambda - H).$$

Therefore, we obtain

$$\tfrac{1}{2}\mathcal{L}H^2 = \tfrac{1}{2}\Delta H^2 - \frac{1}{4}\sum_i \nabla_i H^2\langle X, e_i\rangle = |\nabla H|^2 + \tfrac{1}{2}H^2 + |A|^2(\lambda - H)H.$$

By using the Ricci formulas, the Gauss equations and the Codazzi equations, we have

$$\begin{aligned}\mathcal{L}h_{ij} &= \Delta h_{ij} - \frac{1}{2}\sum_k \langle X, e_k\rangle h_{ijk}\\ &= \sum_k h_{ijkk} - \frac{1}{2}\sum_k \langle X, e_k\rangle h_{ijk}\\ &= \big(\tfrac{1}{2} - |A|^2\big)h_{ij} + \lambda\sum_k h_{ik}h_{kj}.\end{aligned}$$

Then it follows that

$$\begin{aligned}\tfrac{1}{2}\mathcal{L}|A|^2 &= \tfrac{1}{2}\Delta\Big(\sum_{ij} h_{ij}^2\Big) - \frac{1}{4}\sum_k \langle X, e_k\rangle\nabla_k\Big(\sum_{ij} h_{ij}^2\Big)\\ &= \sum_{i,j,k} h_{ijk}^2 + \big(\tfrac{1}{2} - |A|^2\big)\sum_{ij} h_{ij}^2 + \lambda\sum_{i,j,k} h_{ik}h_{kj}h_{ji}\\ &= |\nabla A|^2 + \big(\tfrac{1}{2} - |A|^2\big)|A|^2 + \lambda f_3,\end{aligned}$$

where $f_3 = \sum_{i,j,k} h_{ij}h_{jk}h_{ki}$. □

We need the following Sobolev inequality for submanifolds in the Euclidean space.

Lemma 6 [Xu and Gu 2007a; Hoffman and Spruck 1974]. *Let M^n $(n \geq 3)$ be an n-dimensional complete submanifold in the Euclidean space $\mathbb{R}^{n+p}$. Let f be a* *nonnegative C^1 function with compact support. Then we have*

$$\|f\|_{2n/(n-2)}^2 \le D^2(n)\left[\frac{4(n-1)^2(1+s)}{(n-2)^2}\|\nabla f\|_2^2 + \left(1+\frac{1}{s}\right)\frac{1}{n^2}\big\||H|f\big\|_2^2\right],$$

where

$$D(n) = 2^n(1+n)^{(n+1)/n}(n-1)^{-1}\sigma_n^{-1/n},$$

and σ_n denotes the volume of the unit ball in $\mathbb{R}^n$.

3. Gap theorems for λ-hypersurfaces

Proof of Theorem 1. It follows from (2) and the inequality $|\nabla A|^2 \ge \big|\nabla|A|\big|^2$, which is an easy consequence of the Schwartz inequality, that

$$\begin{aligned}\mathcal{L}|A|^2 &= 2|\nabla A|^2 + 2\big(\tfrac{1}{2} - |A|^2\big)|A|^2 + 2\lambda f_3\\ &\ge 2\big|\nabla|A|\big|^2 + 2\big(\tfrac{1}{2} - |A|^2\big)|A|^2 - 2|\lambda||A|^3.\end{aligned}$$

Let η be a smooth function with compact support on M. Multiplying $\eta^2|A|^{n-2}$ on both sides of the inequality above and integrating by parts with respect to the measure $\rho\,\mathrm{d}\mu$ on M yields that for any $\tau > 0$

$$\begin{aligned}0 &\ge 2\int_M \big|\nabla|A|\big|^2|A|^{n-2}\eta^2\rho + \int_M |A|^n\eta^2\rho - 2\int_M |A|^{n+2}\eta^2\rho\\ &\qquad - 2|\lambda|\int_M |A|^{n+1}\eta^2\rho - \int_M \eta^2|A|^{n-2}\rho\mathcal{L}|A|^2\\ &= 2\int_M \big|\nabla|A|\big|^2|A|^{n-2}\eta^2\rho + \int_M |A|^n\eta^2\rho - 2\int_M |A|^{n+2}\eta^2\rho\\ &\qquad - 2|\lambda|\int_M |A|^{n+1}\eta^2\rho + 2\int_M \rho|A|\nabla|A|\cdot\nabla(|A|^{n-2}\eta^2)\\ &= 2(n-1)\int_M \big|\nabla|A|\big|^2|A|^{n-2}\eta^2\rho + \int_M |A|^n\eta^2\rho - 2\int_M |A|^{n+2}\eta^2\rho\\ &\qquad - 2|\lambda|\int_M |A|^{n+1}\eta^2\rho + 4\int_M (\nabla|A|\cdot\nabla\eta)|A|^{n-1}\eta\rho\\ &\ge 2(n-1)\int_M \big|\nabla|A|\big|^2|A|^{n-2}\eta^2\rho + \int_M |A|^n\eta^2\rho - 2\int_M |A|^{n+2}\eta^2\rho\\ &\qquad - 2|\lambda|\left(\frac{\tau}{2}\int_M |A|^n\eta^2\rho + \frac{1}{2\tau}\int_M |A|^{n+2}\eta^2\rho\right) + 4\int_M (\nabla|A|\cdot\nabla\eta)|A|^{n-1}\eta\rho.\end{aligned}$$

By the Cauchy inequality, for any $\varepsilon > 0$, we have

$$\left(\frac{|\lambda|}{\tau}+2\right)\int_M |A|^{n+2}\eta^2\rho + (|\lambda|\tau - 1)\int_M |A|^n\eta^2\rho + \frac{2}{\varepsilon}\int_M |A|^n|\nabla\eta|^2\rho \geq 2(n-1-\varepsilon)\int_M \big|\nabla|A|\big|^2|A|^{n-2}\eta^2\rho. \tag{3}$$

Set $f = |A|^{n/2}\rho^{1/2}\eta$. Integrating by parts, we obtain

$$\begin{aligned}\int_M |\nabla f|^2 &= \int_M \big|\nabla(|A|^{n/2}\eta)\big|^2\rho + \int_M |A|^n\eta^2|\nabla\rho^{1/2}|^2 + \frac{1}{2}\int_M \nabla(|A|^n\eta^2)\nabla\rho \\ &= \int_M \big|\nabla(|A|^{n/2}\eta)\big|^2\rho + \frac{1}{16}\int_M |A|^n\eta^2|X^T|^2\rho - \frac{1}{2}\int_M |A|^n\eta^2\Delta\rho.\end{aligned} \tag{4}$$

Since

$$\Delta|X|^2 = 2|\nabla X|^2 + 2\langle X, \Delta X\rangle = 2n + 2H\langle X, N\rangle = 2n + 2\lambda\langle X, N\rangle - |X^N|^2,$$

where X^N is the normal part of X, we have

$$\begin{aligned}\Delta\rho &= -\tfrac{1}{4}\rho\Delta|X|^2 + \tfrac{1}{16}\rho\big|\nabla|X|^2\big|^2 = -\tfrac{1}{4}\rho\big(2n + 2\lambda\langle X, N\rangle - |X^N|^2\big) + \tfrac{1}{4}\rho|X^T|^2 \\ &= -\tfrac{1}{2}n\rho - \tfrac{1}{2}\lambda\rho\langle X, N\rangle + \tfrac{1}{4}\rho|X|^2.\end{aligned}$$

From (4), we get

$$\begin{aligned}\int_M |\nabla f|^2 = \int_M \big|\nabla(|A|^{n/2}\eta)\big|^2\rho &- \frac{1}{16}\int_M |A|^n\eta^2|X^T|^2\rho - \frac{1}{8}\int_M |A|^n\eta^2|X^N|^2\rho \\ &+ \frac{n}{4}\int_M |A|^n\eta^2\rho + \frac{1}{4}\int_M |A|^n\eta^2\lambda\langle X, N\rangle\rho.\end{aligned} \tag{5}$$

Combining the Sobolev inequality in Lemma 6 and (5), we have

$$\begin{aligned}\left(\int_M |f|^{\frac{2n}{n-2}}\right)^{\frac{n-2}{n}} &\leq D^2(n)\cdot\left[\frac{4(n-1)^2(1+s)}{(n-2)^2}\int_M |\nabla f|^2 + \left(1+\frac{1}{s}\right)\cdot\frac{1}{n^2}\int_M H^2 f^2\right] \\ &= \frac{4D^2(n)(n-1)^2(1+s)}{(n-2)^2}\left[\int_M \big|\nabla(|A|^{n/2}\eta)\big|^2\rho - \frac{1}{16}\int_M |A|^n\eta^2|X^T|^2\rho \right.\\ &\qquad - \frac{1}{8}\int_M |A|^n\eta^2|X^N|^2\rho + \frac{n}{4}\int_M |A|^n\eta^2\rho \\ &\qquad \left. + \frac{1}{4}\int_M |A|^n\eta^2\lambda\langle X, N\rangle\rho\right] \\ &\quad + D^2(n)\left(1+\frac{1}{s}\right)\cdot\frac{1}{n^2}\int_M |A|^n\eta^2\big(\lambda - \tfrac{1}{2}\langle X, N\rangle\big)^2\rho.\end{aligned}$$

We choose

$$s=\frac{(n-2)^2}{2n^2(n-1)^2}\in\mathbb{R}^+$$

such that

$$\frac{4(n-1)^2(1+s)}{(n-2)^2}\cdot\frac{1}{8}=\frac{1}{4}\left(1+\frac{1}{s}\right)\cdot\frac{1}{n^2}.$$

Hence

$$\begin{aligned}\left(\int_M|f|^{\frac{2n}{n-2}}\right)^{\frac{n-2}{n}}&\\ \leq\frac{2D^2(n)[(n-2)^2+2n^2(n-1)^2]}{n^2(n-2)^2}&\left[\int_M\left|\nabla(|A|^{n/2}\eta)\right|^2\rho\right.\\ &\left.+\frac{n}{4}\int_M|A|^n\eta^2\rho+\frac{1}{4}\int_M|A|^n\eta^2\lambda\langle X,N\rangle\rho\right]\\ +\frac{D^2(n)[(n-2)^2+2n^2(n-1)^2]}{n^2(n-2)^2}&\left[\int_M\lambda^2|A|^n\eta^2\rho-\int_M|A|^n\eta^2\lambda\langle X,N\rangle\rho\right].\end{aligned}$$

Now we put

$$\kappa=\frac{2D^2(n)[(n-2)^2+2n^2(n-1)^2]}{n^2(n-2)^2}.$$

It follows from the inequality above that

$$\begin{aligned}(6)\quad\kappa^{-1}&\left(\int_M|f|^{\frac{2n}{n-2}}\right)^{\frac{n-2}{n}}\\ &\leq\int_M\left|\nabla(|A|^{n/2}\eta)\right|^2\rho+\frac{n}{4}\int_M|A|^n\eta^2\rho+\frac{1}{4}\int_M|A|^n\eta^2\lambda\langle X,N\rangle\rho\\ &\quad+\frac{1}{2}\left(\int_M\lambda^2|A|^n\eta^2\rho-\int_M|A|^n\eta^2\lambda\langle X,N\rangle\rho\right)\\ &=\int_M\left|\nabla(|A|^{n/2}\eta)\right|^2\rho+\left(\frac{n+2\lambda^2}{4}\right)\int_M|A|^n\eta^2\rho-\frac{1}{4}\int_M|A|^n\eta^2\lambda\langle X,N\rangle\rho\\ &=\int_M\left(\frac{n^2}{4}\left|\nabla|A|\right|^2|A|^{n-2}\eta^2+n|A|^{n-1}\eta\nabla|A|\cdot\nabla\eta+|A|^n|\nabla\eta|^2\right)\rho\\ &\quad+\left(\frac{n+2\lambda^2}{4}\right)\int_M|A|^n\eta^2\rho-\frac{1}{4}\int_M|A|^n\eta^2\lambda\langle X,N\rangle\rho.\end{aligned}$$

On the other hand, for any $\theta > 0$, we have

$$
\begin{aligned}
(7)\quad -\frac{1}{2}\int_M |A|^n \eta^2 \lambda\langle X, N\rangle\rho &= -\int_M |A|^n\eta^2\lambda(\lambda - H)\rho \\
&= -\int_M |A|^n\eta^2\lambda^2\rho + \int_M |A|^n\eta^2\lambda H\rho \\
&\le -\lambda^2\int_M |A|^n\eta^2\rho + |\lambda|\int_M |A|^n\eta^2\left(\frac{\theta}{2}H^2 + \frac{1}{2\theta}\right)\rho \\
&\le \left(\frac{|\lambda|}{2\theta} - \lambda^2\right)\int_M |A|^n\eta^2\rho + \frac{|\lambda|\theta}{2}\int_M |A|^n\eta^2 H^2\rho \\
&\le \left(\frac{|\lambda|}{2\theta} - \lambda^2\right)\int_M |A|^n\eta^2\rho + \frac{n|\lambda|\theta}{2}\int_M |A|^{n+2}\eta^2\rho.
\end{aligned}
$$

Combining (6) and (7), we get

$$
\begin{aligned}
(8)\quad & \kappa^{-1}\left(\int_M |f|^{\frac{2n}{n-2}}\right)^{\frac{n-2}{n}} \\
&\le \int_M \left(\frac{n^2}{4}\big|\nabla|A|\big|^2|A|^{n-2}\eta^2 + n|A|^{n-1}\eta\nabla|A|\cdot\nabla\eta + |A|^n|\nabla\eta|^2\right)\rho \\
&\quad + \left(\frac{|\lambda|}{4\theta} + \frac{n}{4}\right)\int_M |A|^n\eta^2\rho + \frac{n\theta|\lambda|}{4}\int_M |A|^{n+2}\eta^2\rho.
\end{aligned}
$$

Combining the Cauchy inequality, (3) and (8), we have for any $\delta > 0$

$$
\begin{aligned}
&\kappa^{-1}\left(\int_M |f|^{\frac{2n}{n-2}}\right)^{\frac{n-2}{n}} \\
&\le (1+\delta)\frac{n^2}{4}\int_M \big|\nabla|A|\big|^2|A|^{n-2}\eta^2\rho + \left(1+\frac{1}{\delta}\right)\int_M |A|^n|\nabla\eta|^2\rho \\
&\quad + \left(\frac{|\lambda|}{4\theta} + \frac{n}{4}\right)\int_M |A|^n\eta^2\rho + \frac{n\theta|\lambda|}{4}\int_M |A|^{n+2}\eta^2\rho \\
&\le \frac{(1+\delta)n^2}{8(n-1-\varepsilon)}\Bigg[\left(\frac{|\lambda|}{\tau} + 2\right)\int_M |A|^{n+2}\eta^2\rho \\
&\qquad + (|\lambda|\tau - 1)\int_M |A|^n\eta^2\rho + \frac{2}{\varepsilon}\int_M |A|^n|\nabla\eta|^2\rho\Bigg] \\
&\quad + \left(1+\frac{1}{\delta}\right)\int_M |A|^n|\nabla\eta|^2\rho + \left(\frac{|\lambda|}{4\theta} + \frac{n}{4}\right)\int_M |A|^n\eta^2\rho + \frac{n\theta|\lambda|}{4}\int_M |A|^{n+2}\eta^2\rho.
\end{aligned}
$$

Put

$$
\delta = \frac{2(|\lambda| + n\theta)(n-1+\varepsilon)}{(1-|\lambda|\tau)\theta n^2} - 1 > 0,
$$

where $\varepsilon, \theta, \tau$ are positive constants such that $|\lambda|\tau - 1 < 0$. Then

$$\begin{aligned}
(9)\qquad & \kappa^{-1}\Big(\int_M |f|^{\frac{2n}{n-2}}\Big)^{\frac{n-2}{n}} \\
&\le \Big[\frac{n\theta+|\lambda|}{4\theta(1-|\lambda|\tau)}\cdot\Big(\frac{|\lambda|}{\tau}+2\Big)\frac{n-1+\varepsilon}{n-1-\varepsilon}+\frac{n\theta|\lambda|}{4}\Big]\int_M |A|^{n+2}\eta^2\rho \\
&\quad + \Big[\frac{n\theta+|\lambda|}{2\theta\varepsilon(1-|\lambda|\tau)}\frac{n-1+\varepsilon}{n-1-\varepsilon}+1+\frac{1}{\delta}\Big]\int_M |A|^n|\nabla\eta|^2\rho \\
&\le \frac{(n\theta+|\lambda|)(|\lambda|+2\tau)+n\tau\theta^2(1-|\lambda|\tau)|\lambda|}{4\tau\theta(1-|\lambda|\tau)}\cdot\frac{n-1+\varepsilon}{n-1-\varepsilon} \\
&\qquad \times\Big(\int_M |A|^{2\cdot\frac{n}{2}}\Big)^{\frac{2}{n}}\cdot\Big(\int_M (|A|^n\eta^2\rho)^{\frac{n}{n-2}}\Big)^{\frac{n-2}{n}} \\
&\quad + \Big(\frac{n\theta+|\lambda|}{2\theta\varepsilon(1-|\lambda|\tau)}\cdot\frac{n-1+\varepsilon}{n-1-\varepsilon}+1+\frac{1}{\delta}\Big)\int_M |A|^n|\nabla\eta|^2\rho.
\end{aligned}$$

Set

$$K(n,\lambda,\theta,\tau)=\sqrt{\frac{4\tau\theta(1-|\lambda|\tau)}{\big[(n\theta+|\lambda|)(|\lambda|+2\tau)+n\tau\theta^2(1-|\lambda|\tau)|\lambda|\big]\kappa}}.$$

By a direct computation, $K(n,\lambda,\theta,\tau)$ achieves its maximum

$$K(n,\lambda)=\sqrt{\frac{2(\sqrt{\lambda^2+2}-|\lambda|)}{\big(n|\lambda|+2\sqrt{n}|\lambda|+n\sqrt{\lambda^2+2}\big)\kappa}}$$

when

$$\tau=\tfrac{1}{2}\big(\sqrt{\lambda^2+2}-|\lambda|\big),\quad \theta=\sqrt{\frac{|\lambda|+2\tau}{n\tau-n|\lambda|\tau^2}}=\frac{2}{\sqrt{n}\big(\sqrt{\lambda^2+2}-|\lambda|\big)}=\frac{1}{\sqrt{n}\tau}.$$

Since

$$\Big(\int_M |A|^n\,\mathrm{d}\mu\Big)^{1/n} < K(n,\lambda),$$

we have from (9) that there exists $0<\varepsilon_0<1$ such that

$$\kappa^{-1}\Big(\int_M |f|^{\frac{2n}{n-2}}\Big)^{\frac{n-2}{n}} \le \frac{n-1+\varepsilon}{n-1-\varepsilon}\cdot\frac{1-\varepsilon_0}{\kappa}\Big(\int_M |f|^{\frac{2n}{n-2}}\Big)^{\frac{n-2}{n}}+C(\varepsilon,\lambda)\int_M |A|^n|\nabla\eta|^2\rho,$$

namely,

$$(10)\qquad \frac{(n-1+\varepsilon)\varepsilon_0-2\varepsilon}{(n-1-\varepsilon)\kappa}\Big(\int_M |f|^{\frac{2n}{n-2}}\Big)^{\frac{n-2}{n}} \le C(\varepsilon,\lambda)\int_M |A|^n|\nabla\eta|^2\rho.$$

Let $\eta(X) = \eta_r(X) = \phi(|X|/r)$ for any $r > 0$, where ϕ is a nonnegative function on $[0, +\infty)$ satisfying

$$\phi(x) = \begin{cases} 1 & \text{if } x \in [0, 1), \\ 0 & \text{if } x \in [2, +\infty), \end{cases}$$

and $|\phi'| \le C$ for some absolute constant. Let $\varepsilon = \frac{1}{2}\varepsilon_0$. Since $\int_M |A|^n \, d\mu$ is bounded, the right-hand side of (10) approaches zero as $r \to +\infty$, which implies $|A| \equiv 0$. Hence M is a hyperplane of $\mathbb{R}^{n+1}$. This completes the proof of Theorem 1. □

Setting $\mathring{A} = \sum_{i,j} \mathring{h}_{ij} \omega_i \otimes \omega_j$, we have $\mathring{h}_{ij} = h_{ij} - (H/n) g_{ij}$. Choose $\{e_i\}$ such that $h_{ij} = \lambda_i \delta_{ij}$ at a point p. Then $\mathring{h}_{ij} = \mathring{\lambda}_i \delta_{ij}$, where $\mathring{\lambda}_i = \lambda_i - H/n$, and

$$f_3 = \sum_i \lambda_i^3 = \sum_i \left(\mathring{\lambda}_i + \frac{H}{n}\right)^3 = B_3 + \frac{3}{n} H |\mathring{A}|^2 + \frac{1}{n^2} H^3,$$

where $|\mathring{A}|^2 = \sum_i \mathring{\lambda}_i^2 = |A|^2 - H^2/n$ and $B_3 = \sum_i \mathring{\lambda}_i^3$. Thus, from (1) and (2) we have

$$\begin{aligned} \tfrac{1}{2}\mathcal{L}|\mathring{A}|^2 &= \tfrac{1}{2}\mathcal{L}|A|^2 - \tfrac{1}{2}\mathcal{L}\left(\frac{H^2}{n}\right) \\ &= |\nabla A|^2 - \frac{1}{n}|\nabla H|^2 + \left(\tfrac{1}{2} - |A|^2\right)|A|^2 + \lambda f_3 - \frac{H^2}{2n} - |A|^2(\lambda - H)\frac{H}{n} \\ &= |\nabla \mathring{A}|^2 + \left(\tfrac{1}{2} - |\mathring{A}|^2\right)|\mathring{A}|^2 - \frac{1}{n} H^2 |\mathring{A}|^2 + \lambda B_3 + \frac{2}{n}\lambda H |\mathring{A}|^2. \end{aligned}$$

By using an algebraic inequality in [Okumura 1974], we have

$$|B_3| \le \frac{n-2}{\sqrt{n(n-1)}} |\mathring{A}|^3,$$

and the equality holds if and only if at least $n-1$ of the $\mathring{\lambda}_i$ are equal. Then we get

(11)
$$\begin{aligned} \tfrac{1}{2}\mathcal{L}|\mathring{A}|^2 &\ge |\nabla \mathring{A}|^2 + \left(\tfrac{1}{2} - |\mathring{A}|^2\right)|\mathring{A}|^2 - \frac{1}{n} H^2 |\mathring{A}|^2 - |\lambda| \frac{n-2}{\sqrt{n(n-1)}} |\mathring{A}|^3 + \frac{2}{n}\lambda H |\mathring{A}|^2 \\ &\ge \bigl|\nabla |\mathring{A}|\bigr|^2 + \left(\tfrac{1}{2} - |\mathring{A}|^2\right)|\mathring{A}|^2 - \frac{1}{n}\left(\lambda - \tfrac{1}{2}\langle X, N\rangle\right)^2 |\mathring{A}|^2 \\ &\qquad - |\lambda| \frac{n-2}{\sqrt{n(n-1)}} |\mathring{A}|^3 + \frac{2}{n}\lambda\left(\lambda - \tfrac{1}{2}\langle X, N\rangle\right)|\mathring{A}|^2 \\ &= \bigl|\nabla |\mathring{A}|\bigr|^2 + \left(\frac{1}{2} + \frac{\lambda^2}{n}\right)|\mathring{A}|^2 - \frac{1}{4n}|\mathring{A}|^2 |X^N|^2 - |\lambda| \frac{n-2}{\sqrt{n(n-1)}} |\mathring{A}|^3 - |\mathring{A}|^4. \end{aligned}$$

By using (11), we give the proof of Theorem 2 as follows.

Proof of Theorem 2. Let η be a smooth function with compact support on M. Multiplying $|\mathring{A}|^{n-2}\eta^2$ on both sides of the inequality (11) above and integrating by parts with respect to the measure $\rho\,\mathrm{d}\mu$ on M yields

$$\begin{aligned}
0 &\geq 2\int_M \big|\nabla|\mathring{A}|\big|^2|\mathring{A}|^{n-2}\eta^2\rho+\Big(1+\frac{2\lambda^2}{n}\Big)\int_M|\mathring{A}|^n\eta^2\rho-\frac{1}{2n}\int_M|\mathring{A}|^n|X^N|^2\eta^2\rho\\
&\quad-2|\lambda|\frac{n-2}{\sqrt{n(n-1)}}\int_M|\mathring{A}|^{n+1}\eta^2\rho-2\int_M|\mathring{A}|^{n+2}\eta^2\rho-\int_M|\mathring{A}|^{n-2}\eta^2\mathcal{L}|\mathring{A}|^2\rho\\
&=2\int_M \big|\nabla|\mathring{A}|\big|^2|\mathring{A}|^{n-2}\eta^2\rho+\Big(1+\frac{2\lambda^2}{n}\Big)\int_M|\mathring{A}|^n\eta^2\rho-\frac{1}{2n}\int_M|\mathring{A}|^n|X^N|^2\eta^2\rho\\
&\quad-2|\lambda|\frac{n-2}{\sqrt{n(n-1)}}\int_M|\mathring{A}|^{n+1}\eta^2\rho-2\int_M|\mathring{A}|^{n+2}\eta^2\rho\\
&\quad+2\int_M\rho|\mathring{A}|\nabla|\mathring{A}|\cdot\nabla(|\mathring{A}|^{n-2}\eta^2)\\
&\geq 2(n-1)\int_M \big|\nabla|\mathring{A}|\big|^2|\mathring{A}|^{n-2}\eta^2\rho+\Big[\Big(1+\frac{2\lambda^2}{n}\Big)-|\lambda|\zeta\frac{n-2}{\sqrt{n(n-1)}}\Big]\int_M|\mathring{A}|^n\eta^2\rho\\
&\quad-\Big(2+\frac{|\lambda|}{\zeta}\frac{n-2}{\sqrt{n(n-1)}}\Big)\int_M|\mathring{A}|^{n+2}\eta^2\rho-\frac{1}{2n}\int_M|\mathring{A}|^n|X^N|^2\eta^2\rho\\
&\quad+4\int_M(\nabla|\mathring{A}|\cdot\nabla\eta)|\mathring{A}|^{n-1}\eta\rho
\end{aligned}$$

with constant $\zeta>0$.

From the assumption $|H|\leq\sqrt{\frac{1}{3}n+\lambda^2}-|\lambda|\triangleq C$, we have

$$\int_M|\mathring{A}|^n|X^N|^2\eta^2\rho=4\int_M|\mathring{A}|^n(\lambda-H)^2\eta^2\rho\leq 4(\lambda^2+C^2+2C|\lambda|)\int_M|\mathring{A}|^n\eta^2\rho.$$

This implies

$$\begin{aligned}
0 &\geq 2(n-1)\int_M \big|\nabla|\mathring{A}|\big|^2|\mathring{A}|^{n-2}\eta^2\rho\\
&\quad+\Big[\Big(1+\frac{2\lambda^2}{n}\Big)-|\lambda|\zeta\frac{n-2}{\sqrt{n(n-1)}}-\frac{2}{n}(\lambda^2+C^2+2C|\lambda|)\Big]\int_M|\mathring{A}|^n\eta^2\rho\\
&\quad-\Big(2+\frac{|\lambda|}{\zeta}\frac{n-2}{\sqrt{n(n-1)}}\Big)\int_M|\mathring{A}|^{n+2}\eta^2\rho+4\int_M(\nabla|\mathring{A}|\cdot\nabla\eta)|\mathring{A}|^{n-1}\eta\rho.
\end{aligned}$$

By using the Cauchy inequality, for any $\varepsilon>0$ we obtain

$$\begin{aligned}
&\Big(\frac{|\lambda|}{\zeta}\frac{n-2}{\sqrt{n(n-1)}}+2\Big)\int_M|\mathring{A}|^{n+2}\eta^2\rho\\
&\quad+\Big[|\lambda|\zeta\frac{n-2}{\sqrt{n(n-1)}}+\frac{2}{n}(C^2+2C|\lambda|)-1\Big]\int_M|\mathring{A}|^n\eta^2\rho+\frac{2}{\varepsilon}\int_M|\mathring{A}|^n|\nabla\eta|^2\rho\\
&\qquad\qquad\geq 2(n-1-\varepsilon)\int_M\big|\nabla|\mathring{A}|\big|^2|\mathring{A}|^{n-2}\eta^2\rho.
\end{aligned}\tag{12}$$

Set $f = |\mathring{A}|^{n/2}\rho^{1/2}\eta$. Using the same argument as in the proof of Theorem 1, for any $\delta > 0$ we get

$$\kappa^{-1}\left(\int_M |f|^{\frac{2n}{n-2}}\right)^{\frac{n-2}{n}} \le (1+\delta)\frac{n^2}{4}\int_M \big|\nabla|\mathring{A}|\big|^2|\mathring{A}|^{n-2}\eta^2\rho + \left(1+\frac{1}{\delta}\right)\int_M |\mathring{A}|^n|\nabla\eta|^2\rho + \frac{n+2\lambda^2}{4}\int_M |\mathring{A}|^n\eta^2\rho - \frac{1}{4}\int_M |\mathring{A}|^n\eta^2\lambda\langle X, N\rangle\rho. \tag{13}$$

It is easy to see that

$$\begin{aligned} -\int_M |\mathring{A}|^n\eta^2\lambda\langle X, N\rangle\rho &= -2\int_M |\mathring{A}|^n\eta^2\lambda(\lambda - H)\rho \\ &= -2\int_M |\mathring{A}|^n\eta^2\lambda^2\rho + 2\int_M |\mathring{A}|^n\eta^2\lambda H\rho \\ &\le 2(C|\lambda| - \lambda^2)\int_M |\mathring{A}|^n\eta^2\rho. \end{aligned} \tag{14}$$

Combining (12), (13) and (14), we have

$$\begin{aligned} &\kappa^{-1}\left(\int_M |f|^{\frac{2n}{n-2}}\right)^{\frac{n-2}{n}} \\ &\quad\le (1+\delta)\frac{n^2}{4}\int_M \big|\nabla|\mathring{A}|\big|^2|\mathring{A}|^{n-2}\eta^2\rho + \left(1+\frac{1}{\delta}\right)\int_M |\mathring{A}|^n|\nabla\eta|^2\rho \\ &\qquad + \frac{n+2C|\lambda|}{4}\int_M |\mathring{A}|^n\eta^2\rho \\ &\quad\le \frac{(1+\delta)n^2}{8(n-1-\varepsilon)}\left\{\left(\frac{|\lambda|}{\zeta}\frac{n-2}{\sqrt{n(n-1)}} + 2\right)\int_M |\mathring{A}|^{n+2}\eta^2\rho \right. \\ &\qquad + \left[|\lambda|\zeta\frac{n-2}{\sqrt{n(n-1)}} + \frac{2}{n}(C^2 + 2C|\lambda|) - 1\right]\int_M |\mathring{A}|^n\eta^2\rho \\ &\qquad \left. + \frac{2}{\varepsilon}\int_M |\mathring{A}|^n|\nabla\eta|^2\rho\right\} + \left(1+\frac{1}{\delta}\right)\int_M |\mathring{A}|^n|\nabla\eta|^2\rho + \frac{n+2C|\lambda|}{4}\int_M |\mathring{A}|^n\eta^2\rho. \end{aligned}$$

Let

$$|\lambda|\zeta\frac{n-2}{\sqrt{n(n-1)}} + \frac{2}{n}(C^2 + 2C|\lambda|) - 1 < 0,$$

i.e.,

$$0 < \zeta < \frac{\big[n - 2(C^2 + 2C|\lambda|)\big]\sqrt{n(n-1)}}{n(n-2)|\lambda|}.$$

Putting

$$\delta = \frac{2(n+2C|\lambda|)\sqrt{n-1}\cdot(n-1+\varepsilon)}{n\big[n\sqrt{n-1}-(n-2)\sqrt{n}\,|\lambda|\,\zeta-2\sqrt{n-1}(C^2+2C|\lambda|)\big]}-1>0$$

for some $\varepsilon > 0$ to be defined later, we have

$$\begin{aligned}
(15)\quad & \kappa^{-1}\left(\int_M |f|^{\frac{2n}{n-2}}\right)^{\frac{n-2}{n}} \\
&\le \left\{\frac{\sqrt{n}(n+2C|\lambda|)[(n-2)|\lambda|+2\zeta\sqrt{n(n-1)}]}{4\zeta[n\sqrt{n-1}-(n-2)\sqrt{n}|\lambda|\zeta-2\sqrt{n-1}(C^2+2C|\lambda|)]}\right\} \\
&\qquad\qquad \times\frac{n-1+\varepsilon}{n-1-\varepsilon}\int_M |\mathring{A}|^{n+2}\eta^2\rho \\
&\quad +\left\{\frac{n(n+2C|\lambda|)\sqrt{n-1}\cdot(n-1+\varepsilon)}{2\varepsilon[n\sqrt{n-1}-(n-2)\sqrt{n}|\lambda|\zeta-2\sqrt{n-1}(C^2+2C|\lambda|)]}+1+\frac{1}{\delta}\right\} \\
&\qquad\qquad \times\int_M |\mathring{A}|^n|\nabla\eta|^2\rho \\
&\le \left\{\frac{\sqrt{n}(n+2C|\lambda|)[(n-2)|\lambda|+2\zeta\sqrt{n(n-1)}]}{4\zeta[n\sqrt{n-1}-(n-2)\sqrt{n}|\lambda|\zeta-2\sqrt{n-1}(C^2+2C|\lambda|)]}\right\} \\
&\qquad\qquad \times\frac{n-1+\varepsilon}{n-1-\varepsilon}\cdot\left(\int_M |\mathring{A}|^{2\cdot\frac{n}{2}}\right)^{\frac{2}{n}}\left(\int_M (|\mathring{A}|^n\eta^2\rho)^{\frac{n}{n-2}}\right)^{\frac{n-2}{n}} \\
&\quad +\widetilde{C}(\varepsilon,\lambda,n,\zeta,C)\int_M |\mathring{A}|^n|\nabla\eta|^2\rho.
\end{aligned}$$

Set

$$D(n,\lambda,\zeta,C)=\sqrt{\frac{4\zeta\big[n\sqrt{n-1}-(n-2)\sqrt{n}|\lambda|\zeta-2\sqrt{n-1}(C^2+2C|\lambda|)\big]}{\sqrt{n}(n+2C|\lambda|)\big[(n-2)|\lambda|+2\zeta\sqrt{n(n-1)}\big]\kappa}}.$$

We choose

$$\zeta=\frac{\sqrt{(n-2)^2\lambda^2+2(n-1)\big[n-2(C^2+2C|\lambda|)\big]}-(n-2)|\lambda|}{2\sqrt{n(n-1)}}$$

such that $D(n,\lambda,\zeta,C)$ achieves its maximum $D(n,\lambda)$ with

$$\begin{aligned}
D(n,\lambda)&=\frac{\sqrt{(n-2)^2\lambda^2+2(n-1)\big[n-2(C^2+2C|\lambda|)\big]}-(n-2)|\lambda|}{\sqrt{n(n-1)(n+2C|\lambda|)\kappa}} \\
&=\frac{\sqrt{(n-2)^2\lambda^2+\frac{2}{3}n(n-1)}-(n-2)|\lambda|}{\sqrt{n(n-1)\Big(n+2|\lambda|\sqrt{\frac{1}{3}n+\lambda^2}-2\lambda^2\Big)\kappa}}.
\end{aligned}$$

Combining the assumption

$$\left(\int_M |\mathring{A}|^n \, \mathrm{d}\mu\right)^{1/n} < D(n,\lambda)$$

and (15) implies that there exists $0 < \varepsilon_0 < 1$ such that

$$\kappa^{-1}\left(\int_M |f|^{\frac{2n}{n-2}}\right)^{\frac{n-2}{n}} \le \frac{n-1+\varepsilon}{n-1-\varepsilon}\cdot\frac{1-\varepsilon_0}{\kappa}\left(\int_M |f|^{\frac{2n}{n-2}}\right)^{\frac{n-2}{n}} + \widetilde{C}(\varepsilon,\lambda,n)\int_M |\mathring{A}|^n |\nabla\eta|^2\rho,$$

namely,

$$\frac{(n-1+\varepsilon)\varepsilon_0 - 2\varepsilon}{(n-1-\varepsilon)\kappa}\left(\int_M |f|^{\frac{2n}{n-2}}\right)^{\frac{n-2}{n}} \le \widetilde{C}(\varepsilon,\lambda,n)\int_M |\mathring{A}|^n |\nabla\eta|^2\rho.$$

Let $\varepsilon = \frac{1}{2}\varepsilon_0$ and choose η as in the proof of Theorem 1. Since $\int_M |\mathring{A}|^n \, \mathrm{d}\mu$ is bounded, by using a similar argument we obtain $\mathring{A} \equiv 0$. Therefore, M is totally umbilical, i.e., M is $\mathbb{S}^n(\sqrt{\lambda^2+2n}-\lambda)$ or $\mathbb{R}^n$. Since we have assumed that

$$|H| \le \sqrt{\tfrac{1}{3}n+\lambda^2} - |\lambda|,$$

the first case is excluded. This completes the proof of Theorem 2. □

Remark. In fact, we can prove that if $\sup_M |H| < \sqrt{\frac{1}{2}n+\lambda^2} - |\lambda|$ and if

$$\left(\int_M |\mathring{A}|^n \, \mathrm{d}\mu\right)^{1/n} < D(n,\lambda,\sup_M |H|),$$

then M is a hyperplane. Here $D(n,\lambda,\sup_M |H|)$ is a positive constant depending on n, λ and $\sup_M |H|$.

Remark. In particular, if $\lambda = 0$, Theorem 2 reduces to the rigidity result for self-shrinkers in [Lin 2016]. For the higher codimension case, Cao, Xu and Zhao [Cao et al. 2014] proved some L^n-pinching theorems of $\mathring{A}$ for self-shrinkers.

4. Gap theorems in dimension 2

We need another Sobolev-type inequality in dimension 2, which was proved by Xu and Gu [2007b]:

$$\tilde{c}^{-1}\left(\int f^4 \, \mathrm{d}\mu\right)^{1/2} \le \frac{1}{t}\int |\nabla f|^2 \, \mathrm{d}\mu + t\int f^2 \, \mathrm{d}\mu + \frac{1}{2}\int |H| f^2 \, \mathrm{d}\mu \tag{16}$$

for all $f \in C_c^\infty(M)$ and for all $t \in \mathbb{R}^+$, where $\tilde{c} = 12\sqrt{3\pi}/\pi$.

Proof of Theorem 3. As in the proof of Theorem 1, for any $0 < \varepsilon < 1$, we have

$$\left(\frac{|\lambda|}{\tau}+2\right)\int_M |A|^4\eta^2\rho + (|\lambda|\tau - 1)\int_M |A|^2\eta^2\rho + \frac{2}{\varepsilon}\int_M |A|^2|\nabla\eta|^2\rho \geq 2(1-\varepsilon)\int_M \big|\nabla|A|\big|^2\eta^2\rho. \tag{17}$$

Setting $f = |A|\eta\rho^{1/2}$, we get

$$\int_M |\nabla f|^2 = \int_M \big|\nabla(|A|\eta)\big|^2\rho - \frac{1}{16}\int_M |A|^2\eta^2|X^T|^2\rho - \frac{1}{8}\int_M |A|^2\eta^2|X^N|^2\rho + \frac{1}{2}\int_M |A|^2\eta^2\rho + \frac{1}{4}\int_M |A|^2\eta^2\lambda\langle X, N\rangle\rho. \tag{18}$$

Combining the Sobolev-type inequality (16) and (18), we have

$$\begin{aligned}
\tilde{c}^{-1}&\left(\int_M |f|^4\right)^{1/2} \\
&\leq \frac{1}{t}\bigg[\int_M \big|\nabla(|A|\eta)\big|^2\rho - \frac{1}{16}\int_M |A|^2\eta^2|X^T|^2\rho \\
&\qquad - \frac{1}{8}\int_M |A|^2\eta^2|X^N|^2\rho + \frac{1}{2}\int_M |A|^2\eta^2\rho + \frac{1}{4}\int_M |A|^2\eta^2\lambda\langle X, N\rangle\rho\bigg] \\
&\qquad + t\int_M |A|^2\eta^2\rho + \frac{1}{2}\int |H||A|^2\eta^2\rho \\
&\leq \frac{1}{t}\bigg[\int_M \big|\nabla(|A|\eta)\big|^2\rho + \frac{1}{2}\int_M |A|^2\eta^2\rho + \frac{\lambda^2}{8}\int_M |A|^2\eta^2\rho\bigg] \\
&\qquad + t\int_M |A|^2\eta^2\rho + \frac{1}{2}\int |H||A|^2\eta^2\rho.
\end{aligned}$$

By the Cauchy inequality, for any $\theta > 0$, we get

$$\begin{aligned}
\tilde{c}^{-1}\left(\int_M |f|^4\right)^{1/2} &\leq \frac{1}{t}\bigg[\int_M \big|\nabla(|A|\eta)\big|^2\rho + \frac{1}{2}\int_M |A|^2\eta^2\rho + \frac{\lambda^2}{8}\int_M |A|^2\eta^2\rho\bigg] \\
&\qquad + t\int_M |A|^2\eta^2\rho + \frac{1}{2}\int\left(\frac{\theta}{2}H^2 + \frac{1}{2\theta}\right)|A|^2\eta^2\rho \\
&\leq \frac{1}{t}\bigg[\int_M \big|\nabla(|A|\eta)\big|^2\rho + \frac{1}{2}\int_M |A|^2\eta^2\rho + \frac{\lambda^2}{8}\int_M |A|^2\eta^2\rho\bigg] \\
&\qquad + t\int_M |A|^2\eta^2\rho + \frac{1}{2}\int\left(\theta|A|^2 + \frac{1}{2\theta}\right)|A|^2\eta^2\rho
\end{aligned} \tag{19}$$

$$
\begin{aligned}
&= \frac{1}{t}\int_M \left(\left|\nabla|A|\right|^2\eta^2 + 2|A|\eta\nabla|A|\cdot\nabla\eta + |A|^2|\nabla\eta|^2\right)\rho \\
&\qquad + \left(t + \frac{1}{4\theta} + \frac{1}{2t} + \frac{\lambda^2}{8t}\right)\int |A|^2\eta^2\rho + \frac{\theta}{2}\int |A|^4\eta^2\rho \\
&\le \frac{1}{t}\left[(1+\delta)\int_M \left|\nabla|A|\right|^2\eta^2\rho + \left(1+\frac{1}{\delta}\right)\int_M |A|^2|\nabla\eta|^2\rho\right] \\
&\qquad + \left(t + \frac{1}{4\theta} + \frac{1}{2t} + \frac{\lambda^2}{8t}\right)\int |A|^2\eta^2\rho + \frac{\theta}{2}\int |A|^4\eta^2\rho.
\end{aligned}
$$

Combining (17) and (19), we have

$$
\begin{aligned}
&\tilde{c}^{-1}\left(\int_M |A|^4\eta^4\rho^2\right)^{1/2} \\
&\le \frac{1}{t}(1+\delta)\cdot\frac{1}{2(1-\varepsilon)}\left[\left(\frac{|\lambda|}{\tau}+2\right)\int_M |A|^4\eta^2\rho + (|\lambda|\tau-1)\int_M |A|^2\eta^2\rho \right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left. + \frac{2}{\varepsilon}\int_M |A|^2|\nabla\eta|^2\rho\right] \\
&\quad + \frac{1}{t}\left(1+\frac{1}{\delta}\right)\int_M |A|^2|\nabla\eta|^2\rho + \left(t+\frac{1}{4\theta}+\frac{1}{2t}+\frac{\lambda^2}{8t}\right)\int |A|^2\eta^2\rho + \frac{\theta}{2}\int |A|^4\eta^2\rho.
\end{aligned}
$$

Put

$$
\delta = \frac{(4\theta + 2t + 8\theta t^2 + \theta\lambda^2)(1+\varepsilon)}{4\theta(1-|\lambda|\tau)} - 1 > 0,
$$

where $\varepsilon, \theta, \tau, t$ are positive constants such that $|\lambda|\tau - 1 < 0$. Then

$$
\begin{aligned}
(20)\quad &\tilde{c}^{-1}\left(\int_M |A|^4\eta^4\rho^2\right)^{1/2} \\
&\le \left[\frac{1}{t}\cdot\frac{(1+\varepsilon)}{2(1-\varepsilon)}\cdot\frac{(4\theta+2t+8\theta t^2+\theta\lambda^2)}{4\theta(1-|\lambda|\tau)}\cdot\left(\frac{|\lambda|}{\tau}+2\right)+\frac{\theta}{2}\right]\int_M |A|^4\eta^2\rho \\
&\qquad + \frac{1}{t}\left[\frac{(1+\delta)}{\varepsilon(1-\varepsilon)} + \left(1+\frac{1}{\delta}\right)\right]\int_M |A|^2|\nabla\eta|^2\rho \\
&\le \frac{(4\theta+2t+8\theta t^2+\theta\lambda^2)(|\lambda|+2\tau)+4\theta^2 t\tau(1-|\lambda|\tau)}{8\theta t\tau(1-|\lambda|\tau)}\cdot\frac{1+\varepsilon}{1-\varepsilon} \\
&\qquad\qquad\qquad\qquad \times\left(\int_M |A|^4\eta^4\rho^2\right)^{1/2}\cdot\left(\int_M |A|^4\right)^{1/2} \\
&\qquad + \frac{1}{t}\left[\frac{(1+\delta)}{\varepsilon(1-\varepsilon)} + \left(1+\frac{1}{\delta}\right)\right]\int_M |A|^2|\nabla\eta|^2\rho.
\end{aligned}
$$

Set

$$K(t,\lambda,\theta,\tau)=\frac{8\theta t\tau(1-|\lambda|\tau)}{\big[(4\theta+2t+8\theta t^2+\theta\lambda^2)(|\lambda|+2\tau)+4\theta^2 t\tau(1-|\lambda|\tau)\big]\tilde{c}},$$

where $\tilde{c}=12\sqrt{3\pi}/\pi$. By a direct computation, $K(t,\lambda,\theta,\tau)$ achieves its maximum

$$K(\lambda)=\frac{\sqrt{2}\big(\lambda^2+1-\sqrt{\lambda^2+2}|\lambda|\big)}{\big(2\sqrt{4+\lambda^2}+\sqrt{\lambda^2+2}-|\lambda|\big)\tilde{c}}$$

when

$$t=\sqrt{\tfrac{1}{8}(4+\lambda^2)},\quad \tau=\tfrac{1}{2}\big(\sqrt{\lambda^2+2}-|\lambda|\big),\quad \theta=\sqrt{\frac{|\lambda|+2\tau}{2\tau(1-|\lambda|\tau)}}=\frac{1}{\sqrt{2}\tau}.$$

Since

$$\left(\int_M |A|^4\right)^{1/2}<K(\lambda),$$

we have from (20) that there exists $0<\varepsilon_0<1$ such that

$$\tilde{c}^{-1}\left(\int_M |A|^4\eta^4\rho^2\right)^{1/2}\le\frac{1+\varepsilon}{1-\varepsilon}\cdot\frac{1-\varepsilon_0}{\tilde{c}}\left(\int_M |A|^4\eta^4\rho^2\right)^{1/2}+C(\varepsilon,\lambda)\int_M |A|^2|\nabla\eta|^2\rho.$$

Let $\varepsilon=\frac{1}{2}\varepsilon_0$. Since $\int_M |A|^4\,\mathrm{d}\mu$ is bounded, we choose η as in the proof of Theorem 1 and a similar argument implies $|A|\equiv 0$. □

Using a similar argument, we give the proof of Theorem 4.

Proof of Theorem 4. For $n=2$, we have

$$\tfrac{1}{2}\mathcal{L}|\mathring{A}|^2\ge\big|\nabla|\mathring{A}|\big|^2+\frac{1+\lambda^2}{2}|\mathring{A}|^2-\tfrac{1}{8}|\mathring{A}|^2|X^N|^2-|\mathring{A}|^4,$$

and

$$2\int_M|\mathring{A}|^4\eta^2\rho+(C^2+2C|\lambda|-1)\int_M|\mathring{A}|^2\eta^2\rho+\frac{2}{\varepsilon}\int_M|\mathring{A}|^2|\nabla\eta|^2\rho \ge 2(1-\varepsilon)\int_M\big|\nabla|\mathring{A}|\big|^2\eta^2\rho \tag{21}$$

with $0<\varepsilon<1$.

Set $f=|\mathring{A}|\rho^{1/2}\eta$. By (16) and the hypothesis $|H|\le\sqrt{\frac{2}{3}+\lambda^2}-|\lambda|\triangleq C$, we have

$$\tilde{c}^{-1}\Big(\int_M|f|^4\Big)^{1/2}\le\frac{1}{t}\Big[\int_M\big|\nabla(|\mathring{A}|\eta)\big|^2\rho+\frac{1}{2}\int_M|\mathring{A}|^2\eta^2\rho+\frac{\lambda^2}{8}\int_M|\mathring{A}|^2\eta^2\rho\Big] \tag{22}$$
$$+t\int_M|\mathring{A}|^2\eta^2\rho+\frac{1}{2}\int|H||\mathring{A}|^2\eta^2\rho$$
$$\le\frac{1}{t}\int_M\big(\big|\nabla|\mathring{A}|\big|^2\eta^2+2|\mathring{A}|\eta\nabla|\mathring{A}|\cdot\nabla\eta+|\mathring{A}|^2|\nabla\eta|^2\big)\rho$$
$$+\Big(t+\frac{C}{2}+\frac{1}{2t}+\frac{\lambda^2}{8t}\Big)\int_M|\mathring{A}|^2\eta^2\rho.$$

Combining the Cauchy inequality, (21) and (22), we have for any $\delta>0$

$$\tilde{c}^{-1}\Big(\int_M|f|^4\Big)^{1/2}\le\frac{1}{t}\Big[(1+\delta)\int_M\big|\nabla|\mathring{A}|\big|^2\eta^2\rho+\Big(1+\frac{1}{\delta}\Big)\int_M|\mathring{A}|^2|\nabla\eta|^2\rho\Big]$$
$$+\Big(t+\frac{C}{2}+\frac{1}{2t}+\frac{\lambda^2}{8t}\Big)\int_M|\mathring{A}|^2\eta^2\rho$$
$$\le\frac{1+\delta}{t}\frac{1}{2(1-\varepsilon)}\Big[2\int_M|\mathring{A}|^4\eta^2\rho+(C^2+2C|\lambda|-1)\int_M|\mathring{A}|^2\eta^2\rho$$
$$+\frac{2}{\varepsilon}\int_M|\mathring{A}|^2|\nabla\eta|^2\rho\Big]$$
$$+\frac{1}{t}\Big(1+\frac{1}{\delta}\Big)\int_M|\mathring{A}|^2|\nabla\eta|^2\rho+\Big(t+\frac{C}{2}+\frac{1}{2t}+\frac{\lambda^2}{8t}\Big)\int_M|\mathring{A}|^2\eta^2\rho.$$

Put

$$\delta=\frac{(4+\lambda^2+8t^2+4tC)(1+\varepsilon)}{4[1-(C^2+2C|\lambda|)]}-1>0.$$

Then we get

$$\tilde{c}^{-1}\Big(\int_M|f|^4\Big)^{1/2} \tag{23}$$
$$\le\frac{1}{t}\cdot\frac{4+\lambda^2+8t^2+4tC}{4[1-(C^2+2C|\lambda|)]}\cdot\frac{1+\varepsilon}{1-\varepsilon}\cdot\int_M|\mathring{A}|^4\eta^2\rho$$
$$+\frac{1}{t}\Big[\frac{1+\delta}{\varepsilon(1-\varepsilon)}+1+\frac{1}{\delta}\Big]\int_M|\mathring{A}|^2|\nabla\eta|^2\rho$$
$$\le\frac{1}{t}\cdot\frac{4+\lambda^2+8t^2+4tC}{4[1-(C^2+2C|\lambda|)]}\cdot\frac{1+\varepsilon}{1-\varepsilon}\cdot\Big(\int_M|f|^4\Big)^{1/2}\cdot\Big(\int_M|\mathring{A}|^4\Big)^{1/2}$$
$$+\frac{1}{t}\Big[\frac{1+\delta}{\varepsilon(1-\varepsilon)}+1+\frac{1}{\delta}\Big]\int_M|\mathring{A}|^2|\nabla\eta|^2\rho.$$

Set

$$D(\lambda, C, t) = \frac{4t[1-(C^2+2C|\lambda|)]}{(4+\lambda^2+8t^2+4tC)\tilde{c}}.$$

We choose $t=\sqrt{\frac{1}{8}(4+\lambda^2)}$ such that $D(\lambda, C, t)$ achieves its maximum

$$D(\lambda) = \frac{1}{3\Big(\sqrt{8+2\lambda^2}+\sqrt{\frac{2}{3}+\lambda^2}-|\lambda|\Big)\tilde{c}}.$$

Since

$$\Big(\int_M |\mathring{A}|^4\Big)^{1/2} < D(\lambda),$$

we have from (23) that there exists $0<\varepsilon_0<1$ such that

$$\tilde{c}^{-1}\Big(\int_M |f|^4\Big)^{1/2} \le \frac{1+\varepsilon}{1-\varepsilon}\cdot\frac{1-\varepsilon_0}{\tilde{c}}\cdot\Big(\int_M |f|^4\Big)^{1/2} + C(\varepsilon,\lambda)\int_M |\mathring{A}|^2|\nabla\eta|^2\rho.$$

Let $\varepsilon=\frac{1}{2}\varepsilon_0$. Since $\int_M |\mathring{A}|^4\,\mathrm{d}\mu$ is bounded, we choose η as above and a similar argument implies $\mathring{A}\equiv 0$. Therefore, M is totally umbilical, i.e., M is $\mathbb{S}^2(\sqrt{\lambda^2+4}-\lambda)$ or $\mathbb{R}^2$. Since we have assumed that

$$|H| \le \sqrt{\tfrac{2}{3}+\lambda^2}-|\lambda|,$$

the first case is excluded. This completes the proof of Theorem 4. $\square$

Remark. Similarly, it is seen from the proof of Theorem 4 that we can prove that if $\sup_M |H| < \sqrt{1+\lambda^2}-|\lambda|$ and if $\big(\int_M |\mathring{A}|^4\,\mathrm{d}\mu\big)^{1/2} < D(\lambda, \sup_M |H|)$, then M is a hyperplane. Here $D(\lambda, \sup_M |H|)$ is a positive constant depending on λ and $\sup_M |H|$.

Acknowledgements

The authors would like to thank the referees for their valuable comments and suggestions.

References

[Cao and Li 2013] H.-D. Cao and H. Li, "A gap theorem for self-shrinkers of the mean curvature flow in arbitrary codimension", *Calc. Var. Partial Differential Equations* **46**:3-4 (2013), 879–889. MR Zbl

[Cao et al. 2014] S.-J. Cao, H.-W. Xu, and E.-T. Zhao, "Pinching theorems for self-shrinkers of higher codimensions", preprint, 2014, available at http://www.cms.zju.edu.cn/upload/file/20170320/1489994331903839.pdf.

[Cheng and Peng 2015] Q.-M. Cheng and Y. Peng, "Complete self-shrinkers of the mean curvature flow", *Calc. Var. Partial Differential Equations* **52**:3-4 (2015), 497–506. MR Zbl

[Cheng and Wei 2014a] Q.-M. Cheng and G. Wei, "Complete λ-hypersurfaces of weighted volume-preserving mean curvature flow", preprint, 2014. arXiv

[Cheng and Wei 2014b] Q.-M. Cheng and G. Wei, "The Gauss image of λ-hypersurfaces and a Bernstein type problem", preprint, 2014. arXiv

[Cheng and Wei 2015] Q.-M. Cheng and G. Wei, "A gap theorem of self-shrinkers", *Trans. Amer. Math. Soc.* **367**:7 (2015), 4895–4915. MR Zbl

[Cheng et al. 2016] Q.-M. Cheng, S. Ogata, and G. Wei, "Rigidity theorems of λ-hypersurfaces", *Comm. Anal. Geom.* **24**:1 (2016), 45–58. MR Zbl

[Colding and Minicozzi 2012] T. H. Colding and W. P. Minicozzi, II, "Generic mean curvature flow, I: Generic singularities", *Ann. of Math.* (2) **175**:2 (2012), 755–833. MR Zbl

[Colding et al. 2015] T. H. Colding, T. Ilmanen, and W. P. Minicozzi, II, "Rigidity of generic singularities of mean curvature flow", *Publ. Math. Inst. Hautes Études Sci.* **121** (2015), 363–382. MR Zbl

[Ding and Xin 2013] Q. Ding and Y. L. Xin, "Volume growth, eigenvalue and compactness for self-shrinkers", *Asian J. Math.* **17**:3 (2013), 443–456. MR Zbl

[Ding and Xin 2014] Q. Ding and Y. L. Xin, "The rigidity theorems of self-shrinkers", *Trans. Amer. Math. Soc.* **366**:10 (2014), 5067–5085. MR Zbl

[Guang 2014] Q. Guang, "Gap and rigidity theorems of λ-hypersurfaces", preprint, 2014. arXiv

[Hoffman and Spruck 1974] D. Hoffman and J. Spruck, "Sobolev and isoperimetric inequalities for Riemannian submanifolds", *Comm. Pure Appl. Math.* **27** (1974), 715–727. MR Zbl

[Huisken 1990] G. Huisken, "Asymptotic behavior for singularities of the mean curvature flow", *J. Differential Geom.* **31**:1 (1990), 285–299. MR Zbl

[Ilmanen 1995] T. Ilmanen, "Singularities of mean curvature flow of surfaces", preprint, 1995, available at https://people.math.ethz.ch/~ilmanen/papers/sing.ps.

[Le and Sesum 2011] N. Q. Le and N. Sesum, "Blow-up rate of the mean curvature during the mean curvature flow and a gap theorem for self-shrinkers", *Comm. Anal. Geom.* **19**:4 (2011), 633–659. MR Zbl

[Li and Wei 2014] H. Li and Y. Wei, "Classification and rigidity of self-shrinkers in the mean curvature flow", *J. Math. Soc. Japan* **66**:3 (2014), 709–734. MR Zbl

[Lin 2016] H. Lin, "Some rigidity theorems for self-shrinkers of the mean curvature flow", *J. Korean Math. Soc.* **53**:4 (2016), 769–780. MR Zbl

[McGonagle and Ross 2015] M. McGonagle and J. Ross, "The hyperplane is the only stable, smooth solution to the isoperimetric problem in Gaussian space", *Geom. Dedicata* **178** (2015), 277–296. MR Zbl

[Ogata 2015] S. Ogata, "A global pinching theorem of complete λ-hypersurfaces", preprint, 2015. arXiv

[Okumura 1974] M. Okumura, "Hypersurfaces and a pinching problem on the second fundamental tensor", *Amer. J. Math.* **96** (1974), 207–213. MR Zbl

[White 1997] B. White, "Stratification of minimal surfaces, mean curvature flows, and harmonic maps", *J. Reine Angew. Math.* **488** (1997), 1–35. MR Zbl

[Xu and Gu 2007a] H.-W. Xu and J.-R. Gu, "A general gap theorem for submanifolds with parallel mean curvature in $\mathbb{R}^{n+p}$", *Comm. Anal. Geom.* **15**:1 (2007), 175–193. MR Zbl

[Xu and Gu 2007b] H.-W. Xu and J.-R. Gu, "L^2-isolation phenomenon for complete surfaces arising from Yang–Mills theory", *Lett. Math. Phys.* **80**:2 (2007), 115–126. MR Zbl

Received February 25, 2016. Revised September 22, 2016.

HUIJUAN WANG
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU
ZHEJIANG 310027
CHINA
whjuan@zju.edu.cn

HONGWEI XU
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU
ZHEJIANG 310027
CHINA
xuhw@zju.edu.cn

ENTAO ZHAO
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU
ZHEJIANG 310027
CHINA
zhaoet@zju.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.475

BACH-FLAT h-ALMOST GRADIENT RICCI SOLITONS

GABJIN YUN, JINSEOK CO AND SEUNGSU HWANG

On an n-dimensional complete manifold M, consider an h-almost gradient Ricci soliton, which is a generalization of a gradient Ricci soliton. We prove that if the manifold is Bach-flat and $dh/du > 0$, then the manifold M is either Einstein or rigid. In particular, such a manifold has harmonic Weyl curvature. Moreover, if the dimension of M is four, the metric g is locally conformally flat.

1. Introduction

The notion of an h-almost Ricci soliton was introduced by Gomes, Wang, and Xia [Gomes et al. 2015]. Such a soliton is a generalization of an almost Ricci soliton presented in [Barros and Ribeiro 2012; Pigola et al. 2011]. An h-almost Ricci soliton is a complete Riemannian manifold (M^n, g) with a vector field X on M, a soliton function $\lambda : M \to \mathbb{R}$ and a signal function $h : M \to \mathbb{R}^+$ satisfying the equation

$$r_g + \tfrac{1}{2} h \mathcal{L}_X g = \lambda g,$$

where r_g is the Ricci curvature of g. A function is called signal if it has only one sign; in other words, it is either positive or negative on M. Let (M, g, X, h, λ) denote an h-almost Ricci soliton. In particular, $(M, g, \nabla u, h, \lambda)$ for some smooth function $u : M \to \mathbb{R}$ is called an h-almost gradient Ricci soliton with potential function u. In this case, we have

$$r_g + h\, D_g\, du = \lambda g. \tag{1-1}$$

Here, $D_g\, du$ denotes the Hessian of u. Note that if we take $u = e^{-f/m}$ and $h = -m/u$, then (1-1) becomes

$$\mathrm{Ric}_f^m = r_g + D_g\, df - \frac{1}{m}\, df \otimes df = \lambda g.$$

Yun was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0007465). Hwang (corresponding author) was supported by the Ministry of Education (NRF-2015R1D1A1A01057661).
MSC2010: 53C25, 58E11.
Keywords: h-almost gradient Ricci soliton, Bach-flat, Einstein metric.

In other words, the $(\lambda, n+m)$-Einstein equation is a special case of (1-1). Here, Ric_f^m is called the m-Bakry–Emery tensor. For further details about h-almost Ricci solitons, see [Gomes et al. 2015].

In this paper we consider Bach-flat h-almost gradient Ricci solitons. The Bach tensor was introduced by R. Bach, and this notion plays an important role in conformal relativity. On any n-dimensional Riemannian manifold (M, g), $n \geq 4$, the Bach tensor is defined by

$$B = \frac{1}{n-3}\delta^D \delta\mathcal{W} + \frac{1}{n-2}\mathring{\mathcal{W}}z,$$

where $\mathcal{W}$ is the Weyl tensor, z is the traceless Ricci tensor, and $\mathring{\mathcal{W}}z$ is defined by

$$\mathring{\mathcal{W}}z(X, Y) = \sum_{i=1}^{n} z(\mathcal{W}(X, E_i)Y, E_i)$$

for some orthonormal basis $\{E_i\}_{i=1}^n$. It is easy to see that if (M, g) is either locally conformally flat or Einstein, then it is Bach-flat: $B = 0$. When $n = 4$, it is well known that Bach-flat metrics on a compact manifold M are critical points of the functional

$$g \mapsto \int_M |\mathcal{W}|^2 \, dv_g.$$

It is clear that when $h = 1$ and λ is a positive constant, an h-almost gradient Ricci soliton reduces to a gradient shrinking Ricci soliton. Cao and Chen [2013] proved that a complete Bach-flat gradient shrinking Ricci soliton is either Einstein or rigid. On the other hand, Qing and Yuan [2013] classified Bach-flat static spaces.

Our main result is as follows, which can be considered as a generalization of [Cao and Chen 2013].

Theorem 1.1. *Let $(M, g, \nabla u, h, \lambda)$ be an n-dimensional Bach-flat h-almost gradient Ricci soliton with potential function u. Assume that each level set of u is compact and h is a function of u only. Then $(M, g, \nabla u, h, \lambda)$ is either*

(1) *Einstein with constant functions u and h, or*

(2) *locally isometric to a warped product with $(n-1)$-dimensional Einstein fibers if $dh/du > 0$ on M.*

For example, when $m > 0$, $h = -m/u < 0$ satisfies the condition of Theorem 1.1, since

$$\frac{dh}{du} = \frac{m}{u^2} > 0.$$

This recovers the result of [Chen and He 2013]. It will be interesting if one can weaken the condition of Theorem 1.1.

In the case of (2) in Theorem 1.1, a warped product metric has vanishing Cotton tensor (see (2-4) below) since its fiber is Einstein. Thus, as a consequence of Theorem 1.1, we have the following.

Corollary 1.2. *Let $(M, g, \nabla u, h, \lambda)$ be an n-dimensional Bach-flat h-almost gradient Ricci soliton with potential function u. Assume that each level set of u is compact and h is a function of u only. If $dh/du > 0$ on M, then (M, g) has harmonic Weyl curvature.*

In particular, when $n = 4$, the Einstein fibers in Theorem 1.1 have constant curvature. A computation shows that such a metric is locally conformally flat, which proves the following theorem.

Theorem 1.3. *Let $(M, g, \nabla u, h, \lambda)$ be a 4-dimensional Bach-flat h-almost gradient Ricci soliton with potential function u. Assume each level set of u is compact and h is a function of u only with $dh/du > 0$. Then (M, g) is locally conformally flat.*

As in [Chen and He 2013], Theorem 1.1, Corollary 1.2, and Theorem 1.3 can be extended to the case in which M has a nonempty boundary.

2. Preliminaries

In this section, we derive several useful identities containing various curvatures and the Cotton tensor.

We start with basic definitions of differential operators acting on tensors. Let us denote by $C^\infty(S^2M)$ the space of sections of symmetric 2-tensors on a Riemannian manifold M. Let D be the Levi-Civita connection of (M, g). Then the differential operator d^D from $C^\infty(S^2M)$ into $C^\infty(\Lambda^2 M \otimes T^*M)$ is defined as

$$d^D\omega(X, Y, Z) = (D_X\omega)(Y, Z) - (D_Y\omega)(X, Z)$$

for $\omega \in C^\infty(S^2M)$ and vectors X, Y, and Z. Let us denote by δ^D the formal adjoint operator of d^D.

For a function $f \in C^\infty(M)$ and $\omega \in C^\infty(S^2M)$, $df \wedge \omega$ is defined as

$$(df \wedge \omega)(X, Y, Z) = df(X)\omega(Y, Z) - df(Y)\omega(X, Z).$$

Here, df denotes the usual total differential of f. We also denote by δ the negative divergence operator such that $\Delta f = -\delta\, df$.

Taking the trace of (1-1) gives

$$s_g + h\,\Delta u = n\lambda.$$

Thus,

$$ds_g + \Delta u\, dh + h\, d\Delta u = n\, d\lambda.$$

By taking the divergence of (1-1), we have

$$-\tfrac{1}{2}\, ds_g - D_g\, du(\nabla h, \cdot) - h\, r_g(\nabla u, \cdot) - h\, d\Delta u = -d\lambda.$$

By adding the previous two equations, we have

$$\tfrac{1}{2}\, ds_g - D_g\, du(\nabla h, \cdot) - h\, r_g(\nabla u, \cdot) + \Delta u\, dh = (n-1)\, d\lambda. \tag{2-1}$$

Note that

$$\delta(h r_g(\nabla u, \cdot)) = -r_g(\nabla u, \nabla h) - \tfrac{1}{2} h\langle \nabla s_g, \nabla u\rangle + |r_g|^2 - \lambda s_g. \tag{2-2}$$

Therefore, we have the following equality.

Proposition 2.1. *On M we have*

$$(n-1)\Delta\lambda = \tfrac{1}{2}\,\Delta s_g + |r_g|^2 - \lambda s_g - \tfrac{1}{2} h\langle \nabla s_g, \nabla u\rangle \\ + \left(\Delta u - \frac{\lambda}{h}\right)\Delta h + \frac{1}{h}\langle r_g, D_g\, dh\rangle - 2 r_g(\nabla u, \nabla h).$$

On the other hand, by applying d^D to (1-1), we have

$$d^D r_g - \frac{1}{h}\, dh \wedge r_g + h\,\tilde{\imath}_{\nabla u} R = d\lambda \wedge g - \frac{\lambda}{h}\, dh \wedge g. \tag{2-3}$$

Here, an interior product $\tilde{\imath}$ of the final factor is defined by

$$\tilde{\imath}_\xi R(X, Y, Z) = R(X, Y, Z, \xi),$$

and we used the identity

$$d^D D\, du = \tilde{\imath}_{\nabla u} R.$$

Hereafter, we denote s_g, r_g, and $D_g\, du$ by s, r, and $D\, du$, respectively. From the curvature decomposition, we can compute that

$$\tilde{\imath}_{\nabla u} R = \tilde{\imath}_{\nabla u}\mathcal{W} - \frac{1}{n-2}\, i_{\nabla u} r \wedge g + \frac{s}{(n-1)(n-2)}\, du \wedge g - \frac{1}{n-2}\, du \wedge r,$$

where $i_{\nabla u} r$ denotes the interior product defined by

$$i_{\nabla u} r(X) = r(\nabla u, X).$$

The Cotton tensor C is defined by

$$C = d^D r - \frac{1}{2(n-1)}\, ds \wedge g. \tag{2-4}$$

Then, by (2-1) and (2-3) as well as the fact that

$$s + h\,\Delta u = n\lambda,$$

we have

$$C + h\tilde{\imath}_{\nabla u}\mathcal{W} = h\,D + \frac{h}{n-1}\,i_{\nabla u}r \wedge g + d\lambda \wedge g - \frac{1}{2(n-1)}\,ds \wedge g + \frac{1}{h}\,dh \wedge r - \frac{\lambda}{h}\,dh \wedge g = h\,D + H, \tag{2-5}$$

where D is defined (as usual) by

$$(n-2)\,D = du \wedge r + \frac{1}{n-1} i_{\nabla u}r \wedge g - \frac{s}{n-1}\,du \wedge g, \tag{2-6}$$

and H is defined by

$$\begin{aligned} H &= -\frac{1}{n-1}\,i_{\nabla h}D\,du \wedge g + dh \wedge \left(\frac{1}{h}r + \frac{\Delta u}{n-1}g - \frac{\lambda}{h}\,g\right) \\ &= db \wedge r + \frac{1}{n-1}\,i_{\nabla b}r \wedge g - \frac{s}{n-1}\,db \wedge g. \end{aligned}$$

Here, $b = \log|h|$ with $\nabla b = \nabla h/h$. In particular, $g^{ik}H_{ijk} = -g^{ik}H_{jik} = 0$.

Proposition 2.2. *Let $(M, g, \nabla u, h, \lambda)$ be an h-almost gradient Ricci soliton with potential function u. Then*

$$C + h\tilde{\imath}_{\nabla u}\mathcal{W} = h\,D + H.$$

In particular, if h is constant or $dh/du = 0$, then $H \equiv 0$.

3. Bach-flat metrics

In this section, we assume that g is Bach-flat. Note that

$$\delta\mathcal{W} = -\frac{n-3}{n-2}\,C.$$

Recall that the Bach tensor is given by

$$B = \frac{1}{n-3}\,\delta^D\delta\mathcal{W} + \frac{1}{n-2}\mathring{\mathcal{W}}z = \frac{1}{n-2}(-\delta C + \mathring{\mathcal{W}}z).$$

Since

$$\begin{aligned} &\delta(h\tilde{\imath}_{\nabla u}\mathcal{W})(X, Y) \\ &\quad = -\mathcal{W}(\nabla h, X, Y, \nabla u) + h\,\delta\mathcal{W}(X, Y, \nabla u) + h\,\mathcal{W}(X, E_i, Y, D_{E_i}\,du) \\ &\quad = l\mathcal{W}(X, \nabla h, Y, \nabla u) - \frac{n-3}{n-2}\,h\,C(Y, \nabla u, X) - \mathring{\mathcal{W}}z, \end{aligned}$$

by taking the divergence of (2-5) we have

$$-(n-2)B(X,Y) = -\mathcal{W}(X,\nabla h,Y,\nabla u) + \frac{n-3}{n-2}hC(Y,\nabla u,X) - i_{\nabla h}D(X,Y) + h\delta D(X,Y) + \delta H(X,Y).$$

Hence,

$$-(n-2)B(\nabla u,\nabla u) = -D(\nabla h,\nabla u,\nabla u) + h\delta D(\nabla u,\nabla u) + \delta H(\nabla u,\nabla u).$$

As a result, from the assumption that $B = 0$ and h is a function of u only,

$$0 = \frac{1}{h}D(\nabla h,\nabla u,\nabla u) = \delta D(\nabla u,\nabla u) + \frac{1}{h}\delta H(\nabla u,\nabla u).$$

Let $\{E_i\}_{i=1}^n$ be a normal geodesic frame. Note that, since

$$hD(E_i, D_{E_i}\,du,\nabla u) = -D(E_i,E_k,\nabla u)r_{ik} = 0,$$

we have

$$\operatorname{div}(D(\,\cdot\,,\nabla u,\nabla u)) = -\delta D(\nabla u,\nabla u) + D(E_i,\nabla u, D_{E_i}\,du).$$

Furthermore,

$$\begin{aligned}|D|^2 &= \frac{1}{n-2}\big(du(E_i)r(E_j,E_k) - du(E_j)r(E_i,E_k)\big)D_{ijk}\\ &= -\frac{2}{n-2}D(E_i,\nabla u,E_k)r_{ik}\\ &= \frac{2h}{n-2}D(E_i,\nabla u,D_{E_i}\,du).\end{aligned}$$

Similarly, since

$$hH(E_i, D_{E_i}\,du,\nabla u) = -H(E_i,E_k,\nabla u)r_{ik} = 0$$

and h is a function of u only, we have

$$\operatorname{div}\Big(\frac{1}{h}H(\,\cdot\,,\nabla u,\nabla u)\Big) = -\frac{1}{h}\delta H(\nabla u,\nabla u) + \frac{1}{h}H(E_i,\nabla u,D_{E_i}\,du).$$

Moreover,

$$\begin{aligned}|H|^2 &= -\frac{2}{h}H(E_i,\nabla h,E_k)r_{ik}\\ &= -\frac{2}{h}\frac{dh}{du}H(E_i,\nabla u,E_k)r_{ik}\\ &= 2\frac{dh}{du}H(E_i,\nabla u,D_{E_i}\,du).\end{aligned}$$

Thus,

$$0=\int_{t_1\le u\le t_2}\delta D(\nabla u,\nabla u)+\frac{1}{h}\delta H(\nabla u,\nabla u)$$
$$=\frac{n-2}{2}\int_{t_1\le u\le t_2}\frac{|D|^2}{h}+\frac{1}{2}\int_{t_1\le u\le t_2}\frac{|H|^2}{h\,\frac{dh}{du}}.$$

Since h is signal, h is either positive or negative. For each case, we derive $D=H=0$ when $dh/du>0$. Therefore we have the following result.

Lemma 3.1. *Let $(M, g, \nabla u, h, \lambda)$ be a Bach-flat h-almost gradient Ricci soliton with potential function u. Assume that each level set of u is compact and h is a function of u only. If $dh/du>0$ on M, then on M we have*

$$D=H=0.$$

Now, since $D=H=0$, by (2-4) and (2-5)

$$C=-h\,\tilde{\imath}_{\nabla u}\mathcal{W}. \tag{3-1}$$

By taking the divergence of (3-1), we have

$$\mathcal{W}(X,\nabla h,Y,\nabla u)=\frac{n-3}{n-2}hC(Y,\nabla u,X).$$

By combining these equations,

$$\frac{n-3}{n-2}h^2C(Y,\nabla u,X)=-C(X,\nabla h,Y),$$

and

$$\mathcal{W}(X,\nabla h,Y,\nabla u)=-\frac{n-3}{n-2}h^2\mathcal{W}(X,\nabla u,Y,\nabla u).$$

Therefore, we have the following.

Corollary 3.2. *When $D=H=0$, we have*

$$\mathcal{W}(\,\cdot\,,\nabla u,\,\cdot\,,\nabla u)=C(\cdot\,,\nabla u,\,\cdot\,)=0, \tag{3-2}$$

unless

$$\frac{dh}{du}=-\left(\frac{n-3}{n-2}\right)h^2.$$

For example, when $h=-m/u$, (3-2) holds if $m\neq 0$ or $-(n-2)/(n-3)$. Note that (3-2) also holds if h is constant.

Moreover, we have the following result.

Lemma 3.3. *Suppose that $dh/du>0$. Then, for X orthogonal to ∇u,*

$$r(X,\nabla u)=0. \tag{3-3}$$

In particular,

$$i_{\nabla u} r = \alpha\, du,$$

where $\alpha = r(N, N)$ *with* $N = \nabla u/|\nabla u|$.

Proof. By Lemma 3.1, $D = H = 0$. From (2-3), if X is orthogonal to ∇u,

$$d^D r(X, Y, \nabla u) = -\frac{1}{h}\, dh(Y) r(X, \nabla u) + d\lambda(X)\, du(Y).$$

Since $C(X, Y, \nabla u) = -h\mathcal{W}(X, Y, \nabla u, \nabla u) = 0$ by (3-1), by (2-4) we have

$$d^D r(X, Y, \nabla u) = \frac{1}{2(n-1)}\, ds(X)\, du(Y).$$

Thus, by (2-1)

$$\begin{aligned}\frac{1}{h}\frac{dh}{du} r(X, \nabla u) &= d\lambda(X) - \frac{1}{2(n-1)}\, ds(X)\\ &= \frac{1}{(n-1)h}\left(\frac{dh}{du} - h^2\right) r(X, \nabla u),\end{aligned}$$

which implies that

$$\left((n-2)\frac{dh}{du} + h^2\right) r(X, \nabla u) = 0. \qquad \square$$

Note that Lemma 3.3 holds with the assumptions that $D = H = 0$ and

$$\frac{dh}{du} \neq -\frac{1}{n-2} h^2 \tag{3-4}$$

without $dh/du > 0$. For example, in the case of the m-Bakry–Emery tensor, $h = -m/u$ satisfies (3-4) if $m \neq 2 - n$.

4. Level sets of u

In this section, we investigate the structure of regular level sets of the potential function u. For a regular value c, we denote the level set $u^{-1}(u)$ by L_c. On L_c, let $\{E_i\}$, $1 \le i \le n$, be an orthonormal frame with $E_n = N = \nabla u/|\nabla u|$.

Furthermore, throughout the section we assume that $D = H = 0$ with

$$\frac{dh}{du} \neq -\left(\frac{n-3}{n-2}\right) h^2 \quad \text{and} \quad \frac{dh}{du} \neq -\frac{1}{n-2} h^2.$$

Then, by Corollary 3.2, (3-2) and (3-3) hold. Furthermore, for X orthogonal to ∇u, by the proof of Lemma 3.3,

$$d\lambda(X) = \frac{1}{2(n-1)}\, ds(X).$$

Thus, $s+2(1-n)\lambda$ is constant on each level set of u. Furthermore,

$$\tfrac{1}{2}X(|\nabla u|^2)=\langle D_X\, du,\nabla u\rangle=\frac{1}{h}\big(\lambda\, du(X)-r(X,\nabla u)\big)=0,$$

which implies that $|\nabla u|^2$ is constant on each level set of u. Therefore, we have the following.

Lemma 4.1. *$|\nabla u|^2$ and $s+2(1-n)\lambda$ are constant on each regular level set of u.*

For further investigation, we need the following key lemma.

Lemma 4.2.

$$\begin{aligned}0=\frac{ns-(n-1)^2\lambda-\alpha}{(n-1)h}r-D_{\nabla u}r-\frac{r\circ r}{h}+\frac{n-3}{2(n-1)}\,du\otimes ds\\ +\frac{1}{n-1}\big(ds(u)-\langle\nabla u,\nabla\alpha\rangle\big)g+\frac{s+(1-n)\lambda}{(n-1)h}(\alpha-s)g+\frac{1}{n-1}\,du\otimes d\alpha.\end{aligned}$$

Proof. To find δD, by (2-6), we first compute

$$\delta(du\wedge r)=\frac{s-(n-1)\lambda}{h}r-D_{\nabla u}r-\frac{r\circ r}{h}+\tfrac{1}{2}\,du\otimes ds.$$

By Lemma 3.3, $i_{\nabla u}r=\alpha\, du$. Thus,

$$\delta(i_{\nabla u}r\wedge g)=-\langle\nabla u,\nabla\alpha\rangle g+\frac{s+(1-n)\lambda}{h}\alpha g+du\otimes d\alpha-\frac{\alpha}{h}r.$$

Similarly,

$$-\delta(s\, du\wedge g)=ds(u)\,g-\frac{s^2+(1-n)s\lambda}{h}g-du\otimes ds+\frac{s}{h}r.$$

Hence, by (2-6) together with (3-3), we have

$$\begin{aligned}(n-2)\,\delta D=\frac{ns-(n-1)^2\lambda-\alpha}{(n-1)h}r-D_{\nabla u}r-\frac{r\circ r}{h}\\ +\frac{n-3}{2(n-1)}\,du\otimes ds+\frac{1}{n-1}\,du\otimes d\alpha\\ +\frac{1}{n-1}\left(ds(u)-\langle\nabla u,\nabla\alpha\rangle+\frac{s+(1-n)\lambda}{h}(\alpha-s)\right)g.\end{aligned}$$

Since $D=\delta D=0$, the proof follows. $\square$

Thus, we have the following.

Corollary 4.3. *$(n-3)\,s+2\alpha$ is constant on each regular level set of u.*

Proof. Let X be a vector orthogonal to ∇u. By putting $(X,\nabla u)$ in the equation in Lemma 4.2,

$$D_{\nabla u}r(X,\nabla u)=0. \tag{4-1}$$

Now, by putting $(\nabla u, X)$ in the equation in Lemma 4.2 again, we have

$$0 = \frac{n-3}{2(n-1)} |\nabla u|^2 \, ds(X) + \frac{1}{n-1} |\nabla u|^2 \, d\alpha(X),$$

since $r(X, \nabla u) = 0$ and

$$D_{\nabla u} r(\nabla u, X) = D_{\nabla u} r(X, \nabla u). \qquad \square$$

Lemma 4.4. *$s_g + 2(1-n)\alpha$ is constant on each regular level set of u.*

Proof. For X orthogonal to ∇u, by (3-2) and (4-1)

$$\begin{aligned} 0 &= C(X, \nabla u, \nabla u) \\ &= D_X r(\nabla u, \nabla u) - \frac{1}{2(n-1)} \, ds(X) |\nabla u|^2. \end{aligned}$$

Thus,

$$\begin{aligned} X(\alpha) &= \frac{1}{|\nabla u|^2} X(r(\nabla u, \nabla u)) \\ &= \frac{1}{|\nabla u|^2} \big(D_X r(\nabla u, \nabla u) + 2r(D_X \, du, \nabla u) \big) \\ &= \frac{1}{2(n-1)} \, ds(X), \end{aligned}$$

since

$$r(D_X \, du, \nabla u) = \frac{1}{h} (\lambda r(X, \nabla u) - r \circ r(X, \nabla u)) = 0. \qquad \square$$

By combining Lemma 4.1, Corollary 4.3, and Lemma 4.4, we have the following.

Theorem 4.5. *Let $(M, g, \nabla u, h, \lambda)$ be a Bach-flat h-almost gradient Ricci soliton with potential function u. Assume that each level set of u is compact and h is a function of u only with $dh/du > 0$. Then s_g, α, and λ are constant on each regular level set of u. In particular, if h is constant, the condition on dh/du is not necessary.*

When $D = 0$, the Ricci tensor has the following characterization.

Lemma 4.6. *Suppose that $D = 0$. Then the Ricci curvature tensor has at most two eigenvalues.*

Proof. Let $\{E_i\}$, $1 \le i \le n$, be an orthonormal frame with $E_n = N = \nabla u / |\nabla u|$. Then

$$II_{ij} = \frac{1}{h |\nabla u|} (\lambda g_{ij} - r_{ij}), \tag{4-2}$$

and

$$m = \operatorname{tr} II = \frac{n-1}{h |\nabla u|} \left(\lambda + \frac{\alpha - s}{n-1} \right).$$

Thus, m is constant on each level set of u, and

$$\begin{aligned}\left|II - \frac{m}{n-1} g\right|^2 &= |II|^2 - \frac{m^2}{n-1} \\ &= \frac{1}{h^2|\nabla u|^2}\left(|r|^2 - \alpha^2 - \frac{(s-\alpha)^2}{n-1}\right) \\ &= \frac{1}{h^2|\nabla u|^2}\left(|r|^2 - \frac{n}{n-1}\alpha^2 + \frac{2s\alpha}{n-1} - \frac{s^2}{n-1}\right).\end{aligned}$$

Since $r \circ r(\nabla u, \nabla u) = \alpha^2 |\nabla u|^2$, from the identity

$$\frac{n-2}{2}|D|^2 = |r|^2|\nabla u|^2 - \frac{n}{n-1} r \circ r(\nabla u, \nabla u) + \frac{2s}{n-1} r(\nabla u, \nabla u) - \frac{s^2}{n-1}|\nabla u|^2,$$

we have

$$|D|^2 = \frac{2}{n-2} h^2 |\nabla u|^4 \left| II - \frac{m}{n-1} g\right|^2.$$

Since $D = 0$, we have

$$II_{ij} = \frac{m}{n-1} g_{ij}, \tag{4-3}$$

which implies that

$$r_{ij} = \frac{s-\alpha}{n-1} g_{ij} \tag{4-4}$$

for $i = 1, \dots, n-1$ by (4-2). □

As an immediate consequence, on an open set $\{x \in M \mid \nabla u(x) \neq 0\}$, the Ricci tensor may be written as

$$r_g = \beta\, du \otimes du + \left(\frac{s-\alpha}{n-1}\right) g,$$

where

$$\beta = \frac{n\alpha - s}{(n-1)|\nabla u|^2}.$$

Thus, by (1-1) we have

$$D_g\, du = \frac{1}{h}\left(\lambda + \frac{\alpha - s}{n-1}\right) g - \frac{\beta}{h}\, du \otimes du.$$

Now, we are ready to prove Corollary 1.2, which shows the relationship between Bach-flat metrics and harmonic Weyl metrics.

Proof of Corollary 1.2. Note that, by (3-1) and (3-2)

$$C(\,\cdot\,, \cdot\,, \nabla u) = C(\,\cdot\,, \nabla u, \cdot\,) = 0.$$

On the other hand, by the Codazzi equation,

$$\langle R(X,Y)Z, N\rangle = D_Y II(X,Z) - D_X II(Y,Z).$$

Thus, for $1 \le i, j, k \le n-1$, by (4-3)

$$\begin{aligned}\langle R(E_i, E_j)E_k, N\rangle = E_j(II(E_i,E_k)) - II(D_{E_j}E_i, E_k) - II(E_i, D_{E_j}E_k)\\ - E_i(II(E_j,E_k)) + II(D_{E_i}E_j, E_k) + II(E_j, D_{E_i}E_k)\\ = 0.\end{aligned}$$

Therefore, by (2-3)

$$d^D r(E_i, E_j, E_k) = 0,$$

which implies that

$$C(E_i, E_j, E_k) = d^D r(E_i, E_j, E_k) - \frac{1}{2(n-1)}\, ds \wedge g(E_i, E_j, E_k) = 0.$$

Hence, C is identically zero, and so is $\delta\mathcal{W}$. □

The following is a restatement of Theorem 1.1.

Theorem 4.7. *Let $(M, g, \nabla u, h, \lambda)$ be a Bach-flat h-almost gradient Ricci soliton with potential function u. Assume that each level set of u is compact with $dh/du > 0$ on M. Then, either g is Einstein with constant function u or the metric can be written as*

$$g = dt^2 + \psi^2(t)\,\hat{g}_E,$$

where $\hat{g}_E$ is the Einstein metric on the level set $E = L_{c_0}$ for some c_0.

Proof. Assume that u is not constant. By Lemma 3.1, $D = H = 0$. Since $|\nabla u|^2$ depends only on u by Lemma 4.6, as shown in the proof of Theorem 7.9 of [He et al. 2012] with Remark 3.2 of [Cao and Chen 2013], the metric can be locally written as

$$g = dt^2 + \hat{g}_c.$$

Here, $\hat{g}_c$ denotes the induced metric on the level set $L_c = u^{-1}(c)$ for each regular value c. Furthermore, $(L_c, \hat{g}_c)$ is necessarily Einstein; by the Gauss equation

$$\hat{R}_{ijij} = R_{ijij} + II_{ii}II_{jj} - II_{ij}^2 = R_{ijij} + \frac{m^2}{(n-1)^2}.$$

Thus,

$$\hat{r}_{ii} = r_{ii} - R(N, E_i, N, E_i) + \frac{m^2}{n-1}.$$

By (3-2) and (4-4), we have

$$R(E_i, N, E_i, N) = \frac{1}{n-2}(r_{ii} + \alpha) - \frac{s}{(n-1)(n-2)} = \frac{\alpha}{n-1}.$$

Hence, it follows that

$$\hat{r}_{ii} = r_{ii} + \frac{m^2 - \alpha}{n-1} = \frac{1}{n-1}(s - 2\alpha + m^2) = \hat{\lambda}_0.$$

Since s, α, and m are constant along L_c, this proves that $(L_c, \hat{g}_c)$ has constant Ricci curvature. As a result, by a suitable change of variable, the metric g can be written as in the statement of Theorem 4.7. □

Acknowledgment

The authors express their gratitude to the referee for several valuable comments.

References

[Barros and Ribeiro 2012] A. Barros and E. Ribeiro, Jr., "Some characterizations for compact almost Ricci solitons", *Proc. Amer. Math. Soc.* **140**:3 (2012), 1033–1040. MR Zbl

[Cao and Chen 2013] H.-D. Cao and Q. Chen, "On Bach-flat gradient shrinking Ricci solitons", *Duke Math. J.* **162**:6 (2013), 1149–1169. MR Zbl

[Chen and He 2013] Q. Chen and C. He, "On Bach flat warped product Einstein manifolds", *Pacific J. Math.* **265**:2 (2013), 313–326. MR Zbl

[Gomes et al. 2015] J. N. Gomes, Q. Wang, and C. Xia, "On the h-almost Ricci soliton", preprint, 2015. arXiv

[He et al. 2012] C. He, P. Petersen, and W. Wylie, "On the classification of warped product Einstein metrics", *Comm. Anal. Geom.* **20**:2 (2012), 271–311. MR Zbl

[Pigola et al. 2011] S. Pigola, M. Rigoli, M. Rimoldi, and A. G. Setti, "Ricci almost solitons", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **10**:4 (2011), 757–799. MR Zbl

[Qing and Yuan 2013] J. Qing and W. Yuan, "A note on static spaces and related problems", *J. Geom. Phys.* **74** (2013), 18–27. MR Zbl

Received April 14, 2016. Revised September 29, 2016.

GABJIN YUN
DEPARTMENT OF MATHEMATICS
MYONG JI UNIVERSITY
SAN 38-2 NAMDONG
YONGIN, GYEONGGI 449-728
SOUTH KOREA
gabjin@mju.ac.kr

JINSEOK CO
DEPARTMENT OF MATHEMATICS
CHUNG-ANG UNIVERSITY
84 HEUKSEOK-RO, DONGJAK-GU
SEOUL 06969
SOUTH KOREA
co1010@hanmail.net

SEUNGSU HWANG
DEPARTMENT OF MATHEMATICS
CHUNG-ANG UNIVERSITY
84 HEUKSUK-RO, DONGJAK-GU
SEOUL 06969
SOUTH KOREA

seungsu@cau.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 288, No. 2, 2017
dx.doi.org/10.2140/pjm.2017.288.489

A SHARP HEIGHT ESTIMATE FOR THE SPACELIKE CONSTANT MEAN CURVATURE GRAPH IN THE LORENTZ–MINKOWSKI SPACE

JINGYONG ZHU

Based on the local comparison principle of Chen and Huang (1982), we study the local behavior of the difference of two spacelike graphs in a neighborhood of a second contact point. Then we apply it to the spacelike constant mean curvature graph in 3-dimensional Lorentz–Minkowski space $\mathbb{L}^3$, which can be viewed as a solution to the constant mean curvature equation over a convex domain $\Omega \subset \mathbb{R}^2$. We get the uniqueness of critical points for such a solution, which is an analogue of a result of Sakaguchi (1988). Last, by this uniqueness, we obtain a minimum principle for a functional depending on the solution and its gradient. This gives us a sharp gradient estimate for the solution, which leads to a sharp height estimate.

1. Introduction

Spacelike hypersurfaces of constant mean curvature (CMC) and CMC foliations play an important role in general relativity. Such surfaces are important because they provide Riemannian submanifolds with properties reflecting those of the spacetime. For example, if the weak energy condition is satisfied, a maximal hypersurface has positive scalar curvature. So the geometric properties of such hypersurfaces are worth researching, and finding conditions for their existence is a fundamental problem. Under the graph setting and some assumptions, Robert Bartnik and Leon Simon [1982] got a sufficient and necessary condition for the existence of a solution to

$$\begin{cases} \operatorname{div} \dfrac{Du}{\sqrt{1-|Du|^2}} = H(x,u), \quad |Du|<1 & \text{in } \Omega \subset \mathbb{R}^n, \\ u=\phi & \text{on } \partial\Omega, \end{cases} \tag{1-1}$$

where div stands for the divergence operator in the Euclidean plane $\mathbb{R}^n$ and

$$Du = (u_1, \dots, u_n), \quad u_i = \frac{\partial u}{\partial x_i}. \tag{1-2}$$

MSC2010: 35B38, 35B45, 35J93, 53C42, 53C50.
Keywords: height estimate, critical point, constant mean curvature, a priori estimates, Lorentz–Minkowski space.

In particular, the Theorem 3.6 in [Bartnik and Simon 1982] gives us a solution $u \in C^\infty(\overline{\Omega})$ to

$$\begin{cases} \operatorname{div} \dfrac{Du}{\sqrt{1-|Du|^2}} = nH, \quad |Du| < 1 & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{1-3}$$

over a bounded $C^{2,\alpha}$ domain Ω with H being a positive constant. In this case, they pointed out that $\nu_{n+1} = 1/\sqrt{1-|Du|^2}$ satisfies following elliptic equation

$$\Delta_M \nu_{n+1} = \nu_{n+1} \|A\|^2, \tag{1-4}$$

where Δ_M and A denote the Laplace operator and the second fundamental form of the graph $M = \{(x, u(x)) : x \in \mathbb{R}^n, u \in C^\infty(\mathbb{R}^n)\}$, respectively. The boundary gradient estimate is the most important step leading to the existence of u. To do so, they used the following spherically symmetric barrier functions:

$$w^\pm = w^\pm(\xi) \pm \int_0^{|x-\xi|} \frac{K - Ht^n}{\sqrt{t^{2n-2} + (K - Ht^n)^2}}\, dt, \tag{1-5}$$

where K is a positive constant. From the proof of their Proposition 3.1, one can get following boundary gradient estimate:

$$\max_{\partial\Omega} |Du| \le \frac{1 - H\varepsilon^{n+1}}{\sqrt{\varepsilon^{2n} + (1 - H\varepsilon^{n+1})^2}}, \tag{1-6}$$

where $\varepsilon = \varepsilon(\Omega)$ is a sufficiently small constant. Obviously, this bound is not sharp. Also, the dependence of ε on Ω is not specific. Since the graph is spacelike, they roughly used the diameter of the domain Ω to control the C^0 norm of the solution u. So the question is, can we give a sharp C^0 or C^1 estimate for the solution in terms of the boundary geometry?

Early in 1979, Lawrence E. Payne and Gérard A. Philippin [1979] have used so-called P-functions to derive sharp C^0 and C^1 upper bounds for the solution of the Dirichlet problem

$$\begin{cases} \operatorname{div} \dfrac{Du}{\sqrt{1+|Du|^2}} = -2H & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{1-7}$$

over a strictly convex domain $\Omega \subset \mathbb{R}^2$ with H being a positive constant. The key is a maximum principle for the P-function

$$\Phi(x, \alpha) = \int_0^{q^2} \frac{g(\xi + 2\xi g'(\xi))}{\rho}\, d\xi + \alpha \int_0^u f(\eta)\, d\eta, \tag{1-8}$$

where u, g, ρ, f, q satisfy

$$(g(q^2)u_i)_i+\rho(q^2)f(u)=0,\quad g(\xi)+2\xi g'(\xi)>0,\qquad \text{for all } \xi\geq 0,$$
$$\rho>0,\quad g>0,\qquad q^2=|Du|^2=\sum u_i^2. \tag{1-9}$$

In the same year, by the uniqueness of critical points for a solution and the strict convexity of the domain, G. A. Philippin [1979] also got a minimum principle for $\Phi(x,\alpha)$ provided $\alpha>1$ and used it to derive lower bounds for C^0 and C^1 norms of the solution. But he did not assert the sharpness of the estimates, since he did not have a similar minimum principle for $\Phi(x,1)$ at that time. In 2000, Xi-Nan Ma [2000] solved this issue through uniqueness of critical points and analyticity of the solution. He did a long computation to show that all the derivatives of $\Phi(x,1)$ vanish at the unique critical point if $\Phi(x,1)$ takes its minimum value at that point. By the strong unique continuation of analytic function, $\Phi(x,1)$ is a constant. Once one has this minimum principle, the sharpness is easy to derive.

For our question, the maximum principle in [Payne and Philippin 1979] still works. So the upper bound of the gradient estimate and the lower bound of the minimum value are easy to derive, which we will do later in this paper. However, the minimum principle is not available any more. In this paper, we want to prove a minimum principle for $\Phi(x,1)$ when u is a spacelike CMC graph solving

$$\begin{cases}\operatorname{div}\dfrac{Du}{\sqrt{1-|Du|^2}}=2H,\quad |Du|<1 & \text{in } \Omega,\\ u=0 & \text{on } \partial\Omega,\end{cases} \tag{1-10}$$

and use it to derive sharp C^0 and C^1 bounds for the solution to (1-10).

Not only is the uniqueness of critical point the important ingredient to get the sharpness of the a priori estimate, but is itself worth study. Together with the convexity [Caffarelli and Friedman 1985; Guan and Ma 2003; Chen 2014] and curvature estimates [Ma and Zhang 2013] for level sets, they are the most important geometric properties of solutions to elliptic or parabolic equations. G. A. Philippin [1979] showed that the solution to (1-7) has only one critical point when Ω is strictly convex. His method of proof is based on an idea of L. E. Payne [1973]. Jin-Tzu Chen [1984] proved the uniqueness of the critical point for a solution to

$$\begin{cases}\operatorname{div}\dfrac{Du}{\sqrt{1+|Du|^2}}=2H & \text{in } \Omega,\\ \dfrac{Du}{\sqrt{1+|Du|^2}}\cdot\nu=1 & \text{on } \partial\Omega,\end{cases} \tag{1-11}$$

where $\Omega\subset\mathbb{R}^2$ is a bounded convex domain with outer normal ν on the boundary $\partial\Omega$ and H is a positive constant. His proof is based on a nice comparison technique and the result in [Chen and Huang 1982] and the method of continuity with respect to the contact angle. Later, Shigeru Sakaguchi [1989] showed that the solution to

$$\begin{cases} \operatorname{div} \dfrac{Du}{\sqrt{1+|Du|^2}} = 2H & \text{in } \Omega, \\ u = 0, & \text{on } \partial\Omega, \end{cases} \tag{1-12}$$

or

$$\begin{cases} \operatorname{div} \dfrac{Du}{\sqrt{1+|Du|^2}} = 2H & & \text{in } \Omega, \\ \dfrac{Du}{\sqrt{1+|Du|^2}} \cdot \nu = \cos\gamma, & \gamma \in \left(0, \frac{\pi}{2}\right) & \text{on } \partial\Omega, \end{cases} \tag{1-13}$$

has only one critical point under the hypothesis of the existence of the solution over a bounded convex domain $\Omega \subset \mathbb{R}^2$.

Another motivation for studying uniqueness of critical points for solutions to (1-10) is from a recent paper [Albujer et al. 2015]. As we know, CMC spacelike hypersurfaces are very different from those in Euclidean space. For example, Corollary 12.1.8 in [López 2013] tells us any compact spacelike surface immersed in $\mathbb{L}^3$ spanning a plane simple closed curve is a graph over a spacelike plane, which is not true in $\mathbb{R}^3$. Therefore, up to an isometry, we only need to consider the solution to the Dirichlet problem (1-10). Recently, Alma L. Albujer, Magdalena Caballero and Rafael López proved the following interesting theorem on the convexity of the solutions to (1-10):

Theorem A [Albujer et al. 2015]. *Let Σ be a spacelike compact surface in $\mathbb{L}^3$ with constant mean curvature $H \neq 0$ (H-surface for short), such that its boundary is a planar curve which is pseudoelliptic. Then Σ has negative Gaussian curvature at all its interior points. In particular, Σ is a convex surface.*

In their paper, they also proved that pseudoelliptic curves are convex and provided an example that shows the assumption on the boundary can not be replaced by convex curves, but they did not show whether there is a critical point of the solution to (1-10) with nonnegative Gaussian curvature over a convex domain, which is a so-called saddle point. In this paper, we will show that the nonexistence of such saddle points is equivalent to the uniqueness of the critical point. Notice that the Gaussian curvature in [Sakaguchi 1989] is different from that in the Theorem A, which is defined in the next section.

Theorem 1.1. *Any solution to (1-10) in a convex domain for $H \neq 0$ has only one critical point.*

The proof of this theorem is based on the idea of [Sakaguchi 1989], which mainly relies on the comparison of a cylinder with the given surface and the continuity method. In the present result, our comparison surface is a connected component of a hyperbolic cylinder, which is an entire graph over $\mathbb{R}^2$ and, in contrast with the

Euclidean case, the existence of the solution for any bounded domain is assured by the necessary and sufficient conditions given in [Bartnik and Simon 1982].

As we said before, Theorem 1.1 can be used to derive sharp C^0 and C^1 bounds for the solution to (1-10).

Theorem 1.2. *Let $u \in C^\infty(\bar{\Omega})$ be a solution to (1-10) over a strictly convex domain Ω for $H > 0$ and K be the curvature of the boundary $\partial\Omega$ with respect to the inner normal direction. Then*

$$\max_{\bar{\Omega}} |Du|^2 = \max_{\partial\Omega} |Du|^2 \le \frac{H^2}{H^2 + K_{\min}^2},$$

$$-\frac{1}{H}\left(\frac{\sqrt{H^2 + K_{\min}^2}}{K_{\min}} - 1\right) \le \quad \min_{\Omega} u \quad \le -\frac{1}{H}\left(\frac{\sqrt{H^2 + K_{\max}^2}}{K_{\max}} - 1\right) \tag{1-14}$$

where $K_{\min} = \min_{\partial\Omega} K$, $K_{\max} = \max_{\partial\Omega} K$, and one of the equality signs holds if and only if the boundary $\partial\Omega$ is a circle.

At this point, we should give a remark. When $H \neq 0$ and Ω is a round disc of radius R (which is centered at the origin), then

$$u(x, y) = \sqrt{x^2 + y^2 + \frac{1}{H^2}} - \sqrt{R^2 + \frac{1}{H^2}}, \tag{1-15}$$

whose graph is a so-called hyperbolic cap [López 2013].

This article is organized as follows. In Section 3, we will investigate the local behavior of the difference of two spacelike graphs in a neighborhood of a second contact point. In Section 4, we will prove a necessary and sufficient condition for the uniqueness of the minimal point of the solution to (1-10), which is a key step in the proof of Theorem 1.1 in Section 5. In the end, based on the uniqueness of the critical point, we will prove a minimum principle and use it to get the sharp estimates in Theorem 1.2.

2. Notions and local comparison technique

For easier reading, let us recall some background knowledge of Lorentzian geometry. More details can be found in [López 2013]. Let $\mathbb{L}^3$ be the 3-dimensional Lorentz–Minkowski space, that is $\mathbb{R}^3$ endowed with the flat Lorentzian metric

$$ds^2 = dx_1^2 + dx_2^2 - dx_3^2,$$

where (x_1, x_2, x_3) are the canonical coordinates in $\mathbb{R}^n$. The nondegenerate metric of index one classifies the vectors of $\mathbb{R}^3$ into three types.

Definition 2.1 [López 2013]. A vector $v \in \mathbb{L}^3$ is said to be:

(1) spacelike if $\langle v, v\rangle > 0$ or $v = 0$;

(2) timelike if $\langle v, v\rangle < 0$;

(3) lightlike if $\langle v, v\rangle = 0$ and $v \neq 0$.

The modulus of v is $|v| = \sqrt{|\langle v, v\rangle|}$.

Definition 2.2 [López 2013]. An immersed surface Σ in $\mathbb{L}^3$ is called spacelike if the induced metric on Σ is positive definite.

Given a spacelike immersed surface Σ, by Proposition 12.1.5 in [López 2013], Σ is orientable. We can choose Σ to be future-oriented, which means the unit normal vector field N satisfies $\langle N, e_3\rangle > 0$. Here $e_3 = (0, 0, 1)$. Let $\bar{\nabla}$ and ∇ denote the Levi-Civita connection in $\mathbb{L}^3$ and Σ, respectively. If $X, Y \in \mathfrak{X}(\Sigma)$, the Gauss and Weingarten formulae are

$$\bar{\nabla}_X Y = \nabla_X Y + \sigma(X, Y) = \nabla_X Y - \langle AX, Y\rangle N \quad \text{and} \quad AX = -\bar{\nabla}_X N, \tag{2-1}$$

respectively, where σ is the second fundamental form and $A : \mathfrak{X}(\Sigma) \to \mathfrak{X}(\Sigma)$ stands for the shape operator of Σ with respect to N. The mean curvature and the Gaussian curvature are defined by

$$H = -\tfrac{1}{2}\operatorname{trace}(A) = -\tfrac{1}{2}(\kappa_1 + \kappa_2) \quad \text{and} \quad K = -\det(A) = -\kappa_1\kappa_2. \tag{2-2}$$

Let $u \in C^2(\Omega)$ be a function defined on a domain $\Omega \in \mathbb{R}^2$ and consider the surface $\Sigma_u = (x, y, u(x, y))$. The coefficients of the first fundamental form are

$$E = 1 - u_x^2, \quad F = -u_x u_y \quad \text{and} \quad G = 1 - u_y^2. \tag{2-3}$$

Thus $EG - F^2 = 1 - u_x^2 - u_y^2 = 1 - |\nabla u|^2$ and since the immersion is spacelike, $|\nabla u|^2 < 1$ on Ω. The future-directed normal is given by

$$N(x, y, u(x, y)) = \frac{(u_x, u_y, 1)}{\sqrt{1 - |\nabla u|^2}} = \frac{(\nabla u, 1)}{\sqrt{1 - |\nabla u|^2}}. \tag{2-4}$$

With this normal, the mean curvature H and Gaussian curvature K satisfy

$$\operatorname{div} \frac{\nabla u}{\sqrt{1 - |\nabla u|^2}} = 2H \quad \text{and} \quad K = -\frac{u_{xx}u_{yy} - u_{xy}^2}{(1 - |\nabla u|^2)^2}, \tag{2-5}$$

respectively, where div is the Euclidean divergence in $\mathbb{R}^2$.

As mentioned previously, every compact spacelike surface Σ in $\mathbb{L}^3$ with simple closed boundary contained in a hyperplane can be regarded as the graph of a solution $u(x, y)$ to (1-10). There are more interesting facts on compact spacelike surfaces in $\mathbb{L}^3$ with constant mean curvature spanning a given boundary curve (see [López 2013]).

From now on, we assume u to be a solution to (1-10) with $H > 0$ in a convex domain Ω. For $H < 0$, we can consider $-u$ and our theorem still holds. By the maximum principle, u has a interior minimal point, which is a point of nonpositive Gaussian curvature.

In the rest of this section, based on the local comparison technique found in [Chen and Huang 1982], we will investigate the local behavior of the difference of two spacelike graphs in a neighborhood of the point where they have the second contact.

Lemma 2.3. *Let $u(x, y)$, $v(x, y)$ satisfy the same spacelike constant mean curvature equation (the first equations in (1-10) or (2-5)). Without loss of generality, we assume that u, v have a second order contact at $P_0 = (x_0, y_0, u(x_0, y_0))$ with $(x_0, y_0) = (0, 0)$. Then by changing coordinates from (x, y) to (ξ, η) linearly, the difference $u - v$ around $(\xi, \eta) = (0, 0) = (x, y)$ is given by*

$$u - v = \mathbb{R}\Diamond(\lambda \cdot (\xi + \eta i)^n + o(\xi^2 + \eta^2)^{\frac{n}{2}}), \tag{2-6}$$

where $n \geq 3$, λ is a complex number and $\xi + \eta i$ is the complex coordinate.

Proof. Let $w = u - v$. Since u and v solve the same constant mean curvature equation, we have

$$\begin{aligned} 0 &= (1 - u_x^2 - u_y^2)(u_{xx} + u_{yy}) + (u_x^2 u_{xx} + u_y^2 u_{yy} + 2u_x u_y u_{xy}) \\ &\quad - 2H(\sqrt{1 - |Du|^2})^3 \\ &= (1 - u_y^2)u_{xx} + (1 - u_x^2)u_{yy} + 2u_x u_y u_{xy} - 2H\big(\sqrt{1 - u_x^2 - u_y^2}\big)^3, \end{aligned} \tag{2-7}$$

$$0 = (1 - v_y^2)v_{xx} + (1 - v_x^2)v_{yy} + 2v_x v_y v_{xy} - 2H\big(\sqrt{1 - v_x^2 - v_y^2}\big)^3. \tag{2-8}$$

Define $r(\tau)$, $s(\tau)$, $t(\tau)$, $p(\tau)$, $q(\tau)$ for $0 \leq \tau \leq 1$ by

$$\begin{aligned} &r(\tau) = (1 - \tau)v_{xx} + \tau u_{xx}, \quad s(\tau) = (1 - \tau)v_{xy} + \tau u_{xy}, \\ &t(\tau) = (1 - \tau)v_{yy} + \tau u_{yy}, \quad p(\tau) = (1 - \tau)v_x + \tau u_x, \\ &q(\tau) = (1 - \tau)v_y + \tau u_y, \end{aligned} \tag{2-9}$$

and consider the function

$$F = F(\tau) = (1 - q^2)r + 2pqs + (1 - p^2)t - 2H\big(\sqrt{1 - p^2 - q^2}\big)^3. \tag{2-10}$$

Then we get

$$\begin{aligned} 0 = F(1) - F(0) &= \int_0^1 \frac{\partial F}{\partial \tau}\, d\tau \\ &= a_{11}w_{xx} + 2a_{12}w_{xy} + a_{22}w_{yy} + b_1 w_x + b_2 w_y, \end{aligned} \tag{2-11}$$

with

$$a_{11}=\int_0^1 (1-q^2)\,d\tau,\quad a_{12}=\int_0^1 pq\,d\tau,\quad a_{22}=\int_0^1 (1-p^2)\,d\tau,$$
$$b_1=-2\int_0^1 \big[(pt-qs)-3H\sqrt{1-p^2-q^2}\,p\big]\,d\tau,$$
$$b_2=-2\int_0^1 \big[(qr-ps)-3H\sqrt{1-p^2-q^2}\,q\big]\,d\tau. \tag{2-12}$$

Since $Dw=0$ at $(0,0)$, there exists a neighborhood, say $O(0,0)$, such that (p,q) stays in the unit ball, i.e., $p^2+q^2<1$ over $O(0,0)$. Therefore, we have

$$\begin{aligned} a_{12}^2=\Big(\int_0^1 pq\,d\tau\Big)^2 &\le \int_0^1 (p^2)\,d\tau\int_0^1 (q^2)\,d\tau\\ &< \int_0^1 (p^2)\,d\tau\int_0^1 (1-p^2)\,d\tau\\ &< \int_0^1 (1-q^2)\,d\tau\int_0^1 (1-p^2)\,d\tau=a_{11}a_{22}. \end{aligned} \tag{2-13}$$

Hence, w satisfies a homogeneous elliptic equation

$$Lw=a_{11}w_{xx}+2a_{12}w_{xy}+a_{22}w_{yy}+b_1w_x+b_2w_y, \tag{2-14}$$

in $O(0,0)$.

Now, we transform (x,y) into (ξ,η) such that $\xi(0,0)=0$ and $\eta(0,0)=0$ and at $(0,0)$

$$Lw=\Big(\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\eta^2}+b_1'\frac{\partial}{\partial\xi}+b_2'\frac{\partial}{\partial\eta}\Big)w. \tag{2-15}$$

Since the coefficient of Lw and w itself are analytic in (x,y) as well as in (ξ,η), we have the expansion around $(\xi,\eta)=(0,0)$ as follows,

$$\begin{aligned} Lw=\Big\{&(1+\alpha_{11}\xi+\beta_{11}\eta+O(\xi^2+\eta^2))\frac{\partial^2}{\partial\xi^2}+2(\alpha_{12}\xi+\beta_{12}\eta+O(\xi^2+\eta^2))\frac{\partial^2}{\partial\xi\partial\eta}\\ &+(1+\alpha_{22}\xi+\beta_{22}\eta+O(\xi^2+\eta^2))\frac{\partial^2}{\partial\eta^2}+(\gamma_1+\delta_1\xi+\lambda_1\eta+O(\xi^2+\eta^2))\frac{\partial}{\partial\xi}\\ &+(\gamma_2+\delta_2\xi+\lambda_2\eta+O(\xi^2+\eta^2))\frac{\partial}{\partial\eta}\Big\}w. \end{aligned}$$

By Theorem I in [Bers 1955], we know

$$w=w(\xi,\eta)=P_n(\xi,\eta)+o(\xi^2+\eta^2)^{n/2}, \tag{2-16}$$

where $P_n(\xi,\eta)$ is a nonzero harmonic homogeneous polynomial in (ξ,η) of degree n. We know $n\ge 3$, as u and v have a second contact at $(0,0)$. Thus the argument in page 82 of [Axler et al. 2001] tells us

$$P_n(\xi, \eta) = \mathrm{Re}(\lambda \cdot (\xi + \eta i)^n), \tag{2-17}$$

where λ is a complex number. This, together with the expansion above, completes the proof. □

Let $u - v$ to be defined on $D \in \mathbb{R}^2$ and Z be the zero set of $u - v$ extended to the closure $\bar{D}$ of D. By Lemma 2.3, Z divides a neighborhood U of $(0, 0)$ into at least six components on which the sign of $u - v$ alternate. However, Lemma 2.3 does not tell us that $Z \cap U$ is a union of smooth arcs intersecting at $(0, 0)$. We do not know if Z may contain cusps at $(0, 0)$. To exclude such irregular possibilities, we need a lemma from Chen and Huang:

Lemma 2.4 [Chen and Huang 1982, Lemma 2]. *Let $f = f(x, y)$ be a nonconstant solution of a homogeneous quasilinear elliptic equation of the form*

$$Lf = a_{11} f_{xx} + 2a_{12} f_{xy} + a_{22} f_{yy} + b_1 f_x + b_2 f_y = 0 \tag{2-18}$$

in Ω having analytic coefficients the a_{ij} and b_k in x, y and involving no zero order term. Then every interior critical point of f is an isolated critical point.

Using the previous two lemmas as well as the implicit function theorem, we see that the zero set $Z \cap U$ of $u - v$ consists of at least three smooth arcs intersecting at $(0, 0)$ and dividing U into at least six sectors. Furthermore, the zero set Z is globally a union of smooth arcs.

3. Nonuniqueness of the minimal point

In this section, by using Lemmas 2.3 and 2.4, we will prove a sufficient and necessary condition for the nonuniqueness of minimal points of the solutions v_t ($t \in [0, 1]$) to

$$\begin{cases} \operatorname{div} \dfrac{Dv}{\sqrt{1 - t^2 |Dv|^2}} = 2H, \quad t|Dv| < 1 & \text{in } \Omega, \\ v = 0, & \text{on } \partial\Omega. \end{cases} \tag{3-1}$$

Let $u_t = t v_t$ for $t > 0$. Then u_t satisfies

$$\begin{cases} \operatorname{div} \dfrac{Du}{\sqrt{1 - |Du|^2}} = 2tH, \quad |Du| < 1 & \text{in } \Omega, \\ u = 0, & \text{on } \partial\Omega. \end{cases} \tag{3-2}$$

Proposition 3.1. *There always exists a unique solution v_t to (3-1) satisfying*

$$t|Dv_t| \le 1 - \theta_0 < 1, \quad \textit{in } \bar{\Omega}, \quad \|v_t\|_{C^{2,\alpha}(\bar{\Omega})} \le C, \quad \textit{for all } t \in [0, 1], \tag{3-3}$$

where C, θ_0, α are positive constants independent of t.

Proof. By Theorem 3.6 in [Bartnik and Simon 1982], Theorem 13.8 in [Gilbarg and Trudinger 1983] and Theorem 12.2.2 in [López 2013], there is a unique solution $u_t \in C^{2,\alpha}(\overline{\Omega})$ to the problem (3-2) with

$$|Du_t| < 1 - \theta_0 < 1 \quad \text{in } \overline{\Omega} \qquad \text{and} \qquad \|u_t\|_{C^{2,\alpha}(\overline{\Omega})} \le C, \tag{3-4}$$

where C, θ_0, α are positive constants independent of t.

Put $v_t = t^{-1}u_t$. Then v_t satisfies (3-1). By putting

$$b(x) = (1 - |Du_t|^2)^{-1/2}, \tag{3-5}$$

we regard v_t as a unique solution to the linear elliptic Dirichlet problem:

$$\begin{cases} \operatorname{div}(b(x)Dv_t) = 2H & \text{in } \Omega, \\ v_t = 0 & \text{on } \partial\Omega. \end{cases} \tag{3-6}$$

In view of (3-4), using the Schauder global estimate (see Theorem 6.6 in [Gilbarg and Trudinger 1983]), we get

$$\|v_t\|_{C^{2,\alpha}(\overline{\Omega})} \le C(\sup_{\Omega} |v_t| + 2H). \tag{3-7}$$

Also, it follows from their Theorem 3.7 that

$$\sup_{\Omega} |v_t| \le C. \tag{3-8}$$

Therefore, we get (3-3) for $t \in (0, 1]$. When $t = 0$, (3-1) is a linear problem. Hence there exists a unique solution $v_0 \in C^\infty(\overline{\Omega})$ to (3-1). This completes the proof. □

Before proving the sufficient and necessary condition for nonuniqueness of the minimal point of v_t, we need the following lemmas.

Lemma 3.2. *Let t belong to* $(0, 1]$. *If $Dv_t = 0$ at some point $p \in \Omega$, then the Gaussian curvature $K_t(p)$ of the graph $\Sigma_{v_t} = (x, y, v_t(x, y))$ at p does not vanish.*

Proof. Since t is positive, it suffices to show this for $u_t = tv_t$. Recall that graph of u_t has constant mean curvature tH. Let p be a critical point of u_t with $K_t(p) = 0$.

Consider the upper connected component of a hyperbolic cylinder in $\mathbb{L}^3$, S, with radius $r = 1/(2tH)$, tangent to Σ_{u_t} at p and such that the line generators are parallel to the zero principal curvature direction of Σ_{u_t} at p. Recall that each connected component of a hyperbolic cylinder is an entire graph over $\mathbb{R}^2$ with constant mean curvature $|H| = 1/(2r)$ and zero Gaussian curvature.

In general, the intersection of S and $\mathbb{R}^2$ should be a branch of a hyperbola or two parallel lines. In our case, it should be the latter one, as S touches u_t at its critical point p. Hence, $S \cap \mathbb{R}^2$ divides $\mathbb{R}^2$ into three domains, and suppose that the piece of S with negative height is the graph of a function $v \in C^\infty(\Omega'),\ v < 0$.

Define $D = \Omega \cap \Omega'$. On the one hand, by the convexity of Ω, we see $\partial(\Omega \cap \Omega')$ consists of at most four arcs, each of which belongs to $\partial\Omega$ or $\partial\Omega'$ alternatively. Consider $A = \{(x, y) \in \Omega \cap \Omega' \mid u_t(x, y) > v(x, y)\}$. Since $u_t = 0$ on $\partial\Omega$ and $v = 0$ on $\partial\Omega'$, there are at most two components of A, each of which meets the boundary $\Omega \cap \Omega'$. On the other hand, by the previous construction, u_t and v have a second order contact at p. Lemma 2.3 and Lemma 2.4 tell us A has at least three components, each of which meets $\Omega \cap \Omega'$. Thus we get a contradiction. This completes the proof. □

Now, we see that there is no critical point of v_t with Gaussian curvature vanishing for $t \in (0, 1]$. What about the case of $t = 0$?

Lemma 3.3. *Every critical point p of v_0 is a minimal point, i.e., the Gaussian curvature $K_0(p)$ of the graph Σ_{v_0} is negative at p.*

Proof. Let p be a critical point of v_0. Then $K_0(p) = -((v_0)_{xx}(v_0)_{yy} - (v_0)^2_{xy})$ by the second equation of (2-5). Suppose that $K_0(p) \geq 0$. For simplicity, by translation and rotation of the coordinates, we may assume that $p = (0, 0)$ and $[D_{ij}v_0] = \operatorname{diag}[\lambda_1, \lambda_2]$, where $\lambda_1 + \lambda_2 = 2H > 0$, $\lambda_1 > 0$ and $\lambda_2 \leq 0$. Then $v_0(x, y) = w(x, y) + P(x, y)$, where $w(x, y) = v_0(0, 0) + \frac{1}{2}\lambda_1 x^2 + \frac{1}{2}\lambda_2 y^2$ and $P(x, y)$ is a harmonic function in Ω. Consider

$$A = \{(x, y) \in \Omega \mid P(x, y) > 0\}, \quad B = \{(x, y) \in \Omega \mid P(x, y) < 0\}. \tag{3-9}$$

Since $P(x, y)$ vanishes up to second order derivatives at $(0, 0)$ and $P(x, y)$ is real analytic, it follows from Lemma 2.3 and Lemma 2.4 that both A and B have at least three components, each of which meets the boundary $\partial\Omega$. Put

$$\Omega' = \{(x, y) \in \mathbb{R}^2 \mid w(x, y) < 0\}. \tag{3-10}$$

Since Ω is convex and w is a quadratic function with $\lambda_1 > 0$ and $\lambda_2 \leq 0$, we see that $\partial(\Omega \cap \Omega')$ consists of at most four arcs each of which belongs to $\partial\Omega$ or $\partial\Omega'$ alternatively. Let $A' = \{(x, y) \in \Omega \cap \Omega' \mid P(x, y) > 0\}$. Since $v_0 = 0$ on $\partial\Omega$ and $w = 0$ on $\partial\Omega'$, there are at most two components of A' each of which meets the boundary $\partial(\Omega \cap \Omega')$. This contradicts the fact that both A and B have at least three components which meet the boundary of $\partial\Omega$. This completes the proof. □

Now, we can prove the sufficient and necessary condition for nonuniqueness of the minimal point of v_t.

Theorem 3.4. *Let t belong to* $[0, 1]$. *The solution v_t has more than two minimal points if and only if there exists a saddle point $p \in \Omega$, i.e., $Dv_t(p) = 0$ and $K_t(p) > 0$.*

Proof. It follows from Hopf's boundary point lemma that $Dv_t \cdot \nu$ is positive on $\partial\Omega$. There v_t does not have minimal point on the boundary $\partial\Omega$.

“If” part: Let $p \in \Omega$ be a point with $Dv_t(p) = 0$ and $K_t(p) > 0$. Then there exists an open neighborhood U of p in which the zero set of $\tilde{v}_t = v_t - v_t(p)$ consists of two smooth arcs intersecting at p and divides U into four sections. Consider the open set $E = \{(x, y) \in \Omega \mid \tilde{v}_t > 0\}$. It follows from the maximum principle that each component of E has to meet the boundary $\partial\Omega$. Accordingly, we see that the open set $G = \{(x, y) \in \Omega \mid \tilde{v}_t < 0\}$ has more than two components. This shows that v_t has more than two minimal points.

“Only if” part: Suppose that v_t has more than two minimal points and there is no point p with $Dv_t(p) = 0$ and $K_t(p) > 0$. By Lemma 3.2 and Lemma 3.3, we see that each critical point of v_t is a minimal point. Since Dv_t does not vanish on $\partial\Omega$, then Lemma 3.2 and Lemma 3.3 imply that every critical point of v_t is isolated and the number of critical (minimal) points is finite, say $\{P_1, \dots, P_N\}$. Hence, we have

$$Dv_t(x, y) \neq 0, \quad \text{for all } (x, y) \in \Omega - \{P_1, \dots, P_N\}. \tag{3-11}$$

Put $m_0 = \max\{v_t(P_j) \mid 1 \leq j \leq N\}$. Consider the level set $L_m = \{(x, y) \in \Omega \mid v_t(x, y) < m\}$ for $m_0 < m < 0$. It follows from (3-11) and Theorem 3.1 in [Milnor 1963] that the boundary ∂L_m is a smooth manifold for $m_0 < m < 0$ and $\{\partial L_m\}$ are diffeomorphic to each other. Since $K_t(P_j)$ is negative, if m is near m_0, L_m has more than two components. On the other hand, if m is near 0, ∂L_m is diffeomorphic to $\partial\Omega$ and L_m is connected. This is a contradiction, so the proof is complete. □

Now, Lemma 3.2, Lemma 3.3 and Theorem 3.4 tell us the nonexistence of the critical point described in the first question of the first section is equivalent to the uniqueness for the critical point of the solution to (1-10), which will be proved in the next section.

4. Proof of Theorem 1.1

In view of Lemma 3.2, Lemma 3.3 and Theorem 3.4, it suffices to show that the set of minimal points of the solution consists of only one point. Put $I = [0, 1]$. Divide I into two sets I_1 and I_2 as follows:

$$\begin{aligned} I_1 &= \{t \in I \mid v_t \text{ has only one minimal point in } \Omega\}, \\ I_2 &= \{t \in I \mid v_t \text{ has more than two minimal points in } \Omega\}. \end{aligned} \tag{4-1}$$

Then $I = I_1 + I_2$ and $I_1 \cap I_2 = \varnothing$. Lemma 3.3 and Theorem 3.4 imply that $0 \in I_1$, so I_1 is not empty.

On the one hand, I_2 is open in I. That is, for any $t_0 \in I_2$, there exists a constant $\varepsilon > 0$ such that $(t_0 - \varepsilon, t_0 + \varepsilon) \subset I_2$. If it were not so, we can assume that there exists a sequence of solutions $\{v_{t_n}\}$ with only one minimal point and $t_n \in (t_0 - 1/n, t_0 + 1/n)$ for some positive $t_0 \in I_2$. By Lemma 3.2 and Theorem 3.4, v_{t_n} has only one critical point. By compactness and Lemma 3.2, we can take a subsequence of v_{t_n} such that

$$p_n \to p, \quad Dv_{t_n}(p_n) = 0, \quad K_{t_n}(p_n) < 0, \quad Dv_{t_0}(p) = 0, \quad K_{t_0}(p) < 0. \tag{4-2}$$

Since $t_0 \in I_2$, there exists another point $q \in U(q) \subseteq \Omega$ such that

$$q_n \to q, \quad Dv_{t_n}(q_n) \to Dv_{t_0}(q) = 0. \tag{4-3}$$

By uniqueness of the critical point of v_{t_n}, we can take a subsequence of $\{v_{t_n}\}$ such that v_{t_n} are all monotone in the line $l(p_n, q_n)$. Then there exists a sequence of points $\{s_n \mid s_n \in l(p_n, q_n)\}$ such that

$$|Dv_{t_n}(s_n)| \le |Dv_{t_n}(q_n)| \to 0, \quad |K_{t_n}(s_n)| = \frac{|Dv_{t_n}(q_n)|}{|p_n - q_n|} \to 0. \tag{4-4}$$

Therefore, there should be a point $s \in l(p, q)$ which satisfies

$$Dv_{t_0}(s) = 0, \quad K_{t_0}(s) = 0. \tag{4-5}$$

This is a contradiction with Lemma 3.2.

On the other hand, I_2 is closed in I. In fact, let $\{t_j\}$ be a sequence of points in I_2 such that t_j converges to t_0 as j goes to ∞. Theorem 3.4 and the compactness imply that there exists a subsequence $\{t_k\}$, a sequence $\{p_k\}$ and a point $p \in \Omega$ such that

$$p_k \to p \quad \text{as } k \to \infty, \quad Dv_{t_k}(p_k) = 0, \quad \text{and} \quad K_{t_k}(p_k) > 0. \tag{4-6}$$

By continuity, we have

$$Dv_{t_0}(p) = 0, \quad \text{and} \quad K_{t_0}(p) \ge 0. \tag{4-7}$$

Since $Dv_{t_0} \neq 0$ on $\partial\Omega$, $p \in \Omega$. Therefore it follows from Lemma 3.2 and Lemma 3.3, Theorem 3.4 and (4-7) that $t_0 \in I_2$. This shows that I_2 is closed in I.

Hence, I_2 must be $\varnothing$ or I. Since I_1 is not $\varnothing$, $I_1 = I$. This completes the proof.

5. Sharp C^0 and C^1 estimates

In [Payne and Philippin 1979], the authors derived a maximum principle for a function $\Phi(x; \alpha)$ defined by

$$\Phi(x; \alpha) = \int_0^{q^2} \frac{g(\xi) + 2\xi g'(\xi)}{\rho(\xi)}\, d\xi + \alpha \int_0^u f(\eta)\, d\eta, \tag{5-1}$$

where $g > 0$, $\rho > 0$, f are functions and u satisfies the following elliptic equation:

$$\sum_i (g(q^2)u_i)_i + \rho(q^2) f(u) = 0, \quad q^2 = \sum_i u_i u_i = |Du|^2. \tag{5-2}$$

In our case, we can take $g(\xi) = (1-\xi)^{-1/2}$, $\rho = 1$, $f = -2H$. Then

$$\Phi(x;\alpha) = 2\left(\frac{1}{\sqrt{1-|Du|^2}} - 1 - \alpha H u\right). \tag{5-3}$$

In particular, $\Phi := \Phi(x;1) = 2(1/\sqrt{1-|Du|^2} - 1 - Hu)$.

Theorem 4 in [Payne and Philippin 1979] gives us

$$\sum_{i,j}\left(\delta_{ij} + \frac{u_i u_j}{1-|Du|^2}\right)\Phi_{ij} + \sum_k W_k \Phi_k \geq 0, \tag{5-4}$$

where $W_k s$ are just the components of a vector function uniformly bounded in Ω. It follows that $\Phi(x;1)$ takes its maximum value on $\partial\Omega$. Together with (5-1), we know $\Phi(x;1)$ takes its maximum value where $|Du|^2 = \max_{\partial\Omega}|Du|^2$. It follows that, at any point $x \in \Omega$, we have

$$-Hu \leq \frac{1}{\sqrt{1-\max_{\partial\Omega}|Du|^2}} - \frac{1}{\sqrt{1-|Du|^2}}. \tag{5-5}$$

So, at the critical point, we get

$$-Hu_{\min} \leq \frac{1}{\sqrt{1-q_{\max}^2}} - 1, \tag{5-6}$$

where $u_{\min} = \min_\Omega u$ and $q_{\max} = \max_{\partial\Omega}|Du|$.

Now, we want to derive the upper bound for $|Du|^2_{\max}$. Suppose $\Phi(x;\alpha)$ attains its maximum at $p \in \partial\Omega$. Then $|Du|(p) = q_{\max}$. On the one hand, by the strong maximum principle, we have at p,

$$\frac{\partial\Phi(x;\alpha)}{\partial\nu} = 2\frac{g+2q^2g'}{\rho}u_\nu u_{\nu\nu} + f u_\nu \geq 0, \tag{5-7}$$

where $\partial/\partial\nu$ or a subscript ν denotes the outward directed normal derivative on $\partial\Omega$ and the equality holds if and only if $\Phi(x;\alpha) = \text{constant}$. On the other hand, making use of (5-2) evaluated on $\partial\Omega$, we have

$$(g+2q^2g')u_{\nu\nu} + gKu_\nu + \rho f = 0. \tag{5-8}$$

Together with (5-7), this leads to

$$\frac{\partial\Phi(x;\alpha)}{\partial\nu} = -(2Kgu_\nu^2 + fu_\nu) \geq 0. \tag{5-9}$$

Applying to our case, we get

$$\frac{q_{\max}}{\sqrt{1-q_{\max}^2}} \leq \frac{H}{K(p)} \leq \frac{H}{K_{\min}}. \tag{5-10}$$

So

$$q_{\max}^2 \le \frac{H^2}{H^2 + K_{\min}^2}. \tag{5-11}$$

Therefore, the left inequality in (1-14) follows from (5-6) and (5-11). And the equality holds if and only if the boundary is a circle. In fact, if the equality holds, then $\Phi(x; 1) =$ constant on $\partial\Omega$ from the strong maximum principle. From (5-1), $u_\nu =$ constant on $\partial\Omega$. So $\partial\Omega$ is a circle according to Theorem 2 and Remark 1 in [Serrin 1971]. Conversely, if $\partial\Omega$ is a circle, then the solution u is radially symmetric. So $u_\nu =$ constant on $\partial\Omega$, and then the equality in (5-11) follows from the divergence theorem.

To derive the upper bound of $u_{\min}$ in the same way above, we need a minimum principle for $\Phi(x; 1)$. First, we need the following lemma.

Lemma 5.1 [Payne and Philippin 1979].

$$\sum_{i,j}\Big(\delta_{ij} + \frac{u_i u_j}{1-|Du|^2}\Big)\Phi_{ij}(x,\alpha) + \sum_k \widehat{W_k}\Phi_k(x,\alpha) = 4H^2(\alpha-1)(\alpha-2)\frac{1}{\sqrt{1-|Du|^2}}, \tag{5-12}$$

where $\widehat{W_k}$s are the components of a vector function which is singular at the critical point of u.

From Lemma 5.1 and the Hopf maximum principle, we conclude that $\Phi(x; \alpha)$ takes its minimum value either on the boundary $\partial\Omega$, or at the unique critical point of u in Ω when $\alpha \in [1, 2]$. What if the second alternative happens? We answer this in the following theorem whose Euclidean version was proved by Xi-Nan Ma [2000]:

Theorem 5.2. *Let $u \in C^\infty(\bar\Omega)$ be a solution to (1-10). If $\Phi(x; 1)$ attains its minimum at the unique critical point in Ω, then $\Phi(x; 1)$ is a constant on $\bar\Omega$.*

By Theorem 5.2, we assume $\Phi(x; 1)$ takes its minimum at $p' \in \partial\Omega$, then $|Du|(p') = q_{\min} = \min_{\partial\Omega} |Du|$ and

$$-Hu_{\min} \ge \frac{1}{\sqrt{1-q_{\min}^2}} - 1, \tag{5-13}$$

and

$$\frac{\partial\Phi}{\partial\nu}(p'; 1) \le 0, \tag{5-14}$$

where the equality holds if and only if $\Phi(x;1) = \text{constant}$. As before, one can also get

$$\frac{q_{\min}}{\sqrt{1-q_{\min}^2}} \geq \frac{H}{K(p')} \geq \frac{H}{K_{\max}}. \tag{5-15}$$

So

$$q_{\min}^2 \geq \frac{H^2}{H^2+K_{\max}^2}, \tag{5-16}$$

where the equality holds if and only if the boundary is a circle. Therefore, the right inequality in (1-14) follows from (5-13) and (5-16).

For completeness, we will prove Theorem 5.2 to end this paper. Our proof is similar to that in [Ma 2000] except for the different signs in some places.

Proof of Theorem 5.2. The proof consists of four steps. Assume the unique critical point to be $P \in \Omega$.

Step 1: Derivatives of Φ up to the second order vanish at P. From the proof of Theorem 1.1, we can choose the coordinates at P such that

$$u_1(P) = u_2(P) = 0, \quad u_{11} > 0, \quad u_{22} > 0, \quad u_{12} = 0. \tag{5-17}$$

By direct computation, we have

$$\Phi_1 = 2v^{-\frac{3}{2}} u_i u_{i1} - 2Hu_1 = 0, \tag{5-18}$$

$$\Phi_2 = 2v^{-\frac{3}{2}} u_i u_{i2} - 2Hu_2 = 0, \tag{5-19}$$

$$\begin{aligned}\Phi_{11} &= \tfrac{3}{2}v^{-\frac{5}{2}}(2u_iu_{i1})(2u_ju_{j1}) + 2v^{-\frac{3}{2}}u_{i1}^2 + 2v^{-\frac{3}{2}}u_iu_{i11} - 2Hu_{11}\\ &= 2u_{11}^2 - 2Hu_{11},\end{aligned} \tag{5-20}$$

$$\begin{aligned}\Phi_{12} &= \tfrac{3}{2}v^{-\frac{5}{2}}(2u_iu_{i1})(2u_ju_{j2}) + 2v^{-\frac{3}{2}}u_{i1}u_{i2} + 2v^{-\frac{3}{2}}u_iu_{i12} - 2Hu_{12}\\ &= 0,\end{aligned} \tag{5-21}$$

$$\begin{aligned}\Phi_{22} &= \tfrac{3}{2}v^{-\frac{5}{2}}(2u_iu_{i2})(2u_ju_{j2}) + 2v^{-\frac{3}{2}}u_{i2}^2 + 2v^{-\frac{3}{2}}u_iu_{i22} - 2Hu_{22}\\ &= 2u_{22}^2 - 2Hu_{22},\end{aligned} \tag{5-22}$$

where $v = 1 - |Du|^2$. Since Φ attains its minimum at P, we get

$$\Phi_{11}(P)\Phi_{22}(P) - \Phi_{12}(P) \geq 0. \tag{5-23}$$

Together with (5-17), we know

$$u_{11}(P) = u_{22}(P) = H, \tag{5-24}$$

and

$$\Phi_{11}(P) = \Phi_{22}(P) = 0. \tag{5-25}$$

Step 2: Derivatives of Φ up to the fifth order vanish at P. First we claim

$$\Phi_{x_1^k x_2^{3-k}}(P) = 0, \quad k = 0, 1, 2, 3. \tag{5-26}$$

By (5-17), (5-24) and direct calculations, we have

$$\begin{aligned} \Phi_{x_1^3}(P) &= 4Hu_{x_1^3}, & \Phi_{x_1^2 x_2}(P) &= 4Hu_{x_1^2 x_2}, \\ \Phi_{x_1 x_2^2}(P) &= 4Hu_{x_1 x_2^2}, & \Phi_{x_2^3}(P) &= 4Hu_{x_2^3}. \end{aligned} \tag{5-27}$$

Now, by differentiating (1-10), we obtain

$$u_{x_1^3} = -u_{x_1^2 x_2} \quad \text{and} \quad u_{x_1 x_2^2} = -u_{x_2^3}. \tag{5-28}$$

Together with (5-18), (5-19), (5-21) and (5-25), we can expand Φ in a neighborhood of P:

$$\Phi(x_1, x_2; 1) - \Phi(P; 1) = \frac{r^3}{3!}\big(\Phi_{x_1^3}(P)\cos(3\phi) + \Phi_{x_1^2 x_2}(P)\sin(3\phi)\big) + O(r^4), \tag{5-29}$$

where (r, ϕ) are polar coordinates. Suppose

$$\sqrt{(\Phi_{x_1^3}(P))^2 + (\Phi_{x_1^2 x_2}(P))^2} \neq 0. \tag{5-30}$$

Then (5-29) becomes

$$\Phi(x_1, x_2; 1) - \Phi(P; 1) = A_3(P)\cos[3\phi - \beta_3]r^3 + O(r^4), \tag{5-31}$$

with

$$A_3(P) = \frac{\sqrt{(\Phi_{x_1^3}(P))^2 + (\Phi_{x_1^2 x_2}(P))^2}}{3!}, \quad \cos\beta_3 = \frac{\Phi_{x_1^3}(P)}{\sqrt{(\Phi_{x_1^3}(P))^2 + (\Phi_{x_1^2 x_2}(P))^2}}, \tag{5-32}$$

$$\sin\beta_3 = \frac{\Phi_{x_1^2 x_2}(P)}{\sqrt{(\Phi_{x_1^3}(P))^2 + (\Phi_{x_1^2 x_2}(P))^2}}.$$

From (5-31) we conclude that Φ has at least three nodal lines forming equal angles at P, but Lemma 5.1 tells us that Φ takes its minimum value only on $\partial\Omega$ or at P, which is a contradiction. Thus $A_3(P) = 0$. That is,

$$\Phi_{x_1^k x_2^{3-k}}(P) = 0, \quad k = 0, 1, 2, 3, \tag{5-33}$$

and

$$u_{x_1^k x_2^{3-k}}(P) = 0, \quad k = 0, 1, 2, 3. \tag{5-34}$$

Using a similar argument we can show

$$\begin{aligned} 0 &= \Phi_{x_1^4}(P) = 6H(u_{x_1^4}(P) + 3H^3) \\ &= -\Phi_{x_1^2x_2^2}(P) = 6H(u_{x_1^2x_2^2}(P) + H^3) \\ &= \Phi_{x_2^4}(P) = 6H(u_{x_2^4}(P) + 3H^3), \end{aligned} \tag{5-35}$$

$$0 = \Phi_{x_1^3x_2}(P) = 6Hu_{x_1^3x_2}(P) = -\Phi_{x_1x_2^3}(P) = 6Hu_{x_1x_2^3}(P), \tag{5-36}$$

$$\begin{aligned} &u_{x_1^4}(P) = u_{x_2^4}(P) = -3H^3, \quad u_{x_1^3x_2}(P) = u_{x_1x_2^3}(P) = 0, \\ &u_{x_1^2x_2^2}(P) = -H^3, \end{aligned} \tag{5-37}$$

$$\Phi_{x_1^kx_2^{5-k}}(P) = u_{x_1^kx_2^{5-k}}(P) = 0, \quad k = 0, 1, 2, 3, \tag{5-38}$$

and

$$\Phi_{x_1^5}(P) = -\Phi_{x_1^3x_2^2}(P) = \Phi_{x_1x_2^4}(P), \quad \Phi_{x_1^4x_2}(P) = -\Phi_{x_1^2x_2^3}(P) = \Phi_{x_2^5}(P). \tag{5-39}$$

Step 3: Now we assume all derivatives of Φ up to the n-th order vanish at P, where $n \geq 5$. Using the same argument as in the previous step, we have the following relations.

If $n = 2l$, $l \geq 3$. Then

$$u_{x_1^mx_2^{k-m}}(P) = u_{x_1^{k-m}x_2^m}(P) \qquad \text{for any } m = 0, 1, 2, \dots, k, \text{ if } k = 5, 6, 8, \dots, 2l, \tag{5-40}$$

$$u_{x_1^mx_2^{k-m}}(P) = 0 \qquad \text{for any } m = 0, 1, 2, \dots, k, \text{ if } k = 5, 7, 9, \dots, 2l-1, \tag{5-41}$$

$$u_{x_1^mx_2^{2p-m}}(P) = 0 \qquad \text{for any } m = 1, 3, 5, \dots, 2p-1, \text{ if } p = 3, 4, 5, \dots, l, \tag{5-42}$$

and

$$u_{x_1^{2p}}(P) = (-1)^{p+1}(2p-1)[(2p-3)(2p-5)\cdots 1]^2 H^{2p-1} \tag{5-43}$$

for any $p = 3, 4, 5, \dots, l$. When l is even, we have for any $p = 4, 6, 8, \dots, l$

$$\frac{u_{x_1^{2p}}}{u_{x_1^{2p-2}x_2^2}}(P) = 2p-1, \quad \frac{u_{x_1^{2p-2}x_2^2}}{u_{x_1^{2p-4}x_2^4}}(P) = \frac{2p-3}{3}, \cdots, \frac{u_{x_1^{p+2}x_2^{p-2}}}{u_{x_1^px_2^p}}(P) = \frac{p+1}{p-1}, \tag{5-44}$$

and for any $p = 3, 5, 7, \dots, l-1$, we have

$$\frac{u_{x_1^{2p}}}{u_{x_1^{2p-2}x_2^2}}(P) = 2p-1, \quad \frac{u_{x_1^{2p-2}x_2^2}}{u_{x_1^{2p-4}x_2^4}}(P) = \frac{2p-3}{3}, \cdots, \frac{u_{x_1^{p+3}x_2^{p-3}}}{u_{x_1^{p+1}x_2^{p-1}}}(P) = \frac{p+2}{p-2}. \tag{5-45}$$

When l is odd, we have similar relations to (5-44) and (5-45).

If $n=2l+1,\ l\geq 2$, by a similar argument we have (5-40)–(5-45) and

$$u_{x_1^m x_2^{2l+1-m}}(P)=0, \quad \text{for any } m=0,1,2,\dots,2l+1. \tag{5-46}$$

Step 4: Derivatives of Φ of order n+1 vanish at P. We divide this step into two parts according to whether n is odd or even.

Case A: If $n=2l,\ l\geq 3$. By the inductive assumption, we have

$$v_{x_1^m x_2^{k-m}}(P)=0 \quad \text{for any } m=0,1,2,\dots,k, \quad \text{if } k=1,3,5,\dots,n-1. \tag{5-47}$$

Then for any $m=0,1,2,\dots,n+1$,

$$\begin{aligned}(2v^{-\frac{1}{2}})_{x_1^m x_2^{n+1-m}}(P)&=-v^{\frac{3}{2}}v_{x_1^m x_2^{n+1-m}}(P)\\&=2v^{-\frac{3}{2}}((n+1-m)Hu_{x_1^m x_2^{n+1-m}}+mHu_{x_1^m x_2^{n+1-m}})\\&=2(n+1)Hu_{x_1^m x_2^{n+1-m}}.\end{aligned} \tag{5-48}$$

So

$$\Phi_{x_1^m x_2^{n+1-m}}(P)=2nHu_{x_1^m x_2^{n+1-m}}(P). \tag{5-49}$$

Now, by differentiating (1-10), we obtain

$$u_{x_1^m x_2^{n+1-m}}(P)=-u_{x_1^{m+2} x_2^{n-1-m}}(P), \quad \text{for } m=0,1,2,\dots,n+1. \tag{5-50}$$

Then

$$\Phi_{x_1^m x_2^{n+1-m}}(P)=-\Phi_{x_1^{m+2} x_2^{n-1-m}}(P), \quad \text{for } m=0,1,2,\dots,n+1. \tag{5-51}$$

Using Taylor expansion as in Step 2, we can conclude that the derivatives of Φ of order $n+1$ vanish at P.

Case B: If $n=2l+1,\ l\geq 2$, so $n+1=2(l+1)$ is even. We first look for the relations among $\Phi_{x_1^m x_2^{n+1-m}}(P)$, where $m=0,2,4,\dots,n+1$. Through computations, we have

$$\Phi_{x_1^{n+1}}(P)=2nH(u_{x_1^{n+1}}+(-1)^{l+1}(2l+1)[(2l-1)(2l-3)\cdots 1]^2H^{2l+1}), \tag{5-52}$$

and

$$\Phi_{x_1^{n-1}x_2^2}(P)=2nH(u_{x_1^{n-1}x_2^2}+(-1)^{l+1}[(2l-1)(2l-3)\cdots 1]^2H^{2l+1}). \tag{5-53}$$

Now, by differentiating (1-10), we get

$$(\Delta u+u_iu_ju_{ij}v^{-1})_{x_1^{n-1}}(P)=(2Hv_{\frac{1}{2}})_{x_1^{n-1}}(P). \tag{5-54}$$

Together with the relations in Step 3, this leads to

$$u_{x_1^{n+1}}+u_{x_1^{n-1}x_2^2}=(n+1)(-1)^l[(2l-1)(2l-3)\cdots 1]^2H^{2l+1}. \tag{5-55}$$

So

$$\Phi_{x_1^{n+1}}(P) = -\Phi_{x_1^{n-1}x_2^2}(P). \tag{5-56}$$

By a similar argument, it follows that

$$\Phi_{x_1^m x_2^{n+1-m}}(P) = -\Phi_{x_1^{m+2} x_2^{n-1-m}}(P), \quad \text{for } m = 0, 2, 4, \ldots, n+1. \tag{5-57}$$

Then, using the same argument, we have

$$\Phi_{x_1^m x_2^{n+1-m}}(P) = -\Phi_{x_1^{m+2} x_2^{n-1-m}}(P), \quad \text{for } m = 0, 1, 2, \ldots, n+1. \tag{5-58}$$

Now, as in Case A, we can show the derivatives of Φ of order $n+1$ vanish at P.

By the unique continuation of analytic functions, we know if Φ attains its minimum at P, then it must be a constant. This completes the proof. □

Acknowledgements

The author is fully supported by China Scholarship Council(CSC) for visiting University of California, Santa Cruz. The author is also very grateful to his advisors Professor Xi-Nan Ma and Professor Jie Qing for their expert guidance and useful conversations.

References

[Albujer et al. 2015] A. L. Albujer, M. Caballero, and R. López, "Convexity of the solutions to the constant mean curvature spacelike surface equation in the Lorentz–Minkowski space", *J. Differential Equations* **258**:7 (2015), 2364–2374. MR Zbl

[Axler et al. 2001] S. Axler, P. Bourdon, and W. Ramey, *Harmonic function theory*, 2nd ed., Graduate Texts in Mathematics **137**, Springer, New York, 2001. MR Zbl

[Bartnik and Simon 1982] R. Bartnik and L. Simon, "Spacelike hypersurfaces with prescribed boundary values and mean curvature", *Comm. Math. Phys.* **87**:1 (1982), 131–152. MR Zbl

[Bers 1955] L. Bers, "Local behavior of solutions of general linear elliptic equations", *Comm. Pure Appl. Math.* **8**:4 (1955), 473–496. MR Zbl

[Caffarelli and Friedman 1985] L. A. Caffarelli and A. Friedman, "Convexity of solutions of semilinear elliptic equations", *Duke Math. J.* **52**:2 (1985), 431–456. MR Zbl

[Chen 1984] J.-T. Chen, "Uniqueness of minimal point and its location of capillary free surfaces over convex domain", pp. 137–143 in *Variational methods for equilibrium problems of fluids* (Trento, 1983), Astérisque **118**, Société Mathématique de France, Paris, 1984. MR Zbl

[Chen 2014] C. Chen, "On the microscopic spacetime convexity principle of fully nonlinear parabolic equations, I: Spacetime convex solutions", *Discrete Contin. Dyn. Syst.* **34**:9 (2014), 3383–3402. MR Zbl

[Chen and Huang 1982] J. T. Chen and W. H. Huang, "Convexity of capillary surfaces in the outer space", *Invent. Math.* **67**:2 (1982), 253–259. MR Zbl

[Gilbarg and Trudinger 1983] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Grundlehren Math. Wissenschaften **224**, Springer, Berlin, 1983. MR Zbl

[Guan and Ma 2003] P. Guan and X.-N. Ma, "The Christoffel–Minkowski problem, I: Convexity of solutions of a Hessian equation", *Invent. Math.* **151**:3 (2003), 553–577. MR Zbl

[López 2013] R. López, *Constant mean curvature surfaces with boundary*, Springer, Heidelberg, 2013. MR Zbl

[Ma 2000] X.-N. Ma, "Sharp size estimates for capillary free surfaces without gravity", *Pacific J. Math.* **192**:1 (2000), 121–134. MR Zbl

[Ma and Zhang 2013] X.-N. Ma and W. Zhang, "The concavity of the Gaussian curvature of the convex level sets of p-harmonic functions with respect to the height", *Commun. Math. Stat.* **1**:4 (2013), 465–489. MR Zbl

[Milnor 1963] J. Milnor, *Morse theory*, Annals of Mathematics Studies **51**, Princeton Univ. Press, 1963. MR Zbl

[Payne 1973] L. E. Payne, "On two conjectures in the fixed membrane eigenvalue problem", *Z. Angew. Math. Phys.* **24** (1973), 721–729. MR Zbl

[Payne and Philippin 1979] L. E. Payne and G. A. Philippin, "Some maximum principles for nonlinear elliptic equations in divergence form with applications to capillary surfaces and to surfaces of constant mean curvature", *Nonlinear Anal.* **3**:2 (1979), 193–211. MR Zbl

[Philippin 1979] G. A. Philippin, "A minimum principle for the problem of torsional creep", *J. Math. Anal. Appl.* **68**:2 (1979), 526–535. MR Zbl

[Sakaguchi 1989] S. Sakaguchi, "Uniqueness of critical point of the solution to the prescribed constant mean curvature equation over convex domain in $\mathbb{R}^2$", pp. 129–151 in *Recent topics in nonlinear PDE, IV* (Kyoto, 1988), edited by M. Mimura and T. Nishida, North-Holland Math. Stud. **160**, North-Holland, Amsterdam, 1989. MR

[Serrin 1971] J. Serrin, "A symmetry problem in potential theory", *Arch. Rational Mech. Anal.* **43**:4 (1971), 304–318. MR Zbl

Received May 30, 2016. Revised October 12, 2016.

JINGYONG ZHU
SCHOOL OF MATHEMATICAL SCIENCES
UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA
96 JINZHAI ROAD, HEIFEI
230026 ANHUI
CHINA
zjyjj0224@gmail.com

CONTENTS

Volume 288, 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 288 No. 2 June 2017