Pacific Journal of Mathematics

Volume 294 No. 2 June 2018

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 2018 is US $475/year for the electronic version, and $640/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. 294, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.294.257

A POSITIVE MASS THEOREM AND PENROSE INEQUALITY FOR GRAPHS WITH NONCOMPACT BOUNDARY

EZEQUIEL BARBOSA AND ADSON MEIRA

We prove a version of the positive mass theorem for graph hypersurfaces with a noncompact boundary, in Euclidean space. We also prove a Penrose inequality for such hypersurfaces.

1. Introduction

Let (M^n, g) be an asymptotically flat Riemannian manifold. Suppose that the scalar curvature of g is nonnegative, $S_g \geq 0$. The Riemannian positive mass theorem states that, if either $3 \leq n \leq 7$ or $n \geq 3$ and the manifold is spin, the ADM-mass of g is nonnegative: $m_{\text{ADM}} \geq 0$. Moreover, $m_{\text{ADM}} = 0$ if and only if (M^n, g) is isometric to the Euclidean space $(\mathbb{R}^n, \delta)$. Recently, Almaraz, Barbosa and de Lima [Almaraz et al. 2016] defined a kind of ADM-mass for asymptotically flat manifolds with a noncompact boundary, and they proved that, if either $3 \leq n \leq 7$ or if $n \geq 3$ and the manifold is spin, then that ADM-mass is nonnegative, assuming the scalar curvature of the manifold and the mean curvature of the boundary are nonnegative. A similar positive mass theorem for all dimensions and with no spin condition has not been proved yet.

Although graphical hypersurfaces in Euclidean spaces are spin, Lam [2011] used an elementary method for such manifolds — asymptotically flat graphical hypersurfaces with an empty boundary — without invoking the spin structure, and proved the positive mass conjecture for graphical hypersurfaces with compact boundary. A version of the positive mass theorem for manifolds with compact boundary is known as the Penrose inequality. The main goal of this work is to provide an elementary proof for the positive mass theorem for graph hypersurfaces with noncompact boundary in Euclidean spaces, and a kind of Penrose inequality for such hypersurfaces. For more about the positive mass theorem and the Penrose inequality, see [Almaraz et al. 2016; Huang and Wu 2015; Lee and Sormani 2014; Mirandola and Vitório 2015].

Barbosa was supported by FAPEMIG and CNPq grants. Meira was supported by FAPEMIG and CAPES grants.
MSC2010: 53C24, 83C99.
Keywords: graphical hypersurface, ADM mass, scalar curvature, rigidity, asymptotically flat.

Let us be a little bit more precise with respect to the case where the manifold has a noncompact boundary. Let (M^n, g) be an oriented Riemannian manifold with a noncompact boundary Σ and dimension $n \geq 3$. We denote by S_g the scalar curvature of the manifold (M, g). We also assume that Σ is oriented by an outward pointing unit normal vector η, so that its mean curvature is $H_g = \mathrm{div}_g\, \eta$. We say that (M, g) is asymptotically flat with decay rate $\tau > 0$ if there exists a compact subset $K \subset M$ and a diffeomorphism $\Psi : M \setminus K \to \mathbb{R}^n_- \setminus \overline{B}^-_1(0)$ such that the following asymptotic expansion holds as $r \to +\infty$:

$$|g_{ij}(x) - \delta_{ij}| + r|g_{ij,k}(x)| + r^2|g_{ij,kl}(x)| = O(r^{-\tau}). \tag{1-1}$$

Here, $x = (x_1, \dots, x_n)$ is the coordinate system induced by Ψ; $r = |x|$ and g_{ij} are the coefficients of g with respect to x; the comma denotes partial differentiation; $\mathbb{R}^n_- = \{x \in \mathbb{R}^n;\, x_n \leq 0\}$, and $\overline{B}^-_1(0) = \{x \in \mathbb{R}^n_-;\, |x| \leq 1\}$. The subset $M_\infty = M \backslash K$ is called the end of M. In this paper, we use the Einstein summation convention with the index ranges $i, j, \dots = 1, \dots, n$ and $\alpha, \beta, \dots = 1, \dots, n-1$. Observe that along Σ, $\{\partial_\alpha\}_\alpha$ spans $T\Sigma$ while ∂_n points inwards.

The most important example of a manifold in this class is the half-space $\mathbb{R}^n_-$ endowed with the standard flat metric δ; see Figure 1.

Definition 1.1. Suppose that $\tau > \frac{1}{2}(n-2)$ and S_g and H_g are integrable on M and Σ, respectively. In terms of asymptotically flat coordinates as above, the mass of (M, g) is given by

$$\mathfrak{m}_{(M,g)} = \lim_{r \to +\infty} \left\{ \int_{\mathcal{S}^{n-1}_{r,-}} (g_{ij,j} - g_{jj,i})\mu^i \, \mathrm{d}\mathcal{S}^{n-1}_{r,-} + \int_{\mathcal{S}^{n-2}_r} g_{\alpha n} \vartheta^\alpha \, \mathrm{d}\mathcal{S}^{n-2}_r \right\}, \tag{1-2}$$

where $\mathcal{S}^{n-1}_{r,-} \subset M$ is a large coordinate hemisphere of radius r with outward unit normal μ, and ϑ is the outward pointing unit conormal to $\mathcal{S}^{n-2}_r = \partial \mathcal{S}^{n-1}_{r,-}$, oriented as the boundary of the bounded region $\Sigma_r \subset \Sigma$.

Almaraz–Barbosa–de Lima [Almaraz et al. 2016] showed that the limit on the right-hand side of (1-2) exists and its value does not depend on the particular asymptotically flat coordinates chosen. Thus, $\mathfrak{m}_{(M,g)}$ is an invariant of the asymptotic geometry of (M, g). Moreover, they considered the following conjecture:

Conjecture 1.2. *If (M, g) is asymptotically flat with decay rate $\tau > \frac{1}{2}(n-2)$ as above and satisfies $S_g \geq 0$ and $H_g \geq 0$ then $\mathfrak{m}_{(M,g)} \geq 0$, with the equality occurring if and only if (M, g) is isometric to $(\mathbb{R}^n_-, \delta)$. Here, H_g is the mean curvature of the noncompact boundary, related to the outward pointing unit normal vector.*

This conjecture has been confirmed in some special cases in [Escobar 1992; Raulot 2011]. Finally, Almaraz–Barbosa–de Lima [Almaraz et al. 2016] showed that the following result holds.

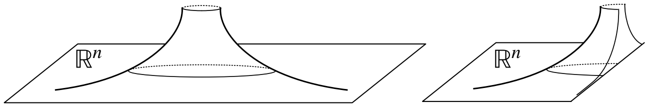


Figure 1. Asymptotically flat manifolds.

Theorem 1.3. *Conjecture 1.2 holds true if either $3 \le n \le 7$ or if $n \ge 3$ and M is spin.*

An immediate consequence of the rigidity statement in Theorem 1.3 is also worth noticing.

Corollary 1.4. *Let (M, g) be as in Theorem 1.3 and assume further that there exists a compact subset $K \subset M$ such that $(M \setminus K, g)$ is isometric to $(\mathbb{R}^n_- \setminus \overline{B}^-_1(0), \delta)$. Then (M, g) is isometric to $(\mathbb{R}^n_-, \delta)$.*

Now, we consider a graphical hypersurface with a noncompact boundary in the Euclidean space.

Definition 1.5. Let $\Omega \subset \mathbb{R}^n$ be a bounded subset. Let $F : \mathbb{R}^n_- \backslash \Omega \to \mathbb{R}^m$, $F(x) = (f^1(x), \dots, f^m(x))$ be a C^2 application. We will denote by $G(F)$ the graph of F. We say that F is asymptotically flat, with order $p > 0$, if the scalar curvature S of the graph of F with the metric of $\mathbb{R}^{n+m}$ is an integrable function over $G(F)$, and if there exists a compact subset $K \subset \mathbb{R}^n$ such that $\Omega \subset K$ and, over $\mathbb{R}^n_- \backslash K$, the partial derivatives $f_i^\alpha = \partial f^\alpha / \partial x_i$, $f_{ij}^\alpha = \partial^2 f^\alpha / (\partial x_i \partial x_j)$ satisfies

$$|f_i^\alpha(x)| = O(|x|^{-p/2}), \quad |f_{ij}^\alpha(x)| = O(|x|^{-p/2-1}), \quad |f_{ijk}^\alpha(x)| = O(|x|^{-p/2-2})$$

for all $\alpha = 1, \dots, m$ and $i, j, k = 1, \dots, n$.

From here, g_f will denote $\delta + df \otimes df$, where δ is the canonical metric of the Euclidean space. Our main result is the following:

Theorem 1.6 (positive mass theorem). *Let $f : \mathbb{R}^n_- \to \mathbb{R}$ be a C^2 function up to boundary, asymptotically flat over $\mathbb{R}^n_- \backslash \Omega$, with order $p > \frac{1}{2}(n-2)$. Let $(\mathbb{R}^n_-, g_f)$ be the graph of f. Suppose that $f_n = 0$ over $\partial \mathbb{R}^n_-$, that $S \in L^1(\mathbb{R}^n_-)$, and $S \ge 0$. We also assume that the mean curvature $\overline{H}$ of the boundary of $(\mathbb{R}^n_-, g_f)$, seen as a submanifold of $(\mathbb{R}^n_-, g_f)$ is such that $\overline{H} \ge 0$ and $\overline{H} \in L^1(\partial \mathbb{R}^n_-)$. Then, the mass of $G(f)$ is nonnegative. Moreover, it is null if and only if $G(f)$ is a half-plane.*

As a consequence of [Lee and Sormani 2014], we obtain the stability of the rigidity supposing that the graph is rotationally symmetric. We can also consider the Penrose inequality for such graphs.

Theorem 1.7 (Penrose inequality). *Let $\Omega \subset \mathbb{R}^n_- \backslash \{x_n = 0\}$, $n \ge 3$, be an open bounded set whose boundary is smooth and mean-convex. Suppose that $\partial\Omega$ is outer-minimizing or each connected component of Ω is star-shaped. Let $f : \mathbb{R}^n_- \backslash \Omega \to \mathbb{R}$*

be a function that is C^2 up to boundary, asymptotically flat, constant over each connected component of $\partial\Omega$ and such that $|Df| \to \infty$ when $x \to \partial\Omega$. We also suppose that $f_n = 0$ over $\partial\mathbb{R}^n_-$, the scalar curvature of the graph is nonnegative, the mean curvatures of the compact boundaries are nonnegative, and that the mean curvature of the noncompact boundary (viewed as a submanifold of the graph) is nonnegative. Then,

$$m_{(M,g)} \geq \frac{1}{2}\left(\frac{|\partial\Omega|}{\omega_{n-1}}\right)^{(n-2)/(n-1)},$$

where $|\partial\Omega|$ is the $(n-1)$-volume of $\partial\Omega$.

2. Proof of the positive mass theorem

We start this section considering a very important proposition.

Proposition 2.1. *Let $\Omega \subset \mathbb{R}^n$ be an open set and $f : \Omega \to \mathbb{R}$ be a C^2 function. Let $X : \Omega \to \mathbb{R}^n$ be the vector field given by*

$$X = \overline{U}(f_i f_{kk} - f_k f_{ik})e_i, \tag{2-1}$$

where $\overline{U} = 1/U$, $U = 1 + \langle Df, Df\rangle$, and $\langle\cdot,\cdot\rangle$ is the Euclidean metric. Consider the function $s : \Omega \to \mathbb{R}$ given by

$$s = \overline{U}(f_{ii} f_{kk} - f_{ik} f_{ik}) - \overline{U}^2 2 f_l f_{li}(f_i f_{kk} - f_k f_{ik}).$$

Then, $s = \operatorname{div} X$ and the scalar curvature of $(G(f), g)$ is s. Here, g is the induced metric and $G(f)$ is the graph of f.

Proof. See [Huang and Wu 2013; Lam 2011; Mirandola and Vitório 2015; Reilly 1973]. □

Now we can prove the following result.

Theorem 2.2. *Let $f : \mathbb{R}^n_- \to \mathbb{R}$ be an asymptotically flat function over $\mathbb{R}^n_- \backslash \Omega$ of class C^2 up to boundary, with order $p > \frac{1}{2}(n-2)$. Let $(\mathbb{R}^n_-, g_f)$ be the graph of f. We suppose the following items:*

- *$f_n = \partial f/\partial x_n \geq 0$ over $\partial\mathbb{R}^n_-$ and $\partial_n = (e_n, f_n)$ is normal to the noncompact boundary (this occurs when $f_n = 0$ over $\partial\mathbb{R}^n_-$, for example);*
- *$S \in L^1(\mathbb{R}^n_-)$ and $S \geq 0$;*
- *The mean curvature $\overline{H}$ of the boundary of $(\mathbb{R}^n_-, g_f)$, viewed as a submanifold of $(\mathbb{R}^n_-, g_f)$ be such that $\overline{H} \geq 0$ (with respect to the unit normal inward pointing vector field) and $\overline{H} \in L^1(\partial\mathbb{R}^n_-)$;*
- *The scalar second fundamental form $\tilde{h}$ (with respect to the unit normal upward pointing vector field) of the boundary of $(\mathbb{R}^n_-, g_f)$, viewed as a submanifold of $(\mathbb{R}^n, \delta) = (\partial\mathbb{R}^n_- \times \mathbb{R}, \delta)$, be such that $\sum_{i=1}^{n-1} \tilde{h}(\bar{\partial}_i, \bar{\partial}_i) \geq 0$ and $\sum_{i=1}^{n-1} \tilde{h}(\bar{\partial}_i, \bar{\partial}_i) \in$*

$L^1(\partial\mathbb{R}^n_-)$. Here, $\bar{e}_i \in \mathbb{R}^{n-1}$ is a canonical vector and $\bar{\partial}_i = (\bar{e}_i, f_i)$ is a tangent vector field.

Then,

$$c(n)m_{(M,g)} = \int_{\mathbb{R}^n_-} S \,\mathrm{d}x_\delta + \int_{\partial\mathbb{R}^n_-} \frac{\overline{H}}{\sqrt{1+(f_n)^2}}(2+|Df|^2)\,\mathrm{d}x_\delta + \int_{\partial\mathbb{R}^n_-} f_n\sqrt{1+|\overline{D}f|^2}\biggl(\sum_{i=1}^{n-1}\tilde{h}(\bar{\partial}_i,\bar{\partial}_i)\biggr)\frac{|Df|^2}{1+|Df|^2}\,\mathrm{d}x_\delta, \tag{2-2}$$

where $\overline{D}f = (f_1, \dots, f_{n-1})$. In particular, the mass $m_{(M,g)}$ is nonnegative.

Proof. Note that $\partial B_r^- = S_r^- \cup D_r$, where $S_r^- = \{x \in \mathbb{R}^n_- \mid \|x\| = r\}$, $D_r = \{x \in \partial\mathbb{R}^n_- \mid \|x\| \le r\}$ and $S_r^{n-2} = \partial D_r$. Remembering that $S = \operatorname{div} X$, we have

$$\begin{aligned}\int_{\mathbb{R}^n_-} S\,\mathrm{d}x_\delta &= \lim_{r\to\infty}\int_{B_r^-} S\,\mathrm{d}x_\delta = \lim_{r\to\infty}\int_{B_r^-}\operatorname{div} X\,\mathrm{d}x_\delta = \lim_{r\to\infty}\int_{\partial B_r^-}\langle X, N\rangle\,\mathrm{d}A_r\\ &= \lim_{r\to\infty}\int_{S_r^-}\Bigl\langle X, \frac{x}{|x|}\Bigr\rangle\,\mathrm{d}\sigma_r + \lim_{r\to\infty}\int_{D_r}\langle X, e_n\rangle\,\mathrm{d}x_\delta.\end{aligned}$$

By hypothesis, $f_i = O(|x|^{-p/2})$ and $f_{ik} = O(|x|^{-p/2-1})$ for all $i, k = 1, \dots, n$. Since $U - 1 = \langle Df, Df\rangle = O(|x|^{-p})$, we have $\lim_{|x|\to\infty} U = 1$, therefore we have $\lim_{|x|\to\infty}\overline{U} = 1$, where $\overline{U} = 1/U$. Therefore, $\overline{U} - 1 = -\overline{U}\langle Df, Df\rangle = O(|x|^{-p})$. With this, we conclude that

$$(\overline{U}-1)(f_i f_{kk} - f_k f_{ik}) = O(|x|^{-2p-1}).$$

Since $p > \frac{1}{2}(n-2)$, we have $2p+1 > n-1 = \dim S_r^-$. Thus,

$$\begin{aligned}\lim_{r\to\infty}\int_{S_r^-}\Bigl|(\overline{U}-1)(f_i f_{kk} - f_k f_{ik})\frac{x^i}{|x|}\Bigr|\,\mathrm{d}\sigma_r &\le \lim_{r\to\infty}\int_{S_r^-} C\cdot|x|^{-2p-1}\,\mathrm{d}\sigma_r\\ &\le C\lim_{r\to\infty} r^{-2p-1}|S_r^-| = 0.\end{aligned}$$

Then

$$\lim_{r\to\infty}\int_{S_r^-}(\overline{U}-1)(f_i f_{kk} - f_k f_{ik})\frac{x^i}{|x|}\,\mathrm{d}\sigma_r = 0.$$

Therefore,

$$\lim_{r\to\infty}\int_{S_r^-}\overline{U}(f_i f_{kk} - f_k f_{ik})\frac{x^i}{|x|}\,\mathrm{d}\sigma_r = \lim_{r\to\infty}\int_{S_r^-}(f_i f_{kk} - f_k f_{ik})\frac{x^i}{|x|}\,\mathrm{d}\sigma_r.$$

Thus,

$$\lim_{r\to\infty}\int_{S_r^-}\Bigl\langle X, \frac{x}{|x|}\Bigr\rangle\,\mathrm{d}\sigma_r = \lim_{r\to\infty}\int_{S_r^-}(f_i f_{kk} - f_k f_{ik})\frac{x_i}{|x|}\,\mathrm{d}\sigma_r.$$

Now, see that $\langle X, e_n\rangle = \overline{U}(f_n f_{kk} - f_k f_{nk}) = \overline{U}\sum_{k=1}^{n-1}(f_n f_{kk} - f_k f_{nk})$, because when $k = n$ the terms are canceled. On the other hand, since $\overline{U} = 1 - |Df|^2/(1+|Df|^2)$, we obtain

$$\langle X, e_n\rangle = \sum_{k=1}^{n-1}(f_n f_{kk} - f_k f_{nk}) - \frac{|Df|^2}{1+|Df|^2}\sum_{k=1}^{n-1}(f_n f_{kk} - f_k f_{nk}).$$

This implies that

$$\begin{aligned}\int_{\mathbb{R}^n_-} S\,\mathrm{d}x_\delta &= \lim_{r\to\infty}\int_{S_r^-}(f_i f_{kk} - f_k f_{ik})\frac{x_i}{|x|}\,\mathrm{d}\sigma_r + \lim_{r\to\infty}\int_{D_r}\sum_{k=1}^{n-1}(f_n f_{kk} - f_k f_{nk})\,\mathrm{d}x_\delta \\ &\qquad - \lim_{r\to\infty}\int_{D_r}\frac{|Df|^2}{1+|Df|^2}\sum_{k=1}^{n-1}(f_n f_{kk} - f_k f_{nk})\,\mathrm{d}x_\delta \\ &= \lim_{r\to\infty}\int_{S_r^-}(g_{ki,k} - g_{kk,i})(\nu_r)^i\,\mathrm{d}\sigma_r \\ &\qquad + \lim_{r\to\infty}\int_{D_r}\big(\mathrm{div}_{\mathbb{R}^{n-1}}(f_n\overline{D}f) - 2\langle\overline{D}f, \overline{D}f_n\rangle\big)\,\mathrm{d}x_\delta \\ &\qquad - \lim_{r\to\infty}\int_{D_r}\frac{|Df|^2}{1+|Df|^2}\sum_{k=1}^{n-1}(f_n f_{kk} - f_k f_{nk})\,\mathrm{d}x_\delta.\end{aligned}$$

Here, $\overline{D}f = (\partial f/\partial x_1, \dots, \partial f/\partial x_{n-1})$, ν_r is the normal vector field to S_r^-, and η_r is the normal vector field to S_r^{n-2}. Using that

$$\int_{D_r}\mathrm{div}_{\mathbb{R}^{n-1}}(f_n\overline{D}f)\,\mathrm{d}x_\delta = \int_{S_r^{n-2}} f_n\langle\overline{D}f, \eta_r\rangle\,\mathrm{d}\sigma_r,$$

we find

$$\begin{aligned}\int_{\mathbb{R}^n_-} S\,\mathrm{d}x_\delta &= \lim_{r\to\infty}\left\{\int_{S_r^-}(g_{ki,k} - g_{kk,i})(\nu_r)^i\,\mathrm{d}\sigma_r + \int_{S_r^{n-2}} f_n f_k(\eta_r)^k\,\mathrm{d}\sigma_r\right\} \\ &\qquad \lim_{r\to\infty}\left\{-2\int_{D_r}\langle\overline{D}f, \overline{D}f_n\rangle\,\mathrm{d}x_\delta - \int_{D_r}(f_n f_{kk} - f_k f_{nk})\frac{|Df|^2}{1+|Df|^2}\,\mathrm{d}x_\delta\right\} \\ &= \lim_{r\to\infty}\left\{\int_{S_r^-}(g_{ki,k} - g_{kk,i})(\nu_r)^i\,\mathrm{d}\sigma_r + \int_{S_r^{n-2}} g_{nk}(\eta_r)^k\,\mathrm{d}\sigma_r\right\} \\ &\qquad \lim_{r\to\infty}\left\{-2\int_{D_r}\langle\overline{D}f, \overline{D}f_n\rangle\,\mathrm{d}x_\delta - \int_{D_r}(f_n f_{kk} - f_k f_{nk})\frac{|Df|^2}{1+|Df|^2}\,\mathrm{d}x_\delta\right\}.\end{aligned}$$

Thus,

$$\begin{aligned}c(n)m_{(M,g)} &= \int_{\mathbb{R}^n_-} S\,\mathrm{d}x_\delta + 2\lim_{r\to\infty}\int_{D_r}\sum_{i=1}^{n-1} f_{ni} f_i\,\mathrm{d}x_\delta \\ &\qquad + \lim_{r\to\infty}\int_{D_r}(f_n f_{kk} - f_k f_{nk})\frac{|Df|^2}{1+|Df|^2}\,\mathrm{d}x_\delta.\end{aligned}$$

Now we will calculate the second fundamental form of the boundary viewed as a submanifold of the graph of f. Note that $N = -\partial_n = -(e_n, f_n)$ is a normal field to the boundary, and moreover, inward pointing and tangent to the graph. We have that

$$\begin{aligned}\overline{\nabla}_{\partial_i}\partial_n &= (0, \dots, 0, \partial_i f_n) = (0, \dots, 0, \langle Df_n, \partial_i\rangle) \\ &= (0, \dots, 0, \langle (f_{n1}, \dots, f_{nn}, 0), (e_i, f_i)\rangle) = (0, \dots, 0, f_{ni}).\end{aligned}$$

Since $\partial_j = (e_j, f_j)$, we have that $\langle \overline{\nabla}_{\partial_i}\partial_n, \partial_j\rangle = f_{ni} f_j$. Thus,

$$\langle \overline{\nabla}_{\partial_i} N, \partial_j\rangle = -\langle \overline{\nabla}_{\partial_i}\partial_n, \partial_j\rangle = -f_{ni} f_j.$$

Here we used the fact that $\langle \overline{\nabla}_{\partial_i} N, \partial_j\rangle = -\langle N, II(\partial_i, \partial_j)\rangle$, where II is the second fundamental form of the boundary viewed as a submanifold of the graph, and we find that $\langle II(\partial_i, \partial_j), N)\rangle = f_{ni} f_j$. Therefore, denoting the scalar second fundamental form of the boundary viewed as a submanifold of the graph by $\bar{h}$, we find

$$\bar{h}(\partial_i, \partial_j) = \langle II(\partial_i, \partial_j), N/|N|\rangle = \frac{f_{ni} f_j}{\sqrt{1 + (f_n)^2}}.$$

With this, we see that the mean curvature of $\partial G(f)$ viewed as a submanifold of $G(f)$ is

$$\begin{aligned}\overline{H} &= \sum_{i,j=1}^{n-1} g^{ij}\bar{h}(\partial_i, \partial_j) = \sum_{i,j=1}^{n-1}\left(\delta_{ij} - \frac{f_i f_j}{1 + |Df|^2}\right)\frac{f_{ni} f_j}{\sqrt{1 + (f_n)^2}} \\ &= \sum_{i,j=1}^{n-1}\left(\frac{\delta_{ij} f_{ni} f_j}{\sqrt{1 + (f_n)^2}} - \frac{f_i f_j f_{ni} f_j}{(1 + |Df|^2)\sqrt{1 + (f_n)^2}}\right) \\ &= \sum_{i=1}^{n-1}\frac{f_{ni} f_i}{\sqrt{1 + (f_n)^2}} - \sum_{i,j=1}^{n-1}\frac{f_{ni} f_i (f_j)^2}{(1 + |Df|^2)\sqrt{1 + (f_n)^2}} \\ &= \sum_{i=1}^{n-1}\frac{f_{ni} f_i}{\sqrt{1 + (f_n)^2}} - \sum_{i=1}^{n-1}\frac{f_{ni} f_i}{\sqrt{1 + (f_n)^2}}\sum_{j=1}^{n-1}\frac{(f_j)^2}{1 + |Df|^2} \\ &= \sum_{i=1}^{n-1}\frac{f_{ni} f_i}{\sqrt{1 + (f_n)^2}}\left[1 - \sum_{j=1}^{n-1}\frac{(f_j)^2}{1 + |Df|^2}\right].\end{aligned}$$

Hence,

$$\sum_{i=1}^{n-1} f_{ni} f_i = \frac{\overline{H}\sqrt{1 + (f_n)^2}}{1 - \sum_{j=1}^{n-1}(f_j)^2/(1 + |Df|^2)}.$$

Therefore,

$$
\begin{aligned}
c(n)m_{(M,g)}&\\
&= \int_{\mathbb{R}^n_-} S\,\mathrm{d}x_\delta + 2\lim_{r\to\infty}\int_{D_r} \frac{\overline{H}\sqrt{1+(f_n)^2}}{1-\sum_{j=1}^{n-1}(f_j)^2/(1+|Df|^2)}\,\mathrm{d}x_\delta\\
&\qquad\qquad + \lim_{r\to\infty}\int_{D_r}(f_n f_{kk} - f_k f_{nk})\frac{|Df|^2}{1+|Df|^2}\,\mathrm{d}x_\delta\\
&= \int_{\mathbb{R}^n_-} S\,\mathrm{d}x_\delta + \lim_{r\to\infty}\int_{D_r} \frac{\overline{H}\sqrt{1+(f_n)^2}}{1-\sum_{j=1}^{n-1}(f_j)^2/(1+|Df|^2)}\left\{2-\frac{|Df|^2}{1+|Df|^2}\right\}\mathrm{d}x_\delta\\
&\qquad\qquad + \lim_{r\to\infty}\int_{D_r} f_n f_{kk}\frac{|Df|^2}{1+|Df|^2}\,\mathrm{d}x_\delta.
\end{aligned}
$$

By hypothesis, $f_n \geq 0$ over $\partial\mathbb{R}^n_-$. Also, $\sum_{k=1}^{n-1} f_{kk} = \sqrt{1+|D\bar{f}|^2}\sum_{i=1}^{n-1}\tilde{h}(\bar{\partial}_i, \bar{\partial}_i)$. In this way,

$$
\begin{aligned}
c(n)m_{(M,g)} = \int_{\mathbb{R}^n_-} S\,\mathrm{d}x_\delta &+ \int_{\partial\mathbb{R}^n_-}\frac{\overline{H}}{\sqrt{1+(f_n)^2}}(2+|Df|^2)\,\mathrm{d}x_\delta\\
&+ \int_{\partial\mathbb{R}^n_-} f_n\sqrt{1+|D\bar{f}|^2}\left(\sum_{i=1}^{n-1}\tilde{h}(\bar{\partial}_i,\bar{\partial}_i)\right)\frac{|Df|^2}{1+|Df|^2}\,\mathrm{d}x_\delta \geq 0.
\end{aligned}
$$

In order to conclude, we will show that $\sum_{k=1}^{n-1} f_{kk} = \sqrt{1+|D\bar{f}|^2}\sum_{i=1}^{n-1}\tilde{h}(\bar{\partial}_i, \bar{\partial}_i)$. Viewed as a submanifold of $\partial\mathbb{R}^n_- \times \mathbb{R}$, the boundary is the graph of $\bar{f} = f|_{\partial\mathbb{R}^n_-}$. Thus, $\bar{\partial}_i = (\bar{e}_i, f_i)$, $i = 1, \dots, n-1$, are tangent vector fields. Here $\bar{e}_i \in \mathbb{R}^{n-1}$. Moreover, $\bar{\eta} = (-D\bar{f}, 1)$ is a normal field. We have

$$
\overline{\nabla}_{\bar{\partial}_i}\bar{\eta} = (-\bar{\partial}_i\bar{f}_1, \dots, -\bar{\partial}_i\bar{f}_{n-1}, 0) = (-\langle D\bar{f}_1, \bar{\partial}_i\rangle, \dots, -\langle D\bar{f}_{n-1}, \bar{\partial}_i\rangle, 0).
$$

Using that

$$
D\bar{f}_j = (\bar{f}_{j1}, \dots, \bar{f}_{j(n-1)}) = (\bar{f}_{j1}, \dots, \bar{f}_{j(n-1)}, 0)
$$

and $\bar{\partial}_i = (\bar{e}_i, \bar{f}_i)$, we obtain that $-\langle D\bar{f}_j, \bar{\partial}_i\rangle = \bar{f}_{ji}$. Thus

$$
\overline{\nabla}_{\bar{\partial}_i}\bar{\eta} = (-\bar{f}_{1i}, \dots, -\bar{f}_{(n-1)i}, 0) = (-D\bar{f}_i, 0).
$$

With this, $\langle\overline{\nabla}_{\bar{\partial}_i}\bar{\eta}, \bar{\partial}_j\rangle = -\bar{f}_{ji}$. Using the Weingarten equation, we find

$$
\langle\widetilde{II}(\bar{\partial}_i, \bar{\partial}_j), \bar{\eta}\rangle = \bar{f}_{ji} = f_{ji}.
$$

Therefore,

$$
\sum_{i=1}^{n-1}\tilde{h}(\bar{\partial}_i, \bar{\partial}_i) = \sum_{k=1}^{n-1}\left\langle\widetilde{II}(\bar{\partial}_k, \bar{\partial}_k), \frac{\bar{\eta}}{|\bar{\eta}|}\right\rangle = \sum_{k=1}^{n-1}\frac{f_{kk}}{\sqrt{1+|D\bar{f}|^2}}. \qquad \square
$$

In the next theorem, we will use a doubling argument to produce an asymptotically flat manifold without a compact boundary; more specifically, given $f : \mathbb{R}^n_- \to \mathbb{R}$, we will consider $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ given by $\tilde{f}(x_1, \dots, x_n) = f(x_1, \dots, x_n)$ if $x_n \le 0$, and $\tilde{f}(x_1, \dots, x_n) = f(x_1, \dots, -x_n)$ if $x_n > 0$. Then, using (2-2) and some results and arguments of [Huang and Wu 2013], we obtain the following result:

Theorem 2.3. *Let $f : \mathbb{R}^n_- \to \mathbb{R}$ be a C^{n+1} function up to boundary, asymptotically flat, over $\mathbb{R}^n_- \backslash \Omega$, with order $p > \frac{1}{2}(n-2)$ and such that $\tilde{f}$ defined as above is C^{n+1}. Let $(\mathbb{R}^n_-, g_f)$ be the graph of f. We suppose that $f_n = 0$ over $\partial \mathbb{R}^n_-$, that $S \in L^1(\mathbb{R}^n_-)$ and $S \ge 0$. We also suppose that the mean curvature $\overline{H}$ of the boundary of $(\mathbb{R}^n_-, g_f)$, viewed as a submanifold of $(\mathbb{R}^n_-, g_f)$ is such that $\overline{H} \ge 0$ and $\overline{H} \in L^1(\partial \mathbb{R}^n_-)$. If the mass of $G(f)$ is null, then $G(f)$ is a half-plane.*

Proof. We will assume that f is asymptotic to $\{x_{n+1} = 0, x_n \le 0\}$ and that $f \ne 0$; if not the result will be trivially true. Consider $\tilde{f}$ and note that $\tilde{f}$ is asymptotically flat, and its graph has integrable and nonnegative scalar curvature. By [Huang and Wu 2013, Theorem 4], we can suppose that the mean curvature of the graph of $\tilde{f}$ is nonnegative, $H(\tilde{f}) \ge 0$, with respect to ν, where ν is the vector field

$$\frac{(-D\tilde{f}, 1)}{\sqrt{1 + |D\tilde{f}|^2}}$$

(because we can reflect the graph over $\{x_{n+1} = 0\}$). Let B_r be an open ball in $\mathbb{R}^n$ centered at the origin of radius r; by [Huang and Wu 2013, Lemma 3.10],

$$\max_{\overline{B}_{r_2} \backslash B_{r_1}} \tilde{f} = \max_{\partial B_{r_2}} \tilde{f} \quad \forall\, r_2 > r_1 > 0.$$

Because $\max_{\partial B_{r_2}} \tilde{f} \to 0$ when $r_2 \to \infty$ (since $\tilde{f}$ is asymptotic to $\{x_{n+1} = 0\}$), we conclude that $\tilde{f} \le 0$ outside of B_{r_1}. Moreover, applying the strong maximum principle to $H(\tilde{f}) \ge 0$, we have $\tilde{f} < 0$ outside of B_{r_1}, unless $\tilde{f} \equiv 0$. In the latter case we can, moreover, conclude that $G(\tilde{f})$ is identical to $\{x_{n+1} = 0\}$, repeating the argument over $B_{r_2} \backslash B_{r_0}$, for $0 < r_0 < r_1$, and making $r_0 \to 0$. With this we conclude that if $\tilde{f} \ne 0$, then $\tilde{f} < 0$, i.e., $G(\tilde{f}) \subset \{x_{n+1} < 0\}$. Therefore, for $\epsilon > 0$ sufficiently small, some connected components of the level set $\{x \in \{x_{n+1} = 0\} \mid \tilde{f}(x) = -\epsilon\}$ lie over $G(\tilde{f})$ and have no boundary. We define $\Sigma_{-\epsilon}$ as being the connected component outermost, i.e., $\Sigma_{-\epsilon}$ is not enclosed by the others components. By Sard's theorem, $\Sigma_{-\epsilon}$ is smooth for almost all ϵ. Moreover, because $\tilde{f}$ tends to zero, for some small $\epsilon > 0$, we see that $\eta = -D\tilde{f}/|D\tilde{f}|$ is the unit normal vector on $\Sigma_{-\epsilon}$ pointing inward to the limited region in $\{x_{n+1} = 0\}$, which is delimited by $\Sigma_{-\epsilon}$. Let $H_{\Sigma_{-\epsilon}}$ be the mean curvature of $\Sigma_{-\epsilon}$ defined by η. Then, using that $H(\tilde{f}) \ge 0$ and $S(\tilde{f}) \ge 0$, by [Huang and Wu 2013, Theorem 2.2] we have $H_{\Sigma_{-\epsilon}} \ge 0$. Since $S(f) \ge 0$ and $c(n)m(g)(G(f)) = 0$, by (2-2) we conclude that $S(f) = 0$; then $S(\tilde{f}) = 0$ and $c(n)m_{(M,g)}(G(\tilde{f})) = 0$. This implies that $G(\tilde{f}) = \{x_{n+1} = 0\}$. If not, by [Huang and Wu 2013, Lemma 5.6] (or more generally by [Mirandola and Vitório 2015, Theorem 1.2]) we will have $H_{\Sigma_{-\epsilon}} = 0$ and therefore $\Sigma_{-\epsilon}$ will be a compact minimal hypersurface without boundary, and embedded in $\mathbb{R}^n$; this will be a contradiction. $\square$

3. Penrose inequality

Let we start this section with a simple proposition that will be very useful in the next one.

Proposition 3.1. *Let $(M, \langle\cdot,\cdot\rangle)$ be an $(n+1)$-Riemannian manifold and $f : M \to \mathbb{R}$ a differentiable function. Let $\Sigma \subset M$ be an embedded hypersurface and ν a unitary normal field to Σ. Suppose that f is constant on Σ; then on Σ we have*

$$\Delta f = \operatorname{Hess} f(\nu, \nu) - H^{\Sigma}\langle Df, \nu\rangle.$$

Here, H^{Σ} is related to ν.

Proof. Given a point $x \in \Sigma$, we have

$$\begin{aligned}\Delta f &= \operatorname{div} Df = \operatorname{div}(\langle Df, \nu\rangle\nu) = \langle D\langle Df, \nu\rangle, \nu\rangle + \langle Df, \nu\rangle \operatorname{div} \nu \\ &= \langle \overline{\nabla}_{\nu} Df, \nu\rangle + \langle Df, \overline{\nabla}_{\nu}\nu\rangle + \langle Df, \nu\rangle \operatorname{div} \nu\end{aligned}$$

Using that Σ is embedded in M, we take a neighborhood U of x in Σ such that $U = g^{-1}(a)$, where $g : V \subset M \to \mathbb{R}$ is differentiable, $U \subset V$, and $a \in \mathbb{R}$ is a regular value of g. Take an orthonormal referential $E_1, \dots, E_{n+1}$ on V such that $E_{n+1} = Dg/|Dg| = \nu$. Denote by H the mean curvature and by A the second fundamental form of Σ; then

$$\begin{aligned}H &= \sum_{i=1}^{n}\langle A(E_i, E_i), \nu\rangle = \sum_{i=1}^{n}\langle \overline{\nabla}_{E_i} E_i, \nu\rangle \\ &= -\sum_{i=1}^{n}\langle \overline{\nabla}_{E_i}\nu, E_i\rangle - \langle \overline{\nabla}_{\nu}\nu, \nu\rangle = -\sum_{i=1}^{n+1}\langle \overline{\nabla}_{E_i}\nu, E_i\rangle = -\operatorname{div} \nu.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\Delta f &= \langle \overline{\nabla}_{\nu} Df, \nu\rangle + \langle Df, \overline{\nabla}_{\nu}\nu\rangle - H^{\Sigma}\langle Df, \nu\rangle \\ &= \operatorname{Hess} f(\nu, \nu) + \langle Df, \nu\rangle\langle \nu, \overline{\nabla}_{\nu}\nu\rangle - H^{\Sigma}\langle Df, \nu\rangle \\ &= \operatorname{Hess} f(\nu, \nu) - H^{\Sigma}\langle Df, \nu\rangle.\end{aligned}$$

$\square$

The next proposition will be useful in the proof of the Penrose inequality.

Proposition 3.2. *Let $\Omega \subset \mathbb{R}^n$ be an open bounded set such that the boundary of $\mathbb{R}^n_{-}\backslash\Omega$ is smooth. Let $f : \mathbb{R}^n_{-}\backslash\Omega \to \mathbb{R}$ be a C^2 function up to boundary, asymptotically flat, constant on each connected component of $\partial\Omega$, and such that $|Df| \to \infty$ when $x \to \partial\Omega$. We suppose that the graph of f has the induced metric from*

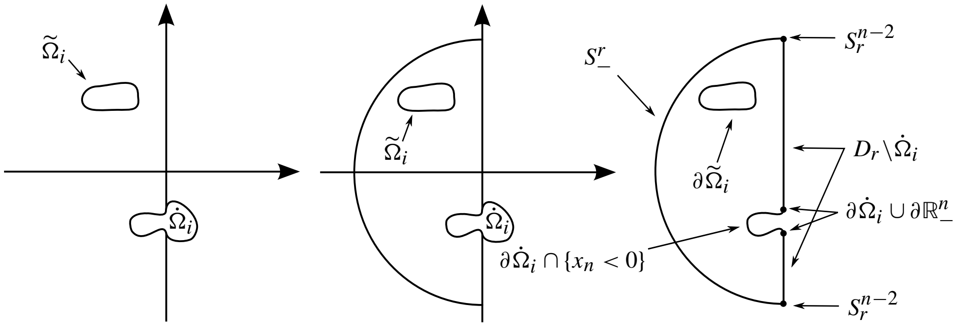


Figure 2. Sets of Proposition 3.2.

$\mathbb{R}^{n+1}$. *Denoting by* $\dot{\Omega}_i$, $i = 1, \dots, \dot{n}$, *the connected components of* Ω *such that* $\dot{\Omega}_i \cap \{x_n < 0\} \neq \varnothing$ *and* $\dot{\Omega}_i \cap \{x_n > 0\} \neq \varnothing$; *and by* $\widetilde{\Omega}_i$, $i = 1, \dots, \tilde{n}$, *the connected components of* Ω *such that* $\widetilde{\Omega}_i \subset \mathbb{R}^n_-$, *then*

$$\begin{aligned} c(n)m_{(M,g)} = \int_{\mathbb{R}^n_- \setminus \Omega} S \,\mathrm{d}x_\delta &+ \int_{\partial\mathbb{R}^n_- \setminus \cup\dot{\Omega}_i} \frac{\overline{H}}{\sqrt{1+(f_n)^2}}(2+|Df|^2)\,\mathrm{d}x_\delta \\ &+ \int_{\partial\mathbb{R}^n_- \setminus \cup\dot{\Omega}_i} f_n\sqrt{1+|\overline{D}f|^2}\Big(\sum_{i=1}^{n-1}\tilde{h}(\bar{\partial}_i, \bar{\partial}_i)\Big)\frac{|Df|^2}{1+|Df|^2}\,\mathrm{d}x_\delta \\ &- \sum_{i=1}^{\dot{n}}\int_{\partial\dot{\Omega}_i\cap\partial\mathbb{R}^n_-} f_n\langle\overline{D}f, \dot{\eta}\rangle\,\mathrm{d}\dot{\sigma}_i + \int_{\cup\partial\dot{\Omega}_i\cap\{x_n<0\}} H^{\partial\Omega}\,\mathrm{d}\dot{\sigma} \\ &+ \int_{\cup\partial\widetilde{\Omega}_i} H^{\partial\Omega}\,\mathrm{d}\tilde{\sigma}. \end{aligned}$$

Here $c(n) = 2(n-1)\omega_{n-1}$ *and* $\dot{\eta}$ *is the unit normal vector field on* $\partial\dot{\Omega}$ *pointing inward to* $\dot{\Omega}$. *Here*, $\overline{H}$ *and* $\tilde{h}$ *are the mean curvature and the scalar second fundamental form, respectively defined in Theorem 2.2 and* $H^{\partial\Omega}$ *is the mean curvature of* $\partial\Omega$ *viewed as a submanifold of the hyperplanes containing them.*

Proof. We have

$$\begin{aligned} \text{(3-1)} \quad \int_{\mathbb{R}^n_-\setminus\Omega} S\,\mathrm{d}x_\delta &= \lim_{r\to\infty}\int_{B^-_r\setminus\Omega}\operatorname{div}X\,\mathrm{d}x_\delta \\ &= \lim_{r\to\infty}\int_{S^-_r}\Big\langle X, \frac{x}{|x|}\Big\rangle\,\mathrm{d}\sigma_r + \lim_{r\to\infty}\int_{D_r\setminus\cup\dot{\Omega}_i}\langle X, e_n\rangle\,\mathrm{d}x_\delta \\ &\qquad + \int_{\cup\partial\dot{\Omega}_i\cap\{x_n<0\}}\langle X, \dot{\eta}_i\rangle\,\mathrm{d}\dot{\sigma} + \int_{\cup\partial\widetilde{\Omega}_i}\langle X, \tilde{\eta}_i\rangle\,\mathrm{d}\tilde{\sigma}, \end{aligned}$$

where $X = 1/(1+|Df|^2)\,(f_i f_{kk} - f_k f_{ik})e_i$. Like in the Theorem 2.2, we find

$$\lim_{r\to\infty}\int_{S_r^-}\Big\langle X, \frac{x}{|x|}\Big\rangle\,\mathrm{d}\sigma_r = \lim_{r\to\infty}\int_{S_r^-}(f_i f_{kk} - f_k f_{ik})\frac{x_i}{|x|}\,\mathrm{d}\sigma_r. \tag{3-2}$$

On the other hand, since $\overline{U} = 1 - |Df|^2/(1+|Df|^2)$:

$$\langle X, e_n\rangle = \sum_{k=1}^{n-1}(f_n f_{kk} - f_k f_{nk}) - \frac{|Df|^2}{1+|Df|^2}\sum_{k=1}^{n-1}(f_n f_{kk} - f_k f_{nk}).$$

Thus,

$$\begin{aligned}\lim_{r\to\infty}\int_{D_r\setminus\cup\dot\Omega_i}\langle X, e_n\rangle\,\mathrm{d}x_\delta &= \lim_{r\to\infty}\int_{D_r\setminus\cup\dot\Omega_i}\sum_{k=1}^{n-1}(f_n f_{kk} - f_k f_{nk})\,\mathrm{d}x_\delta \\ &\qquad - \lim_{r\to\infty}\int_{D_r\setminus\cup\dot\Omega_i}\frac{|Df|^2}{1+|Df|^2}\sum_{k=1}^{n-1}(f_n f_{kk} - f_k f_{nk})\,\mathrm{d}x_\delta \\ &= \lim_{r\to\infty}\int_{D_r\setminus\cup\dot\Omega_i}(\mathrm{div}_{\mathbb{R}^{n-1}}(f_n\overline{D}f) - 2\langle\overline{D}f, \overline{D}f_n\rangle)\,\mathrm{d}x_\delta \\ &\qquad - \lim_{r\to\infty}\int_{D_r\setminus\cup\dot\Omega_i}\frac{|Df|^2}{1+|Df|^2}\sum_{k=1}^{n-1}(f_n f_{kk} - f_k f_{nk})\,\mathrm{d}x_\delta.\end{aligned}$$

Here, $\overline{D}f = (\partial f/\partial x_1, \dots, \partial f/\partial x_{n-1})$. Using that

$$\begin{aligned}&\lim_{r\to\infty}\int_{D_r\setminus\cup\dot\Omega_i}\mathrm{div}_{\mathbb{R}^{n-1}}(f_n\overline{D}f)\,\mathrm{d}x_\delta \\ &\qquad = \lim_{r\to\infty}\int_{S_r^{n-2}} f_n\langle\overline{D}f, \eta_r\rangle\,\mathrm{d}\sigma_r + \sum_{i=1}^{\dot n}\int_{\partial\dot\Omega_i\cap\partial\mathbb{R}^n_-} f_n\langle\overline{D}f, \dot\eta\rangle\,\mathrm{d}\dot\sigma_i,\end{aligned}$$

where η_r is the unit normal vector field on S_r^{n-2} pointing outward to D_r, we find

$$\begin{aligned}\lim_{r\to\infty}\int_{D_r\setminus\cup\dot\Omega_i}\langle X, e_n\rangle\,\mathrm{d}x_\delta &= \lim_{r\to\infty}\int_{S_r^{n-2}} f_n\langle\overline{D}f, \eta_r\rangle\,\mathrm{d}\sigma_r \\ &\quad + \sum_{i=1}^{\dot n}\int_{\partial\dot\Omega_i\cap\partial\mathbb{R}^n_-} f_n\langle\overline{D}f, \dot\eta\rangle\,\mathrm{d}\dot\sigma_i \\ &\quad -2\lim_{r\to\infty}\int_{D_r\setminus\cup\dot\Omega_i}\langle\overline{D}f, \overline{D}f_n\rangle\,\mathrm{d}x_\delta \\ &\quad - \lim_{r\to\infty}\int_{D_r\setminus\cup\dot\Omega_i}\frac{|Df|^2}{1+|Df|^2}\sum_{k=1}^{n-1}(f_n f_{kk} - f_k f_{nk})\,\mathrm{d}x_\delta.\end{aligned} \tag{3-3}$$

Using (3-1), (3-2), and (3-3), we find

$$\int_{\mathbb{R}^n_-\setminus\Omega} S\,\mathrm{d}x_\delta = \lim_{r\to\infty}\int_{S_r^-}(f_i f_{kk}-f_k f_{ik})\frac{x_i}{|x|}\,\mathrm{d}\sigma_r + \lim_{r\to\infty}\int_{S_r^{n-2}} f_n\langle \overline{D}f,\eta_r\rangle\,\mathrm{d}\sigma_r$$
$$+\sum_{i=1}^{\dot{n}}\int_{\partial\dot{\Omega}_i\cap\partial\mathbb{R}^n_-} f_n\langle\overline{D}f,\dot{\eta}\rangle\,\mathrm{d}\dot{\sigma}_i - 2\lim_{r\to\infty}\int_{D_r\setminus\cup\dot{\Omega}_i}\langle\overline{D}f,\overline{D}f_n\rangle\,\mathrm{d}x_\delta$$
$$-\lim_{r\to\infty}\int_{D_r\setminus\cup\dot{\Omega}_i}\frac{|Df|^2}{1+|Df|^2}\sum_{k=1}^{n-1}(f_n f_{kk}-f_k f_{nk})\,\mathrm{d}x_\delta$$
$$+\int_{\cup\partial\dot{\Omega}_i\cap\{x_n<0\}}\langle X,\dot{\eta}_i\rangle\,\mathrm{d}\dot{\sigma}+\int_{\cup\partial\widetilde{\Omega}_i}\langle X,\tilde{\eta}_i\rangle\,\mathrm{d}\tilde{\sigma}.$$

That is,

$$c(n)m_{(M,g)} = \int_{\mathbb{R}^n_-\setminus\Omega} S\,\mathrm{d}x_\delta - \sum_{i=1}^{\dot{n}}\int_{\partial\dot{\Omega}_i\cap\partial\mathbb{R}^n_-} f_n\langle\overline{D}f,\dot{\eta}\rangle\,\mathrm{d}\dot{\sigma}_i$$
$$+2\lim_{r\to\infty}\int_{D_r\setminus\cup\dot{\Omega}_i}\langle\overline{D}f,\overline{D}f_n\rangle\,\mathrm{d}x_\delta$$
$$+\lim_{r\to\infty}\int_{D_r\setminus\cup\dot{\Omega}_i}\frac{|Df|^2}{1+|Df|^2}\sum_{k=1}^{n-1}(f_n f_{kk}-f_k f_{nk})\,\mathrm{d}x_\delta$$
$$-\int_{\cup\partial\dot{\Omega}_i\cap\{x_n<0\}}\langle X,\dot{\eta}_i\rangle\,\mathrm{d}\dot{\sigma}-\int_{\cup\partial\widetilde{\Omega}_i}\langle X,\tilde{\eta}_i\rangle\,\mathrm{d}\tilde{\sigma}. \tag{3-4}$$

Like in the Theorem 2.2 we have

$$\lim_{r\to\infty}\int_{D_r\setminus\cup\dot{\Omega}_i} f_k f_{nk}\left\{2-\frac{|Df|^2}{1+|Df|^2}\right\}\mathrm{d}x_\delta = \int_{\partial\mathbb{R}^n_-\setminus\cup\dot{\Omega}_i}\frac{\overline{H}}{\sqrt{1+(f_n)^2}}(2+|Df|^2)\,\mathrm{d}x_\delta, \tag{3-5}$$

where $\overline{H}$ is the mean curvature of $\partial G(f)$, viewed as a submanifold of $G(f)$, and we also have

$$\lim_{r\to\infty}\int_{D_r\setminus\cup\dot{\Omega}_i} f_n f_{kk}\frac{|Df|^2}{1+|Df|^2}\,\mathrm{d}x_\delta = \int_{\partial\mathbb{R}^n_-\setminus\cup\dot{\Omega}_i} f_n\sqrt{1+|\overline{D}f|^2}\left(\sum_{i=1}^{n-1}\tilde{h}(\bar{\partial}_i,\bar{\partial}_i)\right)\frac{|Df|^2}{1+|Df|^2}\,\mathrm{d}x_\delta, \tag{3-6}$$

where $\tilde{h}$ is the scalar second fundamental form of $\partial G(f)$, viewed as a submanifold of $\partial\mathbb{R}^n_-$. Therefore,

$$
\begin{aligned}
\text{(3-7)}\quad c(n)m(g) = &\int_{\mathbb{R}^n_-\setminus\Omega} S\,\mathrm{d}x_\delta + \int_{\partial\mathbb{R}^n_-\setminus\cup\dot{\Omega}_i} \frac{\overline{H}}{\sqrt{1+(f_n)^2}}(2+|Df|^2)\,\mathrm{d}x_\delta \\
&+ \int_{\partial\mathbb{R}^n_-\setminus\cup\dot{\Omega}_i} f_n\sqrt{1+|\overline{D}f|^2}\left(\sum_{i=1}^{n-1}\tilde{h}(\bar{\partial}_i,\bar{\partial}_i)\right)\frac{|Df|^2}{1+|Df|^2}\,\mathrm{d}x_\delta \\
&- \sum_{i=1}^{\dot{n}}\int_{\partial\dot{\Omega}_i\cap\partial\mathbb{R}^n_-} f_n\langle\overline{D}f,\dot{\eta}\rangle\,\mathrm{d}\dot{\sigma}_i \\
&- \int_{\cup\partial\dot{\Omega}_i\cap\{x_n<0\}}\langle X,\dot{\eta}\rangle\,\mathrm{d}\dot{\sigma} - \int_{\cup\partial\widetilde{\Omega}_i}\langle X,\tilde{\eta}\rangle\,\mathrm{d}\tilde{\sigma}.
\end{aligned}
$$

The next computations of the mean curvature of the level set $\partial\Omega$ can also be found in [Lam 2011, Equation 5.3] or [Mirandola and Vitório 2015, Equation 33]. Since f is constant on $\partial\Omega$, we have that Df is normal to this set. Denote by Ω^c a component of Ω such that f increases when $x\to\partial\Omega^c$, thus $Df/|Df|$ is the unity normal vector field outward pointing to the graph on $\partial\Omega^c$ ($Df/|Df|$ points inward to Ω^c in the hyperplane containing Ω^c). Denote by Ω^d a component of Ω such that f decreases when $x\to\partial\Omega^d$, thus $-Df/|Df|$ is the unity normal vector field outward pointing to the graph on $\partial\Omega^d$ ($-Df/|Df|$ points inward to Ω^d in the hyperplane containing Ω^d). For an illustration, see Figure 3.

We have

$$
\begin{aligned}
\left\langle X,\frac{Df}{|Df|}\right\rangle &= \left\langle \overline{U}(f_i f_{kk}-f_k f_{ik})e_i,\frac{f_j}{|Df|}e_j\right\rangle = \frac{f_i\overline{U}}{|Df|}(f_i f_{kk}-f_k f_{ik}) \\
&= \frac{\overline{U}}{|Df|}(f_i^2 f_{kk}-f_k f_i f_{ik}) = \frac{\overline{U}}{|Df|}(|Df|^2\Delta f-\operatorname{Hess} f(Df,Df)) \\
&= \frac{|Df|^2}{|Df|}\overline{U}\left(\Delta f-\operatorname{Hess} f\left(\frac{Df}{|Df|},\frac{Df}{|Df|}\right)\right).
\end{aligned}
$$

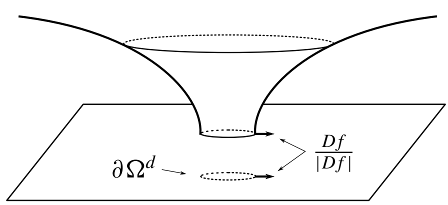


Figure 3. Illustration of the argument above.

Since $\nu = Df/|Df|$ is the unity normal vector field pointing inward to Ω^c, using Proposition 3.1 and that $\overline{U} = 1/(1+|Df|^2)$, we find

$$\begin{aligned}\left\langle X, \frac{Df}{|Df|}\right\rangle &= \frac{|Df|}{1+|Df|^2}\left(-H^{\partial\Omega^c}\left\langle Df, \frac{Df}{|Df|}\right\rangle\right)\\ &= -\frac{|Df|^2}{1+|Df|^2}H^{\partial\Omega^c}\left\langle \frac{Df}{|Df|}, \frac{Df}{|Df|}\right\rangle = -\frac{|Df|^2}{1+|Df|^2}H^{\partial\Omega^c}.\end{aligned}$$

Since $\nu = -Df/|Df|$ is the unity normal vector field pointing inward to Ω^d, using the Proposition 3.1 and that $\overline{U} = 1/(1+|Df|^2)$, we find

$$\begin{aligned}\left\langle X, -\frac{Df}{|Df|}\right\rangle &= -\frac{|Df|}{1+|Df|^2}\left(-H^{\partial\Omega^d}\left\langle Df, -\frac{Df}{|Df|}\right\rangle\right)\\ &= -\frac{|Df|^2}{1+|Df|^2}H^{\partial\Omega^d}\left\langle -\frac{Df}{|Df|}, -\frac{Df}{|Df|}\right\rangle = -\frac{|Df|^2}{1+|Df|^2}H^{\partial\Omega^d}.\end{aligned}$$

Here, $H^{\partial\Omega}$ is related with the vector field pointing inward to Ω. We know that $\lim_{x\to\partial\Omega}|Df(x)| = \infty$, thus $\lim_{x\to\partial\Omega}|Df|^2/(1+|Df|^2) = 1$, therefore, supposing that ν is the unity normal vector field pointing inward to Ω on $\partial\Omega$, on $\partial\Omega$ we have

$$\langle X, \nu\rangle = -H^{\partial\Omega}.$$

Using (3-7), we have

$$\begin{aligned}c(n)m_{(M,g)} = \int_{\mathbb{R}^n_-\setminus\Omega} S\,\mathrm{d}x_\delta &+ \int_{\partial\mathbb{R}^n_-\setminus\cup\dot{\Omega}_i}\frac{\overline{H}}{\sqrt{1+(f_n)^2}}(2+|Df|^2)\,\mathrm{d}x_\delta\\ &+ \int_{\partial\mathbb{R}^n_-\setminus\cup\dot{\Omega}_i} f_n\sqrt{1+|\overline{D}f|^2}\left(\sum_{i=1}^{n-1}\tilde{h}(\bar{\partial}_i,\bar{\partial}_i)\right)\frac{|Df|^2}{1+|Df|^2}\,\mathrm{d}x_\delta\\ &- \sum_{i=1}^{\dot{n}}\int_{\partial\dot{\Omega}_i\cap\partial\mathbb{R}^n_-} f_n\langle\overline{D}f, \dot{\eta}\rangle\,\mathrm{d}\dot{\sigma}_i + \int_{\cup\partial\dot{\Omega}_i\cap\{x_n<0\}} H^{\partial\Omega}\,\mathrm{d}\dot{\sigma}\\ &+ \int_{\cup\partial\widetilde{\Omega}_i} H^{\partial\Omega}\,\mathrm{d}\tilde{\sigma}.\end{aligned}$$ □

Now we will enunciate some auxiliary results.

Proposition 3.3 [Guan and Li 2009, Theorem 2]. *Let $\Omega\subset\mathbb{R}^{n+1}$ be a limited and star-shaped set. We also suppose that $\partial\Omega$ is smooth and mean-convex. Denote by $H^{\partial\Omega}$ the mean curvature of $\partial\Omega$ with respect to the normal unit vector field inward pointing to Ω and by $B\subset\mathbb{R}^{n+1}$ a unit ball. Then,*

$$\frac{1}{2n\omega_n}\int_{\partial\Omega} H^{\partial\Omega}\,\mathrm{d}\mu_{\partial\Omega} \geq \frac{1}{2}\left(\frac{|\partial\Omega|}{\omega_n}\right)^{(n-1)/n}.$$

Moreover, the equality will occur if and only if Ω is a ball.

Proposition 3.4 [Freire and Schwartz 2014, Theorem 5, item (V)]. *Let $\Omega \subset \mathbb{R}^n$ be a set (not necessarily connected) limited, with a smooth mean-convex and outer-minimizing boundary. Denote by $H^{\partial\Omega}$ the mean curvature of $\partial\Omega$ with respect to the normal unit vector field inward pointing to Ω. Then,*

$$\frac{1}{2(n-1)\omega_{n-1}} \int_{\partial\Omega} H^{\partial\Omega}\,\mathrm{d}\partial\Omega \geq \frac{1}{2}\Big(\frac{|\partial\Omega|}{\omega_{n-1}}\Big)^{(n-2)/(n-1)}.$$

Moreover, the equality will occur if and only if each connected component of Ω is a rounded ball.

Lemma 3.5 [Huang and Wu 2015, Proposition 5.2]. *Let $a_1, \dots, a_k$ be nonnegative real numbers and $0 \leq \beta \leq 1$. Then,*

$$\sum_{i=1}^{k} a_i^{\beta} \geq \Big(\sum_{i=1}^{k} a_i\Big)^{\beta}.$$

If $0 \leq \beta < 1$, then the equality holds if and only if at most one element of $\{a_1, \dots, a_k\}$ is nonzero.

Using these results we obtain the following theorem:

Theorem 3.6. *Let $\Omega \subset \mathbb{R}^n_- \setminus \{x_n = 0\}$, $n \geq 3$, be an open bounded set whose boundary is smooth and mean-convex. Suppose that $\partial\Omega$ is outer-minimizing or each connected component of Ω is star-shaped. Let $f : \mathbb{R}^n_- \setminus \Omega \to \mathbb{R}$ be a C^2 function up to boundary, asymptotically flat, constant on each connected component of $\partial\Omega$ and such that $|Df| \to \infty$ when $x \to \partial\Omega$. We also suppose that $f_n = 0$ on $\partial\mathbb{R}^n_-$, and the curvatures that appears at the Proposition 3.2 are nonnegative. Then,*

$$m_{(M,g)} \geq \frac{1}{2}\Big(\frac{|\partial\Omega|}{\omega_{n-1}}\Big)^{(n-2)/(n-1)},$$

where $|\partial\Omega|$ denotes the area measure on $\partial\Omega$.

Proof. By the Proposition 3.2, we have

$$2(n-1)\omega_{n-1} m(g) = \int_{\mathbb{R}^n_- \setminus \Omega} S\,\mathrm{d}x_\delta + \int_{\partial\mathbb{R}^n_-} \frac{\overline{H}}{\sqrt{1+(f_n)^2}}(2+|Df|^2)\,\mathrm{d}x_\delta + \int_{\partial\Omega} H^{\partial\Omega}\,\mathrm{d}\partial\Omega.$$

Denoting by Ω_i, $i = 1, \dots, k$, the connected components of Ω, we have

$$\begin{aligned} m_{(M,g)} &\geq \frac{1}{2(n-1)\omega_{n-1}} \sum_i \int_{\partial\Omega_i} H^{\partial\Omega_i}\,\mathrm{d}\partial\Omega_i \geq \frac{1}{2}\sum_i \Big(\frac{|\partial\Omega_i|}{\omega_{n-1}}\Big)^{(n-2)/(n-1)} \\ &\geq \frac{1}{2}\Big(\sum_i \frac{|\partial\Omega_i|}{\omega_{n-1}}\Big)^{(n-2)/(n-1)} = \frac{1}{2}\Big(\frac{|\partial\Omega|}{\omega_{n-1}}\Big)^{(n-2)/(n-1)}. \end{aligned}$$

□

References

[Almaraz et al. 2016] S. Almaraz, E. Barbosa, and L. L. de Lima, "A positive mass theorem for asymptotically flat manifolds with a non-compact boundary", *Comm. Anal. Geom.* **24**:4 (2016), 673–715. MR Zbl

[Escobar 1992] J. F. Escobar, "The Yamabe problem on manifolds with boundary", *J. Differential Geom.* **35**:1 (1992), 21–84. MR Zbl

[Freire and Schwartz 2014] A. Freire and F. Schwartz, "Mass-capacity inequalities for conformally flat manifolds with boundary", *Comm. Partial Differential Equations* **39**:1 (2014), 98–119. MR

[Guan and Li 2009] P. Guan and J. Li, "The quermassintegral inequalities for k-convex starshaped domains", *Adv. Math.* **221**:5 (2009), 1725–1732. MR Zbl

[Huang and Wu 2013] L.-H. Huang and D. Wu, "Hypersurfaces with nonnegative scalar curvature", *J. Differential Geom.* **95**:2 (2013), 249–278. MR Zbl

[Huang and Wu 2015] L.-H. Huang and D. Wu, "The equality case of the Penrose inequality for asymptotically flat graphs", *Trans. Amer. Math. Soc.* **367**:1 (2015), 31–47. MR Zbl

[Lam 2011] M.-K. G. Lam, *The Graph Cases of the Riemannian Positive Mass and Penrose Inequalities in All Dimensions*, Ph.D. thesis, Duke University, 2011, available at https://search.proquest.com/docview/868529318. MR

[Lee and Sormani 2014] D. A. Lee and C. Sormani, "Stability of the positive mass theorem for rotationally symmetric Riemannian manifolds", *J. Reine Angew. Math.* **686** (2014), 187–220. MR Zbl

[Mirandola and Vitório 2015] H. Mirandola and F. Vitório, "The positive mass theorem and Penrose inequality for graphical manifolds", *Comm. Anal. Geom.* **23**:2 (2015), 273–292. MR

[Raulot 2011] S. Raulot, "Green functions for the Dirac operator under local boundary conditions and applications", *Ann. Global Anal. Geom.* **39**:4 (2011), 337–359. MR Zbl

[Reilly 1973] R. C. Reilly, "On the Hessian of a function and the curvatures of its graph", *Michigan Math. J.* **20** (1973), 373–383. MR

Received September 8, 2016. Revised June 27, 2017.

EZEQUIEL BARBOSA
DEPARTAMENTO DE MATEMÁTICA
FEDERAL UNIVERSITY OF MINAS GERAIS
BELO HORIZONTE
BRAZIL
ezequiel@mat.ufmg.br

ADSON MEIRA
DEPARTAMENTO DE MATEMÁTICA
FEDERAL UNIVERSITY OF MINAS GERAIS
BELO HORIZONTE
BRAZIL
nosdariem@ufmg.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 294, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.294.275

DIAGRAMS FOR RELATIVE TRISECTIONS

NICKOLAS A. CASTRO, DAVID T. GAY AND JUANITA PINZÓN-CAICEDO

We establish a correspondence between trisections of smooth, compact, oriented 4-manifolds with connected boundary and diagrams describing these trisected 4-manifolds. Such a diagram comes in the form of a compact, oriented surface with boundary together with three tuples of simple closed curves, with possibly fewer curves than the genus of the surface, satisfying a pairwise condition of being standard. This should be thought of as the 4-dimensional analog of a sutured Heegaard diagram for a sutured 3-manifold. We also give many foundational examples.

1. Introduction

Gay and Kirby [2016] defined, and proved existence and uniqueness statements for, trisections of both closed 4-manifolds and compact 4-manifolds with connected boundary. In the latter, relative case, the trisections restrict to open book decompositions on the bounding 3-manifolds. In the closed case, they discuss trisection diagrams in the same paper: these are diagrams involving curves on surfaces which uniquely determine closed, trisected 4-manifolds up to diffeomorphism. The aim of this paper is to complete the story by defining relative trisection diagrams and showing that they uniquely determine trisected 4-manifolds with connected boundary, as well as to present a series of fundamental examples.

Before recalling the background definitions in [Gay and Kirby 2016], we introduce some basic definitions and state the main result of the present article.

Definition 1. Two $(n+1)$-tuples of the form $(\Sigma, \alpha^1, \dots, \alpha^n)$, where each α^i is a collection $\alpha^i = \{\alpha^i_1, \dots, \alpha^i_k\}$ of k disjoint simple closed curves on the surface Σ, are *diffeomorphism and handle slide equivalent* if they are related by a diffeomorphism between the surfaces and a sequence of handle slides within each α^i; i.e., one is only allowed to slide curves from α^i over other curves from α^i, but not over curves from α^j when $j \neq i$.

This work was supported by a grant from the Simons Foundation (# 359873, David Gay).
MSC2010: 57M99.
Keywords: trisection, open book, Heegaard diagram, monodromy, 3-manifold, 4-manifold.

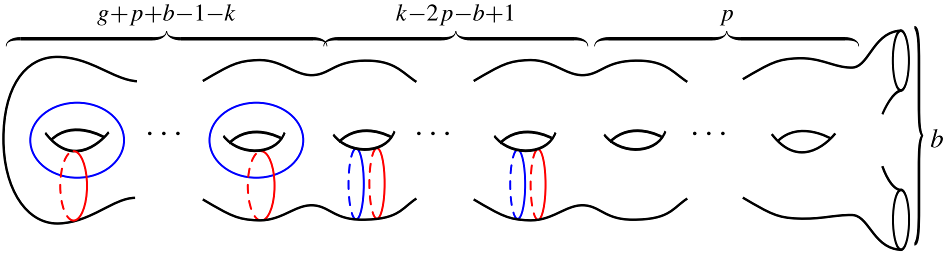


Figure 1. The standard model $(\Sigma, \delta, \epsilon)$.

Definition 2. A $(g, k;\, p, b)$*-trisection diagram*, where

$$2p+b-1 \leq k \leq g+p+b-1,$$

is a 4-tuple $(\Sigma, \alpha, \beta, \gamma)$, where Σ is a surface of genus g with b boundary components and each of α, β and γ is a collection of $g-p$ simple closed curves such that each triple (Σ, α, β), (Σ, β, γ), and (Σ, γ, α) is diffeomorphism and handle slide equivalent to the triple $(\Sigma, \delta, \epsilon)$ shown in Figure 1.

The following theorem, the main result of this paper, references trisections of 4-manifolds with boundary, but we defer the definition of this concept to a later section. If this is new to the reader, the main thing to know at the moment is that a trisection of a 4-manifold X is a decomposition into three codimension–0 submanifolds $X = X_1 \cup X_2 \cup X_3$, and that in the relative case a trisection induces an open book decomposition on ∂X.

Theorem 3. *For every* $(g, k;\, p, b)$*-trisection diagram* $(\Sigma, \alpha, \beta, \gamma)$ *there is a unique* (*up to diffeomorphism*) *trisected* 4*-manifold* $X = X_1 \cup X_2 \cup X_3$ *with connected boundary, such that, with respect to a fixed identification* $\Sigma \cong X_1 \cap X_2 \cap X_3$, *the* α, β *and* γ *curves, respectively, bound disks in* $X_1 \cap X_2$, $X_2 \cap X_3$ *and* $X_3 \cap X_1$. *In particular, the open book decomposition on* ∂X *has* b *binding components and pages of genus* p. *Furthermore, any trisected* 4*-manifold with connected boundary is determined in this way by some relative trisection diagram, and any two relative trisection diagrams for the same* 4*-manifold trisection are diffeomorphism and handle slide equivalent.*

As a consequence, the monodromy of the open book decomposition on ∂X is also completely determined by the diagram $(\Sigma, \alpha, \beta, \gamma)$. We now describe how to read off the monodromy from the diagram.

Definition 4. Given a compact oriented surface Σ, consider a pair

$$(\alpha = (\alpha_1, \ldots, \alpha_k), a = (a_1, \ldots, a_l)),$$

where each α_i is a simple closed curve in Σ, each a_j is a properly embedded arc in Σ, and $\{\alpha_1, \ldots, \alpha_k, a_1, \ldots, a_l\}$ are disjoint. We say that another such pair (α', a') is *handle slide equivalent* to (α, a) if (α', a') is obtained from (α, a) by a sequence of the following two operations: (1) Slide one simple closed curve in α over another simple closed curve in α. (2) Slide one arc in a over a simple closed curve in α.

Note that we do not allow "arc slides", in which arcs in a slide over other arcs in a.

We adopt the following notation: Given a surface Σ and a collection of simple closed curves α, Σ_α denotes the surface obtained by performing surgery along α. This comes with an embedding $\phi_\alpha : \Sigma \setminus \alpha \to \Sigma_\alpha$, the image of which is the complement of a collection of pairs of points, one for each component of α.

Theorem 5. *A relative trisection diagram* $(\Sigma, \alpha, \beta, \gamma)$ *encodes an open book decomposition on* ∂X *with page given by* Σ_α, *the surface resulting from* Σ *by performing surgery along the* α *curves, and monodromy* $\mu : \Sigma_\alpha \to \Sigma_\alpha$ *determined by the following algorithm*:

(1) *Choose an ordered collection of arcs* a *on* Σ, *disjoint from* α *and such that its image* $\phi_\alpha(a)$ *in* Σ_α *cuts* Σ_α *into a disk.*

(2) *There exists a collection of arcs* a_1 *and simple closed curves* β' *in* Σ *such that* (α, a_1) *is handle slide equivalent to* (α, a), β' *is handle slide equivalent to* β, *and* a_1 *and* β' *are disjoint.* (*We claim that in this step we do not need to slide* α *curves over* α *curves, only* a *arcs over* α *curves and* β *curves over* β *curves.*) *Choose such an* a_1 *and* β'

(3) *There exists a collection of arcs* a_2 *and simple closed curves* γ' *in* Σ *such that* (β', a_2) *is handle slide equivalent to* (β', a_1), γ' *is handle slide equivalent to* γ, *and* a_2 *and* γ' *are disjoint.* (*Again we claim that we do not need to slide* β' *curves over* β' *curves.*) *Choose such an* a_2 *and* γ'

(4) *There exists a collection of arcs* a_3 *and simple closed curves* α' *in* Σ *such that* (γ', a_3) *is handle slide equivalent to* (γ', a_2), α' *is handle slide equivalent to* α, *and* a_3 *and* α' *are disjoint.* (*Again we do not need to slide* γ' *curves over* γ' *curves.*) *Choose such an* a_3 *and* α'.

(5) *The pair* (α', a_3) *is handle slide equivalent to* (α, a_*) *for some collection of arcs* a_*. *Choose such an* a_*. *Note that now* a *and* a_* *are both disjoint from* α *and thus we can compare* $\phi_\alpha(a)$ *and* $\phi_\alpha(a_*)$ *in* Σ_α.

(6) *The monodromy* μ *is the unique* (*up to isotopy*) *map such that*

$$\mu(\phi_\alpha(a)) = \phi_\alpha(a_*),$$

respecting the ordering of the collections of arcs.

Of course there are choices in the above algorithm each time we perform handleslides to arrange disjointness from the next system of curves, but part of the content of the theorem is that the resulting μ is independent of these choices.

Note that this, together with the existence of trisections relative to given open books [Gay and Kirby 2016], gives us a purely 2-dimensional result, namely that there is a way to encode mapping classes of surfaces with boundary via trisection diagrams (on higher genus surfaces).

An alternative definition of a relative trisection diagram includes both the systems of curves α, β and γ and the systems of arcs a_1, a_2, a_3; from such a definition it is easier to see that a diagram determines a trisected 4-manifold. The nontriviality of both Theorem 3 and Theorem 5 is that one does not in fact need the arcs to uniquely determine the 4-manifold and the open book on its boundary.

2. Trisections of closed manifolds and their diagrams

Let $Z_k = \natural^k(S^1 \times B^3)$ with $Y_k = \partial Z_k = \#^k(S^1 \times S^2)$. Given an integer $g \geq k$, let $Y_k = Y^-_{g,k} \cup Y^+_{g,k}$ be the standard genus g Heegaard splitting of Y_k obtained by stabilizing the standard genus k Heegaard splitting $g - k$ times.

Definition 6. A (g, k)-trisection of a closed, connected, oriented 4-manifold X is a decomposition of X into three submanifolds $X = X_1 \cup X_2 \cup X_3$ satisfying the following properties:

(1) For each $i = 1, 2, 3$, there is a diffeomorphism $\phi_i : X_i \to Z_k$.

(2) For each $i = 1, 2, 3$, taking indices mod 3,

$$\phi_i(X_i \cap X_{i+1}) = Y^-_{g,k} \quad \text{and} \quad \phi_i(X_i \cap X_{i-1}) = Y^+_{g,k}.$$

Theorem 7 [Gay and Kirby 2016]. *Every smooth closed oriented connected 4-manifold has a trisection.*

Definition 8. A (g, k)-*trisection diagram* is a tuple $(\Sigma, \alpha, \beta, \gamma)$ such that Σ is a closed oriented surface of genus g and each triple (Σ, α, β), (Σ, β, γ) and (Σ, γ, α) is diffeomorphism and handle slide equivalent to the triple $(\Sigma, \delta, \epsilon)$ shown in Figure 2.

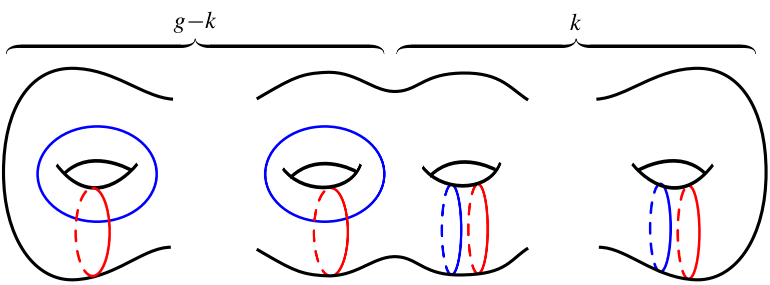


Figure 2. The standard model $(\Sigma, \delta, \epsilon)$ in the closed case.

The following result is straightforward, and we present the proof here only to set the stage for the more subtle relative case.

Theorem 9 [Gay and Kirby 2016]. *For every (g, k)-trisection diagram $(\Sigma, \alpha, \beta, \gamma)$ there is a unique (up to diffeomorphism) closed trisected 4-manifold $X = X_1 \cup X_2 \cup X_3$ such that, with respect to a fixed identification $\Sigma \cong X_1 \cap X_2 \cap X_3$, the α, β and γ curves, respectively, bound disks in $X_1 \cap X_2$, $X_2 \cap X_3$ and $X_3 \cap X_1$. Furthermore, any closed trisected 4-manifold is determined in this way by some trisection diagram, and any two trisection diagrams for the same 4-manifold trisection are diffeomorphism and handle slide equivalent.*

Proof. Note that the diagram in Figure 2 is a standard genus g Heegaard diagram for $\#^k S^1 \times S^2 = Y_k$, describing the standard genus g splitting $Y_k = Y^-_{g,k} \cup Y^+_{g,k}$. Fix an identification of Σ with $Y^-_{g,k} \cap Y^+_{g,k}$ such that the δ curves bound disks in $Y^-_{g,k}$ and the ϵ curves bound disks in $Y^+_{g,k}$.

Given a trisected 4-manifold $X = X_1 \cup X_2 \cup X_3$, let $\phi_i : X_i \to Z_k$, for $i = 1, 2, 3$, be the diffeomorphisms from Definition 6. The associated diagram is then $(X_1 \cap X_2 \cap X_3, \phi_1^{-1}(\delta), \phi_2^{-1}(\delta), \phi_3^{-1}(\delta))$. Equivalently one could replace any $\phi_i^{-1}(\delta)$ with $\phi_{i+1}^{-1}(\epsilon)$, or in fact any other cut system of g curves bounding disks in $X_i \cap X_{i+1}$; the resulting diagrams would be handle slide equivalent [Johannson 1995].

Conversely, given a trisection diagram $(\Sigma, \alpha, \beta, \gamma)$, let H_α, H_β and H_γ, be handlebodies bounded by Σ and determined by α, β and γ, respectively. Then build X by starting with $B^2 \times \Sigma$, attaching $I \times H_\alpha$, $I \times H_\beta$ and $I \times H_\gamma$ to $\partial B^2 \times \Sigma = S^1 \times F_g$ along successive arcs in S^1 crossed with Σ. This produces a 4-manifold with three boundary components, but because each pair of systems of curves is a Heegaard diagram for $\#^k S^1 \times S^2$, each boundary component is diffeomorphic to $\#^k S^1 \times S^2$, and hence can be capped off uniquely with $\natural^k S^1 \times B^3$ [Laudenbach and Poénaru 1972]. $\square$

3. Relative trisections

Here we rephrase the definition of relative trisection from [Gay and Kirby 2016]. Given integers $(g, k; p, b)$ with $g \geq p$ and $g + p + b - 1 \geq k \geq 2p + b - 1$, we begin as in the closed case with $Z_k = \natural^k S^1 \times B^3$ and $Y_k = \partial Z_k = \#^k S^1 \times S^2$, but in this case we describe a certain decomposition of Y_k as $Y_k = Y^-_{g,k;p,b} \cup Y^0_{g,k;p,b} \cup Y^+_{g,k;p,b}$ needed for the definition. This decomposition is illustrated in Figure 3 as a lower-dimensional analog.

Let D be a third of a unit 2-dimensional disk. Namely, use polar coordinates and set

$$D = \{(r, \theta) \mid r \in [0, 1], \theta \in [-\pi/3, \pi/3]\}.$$

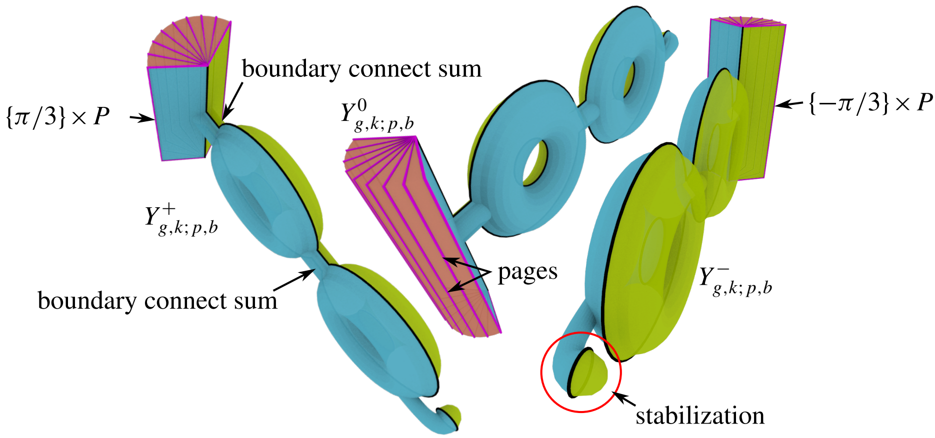


Figure 3. Several views of a lower-dimensional analog of the standard model Z_k for a sector of a relative trisection, with the decomposition of the boundary $Y_k = Y^-_{g,k;p,b} \cup Y^0_{g,k;p,b} \cup Y^+_{g,k;p,b}$. The page P is represented as a straight line segment, in purple.

Decompose ∂D as $\partial D = \partial^- D \cup \partial^0 D \cup \partial^+ D$, where

$$\begin{aligned} \partial^- D &= \{r \in [0,1], \theta = -\pi/3\}, \\ \partial^0 D &= \{r = 1, \theta \in [-\pi/3, \pi/3]\}, \text{ and} \\ \partial^+ D &= \{r \in [0,1], \theta = \pi/3\}. \end{aligned} \tag{3-1}$$

Now let P be a compact surface of genus p with b boundary components and consider $U = D \times P$. Note that $U \cong \natural^{2p+b-1} S^1 \times B^3$ and that the decomposition (3-1) induces a decomposition of ∂U as

$$\partial U = \partial^- U \cup \partial^0 U \cup \partial^+ U,$$

where $\partial^\pm U = \partial^\pm D \times P$ and $\partial^0 U = (\partial^0 D \times P) \cup (D \times \partial P)$. Similarly, notice that if we regroup the sets involved in the decomposition of ∂U into $\partial^- U \cup \partial^0 U$ and $\partial^+ U$, we obtain the standard genus $2p+b-1$ Heegaard splitting of $\#^{2p+b-1} S^1 \times S^2$.

Next, decompose $\partial(S^1 \times B^3) = S^1 \times S^2$ as $\partial^-(S^1 \times B^3) \cup \partial^+(S^1 \times B^3)$, where $\partial^\pm(S^1 \times B^3) = S^1 \times S^2_\pm$ and $S^2_\pm$ are the northern and southern hemispheres. This is the standard genus 1 Heegaard splitting of $S^1 \times S^2$. For a positive integer n, let $V_n = \natural^n (S^1 \times B^3)$, with the boundary connect sums all occurring in neighborhoods of points in the Heegaard surface of each copy of $\partial(S^1 \times B^3)$, so that the induced decomposition $\partial V = \partial^- V \cup \partial^+ V$ is the standard genus n Heegaard splitting of $\#^n (S^1 \times S^2)$. Now, given an integer $s \geq n$, let $\partial V_n = \partial^-_s V_n \cup \partial^+_s V_n$ be the result of stabilizing this Heegaard splitting exactly s times. In what follows, to simplify notation, let $V = V_n$, where $n = k - 2p - b + 1$, and take s to be $g - k + p + b - 1$.

Finally, identify Z_k with $U \natural V$, with the boundary connect sum connecting a neighborhood of a point in the interior of $\partial^- U \cap \partial^+ U$ with a neighborhood of a point in the Heegaard surface $\partial_s^- V \cap \partial_s^+ V$. The induced decomposition of $Y_k = \partial Z_k$ is the advertised decomposition $Y_k = Y^-_{g,k;p,b} \cup Y^0_{g,k;p,b} \cup Y^+_{g,k;p,b}$. To be more specific,

$$Y^{\pm}_{g,k;p,b} = \partial^{\pm} U \natural \partial_s^{\pm} V \quad \text{and} \quad Y^0_{g,k;p,b} = \partial^0 U. \tag{3-2}$$

Before presenting the definition of a trisection relative to the boundary, we make a brief comment on the schematic representation of the stabilization in Figure 3: The illustration shows a "Heegaard splitting" of a 2-manifold, not a 3-manifold, in which case "stabilization" corresponds to introducing a canceling 0-1-handle pair, or 1-2-pair, depending on your perspective, and this is of course not as symmetric as stabilization in dimension 3. In particular, the result is that one half of the splitting becomes disconnected while the other half remains connected. This is the best representation we can give when embedding the schematic in $\mathbb{R}^3$.

Definition 10. A $(g, k; p, b)$*-trisection* of a compact, connected, oriented 4-manifold X with connected boundary is a decomposition of X into three submanifolds $X = X_1 \cup X_2 \cup X_3$ satisfying the following properties:

(1) For each $i = 1, 2, 3$, there is a diffeomorphism $\phi_i : X_i \to Z_k$.

(2) For each $i = 1, 2, 3$, taking indices mod 3, $\phi_i(X_i \cap X_{i+1}) = Y^-_{g,k;p,b}$ and $\phi_i(X_i \cap X_{i-1}) = Y^+_{g,k;p,b}$, while $\phi_i(X_i \cap \partial X) = Y^0_{g,k;p,b}$.

Lemma 11. *A* $(g, k; p, b)$-trisection *of a compact, connected, oriented* 4*-manifold* X *with connected boundary induces an open book decomposition on* ∂X *with pages of genus* p *with* b *boundary components.*

Proof. Each $X_i \cap \partial X$ is diffeomorphic to $Y^0_{g,k;p,b}$, which is diffeomorphic to $([-\pi/3, \pi/3] \times P) \cup (D \times \partial P)$. These three pieces fit together to form ∂X precisely so that the three copies of $[-\pi/3 \times \pi/3] \times P$ form a bundle over S^1 with fiber P, and so that the three copies of $D \times \partial P$ form a $B^2 \times \partial P$, a disjoint union of solid tori that fill the boundary components of the bundle as neighborhoods of the binding components of an open book. □

Theorem 12 [Gay and Kirby 2016]. *Every smooth, compact, oriented, connected* 4*-manifold with connected boundary, with a fixed open book decomposition on the boundary, has a trisection inducing the given open book.*

4. Relative trisections and sutured 3-manifolds, and proofs of the main theorems

In this section we make several observations about our model (Z_k, Y_k). These observations will help us analyze the topology of the corresponding pieces of a

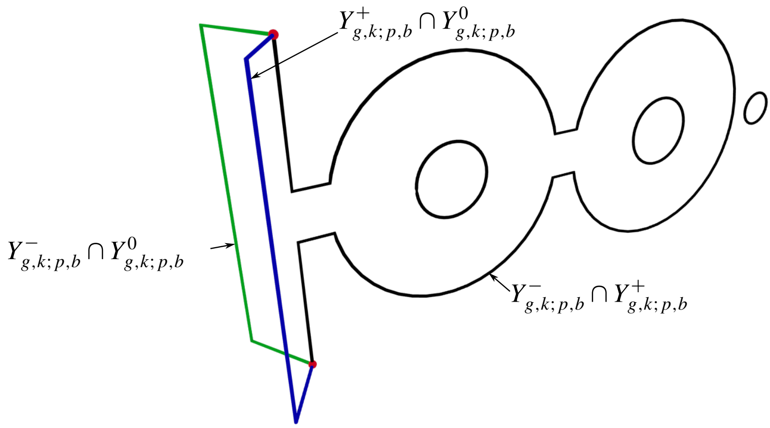


Figure 4. Three "surfaces" in the standard model, as represented in the lower-dimensional schematic. Their common intersection, here shown as a red S^0, is really a disjoint union of b copies of the circle S^1.

relative trisection $X = X_1 \cup X_2 \cup X_3$ and will allow us to identify these spaces with more familiar ones.

(1) The intersection $Y^-_{g,k;p,b} \cap Y^+_{g,k;p,b}$, and hence the triple intersection $X_1 \cap X_2 \cap X_3$, is a surface of genus g with b boundary components. This is schematically illustrated in Figure 3 as a black 1-manifold, see Figure 4.

(2) The intersection $Y^\pm_{g,k;p,b} \cap Y^0_{g,k;p,b}$, and hence $X_i \cap X_{i\mp1} \cap \partial X$, is a surface of genus p with b boundary components, and so diffeomorphic to P. For $i = 1, 2, 3$, these become three pages of the induced open book decomposition of ∂X. In Figure 3, these appear as the two gray ends of the "fan" of pages; Figure 4 isolates the schematic representations of these two surfaces.

(3) The 3-dimensional triple intersection $Y^-_{g,k;p,b} \cap Y^0_{g,k;p,b} \cap Y^+_{g,k;p,b}$, and hence the 4-dimensional intersection $X_1 \cap X_2 \cap X_3 \cap \partial X$, is a disjoint union of b circles. These circles are precisely the components of ∂P, and as such, the binding of the induced open book. This appears schematically in Figure 4 as a red pair of points.

(4) Both $Y^-_{g,k;p,b}$ and $Y^+_{g,k;p,b}$, and hence $X_i \cap X_{i\pm1}$, are 3-dimensional relative compression bodies starting from a surface Σ of genus g with b boundary components and compressing along $g - p$ disjoint simple closed curves to get to a surface P of genus p with b boundary components. Here, by "relative compression body", we mean a cobordism with sides from a high genus surface at the bottom to a low genus surface at the top, each with the same number of boundary components, with a Morse function with critical points only of

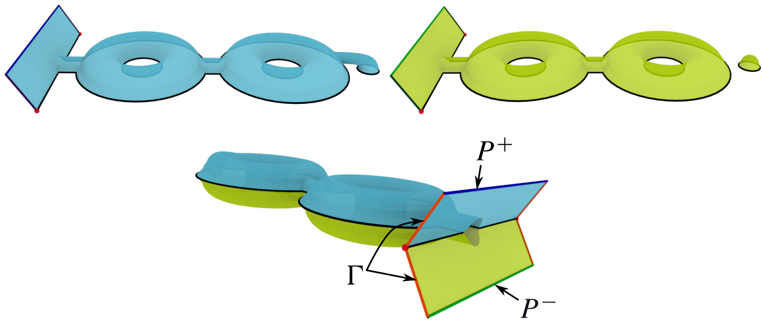


Figure 5. Diagrams concerning relative compression bodies. Top: The two relative compression bodies $Y^-_{g,k;p,b}$ and $Y^+_{g,k;p,b}$, each shown with the high genus "surface" Σ on the bottom, the sides of the cobordism, slanted up and to the left, and the low genus "page" P on the top. Bottom: The two relative compression bodies fit together to form a sutured 3-manifold, depicted here with "sutures" vertical and bent at a $2\pi/3$ angle along the core binding.

index 2. The schematic representations of these two relative compression bodies are illustrated side by side in Figure 5, top.

(5) The union $Y^-_{g,k;p,b} \cup Y^+_{g,k;p,b} = \overline{Y_k \setminus Y^0_{g,k;p,b}}$, and hence each $\overline{\partial X_i \setminus \partial X}$, is a *balanced sutured* 3-*manifold*, with suture equal to a disjoint union of annuli described, in the explicit construction of (Z_k, Y_k) described in Equation (3-2), as $\{r \in [0,1], \theta = \pm\pi/3\} \times \partial P$ with the first factor as in Equation (3-1). Thus each $\overline{\partial X_i \setminus \partial X}$ is a *balanced sutured* 3-*manifold*, with suture Γ equal to a regular neighborhood in $\partial(\overline{\partial X_i \setminus \partial X})$ of the binding. The suture divides the boundary into two remaining pieces P^- and P^+ which, in our case, are, respectively, $\{-\pi/3\} \times P$ and $\{\pi/3\} \times P$. See Figure 5, bottom. Note that, in this paper, annular sutures of a sutured manifold are considered to be *parametrized annuli*, i.e., parametrized as $[-1,1] \times \partial P^-$.

(6) In fact the sutured manifold $Y^-_{g,k;p,b} \cup Y^+_{g,k;p,b} = \overline{Y_k \setminus Y^0_{g,k;p,b}}$, and hence each $\overline{\partial X_i \setminus \partial X} = (X_i \cap X_{i-1}) \cup (X_i \cap X_{i+1})$, is diffeomorphic to

$$([-1,1] \times P) \,\#\, (\#^{k-2p-b+1} S^1 \times S^2),$$

with suture $\Gamma = [-1,1] \times \partial P$ and boundary pieces $P^\pm = \{\pm 1\} \times P$. The decomposition as $Y^+_{g,k;p,b} \cup Y^-_{g,k;p,b}$ is the connected sum of the decomposition of $[-1,1] \times P$ as $([-1,0] \times P) \cup ([0,1] \times P)$ with a $(g-k+b-1)$-times stabilized standard Heegaard splitting of $\#^{k-2p-b+1} S^1 \times S^2$. This gives a standard genus g *sutured Heegaard splitting* of $\overline{\partial X_i \setminus \partial X}$.

(7) There is a diffeomorphism between the surface Σ in Figure 1 and $Y^+_{g,k;p,b} \cap Y^-_{g,k;p,b}$ such that the δ curves in Figure 1 bound disks in $Y^-_{g,k;p,b}$ and the ϵ curves in Figure 1 bound disks in $Y^+_{g,k;p,b}$. Thus $(\Sigma, \delta, \epsilon)$ is a *sutured Heegaard diagram* for

$$Y^+_{g,k;p,b} \cup Y^-_{g,k;p,b} = \overline{Y_k \setminus Y^0_{g,k;p,b}}.$$

(A sutured Heegaard diagram is a triple $(\Sigma, \delta, \epsilon)$ such that Σ is a surface with boundary and each of δ and ϵ is a nonseparating collection of simple closed curves in Σ; such a diagram determines a sutured 3-manifold, balanced if $|\delta| = |\epsilon|$.)

Notice that the decomposition of $Y_{g,k;p,b}$ into $Y^-_{g,k;p,b} \cup Y^0_{g,k;p,b} \cup Y^+_{g,k;p,b}$ can be modified into a decomposition with pieces $Y^-_{g,k;p,b} \cup Y^0_{g,k;p,b}$ and $Y^+_{g,k;p,b}$, by grouping together the first two pieces. This decomposition is the standard genus k Heegaard splitting of $\#^k S^1 \times S^2$ stabilized $g - k + p + b - 1$ times. Notice also that $Y^0_{g,k;p,b}$ can be identified with a collar of the surface P in $Y_{k,g+p+b-1}$. Thus, we can think of the space $Y^-_{g,k;p,b} \cup Y^+_{g,k;p,b} = Y_k \setminus Y^0_{g,k;p,b}$ as the complement of a surface with boundary in a Heegaard splitting and so it is only natural to expect arcs to be part of a notion of diagram for $Y^-_{g,k;p,b} \cup Y^+_{g,k;p,b}$. However, the last two observations indicate that it is possible to avoid the arcs. All this sets the stage for our main technical lemma.

Lemma 13. *Consider a diffeomorphism*

$$\phi : ([-1,1] \times P) \# (\#^l S^1 \times S^2) \to ([-1,1] \times P) \# (\#^l S^1 \times S^2),$$

where the domain and range here are equipped with the sutured structure $\Gamma = [-1,1] \times \partial P$ *and* $P^\pm = \{\pm 1\} \times P$ *discussed above. Suppose that* $\phi|_{\Gamma \cup P^-} = \mathrm{id}$. *Then* $\phi|_{P^+}$ *is isotopic rel. boundary to the identity function* $\mathrm{id} : P^+ \to P^+$.

Proof. To simplify notation, let $M = ([-1,1] \times P) \# (\#^l S^1 \times S^2)$ and consider a properly embedded arc $a \subset P$; this gives rise to a simple closed curve $\gamma_a = (\{0\} \times a) \cup ([0,1] \times \partial a) \cup (\{1\} \times a) \subset \partial M$. Since $\phi|_{\Gamma \cup P^-} = \mathrm{id}$, then $\phi(\gamma_a) = (\{0\} \times a) \cup ([0,1] \times \partial a) \cup (\{1\} \times a')$ for some other arc $a' \subset P$ with the same endpoints as a. Since γ_a bounds a disk in M, so does $\phi(\gamma_a)$ and thus, in fact $\phi(\gamma_a)$ is homotopically trivial in $[-1,1] \times P$. Therefore the loop $\tau_a = a * (a')^{-1}$ obtained by concatenating a and $(a')^{-1}$ is homotopically trivial in P. So a and a' are homotopic rel. endpoints, and thus by a result of Baer [1928], see [Epstein 1966, Theorem 3.1], a and a' are actually isotopic. Apply this to a collection of arcs cutting P into a disk to conclude that $\phi|_{P^+}$ is isotopic rel. boundary to $\mathrm{id} : P^+ \to P^+$. $\square$

In what follows we use this lemma in the following form:

Corollary 14. *Consider the model sutured* 3*-manifold*

$$(([-1,1]\times P)\,\#\,(\#^l S^1\times S^2),\Gamma,P^-,P^+)$$

discussed above, and note that there is an "identity" map $\mathrm{id}: P^-\to P^+$ *defined by* $\mathrm{id}(-1,p)=(1,p)$. *Given any sutured* 3*-manifold*

$$(M,\Gamma_M,P^-_M,P^+_M)$$

diffeomorphic to $(([-1,1]\times P)\,\#\,(\#^l S^1\times S^2),\Gamma,P^-,P^+)$ *there is a unique* (*up to isotopy rel. boundary*) *diffeomorphism* $\mathrm{id}_M : P^-_M\to P^+_M$ *such that*, *for any diffeomorphism*

$$\phi:(M,\Gamma_M,P^-_M,P^+_M)\to(([-1,1]\times P)\,\#\,(\#^l S^1\times S^2),\Gamma,P^-,P^+),$$

we have $\mathrm{id}_M=\phi^{-1}\circ\mathrm{id}\circ\phi$.

We are finally ready to prove the main results of this paper, namely Theorem 3 and Theorem 5. We include the statements of both theorems again to make it easier for the reader to follow our proofs.

Theorem 3. *For every* $(g,k;\,p,b)$*-trisection diagram* $(\Sigma,\alpha,\beta,\gamma)$ *there is a unique* (*up to diffeomorphism*) *trisected* 4*-manifold* $X=X_1\cup X_2\cup X_3$ *with connected boundary*, *such that*, *with respect to a fixed identification* $\Sigma\cong X_1\cap X_2\cap X_3$, *the* α, β *and* γ *curves*, *respectively*, *bound disks in* $X_1\cap X_2$, $X_2\cap X_3$ *and* $X_3\cap X_1$. *In particular*, *the open book decomposition on* ∂X *has* b *binding components and pages of genus* p. *Furthermore*, *any trisected* 4*-manifold with connected boundary is determined in this way by some relative trisection diagram*, *and any two relative trisection diagrams for the same* 4*-manifold trisection are diffeomorphism and handle slide equivalent.*

Proof of Theorem 3. We parallel as much as possible the proof of Theorem 9.

As mentioned above, the diagram (Σ,δ,ϵ) in Figure 1 is a *sutured Heegaard diagram* for $Y^+_{g,k;p,b}\cup Y^-_{g,k;p,b}=Y_k\setminus Y^0_{g,k;p,b}$. Fix an identification of Σ with $Y^-_{g,k;p,b}\cap Y^+_{g,k;p,b}$ such that the δ curves bound disks in $Y^-_{g,k;p,b}$ and the ϵ curves bound disks in $Y^+_{g,k;p,b}$.

Given a trisected 4-manifold $X=X_1\cup X_2\cup X_3$, for $i=1,2,3$, let $\phi_i: X_i\to Z_k$ be the diffeomorphisms from Definition 10. As before, the associated diagram is then $(X_1\cap X_2\cap X_3,\phi_1^{-1}(\delta),\phi_2^{-1}(\delta),\phi_3^{-1}(\delta))$. Equivalently one could replace any $\phi_i^{-1}(\delta)$ with $\phi_{i+1}^{-1}(\epsilon)$, or in fact any other complete nonseparating system of curves bounding disks in $X_i\cap X_{i+1}$; the resulting diagrams would again be handle slide equivalent [Johannson 1995; Casson and Gordon 1987].

Conversely, given a relative $(g,k;\,p,b)$-trisection diagram $(\Sigma,\alpha,\beta,\gamma)$, let C_α, C_β and C_γ, be relative compression bodies built by starting with $I\times\Sigma$ and attaching 3-dimensional 2-handles along α, β and γ, respectively. The boundary of C_α, for

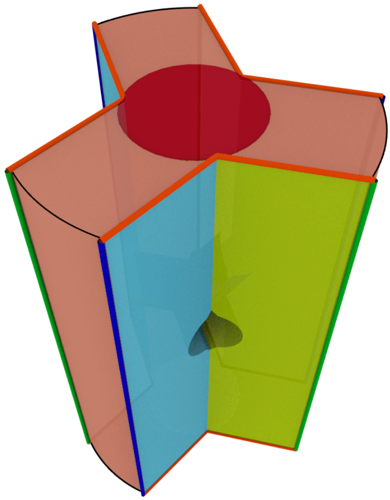


Figure 6. $B^2 \times \Sigma$ with $I\times$ three relative compression bodies.

example, is naturally identified with $\Sigma \cup (I \times \partial\Sigma) \cup \Sigma_\alpha$, where Σ_α is the result of surgery applied to Σ along α. Let $P = \Sigma_\alpha \cong \Sigma_\beta \cong \Sigma_\gamma$.

Build X by starting with $B^2 \times \Sigma$, attaching $I \times C_\alpha$, $I \times C_\beta$ and $I \times C_\gamma$ to $\partial B^2 \times \Sigma = S^1 \times \Sigma$ along the product of successive arcs in S^1 with Σ. This produces a 4-manifold with boundary naturally divided into $B^2 \times \Sigma$, three copies of $(I \times P) \cup (I \times I \times \partial P)$ and three sutured 3-manifolds diffeomorphic to

$$(([-1, 1] \times P) \,\#\, (\#^l S^1 \times S^2), \Gamma, P^-, P^+).$$

The three sutured manifolds are as advertised because each of (Σ, α, β), (Σ, β, γ) and (Σ, γ, α) is handle slide and diffeomorphism equivalent to the standard sutured Heegaard diagram $(\Sigma, \delta, \epsilon)$ discussed above. This is illustrated in Figure 6. Using Corollary 14, there is a unique way to glue $([-1, 1] \times P) \cup (D \times \partial P)$, that is one third of an open book, to each of these sutured 3-manifolds. Thickening the three pieces we have glued on to be 4-dimensional, we get a 4-manifold with four boundary components: one on the "outside", equal to an open book decomposition with page P, and three "inside" boundary components each diffeomorphic to $\#^k S^1 \times S^2$. This is illustrated in Figure 7, in which at the last stage we only see the outer boundary. Cap off each of the inside boundary components with $\natural^k S^1 \times B^3$ (uniquely, by [Laudenbach and Poénaru 1972]). The end result is our trisected 4-manifold $X = X_1 \cup X_2 \cup X_3$. (Each X_i is the union of a third of $B^2 \times \Sigma$, half of I cross one relative compression body, half of I cross the next relative compression body, the thickened copy of $([-1, 1] \times P) \cup (D \times \partial P)$ glued in to this third, and the corresponding copy of $\natural^k S^1 \times B^3$.) $\square$

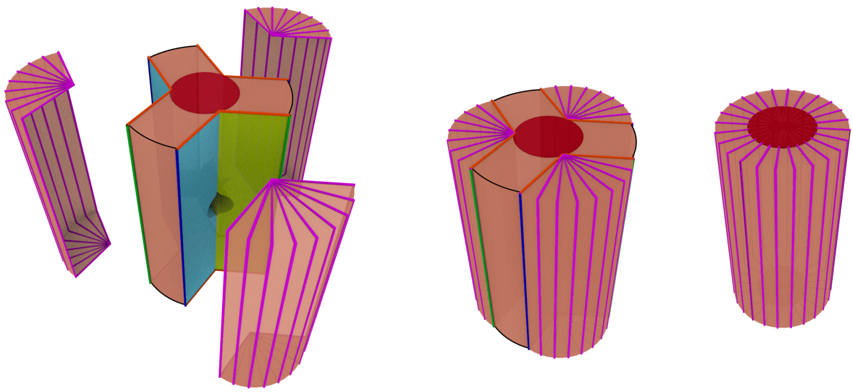


Figure 7. Gluing on three groups of pages and closing up.

Theorem 5. *A relative trisection diagram* $(\Sigma, \alpha, \beta, \gamma)$ *encodes an open book decomposition on* ∂X *with page given by* Σ_α*, the surface resulting from* Σ *by performing surgery along the* α *curves, and monodromy* $\mu : \Sigma_\alpha \to \Sigma_\alpha$ *determined by the following algorithm*:

(1) *Choose an ordered collection of arcs* a *on* Σ*, disjoint from* α *and such that its image* $\phi_\alpha(a)$ *in* Σ_α *cuts* Σ_α *into a disk.*

(2) *There exists a collection of arcs* a_1 *and simple closed curves* β' *in* Σ *such that* (α, a_1) *is handle slide equivalent to* (α, a), β' *is handle slide equivalent to* β, *and* a_1 *and* β' *are disjoint.* (*We claim that in this step we do not need to slide* α *curves over* α *curves*, *only* a *arcs over* α *curves and* β *curves over* β *curves.*) *Choose such an* a_1 *and* β'

(3) *There exists a collection of arcs* a_2 *and simple closed curves* γ' *in* Σ *such that* (β', a_2) *is handle slide equivalent to* (β', a_1), γ' *is handle slide equivalent to* γ, *and* a_2 *and* γ' *are disjoint.* (*Again we claim that we do not need to slide* β' *curves over* β' *curves.*) *Choose such an* a_2 *and* γ'

(4) *There exists a collection of arcs* a_3 *and simple closed curves* α' *in* Σ *such that* (γ', a_3) *is handle slide equivalent to* (γ', a_2), α' *is handle slide equivalent to* α, *and* a_3 *and* α' *are disjoint.* (*Again we do not need to slide* γ' *curves over* γ' *curves.*) *Choose such an* a_3 *and* α'.

(5) *The pair* (α', a_3) *is handle slide equivalent to* (α, a_*) *for some collection of arcs* a_*. *Choose such an* a_*. *Note that now* a *and* a_* *are both disjoint from* α *and thus we can compare* $\phi_\alpha(a)$ *and* $\phi_\alpha(a_*)$ *in* Σ_α.

(6) *The monodromy* μ *is the unique* (*up to isotopy*) *map such that*

$$\mu(\phi_\alpha(a)) = \phi_\alpha(a_*),$$

respecting the ordering of the collections of arcs.

Proof of Theorem 5. The fact that each of (Σ, α, β), (Σ, β, γ) and (Σ, γ, α) is handle slide and diffeomorphism equivalent to the sutured Heegaard diagram $(\Sigma, \delta, \epsilon)$ in Figure 1 tells us that we can in fact find the collections of arcs and sequences of slides advertised. Each time we find a collection of arcs which is disjoint from, for example, both β and γ, this describes a diffeomorphism from Σ_β to Σ_γ, which is the "identity" map coming from Corollary 14. Thus we have the following steps:

(1) Note that $\phi_\alpha(a)$ is isotopic to $\phi_\alpha(a_1)$ in Σ_α because a_1 was produced from a by sliding over α curves.

(2) Map Σ_α to $\Sigma_{\beta'}$ so as to send $\phi_\alpha(a_1) \subset \Sigma_\alpha$ to $\phi_{\beta'}(a_1) \subset \Sigma_{\beta'}$.

(3) Note that $\phi_{\beta'}(a_1)$ is isotopic to $\phi_{\beta'}(a_2)$ in $\Sigma_{\beta'}$ because a_2 was produced from a_1 by sliding over β' curves.

(4) Map $\Sigma_{\beta'}$ to $\Sigma_{\gamma'}$ so as to send $\phi_{\beta'}(a_2) \subset \Sigma_{\beta'}$ to $\phi_{\gamma'}(a_2) \subset \Sigma_{\gamma'}$.

(5) Note that $\phi_{\gamma'}(a_2)$ is isotopic to $\phi_{\gamma'}(a_3)$ in $\Sigma_{\gamma'}$ because a_3 was produced from a_2 by sliding over γ' curves.

(6) Map $\Sigma_{\gamma'}$ to $\Sigma_{\alpha'}$ so as to send $\phi_{\gamma'}(a_3) \subset \Sigma_{\gamma'}$ to $\phi_{\alpha'}(a_3) \subset \Sigma_{\alpha'}$.

(7) Map $\Sigma_{\alpha'}$ to Σ_α so as to send $\phi_{\alpha'}(a_3)$ to $\phi_\alpha(a_*)$.

The fact that each of the maps in the above sequence of maps is independent of the choices is a restatement of Corollary 14, and thus we see the monodromy expressed as a composition $\Sigma_\alpha \to \Sigma_{\beta'} \to \Sigma_{\gamma'} \to \Sigma_{\alpha'} \to \Sigma_\alpha$. □

5. Examples

5.1. ***Disk bundles over the 2-sphere S^2.*** Consider $p : E_n \to S^2$ the oriented disk bundle over S^2 with Euler number n. Decompose S^2 as the union of three wedges B_1, B_2, B_3 that intersect pairwise in arcs joining the north and south pole and whose triple intersection consists precisely of the north and south poles as shown in Figure 8. Ideally, we would just lift this trisection of S^2 to get a trisection for E_n. However, although each $p^{-1}(B_i)$ is in fact a 4-dimensional 1-handlebody, the triple intersection of these pieces is not connected and so this naive decomposition of E_n is not really a trisection. To fix this, for $i, j = 1, 2, 3$ let $\varphi_i : B_i \times D^2 \to p^{-1}(B_i)$ be a trivialization over B_i and let $g_{ij} : B_i \cap B_j \to \mathrm{SO}(2)$ be the transition function for $\varphi_i^{-1} \circ \varphi_j$. Next, parametrize each arc $B_i \cap B_{i+1}$ by $t \in [0, 1]$ and use the cocycle condition to set

$$g_{12}, g_{23} : t \to 1, \quad \text{and} \quad g_{31} : t \to e^{2\pi int}. \tag{5-1}$$

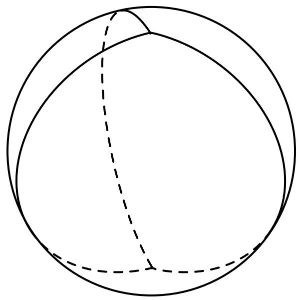


Figure 8. Decomposition of $S^2 = B_1 \cup B_2 \cup B_3$.

Here we are using the identification $e^{i\theta} = \left(\begin{smallmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{smallmatrix}\right)$ and the notion of cocycle condition from [Davis and Kirk 2001]. In addition, choose sections σ_i over B_i ($i = 1, 2, 3$) disjoint from one another and so that at each point $b \in B_i$, $\sigma_i(b)$ lies in the interior of the fiber $p^{-1}(\{b\})$. Let $\nu_i \cong B_i \times N_i$ be a tubular neighborhood of $\sigma_i(B_i)$ in $p^{-1}(B_i)$, and also assume that these tubular neighborhoods are pairwise disjoint and that at each point $b \in B_i$, the vertical direction of ν_i at b lies in the interior of $p^{-1}(b)$. Finally, set

$$X_i = \overline{p^{-1}(B_i) \setminus \nu_i} \underset{\varphi_i \circ \varphi_{i+1}^{-1}}{\cup} \nu_{i+1},$$

where the gluing is done via

$$\varphi_i \circ \varphi_{i+1}^{-1} : \nu_{i+1} \cap p^{-1}(B_i) \to \nu_{i+1} \cap p^{-1}(B_i).$$

Notice that since ν_i is a 2-handle, removing it from $p^{-1}(B_i)$ results in a space diffeomorphic to $S^1 \times B^3$. In addition, since ν_{i+1} is attached along $\nu_{i+1} \cap p^{-1}(B_i)$ and this set is a 3-ball, attaching ν_{i+1} does not change the diffeomorphism type and thus X_i is diffeomorphic to $S^1 \times B^3$.

For the X_i's to define a trisection of E_n, we need to check that the intersections between them behave in the way stipulated in Definition 10. With this in mind, consider first the pairwise intersection $X_{i-1} \cap X_i$ and notice that this intersection is such that

$$\begin{aligned} X_{i-1} \cap X_i & \\ &= \overline{\left(p^{-1}(B_{i-1} \cap B_i) \setminus (\nu_{i-1} \cup \nu_i)\right)} \underset{\varphi_{i-1} \circ \varphi_i^{-1}}{\cup} \partial_i \nu_i \underset{\varphi_i \circ \varphi_{i+1}^{-1}}{\cup} (\nu_{i+1} \cap p^{-1}(B_{i-1})). \end{aligned} \tag{5-2}$$

Here $\left(p^{-1}(B_{i-1} \cap B_i) \setminus (\nu_{i-1} \cup \nu_i)\right)$ is diffeomorphic to a 3-ball with two 2-handles removed, and $\nu_{i+1} \cap p^{-1}(B_{i-1})$ is a 1-handle. Moreover, the set $\partial_i \nu_i \cong B_i \times \partial N_i$, the boundary of ν_i as a subspace of $p^{-1}(B_i)$, is a solid torus attached to the 3-ball with two 2-handles removed along a cylinder in its boundary and thus is simply a thickening of one of the holes left by the 2-handles. We can then conclude that $X_{i-1} \cap X_i$ is diffeomorphic to a handlebody of genus 3. An extension of the previous argument then shows that the triple intersection is given by

$$\begin{aligned} X_1 \cap X_2 \cap X_3 & \\ &= p^{-1}(B_1 \cap B_2 \cap B_3) \setminus (\nu_1 \cup \nu_2 \cup \nu_3) \bigcup_{\substack{\varphi_i \circ \varphi_{i+1}^{-1} \\ i=1,2,3}} \left[\bigcup_{i=1}^{3} \partial_i \nu_i \cap p^{-1}(B_{i+1}) \right], \end{aligned} \tag{5-3}$$

where $p^{-1}(B_1 \cap B_2 \cap B_3) \setminus (\nu_1 \cup \nu_2 \cup \nu_3)$ consists of the disjoint union of two 2-disks with three interior disks removed, and each $\partial_i \nu_i \cap p^{-1}(B_{i+1})$ is diffeomorphic to the cylinder $B_i \cap B_{i+1} \times \partial N_i$ and is glued to the first space in such a way that it joins internal boundary components of the two different disks. From this it follows that the triple intersection is a twice punctured genus two surface. The last intersections to consider are those that involve the boundary, namely, $X_i \cap E_n$ and $X_{i-1} \cap X_i \cap \partial E_n$. In this case we have

$$X_i \cap \partial E_n = \partial p^{-1}(B_i) \setminus p^{-1}(\partial B_i) \cong B_i \times \partial D^2,$$

and

$$X_{i-1} \cap X_i \cap \partial E_n = \partial p^{-1}(B_{i-1} \cap B_i) \setminus p^{-1}(\partial B_{i-1} \cap \partial B_i) \cong B_{i-1} \cap B_i \times \partial D^2.$$

From this we see that $X_i \cap E_n$ is diffeomorphic to $I \times X_{i-1} \cap X_i \cap \partial E_n$ with the space $\partial I \times X_{i-1} \cap X_i \cap \partial E_n$ identified, or, using the terminology of Equation (3-1), that $X_i \cap E_n$ is diffeomorphic to $\partial^0 D \times (X_{i-1} \cap X_i \cap \partial E_n) \cup D \times \partial(X_1 \cap X_2 \cap X_3)$.

In sum, the previous paragraphs describe a (2, 1; 0, 2) relative trisection of E_n whose relative trisection diagram $(\Sigma, \alpha, \beta, \gamma)$ has yet to be exhibited. To this end, notice that by Equation (5-3), Σ is a surface decomposed as the union of two copies of a three times punctured disk with three cylinders joining the punctures of the two disks. To finish the description of the diagram, it is enough to find three sets of curves in $F = X_1 \cap X_2 \cap X_3$ that bound disks in the double intersections $X_{i-1} \cap X_i$, and draw their images in Σ. For example, in $X_3 \cap X_1$, the 1-handle $\nu_2 \cap p^{-1}(B_3)$ has the cylinder $\partial_2 \nu_2 \cap p^{-1}(B_3)$ as its boundary and so the central circle in the latter is one of the curves in the collection γ. A similar argument applied to the other two pairwise intersections shows that the central circle in $\partial_3 \nu_3 \cap p^{-1}(B_1)$ is a curve in α and $\partial_1 \nu_1 \cap p^{-1}(B_2)$ is a curve in β. Next, consider the disk $\mathcal{D}$ in $X_3 \cap X_1$ constructed as the union of a disk in $p^{-1}(B_3 \cap B_1) \setminus \nu_3 \cup \nu_1$ that lies between the holes left by ν_3, ν_1 with a meridional disk in $\partial_3 \nu_3$. Then, the curve $\partial \mathcal{D}$ can be realized as the union of:

(i) a properly embedded arc in $\partial_3 \nu_3 \cap p^{-1}(B_1)$ with one endpoint in each boundary component,

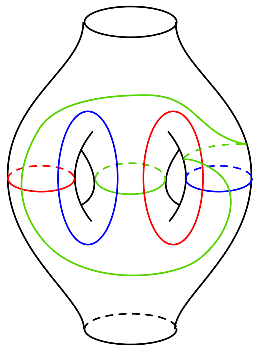


Figure 9. A $(2, 1; 0, 2)$ relative trisection diagram for the disk bundle over S^2 corresponding to the integer -1. The monodromy of the open book in the boundary is a left handed twist.

(ii) a properly embedded arc in $\partial_1 \nu_1 \cap p^{-1}(B_2)$ with one endpoint in each boundary component, and

(iii) two horizontal arcs that lie in different components of the disjoint union of disks $(p^{-1} \setminus \nu_1 \cup \nu_2 \cup \nu_3)(B_1 \cap B_2 \cap B_3)$.

This curve $\partial \mathcal{D}$ is the second curve in the collection γ and to draw it in Σ we have to proceed with caution since by assumption the gluing map $\varphi_3 \circ \varphi_1^{-1}$ depends on n. Indeed, using Equation (5-1) we see that the two disks that make up $\mathcal{D}$ align only if the second one is twisted. Thus, the arc described in (ii) appears in $\partial_1 \nu_1 \cap p^{-1}(B_2)$ as an arc with n-twists. Lastly, to get the remaining curves in α and β, we proceed in a similar manner noticing that in these cases the gluing maps are trivial and thus the analogous arcs to the one from (ii) are not twisted. This shows that the trisection diagram corresponding to the decomposition $E_n = X_1 \cup X_2 \cup X_3$ can be obtained from the one shown in Figure 9 by replacing the single left handed twist on the green curve appearing in the right, with n full twists around the cylinder.

5.2. ***Local modifications of diagrams, Lefschetz fibrations and Hopf plumbings.*** Throughout this section, suppose that we are given a relative trisection diagram $(\Sigma, \alpha, \beta, \gamma)$ for a trisected 4-manifold $X = X_1 \cup X_2 \cup X_3$, with induced open book on ∂X with page $P = \Sigma_\alpha$ and monodromy $\mu : P \to P$.

Lemma 15. *Let $\Sigma' \supset \Sigma$ be the result of attaching a 2-dimensional 1-handle to Σ along some $S^0 \subset \partial \Sigma$. Then the tuple $(\Sigma', \alpha, \beta, \gamma)$ is a relative trisection diagram for a trisected 4-manifold $X' = X_1' \cup X_2' \cup X_3'$ such that X' is the result of attaching a 4-dimensional 1-handle H to X along the same $S^0 \subset \partial \Sigma$, seeing $\partial \Sigma \subset \partial X$ as the binding of the open book on ∂X. Furthermore, $H = H_1 \cup H_2 \cup H_3$, where each H_i is a 4-dimensional 1-handle attached to X_i to form X_i'. The open book on $\partial X'$ has*

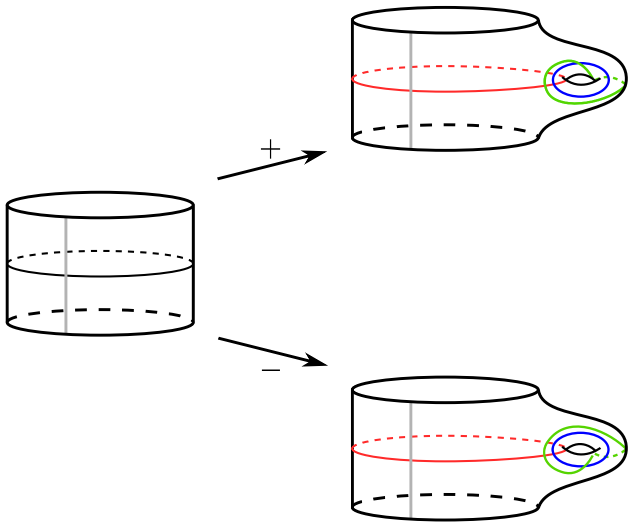


Figure 10. Local modification of $(\Sigma, \alpha, \beta, \gamma)$ near a curve C disjoint from α and transverse to β and γ. The gray transverse arc represents a collection of parallel β and γ arcs.

page $P' = P \cup h$, the result of attaching the 2-dimensional 1-handle h to P, and monodromy μ' equal to μ extended by the identity across h.

Proof. Let h be the 2-dimensional 1-handle attached to Σ to form Σ'. In the construction of X and X', we see that X is naturally a subset of X' and that $X' \setminus X$ is precisely a 1-handle $H = B^2 \times h$. Splitting B^2 into three thirds $B^2 = D_1 \cup D_2 \cup D_3$ gives the three 1-handles $H_i = D_i \times h$. □

Lemma 16. *Consider a simple closed curve $C \subset \Sigma$ disjoint from α and transverse to β and γ. Let $(\Sigma^\pm, \alpha^\pm, \beta^\pm, \gamma^\pm)$ be the result of removing a cylinder neighborhood of C, together with the β and γ arcs running across this neighborhood, and replacing it with a twice-punctured torus as in Figure 10 with β and γ arcs as drawn, and with one new α, β and γ curve as drawn. Then $(\Sigma^\pm, \alpha^\pm, \beta^\pm, \gamma^\pm)$ is a relative trisection diagram for a trisected 4-manifold $X' = X_1' \cup X_2' \cup X_3'$ such that X' is the result of attaching a 2-handle to X along $C \subset P$ with framing ∓ 1 relative to P, and such that the open book on $\partial X'$ has page P with monodromy $\tau_C^{\pm 1} \circ \mu$, where τ_C is a right-handed Dehn twist about C.*

Proof. Since $(\Sigma, \alpha, \beta, \gamma)$ is a trisection diagram, we know that there is an arc A connecting C to ∂X avoiding α and transverse to β and γ; we draw a neighborhood of $C \cup A$ as on the left in Figure 11. In this picture there are two groups of β and γ arcs: those transverse to C and those transverse to A. The modification drawn in Figure 10 is then redrawn in Figure 11 so that we see the new genus in Σ' as

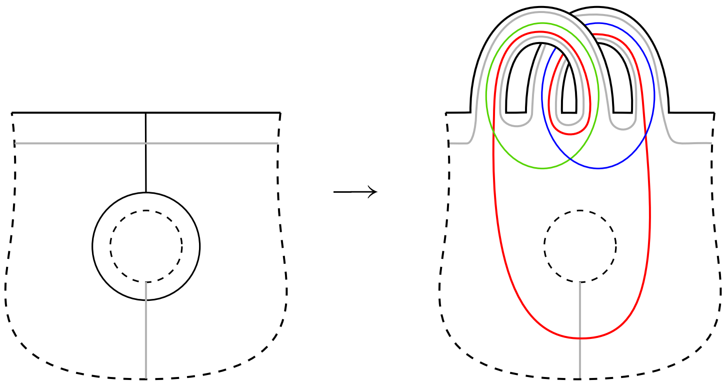


Figure 11. A different perspective of the local modification of $(\Sigma, \alpha, \beta, \gamma)$, taking into account an arc A connecting C to $\partial\Sigma$. Again, the gray arcs represent collections of parallel β and γ arcs; now one collection of such arcs is transverse to the closed curve C and one collection is transverse to the arc A.

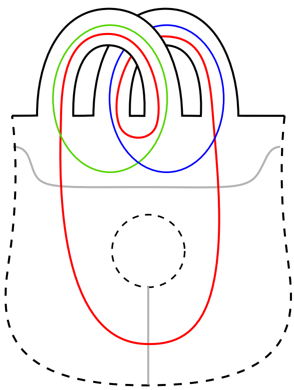


Figure 12. After some handle slides.

arising from Σ by attaching two 2-dimensional 1-handles h_1 and h_2. The β and γ arcs that were transverse to A avoid the new α, β and γ curves by running parallel to $\partial\Sigma'$. Note that we can slide these boundary-parallel β and γ arcs over the new β or, respectively, γ curve to get Figure 12. (Each β (resp. γ) arc slides twice over the β (resp. γ) curve.) Thus we can take Figure 12 to be the modification of the trisection diagram which we work with; i.e., $(\Sigma^\pm, \alpha^\pm, \beta^\pm, \gamma^\pm)$ is obtained from $(\Sigma, \alpha, \beta, \gamma)$ by replacing the figure on the left in Figure 11 with Figure 12.

Now, recalling the construction of X from the diagram $(\Sigma, \alpha, \beta, \gamma)$ and of X' from $(\Sigma^\pm, \alpha^\pm, \beta^\pm, \gamma^\pm)$, we see that X' is naturally built by adding two 4-dimensional 1-handles to X (as in Lemma 15) followed by three 4-dimensional

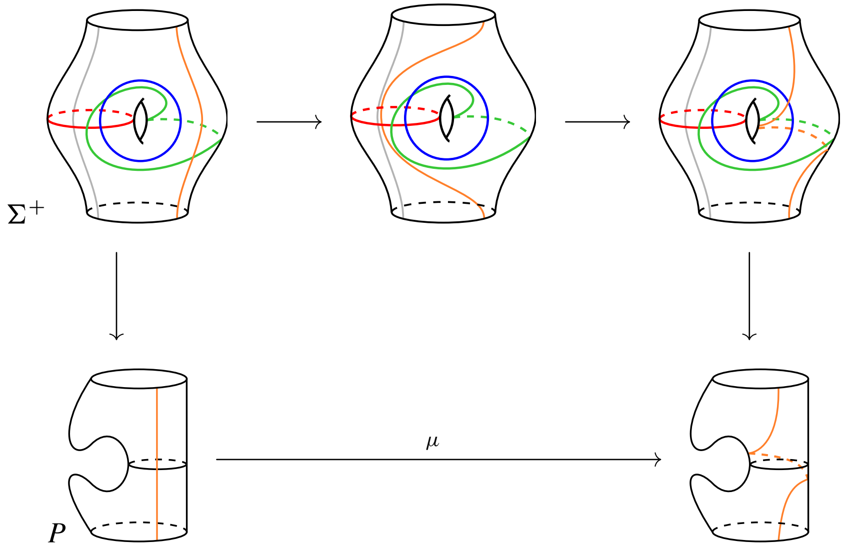


Figure 13. Local effect on the monodromy.

2-handles, one along the new α curve in Σ'_α, one along the new β curve in Σ'_β and one along the new γ curve in Σ'_γ, with 0-framings relative to the pages in which they sit. The β and γ 2-handles each, topologically, cancel one of the new 1-handles, and when this cancellation is performed, we see that the α curve now sits in Σ_α with framing equal to ± 1 with respect to Σ_α.

Figure 13 shows a local implementation of the algorithm from Theorem 5 to show the effect of the new monodromy on a single arc transverse to C, thus completing the proof of the lemma. □

Note that the roles of α, β and γ in Lemma 16 can obviously be cyclically permuted; in some of the following applications, γ will play the role that α plays here.

Notice also that if $(\Sigma, \alpha, \beta, \gamma)$ is a relative trisection diagram for a $(g, k;\, p, b)$ trisection, then the tuple $(\Sigma', \alpha, \beta, \gamma)$ from Lemma 15 is a relative trisection diagram of a $(g+1, k+1;\, p+1, b)$ trisection or a $(g, k+1;\, p, b+1)$ trisection depending on whether the chosen 0-sphere $S^0 \subset \partial\Sigma$ is contained in different components of $\partial\Sigma$ or in the same one. Similarly, the tuple $(\Sigma^\pm, \alpha^\pm, \beta^\pm, \gamma^\pm)$ from Lemma 16 is a relative trisection diagram of a $(g+1, k;\, p, b)$ trisection.

We have two immediate corollaries. The first describes a stabilization operation on trisection diagrams corresponding to Hopf plumbing on the bounding open book decomposition, and is the diagrammatic version of the construction described in Section 3.3 of [Castro 2016].

Corollary 17. *Suppose that X has a trisection T with induced open book decomposition D on ∂X, and that D^+ (resp. D^-) is an open book decomposition of ∂X obtained from D by plumbing a left-handed (resp. right-handed) Hopf band along a properly embedded arc A in a page P of the open book D. If T is described by the relative trisection diagram $(\Sigma, \alpha, \beta, \gamma)$ such that P is identified with Σ_α, consider the new diagram $(\Sigma'^{\pm}, \alpha^{\pm}, \beta^{\pm}, \gamma^{\pm})$ obtained by first attaching a 2-dimensional 1-handle to Σ at the end points of A, as in Lemma 15, producing $(\Sigma', \alpha, \beta, \gamma)$ and then modifying this as in Lemma 16 in a neighborhood of the curve C obtained by attaching the core of the 1-handle to the arc A. Then $(\Sigma'^{\pm}, \alpha^{\pm}, \beta^{\pm}, \gamma^{\pm})$ is again a trisection of X inducing the open book decomposition $D^{\pm}$ on ∂X.*

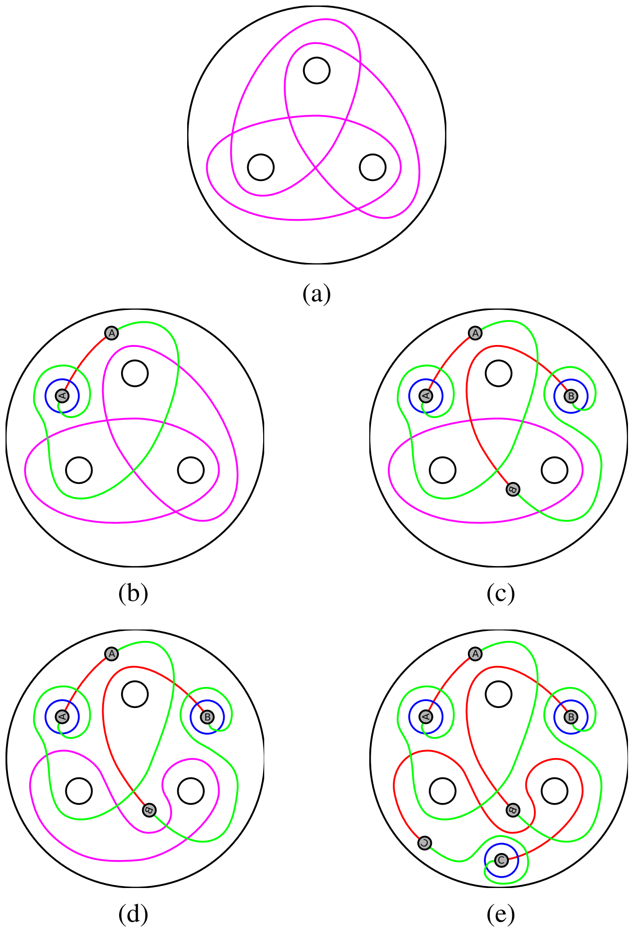
We leave the proof of this corollary to the reader.

For the next corollary, let P be a smooth orientable surface with boundary and for c a curve embedded in P, denote by τ_c the right handed twist of P along c. Given a 3-manifold Y with open book decomposition given by (P, μ) with μ factored as $\mu = \tau_{c_n}^{\epsilon_n} \circ \cdots \circ \tau_{c_1}^{\epsilon_1}$ with $\epsilon_i \in \{-1, 1\}$, and c_i a curve in P, $i = 1, \ldots, n$ it is well known that Y is the boundary of a 4-manifold X admitting an achiral Lefschetz fibration over D^2 with vanishing cycles $c_1, \ldots, c_n$. Moreover [Kas 1980], X admits a handle decomposition diffeomorphic to the result of attaching n 2-handles $h_1^2, \ldots, h_n^2$, to $D^2 \times P$ along the circles $\{1\} \times c_i$ with framing given by the surface framing minus ϵ_i.

Corollary 18. *Let $\pi : X \to D^2$ be an achiral Lefschetz fibration with regular fiber a surface P of genus p and b boundary components, and with n vanishing cycles. The manifold X admits a $(p+n, 2p+b-1; p, b)$ trisection.*

Proof. Build X and its trisection beginning with the standard $(0, 0; 0, 1)$ trisection of B^4 and attaching 1-handles as in Lemma 15 to produce $P \times D^2$ with a trisection inducing the standard open book on $P \times S^1$ with page P and identity monodromy. At this stage the central surface Σ^0 is P, and there are no α, β or γ curves. Attach a 2-handle along c_1 as in Lemma 16 to get a new $(\Sigma^1, \alpha^1, \beta^1, \gamma^1)$, such that each of α^1, β^1 and γ^1 consists of a single curve, and P is identified with $\Sigma^1_{\alpha^1}$. Now, as i goes from 2 to n repeat the following process: Pull c_i back from P to Σ^{i-1}, using the fact that P is identified with $\Sigma^{i-1}_{\alpha^{i-1}}$, and then apply Lemma 16 to $c_i \subset \Sigma^{i-1}$ to produce $(\Sigma^i, \alpha^i, \beta^i, \gamma^i)$, with P again identified with $\Sigma^i_{\alpha^i}$. □

The subtlety in implementing the method of proof above in a particular example arises when the vanishing cycles intersect. The images in Figure 14 illustrate a slightly nontrivial example, in which the vanishing cycles correspond to one side of the lantern relation in the mapping class group of a genus 0 surface with four boundary components. The end result is a relative trisection diagram for a well known rational homology 4-ball with boundary $L(4, 1)$; see [Endo and Gurtas 2010; Fintushel and Stern 1997]. Note that from Figure 14(c) to Figure 14(d) we need to isotope the third vanishing cycle so as to be disjoint from a red α curve before



(a) (b) (c) (d) (e)

Figure 14. A relative trisection diagram for a rational homology 4-ball with boundary $L(4, 1)$. (a) Three vanishing cycles on a genus 0 surface with 4 boundary components. (b) One vanishing cycle turned into α, β and γ curves, genus now equal to 1. (c) Two vanishing cycles done, genus equals 2; note that C_3 now intersects α curves. (d) C_3 isotoped to intersect only γ curves. (e) A rational homology B^4.

proceeding to Figure 14(e). This corresponds to adjusting our drawing so that the third vanishing cycle does in fact live in the page obtained by surgering the central surface along the α curves.

5.3. ***Plumbings.*** In this section, we explain how to combine the method to obtain a diagram for achiral Lefschetz fibrations with well-known facts about plumbings of disk bundles over surfaces to describe trisection diagrams for plumbings of disk bundles. Notice however that for a single disk bundle of large Euler class, this method gives a much higher genus trisection than the method in Section 5.1.

Definition 19. A plumbing graph is a finite connected graph Γ whose vertices and edges are assigned weights as follows:

- each vertex v of Γ carries two integer weights e_v, and g_v, with $g_v \geq 0$,
- each edge of Γ is assigned a sign $+1$ or -1.

To simplify notation, denote by $V(\Gamma)$ the set of vertices, $E(\Gamma)$ the set of edges, and $Q(\Gamma)$ the incidence matrix of Γ, that is, the matrix whose q_{vw} entry is given by the signed count of edges joining the vertices v and w if $v \neq w$, and $q_{vv} = e_v$. In addition, for every vertex v let $s_v = \sum_{w \in V(\Gamma)} q_{vw}$. Then, if d_v is the degree of v, or in other words the weighted sum of edges that intersect v, we have $s_v = e_v + d_v$.

Definition 20. Given a plumbing graph Γ, its modified plumbing graph is the connected graph Γ^* that results from adding loose edges (edges with only one end at a vertex and the other end "loose") to Γ as follows:

- at each vertex v of Γ attach $|s_v|$ loose edges,
- to each loose edge assign the sign of $-s_v$.

If we call $\mathcal{L}(\Gamma^*)$ the set of loose edges and if we let D be the diagonal matrix with entries given by the sums s_v, using the notation introduced after Definition 19 we have

$$\begin{aligned} V(\Gamma^*) &= V(\Gamma), \\ E(\Gamma^*) &= E(\Gamma) \cup \mathcal{L}(\Gamma^*), \\ Q(\Gamma^*) &= Q(\Gamma). \end{aligned}$$

To a modified plumbing graph Γ^* with underlying plumbing graph Γ, one can associate a surface $F(\Gamma^*)$ and a set of vanishing cycles as follows: Assign to each vertex v the closed orientable surface of genus g_v and to each loose end of a loose edge a disk D^2 and connect these surfaces with tubes according to Γ^* to obtain the surface $F(\Gamma^*)$ (i.e., for each edge, replace two disks, one in the interior of each surface corresponding to the ends of the edge, with $[0, 1] \times S^1$). The vanishing cycles are simply the necks of the tubes (explicitly, the curves $\{1/2\} \times S^1 \subset [0, 1] \times S^1$) used in the construction of $F(\Gamma^*)$ and each vanishing cycle's framing is equal to the sign ± 1 of the edge of Γ^* giving rise to that tube.

Lemma 21. *Let Γ be a plumbing graph. Then there exists an* (*achiral*) *Lefschetz fibration $\pi : L(\Gamma) \to D^2$ with the following properties*:

(i) *the regular fiber of π is diffeomorphic to $F(\Gamma^*)$*,

(ii) *the vanishing cycles and their framings correspond to edges in Γ^* and their signs*,

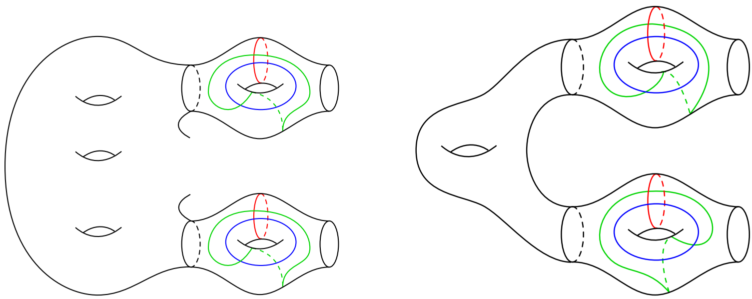


Figure 15. Relative trisection diagrams for the disk bundles over closed orientable surfaces. Left: Disk bundle over a closed surface with Euler number $n < 0$. Right: Disk bundle over a torus with Euler number 0.

(iii) *the monodromy μ is equal to the signed product of Dehn twists along the vanishing cycles.*

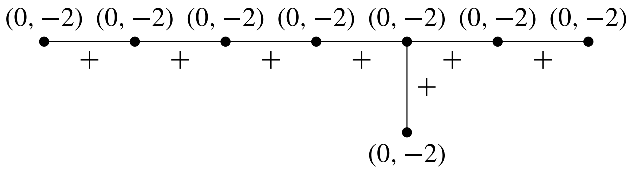
Furthermore, the 4*-manifold $P(\Gamma)$ obtained as a plumbing of disk bundles of surfaces according to a plumbing graph Γ and $L(\Gamma)$ constructed from the given vanishing cycle data are diffeomorphic.*

Proof. To see that $L(\Gamma)$ is diffeomorphic to $P(\Gamma)$ we need to show that $L(\Gamma)$ is a regular neighborhood of a collection of surfaces of the right genus transversely- and self-intersecting according to Γ. Since all the vanishing cycles are disjoint on $F(\Gamma^*)$, we can see $L(\Gamma)$ as a Lefschetz fibration with exactly one singular fiber containing all the singularities. Since each vanishing cycle becomes a transverse intersection point in the singular fiber, with sign given by the sign of the vanishing cycle, we immediately get the correct configuration of surfaces. Since there is only one singular value, $L(\Gamma)$ is a neighborhood of that singular fiber. □

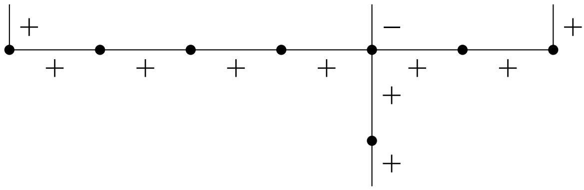
Lemma 21 can be combined with Corollary 18 to obtain trisections and trisection diagrams for plumbing manifolds. For example, if Σ is the closed orientable surface of genus $G > 1$ and $p : E_n \to \Sigma$ is the disk bundle over Σ with Euler number n, let $\pi : E_n \to D^2$ be the (achiral) Lefschetz fibration described in Lemma 21. If $n \neq 0$, there is a $(|n|+G, |n|+2G-1; G, |n|)$ trisection of E_n with diagram given by Figure 15, left. If $n = 0$, there is a $(G+2, 2G+1; G, 2)$ trisection of E_n with diagram given by Figure 15, right.

A less trivial example is the negative definite E_8 manifold. The plumbing graph, the modified plumbing graph, the regular surface, and the trisection diagram are shown in Figure 16.

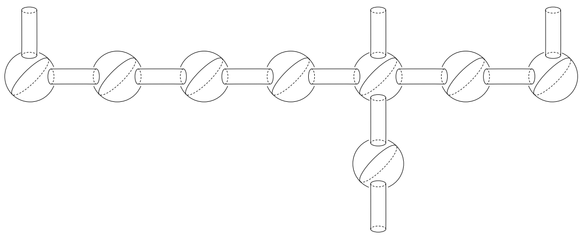
5.4. ***The product of the circle with knot complements.*** In this section we show that if a knot $K \subset S^3$ is in bridge position with B bridges, then $X = S^1 \times S^3 \setminus N(K)$



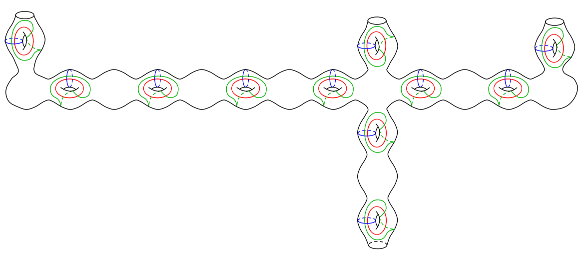
The plumbing graph E_8.



The modified plumbing graph E_8^*.



The regular fiber of $L(E_8)$.



The trisection diagram of $P(E_8)$.

Figure 16. The negative definite E_8 manifold. Its boundary is the Poincaré homology sphere.

admits a $(6B-1, 2B+1; 1, 4)$ trisection. The description of the trisection and the trisection diagram will depend on the notion of doubly pointed diagrams for knots in S^3 and so we begin the section with its definition. For the details regarding this construction we refer the reader to [Rasmussen 2003, Section 3.2; Manolescu 2016, Example 3.4].

Definition 22. A doubly-pointed diagram for a knot $K \subset S^3$, is a tuple

$$(\Sigma, \mathcal{E}, \mathcal{F}, z_1, z_2),$$

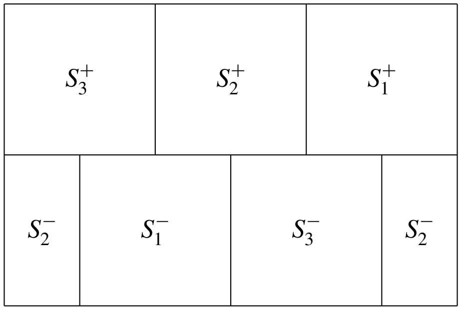


Figure 17. A projection of $S^1 \times S^3 \setminus N(K)$ into $[0, 6] \times [0, 3]$ using a factor of the angle of S^1 and the restriction of a Morse function on S^3 to the knot complement.

where $(\Sigma, \mathcal{E}, \mathcal{F})$ is a Heegaard diagram for S^3 and z_1 and z_2 are distinct points on Σ in the complement of $\mathcal{E}$ and $\mathcal{F}$, such that, in the associated handle decomposition of S^3, K is the union of two arcs connecting the index 0 and 3 critical points, avoiding the cocores of the 1-handles and the cores of the 2-handles, intersecting Σ at z_1 and z_2.

Note that if K is given in bridge position with B bridges, stabilizing the genus 0 Heegaard splitting $B-1$ times gives a genus $B-1$ doubly pointed diagram describing K.

This description can then be translated into a Morse function $f : S^3 \to [0, 3]$ such that the knot K is obtained as the union of the gradient flow lines of f joining the unique index 3 critical point with the unique index 0 critical point and passing through the points z_1 and z_2. After a small perturbation we may assume that $f|_{\partial N(K)}$ is a standard Morse function on T^2; the only feature we really care about is that $f^{-1}(3/2)$ intersect $N(K)$ as meridional disks and thus splits $\partial N(K)$ into two annuli.

Identify S^1 with $[0, 6]/0 \sim 6$, draw a grid on $[0, 6] \times [0, 3]$ as in Figure 17 and label the squares $S_i^{\pm}$, $i = 1, 2, 3$ with the sign chosen depending on the position of the square relative to the horizontal line $[0, 1] \times \{3/2\}$. Notice that the left and right ends of the figure should be identified since $[0, 6]$ is actually $[0, 6]/0 \sim 6 = S^1$. Consider the projection $\pi : S^1 \times S^3 \setminus N(K) \to S^1 \times [0, 3]$ given by the identity in the first component, and the restriction of the Morse function f to the knot complement in the second component. Over each vertical line segment in Figure 17 is a 3-dimensional handlebody with B 1-handles, realized as the intersection of the genus $(B-1)$ handlebody $U^{\pm}$ with the knot complement $S^3 \setminus N(K)$. Therefore, over each square lies a 4-dimensional space diffeomorphic to $\natural^B S^1 \times B^3$. Similarly, over each interior vertex lies the punctured surface $\Sigma' = \Sigma \setminus (D(z_1) \sqcup D(z_2))$, where z_1, z_2 are the points in Σ that describe the knot $K \subset S^3$, and $D(z_j)$ $(j = 1, 2)$, is a

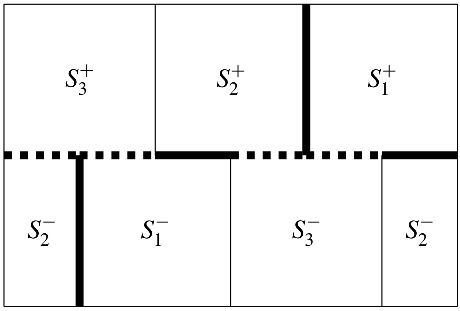


Figure 18. The pieces involved in the pairwise intersection $X_1 \cap X_2$.

disk neighborhood of z_j in Σ. We thus see that over each interior and horizontal edge of a square lies the genus $2B-1$, 3-dimensional handlebody $I \times \Sigma'$.

We will obtain a trisection of $X = S^1 \times S^3 \setminus N(K)$ by connecting the preimage of S_i^+ to the preimage of S_i^- using 4-dimensional 1-handles realized as tubular neighborhoods of appropriately chosen arcs. Let $k = 1, \ldots, 6$ and $j = 1, 2$, and to simplify notation identify $\partial D(z_j)$ with the unit circle in $\mathbb{C}$, and denote by ξ_j^k the k-th power of a third root of unity $\xi \in S^1$, regarded as a point in $\partial D(z_j)$. Consider the arcs a_{kj} obtained by taking the product of the preimage of $[k-1, k]$ in $S^1 = [0, 6]/(0 \sim 6)$ with the point ξ_j^k in $S^3 \setminus N(K)$. The i-th piece of the trisection of X into $X_1 \cup X_2 \cup X_3$ will be obtained by connecting $\pi^{-1}(S_i^+)$ to $\pi^{-1}(S_i^-)$ using the 1-handles whose cores project into the grid as a horizontal edge disjoint from the squares S_i^+ and S_i^-, and removing from it the other 1-handles. Specifically, if we denote the tubular neighborhood of a_{kj} $(k = 1, \ldots, 6,\ j = 1, 2)$ in X by ν_{kj}, then

$$X_i = \left(\pi^{-1}(S_i^+ \sqcup S_i^-) \Big\backslash \bigcup_{\substack{l \neq i, i+3 \\ j=1,2}} \nu_{lj} \right) \cup \left(\bigcup_{j=1,2} \nu_{ij} \sqcup \nu_{i+3,j} \right).$$

Since for $k \not\equiv i \mod 3$ the cores of the tubes ν_{kj} lie in the boundary of the squares $S_i^{\pm}$, removing them from their preimages does not change the diffeomorphism type of this space. Thus, X_i is a connected space and since $\pi^{-1}(S_i^{\pm}) \cong \natural^B S^1 \times B^3$, we see that X_i is diffeomorphic to $\natural^{2B+3} S^1 \times B^3$.

Next we analyze the pairwise intersections of the pieces, and since the calculations are analogous for any pair $(i, i+1)$, we present the details for $X_1 \cap X_2$ and leave out those concerning the other cases. There are three different types of spaces involved in the double intersection: the preimages of the vertical segments of the intersections $S_1 \cap S_2$, the preimages of the horizontal intersections, and 3-dimensional tubular neighborhoods of some of the arcs a_{kj}. These sets are highlighted in Figure 18, with the dotted line representing the presence of tubular neighborhood of two arcs.

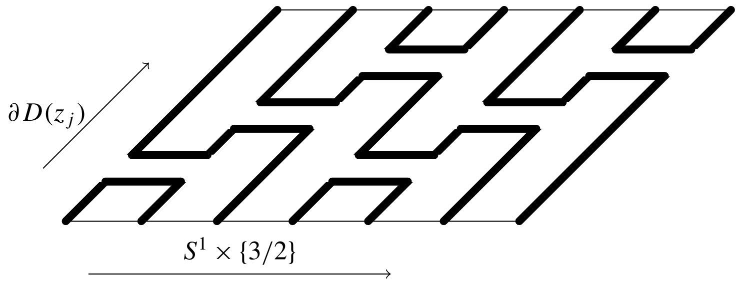


Figure 19. Two of the four components of $X_1 \cap X_2 \cap X_3 \cap \partial X$. This parallelogram represents a torus as follows: the horizontal component represents the S^1 direction in the middle of Figure 17 and the slanted direction represents the direction of $\partial D(z_j)$ which is "internal" to Σ' and therefore not represented in Figure 17.

We then see that the space $X_1 \cap X_2$ is diffeomorphic to the disjoint union of two 3-dimensional handlebodies of genus B and two 3-dimensional handlebodies of genus $2B-1$ (two copies of $I \times \Sigma'$), connected to one another using eight 3-dimensional 1-handles. Therefore, $X_1 \cap X_2$ is diffeomorphic to $\natural^{6B+3} S^1 \times D^2$.

The triple intersection $F = X_1 \cap X_2 \cap X_3$ is the union of six copies of the punctured surface $\Sigma' = \Sigma \setminus (N(z_1) \sqcup N(z_2))$ realized as the preimages of the six interior vertices in Figure 17, connected to one another using band neighborhoods of the arcs a_{kj} in $S^1 \times \Sigma'$. A simple computation shows that a surface so decomposed has Euler characteristic equal to $-12B$ and so, to establish the diffeomorphism type of this central surface F, it is enough to calculate the number of boundary components. With that in mind, notice that ∂F is precisely the space $X_1 \cap X_2 \cap X_3 \cap \partial X$, and that this space is the result of joining the copies of $\partial D(z_j)$ lying above the six internal vertices to one another using band neighborhoods of the six arcs a_{jk} for $j = 1, 2$. For each $j = 1, 2$ this results in two circles, for a total of four boundary components. A schematic picture that describes these components can be found in Figure 19. A simple Euler characteristic argument then shows that the surface $X_1 \cap X_2 \cap X_3$ has genus $6B - 1$.

Next, to understand $X_1 \cap X_2 \cap \partial X$ intersect the highlighted pieces in Figure 18 with $\partial N(K)$. Above the vertical edges lies a cylinder, above each horizontal edge two disks realized as $I \times (\partial D(z_1) \setminus N(\xi_1^3))$ and $I \times (\partial D(z_2) \setminus N(\xi_2^3))$, and above each dotted line two band neighborhoods of the arcs (one for each of z_1 and z_2). Thus, $X_1 \cap X_2 \cap \partial X$ is diffeomorphic to the disjoint union of six cylinders connected to one another using eight bands. A surface with this decomposition has Euler characteristic equal to -4, and since its boundary is the same as the boundary of the central surface F we conclude that $X_1 \cap X_2 \cap \partial X$ is a surface of genus 1 and 4 boundary components.

The last intersection to consider is $X_1 \cap \partial X$. This space consists of two solid tori, one above each one of $S_1^{\pm}$, and two 3-dimensional 1-handles that lie above $[0, 1] \times \{3/2\}$ and $[3, 4] \times \{3/2\}$. This shows that $X_1 \cap \partial X$ is a genus 5 handlebody. Moreover, notice that each solid torus is a relative compression body from one of the cylinders in $X_1 \cap X_2 \cap \partial X$ to a cylinder in $X_1 \cap X_3 \cap \partial X$, and that each solid torus contains one of the disks in each one of $X_1 \cap X_2 \cap \partial X$ and $X_1 \cap X_3 \cap \partial X$. In addition, the 3-dimensional 1-handles are relative compression bodies between the band neighborhoods of the arcs, and so $X_1 \cap \partial X$ is diffeomorphic to the product of an interval and the surface $X_1 \cap X_2 \cap \partial X$.

Finally, to obtain a trisection diagram for $S^1 \times S^3 \setminus N(K)$ all that is left to do is understand the collection of disks in the pairwise intersections $X_i \cap X_{i+1}$ that are bounded by curves that lie entirely in the triple intersection $F = X_1 \cap X_2 \cap X_3$. One more time we focus only on the intersection $X_1 \cap X_2$. In this case we have:

- The collection $\mathcal{F}$ of $B-1$ curves that bound disks D_i^+ at $\{4\} \times U^+$.
- The collection $\mathcal{E}$ of $B-1$ curves that bound disks D_i^- at $\{1\} \times U^-$.
- A collection of $2B$ curves stemming from a handle decomposition of $[2, 3] \times \Sigma'$ relative to the union of $\{2, 3\} \times \Sigma'$ with band neighborhoods of the arcs $[2, 3] \times \{\xi_j^2\}$, $j = 1, 2$. The curves are realized as the union of arcs in $\{2\} \times \Sigma'$ with arcs in $\{3\} \times \Sigma'$ going through the bands; $2(B-1)$ of the arcs arise from some 1-handles in Σ' that give rise to genus, one other from a 1-handle in Σ' that gives rise to the boundary components, and one other that connects the two bands.
- A collection analogous to the one above but related to $[5, 6] \times \Sigma'$.

Thus, the trisection diagram consists of a surface of genus $6B-1$ with 4 boundary components, realized as the union of six copies of Σ' joined to one another using twelve bands, and curves coming either from the Heegaard splitting of S^3 that corresponds to the doubly pointed diagram of K, or from the handlebody structure of $I \times \Sigma'$ and distributed along the pieces of Σ' as shown in Figure 20, top.

References

[Baer 1928] R. Baer, "Isotopie von Kurven auf orientierbaren, geschlossenen Flächen und ihr Zusammenhang mit der topologischen Deformation der Flächen", *J. Reine Angew. Math.* **159** (1928), 101–116. MR Zbl

[Casson and Gordon 1987] A. J. Casson and C. M. Gordon, "Reducing Heegaard splittings", *Topology Appl.* **27**:3 (1987), 275–283. MR Zbl

[Castro 2016] N. A. Castro, *Relative trisections of smooth 4-manifolds with boundary*, Ph.D. thesis, University of Georgia, 2016, Available at https://tinyurl.com/castrothes.

[Davis and Kirk 2001] J. F. Davis and P. Kirk, *Lecture notes in algebraic topology*, Graduate Studies in Mathematics **35**, American Mathematical Society, Providence, RI, 2001. MR Zbl

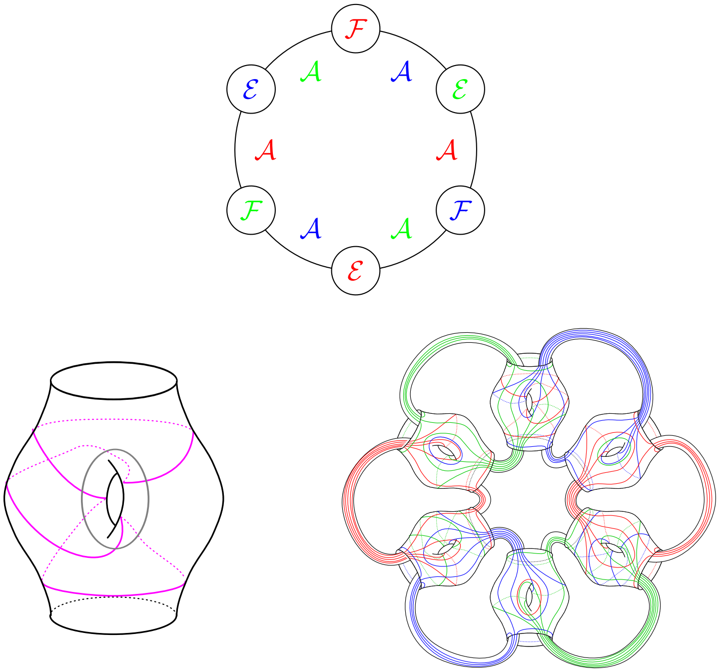


Figure 20. Relative trisection of $S^1 \times S^3 \setminus N(K)$. Top: Recipe for drawing a trisection diagram for $S^1 \times S^3 \setminus N(K)$. The smaller circles with a letter inside represent copies of Σ', the subarcs of the larger circle represent the bands connecting the different copies of Σ'. Here $\mathcal{A}$ denotes the curves obtained as union of arcs arising from $I \times \Sigma'$, whereas $\mathcal{F}, \mathcal{E}$ denote the curves in the doubly pointed diagram for the knot K. Additionally, each color represents one collection of α, β, γ. Bottom left: A Heegaard diagram for $S^3 \setminus N(T_2, 3)$. Denote by $\mathcal{F}$ the pink curve and by $\mathcal{E}$ the gray curve. Bottom right: A trisection diagram for $S^1 \times S^3 \setminus N(T_{2,3})$.

[Endo and Gurtas 2010] H. Endo and Y. Z. Gurtas, "Lantern relations and rational blowdowns", *Proc. Amer. Math. Soc.* **138**:3 (2010), 1131–1142. MR Zbl

[Epstein 1966] D. B. A. Epstein, "Curves on 2-manifolds and isotopies", *Acta Math.* **115** (1966), 83–107. MR Zbl

[Fintushel and Stern 1997] R. Fintushel and R. J. Stern, "Rational blowdowns of smooth 4-manifolds", *J. Differential Geom.* **46**:2 (1997), 181–235. MR Zbl

[Gay and Kirby 2016] D. Gay and R. Kirby, "Trisecting 4-manifolds", *Geom. Topol.* **20**:6 (2016), 3097–3132. MR Zbl

[Johannson 1995] K. Johannson, *Topology and combinatorics of 3-manifolds*, Lecture Notes in Mathematics **1599**, Springer, 1995. MR Zbl

[Kas 1980] A. Kas, "On the handlebody decomposition associated to a Lefschetz fibration", *Pacific J. Math.* **89**:1 (1980), 89–104. MR Zbl

[Laudenbach and Poénaru 1972] F. Laudenbach and V. Poénaru, "A note on 4-dimensional handlebodies", *Bull. Soc. Math. France* **100** (1972), 337–344. MR Zbl

[Manolescu 2016] C. Manolescu, "An introduction to knot Floer homology", pp. 99–135 in *Physics and mathematics of link homology*, edited by M. Khovanov and J. Walcher, Contemp. Math. **680**, American Mathematical Society, Providence, RI, 2016. MR Zbl

[Rasmussen 2003] J. Rasmussen, *Floer homology and knot complements*, Ph.D. thesis, Harvard University, 2003, Available at https://tinyurl.com/rasmusthesis.

Received January 18, 2017. Revised October 16, 2017.

NICKOLAS A. CASTRO
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, DAVIS
ONE SHIELDS AVE
DAVIS, CA 95616
UNITED STATES
ncastro@math.ucdavis.edu

DAVID T. GAY
EUCLID LAB
160 MILLEDGE TERRACE
ATHENS, GA 30606
UNITED STATES
d.gay@euclidlab.org

JUANITA PINZÓN-CAICEDO
DEPARTMENT OF MATHEMATICS
NC STATE UNIVERSITY
2108 SAS HALL
RALEIGH, NC 27695
UNITED STATES
jpinzon@ncsu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 294, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.294.307

LINKAGE OF MODULES WITH RESPECT TO A SEMIDUALIZING MODULE

MOHAMMAD T. DIBAEI AND ARASH SADEGHI

The notion of linkage with respect to a semidualizing module is introduced. This notion enables us to study the theory of linkage for modules in the Bass class with respect to a semidualizing module. It is shown that over a Cohen–Macaulay local ring with canonical module, every Cohen–Macaulay module of finite Gorenstein injective dimension is linked with respect to the canonical module. For a linked module M with respect to a semidualizing module, the connection between the Serre condition (S_n) on M and the vanishing of certain local cohomology modules of its linked module is discussed.

1. Introduction

The theory of linkage of ideals in commutative algebra was introduced by Peskine and Szpiro [1974]. Recall that two ideals I and J in a Cohen–Macaulay local ring R are said to be linked if there is a regular sequence α in their intersection such that $I = (\alpha : J)$ and $J = (\alpha : I)$. One of the main results in the theory of linkage, due to C. Peskine and L. Szpiro, indicates that the Cohen–Macaulay-ness property is preserved under linkage over Gorenstein local rings. They also give a counterexample to show that the above result is no longer true if the base ring is Cohen–Macaulay but non-Gorenstein. Attempts to generalize this theorem have led to several developments in linkage theory, especially by C. Huneke and B. Ulrich [Huneke 1982; Huneke and Ulrich 1987]. Schenzel [1982] used the theory of dualizing complexes to extend the basic properties of linkage to the linkage by Gorenstein ideals.

The classical linkage theory has been extended to modules by Martin [2000], Yoshino and Isogawa [2000], Martsinkovsky and Strooker [2004], and Nagel [2005], in different ways. Based on these generalizations, several works have been done on studying the linkage theory in the context of modules; see for example [Dibaei

Dibaei was supported in part by a grant from the IPM (No. 95130110). Sadeghi was supported by a grant from the IPM.
MSC2010: primary 13C40; secondary 13D05.
Keywords: linkage of modules, Auslander classes, Bass classes, semidualizing modules, G_C-dimensions.

et al. 2011; Dibaei and Sadeghi 2013; 2015; Iima and Takahashi 2016; Sadeghi 2017; Celikbas et al. 2017]. In this paper, we introduce the notion of linkage with respect to a semidualizing module. This is a new notion of linkage for modules and includes the concept of linkage due to Martsinkovsky and Strooker.

To be more precise, let M and N be R-modules and let α be an ideal of R which is contained in $\operatorname{Ann}_R(M) \cap \operatorname{Ann}_R(N)$. Assume that K is a semidualizing R/α-module. We say that M is linked to N with respect to K if $M \cong \lambda_{R/\alpha}(K, N)$ and $N \cong \lambda_{R/\alpha}(K, M)$, where

$$\lambda_{R/\alpha}(K, -) := \Omega_K \mathsf{Tr}_K \mathsf{Hom}_{R/\alpha}(K, -),$$

where Ω_K, Tr_K are the syzygy and transpose operators, respectively, with respect to K. This notion enables us to study the theory of linkage for modules in the Bass class with respect to a semidualizing module. In the first main result of this paper, over a Cohen–Macaulay local ring with canonical module, it is proved that every Cohen–Macaulay module of finite Gorenstein injective dimension is linked with respect to the canonical module (see Theorem 3.12). More precisely:

Theorem A. *Let R be a Cohen–Macaulay local ring of dimension d with canonical module ω_R. Assume that $\mathfrak{a}$ is a Cohen–Macaulay quasi-Gorenstein ideal of grade n and that M is a Cohen–Macaulay R-module of grade n and of finite Gorenstein injective dimension (equivalently $M \in \mathscr{B}_{\omega_R}$). If $\mathfrak{a} \subseteq \operatorname{Ann}_R(M)$ and M is $\omega_{R/\mathfrak{a}}$-stable, then the following statements hold true*:

(i) *M is linked by ideal $\mathfrak{a}$ with respect to $\omega_{R/\mathfrak{a}}$.*

(ii) *$\lambda_{R/\mathfrak{a}}(\omega_{R/\mathfrak{a}}, M)$ has finite Gorenstein injective dimension.*

(iii) *$\lambda_{R/\mathfrak{a}}(\omega_{R/\mathfrak{a}}, M)$ is Cohen–Macaulay of grade n.*

Martsinkovsky and Strooker [2004, Corollary 2] proved that horizontal linkage preserves the maximal Cohen–Macaulay-ness property over Gorenstein rings, while it may not preserve this property over non-Gorenstein rings. Theorem A shows that, over a Cohen–Macaulay local ring with the canonical module, horizontal linkage with respect to canonical module preserves maximal Cohen–Macaulay-ness for every module of finite Gorenstein injective dimension. Note that over a Gorenstein ring, every module has finite Gorenstein injective dimension. Therefore, Theorem A can be viewed as a generalization of [Martsinkovsky and Strooker 2004, Corollary 2].

Recall that an R-module M is called G-perfect if $\mathsf{grade}_R(M) = \mathsf{G}\text{-}\mathsf{dim}_R(M)$. If R is Cohen–Macaulay then M is G-perfect if and only if M is Cohen–Macaulay and $\mathsf{G}\text{-}\mathsf{dim}_R(M) < \infty$. Let us denote the category of G-perfect R-modules by $\mathscr{X}$, and the category of Cohen–Macaulay R-modules of finite Gorenstein injective dimension is by $\mathscr{Y}$. Theorem A enables us to obtain the following adjoint equivalence (see Theorems 3.13 and 3.14).

Theorem B. *Let R be a Cohen–Macaulay local ring with canonical module ω_R and let $\mathfrak{a}$ be a Cohen–Macaulay quasi-Gorenstein ideal, $\bar{R} = R/\mathfrak{a}$. There is an adjoint equivalence*

$$\left\{ M \in \mathscr{X} \;\middle|\; \begin{matrix} M \text{ is linked by} \\ \text{the ideal } \mathfrak{a} \end{matrix} \right\} \underset{\mathsf{Hom}_{\bar{R}}(\omega_{\bar{R}},-)}{\overset{-\otimes_{\bar{R}}\omega_{\bar{R}}}{\rightleftarrows}} \left\{ N \in \mathscr{Y} \;\middle|\; \begin{matrix} N \text{ is linked by the ideal} \\ \mathfrak{a} \text{ with respect to } \omega_{\bar{R}} \end{matrix} \right\}.$$

Let R be a Cohen–Macaulay local ring with canonical module ω_R. For a linked R-module M, with respect to the canonical module, we study the connection between the Serre condition on M with vanishing of certain local cohomology modules of its linked module. We also establish a duality on local cohomology modules of a linked module which is a generalization of [Schenzel 1982, Theorem 4.1; Martsinkovsky and Strooker 2004, Theorem 10] (see Corollaries 4.9 and 4.12).

Theorem C. *Let $(R, \mathfrak{m}, k)$ be a Cohen–Macaulay local ring of dimension $d > 1$ with canonical module ω_R. Assume that an R-module M is horizontally linked to an R-module N with respect to ω_R and that M has finite Gorenstein injective dimension. Then the following statements hold true*:

(i) *M satisfies (S_n) if and only if $\mathrm{H}^i_{\mathfrak{m}}(N) = 0$ for $d - n < i < d$.*

(ii) *If M is generalized Cohen–Macaulay then*

$$\mathrm{H}^i_{\mathfrak{m}}(\mathsf{Hom}_R(\omega_R, M)) \cong \mathsf{Hom}_R(\mathrm{H}^{d-i}_{\mathfrak{m}}(N), \boldsymbol{E}_R(k)) \quad \textit{for } 0 < i < d.$$

In particular, N is generalized Cohen–Macaulay.

The organization of the paper is as follows. In Section 2, we collect preliminary notions, definitions and some known results which will be used in this paper. In Section 3, the precise definition of linkage with respect to a semidualizing is given. We obtain some necessary conditions for an R-module to be linked with respect to a semidualizing (see Theorem 3.7). As a consequence, we prove Theorems A and B in this section. In Section 4, for a linked R-module M, with respect to a semidualizing, the relation between the Serre condition $\widetilde{S}_n$ on M with vanishing of certain relative cohomology modules of its linked module is studied. As a consequence, we prove Theorem C.

2. Preliminaries

Throughout the paper, R is a commutative Noetherian semiperfect ring and all R-modules are finitely generated. Note that a commutative ring R is semiperfect if and only if it is a finite direct product of commutative local rings [Lam 1991, Theorem 23.11]. Whenever, R is assumed to be local, its unique maximal ideal is denoted by $\mathfrak{m}$. The canonical module of R is denoted by ω_R.

Let M be an R-module. For a finite projective presentation $P_1 \xrightarrow{f} P_0 \to M \to 0$ of M, its transpose $\mathsf{Tr}M$ is defined as Coker f^*, where $(-)^* := \mathsf{Hom}_R(-, R)$, which satisfies the exact sequence

$$0 \to M^* \to P_0^* \xrightarrow{f^*} P_1^* \to \mathsf{Tr}M \to 0.$$

Moreover, $\mathsf{Tr}M$ is unique up to projective equivalence. Thus all minimal projective presentations of M represent isomorphic transposes of M. The syzygy module ΩM of M is the kernel of an epimorphism $P \xrightarrow{\alpha} M$, where P is a projective R-module which is unique up to projective equivalence. Thus ΩM is uniquely determined, up to isomorphism, by a projective cover of M.

Martsinkovsky and Strooker [2004] generalized the notion of linkage for modules over noncommutative semiperfect Noetherian rings (i.e., finitely generated modules over such rings have projective covers). In Proposition 1 of that paper, they introduced the operator $\lambda := \Omega\mathsf{Tr}$ and showed that ideals $\mathfrak{a}$ and $\mathfrak{b}$ are linked by zero ideal if and only if $R/\mathfrak{a} \cong \lambda(R/\mathfrak{b})$ and $R/\mathfrak{b} \cong \lambda(R/\mathfrak{a})$.

Definition 2.1 [Martsinkovsky and Strooker 2004, Definition 3]. Two R-modules M and N are said to be *horizontally linked* if $M \cong \lambda N$ and $N \cong \lambda M$. Equivalently, M is horizontally linked (to λM) if and only if $M \cong \lambda^2 M$.

A *stable* module is a module with no nonzero projective direct summands. An R-module M is called a *syzygy module* if it is embedded in a projective R-module. Let i be a positive integer, an R-module M is said to be an i-th syzygy if there exists an exact sequence

$$0 \to M \to P_{i-1} \to \cdots \to P_0,$$

where the $P_0, \ldots, P_{i-1}$ are projective. By convention, every module is a 0-th syzygy.

Here is a characterization of horizontally linked modules.

Theorem 2.2 [Martsinkovsky and Strooker 2004, Theorem 2 and Proposition 3]. *An R-module M is horizontally linked if and only if it is stable and* $\mathsf{Ext}^1_R(\mathsf{Tr}M, R) = 0$, *equivalently M is stable and is a syzygy module.*

Semidualizing modules were initially studied in [Foxby 1972; Golod 1984].

Definition 2.3. An R-module C is called a *semidualizing* module if the homothety morphism $R \to \mathsf{Hom}_R(C, C)$ is an isomorphism and $\mathsf{Ext}^i_R(C, C) = 0$ for all $i > 0$.

It is clear that R itself is a semidualizing R-module. Over a Cohen–Macaulay local ring R, a canonical module ω_R of R, if it exists, is a semidualizing module with finite injective dimension.

Conventions 2.4. Throughout let C denote a semidualizing R-module. We set $(-)^{\triangledown} = \mathsf{Hom}_R(-, C)$ and $(-)^{\curlyvee} = \mathsf{Hom}_R(C, -)$. The notation $(-)^*$ stands for the R-dual functor $\mathsf{Hom}_R(-, R)$. The canonical module of a Cohen–Macaulay local ring, if it exists, is denoted as ω_R; then we set $(-)^{\dagger} = \mathsf{Hom}_R(-, \omega_R)$.

Let $P_1 \xrightarrow{f} P_0 \to M \to 0$ be a projective presentation of an R-module M. The transpose of M with respect to C, denoted by $\mathrm{Tr}_C M$, is defined to be Coker $f^{\triangledown}$, which satisfies the exact sequence

$$0 \to M^{\triangledown} \to P_0^{\triangledown} \xrightarrow{f^{\triangledown}} P_1^{\triangledown} \to \mathrm{Tr}_C M \to 0. \tag{2.4.1}$$

By [Foxby 1972, Proposition 3.1], there exists the exact sequence

$$0 \to \mathrm{Ext}_R^1(\mathrm{Tr}_C M, C) \to M \to M^{\triangledown\triangledown} \to \mathrm{Ext}_R^2(\mathrm{Tr}_C M, C) \to 0. \tag{2.4.2}$$

The Gorenstein dimension has been extended to G_C-dimension in [Foxby 1972; Golod 1984].

Definition 2.5. An R-module M is said to have G_C*-dimension zero* if M is C-reflexive, i.e., the canonical map $M \to M^{\triangledown\triangledown}$ is bijective, and $\mathrm{Ext}_R^i(M, C) = 0 = \mathrm{Ext}_R^i(M^{\triangledown}, C)$ for all $i > 0$.

A G_C-resolution of an R-module M is a right acyclic complex of G_C-dimension zero modules whose 0-th homology is M. The module M is said to have finite G_C-dimension, denoted by $\mathrm{G}_C\text{-}\mathrm{dim}_R(M)$, if it has a G_C-resolution of finite length.

Note that, over a local ring R, a semidualizing R-module C is a canonical module if and only if $\mathrm{G}_C\text{-}\mathrm{dim}_R(M) < \infty$ for all finitely generated R-modules M; see [Gerko 2001, Proposition 1.3].

In the following, we summarize some basic facts about G_C-dimension; see [Auslander and Bridger 1969; Golod 1984] for more details.

Theorem 2.6. *For an R-module M, the following statements hold true*:

(i) $\mathrm{G}_C\text{-}\mathrm{dim}_R(M) = 0$ *if and only if* $\mathrm{Ext}_R^i(M, C) = 0 = \mathrm{Ext}_R^i(\mathrm{Tr}_C M, C)$ *for all* $i > 0$.

(ii) $\mathrm{G}_C\text{-}\mathrm{dim}_R(M) = 0$ *if and only if* $\mathrm{G}_C\text{-}\mathrm{dim}_R(\mathrm{Tr}_C M) = 0$.

(iii) *If* $\mathrm{G}_C\text{-}\mathrm{dim}_R(M) < \infty$ *then* $\mathrm{G}_C\text{-}\mathrm{dim}_R(M) = \sup\{i \mid \mathrm{Ext}_R^i(M, C) \neq 0,\ i \geq 0\}$.

(iv) *If R is local and* $\mathrm{G}_C\text{-}\mathrm{dim}_R(M) < \infty$, *then* $\mathrm{G}_C\text{-}\mathrm{dim}_R(M) = \mathrm{depth}\, R - \mathrm{depth}_R(M)$.

The Gorenstein injective dimension was introduced by Enochs and Jenda [1995].

Definition 2.7 [Enochs and Jenda 1995; Christensen 2000, Definition 6.2.2]. An R-module M is said to be *Gorenstein injective* if there is an exact sequence

$$I_\bullet = \cdots \to I_1 \xrightarrow{\partial_1} I_0 \xrightarrow{\partial_0} I_{-1} \to \cdots$$

of injective R-modules such that $M \cong \mathrm{Ker}(\partial_0)$ and $\mathrm{Hom}_R(E, I_\bullet)$ is exact for any injective R-module E. The Gorenstein injective dimension of M, denoted by $\mathrm{Gid}(M)$, is defined as the infimum of n for which there exists an exact sequence as $I_\bullet$ with $M \cong \mathrm{Ker}(I_0 \to I_{-1})$ and $I_i = 0$ for all $i < -n$. The Gorenstein injective dimension is a refinement of the classical injective dimension, $\mathrm{Gid}(M) \leq \mathrm{id}(M)$, with equality if $\mathrm{id}(M) < \infty$; see [Christensen 2000, Definition 6.2.6]. It follows that every module over a Gorenstein ring has finite Gorenstein injective dimension.

Definition 2.8. The *Auslander class with respect to* C, denoted by $\mathscr{A}_C$, consists of all R-modules M satisfying the following conditions:

(i) The natural map $\mu : M \to \mathsf{Hom}_R(C, M \otimes_R C)$ is an isomorphism.

(ii) $\mathsf{Tor}_i^R(M, C) = 0 = \mathsf{Ext}_R^i(C, M \otimes_R C)$ for all $i > 0$.

Dually, the *Bass class* with respect to C, denoted by $\mathscr{B}_C$, consists of all R-modules M satisfying the following conditions:

(i) The natural evaluation map $\mu : C \otimes_R \mathsf{Hom}_R(C, M) \to M$ is an isomorphism.

(ii) $\mathsf{Tor}_i^R(\mathsf{Hom}_R(C, M), C) = 0 = \mathsf{Ext}_R^i(C, M)$ for all $i > 0$.

In the following we collect some basic properties and examples of modules in the Auslander class, respectively in the Bass class, with respect to C, which will be used in the rest of this paper.

Fact 2.9. The following statements hold:

(i) If any two R-modules in a short exact sequence are in $\mathscr{A}_C$, respectively $\mathscr{B}_C$, then so is the third one [Foxby 1972, Lemma 1.3]. Hence, every module of finite projective dimension is in the Auslander class $\mathscr{A}_C$. Also the class $\mathscr{B}_C$ contains all modules of finite injective dimension.

(ii) Over a Cohen–Macaulay local ring R with canonical module ω_R, we have $M \in \mathscr{A}_{\omega_R}$ if and only if $\mathsf{G}\text{-}\mathsf{dim}_R(M) < \infty$ [Foxby 1975, Theorem 1]. Similarly, $M \in \mathscr{B}_{\omega_R}$ if and only if $\mathsf{Gid}_R(M) < \infty$ [Christensen et al. 2006, Theorem 4.4].

(iii) The $\mathscr{P}_C$-projective dimension of M, denoted by $\mathscr{P}_C\text{-}\mathsf{pd}_R(M)$, is less than or equal to n if and only if there is an exact sequence

$$0 \to P_n \otimes_R C \to \cdots \to P_0 \otimes_R C \to M \to 0$$

such that each P_i is a projective R-module [Takahashi and White 2010, Corollary 2.10]. Note that if M has a finite $\mathscr{P}_C$-projective dimension, then $M \in \mathscr{B}_C$ by Corollary 2.9 of the same paper.

(iv) $M \in \mathscr{A}_C$ if and only if $M \otimes_R C \in \mathscr{B}_C$. Similarly, $M \in \mathscr{B}_C$ if and only if $M^{\curlyvee} \in \mathscr{A}_C$ [Takahashi and White 2010, Theorem 2.8].

Definition 2.10. Let M and N be R-modules. Denote by $\beta(M, N)$ the set of R-homomorphisms of M to N which pass through projective modules. That is, an R-homomorphism $f : M \to N$ lies in $\beta(M, N)$ if and only if it is factored as $M \to P \to N$, where P is projective. We denote the stable homomorphisms from M to N as the quotient module

$$\underline{\mathsf{Hom}}_R(M, N) = \mathsf{Hom}_R(M, N)/\beta(M, N).$$

By [Yoshino 1990, Lemma 3.9], there is a natural isomorphism

$$\underline{\mathsf{Hom}}_R(M, N) \cong \mathsf{Tor}_1^R(\mathsf{Tr}M, N). \tag{2.10.1}$$

The class of C-projective modules is defined as

$$\mathscr{P}_C = \{P \otimes_R C \mid P \text{ is projective}\}.$$

Two R-modules M and N are said to be *stably equivalent* with respect to C, denoted by $M \underset{C}{\approx} N$, if $C_1 \oplus M \cong C_2 \oplus N$ for some C-projective modules C_1 and C_2. We write $M \approx N$ when M and N are stably equivalent with respect to R. An R-module M is called *C-stable* if M does not have a direct summand isomorphic to a C-projective module. An R-module M is called a *C-syzygy module* if it is embedded in a C-projective R-module.

Remark 2.11. Let M be an R-module.

(i) Let $P_1 \xrightarrow{f} P_0 \to M \to 0$ be the minimal projective presentation of M. Then $\mathsf{Tr}M \otimes_R C \cong \mathsf{Tr}_C M$; see [Dibaei and Sadeghi 2015, Remark 2.1(i)].

(ii) Note that, by [Martsinkovsky 2010, Proposition 3(a)], $(P_1)^* \to \mathsf{Tr}M \to 0$ is minimal. Therefore, by (i), we get the exact sequence

$$0 \to \Omega_C \mathsf{Tr}_C M \to (P_1)^* \otimes_R C \to \mathsf{Tr}_C M \to 0,$$

where $\Omega_C \mathsf{Tr}_C M := \operatorname{Im} f^{\triangledown}$.

(iii) It follows, by (2.10.1), that if $\underline{\mathsf{Hom}}_R(M, C) = 0$, then $\Omega_C \mathsf{Tr}_C M \cong \lambda M \otimes_R C$.

Definition 2.12 [Maşek 2000]. An R-module M is said to satisfy the property $\widetilde{S}_k$ if $\mathsf{depth}_{R_\mathfrak{p}}(M_\mathfrak{p}) \geq \min\{k, \mathsf{depth}\, R_\mathfrak{p}\}$ for all $\mathfrak{p} \in \operatorname{Spec} R$.

Note that, for a horizontally linked module M over a Cohen–Macaulay local ring R, the properties $\widetilde{S}_k$ and (S_k) are identical.

3. Horizontal linkage with respect to a semidualizing

In this section C stands for a semidualizing R-module and M is an R-module. Set $(M)^{\curlyvee} := \mathsf{Hom}_R(C, M)$ as in Conventions 2.4. In order to develop the notion of linkage with respect to C, we give the following definition.

Definition 3.1. The *linkage of M with respect to C* is defined as the module $\lambda_R(C, M) := \Omega_C \mathsf{Tr}_C(M^{\curlyvee})$. The module M is said to be horizontally linked to an R-module N with respect to C if $\lambda_R(C, M) \cong N$ and $\lambda_R(C, N) \cong M$. Equivalently, M is horizontally linked (to $\lambda_R(C, M)$) with respect to C if and only if $M \cong \lambda_R^2(C, M) (= \lambda_R(C, \lambda_R(C, M)))$. In this situation M is called a *horizontally linked module with respect to C*.

Assume that $P_1 \xrightarrow{f} P_0 \to M^{\curlyvee} \to 0$ is the minimal projective presentation of $M^{\curlyvee}$. By Remark 2.11, $\lambda_R(C, M) = \mathrm{Im}(f^{\triangledown})$ and we obtain the exact sequence

$$0 \to \lambda_R(C, M) \to (P_1)^* \otimes_R C \to \mathsf{Tr}_C(M^{\curlyvee}) \to 0. \tag{3.1.1}$$

Therefore $\lambda_R(C, M)$ is unique, up to isomorphism. Having defined the horizontal linkage with respect to a semidualizing module C, the general linkage for modules is defined as follows.

Definition 3.2. Let $\mathfrak{a}$ be an ideal of R and let K be a semidualizing $R/\mathfrak{a}$-module. An R-module M is said to be *linked* to an R-module N by the ideal $\mathfrak{a}$, with respect to K, if $\mathfrak{a} \subseteq \mathrm{Ann}_R(M) \cap \mathrm{Ann}_R(N)$ and M and N are horizontally linked with respect to K as $R/\mathfrak{a}$-modules. In this situation we write $M \underset{\mathfrak{a}}{\overset{K}{\sim}} N$.

Lemma 3.3. *Assume that an R-module M satisfies the following conditions*:

(i) *M is a C-stable and C-syzygy.*

(ii) $\underline{\mathsf{Hom}}_R(M^{\curlyvee}, C) = 0 = \underline{\mathsf{Hom}}_R(\lambda(M^{\curlyvee}), C)$.

(iii) *$M \cong C \otimes_R M^{\curlyvee}$ and $\lambda(M^{\curlyvee}) \cong \mathsf{Hom}_R(C, C \otimes_R \lambda(M^{\curlyvee}))$.*

Then M is a horizontally linked R-module with respect to C.

Proof. As M is C-stable, by (iii), $M^{\curlyvee}$ is stable. By (i), we have the exact sequence $0 \to M \to P \otimes_R C$ for some projective R-module P. By applying the functor $(-)^{\curlyvee}$ to the above exact sequence, it is easy to see that $M^{\curlyvee}$ is a first syzygy. It follows from Theorem 2.2 that $M^{\curlyvee}$ is horizontally linked. In other words, $M^{\curlyvee} \cong \lambda^2(M^{\curlyvee})$. Therefore, we obtain the isomorphisms

$$\begin{aligned} M \cong C \otimes_R M^{\curlyvee} &\cong C \otimes_R \lambda^2(M^{\curlyvee}) \\ &\cong \Omega_C \mathsf{Tr}_C(\lambda(M^{\curlyvee})) \\ &\cong \Omega_C \mathsf{Tr}_C \,\mathsf{Hom}_R(C, C \otimes_R \lambda(M^{\curlyvee})) \\ &\cong \lambda_R(C, \lambda_R(C, M)), \end{aligned}$$

by Remark 2.11(iii) and our assumptions. □

For an integer n, set $X^n(R) := \{\mathfrak{p} \in \mathsf{Spec}(R) \mid \mathsf{depth}\, R_{\mathfrak{p}} \leq n\}$.

Lemma 3.4. *Let M be an R-module. Consider the natural map*

$$\mu : M \to \mathsf{Hom}_R(C, M \otimes_R C).$$

Then the following statements hold true:

(i) *If M satisfies $\widetilde{S}_1$ and $\mu_{\mathfrak{p}}$ is a monomorphism for all $\mathfrak{p} \in X^0(R)$, then μ is a monomorphism.*

(ii) *If M satisfies $\widetilde{S}_2$, $M \otimes_R C$ satisfies $\widetilde{S}_1$ and $\mu_{\mathfrak{p}}$ is an isomorphism for all $\mathfrak{p} \in X^1(R)$, then μ is an isomorphism.*

Proof. (i) Set $L = \operatorname{Ker}(\mu)$ and let $\mathfrak{p} \in \operatorname{Ass}_R(L)$. Therefore, $\operatorname{depth}_{R_\mathfrak{p}}(M_\mathfrak{p}) = 0$. As M satisfies $\widetilde{S}_1$, we know $\mathfrak{p} \in X^0(R)$ and so $L_\mathfrak{p} = 0$, which is a contradiction. Therefore, μ is a monomorphism.

(ii) By (i), μ is a monomorphism. Consider the exact sequence

$$0 \to M \xrightarrow{\mu} \operatorname{Hom}_R(C, M \otimes_R C) \to L' \to 0,$$

where $L' := \operatorname{Coker}(\mu)$. Let $\mathfrak{p} \in \operatorname{Ass}_R(L')$. If $\mathfrak{p} \in \operatorname{Ass}_R(\operatorname{Hom}_R(C, M \otimes_R C)) \subseteq \operatorname{Ass}_R(M \otimes_R C)$, then $\operatorname{depth}_{R_\mathfrak{p}}(M \otimes_R C)_\mathfrak{p} = 0$. As $M \otimes_R C$ satisfies $\widetilde{S}_1$, one obtains $\mathfrak{p} \in X^0(R)$, which is a contradiction, because $\mu_\mathfrak{p}$ is an isomorphism for all $\mathfrak{p} \in X^0(R)$. Now let $\operatorname{depth}_{R_\mathfrak{p}}(\operatorname{Hom}_R(C, M \otimes_R C)_\mathfrak{p}) > 0$. It follows easily from the above exact sequence that $\operatorname{depth}_{R_\mathfrak{p}}(M_\mathfrak{p}) = 1$. As $M_\mathfrak{p}$ satisfies $\widetilde{S}_2$, we know $\mathfrak{p} \in X^1(R)$, which is a contradiction because $\mu_\mathfrak{p}$ is an isomorphism for all $\mathfrak{p} \in X^1(R)$. Therefore $L' = 0$ and μ is an isomorphism. □

The proof of the following lemma is dual to the proof of [Dibaei and Sadeghi 2015, Lemma 2.11].

Lemma 3.5. *Let R be a local ring, $n \geq 0$ an integer, and M an R-module. If $M \in \mathscr{B}_C$, then the following statements hold true*:

(i) $\operatorname{depth}_R(M) = \operatorname{depth}_R(M^\curlyvee)$ *and* $\dim_R(M) = \dim_R(M^\curlyvee)$.

(ii) *M satisfies $\widetilde{S}_n$ if and only if $M^\curlyvee$ does.*

(iii) *M is Cohen–Macaulay if and only if $M^\curlyvee$ is Cohen–Macaulay.*

Lemma 3.6 [Sather-Wagstaff et al. 2010, Lemma 2.8]. *Let M be an R-module that is in the Bass class $\mathscr{B}_C$. Then $\mathsf{G}_C\text{-}\mathsf{dim}_R(M) = 0$ if and only if $\mathsf{G}\text{-}\mathsf{dim}_R(M^\curlyvee) = 0$.*

In the following result, we give sufficient conditions for an element $M \in \mathscr{B}_C$ to be a horizontally linked module with respect to C.

Theorem 3.7. *Assume that $M \in \mathscr{B}_C$ is a C-syzygy and that $\operatorname{id}_{R_\mathfrak{p}}(C_\mathfrak{p}) < \infty$ for all $\mathfrak{p} \in X^1(R)$. If M is C-stable and $\underline{\operatorname{Hom}}_R(M^\curlyvee, C) = 0 = \operatorname{Ext}^1_R(M, C)$, then M is a horizontally linked module with respect to C.*

Proof. We shall prove that the conditions of Lemma 3.3 are satisfied. First note that

$$M \cong M^\curlyvee \otimes C, \tag{3.7.1}$$

because $M \in \mathscr{B}_C$. As seen in the proof of Lemma 3.3, $M^\curlyvee$ is horizontally linked. In other words, $M^\curlyvee \cong \lambda^2(M^\curlyvee)$ and so we obtain the exact sequence

$$0 \to M^\curlyvee \to P \to \operatorname{Tr}\lambda(M^\curlyvee) \to 0, \tag{3.7.2}$$

where P is a projective module. Applying $- \otimes_R C$ gives the exact sequence

$$0 \to \operatorname{Tor}_1^R(\operatorname{Tr}\lambda(M^\curlyvee), C) \to M^\curlyvee \otimes_R C \to P \otimes_R C \to \operatorname{Tr}\lambda(M^\curlyvee) \otimes_R C \to 0. \tag{3.7.3}$$

Let $\mathfrak{p} \in \mathrm{Ass}_R(\mathrm{Tor}_1^R(\mathrm{Tr}\lambda(M^\curlyvee), C))$. It follows from (3.7.1) and the exact sequence (3.7.3) that $\mathrm{depth}_{R_\mathfrak{p}}(M_\mathfrak{p}) = \mathrm{depth}_{R_\mathfrak{p}}((M^\curlyvee \otimes_R C)_\mathfrak{p}) = 0$. As M is a C-syzygy module, $\mathfrak{p} \in X^0(R)$. Note that, by Fact 2.9(iv), $M^\curlyvee \in \mathscr{A}_C$ and so, $\mathrm{G\text{-}dim}_{R_\mathfrak{q}}((M^\curlyvee)_\mathfrak{q}) = 0$ for all $\mathfrak{q} \in X^0(R)$ by Fact 2.9(ii) and Theorem 2.6(iv). As $\lambda(M^\curlyvee)$ is a syzygy, one has

$$\mathrm{Ext}_R^1(\mathrm{Tr}\lambda(M^\curlyvee), R) = 0. \tag{3.7.4}$$

It follows from (3.7.4), Theorem 2.6 and the exact sequence (3.7.2) that

$$\mathrm{G\text{-}dim}_{R_\mathfrak{p}}((\mathrm{Tr}\lambda(M^\curlyvee))_\mathfrak{p}) = 0.$$

In other words, by Fact 2.9(ii), $(\mathrm{Tr}\lambda(M^\curlyvee))_\mathfrak{p} \in \mathscr{A}_{C_\mathfrak{p}}$. Hence $\mathrm{Tor}_1^R(\mathrm{Tr}\lambda(M^\curlyvee), C)_\mathfrak{p} = 0$, which is a contradiction. Therefore, $\underline{\mathrm{Hom}}_R(\lambda(M^\curlyvee), C) \cong \mathrm{Tor}_1^R(\mathrm{Tr}\lambda(M^\curlyvee), C) = 0$ by (2.10.1).

Now we prove that the natural map $\mu : \lambda(M^\curlyvee) \to \mathrm{Hom}_R(C, C \otimes_R \lambda(M^\curlyvee))$ is an isomorphism. To this end, we concentrate on Lemma 3.4. As $M^\curlyvee$ is horizontally linked, we obtain the isomorphisms

$$\begin{aligned} \mathrm{Ext}_R^2(\mathrm{Tr}\lambda(M^\curlyvee), R) &\cong \mathrm{Ext}_R^1(\lambda^2(M^\curlyvee), R) \\ &\cong \mathrm{Ext}_R^1(M^\curlyvee, R) \\ &\cong \mathrm{Ext}_R^1(M^\curlyvee, C^\curlyvee) \\ &\cong \mathrm{Ext}_R^1(M, C) = 0, \end{aligned} \tag{3.7.5}$$

by [Takahashi and White 2010, Theorem 4.1 and Corollary 4.2]. It follows from (3.7.4) and (3.7.5) that $\lambda(M^\curlyvee)$ is second syzygy and so it satisfies $\widetilde{S}_2$ by [Maşek 2000, Proposition 11]. By the exact sequence $0 \to \lambda(M^\curlyvee) \to P' \to \mathrm{Tr}(M^\curlyvee) \to 0$ and the fact that $\mathrm{Tor}_1^R(\mathrm{Tr}(M^\curlyvee), C) \cong \underline{\mathrm{Hom}}_R(M^\curlyvee, C) = 0$, it follows that $\lambda(M^\curlyvee) \otimes_R C$ satisfies $\widetilde{S}_1$. As M satisfies $\widetilde{S}_1$, by Fact 2.9(ii), (iv), Lemma 3.5 and Theorem 2.6(iv), $\mathrm{G\text{-}dim}_{R_\mathfrak{p}}((M^\curlyvee)_\mathfrak{p}) = 0$ for all $\mathfrak{p} \in X^1(R)$. Therefore, $\mathrm{G\text{-}dim}_{R_\mathfrak{p}}((\lambda(M^\curlyvee))_\mathfrak{p}) = 0$ for all $\mathfrak{p} \in X^1(R)$ by [Martsinkovsky and Strooker 2004, Theorem 1] and so $(\lambda(M^\curlyvee))_\mathfrak{p} \in \mathscr{A}_{C_\mathfrak{p}}$ for all $\mathfrak{p} \in X^1(R)$ by Fact 2.9(ii). Hence μ is an isomorphism by Lemma 3.4. Now the assertion is clear by Lemma 3.3. $\square$

Martsinkovsky and Strooker [2004, Corollary 2] proved that, over a Gorenstein ring, horizontal linkage preserves the property of a module to be maximal Cohen–Macaulay, while they showed in the example on page 601 of the same paper that over non-Gorenstein rings, being maximal Cohen–Macaulay need not be preserved under horizontal linkage. In the following, it is shown that, over a Cohen–Macaulay local ring with the canonical module, horizontal linkage with respect to canonical module preserves maximal Cohen–Macaulay-ness. Note that over a Gorenstein ring, every module has finite Gorenstein injective dimension. Therefore, the following result can be viewed as a generalization of [Martsinkovsky and Strooker 2004, Corollary 2].

Corollary 3.8. *Let R be a Cohen–Macaulay local ring with canonical module ω_R. Assume that M is a maximal Cohen–Macaulay R-module of finite Gorenstein injective dimension. If M is ω_R-stable then the following statements hold true*:

(i) *M is horizontally linked with respect to ω_R.*

(ii) *$\lambda_R(\omega_R, M)$ has finite Gorenstein injective dimension.*

(iii) *$\lambda_R(\omega_R, M)$ is maximal Cohen–Macaulay.*

Proof. (i) By Fact 2.9(ii), $M \in \mathscr{B}_{\omega_R}$. As M is maximal Cohen–Macaulay, it is a ω_R-syzygy and also $\operatorname{Ext}^1_R(M, \omega_R) = 0$. Therefore, by Theorem 3.7, it is enough to prove that $\underline{\operatorname{Hom}}_R(\operatorname{Hom}_R(\omega_R, M), \omega_R) = 0$. Note that $\operatorname{G-dim}_R(\operatorname{Hom}_R(\omega_R, M)) = 0$ by Theorem 2.6 and Lemma 3.6. Hence $\operatorname{G-dim}_R(\operatorname{Tr}\operatorname{Hom}_R(\omega_R, M)) = 0$ and $\operatorname{Tr}\operatorname{Hom}_R(\omega_R, M) \in \mathscr{A}_{\omega_R}$ by Fact 2.9(ii) so that $\operatorname{Tor}^R_i(\operatorname{Tr}\operatorname{Hom}_R(\omega_R, M), \omega_R) = 0$ for all $i > 0$. Indeed, by (2.10.1),

$$\underline{\operatorname{Hom}}_R(\operatorname{Hom}_R(\omega_R, M), \omega_R) \cong \operatorname{Tor}^R_1(\operatorname{Tr}\operatorname{Hom}_R(\omega_R, M), \omega_R) = 0.$$

Therefore, by Theorem 3.7, M is horizontally linked with respect to ω_R.

(ii) As we have seen in part (i), $\operatorname{Tr}(\operatorname{Hom}_R(\omega_R, M)) \in \mathscr{A}_{\omega_R}$. Hence by Fact 2.9(iv) and Remark 2.11(i), $\operatorname{Tr}_{\omega_R}(\operatorname{Hom}_R(\omega_R, M)) \in \mathscr{B}_{\omega_R}$. Therefore, $\operatorname{Gid}_R(\lambda_R(\omega_R, M)) < \infty$ by Fact 2.9(i) and the exact sequence (3.1.1).

(iii) By Lemma 3.5, $\operatorname{Hom}_R(\omega_R, M)$ is maximal Cohen–Macaulay. Therefore $\operatorname{Tr}_{\omega_R}(\operatorname{Hom}_R(\omega_R, M))$ is maximal Cohen–Macaulay by Theorem 2.6(ii). It follows from the exact sequence (3.1.1) that $\lambda_R(\omega_R, M)$ is maximal Cohen–Macaulay. $\square$

To prove Theorem A, we first bring the following lemma and recall a definition.

Lemma 3.9. *Let R be a Cohen–Macaulay local ring and let I be an unmixed ideal of R. Assume that K is a semidualizing R/I-module and that M is an R-module which is linked by I with respect to K. Then* $\operatorname{grade}_R(M) = \operatorname{grade}_R(I)$.

Proof. First note that $\operatorname{grade}_R(M) = \inf\{\operatorname{depth} R_{\mathfrak{q}} \mid \mathfrak{q} \in \operatorname{Supp}_R(M)\}$. Therefore, $\operatorname{grade}_R(M) = \operatorname{depth} R_{\mathfrak{p}}$ for some $\mathfrak{p} \in \operatorname{Min}_R(M)$ and so $\mathfrak{p}/I \in \operatorname{Min}_{R/I}(M)$. As M is linked by I with respect to K, it is a first K-syzygy module and so $\mathfrak{p}/I \in \operatorname{Ass}_{R/I}(R/I)$, because $\operatorname{Ass}_{R/I}(K) = \operatorname{Ass}_{R/I}(R/I)$. As I is unmixed, $\operatorname{grade}(I) = \operatorname{depth} R_{\mathfrak{p}}$. $\square$

Let $(R, \mathfrak{m}, k)$ be a local ring and let M be an R-module. For every integer $n \geq 0$ the n-th Bass number $\mu^n_R(M)$ is the dimension of the k-vector space $\operatorname{Ext}^n_R(k, M)$.

Definition 3.10 [Avramov and Foxby 1997]. An ideal $\mathfrak{a}$ of a local ring R is called *quasi-Gorenstein* if $\operatorname{G-dim}_R(R/\mathfrak{a}) < \infty$ and for every $i \geq 0$ there is an equality of Bass numbers

$$\mu_R^{i+\operatorname{depth} R}(R) = \mu_{R/\mathfrak{a}}^{i+\operatorname{depth} R/\mathfrak{a}}(R/\mathfrak{a}).$$

Theorem 3.11 [Avramov and Foxby 1997, Corollary 7.9]. *Let R be a Cohen–Macaulay local ring with canonical module ω_R and let $\mathfrak{a}$ be a quasi-Gorenstein ideal of R. For an $R/\mathfrak{a}$-module M, we have $\mathsf{Gid}_R(M) < \infty$ if and only if $\mathsf{Gid}_{R/\mathfrak{a}}(M) < \infty$. Also, $\mathsf{G\text{-}dim}_R(M) < \infty$ if and only if $\mathsf{G\text{-}dim}_{R/\mathfrak{a}}(M) < \infty$.*

We now present Theorem A.

Theorem 3.12. *Let R be a Cohen–Macaulay local ring of dimension d with canonical module ω_R and let $\mathfrak{a}$ be a Cohen–Macaulay quasi-Gorenstein ideal of grade n, $\overline{R} = R/\mathfrak{a}$. Assume that M is a Cohen–Macaulay R-module of grade n and of finite Gorenstein injective dimension such that $\mathfrak{a} \subseteq \operatorname{Ann}_R(M)$. If M is $\omega_{\overline{R}}$-stable then the following statements hold true*:

(i) *M is linked by ideal $\mathfrak{a}$ with respect to $\omega_{\overline{R}}$.*

(ii) *$\lambda_{\overline{R}}(\omega_{\overline{R}}, M)$ has finite Gorenstein injective dimension.*

(iii) *$\lambda_{\overline{R}}(\omega_{\overline{R}}, M)$ is Cohen–Macaulay of grade n.*

Proof. (i) As R is Cohen–Macaulay,

$$d - n = d - \mathsf{grade}(\mathfrak{a}) = \dim(R/\mathfrak{a}).$$

On the other hand, as M is Cohen–Macaulay of grade n,

$$\mathsf{depth}_{\overline{R}}(M) = \mathsf{depth}_R(M) = \dim_R(M) = d - n.$$

Therefore, M is a maximal Cohen–Macaulay $\overline{R}$-module. By Theorem 3.11, $\mathsf{Gid}_{\overline{R}}(M)$ is finite and so M is horizontally linked with respect to $\omega_{\overline{R}}$ as an $\overline{R}$-module by Corollary 3.8.

(ii) By Corollary 3.8, $\mathsf{Gid}_{\overline{R}}(\lambda_{\overline{R}}(\omega_{\overline{R}}, M)) < \infty$, which by Theorem 3.11 is equivalent to $\mathsf{Gid}_R(\lambda_{\overline{R}}(\omega_{\overline{R}}, M)) < \infty$.

(iii) By Corollary 3.8, $\lambda_{\overline{R}}(\omega_{\overline{R}}, M)$ is a maximal Cohen–Macaulay $\overline{R}$-module. Hence

$$\mathsf{depth}_R(\lambda_{\overline{R}}(\omega_{\overline{R}}, M)) = \mathsf{depth}_{\overline{R}}(\lambda_{\overline{R}}(\omega_{\overline{R}}, M)) = \dim(R/\mathfrak{a}).$$

Also, by Lemma 3.9, $\mathsf{grade}_R(\lambda_{\overline{R}}(\omega_{\overline{R}}, M)) = n$. Hence,

$$\dim_R(\lambda_{\overline{R}}(\omega_{\overline{R}}, M)) = d - n = \dim R/\mathfrak{a}.$$

Therefore, $\lambda_{\overline{R}}(\omega_{\overline{R}}, M)$ is Cohen–Macaulay as an R-module. $\square$

Let R be a Cohen–Macaulay local ring with canonical module ω_R. Set

$$\mathscr{X} := \mathrm{CM}(R) \cap \mathscr{A}_{\omega_R} \quad \text{and} \quad \mathscr{Y} := \mathrm{CM}(R) \cap \mathscr{B}_{\omega_R},$$

where $\mathrm{CM}(R)$ is the category of Cohen–Macaulay R-module. Now we prove Theorem B.

Theorem 3.13. *Let R be a Cohen–Macaulay local ring with canonical module ω_R and let $\mathfrak{a}$ be a Cohen–Macaulay quasi-Gorenstein ideal of grade n, $\overline{R} = R/\mathfrak{a}$. There is an adjoint equivalence*

$$\left\{ M \in \mathscr{X} \;\middle|\; \begin{array}{c} M \text{ is linked} \\ \text{by the ideal } \mathfrak{a} \end{array} \right\} \underset{\mathrm{Hom}_{\overline{R}}(\omega_{\overline{R}},-)}{\overset{-\otimes_{\overline{R}}\omega_{\overline{R}}}{\rightleftarrows}} \left\{ N \in \mathscr{Y} \;\middle|\; \begin{array}{c} N \text{ is linked by the ideal} \\ \mathfrak{a} \text{ with respect to } \omega_{\overline{R}} \end{array} \right\}.$$

Proof. Let $M \in \mathscr{X}$, which is linked by the ideal $\mathfrak{a}$. By Theorem 3.11, $M \in \mathscr{A}_{\omega_{\overline{R}}}$. Note that $\mathfrak{a}$ is a G-perfect ideal and so $\mathrm{grade}_R(M) = \mathrm{grade}_R(\mathfrak{a})$ by [Sadeghi 2017, Lemma 3.16]. Therefore

$$\mathrm{depth}_R(M) = \dim_R(M) = \dim R - \mathrm{grade}_R(M) = \dim R - \mathrm{grade}_R(\mathfrak{a}).$$

Hence M is a maximal Cohen–Macaulay $\overline{R}$-module. Set $N = M \otimes_{\overline{R}} \omega_{\overline{R}}$. By [Dibaei and Sadeghi 2015, Lemma 2.11], N is a maximal Cohen–Macaulay $\overline{R}$-module. Therefore $N \in \mathrm{CM}(R)$. Also, by Fact 2.9(iv) and Theorem 3.11, $N \in \mathscr{B}_{\omega_R}$. Hence $N \in \mathscr{Y}$. As $M \in \mathscr{A}_{\omega_{\overline{R}}}$,

$$M \cong \mathrm{Hom}_{\overline{R}}(\omega_{\overline{R}}, N). \tag{3.13.1}$$

Note that M is stable $\overline{R}$-module by Theorem 2.2. It follows from (3.13.1) that N is $\omega_{\overline{R}}$-stable. Hence, by Theorem 3.12, N is linked by the ideal $\mathfrak{a}$ with respect to $\omega_{\overline{R}}$.

Conversely, assume that $N \in \mathscr{Y}$, which is linked by the ideal $\mathfrak{a}$ with respect to $\omega_{\overline{R}}$. As N is Cohen–Macaulay, by Lemma 3.9,

$$\mathrm{depth}_R(N) = \dim_R(N) = \dim R - \mathrm{grade}_R(N) = \dim R - \mathrm{grade}_R(\mathfrak{a}).$$

Therefore N is a maximal Cohen–Macaulay $\overline{R}$-module. Set $M = \mathrm{Hom}_{\overline{R}}(\omega_{\overline{R}}, N)$. Note that by Theorem 3.11, $N \in \mathscr{B}_{\omega_{\overline{R}}}$. Hence $M \in \mathscr{A}_{\omega_R}$ by Fact 2.9(iv), and Theorem 3.11. Also, by Lemma 3.5, M is a maximal Cohen–Macaulay $\overline{R}$-module. Thus $M \in \mathscr{X}$. Set $X = \mathrm{Hom}_{\overline{R}}(\omega_{\overline{R}}, \lambda_{\overline{R}}(\omega_{\overline{R}}, N))$. It follows from Theorem 3.12(ii), Fact 2.9(ii), (iv), and Theorem 3.11 that $X \in \mathscr{A}_{\omega_{\overline{R}}}$. Also, by Theorem 3.12(iii) and Lemma 3.5, X is a maximal Cohen–Macaulay $\overline{R}$-module. Therefore, by Theorem 2.6(ii), (iv) and Fact 2.9(ii), $\mathrm{G}\text{-}\dim_{\overline{R}}(\lambda_{\overline{R}} X) = 0$. In other words, $\lambda_{\overline{R}} X \in \mathscr{A}_{\omega_{\overline{R}}}$. Hence,

$$\lambda_{\overline{R}} X \cong \mathrm{Hom}_{\overline{R}}(\omega_{\overline{R}}, \lambda_{\overline{R}} X \otimes_{\overline{R}} \omega_{\overline{R}}). \tag{3.13.2}$$

As $\lambda_{\overline{R}} X$ is a first syzygy of $\mathrm{Tr}_{\overline{R}} X$, by Fact 2.9(i), $\mathrm{Tr}_{\overline{R}} X \in \mathscr{A}_{\omega_{\overline{R}}}$. Therefore $\underline{\mathrm{Hom}}_{\overline{R}}(X, \omega_{\overline{R}}) \cong \mathrm{Tor}_1^{\overline{R}}(\mathrm{Tr}_{\overline{R}} X, \omega_{\overline{R}}) = 0$. As N is linked by the ideal $\mathfrak{a}$ with respect to $\omega_{\overline{R}}$, it follows from Remark 2.11(iii) that

$$N \cong \Omega_{\omega_{\overline{R}}} \mathrm{Tr}_{\omega_{\overline{R}}} X \cong \lambda_{\overline{R}} X \otimes_{\overline{R}} \omega_{\overline{R}}. \tag{3.13.3}$$

It follows from (3.13.2) and (3.13.3) that $M \cong \lambda_{\overline{R}} X$. Hence, by [Avramov 1998, Corollary 1.2.5], M is a stable $\overline{R}$-module. By [Martsinkovsky and Strooker 2004, Theorem 1], M is linked by the ideal $\mathfrak{a}$. $\square$

Let $\mathfrak{a}$ be an ideal of R an let M be an $R/\mathfrak{a}$-module. Recall that M is said to be *self-linked* by the ideal $\mathfrak{a}$ if $M \cong \lambda_{R/\mathfrak{a}}M$. Let K be a semidualizing $R/\mathfrak{a}$-module. An $R/\mathfrak{a}$-module N is called *self-linked* by the ideal $\mathfrak{a}$ with respect to K if $N \cong \lambda_{R/\mathfrak{a}}(K, N)$.

Theorem 3.14. *Let R be a Cohen–Macaulay local ring with canonical module ω_R and let $\mathfrak{a}$ be a Cohen–Macaulay quasi-Gorenstein ideal of grade n, $\bar{R} = R/\mathfrak{a}$. There is an adjoint equivalence*

$$\left\{ M \in \mathscr{A}_{\omega_R} \,\middle|\, \begin{array}{c} M \text{ is self-linked} \\ \text{by the ideal } \mathfrak{a} \end{array} \right\} \underset{\mathsf{Hom}_{\bar{R}}(\omega_{\bar{R}},-)}{\overset{-\otimes_{\bar{R}}\omega_{\bar{R}}}{\rightleftarrows}} \left\{ N \in \mathscr{B}_{\omega_R} \,\middle|\, \begin{array}{c} N \text{ is self-linked by the} \\ \text{ideal } \mathfrak{a} \text{ with respect to } \omega_{\bar{R}} \end{array} \right\}.$$

Proof. Let $M \in \mathscr{A}_{\omega_R}$ and let $M \cong \lambda_{\bar{R}}M$. It follows from Theorem 3.11 that $M \in \mathscr{A}_{\omega_{\bar{R}}}$. Set $N = M \otimes_{\bar{R}} \omega_{\bar{R}}$. Therefore,

$$M \cong \mathsf{Hom}_{\bar{R}}(\omega_{\bar{R}}, N). \tag{3.14.1}$$

As $M \cong \Omega_{\bar{R}}\mathsf{Tr}_{\bar{R}}M$, we have $\mathsf{Tr}_{\bar{R}}M \in \mathscr{A}_{\omega_{\bar{R}}}$. Hence,

$$\underline{\mathsf{Hom}}_{\bar{R}}(M, \omega_{\bar{R}}) \cong \mathsf{Tor}_1^{\bar{R}}(\mathsf{Tr}_{\bar{R}}M, \omega_{\bar{R}}) = 0.$$

It follows from (3.14.1) and Remark 2.11(iii) that

$$\begin{aligned} \lambda_{\bar{R}}(\omega_{\bar{R}}, N) &= \Omega_{\omega_{\bar{R}}}\mathsf{Tr}_{\omega_{\bar{R}}}(\mathsf{Hom}_{\bar{R}}(\omega_{\bar{R}}, N)) \\ &\cong \Omega_{\omega_{\bar{R}}}\mathsf{Tr}_{\omega_{\bar{R}}}(M) \\ &\cong \lambda_{\bar{R}}M \otimes_{\bar{R}} \omega_{\bar{R}} \\ &\cong M \otimes_{\bar{R}} \omega_{\bar{R}} = N. \end{aligned}$$

In other words, N is self-linked by the ideal $\mathfrak{a}$ with respect to $\omega_{\bar{R}}$. Also, by Fact 2.9(iv), Theorem 3.11, $N \in \mathscr{B}_{\omega_R}$.

Conversely, assume that $N \in \mathscr{B}_{\omega_R}$ which is self-linked by the ideal $\mathfrak{a}$ with respect to $\omega_{\bar{R}}$. Set $M = \mathsf{Hom}_{\bar{R}}(\omega_{\bar{R}}, N)$. It follows from Fact 2.9(iv), Theorem 3.11 that $M \in \mathscr{A}_{\omega_R}$. As $N \cong \lambda_{\bar{R}}(\omega_{\bar{R}}, N)$, we have $\mathsf{Tr}_{\omega_{\bar{R}}}(M) \in \mathscr{B}_{\omega_{\bar{R}}}$ by the exact sequence (3.1.1), Fact 2.9(i) and Theorem 3.11. It follows from Remark 2.11(i) and Fact 2.9(iv) that $\mathsf{Tr}_{\bar{R}}(M) \in \mathscr{A}_{\omega_{\bar{R}}}$. Therefore $\underline{\mathsf{Hom}}_{\bar{R}}(M, \omega_{\bar{R}}) \cong \mathsf{Tor}_1^{\bar{R}}(\mathsf{Tr}_{\bar{R}}(M)\omega_{\bar{R}}) = 0$. Hence, by Remark 2.11(iii),

$$N \cong \lambda_{\bar{R}}(\omega_{\bar{R}}, N) \cong \lambda_{\bar{R}}(M) \otimes_{\bar{R}} \omega_{\bar{R}}. \tag{3.14.2}$$

As $\mathsf{Tr}_{\bar{R}}(M) \in \mathscr{A}_{\omega_{\bar{R}}}$, we have $\lambda_{\bar{R}}M \in \mathscr{A}_{\omega_{\bar{R}}}$. Hence

$$\lambda_{\bar{R}}M \cong \mathsf{Hom}_{\bar{R}}(\omega_{\bar{R}}, \lambda_{\bar{R}}M \otimes_{\bar{R}} \omega_{\bar{R}}). \tag{3.14.3}$$

It follows from (3.14.2) and (3.14.3) that $M \cong \lambda_{\bar{R}}M$. $\square$

4. Serre condition and vanishing of local cohomology

In this section, for a linked module, we study the relation between the Serre condition $\widetilde{S}_n$ and the vanishing of certain relative cohomology modules of its linked module. As a consequence, [Schenzel 1982, Theorem 4.1] is generalized. We start with the following lemma, which will be used in the proof of Theorem 4.2.

Lemma 4.1. *Let M be a C-syzygy module. Then $\mathsf{Ext}^1_R(\mathsf{Tr}_C(M^{\curlyvee}), C) = 0$. In particular, if M is horizontally linked with respect to C, then $\mathsf{Ext}^1_R(\mathsf{Tr}_C(M^{\curlyvee}), C) = 0$.*

Proof. Consider the exact sequence $0 \to M \to P \otimes_R C$, where P is a projective R-module. Applying the functor $(-)^{\curlyvee}$ to the above exact sequence, we get the exact sequence $0 \to M^{\curlyvee} \to P$. Therefore, $\mathsf{Ext}^1_R(\mathsf{Tr}M^{\curlyvee}, R) = 0$. By [Rotman 2009, Theorem 10.62], there is a third quadrant spectral sequence

$$\mathrm{E}_2^{p,q} = \mathsf{Ext}^p_R(\mathsf{Tor}^R_q(\mathsf{Tr}(M^{\curlyvee}), C), C) \Rightarrow \mathsf{Ext}^{p+q}_R(\mathsf{Tr}(M^{\curlyvee}), R).$$

Hence we obtain the following exact sequence

$$0 \to \mathsf{Ext}^1_R(\mathsf{Tr}(M^{\curlyvee}) \otimes_R C, C) \to \mathsf{Ext}^1_R(\mathsf{Tr}(M^{\curlyvee}), R),$$

by [Rotman 2009, Theorem 10.33]. Hence, by Remark 2.11,

$$\mathsf{Ext}^1_R(\mathsf{Tr}_C(M^{\curlyvee}), C) \cong \mathsf{Ext}^1_R(\mathsf{Tr}(M^{\curlyvee}) \otimes_R C, C) = 0. \qquad \square$$

The following is a generalization of [Martsinkovsky and Strooker 2004, Theorem 1].

Theorem 4.2. *Let M be an R-module which is horizontally linked with respect to C. Assume $M \in \mathscr{B}_C$. Then $\mathsf{G}_C\text{-}\mathsf{dim}_R(M) = 0$ if and only if $\mathsf{G}\text{-}\mathsf{dim}_R((\lambda_R(C, M))^{\curlyvee}) = 0$.*

Proof. Set $N = \lambda_R(C, M)$. Consider the exact sequence

$$0 \to M \to P^* \otimes_R C \to \mathsf{Tr}_C(N^{\curlyvee}) \to 0, \tag{4.2.1}$$

where P is a projective R-module; see (3.1.1). As $M \in \mathscr{B}_C$, we know $\mathsf{Tr}_C(N^{\curlyvee}) \in \mathscr{B}_C$ by the exact sequence (4.2.1) and Fact 2.9(i). Hence $\mathsf{Tr}(N^{\curlyvee}) \in \mathscr{A}_C$ by Remark 2.11 and Fact 2.9(iv). In particular,

$$\mathsf{Tr}(N^{\curlyvee}) \cong \mathsf{Hom}_R(C, \mathsf{Tr}(N^{\curlyvee}) \otimes_R C) \cong \mathsf{Hom}_R(C, \mathsf{Tr}_C(N^{\curlyvee})). \tag{4.2.2}$$

It follows from Theorem 2.6(ii), Lemma 3.6 and (4.2.2) that

$$\mathsf{G}\text{-}\mathsf{dim}_R(N^{\curlyvee}) = 0 \iff \mathsf{G}_C\text{-}\mathsf{dim}_R(\mathsf{Tr}_C(N^{\curlyvee})) = 0. \tag{4.2.3}$$

On the other hand, by the exact sequence (4.2.1)

(4.2.4)

$$\mathsf{G}_C\text{-}\mathsf{dim}_R(M) = 0 \text{ and } \mathsf{Ext}^1_R(\mathsf{Tr}_C(N^{\curlyvee}), C) = 0 \iff \mathsf{G}_C\text{-}\mathsf{dim}_R(\mathsf{Tr}_C(N^{\curlyvee})) = 0.$$

Now the assertion is clear by (4.2.3), (4.2.4) and Lemma 4.1. $\square$

The class $\mathscr{P}_C$ is precovering and then each R-module M has an augmented proper $\mathscr{P}_C$-resolution; that is, there is an R-complex

$$X^+ = \cdots \to C \otimes_R P_1 \to C \otimes_R P_0 \to M \to 0$$

such that $\mathrm{Hom}_R(Y, X^+)$ is exact for all $Y \in \mathscr{P}_C$. The truncated complex

$$X = \cdots \to C \otimes_R P_1 \to C \otimes_R P_0 \to 0$$

is called a proper $\mathscr{P}_C$-projective resolution of M. Proper $\mathscr{P}_C$-projective resolutions are unique up to homotopy equivalence.

Definition 4.3 [Takahashi and White 2010]. Let M and N be R-modules. The n-th relative cohomology module is defined as $\mathrm{Ext}^n_{\mathscr{P}_C}(M, N) = \mathrm{H}^n \mathrm{Hom}_R(X, N)$, where X is a proper $\mathscr{P}_C$-projective resolution of M.

Theorem 4.4 [Takahashi and White 2010, Theorem 4.1 and Corollary 4.2]. *Let M and N be R-modules. Then there exists an isomorphism*

$$\mathrm{Ext}^i_{\mathscr{P}_C}(M, N) \cong \mathrm{Ext}^i_R(M^\curlyvee, N^\curlyvee)$$

for all $i \geq 0$. Moreover, if M and N are in $\mathscr{B}_C$ then $\mathrm{Ext}^i_{\mathscr{P}_C}(M, N) \cong \mathrm{Ext}^i_R(M, N)$ for all $i \geq 0$.

For a positive integer n, a module M is called an n-th C-syzygy module if there is an exact sequence $0 \to M \to C_1 \to C_2 \to \cdots \to C_n$, where $C_i \in \mathscr{P}_C$ for all i. The following results will be used in the proof of Theorem 4.7.

Lemma 4.5. *Let M be an R-module such that $\mathrm{G}_{C_p}\text{-}\mathrm{dim}_{R_p}(M_\mathfrak{p}) < \infty$ for all $\mathfrak{p} \in X^{n-2}(R)$. Then the following statements are equivalent*:

(i) *M is an n-th C-syzygy module.*

(ii) $\mathrm{Ext}^i_R(\mathrm{Tr}_C M, C) = 0$ *for* $0 < i < n$.

Proof. The proof is analogous to [Maşek 2000, Theorem 43]. □

Theorem 4.6 [Dibaei and Sadeghi 2015, Proposition 2.4]. *Let C be a semidualizing R-module and M an R-module. For a positive integer n, consider the following statements*:

(i) $\mathrm{Ext}^i_R(\mathrm{Tr}_C M, C) = 0$ *for all i, $1 \leq i \leq n$.*

(ii) *M is an n-th C-syzygy module.*

(iii) *M satisfies $\widetilde{S}_n$.*

Then the following implications hold true:

(a) (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii).

(b) *If M has finite G_C-dimension on $X^{n-1}(R)$, then* (iii) $\Longrightarrow$ (i).

The following is a generalization of [Schenzel 1982, Theorem 4.1].

Theorem 4.7. *Let M be an R-module which is horizontally linked with respect to C. Assume that $M \in \mathscr{B}_C$. For a positive integer n, consider the following statements*:

(i) $\mathsf{Ext}^i_{\mathscr{P}_C}(\lambda_R(C, M), C) = 0$ *for* $0 < i < n$.

(ii) *M is an n-th C-syzygy module.*

(iii) *M satisfies $\widetilde{S}_n$.*

Then the following implications hold true:

(a) (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii).

(b) *If M has finite G_C-dimension on $X^{n-2}(R)$, then the statements* (i) *and* (ii) *are equivalent.*

(c) *If M has finite G_C-dimension on $X^{n-1}(R)$, then all the statements* (i)–(iii) *are equivalent.*

Proof. Set $N = \lambda_R(C, M)$. Consider the exact sequence

$$0 \to M \to P \otimes_R C \to \mathsf{Tr}_C(N^{\curlyvee}) \to 0, \tag{4.7.1}$$

where P is a projective R-module. By Lemma 4.1,

$$\mathsf{Ext}^1_R(\mathsf{Tr}_C(N^{\curlyvee}), C) = 0. \tag{4.7.2}$$

Therefore, by [Dibaei and Sadeghi 2015, Lemma 2.2], the exact sequence (4.7.1) induces the exact sequence

$$0 \to \mathsf{Tr}_C\mathsf{Tr}_C(N^{\curlyvee}) \to Q \otimes_R C \to \mathsf{Tr}_C M \to 0, \tag{4.7.3}$$

where Q is a projective R-module. Moreover, by [Sadeghi 2017, Lemma 2.12], there exists the following exact sequence

$$0 \to N^{\curlyvee} \to \mathsf{Tr}_C\mathsf{Tr}_C(N^{\curlyvee}) \to X \to 0, \tag{4.7.4}$$

where $\mathsf{G}_C\text{-}\mathsf{dim}_R(X) = 0$. As M is horizontally linked with respect to C, it is a C-syzygy module and so $\mathsf{Ext}^1_R(\mathsf{Tr}_C M, C) = 0$. Therefore, by the exact sequences (4.7.3) and (4.7.4), we obtain

$$\mathsf{Ext}^i_R(\mathsf{Tr}_C M, C) = 0 \text{ for } 1 \le i \le n \iff \mathsf{Ext}^i_R(N^{\curlyvee}, C) = 0 \text{ for } 1 \le i \le n-1. \tag{4.7.5}$$

As $M \in \mathscr{B}_C$, by Fact 2.9(i) and the exact sequence (4.7.1), $\mathsf{Tr}_C(N^{\curlyvee}) \in \mathscr{B}_C$. Hence, by Fact 2.9(iv) and Remark 2.11(i), $\mathsf{Tr}(N^{\curlyvee}) \in \mathscr{A}_C$. It follows from [Sadeghi 2017, Theorem 4.1] that

$$\mathsf{Ext}^i_R(N^{\curlyvee}, C) = 0 \text{ for } 0 < i < n \iff \mathsf{Ext}^i_R(N^{\curlyvee}, R) = 0 \text{ for } 0 < i < n. \tag{4.7.6}$$

Note that, by Theorem 4.4, we have the isomorphism

(4.7.7) $$\operatorname{Ext}^i_{\mathscr{P}_C}(N, C) \cong \operatorname{Ext}^i_R(N^{\curlyvee}, R) \quad \text{for all } i \geq 0.$$

Implications (a) and (c) follow from (4.7.5), (4.7.6), (4.7.7) and Theorem 4.6, and (b) follows from (4.7.5), (4.7.6), (4.7.7) and Lemma 4.5. □

Corollary 4.8. *Let C be a semidualizing R-module with $\operatorname{id}_{R_\mathfrak{p}}(C_\mathfrak{p}) < \infty$ for all $\mathfrak{p} \in X^{n-1}(R)$. Assume that M is an R-module which is horizontally linked with respect to C and that $M \in \mathscr{B}_C$. Then the following are equivalent*:

(i) *M satisfies $\widetilde{S}_n$.*

(ii) *M is an n-th C-syzygy module.*

(iii) $\operatorname{Ext}^i_R(\lambda_R(C, M), C) = 0$ *for* $0 < i < n$.

(iv) $\operatorname{Ext}^i_{\mathscr{P}_C}(\lambda_R(C, M), C) = 0$ *for* $0 < i < n$.

Proof. (i) $\Longrightarrow$ (iii): Set $N = \lambda_R(C, M)$. By Lemma 3.5,

(4.8.1) $$M \text{ satisfies } \widetilde{S}_n \quad \Longleftrightarrow \quad M^{\curlyvee} \text{ satisfies } \widetilde{S}_n.$$

By Lemma 4.1, $\operatorname{Ext}^1_R(\operatorname{Tr}_C(M^{\curlyvee}), C) = 0$. It follows from the exact sequence (3.1.1) that

(4.8.2) $$\operatorname{Ext}^i_R(N, C) = 0 \text{ for } 0 < i < n \quad \Longleftrightarrow \quad \operatorname{Ext}^i_R(\operatorname{Tr}_C(M^{\curlyvee}), C) = 0 \text{ for } 0 < i < n+1.$$

Now the assertion follows from (4.8.1), (4.8.2) and Theorem 4.6.

The equivalence of (i), (ii) and (iv) follows from Theorem 4.7. □

Now we are ready to present the first part of Theorem C.

Corollary 4.9. *Let $(R, \mathfrak{m})$ be a Cohen–Macaulay local ring of dimension $d > 0$ with canonical module ω_R. Assume that M is an R-module of finite Gorenstein injective dimension which is horizontally linked with respect to ω_R. The following are equivalent*:

(i) *M satisfies (S_n).*

(ii) $\mathrm{H}^i_{\mathfrak{m}}(\lambda_R(\omega_R, M)) = 0$ *for* $d - n < i < d$.

In particular, M is maximal Cohen–Macaulay if and only if $\lambda_R(\omega_R, M)$ is maximal Cohen–Macaulay.

Proof. This is an immediate consequence of Corollary 4.8, Fact 2.9(ii) and the local duality theorem. □

One may translate Corollary 4.9 to a change-of-rings result.

Corollary 4.10. *Let $(R, \mathfrak{m})$ be a Cohen–Macaulay local ring with canonical module ω_R and let $\mathfrak{a}$ be a Cohen–Macaulay quasi-Gorenstein ideal of R of grade n, $\bar{R} = R/\mathfrak{a}$. Assume that M is an R-module of finite Gorenstein injective dimension which is linked by the ideal $\mathfrak{a}$ with respect to $\omega_{\bar{R}}$. The following are equivalent*:

(i) *M satisfies (S_n).*

(ii) $\mathrm{H}^i_{\mathfrak{m}}(\lambda_{\bar{R}}(\omega_{\bar{R}}, M)) = 0$ *for* $\dim R/\mathfrak{a} - n < i < \dim R/\mathfrak{a}$.

Proof. This is an immediate consequence of Corollary 4.9 and Theorem 3.11. □

Recall that an R-module M of dimension $d \geq 1$ is called a *generalized Cohen–Macaulay* module if $\ell(\mathrm{H}^i_{\mathfrak{m}}(M)) < \infty$ for all i, $0 \leq i \leq d-1$, where ℓ denotes the length. For an R-module M and positive integer n, set $\mathcal{T}^C_n M := \mathrm{Tr}_C \Omega^{n-1} M$.

Theorem 4.11. *Let R be a Cohen–Macaulay local ring of dimension $d > 1$ and let C be a semidualizing R-module with $\mathrm{id}_{R_\mathfrak{p}}(C_\mathfrak{p}) < \infty$ for all $\mathfrak{p} \in \mathrm{Spec}(R) \setminus \{\mathfrak{m}\}$. Assume that M is a generalized Cohen–Macaulay R-module which is horizontally linked with respect to C and that $M \in \mathcal{B}_C$. Then $\mathrm{Ext}^i_R(M^\curlyvee, C) \cong \mathrm{H}^i_{\mathfrak{m}}(\lambda_R(C, M))$ for $0 < i < d$. In particular, $\lambda_R(C, M)$ is generalized Cohen–Macaulay.*

Proof. Set $X = M^\curlyvee$ and $N = \lambda_R(C, M)$. As M is generalized Cohen–Macaulay, by [Trung 1986, Lemmas 1.2 and 1.4] and Theorem 2.6(iv), $\mathrm{G}_{C_p}\text{-}\mathrm{dim}_{R_\mathfrak{p}}(M_\mathfrak{p}) = 0$ for all $\mathfrak{p} \in \mathrm{Spec}\, R \setminus \{\mathfrak{m}\}$. Therefore $\mathrm{G}\text{-}\mathrm{dim}_{R_\mathfrak{p}}(X_\mathfrak{p}) = 0$ for all $\mathfrak{p} \in \mathrm{Spec}\, R \setminus \{\mathfrak{m}\}$ by Lemma 3.6. Hence, $\mathrm{Ext}^i_R(X, C)$ has finite length for all $i > 0$. Consider the exact sequences

$$0 \to \mathrm{Ext}^i_R(X, C) \to \mathcal{T}^C_i X \to L_i \to 0, \tag{4.11.1}$$

$$0 \to L_i \to \oplus^{n_i} C \to \mathcal{T}^C_{i+1} X \to 0 \tag{4.11.2}$$

for all $i > 0$. By applying the functor $\Gamma_{\mathfrak{m}}(-)$ on the exact sequences (4.11.1) and (4.11.2), we get

$$\mathrm{H}^j_{\mathfrak{m}}(\mathcal{T}^C_{i-1} X) \cong \mathrm{H}^j_{\mathfrak{m}}(L_{i-1}) \quad \text{for all } i \text{ and } j, \text{ with } j \geq 1,\ i \geq 2, \tag{4.11.3}$$

$$\mathrm{Ext}^i_R(X, C) = \Gamma_{\mathfrak{m}}(\mathrm{Ext}^i_R(X, C)) \cong \Gamma_{\mathfrak{m}}(\mathcal{T}^C_i X) \quad \text{for all } i \geq 1, \tag{4.11.4}$$

and also

$$\mathrm{H}^j_{\mathfrak{m}}(\mathcal{T}^C_i X) \cong \mathrm{H}^{j+1}_{\mathfrak{m}}(L_{i-1}) \quad \text{for all } i \text{ and } j,\ 0 \leq j < d-1,\ i \geq 2. \tag{4.11.5}$$

As M is horizontally linked with respect to C, we have the exact sequence

$$0 \to N \to \oplus^n C \to \mathcal{T}^C_1 X \to 0$$

for some integer $n > 0$. By applying the functor $\Gamma_{\mathfrak{m}}(-)$ to the above exact sequence, we get the isomorphism

$$\mathrm{H}^j_{\mathfrak{m}}(\mathcal{T}^C_1 X) \cong \mathrm{H}^{j+1}_{\mathfrak{m}}(N) \quad \text{for all } j,\ 0 \leq j \leq d-2. \tag{4.11.6}$$

Now by using (4.11.3), (4.11.4), (4.11.5) and (4.11.6) we obtain the result. □

Now we give a proof for part (ii) of Theorem C as the following corollary.

Corollary 4.12. *Let $(R, \mathfrak{m}, k)$ be a Cohen–Macaulay local ring of dimension $d > 1$ with canonical module ω_R. Assume that M is an R-module of finite Gorenstein injective dimension which is horizontally linked with respect to ω_R. If M is generalized Cohen–Macaulay, then the following statements hold true*:

(i) $\mathrm{H}^i_{\mathfrak{m}}(\mathsf{Hom}_R(\omega_R, M)) \cong \mathsf{Hom}_R(\mathrm{H}^{d-i}_{\mathfrak{m}}(\lambda_R(\omega_R, M)), \boldsymbol{E}_R(k))$ *for* $0 < i < d$.

(ii) $\lambda_R(\omega_R, M)$ *is generalized Cohen–Macaulay.*

(iii) *If M is not maximal Cohen–Macaulay, then*

$$\mathsf{depth}_R(M) = \sup\{i < d \mid \mathrm{H}^i_{\mathfrak{m}}(\lambda_R(\omega_R, M)) \neq 0\}.$$

Proof. Parts (i) and (ii) follow immediately from Theorem 4.11 and the local duality theorem. Part (iii) follows from part (i) and Lemma 3.5. □

We end the paper with the following result, which is an immediate consequence of Corollary 4.12 and Theorem 3.11.

Corollary 4.13. *Let $(R, \mathfrak{m}, k)$ be a Cohen–Macaulay local ring with canonical module ω_R, let $\mathfrak{c}$ be a Cohen–Macaulay quasi-Gorenstein ideal of R, $\bar{R} = R/\mathfrak{c}$ and $\dim \bar{R} = d > 1$. Assume that M is an R-module of finite Gorenstein injective dimension which is linked by the ideal $\mathfrak{c}$ with respect to $\omega_{\bar{R}}$. If M is generalized Cohen–Macaulay, then*

$$\mathrm{H}^i_{\mathfrak{m}}(\mathsf{Hom}_{\bar{R}}(\omega_{\bar{R}}, M)) \cong \mathsf{Hom}_R(\mathrm{H}^{d-i}_{\mathfrak{m}}(\lambda_{\bar{R}}(\omega_{\bar{R}}, M)), \boldsymbol{E}_R(k))$$

for $0 < i < d$.

References

[Auslander and Bridger 1969] M. Auslander and M. Bridger, *Stable module theory*, Memoirs of the American Mathematical Society **94**, Amer. Math. Soc., Providence, RI, 1969. MR Zbl

[Avramov 1998] L. L. Avramov, "Infinite free resolutions", pp. 1–118 in *Six lectures on commutative algebra* (Bellaterra, 1996), edited by J. Elias et al., Progr. Math. **166**, Birkhäuser, Basel, 1998. MR Zbl

[Avramov and Foxby 1997] L. L. Avramov and H.-B. Foxby, "Ring homomorphisms and finite Gorenstein dimension", *Proc. London Math. Soc.* (3) **75**:2 (1997), 241–270. MR Zbl

[Celikbas et al. 2017] O. Celikbas, M. T. Dibaei, M. Gheibi, A. Sadeghi, and R. Takahashi, "Associated primes and syzygies of linked modules", preprint, 2017. To appear in *J. Commut. Algebra*. arXiv

[Christensen 2000] L. W. Christensen, *Gorenstein dimensions*, Lecture Notes in Mathematics **1747**, Springer, 2000. MR Zbl

[Christensen et al. 2006] L. W. Christensen, A. Frankild, and H. Holm, "On Gorenstein projective, injective and flat dimensions: a functorial description with applications", *J. Algebra* **302**:1 (2006), 231–279. MR Zbl

[Dibaei and Sadeghi 2013] M. T. Dibaei and A. Sadeghi, "Linkage of finite Gorenstein dimension modules", *J. Algebra* **376** (2013), 261–278. MR Zbl

[Dibaei and Sadeghi 2015] M. T. Dibaei and A. Sadeghi, "Linkage of modules and the Serre conditions", *J. Pure Appl. Algebra* **219**:10 (2015), 4458–4478. MR Zbl

[Dibaei et al. 2011] M. T. Dibaei, M. Gheibi, S. H. Hassanzadeh, and A. Sadeghi, "Linkage of modules over Cohen–Macaulay rings", *J. Algebra* **335** (2011), 177–187. MR Zbl

[Enochs and Jenda 1995] E. E. Enochs and O. M. G. Jenda, "Gorenstein injective and projective modules", *Math. Z.* **220**:4 (1995), 611–633. MR Zbl

[Foxby 1972] H.-B. Foxby, "Gorenstein modules and related modules", *Math. Scand.* **31** (1972), 267–284. MR Zbl

[Foxby 1975] H.-B. Foxby, "Quasi-perfect modules over Cohen–Macaulay rings", *Math. Nachr.* **66** (1975), 103–110. MR Zbl

[Gerko 2001] A. A. Gerko, "On homological dimensions", *Mat. Sb.* **192**:8 (2001), 79–94. In Russian; translated in *Sb. Math.* **192**:8 (2001), 1165–1179. MR Zbl

[Golod 1984] E. S. Golod, "G-dimension and generalized perfect ideals", *Trudy Mat. Inst. Steklov.* **165** (1984), 62–66. In Russian; translated in *Proc. Steklov Inst. Math.* **165** (1985), 67–71. MR Zbl

[Huneke 1982] C. Huneke, "Linkage and the Koszul homology of ideals", *Amer. J. Math.* **104**:5 (1982), 1043–1062. MR Zbl

[Huneke and Ulrich 1987] C. Huneke and B. Ulrich, "The structure of linkage", *Ann. of Math.* (2) **126**:2 (1987), 277–334. MR Zbl

[Iima and Takahashi 2016] K. Iima and R. Takahashi, "Perfect linkage of Cohen–Macaulay modules over Cohen–Macaulay rings", *J. Algebra* **458** (2016), 134–155. MR Zbl

[Lam 1991] T. Y. Lam, *A first course in noncommutative rings*, Graduate Texts in Mathematics **131**, Springer, 1991. MR Zbl

[Maşek 2000] V. Maşek, "Gorenstein dimension and torsion of modules over commutative Noetherian rings", *Comm. Algebra* **28**:12 (2000), 5783–5811. MR Zbl

[Martin 2000] H. M. Martin, "Linkage and the generic homology of modules", *Comm. Algebra* **28**:9 (2000), 4285–4301. MR Zbl

[Martsinkovsky 2010] A. Martsinkovsky, "1-torsion of finite modules over semiperfect rings", *J. Algebra* **324**:10 (2010), 2595–2607. MR Zbl

[Martsinkovsky and Strooker 2004] A. Martsinkovsky and J. R. Strooker, "Linkage of modules", *J. Algebra* **271**:2 (2004), 587–626. MR Zbl

[Nagel 2005] U. Nagel, "Liaison classes of modules", *J. Algebra* **284**:1 (2005), 236–272. MR Zbl

[Peskine and Szpiro 1974] C. Peskine and L. Szpiro, "Liaison des variétés algébriques, I", *Invent. Math.* **26** (1974), 271–302. MR Zbl

[Rotman 2009] J. J. Rotman, *An introduction to homological algebra*, 2nd ed., Springer, 2009. MR Zbl

[Sadeghi 2017] A. Sadeghi, "Linkage of finite G_C-dimension modules", *J. Pure Appl. Algebra* **221**:6 (2017), 1344–1365. MR Zbl

[Sather-Wagstaff et al. 2010] S. Sather-Wagstaff, T. Sharif, and D. White, "Tate cohomology with respect to semidualizing modules", *J. Algebra* **324**:9 (2010), 2336–2368. MR Zbl

[Schenzel 1982] P. Schenzel, "Notes on liaison and duality", *J. Math. Kyoto Univ.* **22**:3 (1982), 485–498. MR Zbl

[Takahashi and White 2010] R. Takahashi and D. White, "Homological aspects of semidualizing modules", *Math. Scand.* **106**:1 (2010), 5–22. MR Zbl

[Trung 1986] N. V. Trung, "Toward a theory of generalized Cohen–Macaulay modules", *Nagoya Math. J.* **102** (1986), 1–49. MR Zbl

[Yoshino 1990] Y. Yoshino, *Cohen–Macaulay modules over Cohen–Macaulay rings*, London Mathematical Society Lecture Note Series **146**, Cambridge University Press, 1990. MR Zbl

[Yoshino and Isogawa 2000] Y. Yoshino and S. Isogawa, "Linkage of Cohen–Macaulay modules over a Gorenstein ring", *J. Pure Appl. Algebra* **149**:3 (2000), 305–318. MR Zbl

Received July 1, 2017. Revised November 26, 2017.

MOHAMMAD T. DIBAEI

dibaeimt@ipm.ir
SCHOOL OF MATHEMATICS
INSTITUTE FOR RESEARCH IN FUNDAMENTAL SCIENCES (IPM)
TEHRAN
IRAN

and

FACULTY OF MATHEMATICAL AND COMPUTER SCIENCES
KHARAZMI UNIVERSITY
TEHRAN
IRAN

ARASH SADEGHI

sadeghiarash61@gmail.com
SCHOOL OF MATHEMATICS
INSTITUTE FOR RESEARCH IN FUNDAMENTAL SCIENCES (IPM)
TEHRAN
IRAN

PACIFIC JOURNAL OF MATHEMATICS
Vol. 294, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.294.329

BIHARMONIC HYPERSURFACES WITH CONSTANT SCALAR CURVATURE IN SPACE FORMS

YU FU AND MIN-CHUN HONG

Let M^n be a biharmonic hypersurface with constant scalar curvature in a space form $\mathbb{M}^{n+1}(c)$. We show that M^n has constant mean curvature if $c > 0$ and M^n is minimal if $c \leq 0$, provided that the number of distinct principal curvatures is no more than 6. This partially confirms Chen's conjecture and the generalized Chen's conjecture. As a consequence, we prove that there exist no proper biharmonic hypersurfaces with constant scalar curvature in Euclidean space $\mathbb{E}^{n+1}$ or hyperbolic space $\mathbb{H}^{n+1}$ for $n < 7$.

1. Introduction

In 1983, Eells and Lemaire [1983] introduced the concept of *biharmonic maps* in order to generalize classical theory of harmonic maps. A biharmonic map ϕ between an n-dimensional Riemannian manifold (M^n, g) and an m-dimensional Riemannian manifold (N^m, h) is a critical point of the bienergy functional

$$E_2(\phi) = \frac{1}{2} \int_M |\tau(\phi)|^2 \, dv_g,$$

where $\tau(\phi) = \operatorname{trace} \nabla d\phi$ is the tension field of ϕ that vanishes for a harmonic map. More clearly, the Euler–Lagrange equation associated to the bienergy is given by

$$\tau_2(\phi) = -\Delta\tau(\phi) - \operatorname{trace} R^N(d\phi, \tau(\phi))d\phi = 0,$$

where R^N is the curvature tensor of N^m (see, e.g., [Jiang 1987]). We call ϕ to be a biharmonic map if its bitension field $\tau_2(\phi)$ vanishes.

Biharmonic maps between Riemannian manifolds have been extensively studied by geometers. In particular, many authors investigated a special class of biharmonic maps named *biharmonic immersions*. An immersion $\phi : (M^n, g) \to (N^m, h)$ is biharmonic if and only if its mean curvature vector field $\vec{H}$ fulfills the fourth-order semilinear elliptic equations (see, e.g., [Caddeo et al. 2001]):

(1-1) $$\Delta\vec{H} + \operatorname{trace} R^N(d\phi, \vec{H})d\phi = 0.$$

MSC2010: primary 53C40, 53D12; secondary 53C42.
Keywords: biharmonic maps, biharmonic submanifolds, Chen's conjecture, generalized Chen's conjecture.

It is well known that any minimal immersion (satisfying $\vec{H}=0$) is harmonic. The nonharmonic biharmonic immersions are called proper biharmonic.

We should mention that *biharmonic submanifolds* in a Euclidean space $\mathbb{E}^m$ were independently defined by B. Y. Chen in the middle of 1980s (see [Chen 1991]) with the geometric condition $\Delta\vec{H}=0$, or equivalently $\Delta^2\phi=0$. Interestingly, both biharmonic submanifolds and biharmonic immersions in Euclidean spaces coincide with each other.

In recent years, the classification problem of biharmonic submanifolds has attracted great attention in geometry. In particular, there is a longstanding conjecture on biharmonic submanifolds due to Chen:

Chen's conjecture [1991]. *Every biharmonic submanifold in Euclidean space $\mathbb{E}^m$ is minimal.*

Chen's conjecture still remains open, even for hypersurfaces. In three decades, only partial answers to Chen's conjecture have been obtained, e.g., [Akutagawa and Maeta 2013; Alías et al. 2013; Chen 2015; Ou 2012]. In the case of hypersurfaces, Chen's conjecture is true for the following special cases:

- Surfaces in $\mathbb{E}^3$ [Chen 1991; Jiang 1987].
- Hypersurfaces with at most two distinct principal curvatures in $\mathbb{E}^m$ [Dimitrić 1992].
- Hypersurfaces in $\mathbb{E}^4$ [Hasanis and Vlachos 1995] (see also [Defever 1998]).
- $\delta(2)$-ideal and $\delta(3)$-ideal hypersurfaces in $\mathbb{E}^m$ [Chen and Munteanu 2013].
- Weakly convex hypersurfaces in $\mathbb{E}^m$ [Luo 2014].
- Hypersurfaces with at most three distinct principal curvatures in $\mathbb{E}^m$ [Fu 2015a].
- Generic hypersufaces with irreducible principal curvature vector fields in $\mathbb{E}^m$ [Koiso and Urakawa 2014].
- Invariant hypersurfaces of cohomogeneity one in $\mathbb{E}^m$ [Montaldo et al. 2016].

In 2001, Caddeo, Montaldo and Oniciuc [Caddeo et al. 2001] proposed the following generalized Chen's conjecture:

Generalized Chen's conjecture. *Every biharmonic submanifold in a Riemannian manifold with nonpositive sectional curvature is minimal.*

Recently, Ou and Tang [2012] constructed a family of counterexamples, where the generalized Chen's conjecture is false when the ambient space has nonconstant negative sectional curvature. However, the generalized Chen's conjecture remains open when the ambient spaces have constant sectional curvature. For more recent developments of the generalized Chen's conjecture, we refer to [Chen 2014; 2015; Montaldo and Oniciuc 2006; Oniciuc 2012; Nakauchi and Urakawa 2011; Ou 2016].

The classification of proper biharmonic submanifolds in Euclidean spheres is rather rich and interesting. The first example of proper biharmonic hypersurfaces is a generalized Clifford torus $S^p(1/\sqrt{2}) \times S^q(1/\sqrt{2}) \hookrightarrow \mathbb{S}^{n+1}$ with $p \neq q$ and $p+q=n$, given by Jiang [1986]. The complete classifications of biharmonic hypersurfaces in $\mathbb{S}^3$ and $\mathbb{S}^4$ were obtained in [Caddeo et al. 2001; Balmuş et al. 2010]. Moreover, biharmonic hypersurfaces with at most three distinct principal curvatures in $\mathbb{S}^n$ were classified in [Balmuş et al. 2010; Fu 2015b]. For more details, we refer the readers to [Balmuş et al. 2013; Loubeau and Oniciuc 2014; Oniciuc 2002; 2012; Ichiyama et al. 2010].

In general, the classification problem of proper biharmonic hypersurfaces in space forms becomes more complicated when the number of distinct principal curvatures is 4 or more.

In view of the above aspects, it is reasonable to study biharmonic submanifolds with some geometric conditions. In geometry, hypersurfaces with constant scalar curvature have been intensively studied by many geometers for the rigidity problem and classification problem, for instance, see [Cheng and Yau 1977]. Some estimate for scalar curvature of compact proper biharmonic hypersurfaces with constant scalar curvature in spheres was obtained in [Balmuş et al. 2008]. Recently, it was proved in [Fu 2015c] that a biharmonic hypersurface with constant scalar curvature in the 5-dimensional space forms $\mathbb{M}^5(c)$ necessarily has constant mean curvature.

Motivated by above results, in this paper we consider biharmonic hypersurfaces M^n with constant scalar curvatures in a space form $\mathbb{M}^n(c)$. More precisely, we get:

Theorem 1.1. *Let M^n be an orientable biharmonic hypersurface with at most six distinct principal curvatures in $\mathbb{M}^{n+1}(c)$. If the scalar curvature R is constant, then M^n has constant mean curvature.*

In general, it is difficult to deal with the biharmonic immersion equation (1-1) due to its high nonlinearity. In order to prove Theorem 1.1, we use some new ideas to overcome the difficulty of treating the equation of a biharmonic hypersurface. More precisely, we transfer the problem into a system of algebraic equations (see Lemma 3.3), so we can determine the behavior of the principal curvature functions by investigating the solution of the system of algebraic equations (see Lemma 3.4). Then, we are able to prove that a biharmonic hypersurface with constant scalar curvatures in a space form $\mathbb{M}^n(c)$ must have constant mean curvature, provided that the number of distinct principal curvature is no more than 6. We would like to point out that our approach in this paper is different from those in [Fu 2015b; 2015c; Defever 1998; Balmuş et al. 2010].

Remark 1.2. Balmuş, Montaldo and Oniciuc in [Balmuş et al. 2008] conjectured that the proper biharmonic hypersurfaces in $\mathbb{S}^{n+1}$ must have constant mean curvature. Theorem 1.1 with $c=1$ gives a partial answer to this conjecture.

We should point out that the complete classification of proper biharmonic hypersurfaces with constant mean curvature in a sphere is still open in the case where the number of distinct principal curvatures is more than 3 (see [Oniciuc 2012]).

Moreover, combining these results with the biharmonic equations in Section 2, we have:

Corollary 1.3. *Any biharmonic hypersurface with constant scalar curvature and with at most six distinct principal curvatures in Euclidean space $\mathbb{E}^{n+1}$ or hyperbolic space $\mathbb{H}^{n+1}$ is minimal.*

Thus, this result gives a partial answer to Chen's conjecture and the generalized Chen's conjecture.

Further, as a direct consequence, we get the following characterization result:

Corollary 1.4. *Any biharmonic hypersurface with constant scalar curvature in Euclidean space $\mathbb{E}^{n+1}$ or hyperbolic space $\mathbb{H}^{n+1}$ for $n < 7$ has to be minimal.*

Remark 1.5. We could replace or weaken the *constant scalar curvature* condition in Theorem 1.1 by *constant length of the second fundamental form* or *linear Weingarten type*, i.e., the scalar curvature R satisfying $R = aH + b$ for some constants a and b. In fact, the discussion is extremely similar to the proof of Theorem 1.1 and the same conclusion holds true as well.

The paper is organized as follows. In Section 2, we recall some necessary background theory for hypersurfaces and equivalent conditions for biharmonic hypersurfaces. In Section 3, we prove some useful lemmas (Lemmas 3.1–3.6), which are crucial to prove the main theorem. Finally, in Section 4, we give a proof of Theorem 1.1.

2. Preliminaries

In this section, we recall some basic material for the theory of hypersurfaces immersed in a Riemannian space form.

Let $\phi : M^n \to \mathbb{M}^{n+1}(c)$ be an isometric immersion of a hypersurface M^n into a space form $\mathbb{M}^{n+1}(c)$ with constant sectional curvature c. Denote the Levi-Civita connections of M^n and $\mathbb{M}^{n+1}(c)$ by ∇ and $\tilde{\nabla}$, respectively. Let X and Y denote the vector fields tangent to M^n and let ξ be a unit normal vector field. Then the Gauss and Weingarten formulas (see [Chen 2015]) are given, respectively, by

$$\tilde{\nabla}_X Y = \nabla_X Y + h(X, Y), \tag{2-1}$$

$$\tilde{\nabla}_X \xi = -AX, \tag{2-2}$$

where h is the second fundamental form and A is the Weingarten operator. Note that the second fundamental form h and the Weingarten operator A are related by

$$\langle h(X, Y), \xi\rangle = \langle AX, Y\rangle. \tag{2-3}$$

The mean curvature vector field $\vec{H}$ is defined by

$$\vec{H} = \frac{1}{n}\,\mathrm{trace}\,h. \tag{2-4}$$

Moreover, the Gauss and Codazzi equations are given, respectively, by

$$R(X,Y)Z = c\big(\langle Y,Z\rangle X - \langle X,Z\rangle Y\big) + \langle AY,Z\rangle AX - \langle AX,Z\rangle AY,$$
$$(\nabla_X A)Y = (\nabla_Y A)X,$$

where R is the curvature tensor of M^n and $(\nabla_X A)Y$ is given by

$$(\nabla_X A)Y = \nabla_X(AY) - A(\nabla_X Y) \tag{2-5}$$

for all X, Y, Z tangent to M^n.

Assume that $\vec{H} = H\xi$ and H denotes the mean curvature.

By identifying the tangent and the normal parts of the biharmonic condition (1-1) for hypersurfaces in a space form $\mathbb{M}^{n+1}(c)$, the following characterization result for M^n to be biharmonic was obtained (see also [Caddeo et al. 2002; Balmuş et al. 2010]):

Proposition 2.1. *The immersion $\phi : M^n \to \mathbb{M}^{n+1}(c)$ of a hypersurface M^n in an $n+1$-dimensional space form $\mathbb{M}^{n+1}(c)$ is biharmonic if and only if*

$$\begin{cases} \Delta H + H\,\mathrm{trace}\,A^2 = ncH, \\ 2A\ \mathrm{grad}\,H + nH\,\mathrm{grad}\,H = 0. \end{cases} \tag{2-6}$$

The Laplacian operator Δ on M^n acting on a smooth function f is given by

$$\Delta f = -\mathrm{div}(\nabla f) = -\sum_{i=1}^{n} < \nabla_{e_i}(\nabla f), e_i >= -\sum_{i=1}^{n}(e_i e_i - \nabla_{e_i} e_i) f. \tag{2-7}$$

The following result was obtained in [Fu 2015b]:

Theorem 2.2. *Let M^n be an orientable proper biharmonic hypersurface with at most three distinct principal curvatures in $\mathbb{M}^{n+1}(c)$. Then M^n has constant mean curvature.*

3. Some lemmas

We now consider an orientable biharmonic hypersurface M^n $(n > 3)$ in a space form $\mathbb{M}^{n+1}(c)$.

In general, the set M_A of all points of M^n, at which the number of distinct eigenvalues of the Weingarten operator A (i.e., the principal curvatures) is locally constant, is open and dense in M^n. Since M^n with at most three distinct principal curvatures everywhere in a space form $\mathbb{M}^{n+1}(c)$ is CMC, i.e., the mean curvature is constant (Theorem 2.2), one can work only on the connected component of M_A consisting of points where the number of principal curvatures is more than 3 (by passing to the limit, H will be constant on the whole M^n). On that connected component, the principal curvature functions of A are always smooth.

Suppose that, on the component, the mean curvature H is not constant. Thus, there is a point p where $\operatorname{grad} H(p) \neq 0$. In the following, we will work on a neighborhood of p where $\operatorname{grad} H(p) \neq 0$ at any point of M^n.

The second equation of (2-6) shows that $\operatorname{grad} H$ is an eigenvector of the Weingarten operator A with the corresponding principal curvature $-nH/2$. We may choose e_1 such that e_1 is parallel to $\operatorname{grad} H$, and with respect to some suitable orthonormal frame $\{e_1, \dots, e_n\}$, the Weingarten operator A of M takes the form

$$A = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \tag{3-1}$$

where λ_i are the principal curvatures and $\lambda_1 = -nH/2$. Therefore, it follows from (2-4) that $\sum_{i=1}^n \lambda_i = nH$, and hence

$$\sum_{i=2}^n \lambda_i = -3\lambda_1. \tag{3-2}$$

Denote by R the scalar curvature and by B the squared length of the second fundamental form h of M. It follows from (3-1) that B is given by

$$B = \operatorname{trace} A^2 = \sum_{i=1}^n \lambda_i^2 = \sum_{i=2}^n \lambda_i^2 + \lambda_1^2. \tag{3-3}$$

From the Gauss equation, the scalar curvature R is given by

$$R = n(n-1)c + n^2H^2 - B = n(n-1)c + 3\lambda_1^2 - \sum_{i=2}^n \lambda_i^2. \tag{3-4}$$

Hence

$$\sum_{i=2}^n \lambda_i^2 = n(n-1)c - R + 3\lambda_1^2. \tag{3-5}$$

Since $\operatorname{grad} H = \sum_{i=1}^n e_i(H)e_i$ and e_1 is parallel to $\operatorname{grad} H$, it follows that

$$e_1(H) \neq 0, \qquad e_i(H) = 0, \quad 2 \le i \le n,$$

and hence

$$e_1(\lambda_1) \neq 0, \qquad e_i(\lambda_1) = 0, \quad 2 \le i \le n. \tag{3-6}$$

Put $\nabla_{e_i} e_j = \sum_{k=1}^n \omega_{ij}^k e_k$ $(1 \le i, j \le n)$. A direct computation concerning the compatibility conditions $\nabla_{e_k}\langle e_i, e_i\rangle = 0$ and $\nabla_{e_k}\langle e_i, e_j\rangle = 0$ $(i \neq j)$ yields, respectively, that

$$\omega_{ki}^i = 0, \qquad \omega_{ki}^j + \omega_{kj}^i = 0, \quad i \neq j. \tag{3-7}$$

The Codazzi equation yields

$$e_i(\lambda_j) = (\lambda_i - \lambda_j)\omega_{ji}^j, \tag{3-8}$$

$$(\lambda_i - \lambda_j)\omega_{ki}^j = (\lambda_k - \lambda_j)\omega_{ik}^j, \tag{3-9}$$

for distinct i, j, k.

Moreover, from (3-6) we have

$$[e_i, e_j](\lambda_1) = 0,$$

which yields directly

$$\omega_{ij}^1 = \omega_{ji}^1, \qquad 2 \le i, j \le n \quad \text{and} \quad i \ne j. \tag{3-10}$$

Lemma 3.1. *Let M^n be an orientable biharmonic hypersurface with nonconstant mean curvature in $\mathbb{M}^{n+1}(c)$. Then the multiplicity of the principal curvature λ_1 (which equals $-nH/2$) is 1, i.e., $\lambda_j \ne \lambda_1$ for $2 \le j \le n$.*

Proof. If $\lambda_j = \lambda_1$ for $j \ne 1$, by putting $i = 1$ in (3-8), we get

$$0 = (\lambda_1 - \lambda_j)\omega_{j1}^j = e_1(\lambda_j) = e_1(\lambda_1),$$

which contradicts (3-6). □

Lemma 3.2. *The smooth real-valued functions λ_i and ω_{ii}^1 $(2 \le i \le n)$ satisfy the following differential equations*:

$$e_1e_1(\lambda_1) = e_1(\lambda_1)\Big(\sum\nolimits_{i=2}^n \omega_{ii}^1\Big) + \lambda_1(n(n-2)c - R + 4\lambda_1^2), \tag{3-11}$$

$$e_1(\lambda_i) = \lambda_i\omega_{ii}^1 - \lambda_1\omega_{ii}^1, \tag{3-12}$$

$$e_1(\omega_{ii}^1) = (\omega_{ii}^1)^2 + \lambda_1\lambda_i + c. \tag{3-13}$$

Proof. Substituting $H = -2\lambda_1/n$ into the first equation of (2-6), and using (2-7), (3-6), (3-3) and (3-5), we get (3-11). By putting $i = 1$ in (3-8), combining this with (3-9) gives (3-12).

Next, we will prove (3-13).

For $j = 1$ and $i \ne 1$ in (3-8), by (3-6) we have $\omega_{1i}^1 = 0$ $(i \ne 1)$. Combining this with (3-7), we have

$$\omega_{11}^i = 0, \quad \text{for } 1 \le i \le n. \tag{3-14}$$

For $j = 1$, and $k, i \ne 1$ in (3-9) we have

$$(\lambda_i - \lambda_1)\omega_{ki}^1 = (\lambda_k - \lambda_1)\omega_{ik}^1,$$

which together with (3-10) yields

$$\omega_{ki}^1 = 0, \quad k \ne i, \quad \text{if } \lambda_k \ne \lambda_i. \tag{3-15}$$

For $i \neq j$ and $2 \leq i, j \leq n$, if $\lambda_i = \lambda_j$, then by putting $k = 1$ in (3-9) we have

$$(\lambda_1 - \lambda_i)\omega_{i1}^j = 0,$$

which together with Lemma 3.1, (3-15) and (3-7) yields

$$\omega_{i1}^j = 0, \quad i \neq j, \quad \text{and} \quad 2 \leq i, j \leq n. \tag{3-16}$$

From the Gauss equation and (3-1), we have $\langle R(e_1, e_i)e_1, e_i\rangle = -\lambda_1\lambda_i - c$. On the other hand, the Gauss curvature tensor R is defined by

$$R(X, Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z.$$

Using (3-14), (3-16) and (3-7), a direct computation gives

$$\langle R(e_1, e_i)e_1, e_i\rangle = -e_1(\omega_{ii}^1) + (\omega_{ii}^1)^2.$$

Thus, we obtain differential equation (3-13), completing the proof of Lemma 3.2. □

Consider an integral curve of e_1 passing through $p = \gamma(t_0)$ as $\gamma(t)$, $t \in I$. Since $e_i(\lambda_1) = 0$ for $2 \leq i \leq n$ and $e_1(\lambda_1) \neq 0$, it is easy to show that there exists a local chart $(U; t = x^1, x^2, \ldots, x^m)$ around p, such that $\lambda_1(t, x^2, \ldots, x^m) = \lambda_1(t)$ on the whole neighborhood of p.

In the following, we begin our arguments under the assumption that the scalar curvature R is always constant. The following system of algebraic equations is important for us to proceed further.

Lemma 3.3. *Assume that R is constant. We have*

$$\sum_{i=2}^{n}(\omega_{ii}^1)^k = f_k(t), \quad \textit{for } k = 1, \ldots, 5, \tag{3-17}$$

where $f_k(t)$ are some smooth real-valued functions with respect to t.

Proof. Since $e_1(\lambda_1) \neq 0$, $\lambda_1 = \lambda_1(t)$ and R is constant, (3-11) becomes

$$\sum_{i=2}^{n}\omega_{ii}^1 = f_1(t), \tag{3-18}$$

where

$$f_1(t) = \frac{e_1e_1(\lambda_1) - \lambda_1\big(n(n-2)c + 4\lambda_1^2 - R\big)}{e_1(\lambda_1)}.$$

Taking the sum of (3-13) and (3-12) for i and taking into account (3-2) and (3-18), respectively, we have

$$\sum_{i=2}^{n}\big(\omega_{ii}^1\big)^2 = f_2(t), \tag{3-19}$$

$$\sum_{i=2}^{n}\lambda_i\omega_{ii}^1 = g_1(t), \tag{3-20}$$

where $f_2 = 3\lambda_1^2 - (n-1)c + e_1(f_1)$ and $g_1(t) = \lambda_1 f_1 - 3e_1(\lambda_1)$.

Multiplying by ω_{ii}^1 on both sides of (3-13), we have

$$\tfrac{1}{2}e_1((\omega_{ii}^1)^2) = (\omega_{ii}^1)^3 + \lambda_1\lambda_i\omega_{ii}^1 + c\omega_{ii}^1.$$

Using this and (3-18)–(3-20), we obtain

$$\sum_{i=2}^{n}(\omega_{ii}^1)^3 = f_3(t), \tag{3-21}$$

where $f_3 = \frac{1}{2}e_1(f_2) - \lambda_1 g_1 - cf_1$.

Differentiating (3-20) with respect to e_1 and using (3-12) and (3-13), we have

$$e_1(g_1) = 2\sum_{i=2}^{n}\lambda_i(\omega_{ii}^1)^2 + \lambda_1\sum_{i=2}^{n}\lambda_i^2 + c\sum_{i=2}^{n}\lambda_i - \lambda_1\sum_{i=2}^{n}(\omega_{ii}^1)^2. \tag{3-22}$$

From (3-2), (3-5) and (3-19), this yields

$$\sum_{i=2}^{n}\lambda_i(\omega_{ii}^1)^2 = g_2(t), \tag{3-23}$$

where $g_2 = \frac{1}{2}\big\{e_1(g_1) - \lambda_1\big(n(n-1)c - R + 3\lambda_1^2\big) + 3c\lambda_1 + \lambda_1 f_2\big\}$.

Multiplying by $(\omega_{ii}^1)^2$ on both sides of (3-13), we have

$$\tfrac{1}{3}e_1((\omega_{ii}^1)^3) = (\omega_{ii}^1)^4 + \lambda_1\lambda_i(\omega_{ii}^1)^2 + c(\omega_{ii}^1)^2.$$

Applying (3-19), (3-21) and (3-23) to this, we obtain

$$\sum_{i=2}^{n}(\omega_{ii}^1)^4 = f_4(t), \tag{3-24}$$

where $f_4 = \frac{1}{3}e_1(f_3) - \lambda_1 g_2 - cf_2$.

Multiplying by λ_i on both sides of (3-12) gives

$$\lambda_i^2\omega_{ii}^1 = \tfrac{1}{2}e_1(\lambda_i^2) + \lambda_1\lambda_i\omega_{ii}^1,$$

which together with (3-5) and (3-20) yields

$$\sum_{i=2}^{n}\lambda_i^2\omega_{ii}^1 = g_3(t), \tag{3-25}$$

where $g_3 = 3\lambda_1 e_1(\lambda_1) + \lambda_1 g_1$.

Differentiating (3-23) with respect to e_1 and using (3-12) and (3-13), we have

$$e_1(g_2) = 3\sum_{i=2}^{n}\lambda_i(\omega_{ii}^1)^3 - \lambda_1\sum_{i=2}^{n}(\omega_{ii}^1)^3 + 2\lambda_1\sum_{i=2}^{n}\lambda_i^2\omega_{ii}^1 + 2c\sum_{i=2}^{n}\lambda_i\omega_{ii}^1. \tag{3-26}$$

Substituting (3-20), (3-21) and (3-25) into (3-26) gives

$$\sum_{i=2}^{n}\lambda_i(\omega_{ii}^1)^3 = g_4(t), \tag{3-27}$$

where

$$g_4 = \tfrac{1}{3}(e_1(g_2)+\lambda_1 f_3 - 2\lambda_1 g_3 - 2cg_1).$$

Multiplying by $(\omega_{ii}^1)^3$ on both sides of (3-13), we have

$$\tfrac{1}{4}e_1((\omega_{ii}^1)^4) = (\omega_{ii}^1)^5 + \lambda_1\lambda_i(\omega_{ii}^1)^3 + c(\omega_{ii}^1)^3.$$

Applying (3-21), (3-24) and (3-27) to this, we have

$$\sum_{i=2}^{n}\left(\omega_{ii}^1\right)^5 = f_5(t), \tag{3-28}$$

where $f_5 = \frac{1}{4}e_1(f_4) - \lambda_1 g_4 - cf_3$. □

Lemma 3.4. *Assume that R is constant. If the number m of distinct principal curvatures satisfies $m \le 6$, then $e_i(\lambda_j) = 0$ for $2 \le i, j \le n$, i.e., all principal curvature λ_i depend only on one variable t.*

Proof. Since the number m of distinct principal curvatures satisfies $m \le 6$, there are at most five distinct principal curvatures for λ_i $(2 \le i \le n)$ except λ_1. It follows easily from (3-12) and (3-13) that

$$\lambda_i \ne \lambda_j \quad \Leftrightarrow \quad \omega_{ii}^1 \ne \omega_{jj}^1.$$

We now distinguish the following two cases:

Case A: Suppose that $m = 6$. We denote by $\tilde{\lambda}_i$ the five distinct principal curvatures with the corresponding multiplicities n_i for $1 \le i \le 5$. Note that here n_i are positive integers and $\sum_{i=1}^{5} n_i = n-1$ (see Lemma 3.1). According to (3-12), let

$$u_i := \frac{e_1(\tilde{\lambda}_i)}{\tilde{\lambda}_i - \lambda_1}.$$

Thus, the u_i are mutually different for $1 \le i \le 5$.

In this case, the system of polynomial equations (3-17) becomes

$$\begin{aligned}
n_1u_1 + n_2u_2 + n_3u_3 + n_4u_4 + n_5u_5 &= f_1,\\
n_1u_1^2 + n_2u_2^2 + n_3u_3^2 + n_4u_4^2 + n_5u_5^2 &= f_2,\\
n_1u_1^3 + n_2u_2^3 + n_3u_3^3 + n_4u_4^3 + n_5u_5^3 &= f_3,\\
n_1u_1^4 + n_2u_2^4 + n_3u_3^4 + n_4u_4^4 + n_5u_5^4 &= f_4,\\
n_1u_1^5 + n_2u_2^5 + n_3u_3^5 + n_4u_4^5 + n_5u_5^5 &= f_5.
\end{aligned} \tag{3-29}$$

Since $e_i(f_1) = 0$ for $2 \le i \le n$, differentiating both sides of the equations in (3-29) with respect to e_i $(2 \le i \le n)$, we obtain

$$\begin{aligned} n_1 e_i(u_1) + n_2 e_i(u_2) + n_3 e_i(u_3) + n_4 e_i(u_4) + n_5 e_i(u_5) &= 0, \\ n_1 u_1 e_i(u_1) + n_2 u_2 e_i(u_2) + n_3 u_3 e_i(u_3) + n_4 u_4 e_i((u_4) + n_5 u_5 e_i(u_5) &= 0, \\ n_1 u_1^2 e_i(u_1) + n_2 u_2^2 e_i(u_2) + n_3 u_3^2 e_i(u_3) + n_4 u_4^2 e_i(u_4) + n_5 u_5^2 e_i(u_5) &= 0, \\ n_1 u_1^3 e_i(u_1) + n_2 u_2^3 e_i(u_2) + n_3 u_3^3 e_i(u_3) + n_4 u_4^3 e_i(u_4) + n_5 u_5^3 e_i(u_5) &= 0, \\ n_1 u_1^4 e_i(u_1) + n_2 u_2^4 e_i(u_2) + n_3 u_3^4 e_i(u_3) + n_4 u_4^4 e_i(u_4) + n_5 u_5^4 e_i(u_5) &= 0. \end{aligned} \tag{3-30}$$

Now consider this system of five linear equations with five unknowns $e_i(u_k)$ for $1 \le k \le 5$.

According to Cramer's rule in linear algebra, for any k, $e_i(u_k) \equiv 0$ holds true if and only if the determinant of the coefficient matrix of (3-30) is not vanishing, i.e.,

$$\begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ u_1 & u_2 & u_3 & u_4 & u_5 \\ u_1^2 & u_2^2 & u_3^2 & u_4^2 & u_5^2 \\ u_1^3 & u_2^3 & u_3^3 & u_4^3 & u_5^3 \\ u_1^4 & u_2^4 & u_3^4 & u_4^4 & u_5^4 \end{vmatrix} \ne 0. \tag{3-31}$$

We note that the determinant in (3-31) is the famous Vandermonde determinant with order 5 and hence

$$\begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ u_1 & u_2 & u_3 & u_4 & u_5 \\ u_1^2 & u_2^2 & u_3^2 & u_4^2 & u_5^2 \\ u_1^3 & u_2^3 & u_3^3 & u_4^3 & u_5^3 \\ u_1^4 & u_2^4 & u_3^4 & u_4^4 & u_5^4 \end{vmatrix} = \prod_{1 \le j < i \le 5} (u_i - u_j). \tag{3-32}$$

Since the u_i are mutually different for $1 \le i \le 5$, (3-32) implies that (3-31) holds true identically. Hence, we have $e_i(u_k) = 0$ for any $1 \le k \le 5$ and $2 \le i \le n$.

Therefore, by using $e_i(u_k) = 0$ and

$$e_i e_1(u_k) - e_1 e_i(u_k) = [e_i, e_1](u_k) = \sum_{j=2}^{n} (\omega_{i1}^j - \omega_{1i}^j) e_j(u_k),$$

we get

$$e_i e_1(u_k) = 0.$$

Noting that with the notation u_k, (3-13) becomes

$$e_1(u_k) = (u_k)^2 + \lambda_1 \lambda_k + c.$$

Differentiating the above equation with respect to e_i, by taking into account $e_i(u_k)=0$ and $e_ie_1(u_k)=0$ we derive

$$e_i(\lambda_k)=0$$

for any $1\le k\le 5$ and $2\le i\le n$.

Case B: Suppose $m\le 5$. Denote by $\tilde{\lambda}_i$ the distinct principal curvatures with the corresponding multiplicities n_i for $1\le i\le 4$. Then the number of different u_i is less than or equal to 4. In the case that four of the u_i are mutually different, it is only necessary to consider the system (3-17) for $k=1, 2, 3, 4$. A similar discussion to the one in Case A yields the conclusion. If less than four of the u_i are mutually different, then the conclusion follows by some arguments similar to the above.

Thus, we conclude Lemma 3.4. □

Lemma 3.5. *For three arbitrary distinct principal curvatures λ_i, λ_j and λ_k, where $2\le i, j, k\le n$, we have the following relations*:

$$\omega_{ij}^k(\lambda_j-\lambda_k)=\omega_{ji}^k(\lambda_i-\lambda_k)=\omega_{kj}^i(\lambda_j-\lambda_i), \tag{3-33}$$

$$\omega_{ij}^k\omega_{ji}^k+\omega_{jk}^i\omega_{kj}^i+\omega_{ik}^j\omega_{ki}^j=0, \tag{3-34}$$

$$\omega_{ij}^k(\omega_{jj}^1-\omega_{kk}^1)=\omega_{ji}^k(\omega_{ii}^1-\omega_{kk}^1)=\omega_{kj}^i(\omega_{jj}^1-\omega_{ii}^1). \tag{3-35}$$

Proof. We recall from the beginning of this section that the number m of distinct principal curvatures satisfies $m\ge 4$. Hence, by taking into account the second expression of (3-7) and (3-9) for three distinct principal curvatures λ_i, λ_j and λ_k $(2\le i, j, k\le n)$, we obtain (3-33) and (3-34) immediately.

Let us consider (3-35). It follows from the Gauss equation that

$$\langle R(e_i, e_j)e_k, e_1\rangle=0.$$

Moreover, since $\omega_{ij}^1=0$ for $i\ne j$ from (3-7) and (3-16), from the definition of the curvature tensor we have

$$\omega_{ij}^k(\omega_{jj}^1-\omega_{kk}^1)=\omega_{ji}^k(\omega_{ii}^1-\omega_{kk}^1). \tag{3-36}$$

Similarly, by considering $\langle R(e_j, e_k)e_i, e_1\rangle=0$ one also has

$$\omega_{jk}^i(\omega_{kk}^1-\omega_{ii}^1)=\omega_{kj}^i(\omega_{jj}^1-\omega_{ii}^1),$$

which together with (3-7) and (3-36) gives (3-35). □

Lemma 3.6. *Under the assumptions as above, we have*

$$\omega_{ii}^1\omega_{jj}^1-\sum_{k=2,\, k\ne l_{(i,j)}}^{n}2\omega_{ij}^k\omega_{ji}^k=-\lambda_i\lambda_j-c, \quad \textit{for } \lambda_i\ne\lambda_j, \tag{3-37}$$

where $l_{(i,j)}$ stands for the indexes satisfying $\lambda_{l_{(i,j)}}=\lambda_i$ or λ_j.

Proof. In the following, we consider the case that the number m of distinct principal curvatures is 6.

Without loss of generality, except in the case of λ_1, we assume $\lambda_p, \lambda_q, \lambda_r, \lambda_u, \lambda_v$ are the five distinct principal curvatures in sequence with the corresponding multiplicities n_1, n_2, n_3, n_4, n_5, respectively, i.e.,

$$\lambda_1,\ \underbrace{\lambda_p, \dots, \lambda_p}_{n_1},\ \underbrace{\lambda_q, \dots, \lambda_q}_{n_2},\ \underbrace{\lambda_r, \dots, \lambda_r}_{n_3},\ \underbrace{\lambda_u, \dots, \lambda_u}_{n_4},\ \underbrace{\lambda_v, \dots, \lambda_v}_{n_5}.$$

We now compute $\langle R(e_p, e_q)e_p, e_q\rangle$. On one hand, it follows from the Gauss equation and (3-1) that

$$\langle R(e_p, e_q)e_p, e_q\rangle = -\lambda_p\lambda_q - c. \tag{3-38}$$

On the other hand, since

$$\begin{aligned}
\nabla_{e_p}\nabla_{e_q}e_p &= \sum_{k=1}^{n} e_p(\omega_{qp}^k)e_k + \sum_{k=1}^{n}\omega_{qp}^k\sum_{l=1}^{n}\omega_{pk}^l e_l,\\
\nabla_{e_q}\nabla_{e_p}e_p &= \sum_{k=1}^{n} e_q(\omega_{pp}^k)e_k + \sum_{k=1}^{n}\omega_{pp}^k\sum_{l=1}^{n}\omega_{qk}^l e_l,\\
\nabla_{[e_p,e_q]}e_p &= \sum_{k=1}^{n}(\omega_{pq}^k - \omega_{qp}^k)\sum_{l=1}^{n}\omega_{kp}^l e_l,
\end{aligned}$$

it follows that

$$\begin{aligned}
&\langle R(e_p, e_q)e_p, e_q\rangle = \\
&\quad e_p(\omega_{qp}^q) + \sum_{k=1}^{n}\omega_{qp}^k\omega_{pk}^q - e_q(\omega_{pp}^q) - \sum_{k=1}^{n}\omega_{pp}^k\omega_{qk}^q - \sum_{k=1}^{n}(\omega_{pq}^k - \omega_{qp}^k)\omega_{kp}^q.
\end{aligned} \tag{3-39}$$

Since $\lambda_p \neq \lambda_q$, from (3-8), (3-7) and Lemma 3.4 we have

$$\omega_{qp}^q = \omega_{qq}^p = \omega_{pp}^q = 0 \quad\text{and}\quad \sum_{k=2}^{n}\omega_{pp}^k\omega_{qk}^q = 0. \tag{3-40}$$

Moreover, if $2 \le k \le n_1 + 1$, then $\lambda_k = \lambda_p$, by the second expression of (3-7) and (3-9) we get

$$(\lambda_p - \lambda_k)\omega_{qp}^k = (\lambda_q - \lambda_k)\omega_{pq}^k \quad\text{and}\quad (\lambda_k - \lambda_q)\omega_{pk}^q = (\lambda_p - \lambda_q)\omega_{kp}^q,$$

which imply that

$$\omega_{pq}^k = \omega_{pk}^q = \omega_{kp}^q = 0. \tag{3-41}$$

Similarly, if $n_1+2 \le k \le n_1+n_2+1$, we also have

$$\omega_{pq}^k = \omega_{pk}^q = \omega_{kp}^q = 0. \tag{3-42}$$

Hence, by taking (3-40)–(3-42) into account, (3-39) becomes

$$\langle R(e_p, e_q)e_p, e_q\rangle = \omega_{pp}^1\omega_{qq}^1 + \sum_{k=n_1+n_2+2}^{n} \{\omega_{qp}^k\omega_{pk}^q - (\omega_{pq}^k - \omega_{qp}^k)\omega_{kp}^q\},$$

which together with (3-38), (3-7) and (3-34) gives

$$\omega_{pp}^1\omega_{qq}^1 - \sum_{k=n_1+n_2+2}^{n} 2\omega_{pq}^k\omega_{qp}^k = -\lambda_p\lambda_q - c. \tag{3-43}$$

Similarly, we can deduce other equations for different pairs $\omega_{pp}^1\omega_{rr}^1, \omega_{pp}^1\omega_{uu}^1, \dots$. Hence we get (3-37).

In the case that the number m of distinct principal curvatures is equal to 4 or 5, a very similar argument gives (3-37) as well. □

4. Proof of Theorem 1.1

Assume that the mean curvature H is not constant.

Differentiating (3-2) with respect to e_1 and using (3-12) and (3-13), we obtain

$$3e_1(\lambda_1) = \sum_{i=2}^{n}(\lambda_1 - \lambda_i)\omega_{ii}^1. \tag{4-1}$$

Following the previous section, we only deal with the case where the number of distinct principal curvatures is 6, i.e., $m = 6$. In fact, the proofs for the cases $m = 4, 5$ are very similar, so we omit them here without loss of generality.

According to Lemma 3.5, we consider the following cases:

Case A: $\omega_{pq}^r \ne 0$, $\omega_{pq}^u \ne 0$, and $\omega_{pq}^v \ne 0$. Since $\lambda_p, \lambda_q, \lambda_r, \lambda_u, \lambda_v$ are mutually different, equations (3-33) and (3-35) reduce to

$$\begin{aligned}\frac{\omega_{pp}^1-\omega_{qq}^1}{\lambda_p-\lambda_q} = \frac{\omega_{pp}^1-\omega_{rr}^1}{\lambda_p-\lambda_r} &= \frac{\omega_{qq}^1-\omega_{rr}^1}{\lambda_q-\lambda_r}\\ &= \frac{\omega_{pp}^1-\omega_{uu}^1}{\lambda_p-\lambda_u} = \frac{\omega_{qq}^1-\omega_{uu}^1}{\lambda_q-\lambda_u}\\ &= \frac{\omega_{pp}^1-\omega_{vv}^1}{\lambda_p-\lambda_v} = \frac{\omega_{qq}^1-\omega_{vv}^1}{\lambda_q-\lambda_v}.\end{aligned}$$

Thus, there exist two smooth functions φ and ψ depending on t such that

$$\omega_{ii}^1 = \varphi\lambda_i + \psi. \tag{4-2}$$

Differentiating with respect to e_1 on both sides of (4-2), and using (3-12) and (3-13) we get

$$e_1(\varphi) = \lambda_1(\varphi^2+1)+\varphi\psi, \tag{4-3}$$

$$e_1(\psi) = \psi(\lambda_1\varphi+\psi)+c. \tag{4-4}$$

Taking into account (4-2), and using (3-2), (3-5) one has

$$\sum_{i=2}^{n}\omega_{ii}^1 = -3\lambda_1\varphi+(n-1)\psi,$$

and (4-1) and (3-11), respectively, become

$$3e_1(\lambda_1) = \bigl(R-n(n-1)c-6\lambda_1^2\bigr)\varphi+(n+2)\lambda_1\psi, \tag{4-5}$$

$$e_1e_1(\lambda_1) = e_1(\lambda_1)(-3\lambda_1\varphi+(n-1)\psi)+\lambda_1\bigl(n(n-2)c-R+4\lambda_1^2\bigr). \tag{4-6}$$

Differentiating (4-5) with respect to e_1, we may eliminate $e_1e_1(\lambda_1)$ by (4-6). Using (4-3), (4-4) and (4-6) we have

$$3(n-4)e_1(\lambda_1)\psi = \lambda_1\bigl(6R-(4n^2-12n-3)c-27\lambda_1^2\bigr). \tag{4-7}$$

Note here, $n > 4$ since the number of distinct principal curvatures is 6.

Eliminating $e_1(\lambda_1)$ between (4-5) and (4-7) gives

$$\begin{aligned}(n-4)\bigl\{\bigl(R-n(n-1)c-6\lambda_1^2\bigr)\varphi\psi+(n+2)\lambda_1\psi^2\bigr\}&\\ = \lambda_1\bigl(6R-(4n^2-12n-3)c-27\lambda_1^2\bigr).&\end{aligned} \tag{4-8}$$

Further, differentiating (4-7) with respect to e_1, by (4-4), (4-6), (4-7), (4-5) we have

$$\bigl(432\lambda_1^4+a_1\lambda_1^2+a_2\bigr)\varphi+\bigl\{-54(n+3)\lambda_1^3+a_3\lambda_1\bigr\}\psi = 12(n-4)\lambda_1^3+a_4\lambda_1, \tag{4-9}$$

where

$$\begin{aligned}a_1 &= (97n^2-111n+60)c-105R,\\ a_2 &= ((4n^2-9n+9)c-6R)(n(n-1)c-R),\\ a_3 &= 12R-(4n^2-6n+21)c,\\ a_4 &= 3n(n-4)(n-2)c.\end{aligned}$$

Differentiating (4-9) with respect to e_1 and using (4-3) and (4-4), we get

$$\begin{aligned}&(1728\lambda_1^3+2a_1\lambda_1)\varphi e_1(\lambda_1)+(432\lambda_1^4+a_1\lambda_1^2+a_2)\{\lambda_1(\varphi^2+1)+\varphi\psi\}\\ &\quad+\{-162(n+3)\lambda_1^2+a_3\}\psi e_1(\lambda_1)+\{-54(n+3)\lambda_1^3+a_3\lambda_1\}\{\psi(\lambda_1\varphi+\psi)+c\}\\ &\qquad= (36(n-4)\lambda_1^2+a_4)e_1(\lambda_1).\end{aligned}$$

Multiplying by $3(n-4)$ on both sides of the above equation and using (4-5) and (4-7) we have

$$\begin{aligned}(n-4)\big(1728\lambda_1^3+2a_1\lambda_1\big)\varphi\big\{\big(R-n(n-1)c-6\lambda_1^2\big)\varphi+(n+2)\lambda_1\psi\big\}\\ +3(n-4)\big(432\lambda_1^4+a_1\lambda_1^2+a_2\big)\big\{\lambda_1(\varphi^2+1)+\varphi\psi\big\}\\ +\lambda_1\big\{-162(n+3)\lambda_1^2+a_3\big\}\big\{6R-(4n^2-12n-3)c-27\lambda_1^2\big\}\\ +3(n-4)\big\{-54(n+3)\lambda_1^3+a_3\lambda_1\big\}\big\{\psi(\lambda_1\varphi+\psi)+c\big\}\\ =(n-4)\big(36(n-4)\lambda_1^2+a_4\big)\big\{\big(R-n(n-1)c-6\lambda_1^2\big)\varphi+(n+2)\lambda_1\psi\big\}.\end{aligned}\tag{4-10}$$

Note that Equation (4-10) could be rewritten as

$$q_1(\lambda_1)\varphi^2+q_2(\lambda_1)\varphi\psi+q_3(\lambda_1)\psi^2+q_4(\lambda_1)\varphi+q_5(\lambda_1)\psi+q_6(\lambda_1)=0,\tag{4-11}$$

where q_i are nontrivial polynomials concerning the function λ_1 and given by:

$$\begin{aligned}q_1&=(n-4)\big(1728\lambda_1^3+2a_1\lambda_1\big)\big(R-n(n-1)c-6\lambda_1^2\big)\\ &\qquad+3(n-4)\big(432\lambda_1^4+a_1\lambda_1^2+a_2\big)\lambda_1,\\ q_2&=(n-4)(n+2)\lambda_1\big(1728\lambda_1^3+2a_1\lambda_1\big)\\ &\qquad+3(n-4)\big(432\lambda_1^4+a_1\lambda_1^2+a_2\big)+3(n-4)\big\{-54(n+3)\lambda_1^3+a_3\lambda_1\big\}\lambda_1,\\ q_3&=3(n-4)\big\{-54(n+3)\lambda_1^3+a_3\lambda_1\big\},\\ q_4&=(n-4)\big(36(n-4)\lambda_1^2+a_4\big)\big(R-n(n-1)c-6\lambda_1^2\big),\\ q_5&=-(n-4)(n+2)\big(36(n-4)\lambda_1^2+a_4\big)\lambda_1,\\ q_6&=-3(n-4)\big(432\lambda_1^4+a_1\lambda_1^2+a_2\big)\lambda_1\\ &\qquad+\lambda_1\big(-162(n+3)\lambda_1^2+a_3\big)\big\{6R-(4n^2-12n-3)c-27\lambda_1^2\big\}\\ &\qquad+3c(n-4)\big\{-54(n+3)\lambda_1^3+a_3\lambda_1\big\}.\end{aligned}\tag{4-12}$$

In the same manner, (4-8) and (4-9) could be also rewritten, respectively, as:

$$p_1(\lambda_1)\varphi\psi+p_2(\lambda_1)\psi^2=p_3(\lambda_1),\tag{4-13}$$

$$h_1(\lambda_1)\varphi+h_2(\lambda_1)\psi=h_3(\lambda_1),\tag{4-14}$$

where p_i, h_i $(i=1,2)$ are polynomials concerning the function λ_1 and given by

$$\begin{aligned}p_1&=(n-4)\big(R-n(n-1)c-6\lambda_1^2\big),\\ p_2&=(n-4)(n+2)\lambda_1,\\ p_3&=\lambda_1\big(6R-(4n^2-12n-3)c-27\lambda_1^2\big),\\ h_1&=432\lambda_1^4+a_1\lambda_1^2+a_2,\\ h_2&=-54(n+3)\lambda_1^3+a_3\lambda_1,\\ h_3&=12(n-4)\lambda_1^3+a_4\lambda_1.\end{aligned}\tag{4-15}$$

Multiplying by h_1^2 on both sides of (4-11), by taking into account (4-14) we may eliminate φ and get

$$P_1\psi^2+P_2\psi=P_3, \tag{4-16}$$

where

$$\begin{aligned} P_1&=q_1h_2^2-q_2h_1h_2+q_3h_1^2,\\ P_2&=-2q_1h_2h_3+q_2h_1h_3-q_4h_1h_2+q_5h_1^2,\\ P_3&=-q_1h_3^2-q_4h_1h_3-q_6h_1^2. \end{aligned} \tag{4-17}$$

Similarly, eliminating φ in (4-13) by using (4-14) yields

$$Q_1\psi^2+Q_2\psi=Q_3, \tag{4-18}$$

where

$$\begin{aligned} Q_1&=p_2h_1-p_1h_2,\\ Q_2&=p_1h_3,\\ Q_3&=p_3h_1. \end{aligned} \tag{4-19}$$

Moreover, multiplying by Q_1 and P_1 on both sides of the equations (4-16) and (4-18), respectively, after eliminating the 'ψ^2' part we obtain

$$(P_2Q_1-P_1Q_2)\psi=P_3Q_1-P_1Q_3. \tag{4-20}$$

Multiplying (4-20) by $P_1\psi$ and then combining this with (4-16) gives

$$\big\{P_1(P_3Q_1-P_1Q_3)+P_2(P_2Q_1-P_1Q_2)\big\}\psi=P_3(P_2Q_1-P_1Q_2). \tag{4-21}$$

At last, after eliminating ψ between (4-20) and (4-21) we get

$$\begin{aligned} P_1(P_3Q_1-P_1Q_3)^2+P_2(P_2Q_1-P_1Q_2)&(P_3Q_1-P_1Q_3)\\ &=P_3(P_2Q_1-P_1Q_2)^2. \end{aligned} \tag{4-22}$$

We observe from (4-12), (4-15), (4-17) and (4-19) that both P_i and Q_i $(1\le i\le 3)$ are polynomials concerning λ_1 with constant coefficients. Hence, it follows that

$$\begin{aligned} P_1&=-10077696(n-4)(n+3)(n-1)\lambda_1^{11}+\cdots,\\ P_2&=-839808(n-4)^2(11n+5)\lambda_1^{11}+\cdots,\\ P_3&=-69984(19n+113)\lambda_1^{13}+\cdots,\\ Q_1&=108(n-4)(n-1)\lambda_1^5+\cdots,\\ Q_2&=-72(n-4)^2\lambda_1^5+\cdots,\\ Q_3&=-11664\lambda_1^7+\cdots, \end{aligned}$$

where we only need to write the highest order terms of λ_1.

By substituting P_i and Q_i into (4-22), we get a polynomial equation concerning λ_1 with constant coefficients $c_i = c_i(n, c, R)$:

$$\sum_{i=0}^{47} c_i \lambda_1^i = 0, \tag{4-23}$$

where the coefficient c_{47} of the highest order term satisfies

$$c_{47} = -10077696(n-4)^2(n+3)(n-1)^2\big[69984 \times 108(19n+113) \\ +10077696 \times 11664(n+3)\big]^2 \neq 0.$$

Therefore, λ_1 has to be constant and $H = -2\lambda_1/n$ is a constant, which is a contradiction.

Case B: $\omega^r_{pq} \neq 0$, $\omega^u_{pq} \neq 0$, and $\omega^k_{ij} = 0$ for all other distinct triplets $\{i, j, k\}$ and distinct principal curvatures $\lambda_i, \lambda_j, \lambda_k$. Then, (3-37) implies that

$$\omega^1_{pp}\omega^1_{vv} = -\lambda_p\lambda_v - c, \tag{4-24}$$

$$\begin{aligned} \omega^1_{qq}\omega^1_{vv} &= -\lambda_q\lambda_v - c, \\ \omega^1_{rr}\omega^1_{vv} &= -\lambda_r\lambda_v - c, \\ \omega^1_{uu}\omega^1_{vv} &= -\lambda_u\lambda_v - c. \end{aligned} \tag{4-25}$$

Similar to Case A, since $\omega^r_{pq} \neq 0$ and $\omega^u_{pq} \neq 0$, (3-33) and (3-35) imply that

$$\omega^1_{ii} = \varphi\lambda_i + \psi, \quad \text{for } i = p, q, r, u, \tag{4-26}$$

where φ and ψ satisfy the differential equations (4-3) and (4-4).

Substituting (4-26) into (4-24) and (4-25), we obtain

$$\omega^1_{vv} = -\frac{1}{\varphi}\lambda_v, \tag{4-27}$$

$$\lambda_v\psi = c\varphi, \tag{4-28}$$

which means that ω^1_{vv} and λ_v are determined completely by φ and ψ.

Substitute (4-26)–(4-28) into (4-1), and then differentiate it with respect to e_1. By using (4-3), (4-4) and (3-11), a similar discussion as in Case A gives a polynomial concerning the function λ_1 with constant coefficients. Hence, λ_1 has to be constant, which yields a contradiction as well.

Case C: $\omega^r_{pq} \neq 0$ (or $\omega^r_{pq} = 0$), and all the $\omega^k_{ij} = 0$ for distinct triplets $\{i, j, k\}$ and distinct principal curvatures $\lambda_i, \lambda_j, \lambda_k$. Then, (3-37) implies that

$$\omega^1_{pp}\omega^1_{uu} = -\lambda_p\lambda_u - c, \quad \omega^1_{pp}\omega^1_{vv} = -\lambda_p\lambda_v - c, \tag{4-29}$$

$$\omega^1_{qq}\omega^1_{uu} = -\lambda_q\lambda_u - c, \quad \omega^1_{qq}\omega^1_{vv} = -\lambda_q\lambda_v - c, \tag{4-30}$$

$$\omega^1_{rr}\omega^1_{uu} = -\lambda_r\lambda_u - c, \quad \omega^1_{rr}\omega^1_{vv} = -\lambda_r\lambda_v - c, \tag{4-31}$$

$$\omega^1_{uu}\omega^1_{vv} = -\lambda_u\lambda_v - c. \tag{4-32}$$

We first consider $\lambda_i \neq 0$ for $i = p, q, r, u, v$. Consequently, (4-29)–(4-32) reduce to

$$\frac{\omega_{pp}^1}{\lambda_p} = \frac{\omega_{qq}^1}{\lambda_q} = \frac{\omega_{rr}^1}{\lambda_r} = -\frac{\lambda_u - \lambda_v}{\omega_{uu}^1 - \omega_{vv}^1},$$
$$\frac{\omega_{uu}^1}{\lambda_u} = \frac{\omega_{vv}^1}{\lambda_v} = -\frac{\lambda_p - \lambda_q}{\omega_{pp}^1 - \omega_{qq}^1},$$

and hence

$$\frac{\omega_{pp}^1}{\lambda_p} = \frac{\omega_{qq}^1}{\lambda_q} = \frac{\omega_{rr}^1}{\lambda_r} = \varphi, \tag{4-33}$$

$$\frac{\omega_{uu}^1}{\lambda_u} = \frac{\omega_{vv}^1}{\lambda_v} = \psi \tag{4-34}$$

for two functions φ and ψ.

Substituting (4-33) and (4-34) back to (4-29) gives

$$(1 + \varphi\psi)\lambda_p\lambda_u = -c,$$
$$(1 + \varphi\psi)\lambda_p\lambda_v = -c,$$

which imply that $\lambda_u = \lambda_v$. This is impossible.

If $\lambda_p = 0$, then (3-12) and (4-29) imply that $\omega_{pp}^1 = 0$ and $c = 0$. Then (4-30) and (4-31) yield

$$\frac{\omega_{uu}^1}{\lambda_u} = \frac{\omega_{vv}^1}{\lambda_v} = \gamma \tag{4-35}$$

for some function γ. However, combining (4-35) with (4-32) gives $\gamma^2 = -1$. Hence it is a contradiction.

Lastly, we consider $\lambda_u = 0$. Then (3-12) and (4-29) reduce to $\omega_{uu}^1 = c = 0$. The second equations of (4-29)–(4-31) show that

$$\frac{\omega_{pp}^1}{\lambda_p} = \frac{\omega_{qq}^1}{\lambda_q} = \frac{\omega_{rr}^1}{\lambda_r} = \varphi, \tag{4-36}$$

$$\frac{\omega_{vv}^1}{\lambda_v} = -\frac{1}{\varphi}. \tag{4-37}$$

By taking into account (4-36) and (4-37) together with (3-11) and (4-1), a very similar and direct computation as in Case A also gives a polynomial concerning the function λ_1 with constant coefficients. Hence, this is a contradiction and the mean curvature H must be constant.

In conclusion, the proof of Theorem 1.1 is completed.

Acknowledgements

The authors would like to thank Professor Cezar Oniciuc for fruitful discussions and useful suggestions. The authors also wish to express appreciation to the anonymous referee for helpful suggestions to improve the original version of this paper. The first author was supported by NSFC (No. 11601068), CSC of China (No. 201508210004) and China Postdoctoral Science Foundation (No. 2016T90226, No. 2014M560216). The second author was supported by the Australian Research Council (DP150101275).

References

[Akutagawa and Maeta 2013] K. Akutagawa and S. Maeta, "Biharmonic properly immersed submanifolds in Euclidean spaces", *Geom. Dedicata* **164** (2013), 351–355. MR Zbl

[Alías et al. 2013] L. J. Alías, S. C. García-Martínez, and M. Rigoli, "Biharmonic hypersurfaces in complete Riemannian manifolds", *Pacific J. Math.* **263**:1 (2013), 1–12. MR Zbl

[Balmuş et al. 2008] A. Balmuş, S. Montaldo, and C. Oniciuc, "Classification results for biharmonic submanifolds in spheres", *Israel J. Math.* **168** (2008), 201–220. MR Zbl

[Balmuş et al. 2010] A. Balmuş, S. Montaldo, and C. Oniciuc, "Biharmonic hypersurfaces in 4-dimensional space forms", *Math. Nachr.* **283**:12 (2010), 1696–1705. MR Zbl

[Balmuş et al. 2013] A. Balmuş, S. Montaldo, and C. Oniciuc, "Biharmonic PNMC submanifolds in spheres", *Ark. Mat.* **51**:2 (2013), 197–221. MR Zbl

[Caddeo et al. 2001] R. Caddeo, S. Montaldo, and C. Oniciuc, "Biharmonic submanifolds of S^3", *Internat. J. Math.* **12**:8 (2001), 867–876. MR Zbl

[Caddeo et al. 2002] R. Caddeo, S. Montaldo, and C. Oniciuc, "Biharmonic submanifolds in spheres", *Israel J. Math.* **130** (2002), 109–123. MR Zbl

[Chen 1991] B.-Y. Chen, "Some open problems and conjectures on submanifolds of finite type", *Soochow J. Math.* **17**:2 (1991), 169–188. MR Zbl

[Chen 2014] B.-Y. Chen, "Some open problems and conjectures on submanifolds of finite type: recent development", *Tamkang J. Math.* **45**:1 (2014), 87–108. MR Zbl

[Chen 2015] B.-Y. Chen, *Total mean curvature and submanifolds of finite type*, 2nd ed., Series in Pure Mathematics **27**, World Scientific, Hackensack, NJ, 2015. MR Zbl

[Chen and Munteanu 2013] B.-Y. Chen and M. I. Munteanu, "Biharmonic ideal hypersurfaces in Euclidean spaces", *Differential Geom. Appl.* **31**:1 (2013), 1–16. MR Zbl

[Cheng and Yau 1977] S. Y. Cheng and S. T. Yau, "Hypersurfaces with constant scalar curvature", *Math. Ann.* **225**:3 (1977), 195–204. MR Zbl

[Defever 1998] F. Defever, "Hypersurfaces of $\mathbf{E}^4$ with harmonic mean curvature vector", *Math. Nachr.* **196** (1998), 61–69. MR Zbl

[Dimitrić 1992] I. Dimitrić, "Submanifolds of E^m with harmonic mean curvature vector", *Bull. Inst. Math. Acad. Sinica* **20**:1 (1992), 53–65. MR Zbl

[Eells and Lemaire 1983] J. Eells and L. Lemaire, *Selected topics in harmonic maps*, CBMS Regional Conference Series in Mathematics **50**, American Mathematical Society, Providence, RI, 1983. MR Zbl

[Fu 2015a] Y. Fu, "Biharmonic hypersurfaces with three distinct principal curvatures in Euclidean space", *Tohoku Math. J.* (2) **67**:3 (2015), 465–479. MR Zbl

[Fu 2015b] Y. Fu, "Biharmonic hypersurfaces with three distinct principal curvatures in spheres", *Math. Nachr.* **288**:7 (2015), 763–774. MR Zbl

[Fu 2015c] Y. Fu, "On biharmonic hypersurfaces with constant scalar curvatures in $\mathbb{S}^5$", *Proc. Amer. Math. Soc.* **143**:12 (2015), 5399–5409. MR Zbl

[Hasanis and Vlachos 1995] T. Hasanis and T. Vlachos, "Hypersurfaces in E^4 with harmonic mean curvature vector field", *Math. Nachr.* **172** (1995), 145–169. MR Zbl

[Ichiyama et al. 2010] T. Ichiyama, J. Inoguchi, and H. Urakawa, "Classifications and isolation phenomena of bi-harmonic maps and bi-Yang-Mills fields", *Note Mat.* **30**:2 (2010), 15–48. MR Zbl

[Jiang 1986] G. Y. Jiang, "2-harmonic maps and their first and second variational formulas", *Chinese Ann. Math. Ser. A* **7**:4 (1986), 389–402. In Chinese; English summary in Chinese Ann. Math. Ser. B **7**:4 (1986), 523. MR Zbl

[Jiang 1987] G. Y. Jiang, "Some nonexistence theorems on 2-harmonic and isometric immersions in Euclidean space", *Chinese Ann. Math. Ser. A* **8**:3 (1987), 377–383. In Chinese; English summary in *Chinese Ann. Math. Ser. B* **8**:3 (1987), 389. MR Zbl

[Koiso and Urakawa 2014] N. Koiso and H. Urakawa, "Biharmonic submanifolds in a Riemannian manifold", preprint, 2014. arXiv

[Loubeau and Oniciuc 2014] E. Loubeau and C. Oniciuc, "Biharmonic surfaces of constant mean curvature", *Pacific J. Math.* **271**:1 (2014), 213–230. MR Zbl

[Luo 2014] Y. Luo, "Weakly convex biharmonic hypersurfaces in nonpositive curvature space forms are minimal", *Results Math.* **65**:1-2 (2014), 49–56. MR Zbl

[Montaldo and Oniciuc 2006] S. Montaldo and C. Oniciuc, "A short survey on biharmonic maps between Riemannian manifolds", *Rev. Un. Mat. Argentina* **47**:2 (2006), 1–22. MR Zbl

[Montaldo et al. 2016] S. Montaldo, C. Oniciuc, and A. Ratto, "On cohomogeneity one biharmonic hypersurfaces into the Euclidean space", *J. Geom. Phys.* **106** (2016), 305–313. MR Zbl

[Nakauchi and Urakawa 2011] N. Nakauchi and H. Urakawa, "Biharmonic hypersurfaces in a Riemannian manifold with non-positive Ricci curvature", *Ann. Global Anal. Geom.* **40**:2 (2011), 125–131. MR Zbl

[Oniciuc 2002] C. Oniciuc, "Biharmonic maps between Riemannian manifolds", *An. Ştiinţ. Univ. Al. I. Cuza Iaşi. Mat. (N.S.)* **48**:2 (2002), 237–248. MR Zbl

[Oniciuc 2012] C. Oniciuc, *Biharmonic submanifolds in space forms*, Habilitation Thesis, Universitatea Alexandra Ioan Cuza, 2012.

[Ou 2012] Y.-L. Ou, "Some constructions of biharmonic maps and Chen's conjecture on biharmonic hypersurfaces", *J. Geom. Phys.* **62**:4 (2012), 751–762. MR Zbl

[Ou 2016] Y.-L. Ou, "Some recent progress of biharmonic submanifolds", pp. 127–139 in *Recent advances in the geometry of submanifolds—dedicated to the memory of Franki Dillen (1963–2013)*, edited by B. Suceavă et al., Contemp. Math. **674**, Amer. Math. Soc, Providence, RI, 2016. MR Zbl

[Ou and Tang 2012] Y.-L. Ou and L. Tang, "On the generalized Chen's conjecture on biharmonic submanifolds", *Michigan Math. J.* **61**:3 (2012), 531–542. MR Zbl

Received June 9, 2016. Revised February 7, 2017.

YU FU
SCHOOL OF MATHEMATICS
DONGBEI UNIVERSITY OF FINANCE AND ECONOMICS
DALIAN
CHINA
yufu@dufe.edu.cn

MIN-CHUN HONG
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF QUEENSLAND
BRISBANE
AUSTRALIA
hong@maths.uq.edu.au

PACIFIC JOURNAL OF MATHEMATICS
Vol. 294, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.294.351

NONABELIAN FOURIER TRANSFORMS FOR SPHERICAL REPRESENTATIONS

JAYCE R. GETZ

Braverman and Kazhdan have introduced an influential conjecture on local functional equations for general Langlands L-functions. It is related to L. Lafforgue's equally influential conjectural construction of kernels for functorial transfers. We formulate and prove a version of Braverman and Kazhdan's conjecture for spherical representations over an archimedean field that is suitable for application to the trace formula. We then give a global application related to Langlands' beyond endoscopy proposal. It is motivated by Ngô's suggestion that one combine nonabelian Fourier transforms with the trace formula in order to prove the functional equations of Langlands L-functions in general.

1. Introduction

Let G be a split connected reductive group over an archimedean field F and let

$$r : \widehat{G} \to \mathrm{GL}_n \tag{1-1}$$

be a representation of (the connected component of) its Langlands dual group, which we regard as a connected reductive group over $\mathbb{C}$. For simplicity we assume that the neutral component of the kernel of r is trivial (this is the most interesting case anyway). The local Langlands correspondence is a theorem of Langlands in the archimedean case. Thus for every irreducible admissible representation π of $G(F)$ one has an irreducible admissible representation $r(\pi)$ of $\mathrm{GL}_n(F)$ that is the transfer of (the L-packet of) π. One defines

$$\gamma(s, \pi, r, \psi) := \gamma(s, r(\pi), \psi) := \frac{\varepsilon(s, r(\pi), \psi)L(1-s, r(\pi)^\vee)}{L(s, r(\pi))}$$

for $s \in \mathbb{C}$ and for any additive character $\psi : F \to \mathbb{C}^\times$, where the ε-factor on the right is that defined by Godement and Jacquet [1972]. Let $f \in C_c^\infty(G(F))$. In the case

The author is thankful for partial support provided by NSF grant DMS 1405708. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.
MSC2010: primary 11F70; secondary 11F66, 22E45.
Keywords: nonabelian Fourier transform, L-functions, beyond endoscopy.

where r is the standard representation of GL_n and the representation π is unitary one has an identity of operators

$$\gamma(s,\pi,r,\psi)\pi\,|\det|^{(n-1)/2+s}(f)=\pi^{\mathrm{anti}}|\det|^{(n-1)/2+1-s}(\hat{f}), \tag{1-2}$$

where $\pi^{\mathrm{anti}}(g):=\pi(g^{-1})$ and $\hat{f}$ is the restriction to $\mathrm{GL}_n(F)$ of the $\mathfrak{gl}_n(F)$-Fourier transform of f determined by ψ; see [Godement and Jacquet 1972, (9.5)]. Strictly speaking, that book assumes f to be in a space of Gaussian functions, but there is no need to make this precise here.

A conjecture of Braverman and Kazhdan [2000] states that (1-2) is but the first case of a general phenomenon. To be more precise we need to place an additional assumption on the representation r. Assume that there is a character

$$\omega : G \to \mathbb{G}_m \tag{1-3}$$

such that, if we denote by $\omega^\vee$ the dual cocharacter of the center $Z_{\widehat{G}}$ of $\widehat{G}$, then

$$r\circ\omega^\vee=[N],$$

where $[N]:\mathbb{G}_m\to\mathrm{GL}_n$ is the cocharacter given on points by

$$x\mapsto x^N I_n.$$

Here I_n is the $n\times n$ identity matrix. For complex numbers s, we let

$$\omega_s:=|\omega|^{s/N}\quad\text{and}\quad\pi_s:=\pi\otimes\omega_s.$$

The quasicharacter ω_s plays the role of $|\det|^s$ in the case considered by Godement and Jacquet [1972].

Temporarily let F be an arbitrary local field. Braverman and Kazhdan gave a conjectural construction of a nonabelian Fourier transform,

$$\mathcal{F}_{r,\psi}:C_c^\infty(G(F))\to C^\infty(G(F)), \tag{1-4}$$

such that

$$\gamma(s,\pi,r,\psi)\pi_s(f)=\pi_{1-s}^{\mathrm{anti}}(\mathcal{F}_{r,\psi}(f)). \tag{1-5}$$

In the case where r is the standard representation of $G=\mathrm{GL}_n$ one can take

$$\mathcal{F}_{r,\psi}(f)=|\det|^{(n-1)/2}(|\det|^{(1-n)/2}f)^\wedge.$$

In the nonarchimedean case, L. Lafforgue has given a spectral definition of $\mathcal{F}_{r,\psi}$ using Paley–Weiner theory under suitable assumptions that are implied by the local Langlands correspondence for $G(F)$ [Lafforgue 2014, Définition II.15]. The analytic properties of $\mathcal{F}_{r,\psi}(f)$ (e.g., whether or not it is integrable after a suitable twist by a quasicharacter of $G(F)$) are not obvious from his construction.

Assume again that F is archimedean, and let $K\le G(F)$ be a maximal compact subgroup. In this paper we prove the existence of a transform $\mathcal{F}_{r,\psi}(f)$ such that

(1-5) holds for unitary representations π provided that f is spherical and lies in a naturally defined subspace $C_c^\infty(G(F)//K, r)$ of $C_c^\infty(G(F)//K)$ depending on r. We also prove that suitable twists of $\mathcal{F}_{r,\psi}(f)$ by quasicharacters lie in a space of functions for which the trace formula is valid.

In Section 4 we define the subspace $C_c^\infty(G(F)//K, r) \leq C_c^\infty(G(F)//K)$. For $0 < p \leq 2$ let

$$\mathcal{S}^p(G(F)//K) \leq C^\infty(G(F)//K)$$

be the L^p-Harish-Chandra–Schwartz space; we recall its definition in Section 3C. The following is the main theorem of this paper:

Theorem 1.1. *Let $f \in C_c^\infty(G(F)//K, r)$ and $0 < p \leq 1$. There is a function,*

$$\mathcal{F}_{r,\psi}(f) \in C^\infty(G(F)//K),$$

such that

(a) *one has $\mathcal{F}_{r,\psi}(f)\omega_s \in \mathcal{S}^p(G(F)//K)$ for* $\mathrm{Re}(s)$ *sufficiently large in a sense depending only on G and r, and*

(b) *if π is unitary and irreducible then*

$$\gamma(s, \pi, r, \psi)\,\mathrm{tr}\,\pi_s(f) = \mathrm{tr}\,\pi^\vee_{1-s}(\mathcal{F}_{r,\psi}(f)) \tag{1-6}$$

in the sense of analytic continuation.

Remarks. (1) A more precise version of (a) is proved in Theorem 4.1.

(2) Elements of $\mathcal{S}^p(G(F)//K)$ for $0 < p \leq 1$ are in $L^1(G(F)//K)$, so the fact that π is unitary implies that the operator $\pi_s^\vee(\mathcal{F}_{r,\psi}(f))$ is bounded for $\mathrm{Re}(s)$ sufficiently large.

(3) The functions f and $\mathcal{F}_{r,\psi}(f)$ are assumed to be spherical, so for π unitary and nonspherical (1-6) is just the equality $0 = 0$.

(4) The operator $\pi_s(f)$ is holomorphic as a function of s, and $\gamma(s, \pi, r, \psi)$ is meromorphic. Thus (1-6) provides a meromorphic continuation of $\mathrm{tr}\,\pi^\vee_{1-s}(\mathcal{F}_{r,\psi}(f))$ to the complex plane.

Assertion (a) in the theorem is important because the Arthur–Selberg trace formula is valid for functions in (the global version of) $\mathcal{S}^p(G(F)//K)$ for $0 < p \leq 1$ due to work of Finis, Lapid, and Müller [Finis et al. 2011; Finis and Lapid 2011; 2016]. One can then use their results to provide an absolutely convergent expression for the sum of residues of Langlands L-functions that Langlands has isolated for study in his "beyond endoscopy" proposal. This will be discussed in Section 5.

It would be very interesting to extend the results in this paper to nonspherical representations. Our approach might be applicable in this more general setting provided one can prove a certain analytic result. More precisely, we use the characterization of the image of $\mathcal{S}^p(G(F)//K)$ under the Fourier transform due to Trombi and Varadarajan [1971]. A characterization of the image of the nonspherical analogue $\mathcal{S}^p(G(F))$ under the Fourier transform does not seem to be available in the literature, although partial results are known [1981; 1990]. Hopefully the conjectures of Braverman and Kazhdan [2000], Lafforgue [2014], and of course the beyond endoscopy proposal of Langlands [2004] will provide motivation for giving such a characterization.

Remark. Arthur [1983] has characterized the image of the Fourier transform of $\mathcal{S}^2(G(F))$, but this does not seem to be the right space from the point of view of any of these proposals to move beyond endoscopy. It would be interesting to see if the technique of Anker [1991] could be used to deduce the image of the Fourier transform on $\mathcal{S}^p(G(F))$ for $0 < p < 2$ from this case.

We would be remiss not to recall that the fundamental aim of [Braverman and Kazhdan 2000; Lafforgue 2014] is to provide a definition of the nonabelian Fourier transform for which a version of Poisson summation is valid. As explained in these papers, this would lead to a proof of the functional equation and meromorphic continuation of the L-functions attached to r by Langlands. For this purpose it would probably be desirable to have a definition of the Fourier transform $\mathcal{F}_{r,\psi}(f)$, which, unlike the approach of this paper and [Lafforgue 2014], does not rely on Paley–Wiener theorems. For this we can only point to the hints provided in the works [Altuğ 2015; Bouthier et al. 2016; Cheng and Ngô 2017; Cheng 2014; Frenkel et al. 2010; Getz 2016; 2014; Getz and Herman 2015; Getz and Liu 2017; Li 2017; Langlands 2013; Ngô 2014; Sakellaridis 2012].

We close the introduction by outlining the sections of this paper. In Section 2 we recall the notion of the transfer of a spherical representation. Preliminaries on the characterization of the image of the spherical Fourier transform due to Trombi and Varadarajan are given in Section 3C and in Section 4 we prove Theorem 4.1, which immediately implies our main result, Theorem 1.1. Finally, in Section 5 we give an application of the main theorem to Langlands' beyond endoscopy proposal.

2. Tori and transfers of representations

Let F be a local field. In this section we set some notation and recall the notion of the transfer $r(\pi)$ of an irreducible admissible representation π of $G(F)$ under some simplifying assumptions. These assumptions are always true if π is spherical. Before this we give a definition, from [Cheng and Ngô 2017], of a useful extension W' of the Weyl group W of a split maximal torus $T \leq G$ by a subgroup of $\mathfrak{S}_n$, the symmetric group on n letters. The whole point of the discussion below is to explain how r induces a W'-equivariant map,

$$r^\vee : T_n \to T,$$

where T_n is a maximal torus in GL_n.

Let $T \leq G$ be a split maximal torus with Weyl group W. Moreover let $\widehat{T} \leq \widehat{G}$ be the dual torus; its Weyl group in $\widehat{G}$ is isomorphic to W and we denote it by the same letter. Let V_r be the space of r. We can decompose

$$V_r = \oplus_{i=1}^{m} V_{\lambda_i}, \tag{2-1}$$

where the sum is over the nonzero weights $\lambda_1, \ldots, \lambda_m \in X^*(\widehat{T})$ of $\widehat{T}$ in V_r, V_{λ_i} is the λ_i weight space, and $\dim V_{\lambda_i} = d_i$. Fix a basis $A_i \subset V_r$ for each V_{λ_i} (viewed as a subspace of V_r) and let $A = \coprod_{i=1}^{m} A_i$; this is a basis for V_r. This choice of basis gives an embedding $\mathbb{G}_m^A \to \mathrm{GL}(V_r)$. We let $\widehat{T_n}$ be its image and let $\Lambda = X^*(\widehat{T_n}) = \mathbb{Z}[A]$. It comes equipped with a $\mathbb{Z}$-linear map,

$$\Lambda \to X^*(\widehat{T}) = X_*(T), \tag{2-2}$$

given by extending the set theoretic map sending each basis element in A_i to λ_i. Thus r induces a map,

$$r : \widehat{T} \to \widehat{T_n}, \tag{2-3}$$

where $\widehat{T_n} \leq \mathrm{GL}_n$ is the maximal torus with character group Λ. In fact, upon conjugating the representation r by an element of $\mathrm{GL}_n(\mathbb{C})$ we can and do assume that $\widehat{T_n}$ is the standard maximal torus of diagonal matrices. For F-algebras R we take

$$T_n(R) = \Lambda \otimes R^\times.$$

It is a torus over F with dual torus $\widehat{T_n}$, and by construction there is a morphism

$$r^\vee : T_n \to T \tag{2-4}$$

over F whose dual is r.

The Weyl group W_n of GL_n can be identified with the set of permutations of A (which we also identify with $\mathfrak{S}_n$) and we let

$$\Sigma_{\underline{\lambda}} = \mathfrak{S}_{r_1} \times \cdots \times \mathfrak{S}_{r_m} \leq \mathfrak{S}_n = W_n \tag{2-5}$$

denote the subgroup preserving the decomposition $A = \coprod_{i=1}^{m} A_i$ (i.e., those permutations σ such that $\sigma(V_{\lambda_i}) = V_{\lambda_i}$ for all i). We let

$$\Sigma'_{\underline{\lambda}} := \{\tau \in W_n : \text{ there exists } \xi \in \mathfrak{S}_m \text{ such that } \tau(A_i) = A_{\xi(i)} \text{ for all } 1 \leq i \leq m\}.$$

The map $\tau \mapsto \xi$ implicit in this definition is in fact a homomorphism $\Sigma'_{\underline{\lambda}} \to \mathfrak{S}_m$ whose image is the set of permutations fixing the multiplicity function $i \mapsto d_i$ and whose kernel is $\Sigma_{\underline{\lambda}}$. The Weyl group W acts on $X^*(\widehat{T})$ and this action preserves the weights $\lambda_1, \ldots, \lambda_m$ and the multiplicity function $i \mapsto d_i$. Thus the map $W \to \mathfrak{S}_m$ induces a morphism,

$$\rho_W : W \to \Sigma'_{\underline{\lambda}} / \Sigma_{\underline{\lambda}}$$

(this ρ_W has nothing to do with the sum of positive roots). We define W' to be the following extension of W:

$$W' := \{(w, \xi) \in W \times \Sigma'_{\underline{\lambda}} : \rho_W(w) \equiv \xi \pmod{\Sigma_{\underline{\lambda}}}\}. \tag{2-6}$$

The group W' admits natural homomorphisms to both $\Sigma'_{\underline{\lambda}}$ and W by projection to the two factors. We therefore obtain actions of W' via projection to $\Sigma'_{\underline{\lambda}}$ on Λ, and T_n and $\widehat{T}_n$ and actions of W' via projection to W on $X^*(\widehat{T})$, T and $\widehat{T}$. One checks that (2-2) is W' equivariant with respect to these actions, and hence so are the maps (2-3) and (2-4).

For unramified quasicharacters $\chi : T(F) \to \mathbb{C}^\times$ extend χ to a character of a Borel subgroup B containing T. Let $J(\chi)$ be the unique irreducible spherical subquotient of the unitarily normalized induction $\mathrm{Ind}_{B(F)}^{G(F)}(\chi)$ (see Theorem 3.1 for more details and references).

Suppose that $\pi = J(\chi)$ where $\chi : T(F) \to \mathbb{C}^\times$ is an unramified quasicharacter. One defines $r(\chi)$ to be $\chi \circ r^\vee$ and one defines the transfer of π to be

$$r(\pi) := J(r(\chi)).$$

This is an irreducible admissible representation of $\mathrm{GL}_n(F)$. We note that if

$$r(\chi)\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} = \prod_{i=1}^n \eta_i(a_i)$$

for some quasicharacters $\eta_i : F^\times \to \mathbb{C}^\times$ then

$$\gamma(s, \pi, r, \psi) := \gamma(s, r(\pi), \psi) = \prod_{i=1}^n \gamma(s, \eta_i, \psi);$$

see [Godement and Jacquet 1972, Corollaries 3.6 and 8.9].

3. Preliminaries on the Fourier transform

In this section we collect some notation related to Langlands decompositions of Borel subgroups and recall some basic facts about the spherical Fourier transform. All of these results will be used in Section 4. In this section F is an archimedean local field.

3A. ***Langlands decompositions.*** Let $T \leq G$ and $T_n \leq \mathrm{GL}_n$ be maximal tori as in Section 2 (so T_n is the diagonal torus, T is split and r maps $\widehat{T}$ into $\widehat{T}_n$). Let $B \geq T$ be a Borel subgroup of G and let $B_n \geq T_n$ be the Borel subgroup of upper triangular matrices in GL_n. We let $N \leq B$ and $N_n \leq B_n$ be the unipotent radicals. Let $K \leq G(F)$ be a maximal compact subgroup and let $K_n \leq \mathrm{GL}_n(F)$ be the standard maximal compact subgroup. Let

$$M := T(F) \cap K, \qquad M_n := T_n(F) \cap K_n.$$

We can and do assume that M is the maximal compact subgroup of $T(F)$, and then one has $r^\vee(M_n) \leq M$. Let $\mathfrak{a} = X_*(T) \otimes_{\mathbb{Z}} \mathbb{R}$, $\mathfrak{a}_n := X_*(T_n) \otimes_{\mathbb{Z}} \mathbb{R}$ and let

$$\mathfrak{a}^* = \mathrm{Hom}(T(F)/M, \mathbb{R}^\times_{>0}),$$
$$\mathfrak{a}^*_n = \mathrm{Hom}(T_n(F)/M_n, \mathbb{R}^\times_{>0})$$

be their $\mathbb{R}$-linear duals.

We require a norm on $\mathfrak{a}^*_{\mathbb{C}}$. To construct it, let $(\, , \,)$ be a nondegenerate symmetric bilinear form on $\mathfrak{g} := \mathrm{Lie}(\mathrm{Res}_{F/\mathbb{R}} G)$ whose restriction to the derived algebra is the Killing form. We assume that the $+1$ and -1 eigenspaces of the Cartan involution Θ attached to K are orthogonal under $(\, , \,)$ and that $X \mapsto -(X, \Theta X)$ is a positive definite quadratic form on $\mathfrak{g}$. We then set $\|X\|^2 = -(X, \Theta X)$. It induces a hermitian inner product on $\mathfrak{g}_{\mathbb{C}}$ and $\mathfrak{a}_{\mathbb{C}}$ and we continue to denote by $\|\cdot\|$ the induced form on $\mathfrak{a}^*_{\mathbb{C}}$.

The map $r^\vee$ yields

$$r : \mathfrak{a}^*_{\mathbb{C}} \to \mathfrak{a}^*_{n\mathbb{C}}.$$

One has an isomorphism

$$\mathbb{R}^n \xrightarrow{\sim} \mathfrak{a}^*_n, \qquad (s_1, \ldots, s_n) \mapsto \eta_{s_1,\ldots,s_n},$$

where

$$\eta_{s_1,\ldots,s_n}\begin{pmatrix} t_1 & & \\ & \ddots & \\ & & t_n \end{pmatrix} := |t_1|^{s_1} \cdots |t_n|^{s_n}.$$

We let

$$(\mathfrak{a}^*_n)_+$$

be the image of $\mathbb{R}^n_{>0}$ and let

$$\mathfrak{a}^*_+ := \{\lambda \in \mathfrak{a}^* : r(\lambda) \in (\mathfrak{a}^*_n)_+\}. \tag{3-1-1}$$

By the existence of ω (see (1-3)) this is nonempty. Note that this is not a Weyl chamber.

3B. ***Spherical functions.*** We now recall the definition of the Harish-Chandra map. We define a function $H_T : T(F)/M \to \mathrm{Hom}_{\mathbb{Z}}(X^*(T), \mathbb{R})$ via

$$\langle \chi, H_T(x) \rangle := \log|\chi(x)|.$$

Since there is a canonical identification $\mathrm{Hom}_{\mathbb{Z}}(X^*(T), \mathbb{R}) = X_*(T) \otimes_{\mathbb{Z}} \mathbb{R} =: \mathfrak{a}$ we can regard H_T as taking values in $\mathfrak{a}$. For $(k, t, n) \in K \times T(F) \times N(F)$, the Harish-Chandra map is then defined to be

$$H_B : G(F) \to \mathfrak{a}, \qquad ktn \mapsto H_T(t).$$

We choose Haar measures dk, dt, dn, dg on K, $T(F)$, $N(F)$, and $G(F)$, respectively, such that $\mathrm{meas}_{dk}(K) = 1$ and for $f \in C_c^\infty(G(F))$,

$$\int_{G(F)} f(g)\, dg = \int_{K \times T(F) \times N(F)} e^{\langle 2\rho, H_B(t) \rangle} f(ktn)\, dk\, dt\, dn,$$

where $\rho \in \mathfrak{a}^*$ is half the sum of the positive roots of T in B.

For $\lambda \in \mathfrak{a}^*_{\mathbb{C}}$ we let

$$\varphi_\lambda(g) = \int_K e^{\langle -(\lambda+\rho), H_B(g^{-1}k)\rangle}\, dk \tag{3-2-1}$$

be the usual spherical function. It is a matrix coefficient of the representation $\mathrm{Ind}(e^{\langle \lambda, H_B\rangle})$. One defines the spherical transform of suitable K-biinvariant continuous functions $f : G(F) \to \mathbb{C}$ to be

$$\tilde{f}(\lambda) := \int_{G(F)} f(g)\varphi_{-\lambda}(g)\, dg \tag{3-2-2}$$

(here we are using the convention of [Anker 1991, §1], at least up to multiplication by $\sqrt{-1}$).

3C. ***Spaces of functions.*** Let $\mathfrak{g} := \mathrm{Lie}(\mathrm{Res}_{F/\mathbb{R}}\, G)$. For $0 < p \leq 2$ let

$$\mathcal{S}^p(G(F)/\!/K) \tag{3-3-1}$$

denote the space of K-biinvariant functions $f : G(F) \to \mathbb{C}$ such that

$$\sup_{x \in G(F)} (|x|+1)^n \varphi_0(x)^{-2/p} |X * f * Y(x)| < \infty$$

for all $n \geq 0$ and all invariant differential operators X, Y on $G(F)$, that is, all elements of the universal enveloping algebra $U(\mathfrak{g}_{\mathbb{C}})$ of the complexification of $\mathfrak{g}$. Here φ_0 is the spherical function of (3-2-1) and $|x|$ is the distance of x to K (see, e.g., [Anker 1991, §1]).

It is known that for $0 < p < p' \leq 2$ there are continuous inclusions

$$\mathcal{S}^p(G(F)/\!/K) \leq \mathcal{S}^{p'}(G(F)/\!/K) \leq L^{p'}(G(F));$$

see [Anker 1991, §1]. To check the continuity of the last inclusion one uses the fact that $(\varphi_0^{2/p}(x))/(|x|+1)^n \in L^p(G(F))$ for n sufficiently large [Gangolli and Varadarajan 1988, Proposition 4.6.12].

3D. ***The Fourier transform.*** For $0 < p \leq 2$ let $\mathfrak{a}^*_p$ be the closed tube in $\mathfrak{a}^*_{\mathbb{C}}$ of points whose real part lies in the closed convex hull of

$$W \cdot \left(\tfrac{2}{p} - 1\right)\rho$$

in $\mathfrak{a}^*$. Here ρ denotes half the sum of the positive roots of T in B. Let

$$\mathcal{S}(\mathfrak{a}^*_p)$$

denote the space of all complex valued functions $h : \mathfrak{a}^*_p \to \mathbb{C}$ such that

(a) all derivatives of the function h exist and are continuous on $\mathfrak{a}^*_p$,

(b) the function h is holomorphic in the interior of $\mathfrak{a}^*_p$,

(c) for any polynomial P in the symmetric algebra of $\mathfrak{a}^*$ and any integer $n \geq 0$,

$$\sup_{\lambda \in \mathfrak{a}_p^*} (\|\lambda\| + 1)^n \left| P\Big(\frac{\partial}{\partial \lambda}\Big) h(\lambda) \right| < \infty.$$

Trombi and Varadarajan [1971] proved that the spherical Fourier transform $f \mapsto \tilde{f}(\lambda)$ extends to an isomorphism of Fréchet algebras

$$\mathcal{S}^p(G(F)//K) \xrightarrow{\sim} \mathcal{S}(\mathfrak{a}_p^*)^W. \tag{3-4-1}$$

The seminorms which make $\mathcal{S}^p(G(F)//K)$ and $\mathcal{S}(\mathfrak{a}_p^*)$ into Fréchet spaces are the obvious ones. A simpler proof of the isomorphism (3-4-1) is contained in [Anker 1991], which is a very nice reference for the facts above, though, strictly speaking, it assumes that G is semisimple. Therein, $\mathfrak{a}_p^*$ is denoted $i\mathfrak{a}_{2/p-1}^*$.

For $0 < p \leq 1$ and $f \in \mathcal{S}^p(G(F)//K)$, let

$$f^{(B)}(t) = e^{\langle \rho, H_B(t) \rangle} \int_{N(F)} f(tn)\, dn$$

be the constant term of f along B. It is absolutely convergent [Gangolli and Varadarajan 1988, Theorem 6.2.4].

Let $\chi : T(F) \to \mathbb{C}^\times$ be a character. If $\chi = e^{\langle \lambda, H_B \rangle}$ for some $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$ whose real part is in the closed Weyl chamber defined by the positive roots attached to B we will abbreviate

$$J(\lambda) := J\big(e^{\langle \lambda, H_B \rangle}\big) \quad \text{and} \quad \mathrm{Ind}(\lambda) := \mathrm{Ind}_{B(F)}^{G(F)}\big(e^{\langle \lambda, H_B \rangle}\big).$$

We recall the following special case of the Langlands classification:

Theorem 3.1. *Any irreducible admissible spherical representation of $G(F)$ is infinitesimally equivalent to $J(\lambda)$ for some $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$ whose real part is in the closed positive Weyl chamber attached to B. Conversely, for every $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$ whose real part is in the closed positive Weyl chamber with respect to B, the representation* $\mathrm{Ind}(\lambda)$ *has a unique irreducible quotient* (*usually called the Langlands quotient*). *It is spherical.*

Proof. In the notation of [Vogan 1981] and [Barbasch et al. 2008] take $\delta = \text{triv}$ and $\mu = \text{triv}$. In their terminology, δ is a fine $T(F) \cap K$-type, μ is a fine K-type and $\mu \in A(\delta)$. The stabilizer W_δ of δ under the natural action of the Weyl group W is all of W: $W_\delta = W$. Thus the first assertion of the lemma follows from [Barbasch et al. 2008, Theorem in §2.4].

On the other hand, since δ is trivial the R-group R_δ is trivial (see [Vogan 1981, Definition 4.3.13]), hence the set $A(\delta)$ consists of one element, namely the trivial K-type; see [Vogan 1981, Theorem 4.3.16]. In view of this, the final assertion of the theorem is contained in [Barbasch et al. 2008, §2.11]. □

We also give a proof of the following well-known result for the convenience of the reader.

Lemma 3.2. *Let $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$ have real part in the closed positive Weyl chamber with respect to B. Suppose that $J(\lambda)$ is unitary. If $f \in C_c^\infty(G(F)//K)$ then $J(\lambda)(f)$ acts via the scalar*

$$\tilde{f}(-\lambda) = \operatorname{tr} J(\lambda)(f) = \operatorname{tr} \operatorname{Ind}(\lambda)(f) = \operatorname{tr} e^{\langle \lambda, H_B \rangle}(f^{(B)})$$

on the (*unique*) *spherical vector in $J(\lambda)$. Under the same assumptions on $J(\lambda)$, if $0 < p \leq 1$ and $f \in \mathcal{S}^p(G(F)//K)$ then $J(\lambda)(f)$ acts via the scalar*

$$\tilde{f}(-\lambda) = \operatorname{tr} J(\lambda)(f)$$

on the spherical vector in $J(\lambda)$.

Proof. Assume first that $f \in C_c^\infty(G(F)//K)$. The identity

$$\operatorname{tr} \operatorname{Ind}(\lambda)(f) = \operatorname{tr} e^{\langle \lambda, H_B \rangle}(f^{(B)})$$

is the descent formula; see, e.g., [Knapp 1986, (10.23)]. The vector φ_λ is spherical, satisfies $\varphi_\lambda(1) = 1$, and is a matrix coefficient of the representation $\operatorname{Ind}(\lambda)$ (compare [Anker 1991, §1]). It is also known that this representation, even if it is reducible, contains a unique spherical line [Gangolli and Varadarajan 1988, 3.1.13]. It follows that $\operatorname{Ind}(\lambda)(f)$ acts via the scalar $\tilde{f}(-\lambda) = \operatorname{tr} \operatorname{Ind}(\lambda)(f)$ on this spherical line. On the other hand, one has a nonzero equivariant map

$$\operatorname{Ind}(\lambda) \to J(\lambda). \tag{3-4-2}$$

Since the irreducible representation $J(\lambda)$, being spherical, has a unique spherical line this line must be the image of the unique spherical line in $\operatorname{Ind}(\lambda)$ under (3-4-2). Thus $\operatorname{tr} J(\lambda)(f) = \operatorname{tr} \operatorname{Ind}(\lambda)(f)$.

In the following discussion we use basic facts recalled in [Anker 1991, §1] without further comment. Let $0 < p \leq 1$ and $f \in \mathcal{S}^p(G(F)//K)$. Then $\tilde{f}$ is defined on $\lambda \in \mathfrak{a}_1^*$. Since $C_c^\infty(G(F)//K)$ is dense in $\mathcal{S}^p(G(F)//K)$ we can choose a Cauchy sequence

$$\{f_n\}_{n=1}^\infty \subset C_c^\infty(G(F)//K)$$

converging to f in $\mathcal{S}^p(G(F)//K)$. In particular, $\lim_{n\to\infty} \tilde{f}_n = \tilde{f}$ pointwise on $\mathfrak{a}_1^*$.

Now the inclusion

$$\mathcal{S}^p(G(F)//K) \to L^1(G(F)//K)$$

is continuous. Since $J(\lambda)$ is unitary, the fact that $\lim_{n\to\infty} f_n = f$ in $L^1(G(F))$ implies that

$$\lim_{n\to\infty} \operatorname{tr} J(\lambda)(f_n) = \operatorname{tr} J(\lambda)(f).$$

Thus

$$\tilde{f}(-\lambda) = \lim_{n\to\infty} \tilde{f}_n(-\lambda) = \lim_{n\to\infty} \operatorname{tr} J(\lambda)(f_n) = \operatorname{tr} J(\lambda)(f). \qquad \square$$

4. Proof of Theorem 1.1

Recall that we have normalized $r:\widehat{G}\to \mathrm{GL}_n$ so that it induces a morphism $r:\widehat{T}\to\widehat{T}_n$ and used it to construct a dual morphism,

$$r^\vee : T_n \to T.$$

There are natural isomorphisms

$$T_n(F)/M_n \cong X_*(T_n)\otimes_{\mathbb{Z}}\mathbb{R} =: \mathfrak{a}_n,$$

and $T(F)/M \cong \mathfrak{a}$, so $r^\vee$ induces an $\mathbb{R}$-linear map,

$$r^\vee : \mathfrak{a}_n := T_n(F)/M_n \to T(F)/M =: \mathfrak{a}.$$

It is surjective, as the complexification of its dual is

$$r : \mathfrak{a}^*_{\mathbb{C}} \to \mathfrak{a}^*_{n\mathbb{C}},$$

which is injective since we assumed r has zero-dimensional kernel. Recall the group W' of Section 2. It acts naturally on $\mathfrak{a}$, $\mathfrak{a}_n$, $\mathfrak{a}^*_{\mathbb{C}}$, $\mathfrak{a}^*_{n\mathbb{C}}$ and the maps r and $r^\vee$ are W' equivariant (see Section 2). We therefore obtain a push-forward map,

$$r^\vee_* : C_c^\infty(T_n(F)/M_n)^{W'} \to C_c^\infty(T(F)/M)^{W'} = C_c^\infty(T(F)/M)^{W}. \tag{4-1}$$

We define

$$C_c^\infty(G(F)//K, r) := \left\{ f\in C_c^\infty(G(F)//K) : f^{(B)} \in r^\vee_*(C_c^\infty(T_n(F)/M_n)^{W'})\right\}. \tag{4-2}$$

Let $d_\omega\in\mathfrak{a}^*$ be the point corresponding to $\omega^\vee$. In this section we prove Theorem 4.1. It obviously implies Theorem 1.1; it is simply a version of Theorem 1.1 that makes explicit how large $\mathrm{Re}(s)$ must be in terms of the representation r.

Theorem 4.1. *Let $f\in C_c^\infty(G(F)//K, r)$ and $0<p\le 1$. There is a function,*

$$\mathcal{F}_{r,\psi}(f)\in C^\infty(G(F)//K),$$

such that

(a) *$\mathcal{F}_{r,\psi}(f)\omega_s \in \mathcal{S}^p(G(F)//K)$ provided that $\mathrm{Re}(\mathfrak{a}^*_p)+\mathrm{Re}(s)d_\omega \subset \mathfrak{a}^*_+$, and*

(b) *if π is unitary and irreducible then*

$$\gamma(s,\pi,r,\psi)\,\mathrm{tr}\,\pi_s(f) = \mathrm{tr}\,\pi^\vee_{1-s}(\mathcal{F}_{r,\psi}(f)) \tag{4-3}$$

in the sense of analytic continuation.

Now we give a diagram that outlines the construction of $\mathcal{F}_{r,\psi}(f)$ from f. Let $s_0>0$ be large enough that $\mathrm{Re}(\mathfrak{a}^*_p)+s_0 d_\omega\subset\mathfrak{a}^*_+$. Let $\mathcal{S}(\mathfrak{a}^*_p)(-s_0)$ (respectively, $\mathcal{S}(\mathfrak{a}^*_p)(-s_0)^W$) denote the space of functions of the form

$$\mathfrak{a}^*_p + s_0 d_\omega \to \mathbb{C}, \qquad \lambda\mapsto \tilde{f}(\lambda - s_0 d_\omega)$$

for some $\tilde{f} \in \mathcal{S}(\mathfrak{a}_p^*)$ (respectively, $\mathcal{S}(\mathfrak{a}_p^*)^W$). We note that ω is fixed by W, so this space admits an action of W, and $\mathcal{S}(\mathfrak{a}_p^*)^W(-s_0) = \mathcal{S}(\mathfrak{a}_p^*)(-s_0)^W$. Moreover, (3-4-1) induces an isomorphism

$$\mathcal{S}^p(G(F)//K)\omega_{-s_0} \xrightarrow{\sim} \mathcal{S}(\mathfrak{a}_p^*)(-s_0)^W, \qquad f \longmapsto\!\!\xrightarrow{\sim} 1pt\,\tilde{f}(\lambda),$$

where the $\mathbb{C}$-vector space on the left is the space of functions of the form $f\omega_{-s_0}$ for $f \in \mathcal{S}^p(G(F)//K)$. We will construct $\mathcal{F}_{r,\psi}(f)$ so that the following diagram commutes:

$$\begin{tikzcd}
C_c^\infty(T_n(F)/M_n)^{W'} \arrow[r, "\widehat{\ }"] \arrow[d, "r_*^\vee"] & \mathcal{S}(\mathfrak{t}_n(F)/M_n)^{W'} \arrow[d, "\Psi \mapsto \widetilde{\Psi}\circ r"] \\
C_c^\infty(T(F)/M)^W & \mathcal{S}(\mathfrak{a}_p^*)(-s_0)^W \\
C_c^\infty(G(F)//K) \arrow[u, "f \mapsto f^{(B)}"] \arrow[r, "\mathcal{F}_{r,\psi}"] & \mathcal{S}^p(G(F)//K)\omega_{-s_0} \arrow[u, "f \mapsto \tilde{f}(-\lambda)"']
\end{tikzcd}$$

Here the horizontal arrow marked $\widehat{\ }$ is (4-7), which is just the Fourier transform on $T_n(F) \subset \mathfrak{t}_n(F)$ (the F-points of the Lie algebra of T_n). Moreover $\widetilde{\Psi}$ is a Mellin transform. The proof of Theorem 4.1 we now give amounts to filling in the details of this diagram. The functional equation ultimately reduces to the familiar functional equation of Tate zeta functions.

Before beginning the proof in earnest let us describe the map $\Psi \mapsto \widetilde{\Psi} \circ r$ more precisely and show that it is well defined. Let $\mathfrak{t}_n := \mathrm{Lie}(T_n)$ and denote by $\mathcal{S}(\mathfrak{t}_n(F))$ the usual Schwartz space. Assume that

$$\Psi \in \mathcal{S}(\mathfrak{t}_n(F)).$$

We then have that

$$\widetilde{\Psi}(\lambda) := \int_{T_n(F)} e^{\langle \lambda, H_{T_n}(t)\rangle}\Psi(t)\,dt \tag{4-4}$$

is absolutely convergent and holomorphic if $\mathrm{Re}(\lambda) \in (\mathfrak{a}_n^*)_+$. Here, as above, dt is the Haar measure on $T_n(F)$. In fact something stronger is true:

Lemma 4.2. *For any polynomial P in the symmetric algebra of $\mathfrak{a}_n^*$ and any integer $n \geq 0$ the quantity*

$$(\|\lambda\|+1)^n \left| P\Big(\frac{\partial}{\partial\lambda}\Big)\widetilde{\Psi}(\lambda)\right|$$

is bounded for $\mathrm{Re}(\lambda)$ in a fixed compact subset of $(\mathfrak{a}_n^*)_+$. Here $\|\cdot\|$ is any Hermitian inner product on $\mathfrak{a}_{n\mathbb{C}}^*$.

Proof. The lemma is a consequence of the following claim: for any $f \in \mathcal{S}(\mathbb{R})$, real numbers $0 < a < b$, and nonnegative integers n, k one has

$$|s|^n \left|\frac{d^k}{ds^k}\tilde{f}(s)\right| \ll_{f,a,b,n,k} 1$$

provided that $a \le \mathrm{Re}(s) \le b$. Here $\tilde{f}(s) = \int_0^\infty f(x)x^{s-1}\,dx$ is the usual Mellin transform. To check this, let $D := x\frac{d}{dx}$. It is not difficult to see that $x^\varepsilon D^n(\log^k f)$ is continuous on $\mathbb{R}_{>0}$ and in $L^1(\mathbb{R}_{>0}, \frac{dx}{x})$ for any real number $\varepsilon > 0$, and thus the Mellin transform $(D^n(\log^k f))^\sim(\sigma + it)$ is bounded as a function of $t \in \mathbb{R}$ for $a \le \sigma \le b$. Thus

$$\left| s^n \frac{d^k}{ds^k} \tilde{f}(s) \right| = |(D^n(\log^k f))^\sim(s)|$$

is bounded for $a \le \mathrm{Re}(s) \le b$. □

Corollary 4.3. *Suppose that $s_0 \in \mathbb{R}_{>0}$ is large enough that $\mathfrak{a}_p^* + s_0 d_\omega \subset \mathfrak{a}_+^*$ and that $\Psi \in \mathcal{S}(\mathfrak{t}_n(F)/M_n)$. Then the function*

$$\lambda \mapsto \widetilde{\Psi}(r(\lambda))$$

is in $\mathcal{S}(\mathfrak{a}_p^*)(-s_0)$.

Proof. By assumption and the definition (3-1-1) of $\mathfrak{a}^+$ one has $r(\mathfrak{a}_p^* + s_0 d_\omega) \subset (\mathfrak{a}_n^*)_+$. Therefore the corollary follows from Lemma 4.2 and the injectivity of the map $r : \mathfrak{a}_{\mathbb{C}}^* \to \mathfrak{a}_{n\mathbb{C}}^*$. □

The corollary implies that the map $\Psi \mapsto \widetilde{\Psi} \circ r$ in the top right of the diagram is well defined for s_0 large enough.

We now begin the proof of Theorem 4.1. Let π be a given spherical unitary irreducible representation of $G(F)$. Thus there is a quasicharacter

$$\chi : T(F)/M \to \mathbb{C}^\times$$

so that $\pi \cong J(\chi)$. Let $f \in C_c^\infty(G(F)//K, r)$. We will trace $J(\chi)$ and f along the upper path from $C_c^\infty(G(F)//K)$ to $\mathcal{S}^p(G(F)//K)\omega_{-s_0}$ in the diagram. Lemma 3.2 implies that

$$\mathrm{tr}\, J(\chi)(f) = \chi(f^{(B)}). \tag{4-5}$$

We choose a $\Phi \in C_c^\infty(T_n(F)/M_n)^{W'}$ so that the push-forward $r_*^\vee(\Phi)$ is $f^{(B)}$. Thus

$$\mathrm{tr}\, J(\chi)_s(f) = \chi\omega_s(f^{(B)}) = \chi\omega_s(r_*^\vee(\Phi)) = r(\chi)|\det|^s(\Phi). \tag{4-6}$$

We are now at the top row of the diagram. The usual embedding $\mathrm{GL}_n \hookrightarrow \mathfrak{gl}_n$ of algebraic monoids induces an embedding $T_n \hookrightarrow \mathfrak{t}_n$ where $\mathfrak{t}_n := \mathrm{Lie}(T_n)$. We can therefore regard an element of $C_c^\infty(T_n(F))$ as an element of $C_c^\infty(\mathfrak{t}_n(F))$. The pairing

$$\mathfrak{t}_n(F) \times \mathfrak{t}_n(F) \to \mathbb{C}, \qquad (X, Y) \mapsto \psi(\mathrm{tr}(XY))$$

is perfect. For $t \in T_n(F)$ let

$$\widehat{\Phi}(t) = \int_{\mathfrak{t}_n(F)} \Phi(x)\psi(\mathrm{tr}(tx))\,dx; \tag{4-7}$$

it is just the Fourier transform. We then have

$$\gamma(s, J(r(\chi)), \psi) r(\chi)|\det|^s(\Phi) = r(\chi^{-1})|\det|^{1-s}(\widehat{\Phi}) \tag{4-8}$$

by the local functional equation of Tate's thesis [1979, (3.2.1)]. Here the left-hand side is meromorphic as a function of s, and

$$\operatorname{tr} r(\chi^{-1})|\det|^{1-s}(\widehat{\Phi})$$

is absolutely convergent for $\operatorname{Re}(s)$ sufficiently small in a sense depending on χ.

We are now at the upper right corner of the diagram. The representation $J(\chi)$ is unitary, but χ need not be. However, there is a $\mu \in \mathfrak{a}_{\mathbb{C}}^*$ whose real part lies in the closed convex hull of $W \cdot \rho$ in $\mathfrak{a}^*$ such that $\chi(t) = e^{\langle \mu, H_T(t)\rangle}$ by [Knapp 1986, p. 654]. In particular, since $0 < p \le 1$, one has $\mu \in \mathfrak{a}_p^*$. Then

$$r(\chi^{-1}) = e^{\langle r(-\mu), H_T\rangle}. \tag{4-9}$$

Combining this notation with (4-6) and (4-8) we arrive at

$$\begin{aligned} \gamma(s, J(r(\chi)), \psi) \operatorname{tr} J(\chi)_s(f) &= \gamma(s, J(r(\chi)), \psi) r(\chi)|\det|^s(\Phi) \\ &= r(\chi^{-1})|\det|^{1-s}(\widehat{\Phi}) \\ &= \widehat{\Phi}^{\sim}(r(-\mu + (1-s)d_\omega)). \end{aligned} \tag{4-10}$$

Now $\lambda \mapsto \widehat{\Phi}^{\sim}(r(\lambda))$ is in $\mathcal{S}(\mathfrak{a}_p^*)(-s_0)$ for s_0 sufficiently large by Corollary 4.3. Thus there is a unique $h \in \mathcal{S}^p(G(F)//K)$ so that for all $\lambda \in \mathfrak{a}_p^*$ one has

$$\hat{h}(-\lambda) = \widehat{\Phi}^{\sim}(r(\lambda + s_0 d_\omega))$$

by (3-4-1). In particular, if λ is chosen so that $J(\lambda)$ is spherical and unitary then

$$\operatorname{tr} J(\lambda)(h) = \widehat{\Phi}^{\sim}(r(\lambda + s_0 d_\omega)) \tag{4-11}$$

by Lemma 3.2. Set

$$\mathcal{F}_{r,\psi}(f) := h\omega_{-s_0}.$$

By construction, $h \in \mathcal{S}^p(G(F)//K)$. We have now successfully traversed along the upper path from $C_c^\infty(G(F)//K)$ to $\mathcal{S}^p(G(F)//K)\omega_{-s_0}$ in our diagram.

Combining (4-10) and (4-11) we deduce that (with μ as in (4-9)),

$$\begin{aligned} \gamma(s, J(\chi), r, \psi) \operatorname{tr} J(\chi)_s(f) &= \widehat{\Phi}^{\sim}(r(-\mu + (1-s)d_\omega)) \\ &= \mathcal{F}_{r,\psi}(f)^{\sim}(-\mu + (1-s)d_\omega) \\ &= \operatorname{tr} J(\chi)^\vee_{1-s}(\mathcal{F}_{r,\psi}(f)). \end{aligned} \tag{4-12}$$

This completes the proof of the theorem. □

5. A global application

Let F be a number field and let ∞ be the set of infinite places of F. Let $G = \mathrm{GL}_m$ for some integer m. It is not strictly necessary to take $G = \mathrm{GL}_m$ for what follows, but it makes the discussion simpler. We also restrict ourselves to everywhere unramified representations. We assume $\omega = \det : G \to \mathbb{G}_m$.

Remark. We have already discussed how one might remove the unramified assumption at the archimedean places in the introduction. To treat representations that are ramified at finite places one would have to define nonarchimedean Fourier transforms and give some analytic control on them similar to the control afforded in the archimedean setting by Theorem 1.1. More specifically, one would need to show that a twist of them by ω_s is L^1 for $\mathrm{Re}(s)$ sufficiently large. For $G = \mathrm{GL}_m$ one can define these nonarchimedean Fourier transforms spectrally using the Plancherel formula since the local Langlands correspondence is known. This is the approach of [Lafforgue 2014], where the analytic control is not established. Cheng and Ngô's approach [2017], if it is generalized from the finite field case to the local field case, may also yield the desired Fourier transforms.

We retain the obvious analogues of the notation above in this global setting; for example $T \leq \mathrm{GL}_m$ is a maximal split torus which we take to be the diagonal matrices for simplicity. For $v \nmid \infty$ let

$$\mathcal{S} : C_c^\infty(G(F_v)//G(\mathcal{O}_{F_v})) \to \mathbb{C}[\widehat{T}]^W \tag{5-1}$$

be the Satake isomorphism. Let

$$r : \widehat{G} \to \mathrm{GL}_m$$

be an irreducible representation. Let $\mathbb{A}_F^\infty$ denote the ring of finite adeles of F and let

$$\mathbb{L}_r := \prod_{v \nmid \infty} \mathbb{L}_{r,v} \in C_{\mathrm{ac}}^\infty(G(\mathbb{A}_F^\infty)//G(\widehat{\mathcal{O}}_F)),$$

where

$$\mathbb{L}_{r,v} := \sum_{k=0}^{\infty} \mathcal{S}^{-1}\big(\mathrm{tr}\,\mathrm{Sym}^k(r(t))\big) \in C_{\mathrm{ac}}^\infty(G(F_v)//G(\mathcal{O}_{F_v}))$$

with $t \in \widehat{T}(\mathbb{C})$. Here the subscript "ac" denotes the space of functions that are almost compactly supported, in other words, when restricted to a subset of $G(\mathbb{A}_F^\infty)$ with determinant lying in a compact subset of $(\mathbb{A}_F^\infty)^\times$ they are compactly supported (and similarly in the local setting). Then if π^∞ is an irreducible unramified admissible representation of $G(\mathbb{A}_F^\infty)$ one has

$$\mathrm{tr}\,\pi_s^\infty(\mathbb{L}_r) = L(s, \pi^\infty, r) \tag{5-2}$$

for $\mathrm{Re}(s)$ large enough. Here $\pi_s^\infty := \pi^\infty(|\omega|^\infty)^{s/N}$.

Let A be the connected component of the real points of the maximal $\mathbb{Q}$-split torus in the center of $\mathrm{Res}_{F/\mathbb{Q}} G$ and

$$G(\mathbb{A}_F)^1 := \ker(|\cdot|_{\mathbb{A}_F} \circ \det : G(\mathbb{A}_F) \to \mathbb{R}_{>0}).$$

Langlands [2004] proposed a method to prove Langlands functoriality in general via the trace formula. His point of departure was that for $\bar{f} \in C_c^\infty(A \setminus G(\mathbb{A}_{F\infty}))$ one can use the trace formula to provide an absolutely convergent expression for

$$\sum_\pi \operatorname{tr} \pi_\infty(\bar{f}) \pi_s^\infty(\mathbb{L}_r) \tag{5-3}$$

for $\mathrm{Re}(s)$ large. Here the sum over π is over cuspidal automorphic representations of $A \setminus G(\mathbb{A}_F)$. We note that the sum here is infinite, but since $\operatorname{tr} \pi_s^\infty(\mathbb{L}_r)$ is bounded independently of s and π for $\mathrm{Re}(s)$ sufficiently large (compare the proof of Corollary 5.1 below) and the regular action of any element of $C_c^\infty(A \setminus G(\mathbb{A}_F))$ on the cuspidal spectrum is trace class, the sum is absolutely convergent. Strictly speaking, Langlands used logarithmic derivatives of L-functions. Sarnak [2001] proposed the current formulation because it appears to be more tractable analytically.

One can then hope to use the trace formula to give an expression for

$$\sum_\pi \mathrm{Res}_{s=1} \operatorname{tr} \pi_\infty(\bar{f}) L(s, \pi^\infty, r) \tag{5-4}$$

in terms of orbital integrals (and automorphic representations on Levi subgroups). The residues ought to be nonzero for representations whose L-parameter, upon composition with r, fixes a vector in V. These ought to be transfers from smaller groups, and one hopes to compare the sum of residues (5-4) with corresponding sums on smaller groups. Since every algebraic subgroup of $\widehat{G}$ is the fixed points of a line in some representation of $\widehat{G}$ by a theorem of Chevalley, in principle executing this approach would lead to a proof of functoriality in general.

However, Langlands gave no absolutely convergent geometric expression for (5-4) nor any indication of how to obtain one, even assuming Langlands functoriality. In practice this seems to be an extremely difficult analytic hurdle that has only been overcome in a handful of cases [Altuğ 2015; Getz 2016; Getz and Herman 2015; Herman 2011; Venkatesh 2004; White 2014] that are essentially those isolated as tractable by Sarnak in his letter [2001].

In this section we use Theorem 1.1 and work of Finis, Lapid, and Müller to give an absolutely convergent expression in terms of orbital integrals and automorphic representations of Levi subgroups that is equal to (5-4) if one assumes Langlands functoriality (what we need is given precisely in Conjecture 5.3 below). We emphasize that the expression makes sense without any assumption in place, so one could try to use it to study Langlands' beyond endoscopy proposal. At the very least it allows us to replace (5-4) with a quantity which is well defined without any assumptions.

Remark. For some time Ngô has advocated combining Braverman and Kazhdan's proposal [2000] with the trace formula to prove functional equations of L-functions. The author first learned of this idea from Ngô at IAS in 2010, and Ngô has given a progress report on his perspective at the 2016 Takagi lectures. The construction we now propose, which is due to the author, amounts to understanding the residues that occur when one follows Ngô's suggestion. In particular one can give an absolutely convergent expression for the sum of residues that is the focus of Langlands' beyond endoscopy proposal.

For a compact open subgroup $K \leq G(\mathbb{A}_F^\infty)$, let

$$\mathcal{C}(G(\mathbb{A}_F), K) = \{f : G(\mathbb{A}_F)/K \to \mathbb{C} : |f * X|_{L^1(G(\mathbb{A}_F))} < \infty \text{ for all } X \in U(\mathfrak{g}_\mathbb{C})\}.$$

Here $\mathfrak{g}$ is the Lie algebra of

$$G(\mathbb{A}_{F\infty})$$

(viewed as a real Lie algebra) and $U(\mathfrak{g}_\mathbb{C})$ is its universal enveloping algebra. This is the space of test functions treated in [Finis et al. 2011; Finis and Lapid 2011]. A slightly different space of test functions (called $\mathcal{C}(G(\mathbb{A}_F)^1, K)$) is considered in [Finis and Lapid 2016]. In any case, the main result of these papers is that Arthur's noninvariant trace formula is valid for functions in $\mathcal{C}(G(\mathbb{A}_F), K)$.

Let $K_\infty \leq G(\mathbb{A}_{F\infty})$ be a maximal compact subgroup and let

$$f \in \otimes_{v|\infty} C_c^\infty(G(F_v)//K_v, r);$$

see (4-2). Let $\mathcal{F}_{r,\psi}(f) : G(\mathbb{A}_{F\infty}) \to \mathbb{C}$ be its nonabelian Fourier transform. For $g \in G(\mathbb{A}_F)$ and $s \in \mathbb{C}$ we set

$$\begin{aligned} f\mathbb{L}_r\omega_s(g) &:= f\omega_{s\infty}(g_\infty)\mathbb{L}_r\omega_s(g^\infty), \\ \mathcal{F}_{r,\psi}(f)\mathbb{L}_r\omega_s(g) &:= \mathcal{F}_{r,\psi}(f)\omega_{s\infty}(g_\infty)\mathbb{L}_r\omega_s^\infty(g^\infty). \end{aligned}$$

Fix a nontrivial additive character $\psi : F \backslash \mathbb{A}_F \to \mathbb{C}^\times$ and choose $d \in \mathbb{A}_F^\times$ such that $\psi^\infty(d\,\cdot)$ is unramified and $d_\infty \in A_{\mathbb{G}_m}$, the copy of $\mathbb{R}_{>0}$ embedded diagonally in $F_\infty^\times$. We assume, moreover, that $\omega_s(d^N I_m) = |d^\infty|^{ns}$. Choose $\sigma \in \mathbb{R}$ and let

$$\begin{aligned} f_1(g) &= f\mathbb{L}_r\omega_\sigma, \\ f_2(g) &= |d^\infty|^{-n}\mathcal{F}_{r,\psi}(f)\mathbb{L}_r\omega_\sigma(d^N I_m g). \end{aligned}$$

The role of the d here is explained by Conjecture 5.3 below.

The key consequence of Theorem 4.1 we use here is the following corollary:

Corollary 5.1. *If σ is sufficiently large then $f_1, f_2 \in \mathcal{C}(G(\mathbb{A}_F), G(\widehat{\mathcal{O}}_F))$.*

Proof. For σ sufficiently large one has

$$\int_{G(\mathbb{A}_F^\infty)} |\mathbb{L}_r(g)|\omega_\sigma(g)dg < \infty \tag{5-5}$$

by [Li 2017, Proposition 3.11]. Combining this with Theorem 4.1 we immediately deduce the corollary. □

For a $G(F)$-conjugacy class $\mathfrak{o}$ in $G(F)$ and an $h \in \mathcal{C}(G(\mathbb{A}_F), K)$ Finis and Lapid [2016] have shown that one can define the noninvariant orbital integral $J_{\mathfrak{o}}(h)$. Technically speaking they work with a slightly coarser notion than conjugacy, but it reduces to conjugacy for the case at hand since we have assumed G is a general linear group.

Together with Müller [Finis et al. 2011], they have also shown that for functions in the same space one can define a trace

$$\int_{i\mathfrak{a}_{L_s}^*} \operatorname{tr}(\mathcal{M}_L(P, \lambda) M(P, s) \rho(P, \lambda, h)) d\lambda \tag{5-6}$$

that is again absolutely convergent. Unfortunately, it would take several pages to define the notation used in (5-6); we refer the reader to [Finis et al. 2011, Corollary 1] and the discussion preceding it. This is the contribution of the Levi subgroup L_s to the trace formula.

Call a parabolic subgroup of G standard if it contains T. For each standard parabolic subgroup P, let M_P be the unique Levi subgroup of P containing T. For a cuspidal automorphic representation π of $A \backslash G(\mathbb{A}_F)$, let

$$\pi_s := \pi |\omega|^{s/N}.$$

The transfer $r(\pi)$ of π to $\mathrm{GL}_n(\mathbb{A}_F)$ is an irreducible admissible representation (which of course, we do not yet know to be automorphic). We let $\omega_{r(\pi)}$ be its central character.

Using Corollary 5.1 we can now apply the work of Finis, Lapid and Müller to prove the following theorem:

Theorem 5.2. *Consider*

$$\sum_{\pi} \left(\frac{1}{2\pi i} \int_{\mathrm{Re}(s)=\sigma} \left(\operatorname{tr} \pi_s(f \mathbb{L}_r) - \frac{|d^\infty|^{n(s-1)}}{\omega_{r(\pi)}(d^\infty)} \operatorname{tr}(\pi^\vee)_s(\mathcal{F}_{r,\psi}(f)\mathbb{L}_r) \right) ds \right) \tag{5-7}$$

where the sums are over isomorphism classes of cuspidal automorphic representations of $A \backslash G(\mathbb{A}_F)$. It is equal to

$$\sum_{\mathfrak{o}} J_{\mathfrak{o}}(f_1 - f_2) - \sum_{[P] \neq G} \frac{1}{|W(M_P)|} \sum_{s \in W(M_P)} \iota_s \int_{i\mathfrak{a}_{L_s}} \operatorname{tr}(\mathcal{M}_L(P,\lambda) M(P,s) \rho(P,\lambda, f_1 - f_2) d\lambda, \tag{5-8}$$

where the sum over $\mathfrak{o}$ is over conjugacy classes in $G(F)$, the sum over $[P]$ is over associate classes of standard proper parabolic subgroups of G, $W(M_P)$ is the Weyl group of M_P in G, and ι_s is the normalizing factor of [*Finis et al. 2011, Corollary* 1].

Moreover for each i and each $[P]$ and s,

$$\sum_{\mathfrak{o}} |J_{\mathfrak{o}}(f_i)| < \infty \quad \textit{and} \quad \int_{i\mathfrak{a}_{L_s}} |\mathrm{tr}(\mathcal{M}_L(P,\lambda)M(P,s)\rho(P,\lambda,f_i)|d\lambda < \infty.$$

Remark. The sum over $[P]$ and s above is finite.

Proof. Let $h \in \mathcal{C}(G(\mathbb{A}_F), G(\widehat{\mathcal{O}}_F))$. In [Finis et al. 2011; Finis and Lapid 2016] the authors extended the Arthur–Selberg trace formula to obtain the equality

$$\sum_{\mathfrak{o}} J_{\mathfrak{o}}(h) = \sum_{[P]} \frac{1}{|W(M_P)|} \sum_{s \in W(M_P)} \iota_s \int_{i\mathfrak{a}_{L_s}} \mathrm{tr}(\mathcal{M}_L(P,\lambda)M(P,s)\rho(P,\lambda,h)d\lambda \tag{5-9}$$

together with the absolute convergence of the sum over $\mathfrak{o}$ and the integral over $i\mathfrak{a}_{L_s}$. Here the sum is over all association classes of standard parabolic subgroups of G, including G itself. Thus the absolute convergence statements in the theorem follow.

Provided that the measure on A is chosen appropriately, the contribution of $[P] = G$ to the spectral side of (5-9) here is just

$$\sum_{\pi} \frac{1}{2\pi i} \int_{\mathrm{Re}(s)=0} \mathrm{tr}\,\pi_s(h)ds,$$

where the sum is over isomorphism classes of cuspidal automorphic representations of $A \backslash G(\mathbb{A}_F)$. Thus, pulling the contribution of the $[P] = G$ summand to one side, we see that the quantity (5-8) is equal to

$$\sum_{\pi} \frac{1}{2\pi i} \int_{\mathrm{Re}(s)=0} (\mathrm{tr}\,\pi_s(f_1) - \mathrm{tr}\,\pi_s^{\vee}(f_2))\,ds, \tag{5-10}$$

where the sums are over isomorphism classes of cuspidal automorphic representations of $A \backslash G(\mathbb{A}_F)$. We now note that $\omega_{r(\pi)}(aI_n) = \omega_\pi(a^N I_m)$. It follows from this and our choice of d that (5-10) is equal to (5-7), proving the theorem. □

In the remaining pages of this paper we prove that (5-7) is equal to (5-4), assuming a special case of Langlands' functoriality. Before we do this we emphasize again that it makes perfect sense to study (5-7) using (5-8) without assuming any conjecture, and we hope that some progress towards the conjecture we are about to state can be made by proceeding in this manner.

Here is the conjecture we will invoke:

Conjecture 5.3 (Langlands). *Let $\psi : F \backslash \mathbb{A}_F \to \mathbb{C}^\times$ be a nontrivial character, and choose $d^\infty \in \mathbb{A}_F^{\infty\times}$ such that $x \mapsto \psi(d^\infty x)$ is unramified at every finite place. For each everywhere unramified cuspidal automorphic representation π of $G(\mathbb{A}_F)^1$, the function*

$$L(s, \pi^\infty, r)$$

admits a meromorphic continuation to the plane, holomorphic except for a possible *pole at $s=1$. It satisfies the functional equation*

$$L(s,\pi^{\infty},r)=\omega_{r(\pi)}(d^{\infty})^{-1}|d^{\infty}|^{n(1-s)}\gamma(s,\pi_{\infty},r,\psi_{\infty})L(1-s,\pi^{\infty\vee},r).$$

In stating the conjecture in this manner, we are using the fact that

$$\begin{aligned}\varepsilon(s,\pi^{\infty},r,\psi^{\infty}) &:=\varepsilon(s,r(\pi^{\infty}),\psi^{\infty})\\ &=\omega_{r(\pi)}(d^{\infty})^{-1}|d^{\infty}|^{n(1-s)}\varepsilon(s,r(\pi^{\infty}),\psi^{\infty}(d^{\infty}\cdot))\\ &=\omega_{r(\pi)}(d^{\infty})^{-1}|d^{\infty}|^{n(1-s)}\end{aligned}$$

(compare [Tate 1979, (3.2.3)]), where $r(\pi^{\infty})$ is the transfer of π^{∞} to $\mathrm{GL}_m(\mathbb{A}_F^{\infty})$. It is known to exist as an admissible representation.

Theorem 5.4. *If Conjecture 5.3 is true for all unramified cuspidal automorphic representations of $A\setminus G(\mathbb{A}_F)$ then (5-7) is equal to the absolutely convergent sum*

$$\sum_{\pi}\mathrm{Res}_{s=1}\,\mathrm{tr}\,\pi_{\infty s}(f)L(s,\pi^{\infty},r). \tag{5-11}$$

Remark. We remark that the proof of Theorem 5.4 requires Theorem 1.1 in particular at the archimedean places; working outside a finite set of places including the archimedean ones where one can assume an unramified functional equation is not enough.

For an admissible irreducible representation π of $G(\mathbb{A}_F)$, let $\mathcal{C}(\pi,\mathrm{Re}(s))$ be its analytic conductor as defined by Iwaniec and Sarnak ([Brumley 2006, §1] is a nice reference). We have the following corollary of Conjecture 5.3:

Corollary 5.5. *Assume Conjecture 5.3 for the unramified cuspidal automorphic representation π. For any real numbers $A<B$ and $A\le\mathrm{Re}(s)\le B$ one has an estimate*

$$(s-1)^{\mathrm{ord}_{s=1}L(s,\pi^{\infty},r)}L(s,\pi^{\infty},r)\ll_{A,B,M}\mathcal{C}(\pi,\mathrm{Im}(s))^M$$

for some $M>0$.

Proof. Notice that

$$\mathcal{C}(r(\pi),\mathrm{Im}(s))\ll_N\mathcal{C}(\pi,\mathrm{Im}(s))^N$$

for some integer N depending only on r. Thus, since we are assuming Conjecture 5.3, to prove the corollary it suffices to prove that

$$(s-1)^{\mathrm{ord}_{s=1}L(s,\pi^{\infty},r)}L(s,\pi^{\infty},r)\ll_{A,B,M}\mathcal{C}(r(\pi),\mathrm{Im}(s))^M.$$

This is a standard preconvexity estimate; see, e.g., [Brumley 2006, (10)]. □

Proof of Theorem 5.4. Since $f\in C_c^{\infty}(G(F_{\infty}))$, the trace $\mathrm{tr}\,\pi_{\infty s}(f)$ is entire as a function of s. We also note that if $A\le\mathrm{Re}(s)\le B$ then $\mathrm{tr}\,\pi_{\infty s}(f)$ is rapidly decreasing as a function of $C(\pi, \mathrm{Im}(s))$; see [Getz 2012, Lemma 4.4]. Thus for σ sufficiently large

$$\sum_{\pi} \int_{\mathrm{Re}(s)=\sigma} \left| \operatorname{tr} \pi_s(f\mathbb{L}_r) \right| ds < \infty. \tag{5-12}$$

Applying Corollary 5.5 we see that (5-11) also converges absolutely.

Now consider

$$\sum_{\pi} \frac{1}{2\pi i} \int_{\mathrm{Re}(s)=\sigma} \frac{|d^\infty|^{n(s-1)}}{\omega_{r(\pi)}(d^\infty)} \operatorname{tr}(\pi^\vee)_s(\mathcal{F}_{r,\psi}(f)\mathbb{L}_r)\, ds.$$

By Theorem 4.1 and Conjecture 5.3 this is equal to

$$\begin{aligned} &\sum_{\pi} \frac{1}{2\pi i} \int_{\mathrm{Re}(s)=\sigma} \frac{|d^\infty|^{n(s-1)}}{\omega_{r(\pi)}(d^\infty)} \gamma(1-s, \pi_\infty, r, \psi) \operatorname{tr} \pi_{\infty 1-s}(f) \operatorname{tr}(\pi^{\infty\vee})_s(\mathbb{L}_r)\, ds \\ &\qquad = \sum_{\pi} \frac{1}{2\pi i} \int_{\mathrm{Re}(s)=\sigma} \operatorname{tr} \pi_{\infty 1-s}(f) L(1-s, \pi^\infty, r)\, ds \\ &\qquad = \sum_{\pi} \frac{1}{2\pi i} \int_{\mathrm{Re}(s)=1-\sigma} \operatorname{tr} \pi_{\infty s}(f) L(s, \pi^\infty, r)\, ds. \end{aligned} \tag{5-13}$$

Applying Corollary 5.5 and (5-12) we deduce that this converges absolutely. We now shift the contour in (5-13) to the line $\mathrm{Re}(s) = \sigma$, picking up the contribution of (5-11) from the poles at $s = 1$, and deduce the theorem. □

Acknowledgements

The author thanks W. Casselman and F. Shahidi for useful conversations. In particular, the observation that the conjecture of Braverman and Kazhdan considered in this paper ought to follow from Paley–Wiener theorems on reductive groups was made by Casselman in an email to the author. Moreover, W-W. Li and Shahidi deserve thanks for their comments on an earlier version of this paper. The author appreciates S. Cheng's help with references, D. Barbasch and L. Saper's aid with questions, and H. Hahn's constant encouragement and help with editing. V. Lafforgue also provided many useful suggestions on a previous version of this paper which improved the accuracy and exposition significantly.

References

[Altuğ 2015] S. A. Altuğ, "Beyond endoscopy via the trace formula, I: Poisson summation and isolation of special representations", *Compos. Math.* **151**:10 (2015), 1791–1820. MR Zbl

[Anker 1991] J.-P. Anker, "The spherical Fourier transform of rapidly decreasing functions: a simple proof of a characterization due to Harish-Chandra, Helgason, Trombi, and Varadarajan", *J. Funct. Anal.* **96**:2 (1991), 331–349. MR Zbl

[Arthur 1983] J. Arthur, "A Paley–Wiener theorem for real reductive groups", *Acta Math.* **150**:1-2 (1983), 1–89. MR Zbl

[Barbasch et al. 2008] D. Barbasch, D. Ciubotaru, and A. Pantano, "Unitarizable minimal principal series of reductive groups", pp. 63–136 in *Representation theory of real reductive Lie groups*, edited by J. Arthur et al., Contemp. Math. **472**, American Mathematical Society, Providence, RI, 2008. MR Zbl

[Bouthier et al. 2016] A. Bouthier, B. C. Ngô, and Y. Sakellaridis, "On the formal arc space of a reductive monoid", *Amer. J. Math.* **138**:1 (2016), 81–108. MR Zbl

[Braverman and Kazhdan 2000] A. Braverman and D. Kazhdan, "γ-functions of representations and lifting", *Geom. Funct. Anal.* Special Volume, Part I (2000), 237–278. MR Zbl

[Brumley 2006] F. Brumley, "Effective multiplicity one on GL_N and narrow zero-free regions for Rankin–Selberg L-functions", *Amer. J. Math.* **128**:6 (2006), 1455–1474. MR Zbl

[Cheng 2014] S. Cheng, "A global trace formula for reductive Lie algebras and the Harish-Chandra transform on the space of characteristic polynomials", preprint, 2014. arXiv

[Cheng and Ngô 2017] S. Cheng and B. C. Ngô, "On a conjecture of Braverman and Kazhdan", *Int. Math. Res. Not.* **2017** (2017), art. id. rnx052.

[Finis and Lapid 2011] T. Finis and E. Lapid, "On the continuity of Arthur's trace formula: the semisimple terms", *Compos. Math.* **147**:3 (2011), 784–802. MR Zbl

[Finis and Lapid 2016] T. Finis and E. Lapid, "On the continuity of the geometric side of the trace formula", *Acta Math. Vietnam.* **41**:3 (2016), 425–455. MR Zbl

[Finis et al. 2011] T. Finis, E. Lapid, and W. Müller, "On the spectral side of Arthur's trace formula: absolute convergence", *Ann. of Math.* (2) **174**:1 (2011), 173–195. MR Zbl

[Frenkel et al. 2010] E. Frenkel, R. Langlands, and B. C. Ngô, "Formule des traces et fonctorialité: le début d'un programme", *Ann. Sci. Math. Québec* **34**:2 (2010), 199–243. MR Zbl

[Gangolli and Varadarajan 1988] R. Gangolli and V. S. Varadarajan, *Harmonic analysis of spherical functions on real reductive groups*, Ergebnisse der Mathematik **101**, Springer, 1988. MR Zbl

[Getz 2012] J. R. Getz, "An approach to nonsolvable base change and descent", *J. Ramanujan Math. Soc.* **27**:2 (2012), 143–211. MR Zbl arXiv

[Getz 2014] J. R. Getz, "A summation formula for the Rankin–Selberg monoid and a nonabelian trace formula", submitted, 2014. arXiv

[Getz 2016] J. R. Getz, "A four-variable automorphic kernel function", *Res. Math. Sci.* **3** (2016), art. id. 20. MR Zbl

[Getz and Herman 2015] J. R. Getz and P. E. Herman, "A nonabelian trace formula", *Res. Math. Sci.* **2** (2015), art. id. 14. MR Zbl

[Getz and Liu 2017] J. R. Getz and B. Liu, "A refined Poisson summation formula for certain Braverman–Kazhdan spaces", submitted, 2017. arXiv

[Godement and Jacquet 1972] R. Godement and H. Jacquet, *Zeta functions of simple algebras*, Lecture Notes in Math. **260**, Springer, 1972. MR Zbl

[Herman 2011] P. E. Herman, "Beyond endoscopy for the Rankin–Selberg L-function", *J. Number Theory* **131**:9 (2011), 1691–1722. MR Zbl

[Knapp 1986] A. W. Knapp, *Representation theory of semisimple groups*, Princeton Mathematical Series **36**, Princeton Univ. Press, 1986. MR Zbl

[Lafforgue 2014] L. Lafforgue, "Noyaux du transfert automorphe de Langlands et formules de Poisson non linéaires", *Jpn. J. Math.* **9**:1 (2014), 1–68. MR Zbl

[Langlands 2004] R. P. Langlands, "Beyond endoscopy", pp. 611–697 in *Contributions to automorphic forms, geometry, and number theory* (Baltimore, 2002), edited by H. Hida et al., Johns Hopkins Univ. Press, 2004. MR Zbl

[Langlands 2013] R. P. Langlands, "Singularités et transfert", *Ann. Math. Qué.* **37**:2 (2013), 173–253. MR Zbl

[Li 2017] W.-W. Li, "Basic functions and unramified local L-factors for split groups", *Sci. China Math.* **60**:5 (2017), 777–812. MR Zbl

[Ngô 2014] B. C. Ngô, "On a certain sum of automorphic L-functions", pp. 337–343 in *Automorphic forms and related geometry: assessing the legacy of I. I. Piatetski-Shapiro* (New Haven, CT, 2012), edited by J. W. Cogdell et al., Contemp. Math. **614**, American Mathematical Society, Providence, RI, 2014. MR Zbl

[Sakellaridis 2012] Y. Sakellaridis, "Spherical varieties and integral representations of L-functions", *Algebra Number Theory* **6**:4 (2012), 611–667. MR Zbl

[Sarnak 2001] P. Sarnak, Comments on Robert Langlands' lecture "Endoscopy and beyond", 2001, Available at https://tinyurl.com/sarnakletter.

[Tate 1979] J. Tate, "Number theoretic background", pp. 3–26 in *Automorphic forms, representations and L-functions, II* (Corvallis, OR, 1977), edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, American Mathematical Society, Providence, RI, 1979. MR Zbl

[Trombi 1981] P. C. Trombi, "Harmonic analysis of $\mathscr{C}^p(G: F)(1 \leq p < 2)$", *J. Funct. Anal.* **40**:1 (1981), 84–125. MR Zbl

[Trombi 1990] P. C. Trombi, "Uniform asymptotics for real reductive Lie groups", *Pacific J. Math.* **146**:1 (1990), 131–199. MR Zbl

[Trombi and Varadarajan 1971] P. C. Trombi and V. S. Varadarajan, "Spherical transforms of semisimple Lie groups", *Ann. of Math.* (2) **94** (1971), 246–303. MR Zbl

[Venkatesh 2004] A. Venkatesh, "'Beyond endoscopy' and special forms on GL(2)", *J. Reine Angew. Math.* **577** (2004), 23–80. MR Zbl

[Vogan 1981] D. A. Vogan, Jr., *Representations of real reductive Lie groups*, Progress in Mathematics **15**, Birkhäuser, Boston, 1981. MR Zbl

[White 2014] P.-J. White, "The base change L-function for modular forms and 'Beyond endoscopy'", *J. Number Theory* **140** (2014), 13–37. MR Zbl

Received September 18, 2017.

JAYCE R. GETZ
MATHEMATICS DEPARTMENT
DUKE UNIVERSITY
DURHAM, NC
UNITED STATES

jgetz@math.duke.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 294, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.294.375

ENTROPY OF EMBEDDED SURFACES IN QUASI-FUCHSIAN MANIFOLDS

OLIVIER GLORIEUX

We compare critical exponents for quasi-Fuchsian groups acting on the hyperbolic 3-space and entropy of invariant disks embedded in $\mathbb{H}^3$. We give a rigidity theorem for all embedded surfaces when the action is Fuchsian and a rigidity theorem for negatively curved surfaces when the action is quasi-Fuchsian.

1. Introduction

The aim of this paper is to compare two geometric invariants of Riemannian manifolds: critical exponent and volume entropy. The first one is defined through the action of the fundamental group on the universal cover, the second one is defined for compact manifolds as the exponential growth rate of the volume of balls in the universal cover. These two invariants have been studied in many cases; we pursue this study for quasi-Fuchsian manifolds.

Let Γ be a group acting on a simply connected Riemannian manifold (X, g). If the action on X is discrete we define the *critical exponent* by

$$\delta(\Gamma) := \limsup_{R\to\infty} \frac{1}{R} \operatorname{Card}\{\gamma \in \Gamma \mid d(\gamma \cdot o, o) \leq R\}, \tag{1}$$

where o is any point in X. It does not depend on this particular base point thanks to triangle inequality. If we want to insist on the space on which Γ acts we will write $\delta(\Gamma, X)$.

The volume entropy $h(g)$ of a Riemannian compact manifold (Σ, g) is defined by

$$h(g) := \lim_{R\to\infty} \frac{\log \operatorname{Vol}_g(B_g(o, R))}{R}, \tag{2}$$

where $B_g(o, R)$ is the ball of radius R and center o in the universal cover of Σ. We will also use the notation $h(X)$ for the exponential growth rate of ball volumes in a a simply connected manifold X.

It is a classical fact, using a simple volume argument that the volume entropy coincides with the critical exponent of $\pi_1(\Sigma)$ acting on $\widetilde{\Sigma}$. Moreover, a famous

MSC2010: 32Q45, 51F99, 51K99, 53A35.
Keywords: hyperbolic geometry, entropy, quasi-Fuchsian, length spectrum.

theorem of G. Besson, G. Courtois and S. Gallot [Besson et al. 1995] said that the entropy allows us to distinguish the hyperbolic metric in the set of all metrics, $\mathrm{Met}(\Sigma)$. Note that entropy is sensitive to homothetic transformations: for any $\lambda > 0$ we have $h(\lambda^2 g) = \frac{1}{\lambda} h(g)$. Assume that Σ admits a hyperbolic metric g_0 and let $\mathrm{Met}_0(\Sigma)$ be the set of metrics on Σ whose volume is equal to $\mathrm{Vol}(\Sigma, g_0)$, then the theorem of Besson, Courtois, and Gallot says that for all $g \in \mathrm{Met}_0(\Sigma)$

$$h(g) \geq h(g_0), \tag{3}$$

with equality if and only if $g = g_0$.

Our aim is to study the behavior of the volume entropy for a subset of all the metrics on a surface. This subset is the metrics induced by an incompressible embedding into quasi-Fuchsian manifolds. It has not the cone structure of $\mathrm{Met}(\Sigma)$: it is not invariant by all homothetic transformations. Hence we will look at the behavior of $h(g)$ without normalization by the volume.

Let S be a compact surface of genus $g \geq 2$ and $\Gamma = \pi_1(S)$ its fundamental group. A Fuchsian representation of Γ is a faithful and discrete representation in $\mathrm{PSL}_2(\mathbb{R})$. A quasi-Fuchsian representation is a perturbation of Fuchsian representation in $\mathrm{PSL}_2(\mathbb{C})$. More precisely it is a discrete and faithful representation of Γ into $\mathrm{Isom}(\mathbb{H}^3)$ such that the limit set on $\partial\mathbb{H}^3$ is a Jordan curve. A celebrated theorem of R. Bowen [1979] asserts that for quasi-Fuchsian representations, the critical exponent is minimal and equal to 1 if and only if the representation is Fuchsian.

We choose an isometric, totally geodesic embedding of $\mathbb{H}^2$ in $\mathbb{H}^3$ (the equatorial plane in the ball model for example). This embedding gives an inclusion $i : \mathrm{Isom}(\mathbb{H}^2) \to \mathrm{Isom}(\mathbb{H}^3)$.

Let ρ be a Fuchsian representation of Γ. The group Γ acts naturally on $\mathbb{H}^2$ and $\mathbb{H}^3$ by ρ and $i \circ \rho$, respectively. For every point $o \in \mathbb{H}^2$ we have

$$d_{\mathbb{H}^3}(i \circ \rho(\gamma)o, o) = d_{\mathbb{H}^2}(\rho(\gamma)o, o),$$

since $\mathbb{H}^2$ is totally geodesic in $\mathbb{H}^3$. The critical exponents for these two actions of Γ are then equal

$$\delta(\Gamma, \mathbb{H}^3) = \delta(\Gamma, \mathbb{H}^2) = 1.$$

In light of this trivial example, two questions rise up. What is the entropy of a Γ-invariant disk which is not totally geodesic? What happens when we modify the Fuchsian representation in $\mathrm{PSL}_2(\mathbb{C})$?

We will answer the first question. Since ρ is a Fuchsian representation, the critical exponent of Γ acting on $\mathbb{H}^3$ through $i \circ \rho$ is 1, and we have the following:

Theorem 1.1. *Suppose Γ is Fuchsian. Let Σ be a Γ-invariant disk embedded in $\mathbb{H}^3$. We have*

$$h(\Sigma) \leq \delta(\Gamma, \mathbb{H}^3), \tag{4}$$

with equality if and only if Σ *is the totally geodesic hyperbolic plane preserved by* Γ.

Note that $\delta(\Gamma, \mathbb{H}^3) = h(\Sigma, g_0)$, hence the last theorem can be rewritten as follows:

Theorem 1.2. *For all metrics g obtained as induced metrics by an incompressible embedding in a Fuchsian manifold we have*

$$h(g) \leq h(g_0) \tag{5}$$

with equality if and only if $g = g_0$.

We *did not* renormalize by the volume; this explains the dichotomy between (3) and (5).

We will prove this theorem in the next section. The inequality is trivial since the induced distance between two points is always greater than the distance in $\mathbb{H}^3$: $d_\Sigma \geq d_{\mathbb{H}^3}$, but the rigidity is not. We have no geometrical (curvature) hypothesis on Σ, therefore it is not obvious at all to show that the inequality is strict as soon as Σ is not totally geodesic. Indeed we cannot use the "usual" techniques of negative curvature like Bowen–Margulis measure, or even the uniqueness of geodesic between two points.

We obtain an answer to the second question under a geometrical hypothesis on the curvature:

Theorem 1.3. *Let* Γ *be a quasi-Fuchsian group and* $\Sigma \subset \mathbb{H}^3$ *a* Γ*-invariant embedded disk. We suppose that* Σ *endowed with the induced metric has negative curvature. We then have*

$$h(\Sigma) \leq I(\Sigma, \mathbb{H}^3)\delta(\Gamma, \mathbb{H}^3),$$

where $I(\Sigma, \mathbb{H}^3)$ *is the geodesic intersection between* Σ *and* $\mathbb{H}^3$*. Moreover, equality occurs if and only if the length spectrum of* Σ/Γ *is proportional to that of* $\mathbb{H}^3/\Gamma$.

The *geodesic intersection* will be defined in Section 3A. Roughly, it is the average ratio of the length between two points of Σ for the extrinsic and intrinsic distance. We need the curvature assumption to define and use this invariant.

This theorem implies Theorem 1.1 only for negatively curved embedded disks but not in its full generality. Indeed, when Γ is Fuchsian, and Σ/Γ has the same length spectrum as $\mathbb{H}^3/\Gamma$ it follows directly by the work of J-P. Otal [1990] that $\Sigma = \mathbb{H}^2/\Gamma$. However, using the fact that Σ is embedded in $\mathbb{H}^3$ we will be able to prove without the Fuchsian hypothesis that if the two marked length spectra are equal then Σ is totally geodesic, and therefore we obtain the following corollary of Theorem 1.3:

Corollary 1.4. *Under the assumptions of Theorem 1.3 we have*

$$h(\Sigma) \leq \delta(\Gamma, \mathbb{H}^3),$$

with equality if and only if Γ *is fuchsian and* Σ *is the totally geodesic hyperbolic plane, preserved by* Γ.

The proof of this corollary raises the following question generalizing this result: if a quasi-Fuchsian manifold has the same length spectrum as a negatively curved surface, does it imply that it is in fact Fuchsian? We answer this question using a well-known result of Y. Benoist, showing the following theorem:

Theorem 1.5. *Let M be a quasi-Fuchsian manifold and Σ a hyperbolic* (*in the sense that it has constant curvature* -1) *surface. Suppose that M and Σ have proportional length spectrum* (*i.e., there exists $k \in \mathbb{R}^+$ such that for all $\gamma \in \Gamma$, $\ell_M(\gamma) = k\ell_\Sigma(\gamma)$*), *then M is Fuchsian and Σ is isometric to the totally geodesic surface in M.*

Theorem 1.3 has to be compared to results obtained by G. Knieper who compared entropy for two different metrics on the same manifolds, and our proof of Theorem 1.3 follows his paper [Knieper 1995]. As in his paper, we obtain that the intersection is larger than 1 as soon as Γ is not Fuchsian.

The theorem is also related to the work of M. Bridgeman and E. Taylor [2000]; indeed, we answer in the negative Question 2 of their paper. And finally, we can see our work as an extension of U. Hamenstadt's [2002], where she compared the geodesic intersection between the boundary of convex hulls and $\mathbb{H}^3$ for quasi-Fuchsian manifolds.

As we said, the two proofs are very different from one another. For the Fuchsian case, we give precise estimates for the length of some paths of the hyperbolic plane. We show that in some sense the length between two points on Σ is much greater than the extrinsic distance between those two points. For quasi-Fuchsian manifolds, we use well-known techniques of negative curvature geometry: we compare the Patterson–Sullivan measures for $\mathbb{H}^3$ and for Σ.

2. Fuchsian case

In this section we are going to prove Theorem 1.1. This theorem has a strong condition on Γ, i.e., it is conjugate to a subgroup of $\mathrm{PSL}_2(\mathbb{R})$ but we make no geometrical assumptions on Σ. As we said, there could be more than one geodesic between two points on Σ.

We already remarked that the inequality is trivial, as is the equality when Σ is totally geodesic. Therefore, the only thing left to prove is the strict inequality when Σ is not totally geodesic or in other words if $\Sigma \neq \mathbb{H}^2$ then $h(\Sigma) < 1$.

The proof of the theorem is based on the comparison between the distances on equidistant surfaces of the totally geodesic Γ-invariant hyperbolic plane. We are going to prove several lemmas which together give Theorem 1.1. The strict inequality follows directly from Lemmas 2.2 and 2.8. We denote by $\mathbb{D}$ the totally geodesic, Γ-invariant plane. The induced metric on $\mathbb{D}$ is the usual hyperbolic metric, and we will denote it by $\mathbb{H}^2$. We are first going to see that between all the equidistant surfaces, $\mathbb{H}^2$ has the biggest entropy. Then we will make this argument work when only one part of the surface is "above" $\mathbb{D}$. The idea to prove it, is to consider another distance d_m on $\mathbb{D}$, which will be used as an intermediary between Σ and $\mathbb{H}^2$. We will explain, after the definition of d_m, how the two comparisons will be proved.

Let us begin to parametrize $\mathbb{H}^3$ by $\mathbb{H}^2 \times \mathbb{R}$ as follows: take an orientation for the unit normal tangent space of $\mathbb{H}^2$, then to a point $x \in \mathbb{H}^3$ we associate $s(x)$ the orthogonal projection from $\mathbb{H}^3$ to $\mathbb{H}^2$. This is the first parameter of the parametrization. The oriented distance along this geodesic gives the second one. Hence the parametrization, called Fermi coordinates, is defined by

$$\mathbb{H}^3 \mapsto \mathbb{H}^2 \times \mathbb{R}, \qquad z \to (s(z), \hat{d}(z, s(z))),$$

where $\hat{d}$ is the oriented distance defined by the choice of the orientation on the unit normal tangent of $\mathbb{H}^2$. With this parametrization, the metric on $\mathbb{H}^3$ is

$$g_{\mathbb{H}^3} = \cosh^2(r) g_0 + dr^2.$$

Look at $S(r)$ the equidistant disk at distance r of $\mathbb{H}^2$; its metric, induced by the one on $\mathbb{H}^3$, is $g_r = \cosh^2(r) g_0$. It is isometric to a hyperbolic plane of curvature $1/\cosh(r)$, and its volume entropy is $h(S(r)) = h(0)/\cosh(r) = 1/\cosh(r)$, hence the entropy is maximal if and only if $r = 0$. For the general case, we are going to refine this argument showing that it is sufficient that a small part of Σ is over $\mathbb{H}^2$ for the entropy to be strictly less than 1.

Let Σ be a embedded Γ-invariant disk in $\mathbb{H}^3$. We assume that $\Sigma \neq \mathbb{D}$, and we endow Σ with its induced metric. Let x, y be two points on Σ. Let c_Σ be a geodesic on Σ linking x to y. We parametrize c_Σ by its Fermi coordinates, (c, r). We then have

$$\begin{aligned} d_\Sigma(x, y) &= \int_0^L \|c'_\Sigma(t)\|_\Sigma \, dt \\ &= \int_0^L \sqrt{r'(t)^2 + \cosh^2(r(t)) \|c'(t)\|^2_{g_0}} \, dt. \\ &\geq \int_0^L \cosh(r(t)) \|c'(t)\|_{g_0} \, dt. \end{aligned} \tag{6}$$

We now endow $\mathbb{D}$ with a different distance to the one coming from hyperbolic metric. It will play the role of intermediary to compare $d_\Sigma(x, y)$ on Σ with $d_{g_0}(s(x), s(y))$ on $\mathbb{H}^2$.

We call σ the restriction of s on Σ. Since $\Sigma \neq \mathbb{D}$, there exist $x_0 \in \mathbb{D} \setminus \Sigma$, $\varepsilon > 0$ and $\eta > 0$ such that

$$d_{\mathbb{H}^3}(\sigma^{-1} B(x_0, 2\varepsilon), \mathbb{D}) > \eta.$$

This means that *all* the points in the pre-image of $B(x_0, 2\varepsilon)$ by σ are at distance greater than η from $\mathbb{D}$. We will assume that 2ε is smaller than the injectivity radius of $\mathbb{H}^2/\Gamma$ so that the translations of $B(x_0, 2\varepsilon)$ by Γ are disjoint. We have taken 2ε in order to simplify the proof of Lemma 2.4.

We now consider on $\mathbb{D}$ the metric g_m defined by putting weight on the translations of $B(x_0, 2\varepsilon)$ by Γ.

Definition 2.1. We define g_m by

$$g_m := \begin{cases} \cosh(\eta)^2 g_0 & \text{on } \Gamma \cdot B(x_0, 2\varepsilon), \\ g_0 & \text{elsewhere.} \end{cases}$$

We will index by m objects which depend on this metric. Note that this metric is not continuous but it still defines a length space. Let $c : [0, 1] \to \mathbb{D}$ be a C^1 path, we then have

$$\ell_m(c) = \int_0^1 \|\dot{c}(t)\|_{g_m} \, dt.$$

This gives a distance d_m on $\mathbb{D}$ by choosing

$$d_m(x, y) := \inf_c \{\ell_m(c) \mid c(0) = x,\ c(1) = y\}.$$

In order to prove Theorem 1.1 we will compare the entropy of $(\mathbb{D}, d_m)$ with the one of Σ and the one of $\mathbb{H}^2$. The comparison with the entropy of Σ is quite easy and follows quickly from the definition of d_m and the inequality (6). The comparison with the entropy of $\mathbb{H}^2$ is more subtle. Indeed, there exist geodesics of $\mathbb{H}^2$ which are geodesics for $(\mathbb{D}, d_m)$ (any lift of a closed geodesic which does not cross the ball $B(x_0, 2\varepsilon)/\Gamma$) on $\mathbb{H}^2/\Gamma$). We will first prove that two points of $\mathbb{D}$ which are joined by a geodesic of $\mathbb{H}^2$ which often crosses $\Gamma \cdot B(x_0, 2\varepsilon)$ are much farther away from each other for d_m distance, see Lemma 2.4. Then, we will use a large deviation theorem for the geodesic flow (Theorem 2.6), to show that there are few geodesics which do not cross $\Gamma \cdot B(x_0, 2\varepsilon)$ (Lemma 2.7). It will follow from these two results that the balls of radius R for d_m are almost completely included in balls of radius R/C of $\mathbb{H}^2$ for $C > 1$ (Lemma 2.8). The two comparisons give the proof of Theorem 1.1.

The comparison between $h(\Sigma)$ and the critical exponent of $(\mathbb{D}, d_m)$ follows from the inequality (6) and the definition of d_m.

Lemma 2.2. *We have*

$$h(\Sigma) \le \delta((\mathbb{D}, d_m)).$$

Proof. Let $x \in \Sigma$ and $o = \sigma(x) \in \mathbb{D}$. Since Σ/Γ is compact, we have

$$h(\Sigma) = \lim_{R\to\infty} \frac{1}{R} \log \operatorname{Card}\{\gamma \in \Gamma \mid d_\Sigma(\gamma x, x) \le R\}.$$

And by definition

$$\delta((\mathbb{D}, d_m)) = \lim_{R\to\infty} \frac{1}{R} \log \operatorname{Card}\{\gamma \in \Gamma \mid d_m(\gamma o, o) \le R\}.$$

It is sufficient to prove that $d_\Sigma(x, y) \geq d_m(s(x), s(y))$, for all $x, y \in \Sigma$. Let $c_\Sigma = (c, r)$ be a geodesic on Σ joining x to y. Recall that we have

$$d_\Sigma(x, y) \geq \int_0^L \cosh(r(t)) \|c'(t)\|_{g_0}\, dt.$$

If $c(t) \notin \Gamma \cdot B(x_0, 2\varepsilon)$, then $\|c'(t)\|_{g_m} = \|c'(t)\|_{g_0}$. In particular,

$$\|c'(t)\|_{g_m} \leq \cosh(r(t)) \|c'(t)\|_{g_0}.$$

If $c(t) \in \Gamma \cdot B(x_0, 2\varepsilon)$, then by definition of g_m, $\|c'(t)\|_{g_m} = \cosh(\eta) \|c'(t)\|_{g_0}$ and since Σ is "far" from $\mathbb{D}$, $r(t) > \eta$. In particular,

$$\|c'(t)\|_{g_m} \leq \cosh(r(t)) \|c'(t)\|_{g_0}.$$

Finally,

$$\begin{aligned} d_\Sigma(x, y) &\geq \int_0^L \|c'(t)\|_{g_m}\, dt \\ &\geq l_m(c) \\ &\geq d_m(s(x), s(y)). \end{aligned}$$

□

Our next aim is to compare the distances d_m and $d_{\mathbb{H}^2}$. Let us fix some notations before stating the first lemma. For all $v \in T^1\mathbb{H}^2$, let ζ_R^v be the probability measure on $T^1\mathbb{H}^2$, defined for all Borel sets $E \subset T^1\mathbb{H}^2$ by

$$\zeta_R^v(E) = \frac{1}{R} \int_0^R \chi_E(\phi_t^{\mathbb{H}^2}(v))\, dt,$$

where χ_E is the indicator function of E. For a Borel set E which is a unitary tangent bundle of a subset of $\mathbb{D}$, $E := T^1 A$, we have

$$\zeta_R^v(E) = \frac{1}{R} \operatorname{Leb}\{t \in [0, R] \mid c_v(t) \in A\}$$

since $\phi_t^{\mathbb{H}^2}(v) \in E$ is equivalent to $c_v(t) = \pi \phi_t^{\mathbb{H}^2}(v) \in A$.

Let L be the Liouville measure on the unitary tangent bundle of the quotient surface $T^1\mathbb{H}^2/\Gamma$. Recall that the metric g_m is given by $g_m = \cosh^2(\eta) g_0$ on $T^1\Gamma B(x_0, 2\varepsilon)$. We fix $K := T^1(\Gamma \cdot B(x_0, \varepsilon))$.[1]

Definition 2.3. Let $\kappa > 0$ be such $L(K/\Gamma) - 2\kappa > 0$. We define the sets

$$\mathcal{E}(R) := \{v \in T^1\mathbb{H}^2 \mid \zeta_R^v(K) > L(K/\Gamma) - \kappa\},$$

and for all points $o \in \mathbb{H}^2$, we note

$$\mathcal{E}_o(R) := \{v \in T_o^1\mathbb{H}^2 \mid \zeta_R^v(K) > L(K/\Gamma) - \kappa\}.$$

[1] We use a ball of half the size, for a technical reason that appears at the beginning of the proof of Lemma 2.4.

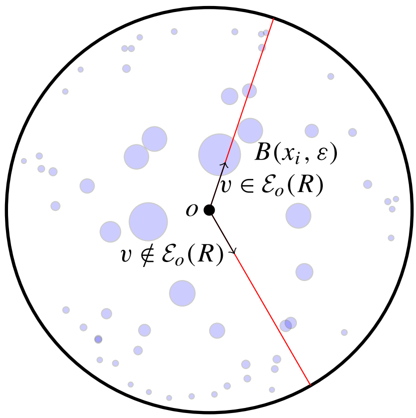


Figure 1. $\Gamma \cdot B(x_0, \varepsilon)$, $\mathcal{E}_o(R)$ and $\mathcal{E}_o^c(R)$.

A geodesic of length R whose direction is given by a vector $v \in \mathcal{E}(R)$ crosses πK "often", that is, at least a number of times proportional to R; see Figure 1. Indeed, if $v \in \mathcal{E}(R)$ we have

$$\frac{1}{R} \operatorname{Leb}\{t \in [0, R] \mid c_0(t) \cap \pi K \neq \varnothing\} > L(K/\Gamma) - \kappa > \kappa > 0,$$

since $\dot{c}_0(t) \in K$ is equivalent to $c_0(t) \in \pi K$ by definition of K.

The next argument is the key in the proof of Theorem 1.1. It shows that we can compare the length of a geodesic in $\mathbb{H}^2$ which often crosses πK with its d_m-length.

Lemma 2.4. *There exists $C > 1$, such that for all $R > 0$, for all $v \in \mathcal{E}_o(R)$ and for all $x \in \{\exp(tv) \mid t \in [R, 2R]\}$, we have*

$$d_m(o, x) \geq C d_{\mathbb{H}^2}(o, x). \tag{7}$$

Proof. Let c_0 be the geodesic for g_0 and c_m be a minimizing geodesic for g_m between o and x. Let d be the hyperbolic distance between o and x, $d = d_{\mathbb{H}^2}(o, x)$, and we parametrize c_0 by unit speed; we thus have $c_0(d) = x$. Let $N(R)$ be the number of intersections between πK and $c_0([0, R])$, that is N is the number of connected components of $c_0([0, R]) \cap \pi K$. On one hand, all components of $c_0([0, R]) \cap \pi K$ are inside balls of radius ε, hence c_0 "stays" at most 2ε in each components. On the other hand, the hypothesis $v \in \mathcal{E}_o(R)$, implies

$$\frac{1}{R} \operatorname{Leb}\{t \in [0, R] \mid c_0(t) \cap \pi K \neq \varnothing\} > L(K/\Gamma) - \kappa = \kappa > 0.$$

These two facts imply that $2\varepsilon N(R) \geq \kappa R$, that is to say,

$$N(R) \geq \frac{\kappa}{2\varepsilon} R. \tag{8}$$

For $i \leq N(R)$, let $t_i \in [0, d]$ such that $c_0(t_i) \in \pi K$ and $c_0[t_{i-1}, t_i] \setminus \pi K$ is connected: we just have chosen a point $x_i = c_0(t_i)$ in each ball of πK crossing c_0.

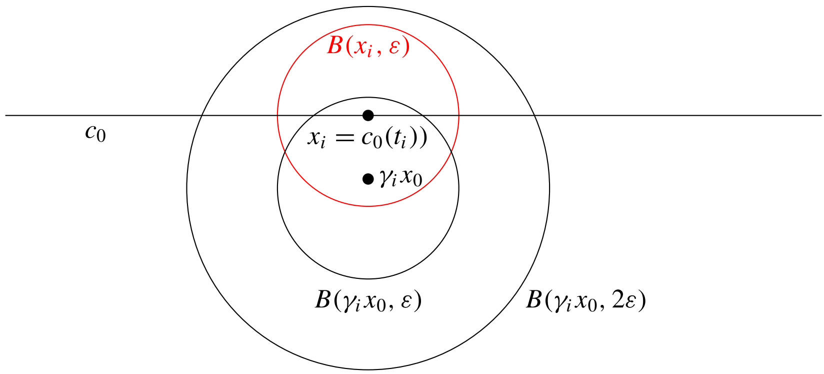


Figure 2. c_0 meets $B(\gamma_i x_0, \varepsilon)$. $B(x_i, \varepsilon) \subset B(\gamma_i x_0, 2\varepsilon)$.

There exists $\gamma_i \in \Gamma$ such that $x_i \in B(\gamma_i x_0, \varepsilon)$, hence $B(x_i, \varepsilon) \subset B(\gamma_i x_0, 2\varepsilon)$ on which the metric g_m is $g_m = \cosh^2(\eta) g_0$. See Figure 2. Therefore the geodesic c_0 is divided into $N(R)$ segments: $[x_i, x_{i+1}]$, such that for every i we know that on the ball $B(x_i, \varepsilon)$ the metric g_m is given by $g_m = \cosh^2(\eta) g_0$. We want a lower bound on $d_m(o, x)$, therefore we can estimate the length of c_m with the metric given by $\cosh^2(\eta) g_0$ on the smaller balls $B(x_i, \varepsilon) \subset B(\gamma_i x_0, 2\varepsilon)$ and g_0 on the rest of the plane.

We call y_i the middle of $[x_i, x_{i+1}]$. We now restrict our attention to one segment $[y_i, y_{i+1}]$. Let $0 < a < 1$ whose dependence on η will be made clear in the rest of the proof. We are going to analyze two different cases.

Case 1: c_m crosses $B(x_i, a\varepsilon)$. Let Δ_i be the lines (geodesics in $\mathbb{H}^2$) orthogonal to c_0 and passing through y_i. Let z_i^1 and z_i^2 be the end points of the diameter of $B(x_i, \varepsilon)$ defined by $z_i^1 = c_0(t_i - \varepsilon)$ and $z_i^2 = c_0(t_i + \varepsilon)$, and call D_i^1 and D_i^2 the lines orthogonal to c_0 and passing through z_i^1 and z_i^2. See Figure 3.

We want to consider the intersections between c_m and the lines Δ_i, D_i^1 and D_i^2. There might be many intersections. We will call the first intersection of c_m with a line D the point $c_m(t_f)$ where $t_f := \inf\{t \mid c_m(t) \in D\}$, and the last intersection of c_m with D the point $c_m(t_l)$, where $t_l := \sup\{t \mid c_m(t) \in D\}$.

Let A_i', B_i' and C_i' be the last intersections of c_m with Δ_i, D_i^1 and D_i^2, respectively. Let B_i, C_i and A_{i+1} be the first intersections of c_m with D_i^1, D_i^2 and Δ_{i+1}, respectively. This divides c_m into five connected components:

$$[A_i', B_i], \quad [B_i, B_i'], \quad [B_i', C_i], \quad [C_i, C_i'], \quad [C_i', A_{i+1}].$$

Our work will be to give a lower bound for the length of each component; see Figure 3. Since it might happen that $B_i = B_i'$ and $C_i = C_i'$ the bound on the length of those two components will be trivial: $d_m(B_i, B_i') \geq 0$ and $d_m(C_i, C_i') \geq 0$.

The g_m-length of c_m from A_i' to B_i is equal to (or larger than) its g_0-length since the metric g_m is equal to the metric g_0 outside K. Moreover the g_0-length of c_m

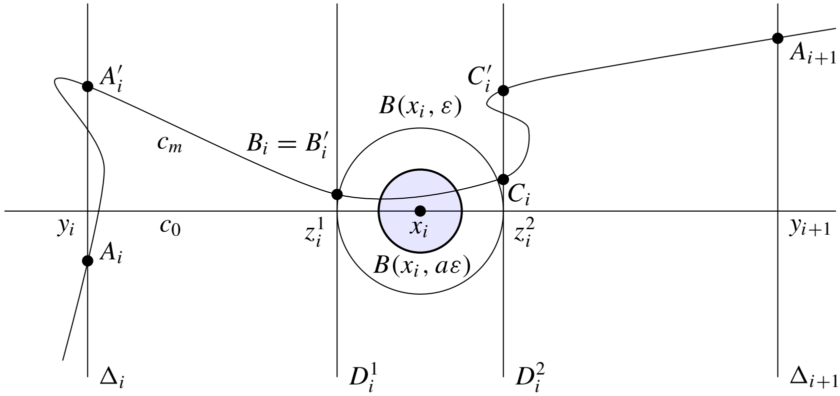


Figure 3. c_m crosses $B(x_i, a\varepsilon)$.

from A_i' to B_i is greater than $d_{g_0}(y_i, z_i^1)$ since the orthogonal projection decreases lengths. We then have

$$d_m(A_i', B_i) \geq d_{g_0}(y_i, z_i^1).$$

For the same reasons we have

$$d_m(C_i', A_{i+1}) \geq d_{g_0}(z_i^2, y_{i+1}).$$

We want to give a lower bound for the g_m-length of c_m between B_i' and C_i. We made the assumption that c_m crosses the ball $B(x_i, a\varepsilon)$ hence c_m stays at least $2\varepsilon - 2a\varepsilon$ in the ball $B(x_i, \varepsilon)$. In other words if c_m is unitary for g_0 we have $\mathrm{Leb}\{t \mid c_m(t) \cap B(x_i, \varepsilon) \neq \varnothing\} \geq 2\varepsilon - 2a\varepsilon$. In the ball $B(x_i, \varepsilon)$, the metric g_m is equal to $\cosh(\eta)^2 g_0$ hence the g_m-length satisfies

$$\begin{aligned} d_m(B_i', C_i) &\geq \int_{\{t \mid c_m(t) \cap B(x_i, \varepsilon) \neq \varnothing\}} \|\dot{c}_m(t)\|_m \, dt = \int_{\{t \mid c_m(t) \cap B(x_i, \varepsilon) \neq \varnothing\}} \cosh(\eta) \\ &\geq \varepsilon \cosh(\eta)(2 - 2a). \end{aligned}$$

Choose $a > 0$ such that $\cosh(\eta)(2\varepsilon - 2a\varepsilon) > 2\varepsilon$, that is to say $a \leq 1 - 1/\cosh(\eta)$. In order to fix the idea we set $a := \frac{1}{2}(1 - 1/\cosh(\eta))$. This implies

$$\begin{aligned} d_m(B_i', C_i) &\geq \varepsilon \cosh(\eta)(2 - 2a) \\ &= \varepsilon \cosh(\eta)\Big(2 - \Big(1 - \frac{1}{\cosh(\eta)}\Big)\Big) \\ &= (\cosh(\eta) + 1)\varepsilon \\ &= 2\varepsilon + \varepsilon[\cosh(\eta) - 1)] \\ &= d_{g_0}(z_i^1, z_i^2) + \varepsilon[\cosh(\eta) - 1)]. \end{aligned}$$

Thus, we have proven

$$d_m(A_i, A_{i+1}) \geq d_m(A_i', A_{i+1}) \geq d_{g_0}(y_i, y_{i+1}) + \varepsilon[\cosh(\eta) - 1]. \tag{9}$$

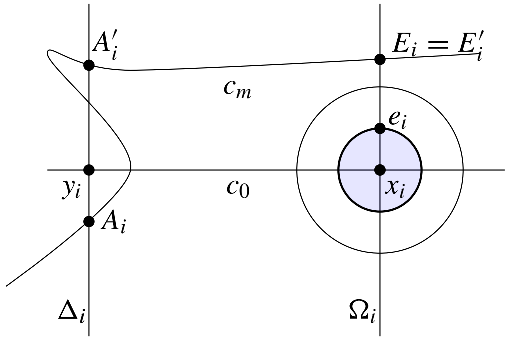


Figure 4. c_m does not cross $B(x_i, a\varepsilon)$.

Case 2: c_m does not cross $B(x_i, a\varepsilon)$. Let Δ_i be the line orthogonal to c_0 and passing through y_i, and Ω_i the one through x_i. Call A'_i the last intersection of c_m and Δ_i and E_i the first intersection of c_m with Ω_i. Since c_m does not cross $B(x_i a\varepsilon)$, E_i is in one of the connected components of $\Omega_i \setminus B(x_i, a\varepsilon)$. Named e_i the intersection of $S(x_i, a\varepsilon)$ (the sphere of center x_i and diameter $a\varepsilon$) and Ω_i in the same connected component as E_i, this is also the orthogonal projection of E_i on $B(x_i, a\varepsilon)$. See Figure 4.

We parametrize the geodesic Ω_i by $\mathbb{R}$; we give $\omega : \mathbb{R} \to \mathbb{H}^2$ such that $\omega(\mathbb{R}) = \Omega_i$. We suppose that $\omega(0) = x_i$ and the orientation is chosen in order to have $\omega(a\varepsilon) = e_i$. The function $t \to d_{g_0}(\omega(t), \Delta_i)$ is convex, and has a minimum at 0; it is hence increasing on $\mathbb{R}^+$. Therefore, $d_{g_0}(\Delta_i, E_i) \geq d_{\mathbb{H}^2}(\Delta_i, e_i)$. It follows that

$$d_m(A'_i, E_i) \geq d_{\mathbb{H}^2}(A'_i, E_i) \geq d_{g_0}(\Delta_i, E_i) \geq d_{g_0}(\Delta_i, e_i).$$

Let us compute $d_{g_0}(\Delta_i, e_i)$. We fix some notation:

$$L = d_{g_0}(\Delta_i, e_i), \qquad l = d_{g_0}(y_i, x_i), \qquad H = d_{g_0}(y_i, e_i).$$

Now Pythagoras' theorem in hyperbolic geometry for the triangle $(y_i x_i e_i)$ gives

$$\cosh(l)\cosh(a\varepsilon) = \cosh(H).$$

Let θ be the angle $\widehat{x_i\, y_i e_i}$. We have

$$\cos(\theta) = \frac{\tanh(l)}{\tanh(H)}$$

and

$$\sin(\pi/2 - \theta) = \frac{\sinh(L)}{\sinh(H)}.$$

Hence

$$\sinh(L) = \sinh(H)\frac{\tanh(l)}{\tanh(H)} = \cosh(H)\tanh(l) = \cosh(a\varepsilon)\sinh(l).$$

From this equation, we *cannot* conclude that $L > l + u$ for some $u > 0$. Indeed if L goes to 0 so does l. To avoid this problem we are going to assume that l is greater than the injectivity radius of S.

Note the following property of sinh which is a consequence of easy calculus. For all $x_0 > 0$ and $\varpi > 1$, there exists $u > 0$, such that for all $x > x_0$, we have $\varpi \sinh(x) \geq \sinh(x + u)$. Now we can choose y_i on c_0 in order to have

$$d_{g_0}(x_i, y_i) \geq s/2,$$

where s is the injectivity radius of $\mathbb{H}^2/\Gamma$. Consequently, applying the previous property with $\varpi = \cosh(a\varepsilon)$ and $x_0 = s/2$, there exists $u > 0$ such that

$$\cosh(a\varepsilon)\sinh(l) \geq \sinh(l + u).$$

Since sinh is increasing we deduce that

$$L \geq l + u.$$

Altogether, we show that there exists $u > 0$ such that

$$d_m(A'_i, E_i) \geq d_{g_0}(y_i, x_i) + u.$$

By the same arguments we can show that

$$d_m(E'_i, A_{i+1}) \geq d_{g_0}(x_i, y_{i+1}) + u.$$

(E'_i is the last intersection of c_m with Ω_i). Hence, if c_m does not meet $B(x_i, a\varepsilon)$, the g_m-length of c_m between A_i and A_{i+1} satisfies, (taking trivial bounds for first and last intersections)

$$d_m(A_i, A_{i+1}) \geq d_{g_0}(y_i, y_{i+1}) + 2u. \tag{10}$$

Now, let $\alpha := \min\{\varepsilon[\cosh(\eta) - 1]; 2u\}$. From (9) and (10) we have

$$d_m(A_i, A_{i+1}) \geq d_{g_0}(y_i, y_{i+1}) + \alpha.$$

Summing on i we get

$$d_m(o, x) \geq d_{g_0}(o, x) + N(R)\alpha.$$

Equation (8) and the fact that $d_{g_0}(o, x) \leq 2R$[2] imply that

$$N(R) \geq \frac{\kappa}{2\varepsilon}R \geq \frac{\kappa}{4\varepsilon}d_{g_0}(o, x).$$

Consequently,

$$d_m(o, x) \geq \left(1 + \frac{\alpha\kappa}{4\varepsilon}\right)d_{g_0}(o, x).$$

This proves the lemma with $C = \left(1 + \frac{\alpha\kappa}{4\varepsilon}\right)$. $\square$

[2]This is where we use the upper bound on $d_{g_0}(o, x)$.

We now compare the entropy of $(\mathbb{D}, d_m)$ with that of $\mathbb{H}^2$. Let us define

$$\mathcal{F}_o(R) = \{\exp(tv) \mid t \in \mathbb{R}^+,\ v \in \mathcal{E}_o(R)\}.$$

We denote by $B_m(o, 2R)$ the ball of radius $2R$ for the d_m distance.

Lemma 2.5. *Let $C' := \min(2, C)$ where C satisfies Lemma 2.4. For all $o \in \mathbb{D}$, and all $R > 0$,*

$$B_m(o, 2R) \subset B_{\mathbb{H}^2}(o, 2R/C') \cup \big(B_{\mathbb{H}^2}(o, 2R) \cap \mathcal{F}_o^c(R)\big).$$

Proof. We have $B_m(o, 2R) = (B_m(o, 2R) \cap \mathcal{F}_o(R)) \cup (B_m(o, 2R) \cap \mathcal{F}_o^c(R))$. Let $x \in B_m(o, 2R) \cap \mathcal{F}_o(R)$. Since $d_{\mathbb{H}^2}(o, x) \leq d_m(o, x)$, it follows that $d_{\mathbb{H}^2}(o, x) \leq 2R$. There are only two possibilities. If $d_{\mathbb{H}^2}(o, x) \leq R$, we have in particular $d_{\mathbb{H}^2}(o, x) \leq 2R/C'$. However, if $d_{\mathbb{H}^2}(o, x) \geq R$, we apply Lemma 2.4 and we get $d_{\mathbb{H}^2}(o, x) \leq 2R/C \leq 2R/C'$. Therefore,

$$B_m(o, 2R) \cap \mathcal{F}_o(R) \subset B_{\mathbb{H}^2}\Big(o, \frac{2R}{C'}\Big) \cap \mathcal{F}_o(R) \subset B_{\mathbb{H}^2}\Big(o, \frac{2R}{C'}\Big).$$

Since we also have for $R > 0$, $B_m(o, 2R) \subset B_{\mathbb{H}^2}(o, 2R)$, this gives

$$B_m(o, 2R) \cap \mathcal{F}_o^c(R) \subset B_{\mathbb{H}^2}(o, 2R) \cap \mathcal{F}_o^c(R),$$

and proves the lemma. □

The Liouville measure on $T^1\mathbb{H}^2$ is the product of the riemannian measure of $\mathbb{H}^2$ with the angular measure on every fiber. We denote this product by $L = d\mu(x) \times d\theta(x)$. Our aim is to show that the set $\mathcal{E}_o^c(R)$ is small and the volume of $(B_{\mathbb{H}^2}(o, 2R) \cap \mathcal{F}_o^c(R))$ is small compared to the one of $B_{\mathbb{H}^2}(o, 2R)$. For this we are going to use a large deviation theorem of Y. Kifer [1990] which gives an upper bound on the mass of the vectors which do not behave as the Liouville measure.

Let $\mathcal{P}$ be the set of probability measures on $T^1\mathbb{H}^2/\Gamma$ endowed with the weak topology. Let $\mathcal{P}^t$ be the subset of $\mathcal{P}$ of probability measures invariant by the geodesic flow. We also denote by L the Liouville measure on the quotient $T^1\mathbb{H}^2/\Gamma$. Recall that for a vector $v \in T^1\mathbb{H}^2/\Gamma$ we denote by ζ_v^R the probability measure given for all Borel subsets $E \subset T^1\mathbb{H}^2/\Gamma$ by

$$\zeta_R^v(E) = \frac{1}{R}\int_0^R \chi_E(\phi_t^{\mathbb{H}^2/\Gamma}(v))\,dt.$$

Theorem 2.6 [Kifer 1990, Theorem 3.4]. *Let $\overline{A}$ be a compact subset of $\mathcal{P}$,*

$$\limsup_{T\to\infty} \frac{1}{T} \log L\big\{v \in T^1\mathbb{H}^2/\Gamma \mid \zeta_v^T \in \overline{A}\big\} \leq -\inf_{\mu\in\overline{A}\cap\mathcal{P}^t} f(\mu),$$

where $f(\mu) = 1 - h_\mu(\phi_t^{\mathbb{H}^2/\Gamma})$ and $h_\mu(\phi_t^{\mathbb{H}^2/\Gamma})$ is the entropy of the geodesic flow $\phi_t^{\mathbb{H}^2/\Gamma}$ with respect to μ.

The fact that the theorem can be applied in this setting is explained after the Theorem 3.4 in [Kifer 1990]. In this reference the function f is given by a formula which seems different. One can look at [Paulin et al. 2015, Chapter 7], where the authors explain in detail why the geodesic flow of negatively curved surfaces satisfies the hypothesis of Kifer's theorem, and that one can take $f(\mu) = 1 - h_\mu(\phi_t^{\mathbb{H}^2/\Gamma})$.

Lemma 2.7. *There exist $o \in \mathbb{H}^2$, $\alpha > 0$ and $R_0 > 0$ such that for all $R > R_0$,*

$$\theta_o(\mathcal{E}_o^c(R)) \leq e^{-\alpha R}.$$

Proof. Let us keep the notations of Lemma 2.4. $K = T^1\Gamma \cdot B(x, \varepsilon)$ and we consider the following subset of $\mathcal{P}$:

$$A := \{\mu \in \mathcal{P} \mid \mu(K/\Gamma) \leq L(K/\Gamma) - \kappa\}.$$

This set is *not* closed for the weak topology. Its closure satisfies

$$\bar{A} \subset \{\mu \in \mathcal{P} \mid \mu(T^1\Gamma \cdot B^\circ(x, \varepsilon)/\Gamma) \leq L(K/\Gamma) - \kappa\},$$

where $B^\circ(x, \varepsilon)$ is the open ball. There might be equality between the two sets, but we won't use it.

However, since the unitary tangent bundle of the sphere $S(x, \varepsilon)$ is transverse to the flow, we have

$$\{v \in T^1\mathbb{H}^2/\Gamma \mid \zeta_v^R \in A\} = \{v \in T^1\mathbb{H}^2/\Gamma \mid \zeta_v^R \in \bar{A}\}.$$

Since $L \notin \bar{A}$ and L is the unique measure of maximal entropy satisfying $h(L) = 1$, we have

$$-\inf_{\mu \in \bar{A}} f(\mu) = -\alpha < 0.$$

Besides, it is clear that the set $\mathcal{E}^c(R) = \{v \in T^1\mathbb{H}^2 \mid \zeta_R^v(K) \leq L(K/\Gamma) - \kappa\}$ is Γ-invariant from the Γ invariance of K. By definition and the previous remark we get

$$\begin{aligned} \mathcal{E}^c(R)/\Gamma &= \{v \in T^1\mathbb{H}^2/\Gamma \mid \zeta_v^R \in A\} \\ &= \{v \in T^1\mathbb{H}^2/\Gamma \mid \zeta_v^R \in \bar{A}\}. \end{aligned}$$

Theorem 2.6 says that there exists $R_0 > 0$ such that for all $R > R_0$ we have

$$L(\mathcal{E}^c(R)/\Gamma) \leq e^{-\alpha R}.$$

The product structure of L implies the existence of a point $o \in \mathbb{H}^2/\Gamma$ such that

$$\theta_o(\mathcal{E}_o^c(R)/\Gamma) \leq e^{-\alpha R}.$$

The lemma follows, choosing any lift of o in $\mathbb{H}^2$. □

We finish the proof of Theorem 1.1 with Lemma 2.8, which compares the critical exponent between d_m and hyperbolic distance. Lemmas 2.2 and 2.8 conclude the proof.

Lemma 2.8. *There exists $u > 0$ such that*

$$\delta((\mathbb{D}, d_m)) \leq 1 - u.$$

Proof. We are going to show that the volume entropy of $(\mathbb{D}, d_m)$ satisfies the inequality, which implies a similar result on the critical exponent.

Let $o \in \mathbb{D}$ be a point satisfying Lemma 2.7. From Lemma 2.5, we have

$$B_m(o, 2R) \subset B_{\mathbb{H}^2}\Big(o, \frac{2R}{C'}\Big) \cup (B_{\mathbb{H}^2}(o, 2R) \cap \mathcal{F}_o^c(R)).$$

On one hand we have the classical upper bound $\mathrm{Vol}(B_{\mathbb{H}^2}(o, 2R/C')) = O(e^{2R/C'})$. On the other hand the volume form on $\mathbb{H}^2$ can be written in polar coordinates as $\sinh(r) dr d\theta$, hence for all $R > R_0$ we get

$$\mathrm{Vol}\left(B_{\mathbb{H}^2}(o, 2R) \cap \mathcal{F}_o^c(R)\right) = \int_0^{2R} \int_{\mathcal{E}_o^c(R)} \sinh(r)\, d\theta\, dr \leq \int_0^{2R} e^{-\alpha R} e^r\, dr$$
$$\leq e^{(2-\alpha)R}.$$

Let $u > 0$, defined by $1 - u = \max(1/C', (1 - \alpha/2)) < 1$. The last two upper bounds give

$$\mathrm{Vol}(B_m(o, 2R)) = O(e^{2R/C'}) + O(e^{(2-\alpha)R}) = O(e^{2(1-u)R}).$$

We finish by taking the log and the limit. $\square$

3. Quasi-Fuchsian case

3A. ***Geodesic intersection.*** Let Σ be an incompressible surface in M. We designate by $\phi_t^{\mathbb{H}^3}$, ϕ_t^{Σ} the geodesic flows on the unitary tangent spaces $T^1\mathbb{H}^3$, $T^1\Sigma$ respectively. We denote by π the projection from $T^1\mathbb{H}^3$ to $\mathbb{H}^3$. The restriction of π to $T^1\Sigma$ will still be denoted by π. There are two distances we can consider on Σ. The intrinsic one, defined as the infimum of the length of curves staying on Σ and the extrinsic one, where we take the distance in $\mathbb{H}^3$. We will denote by d_Σ and d these two distances.

First let us remark that there is no riemanniann metric on Σ which induces d. If such a metric existed, our Theorem 1.3 would be a particular case of [Knieper 1995].

Proposition 3.1. *If Σ is not totally geodesic, there is no riemannian metric on Σ which induces d.*

Proof. Assume there is such a riemannian metric, named g'. Let $\varepsilon > 0$ be such that the exponential map for g' is an embedding at every point. Let $c_{g'} : [0, \varepsilon] \to \Sigma$ be a minimizing geodesic for g' on Σ, then for all $t \in [0, \varepsilon]$,

$$d_{g'}(c_{g'}(0), c_{g'}(t)) + d_{g'}(c_{g'}(t), c_{g'}(\varepsilon)) = d_{g'}(c_{g'}(0), c_{g'}(\varepsilon)).$$

But since we suppose that g' induces d we have the same equality for d,

$$d(c_{g'}(0), c_{g'}(t)) + d(c_{g'}(t), c_{g'}(\varepsilon)) = d(c_{g'}(0), c_{g'}(\varepsilon)),$$

and this implies that $c_{g'}$ is a geodesic for $\mathbb{H}^3$. Hence every point of Σ is included in a totally geodesic disc, therefore Σ is totally geodesic. □

Consider the function a defined by

$$T^1\Sigma \times \mathbb{R} \to \mathbb{R}, \qquad (v, t) \mapsto d(\pi\phi_t^\Sigma(v), \pi(v)).$$

Letting $t_1, t_2 \in \mathbb{R}$ and $v \in T^1\Sigma$, we have by the triangle inequality,

$$\begin{aligned} a(v, t_1 + t_2) &= d(\pi\phi_{t_1+t_2}^\Sigma(v), \pi(v)) \\ &\leq d(\pi\phi_{t_1+t_2}^\Sigma(v), \pi\phi_{t_1}^\Sigma(v)) + d(\pi\phi_{t_1}^\Sigma(v), \pi(v)) \\ &\leq d(\pi\phi_{t_2}^\Sigma(\phi_{t_1}v), \pi\phi_{t_1}^\Sigma(v)) + d(\pi\phi_{t_1}^\Sigma(v), \pi(v)) \\ &\leq a(\phi_{t_1}^\Sigma v, t_2) + a(v, t_1). \end{aligned}$$

Hence a is a subadditive cocycle for the geodesic flow ϕ_t^Σ. Since a is Γ-invariant it defines a subadditive cocycle on $T^1\Sigma$, still denoted by a.

The following is a consequence of Kingman's subadditive ergodic theorem [Kingman 1973].

Theorem 3.2. *Les μ be a ϕ_t^Σ invariant probability measure on $T^1\Sigma$. Then*

$$I_\mu(\Sigma, M, v) := \lim_{t\to\infty} \frac{a(v, t)}{t}$$

exists for μ-almost $v \in T^1\Sigma$ and defines a μ-integrable function on $T^1\Sigma$, invariant under the geodesic flow and we have

$$\int_{T^1\Sigma} I_\mu(\Sigma, M, v)\, d\mu = \lim_{t\to\infty} \int_{T^1\Sigma} \frac{a(v, t)}{t}\, d\mu.$$

Moreover if μ is ergodic, $I_\mu(\Sigma, M, v)$ is constant μ-almost everywhere. In this case, we write $I_\mu(\Sigma, M)$.

3B. *Patterson–Sullivan measures.* We call Λ the limit set of Γ acting on $\mathbb{H}^3$. Since Γ acts cocompactly on Σ, and on the convex core $C(\Lambda)$, the three geometric spaces Γ (seen as its Cayley graph), Σ and $C(\Lambda)$ are quasi-isometric. We assume from now on that (Σ, g) has negative curvature, hence there is a unique geodesic in each homotopy class of curves, and for every pair of points in Σ there is a unique geodesic which joins them. Let c_Σ be a geodesic on Σ, and denote by $c_\Sigma(\pm\infty)$ its limit points on Λ. There is a unique $\mathbb{H}^3$-geodesic $c_{\mathbb{H}^3}$ whose endpoints are $c_\Sigma(\pm\infty)$. Since Σ is quasi-isometric to $C(\Lambda)$, the two geodesics $c_{\mathbb{H}^3}$ and c_Σ are at bounded distance.

Let $p \in \Sigma$ and call pr_p^Σ the projection from Σ to Λ defined as follows. For any point $x \in \Sigma$ call $c_{p,x}^\Sigma$ the geodesic on Σ which joins p to x, then

$$pr_p^\Sigma(x) = c_{p,x}^\Sigma(+\infty).$$

We will denote the equivalent projection in $\mathbb{H}^3$ by $pr_p^{\mathbb{H}^3}$. There are two small distinctions to notice between $pr_p^{\mathbb{H}^3}$ and pr_p^Σ. First, $pr_p^{\mathbb{H}^3}$ is defined for every point in $\mathbb{H}^3$, whereas pr_p^Σ is only defined for points in Σ. Second is that the codomain of pr_p^Σ is exactly Λ whereas the codomain of $pr_p^{\mathbb{H}^3}$ is all S^2.

As we have just stated, for all $\xi \in \Lambda$ the geodesics, $c_{p,\xi}^\Sigma$ and $c_{p,\xi}^{\mathbb{H}^3}$ are at bounded distance, and this bound depends only on the quasi-isometry between Σ and $C(\Lambda)$. There exists C_1 such that for all $\xi \in \Lambda$ the Hausdorff distance between geodesics $c_{p,\xi}^\Sigma$ and $c_{p,\xi}^{\mathbb{H}^3}$ is less than C_1.

Let $x \in \Sigma$, $R > 0$ and consider the ball $B_{\mathbb{H}^3}(x, R)$ in $\mathbb{H}^3$ of center x and radius R. Now take $\xi \in pr_p^{\mathbb{H}^3}(B(x, R - C_1)) \cap \Lambda$; this means that the $\mathbb{H}^3$-geodesic from p to ξ crosses the ball $B_{\mathbb{H}^3}(x, R - C_1)$. This $\mathbb{H}^3$-geodesic is at bounded distance C_1 of the Σ-geodesic joining p to ξ. Hence,

$$c_{p,\xi}^\Sigma \cap (B_{\mathbb{H}^3}(x, R) \cap \Sigma) \neq \varnothing,$$

which proves that

$$\xi \in pr_p^\Sigma(B_{\mathbb{H}^3}(x, R) \cap \Sigma).$$

The same argument shows that

$$pr_p^\Sigma(B_{\mathbb{H}^3}(x, R) \cap \Sigma) \subset pr_p^{\mathbb{H}^3}(B_{\mathbb{H}^3}(x, R + C_1)) \cap \Lambda \subset pr_p^{\mathbb{H}^3}(B_{\mathbb{H}^3}(x, R + C_1)).$$

The distances on Σ and on $\mathbb{H}^3$ are locally equivalent: for every $R > 0$ there exists C_2 such that all balls satisfy

$$B_\Sigma(x, R - C_2) \subset B_{\mathbb{H}^3}(x, R) \cap \Sigma \subset B_\Sigma(x, R + C_2).$$

Set $C = \max(C_1, C_2)$, which leads to the following theorem:

Theorem 3.3.
$$\begin{array}{c} pr_p^\Sigma(B_\Sigma(x, R - C)) \\ \cap \\ pr_p^{\mathbb{H}^3}(B_{\mathbb{H}^3}(x, R - C)) \cap \Lambda \subset pr_p^\Sigma(B_{\mathbb{H}^3}(x, R) \cap \Sigma) \subset pr_p^{\mathbb{H}^3}(B_{\mathbb{H}^3}(x, R + C)) \\ \cap \\ pr_p^\Sigma(B_\Sigma(x, R + C)). \end{array}$$

Before proving Theorem 1.3, we will recall some basic facts about Patterson–Sullivan measure. Some classical references for this are [Patterson 1976] and [Sullivan 1979], the lecture of J-F. Quint [2006] and the monograph of T. Roblin [2003]. Let (X, g) be a simply connected manifold with negative curvature and $X(\infty)$ its geometric boundary. If Γ is a discrete group acting on (X, g) we can associate to it a family of measures $\{\mu^g_p\}_{p\in X}$ on $X(\infty)$ constructed as follows. Let x, y be two points of X and consider the Poincaré series

$$P(s) := \sum_{\gamma\in\Gamma} e^{-sd(\gamma x, y)}.$$

The convergence of $P(s)$ is independent of x and y by the triangle inequality. It converges for $s > \delta(\Gamma)$ and diverges for $s < \delta(\Gamma)$. If the action is cocompact, $\delta(\Gamma) = h(g)$ and the series diverges at $h(g)$. Then we define the probability measure

$$\mu^g_{p,x}(s) := \frac{\sum_{\gamma\in\Gamma} e^{-sd(\gamma x, p)}\delta_{\gamma x}}{\sum_{\gamma\in\Gamma} e^{-sd(\gamma p, p)}}.$$

By compactness of the set of probability measures on $X(\infty)$, we obtain a measure on $X(\infty)$ by taking a weak limit of a sequence $\mu^g_{p,x}(s_n)$[3],

$$\mu^g_p := \lim_{s_n\to h(g)} \mu^g_p(s_n).$$

It is supported on the accumulation points of G, that is to say the limit set.

These measures, called *Patterson–Sullivan measures*, have the following properties. They are quasiconformal, i.e., for all $p \in X$ and all $\xi, \eta \in \Lambda$, we have

$$\frac{d\mu^g_p}{d\mu^g_q}(\xi) = e^{-h(g)\beta_\xi(p,q)},$$

where $\beta_\xi(p, q) = \lim_{z\to\xi} d_g(p, z) - d_g(q, z)$.

They are also Γ-equivariant, i.e., for all $\gamma \in \Gamma$ and all $p \in X$, we have

$$\mu^g_p \circ \gamma = \mu^g_{\gamma^{-1}p}.$$

Moreover we know these measures behave locally like $h(g)-$Hausdorff measures. See [Quint 2006, Lemma 4.10], for example.

Lemma 3.4 (shadowing). *For $R > 0$ sufficiently large, there exists $c > 1$ such that for all $x \in X$,*

$$\frac{1}{c}e^{-h(g)d_g(x,p)} \le \mu^g_p(pr^g_p(B_g(x, R))) \le ce^{-h(g)d_g(x,p)}.$$

Suppose that X/Γ is compact; from the Patterson–Sullivan measure, we can construct an invariant measure on T^1X/Γ. Let $\Lambda^{(2)} := \{(x, y) \in \Lambda^2 \mid x \neq y\}$. There is a natural identification of $\Lambda^{(2)} \times \mathbb{R}$ and T^1X; a vector $v \in T^1X$ is identified with $(c_v(+\infty), c_v(-\infty), \beta_{c_v(+\infty)}(p, \pi v))$. The Bowen–Margulis measure is defined by

$$d\mu_{BM}(\xi, \eta, t) = e^{2h(g)\langle\xi|\eta\rangle_p} d\mu^g_p(\xi)d\mu^g_p(\eta)\,dt,$$

[3] It is a classical result of Sullivan that there is in fact a unique limit, up to normalization. It is equivalent to the ergodicity of Bowen–Margulis measure [Roblin 2003, Chapter 1]

where $\langle \xi \mid \eta \rangle_p$ is the Gromov product:

$$\langle \xi \mid \eta \rangle_p = \tfrac{1}{2}\big(\beta_\xi(z, p) + \beta_\eta(z, p)\big),$$

where z is any point on the geodesic (ξ, η).

Let us recall the classical fact that the measure μ_{BM} is Γ-invariant and define therefore a measure on T^1X/Γ. Letting $z \in (\xi, \eta)$,

$$\begin{aligned}\langle \gamma\xi \mid \gamma\eta \rangle_p &= \tfrac{1}{2}\big(\beta_{\gamma\xi}(\gamma z, p) + \beta_{\gamma\eta}(\gamma z, p)\big)\\ &= \tfrac{1}{2}\big(\beta_{\gamma\xi}(\gamma z, \gamma p) + \beta_{\gamma\xi}(\gamma p, p) + \beta_{\gamma\eta}(\gamma z, \gamma p) + \beta_{\gamma\eta}(\gamma p, p)\big)\\ &= \tfrac{1}{2}\big(\beta_\xi(z, p) + \beta_\eta(z, p) + \beta_{\gamma\xi}(\gamma p, p) + \beta_{\gamma\eta}(\gamma p, p)\big)\\ &= \langle \xi \mid \eta \rangle_p + \tfrac{1}{2}\big(\beta_{\gamma\xi}(\gamma p, p) + \beta_{\gamma\eta}(\gamma p, p)\big).\end{aligned}$$

By the quasiconformal behavior of μ_p^g, we have

$$\begin{aligned}&e^{2h(g)\langle \gamma\xi \mid \gamma\eta \rangle_p} d\mu_p^g(\gamma\xi) d\mu_p^g(\gamma\eta)\\ &\qquad = e^{2h(g)\langle \xi \mid \eta \rangle_p} e^{h(g)\beta_{\gamma\xi}(\gamma p, p))} d\mu_p^g(\gamma\xi) e^{h(g)\beta_{\gamma\eta}(\gamma p, p))} d\mu_p^g(\gamma\eta)\\ &\qquad = e^{2h(g)\langle \xi \mid \eta \rangle_p} d\mu_p^g(\xi) d\mu_p^g(\eta).\end{aligned}$$

The invariance by the geodesic flow is clear by definition and it is shown in [Nicholls 1989] that μ_{BM} is ergodic.

Finally we will need the following theorem, which is classical for compact manifolds endowed with two different negatively curved metrics. Since we treat a slightly different case, we give a proof.

Theorem 3.5. *If μ_p^Σ and $\mu_p^{\mathbb{H}^3}$ are equivalent, then the marked length spectrum of Σ is proportional to the marked length spectrum of M.*

Note that in the Fuchsian case, any surface equidistant to the totally geodesic one has a metric proportional to $\mathbb{H}^2$ and therefore satisfies the hypothesis of the theorem. It seems likely that it is the only case where the length spectrum is proportional to the one of the ambient manifold, however this is still uncertain.

Definition 3.6. For all $\xi, \eta \in \partial X^{(2)}$, we define the function D_X by

$$D_X(\xi, \eta) = \exp(-\langle \xi \mid \eta \rangle_p).$$

It is shown in [Ghys and de la Harpe 1990] that D_X^a for $a > 0$ small enough is a distance, called Gromov distance. However, we do not need such renormalization here.

The proof of Theorem 3.5 is in two steps. In the first, we prove that if the Patterson Sullivan measures are equivalent then the functions D_Σ and $D_{\mathbb{H}^3}$ are Hölder equivalent. In the second, we prove that this last condition implies the proportionality of the length spectrum.

Lemma 3.7. *If μ_p^{Σ} and $\mu_p^{\mathbb{H}^3}$ are equivalent, then the functions $D_{\mathbb{H}^3}$ and D_Σ are Hölder equivalent.*

Proof. Let us consider on $\Lambda^{(2)}$ the Bowen–Margulis currents defined by

$$\nu_\Sigma(\xi,\eta)=\frac{d\mu_\Sigma^p(\xi)d\mu_\Sigma^p(\eta)}{D_\Sigma(\xi,\eta)^{2\delta(\Sigma)}},$$

$$\nu_{\mathbb{H}^3}(\xi,\eta)=\frac{d\mu_{\mathbb{H}^3}^p(\xi)d\mu_{\mathbb{H}^3}^p(\eta)}{D_{\mathbb{H}^3}(\xi,\eta)^{2\delta(\mathbb{H}^3)}}.$$

These two measures are Γ-invariant by the previous computations for the Bowen–Margulis measures.

By assumption, μ_Σ^p and $\mu_p^{\mathbb{H}^3}$ are equivalent, therefore ν_Σ and $\nu_{\mathbb{H}^3}$ are also equivalent. The ergodicity and the Γ-invariance imply the existence of $c>0$ such that

$$\nu_\Sigma=c\nu_{\mathbb{H}^3}.$$

Since μ_p^{Σ} and $\mu_p^{\mathbb{H}^3}$ are equivalent, there exists a function $f:\Lambda\to\mathbb{R}^+$ such that $\mu_p^{\Sigma}(\xi)=f(\xi)\mu_p^{\mathbb{H}^3}$. We have

$$f(\xi)f(\eta)D_{\mathbb{H}^3}^{\delta(\mathbb{H}^3)}(\xi,\eta)=cD_\Sigma^{\delta(\Sigma)}(\xi,\eta).$$

We see that f is equal almost everywhere to a continuous function. We can therefore suppose that f is continuous on Λ and hence strictly positive. By compacity, there exists $C>1$ such that $\frac{1}{C}\le f(\xi)\le C$. Finally we get what we stated:

$$\frac{c}{C^2}D_\Sigma^{\delta(\Sigma)}(\xi,\eta)\le D_{\mathbb{H}^3}^{\delta(\mathbb{H}^3)}(\xi,\eta)\le C^2cD_\Sigma^{\delta(\Sigma)}(\xi,\eta). \qquad \square$$

We now show the second part.

Lemma 3.8. *If D_Σ and $D_{\mathbb{H}^3}$ are Hölder equivalent the marked length spectra of Σ and $M=\mathbb{H}^3/\Gamma$ are proportional.*

Proof. In [Paulin et al. 2015, Section 3.5], the authors show that in a very general setting we have

$$\lim_{n\to\infty}\frac{1}{n}\log[g^-,g+,g^n(\xi),\xi]=\ell(g),$$

where $\ell(g)$ is the displacement of g and

$$[g^-,g^+,g^n(\xi),\xi]=\frac{D(g^-,g^n(\xi))D(g^+,\xi)}{D(g^-,\xi)D(g^+,g^n(\xi))}.$$

In particular, we can apply this result to Σ and $\mathbb{H}^3$ to get

$$\lim_{n\to\infty}\frac{1}{n}\log[g^-,g^+,g^n(\xi),\xi]_\Sigma=\ell_\Sigma(g)$$

and

$$\lim_{n\to\infty}\frac{1}{n}\log[g^-,g^+,g^n(\xi),\xi]_{\mathbb{H}^3}=\ell_{\mathbb{H}^3}(g).$$

By assumption on the distances D_Σ, $D_{\mathbb{H}^3}$, there exists $C > 1$ such that

$$\frac{1}{C}[g^-, g^+, g^n(\xi), \xi]^r_{\mathbb{H}^3} \le [g^-, g^+, g^n(\xi), \xi]_\Sigma \le C[g^-, g^+, g^n(\xi), \xi]^r_{\mathbb{H}^3}.$$

Hence,

$$\ell_\Sigma(g) = r\ell_{\mathbb{H}^3}(g). \qquad \square$$

Theorem 3.5 follows directly from Lemmas 3.7 and 3.8.

We will show at the very end of this article that if Σ has the same length spectrum as $M = \mathbb{H}^3/\Gamma$ then Γ is Fuchsian, to prove Corollary 1.4. It might be also true even when we only suppose that they are proportional, however this does not follow from our proof.

3C. ***Entropy comparison.*** We finally get to the proof of Theorem 1.3. First we prove the inequality using the behavior of Patterson–Sullivan measures and a volume comparison of a subset of Σ; the proof follows the same lines as [Knieper 1995, Theorem 3.4]. Then we prove the equality case using Theorem 3.5.

Theorem 3.9. *Let $\Sigma \subset \mathbb{H}^3$ be a Γ-invariant embedded disk, whose induced metric g has negative curvature, then*

$$h(g) \le I_{\mu_{BM}}(\Sigma, M)\delta(\Gamma).$$

Moreover, the equality occurs if and only if the marked length spectrum of Σ is proportional to the marked length spectrum of M. In this case, the proportionality factor is given by $\ell_\Sigma(g)I(\Sigma, M) = \ell_M(g)$.

Proof. The geodesic flow is ergodic with respect to the Bowen–Margulis measure μ_{BM}, hence for μ_{BM}-almost all $v \in T^1\Sigma$ we have

$$\lim_{t\to\infty} \frac{a(v,t)}{t} = I_\mu(\Sigma, M).$$

Let v and v' be two unit vectors on the same weak stable manifold. Then

$$d(c_{v'}(t), c_{v'}(0)) \le d(c_{v'}(t), (c_v(t)) + d(c_v(t), (c_v(0)) + d(c_v(0), (c_{v'}(0)),$$

and the same inequality holds interchanging the role of v and v'. Moreover $d(c_{v'}(t), (c_v(t))$ decreases exponentially since v and v' are on the same weak stable manifold. Hence $\lim_{t\to\infty} \frac{1}{t}a(v,t)$ exists if and only if $\lim_{t\to\infty} \frac{1}{t}a(v',t)$ does.

Let $v_p(\xi)$ denote the unitary vector in $T^1_p\Sigma$ such that $c_{v_p(\xi)}(\infty) = \xi$. The previous fact and the product structure of $d\mu_{BM}$ ensure that for μ^g_p-almost all $\xi \in \partial\Sigma$,

$$\lim_{t\to\infty} \frac{a(v_p(\xi), t)}{t} = I_\mu(\Sigma, M).$$

For all $\varepsilon > 0$ and $T > 0$, we define the set

$$A_p^{T,\varepsilon} = \left\{ \xi \in \partial\Sigma \;\middle|\; \left| \frac{a(v_p(\xi), t)}{t} - I_\mu(\Sigma, M) \right| \leq \varepsilon,\ t \geq T \right\}.$$

For all $d \in]0, 1[$ and all $\varepsilon > 0$, there exists $T > 0$ such that $\mu_p^\Sigma(A_p^{T,\varepsilon}) \geq d$. For $t > T$, consider the subset $\{c_{p,\xi}(t) \mid \xi \in A_p^{T,\varepsilon}\} \subset S_g(p, t)$ of the geodesic sphere of radius t and center p on Σ.

Choose $\{B_\Sigma(x_i, R) \mid i \in I\}$ a covering of this subset of fixed radius $R > 0$ such that $x_i \in S_\Sigma(p, t)$ and $B_\Sigma(x_i, R/4)$ are pairwise disjoint. Then, by the local behavior of μ_p^Σ, there exists a constant $c > 1$, independent of t, such that

$$\frac{1}{c} e^{-h(g)t} \leq \mu_p^\Sigma(pr_p^\Sigma(B_\Sigma(x_i, R))) \leq ce^{-h(g)t}.$$

It is clear that $A_p^{T,\varepsilon} \subset \bigcup_{i \in I} pr_p^\Sigma(B_\Sigma(x_i, R))$ and therefore,

$$d \leq \mu_p^\Sigma\Big(\bigcup_{i \in I} pr_p^\Sigma(B_\Sigma(x_i, R))\Big) \leq \sum_{i \in I} \mu_p^\Sigma(pr_p^\Sigma(B_\Sigma(x_i, R))) \leq c\,\mathrm{Card}(I)e^{-h(g)t}.$$

Since $\mathbb{H}^3/\Gamma$ is convex cocompact, $C_Q(\Lambda)/\Gamma$ is compact, where $C_Q(\Lambda)$ is the Q neighborhood of the convex core of Λ. Hence for any $Q > 0$,

$$\delta(\Gamma) = \lim_{R \to \infty} \mathrm{Vol}(B_{\mathbb{H}^3}(o, R) \cap C_Q(\Lambda)).$$

Now take Q sufficiently large such that Σ is inside $C_Q(\Lambda)$. There exists K such that $B_\Sigma(x_i, R/4) \subset B_{\mathbb{H}^3}(x_i, R + K) \cap C_Q(\Lambda)$.

From the definition of the set $A_p^{T,\varepsilon}$, we then have that the disjoint union

$$\bigcup_{i \in I} B_\Sigma(x_i, R/4) \subset B_{\mathbb{H}^3}(p, t(I_{\mu_{BM}}(\Sigma, \mathbb{H}^3) + \varepsilon) + R + K) \cap C_Q(\Lambda).$$

It follows that

$$\begin{aligned} e^{h(g)t} &\leq \frac{c}{d}\,\mathrm{Card}(I) \leq \frac{c}{dV} \sum_{i \in I} \mathrm{Vol}_{\mathbb{H}^3}(B_{\mathbb{H}^3}(x_i, R/4)) \cap C_Q(\Lambda)) \\ &\leq \frac{c}{dV} \mathrm{Vol}_{\mathbb{H}^3}\big(B_{\mathbb{H}^3}(p, t(I_{\mu_{BM}}(\Sigma, \mathbb{H}^3) + \varepsilon) + R + K) \cap C_Q(\Lambda)\big). \end{aligned}$$

Hence,

$$h(g) \leq \frac{1}{t}\Big(\log \frac{c}{dV} + \log \mathrm{Vol}_{\mathbb{H}^3}(B_{\mathbb{H}^3}(p, t(I_{\mu_{BM}}(\Sigma, \mathbb{H}^3) + \varepsilon) + R + K) \cap C_Q(\Lambda))\Big).$$

Taking the limit $t \to \infty$, we get

$$h(g) \leq (I_{\mu_{BM}}(\Sigma, \mathbb{H}^3) + \varepsilon)\delta(\Gamma),$$

which concludes the proof since ε is arbitrary. □

For the proof of the equality case in Theorem 1.3 we will use the result equivalent to [Knieper 1995, Corollary 3.6] in our context, that is:

Lemma 3.10 [Knieper 1995]. *Letting $p \in \Sigma$ and μ_p^g be the Patterson–Sullivan measure with respect to p and g, there exists a constant L such that for μ_p^g-almost all $\xi \in \partial\Sigma$ there is a sequence $t_n \to \infty$ such that*

$$|d(p, \pi\phi_{t_n}^{\Sigma} v_p(\xi)) - I_{\mu_{BM}}(\Sigma, \mathbb{H}^3)t_n| \leq L.$$

Proof. It follows from Lemma 3.5 of [Knieper 1995], that our lemma is true provided there exists a constant $C > 0$ such that, for all $t_1, t_2 > 0$ and all $v \in T^1\Sigma$,

$$a(v, t_1) + a(\phi_{t_1}^{\Sigma} v, t_2) \leq C + a(v, t_1 + t_2).$$

Let $v \in T^1\Sigma$ and c_v^{Σ} be the geodesic on Σ directed by v. Recall that there exists C_1 such that the $\mathbb{H}^3$-geodesic from $\pi(v)$ to $c_v^{\Sigma}(t_1 + t_2)$ is at bounded distance C_1 of $c_v^{\Sigma}(t_1 + t_2)$, independent of t_1 and t_2. The $\mathbb{H}^3$-geodesic from p to $c_v^{\Sigma}(t_1)$ and the one from $c_v^{\Sigma}(t_1)$ to $c_v^{\Sigma}(t_1 + t_2)$ are also at bounded distance C_1 of c_v^{Σ}. This implies the desired property with $C = 2C_1$. □

Proof of the equality case in 1.3. Suppose that $h(g) = I_{\mu_{BM}}(\Sigma, \mathbb{H}^3)\delta(\Gamma)$. Choose a point $p \in \Sigma$ and $\xi \in \Lambda$, set $y_n := \pi\phi_{t_n}^{\Sigma} v_p(\xi)$. From the above lemma, for μ_p^{Σ}-almost all ξ we have

$$|d(p, y_n) - I_{\mu_{BM}}(\Sigma, \mathbb{H}^3)t_n| \leq L.$$

Setting a fixed constant, $R > 0$, by the local property of the Patterson–Sullivan measure on $\mathbb{H}^3$, there is c_1 such that

$$\frac{1}{c_1} e^{-\delta(\Gamma)d(p,y_n)} \leq \mu_p^{\mathbb{H}^3}(pr_{\mathbb{H}^3} B_{\mathbb{H}^3}(y_n, R)) \leq c_1 e^{-\delta(\Gamma)d(p,y_n)},$$

and by Theorem 3.3,

$$pr_{\mathbb{H}^3}(B_{\mathbb{H}^3}(x, R - C)) \cap \Lambda \subset pr_{\Sigma}(B_{\mathbb{H}^3}(x, R) \cap \Sigma) \subset pr_{\mathbb{H}^3}(B_{\mathbb{H}^3}(x, R + C)).$$

Hence there is a constant c_2 such that

$$\frac{1}{c_2} e^{-\delta(\Gamma)d(p,y_n)} \leq \mu_p^{\mathbb{H}^3}(pr_{\Sigma} B_{\mathbb{H}^3}(y_n, R) \cap \Sigma) \leq c_1 e^{-\delta(\Gamma)d(p,y_n)}.$$

By the local property of the Patterson–Sullivan measure on Σ, there is c_3 such that

$$\frac{1}{c_3} e^{-h(\Sigma)d_{\Sigma}(p,y_n)} \leq \mu_p^{\Sigma}(pr_{\Sigma} B_{\Sigma}(y_n, R)) \leq c_3 e^{-h(\Sigma)d_{\Sigma}(p,y_n)},$$

and by Theorem 3.3,

$$pr_{\Sigma}(B_{\Sigma}(x, R - C)) \subset pr_{\Sigma}(B_{\mathbb{H}^3}(x, R) \cap \Sigma) \subset pr_{\Sigma}(B_{\Sigma}(x, R + C)).$$

Hence there is c_4 such that

$$\frac{1}{c_4} e^{-h(\Sigma)d_{\Sigma}(p,y_n)} \leq \mu_p^{\Sigma}(pr_{\Sigma} B_{\mathbb{H}^3}(y_n, R) \cap \Sigma) \leq c_4 e^{-h(\Sigma)d_{\Sigma}(p,y_n)}.$$

From the choice of y_n and since $h(\Sigma) = I_{\mu_{BM}}(\Sigma, \mathbb{H}^3)\delta(\Gamma)$,

$$e^{-L} e^{-\delta(\Gamma) d(p, y_n)} \leq e^{-h(g) d_\Sigma(p, y_n)} \leq e^{L} e^{-\delta(\Gamma) d(p, y_n)}.$$

Hence there is $c_5 > 0$ such that

$$\frac{1}{c_5} e^{-\delta(\Gamma) d(p, y_n)} \leq \mu_p^\Sigma(pr_\Sigma B_{\mathbb{H}^3}(y_n, R) \cap \Sigma) \leq c_5 e^{-\delta(\Gamma) d(p, y_n)}.$$

Finally we have a constant c_6 such that

$$c_6 \leq \frac{\mu_p^\Sigma(pr_\Sigma B_{\mathbb{H}^3}(y_n, R) \cap \Sigma)}{\mu_p^{\mathbb{H}^3}(pr_\Sigma B_{\mathbb{H}^3}(y_n, R) \cap \Sigma)} \leq c_6.$$

Since $pr_\Sigma(B_{\mathbb{H}^3}(y_n, R) \cap \Sigma) \to \xi$, the measures μ_p^Σ and $\mu_p^{\mathbb{H}^3}$ are equivalent. Theorem 3.5 concludes the proof. □

We finish this article with the proof of Corollary 1.4:

Corollary 1.4. *Under the assumptions of Theorem 1.3 we have*

$$h(\Sigma) \leq \delta(\Gamma, \mathbb{H}^3),$$

with equality if and only if Γ *is fuchsian and* Σ *is the totally geodesic hyperbolic plane, preserved by* Γ.

Proof. The inequality is obvious. Suppose the equality occurs. Then by Theorem 1.3, we have that the length spectrum is proportional to the one of $\mathbb{H}^3/\Gamma$ and moreover that $I(\Sigma, M) = 1$. In other words the two length spectra are equal.

Since Σ is embedded in $\mathbb{H}^3$, we can prove that the equality between the spectra implies that Σ is totally geodesic by the following argument:

Let $\gamma \in \Gamma$, and consider A its axis in Σ. Then for all $p \in A$, we have

$$\ell_\Sigma(\gamma) = d_\Sigma(\gamma p, p) \geq d_{\mathbb{H}^3}(\gamma p, p) \geq \ell_{\mathbb{H}^3}(\gamma).$$

Since the two spectra are equal, these inequalities are equalities. In particular, it implies that p lies in the axis of γ in $\mathbb{H}^3$. Therefore A is a geodesic of $\mathbb{H}^3$.

Let c be the closed geodesic on Σ represented by g and consider c' any geodesic that intersects c. Let g' be a representative of this closed geodesic such that the axis A' of g' on Σ intersects A. By similar computations as before, we see that A' is a geodesic of $\mathbb{H}^3$.

Since the two geodesics intersect, the endpoints of A and A' are cocyclic on the boundary of $\mathbb{H}^3$, and in particular bound a copy of $\mathbb{H}^2$ inside $\mathbb{H}^3$. By similar arguments for any element $g \in \Gamma$ such that its axis A_g intersects A and A' we see that A_g is a geodesic of $\mathbb{H}^3$ and therefore that A_g is included in the copy of $\mathbb{H}^2$. This last fact implies that Σ is included, therefore equal, to this copy of $\mathbb{H}^2$ and finishes the proof of the corollary. □

3C1. *A remark on length spectrum rigidity.* As we said in the introduction, the proof of the last corollary raises the following question: If a quasi-Fuchsian has the same length spectrum as a negatively curved surface, is it Fuchsian? Or more generally, if the two length spectra are proportional does it imply that it is Fuchsian? The latter question seems to be unanswered even if we suppose that the surface has constant negative curvature equal -1, and the problem in general seems to be quite hard.

We answer the case of constant negative curvature:

Theorem 1.5. *Let M be a quasi-Fuchsian manifold and Σ a hyperbolic (in the sense that it has constant curvature -1) surface. Suppose that M and Σ have proportional length spectrum (i.e., there exists $k \in \mathbb{R}^+$ such that for all $\gamma \in \Gamma$, $\ell_M(\gamma) = k\ell_\Sigma(\gamma)$), then M is Fuchsian, $k = 1$ and Σ is isometric to the totally geodesic surface in M.*

In this case we cannot use the entropy argument that is used when we suppose the equality of the two spectra. Our proof is inspired by the work of F. Dal'bo and I. Kim [2000] and based on the following theorem of Benoist:

Theorem 3.11 [Benoist 1997]. *Let G be a semisimple linear connected Lie group. Let $\Gamma < G$ be a Zariski dense subgroup. Then the limit cone is convex with nonempty interior.*

The limit cone is the smallest closed cone of a Cartan subspace of $\mathfrak{g}$ containing $\log(\lambda(\Gamma))$ where $\lambda(\gamma)$ is the Jordan projection.

Proof of Theorem 1.5. Consider Γ a surface group and ρ_{QF} a quasi-Fuchsian representation into $\mathrm{PSL}_2(\mathbb{C})$ and ρ_0 a Teichmüller representation in $\mathrm{PSL}_2(\mathbb{R})$. Consider the diagonal representation,

$$\rho = (\rho_{QF}, \rho_0) : \Gamma \to \mathrm{PSL}_2(\mathbb{C}) \times \mathrm{PSL}_2(\mathbb{R}).$$

The group $\mathrm{PSL}_2(\mathbb{C}) \times \mathrm{PSL}_2(\mathbb{R})$ is a semisimple linear connected Lie group of rank 2. The Jordan projection of an element (γ_1, γ_2) is given by $(\ell_{\mathbb{H}^3}(\gamma_1), \ell_{\mathbb{H}^2}(\gamma_2))$ where ℓ_X is the translation length in X.

Therefore if the two representations have proportional length spectra, then the limit cone of $\rho(\Gamma)$ is a line, in particular it has empty interior. Using Benoist's theorem we conclude that $\rho(\Gamma)$ is not Zariski dense, which implies that M is Fuchsian. Therefore the length spectrum of Σ is k times the length spectrum of the hyperbolic surface $\Sigma_0 = \mathbb{H}^2/\rho(\Gamma)$. By Otal's theorem [1990] we get

$$(\Sigma, g) = (\Sigma_0, k^2 g_{\mathbb{H}}),$$

hence since Σ is hyperbolic, we have $k = 1$ and $\Sigma = \Sigma_0$. □

Acknowledgements. We want to thank Maxime Wolff for his help in the proof of Theorem 1.1, and the referee for useful comments concerning rigidity questions.

References

[Benoist 1997] Y. Benoist, "Propriétés asymptotiques des groupes linéaires", *Geom. Funct. Anal.* **7**:1 (1997), 1–47. MR Zbl

[Besson et al. 1995] G. Besson, G. Courtois, and S. Gallot, "Entropies et rigidités des espaces localement symétriques de courbure strictement négative", *Geom. Funct. Anal.* **5**:5 (1995), 731–799. MR Zbl

[Bowen 1979] R. Bowen, "Hausdorff dimension of quasicircles", *Inst. Hautes Études Sci. Publ. Math.* **50** (1979), 11–25. MR Zbl

[Bridgeman and Taylor 2000] M. Bridgeman and E. C. Taylor, "Length distortion and the Hausdorff dimension of limit sets", *Amer. J. Math.* **122**:3 (2000), 465–482. MR Zbl

[Dal'Bo and Kim 2000] F. Dal'Bo and I. Kim, "A criterion of conjugacy for Zariski dense subgroups", *C. R. Acad. Sci. Paris Sér. I Math.* **330**:8 (2000), 647–650. MR Zbl

[Ghys and de la Harpe 1990] E. Ghys and P. de la Harpe, "Espaces métriques hyperboliques", pp. 27–45 in *Sur les groupes hyperboliques d'après Mikhael Gromov* ((Bern, 1988)), Progr. Math. **83**, Birkhäuser, Boston, 1990. MR

[Hamenstädt 2002] U. Hamenstädt, "Ergodic properties of function groups", *Geom. Dedicata* **93** (2002), 163–176. MR Zbl

[Kifer 1990] Y. Kifer, "Large deviations in dynamical systems and stochastic processes", *Trans. Amer. Math. Soc.* **321**:2 (1990), 505–524. MR Zbl

[Kingman 1973] J. F. C. Kingman, "Subadditive ergodic theory", *Ann. Probability* **1** (1973), 883–909. MR Zbl

[Knieper 1995] G. Knieper, "Volume growth, entropy and the geodesic stretch", *Math. Res. Lett.* **2**:1 (1995), 39–58. MR Zbl

[Nicholls 1989] P. J. Nicholls, *The ergodic theory of discrete groups*, London Mathematical Society Lecture Note Series **143**, Cambridge University Press, 1989. MR Zbl

[Otal 1990] J.-P. Otal, "Le spectre marqué des longueurs des surfaces à courbure négative", *Ann. of Math.* (2) **131**:1 (1990), 151–162. MR Zbl

[Patterson 1976] S. J. Patterson, "The limit set of a Fuchsian group", *Acta Math.* **136**:3-4 (1976), 241–273. MR Zbl

[Paulin et al. 2015] F. Paulin, M. Pollicott, and B. Schapira, *Equilibrium states in negative curvature*, Astérisque **373**, Société Mathématique de France, Paris, 2015. MR Zbl

[Quint 2006] J. F. Quint, "An overview of Patterson–Sullivan theory", workshop on the barycenter method, Forschungsinstitut für Mathematik , Zurich, 2006, available at https://tinyurl.com/quint-pdf.

[Roblin 2003] T. Roblin, *Ergodicité et équidistribution en courbure négative*, Mém. Soc. Math. Fr. **95**, Société Mathématique de France, Paris, 2003. MR Zbl

[Sullivan 1979] D. Sullivan, "The density at infinity of a discrete group of hyperbolic motions", *Inst. Hautes Études Sci. Publ. Math.* 50 (1979), 171–202. MR Zbl

Received February 2, 2016. Revised November 14, 2017.

OLIVIER GLORIEUX
UNIVERSITY OF LUXEMBOURG
ESCH-SUR-ALZETTE
LUXEMBOURG
olivier.glrx@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 294, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.294.401

SMOOTH SCHUBERT VARIETIES AND GENERALIZED SCHUBERT POLYNOMIALS IN ALGEBRAIC COBORDISM OF GRASSMANNIANS

JENS HORNBOSTEL AND NICOLAS PERRIN

We provide several ingredients towards a generalization of the Littlewood–Richardson rule from Chow groups to algebraic cobordism. In particular, we prove a simple product formula for multiplying classes of smooth Schubert varieties with any Bott–Samelson class in algebraic cobordism of Grassmannians. We also establish some results for generalized Schubert polynomials for hyperbolic formal group laws.

1. Introduction

Throughout the article, we fix an algebraically closed base field k with $\mathrm{char}(\mathsf{k}) = 0$. Recall that for G a reductive group over k and P a parabolic subgroup of G, there exists a Borel-type presentation of the algebraic cobordism ring $\Omega^*(G/P)$ for the homogeneous space G/P; see [Hornbostel and Kiritchenko 2011; Hudson and Matsumura 2016]. For a smooth projective variety X over k, we refer to [Levine and Morel 2007; Levine and Pandharipande 2009] for the foundations on $\Omega^*(X)$.

In this article, we adopt an alternative, more geometric point of view. Namely, it is known that an additive basis of any of these cobordism rings may be described via geometric generators, using resolutions of Schubert varieties; see below. Schubert calculus consists in multiplying these basis elements. One of the new features when passing from Chow groups to cobordism is the need to resolve the singularities of Schubert varieties. There are therefore many possible bases since a given basis element depends on the choice of a resolution of a Schubert variety. In this paper we shall mostly consider Bott–Samelson resolutions. Let us mention that some formulas for the multiplication with divisor classes are already available, see [Calmès et al. 2013; Hornbostel and Kiritchenko 2011], and that in the recent preprints of Hudson and Matsumura [2016; 2017], Giambelli-type formulas are obtained for special

Perrin was supported by a public grant as part of the Investissement d'avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH.
MSC2010: 14F43, 14M15, 14N15, 19L41, 55N22.
Keywords: Schubert calculus, cobordism, Grassmannian, generalized cohomology.

classes and for a group G of type A. There are several other recent preprints on related questions; see, e.g., [Lenart and Zainoulline 2017].

We also focus on groups of type A. In the first part we consider the classes of smooth Schubert varieties in Grassmannians and prove a formula for multiplying the class of a smooth Schubert variety with the class of any Bott–Samelson resolution. Several years ago, Buch [2002] achieved a beautiful generalization of the classical Littlewood–Richardson rule for K-theory instead of Chow groups, building on previous work of Lascoux and Schutzenberger, Fomin and Kirillov and others. In the language of formal group laws (FGL), Buch has generalized the Littlewood–Richardson rule from the additive FGL to the multiplicative FGL. In the second part, we analyse the work of Fomin and Kirillov [1996a; 1996b] used by Buch, and generalize parts of it to other formal group laws. One might hope that ultimately this will be part of a Littlewood–Richardson rule for the universal case, that is, a complete Schubert calculus for algebraic cobordism of Grassmannians.

Recall [Levine and Morel 2007] that algebraic cobordism is the universal oriented algebraic cohomology theory on smooth varieties over k. Its coefficient ring is the Lazard ring $\mathbb{L}$; see [Lazard 1955]. For any homogeneous space $X = G/P$ with G reductive and P a parabolic subgroup of G, we have a cellular decomposition of X given by the B-orbits ($B \subset P$ a Borel subgroup of G) called Schubert cells and denoted by $(\mathring{X}_w)_{w \in W^P}$, where W^P is a subset of the Weyl group W. Choosing resolutions $\widetilde{X}_w \to X_w$ of the closures X_w of $\mathring{X}_w$ defines an additive basis of $\Omega^*(X)$; see [Hornbostel and Kiritchenko 2011, Theorem 2.5]. Schubert calculus aims at understanding the product in terms of these basis elements.

Write $X = \mathrm{Gr}(k, n)$ for the Grassmannian variety of k-dimensional linear subspaces in k^n. This is a homogeneous space of the form G/P with $G = \mathrm{GL}_n(\mathsf{k})$ and P a maximal parabolic subgroup of G. In the first part of the article, we prove some simple product formulas in $\Omega^*(X)$. For Grassmannians, there is another indexing set for Schubert cells and their closures in terms of partitions, and we shall use this notation in the Grassmannian case. In the following statement, λ is a partition associated to a Schubert variety X_λ, that is, the closure of the Schubert cell $\mathring{X}_\lambda$ (see Section 2A). Recall also that for the Grassmannian X, all Bott–Samelson resolutions of the Schubert variety X_λ are isomorphic over X. We denote by $\widetilde{X}_\lambda$ this unique Bott–Samelson resolution. Finally, recall that any smooth Schubert variety in X is of the form X_{b^a} with b^a the partition with a parts of size b.

Before stating the main result of Section 2, recall the definition of the dual partition (see Section 2A for more details): for a partition λ contained in the $k \times (n-k)$ rectangle R, we denote by $\lambda^\vee$ the dual partition obtained by taking the complement of λ in R. For a partition μ in the $a \times b$ rectangle, we write $\mu^{\vee z}$ for its dual partition in the $a \times b$ rectangle.

Theorem 1.1 (Corollary 2.15). *Let $\lambda \in \mathcal{P}(k,n)$. Then in $\Omega^*(X)$, we have*

$$[X_{b^a}]\cdot[\widetilde{X}_\lambda] = \begin{cases} [\widetilde{X}_{(\lambda^\vee)^\vee Z}] & \text{for } \lambda \geq (b^a)^\vee, \\ 0 & \text{for } \lambda \not\geq (b^a)^\vee. \end{cases}$$

Note that for Chow groups or for K-theory, the above results are well known and follow from the Pieri formulas (see for example [Manivel 2001] for the Chow group case and [Buch 2002] for K-theory, by which we always mean K_0).

Note also that there are other natural resolutions of Schubert varieties considered in the literature, such as Zelevinskiĭ's resolutions [1983]. We believe that for those resolutions (which contain as a special case the resolutions considered in the cobordism Giambelli formulas of [Hudson and Matsumura 2016]) similar formulas should exist for the multiplication with the class of a smooth Schubert variety.

In the second part (Sections 3 and 4), inspired by Buch's method for giving a Littlewood–Richardson rule for K-theory, we have a closer look at generalized Schubert polynomials for cobordism. Let us recall first that for the full flag variety $X = G/B$ with $G = \mathrm{GL}_n(\mathsf{k})$ and B a Borel subgroup, there is a Borel-type presentation of the cobordism ring; see [Hornbostel and Kiritchenko 2011, Theorem 1.1]:

Theorem 1.2. *There exists an isomorphism $\Omega^*(X) \simeq \mathbb{L}[x_1,\dots,x_n]/S$, where* $\deg(x_i) = 1$ *for all $i \in [1,n]$ and S is the ideal generated by homogeneous symmetric polynomials of positive degree.*

In particular, given a Schubert variety X_w and a Bott–Samelson resolution $\widetilde{X}_{\underline{w}} \to X_w$ (here $\underline{w}$ is a reduced expression of the permutation w), we may write the class $[\widetilde{X}_{\underline{w}}] \in \Omega^*(X)$ as a polynomial $\mathfrak{L}_{\underline{w}}$ in the $(x_i)_{i\in[1,n]}$. Fomin and Kirillov [1996a; 1996b] gave a very nice description of such polynomials for the K-theory case, and [Buch 2002] builds on these results. In Section 3, we compare the generalized Schubert polynomials for cobordism with those for K-theory (called *Grothendieck polynomials*); see Corollary 3.15. For this, we have to restrict to hyperbolic formal group laws, that is, to elliptic cohomology. Choosing a suitable generalization of the Hecke algebra, we are also able to generalize the main theorem of [Fomin and Kirillov 1996a] from K-theory to elliptic cohomology; see Theorem 3.13.

In the last section, we combine techniques and results from Sections 2 and 3 to compute some explicit generalized Schubert polynomials. In particular, we show that some of the smooth Schubert varieties satisfy a certain symmetry; see Corollary 4.3. For generalized Schubert polynomials associated to other cells, this is no longer true already when looking at Gr(2, 4); see Proposition 4.5.

We have tried to present the first two parts in a way that they can be read essentially independently of each other. However, we emphasize that they both are partial solutions to the quest of a Schubert calculus for arbitrary orientable cohomology theories. Both parts reflect that for general formal group laws with operators not satisfying the naive braid relation, Schubert cells will lead to different elements in the corresponding generalized cohomology theory. On the geometric side, we have different resolutions of a given Schubert variety, and on the combinatorial side we have different reduced words for a given permutation. We hope that forthcoming work will combine these two aspects, leading to a better understanding of general Schubert calculus.

2. Product with smooth Schubert varieties

2A. ***Notation.*** Let $X = \mathrm{Gr}(k, n)$ be the Grassmannian of k-dimensional subspaces in $E = \mathsf{k}^n$. Denote by $(e_i)_{i\in[1,n]}$ the canonical basis of k^n. Denote by B the subgroup of upper-triangular matrices in $\mathrm{GL}_n(\mathsf{k})$, by B^- the subgroup of lower-triangular matrices and by $T = B \cap B^-$ the subgroup of diagonal matrices. For any subset $I \subset [1, n]$ write E_I for the span $\langle e_i \mid i \in I\rangle$. Set $E_i = E_{[1,i]}$ and $E^i = E_{[n+1-i,n]}$ for $i \in [1, n]$.

Call any nonincreasing sequence $\lambda = (\lambda_i)_{i\geq 1}$ of nonnegative integers a partition. The length of a partition is $\ell(\lambda) = \max\{i \mid \lambda_i \neq 0\}$. For λ of length k, we identify λ with its first k parts, i.e., with $(\lambda_i)_{i\in[1,k]}$. The weight of λ is $|\lambda| = \sum_i \lambda_i$. We will also use the pictorial description via Young diagrams, which are left-aligned arrays of $|\lambda|$ boxes with λ_i boxes on the i-th line for all $i \geq 1$. A partition λ fits in the $k \times (n-k)$ rectangle if its Young diagram does or equivalently if $\ell(\lambda) \leq k$ and $\lambda_1 \leq n-k$. Denote by $\mathcal{P}(k, n)$ the set of partitions fitting in the $k \times (n-k)$ rectangle. For $\lambda \in \mathcal{P}(k, n)$ denote by $\lambda^\vee \in \mathcal{P}(k, n)$ its dual partition defined by $\lambda^\vee_i = n-k-\lambda_{k+1-i}$ for $i \in [1, k]$. We have $|\lambda^\vee| = k(n-k) - |\lambda|$. Define $\lambda \leq \mu$ if $\lambda_i \leq \mu_i$ for all i.

Recall the Bruhat decomposition: the B-orbits $(\mathring{X}_\lambda)_{\lambda\in\mathcal{P}(k,n)}$ form a cellular decomposition of X. The same result holds for the B^--orbits $(\mathring{X}^\lambda)_{\lambda\in\mathcal{P}(k,n)}$. Indeed these orbits are isomorphic to affine spaces: $\mathring{X}_w \simeq \mathbb{A}_{\mathsf{k}}^{|\lambda|}$ and $\mathring{X}^\lambda \simeq \mathbb{A}_{\mathsf{k}}^{\dim X-|\lambda|}$. This can easily be deduced from their explicit descriptions:

$$\mathring{X}_\lambda = \{V_k \in X \mid \dim(V_k \cap E_{i+\lambda_{k+1-i}}) = i \text{ for all } i \in [1, k]\},$$
$$\mathring{X}^\lambda = \{V_k \in X \mid \dim(V_k \cap E^{i+n-k-\lambda_i}) = i \text{ for all } i \in [1, k]\}.$$

Note that with this definition we have $\mathring{X}^{\lambda^\vee} = w_X \cdot \mathring{X}_\lambda$, where w_X is the matrix permutation associated to the permutation $i \mapsto n+1-i$ of $[1, n]$. Denote by X_λ the closure of $\mathring{X}_\lambda$ and by X^λ the closure of $\mathring{X}^\lambda$. We have

$$X_\lambda = \{V_k \in X \mid \dim(V_k \cap E_{i+\lambda_{k+1-i}}) \geq i \text{ for all } i \in [1, k]\},$$
$$X^\lambda = \{V_k \in X \mid \dim(V_k \cap E^{i+n-k-\lambda_i}) \geq i \text{ for all } i \in [1, k]\}.$$

Inclusion induces the order on partitions: $X_\lambda \subset X_\mu \iff \lambda \leq \mu$.

Remark 2.1. The bases $([X_\lambda])_{\lambda\in\mathcal{P}(k,n)}$ and $([X^\lambda])_{\lambda\in\mathcal{P}(k,n)}$ are dual bases in $CH^*(X)$; see [Manivel 2001, Proposition 3.2.7]. Since $X^{\lambda^\vee} = w_X \cdot X_\lambda$ we see that $([X_\lambda])_{\lambda\in\mathcal{P}(k,n)}$ and $([X_{\lambda^\vee}])_{\lambda\in\mathcal{P}(k,n)}$ are also dual bases. Note that this is no longer true in K-theory.

2B. ***Smooth Schubert varieties, Bott–Samelson resolution and cobordism.*** The smooth Schubert varieties in X are sub-Grassmannians; see for example [Lakshmibai and Brown 2015, Theorem 6.4.2] or [Gasharov and Reiner 2002, Theorem 1.1], and [Brion and Polo 1999] or [Perrin 2009] for more details on the singular locus and the type of singularities. The partitions corresponding to these smooth Schubert varieties are of the form $\lambda = (\lambda_1, \dots, \lambda_k)$ with $\lambda_i = b$ for $i \in [1, a]$ and $\lambda_i = 0$ for $i > a$ for some integers $a \in [1, k]$ and $b \in [1, n-k]$. Denote this partition by $\lambda = b^a$. As a variety we have

$$X_{b^a} = \{V_k \in X \mid E_{k-a} \subset V_k \subset E_{k+b}\},$$
$$X^{b^{a\vee}} = \{V_k \in X \mid E^{k-a} \subset V_k \subset E^{k+b}\}.$$

Moreover we have $X_{b^a} \simeq \mathrm{Gr}(a, a+b) \simeq X^{b^{a\vee}}$.

As already mentioned, Schubert varieties are in general singular. There exist several resolutions of singularities. We recall here the Bott–Samelson resolutions of Schubert varieties, which were first introduced by Bott and Samelson [1958], as well as by Hansen [1973] and Demazure [1974] for full flag varieties. These constructions and their properties carry over easily to partial flags $G/P = \mathrm{Gr}(k, n)$. See, e.g., [Fulton 1998; Lakshmibai and Brown 2015] for more details. We give here an explicit description of these resolutions in the spirit of configuration spaces; see [Magyar 1998] or [Perrin 2007]. Note also that for Schubert varieties in X, these resolutions are *canonical* in the sense that they do not depend on the choice of a reduced expression.

For a partition λ and a pair of integers (i, j) write $(i, j) \in \lambda$ if $i \in [1, k]$ and $j \in [1, \lambda_i]$ and $(i, j) \notin \lambda$ else. Define $V_{(i,j)} = E_{k+j-i}$ for all $(i, j) \notin \lambda$, where E_i is the zero space for $i \le 0$ and $E_i = \mathsf{k}^n = E_n$ for $i \ge n$. Define $Y_\lambda = \prod_{(i,j)\in\lambda} \mathrm{Gr}(k+j-i, n)$. Set

$$\widetilde{X}_\lambda = \Big\{(V_{(i,j)})_{(i,j)\in\lambda} \in Y_\lambda \;\Big|\; V_{(i+1,j)} \subset V_{(i,j)} \subset V_{(i,j+1)} \text{ for all } (i, j) \in \lambda\Big\}.$$

The projection $\pi_\lambda : \widetilde{X}_\lambda \to X$ defined by $\pi_\lambda((V_{(i,j)})_{(i,j)\in\lambda}) = V_{1,1}$ induces a birational morphism onto X_λ. Furthermore, one easily checks that $\widetilde{X}_\lambda$ has the structure of a tower of $\mathbb{P}^1$-bundles so that $\widetilde{X}_\lambda$ is smooth. The morphisms $\pi_\lambda : \widetilde{X}_\lambda \to X_\lambda$ are called the Bott–Samelson resolutions of X_λ.

These resolutions define classes $[\pi_\lambda : \widetilde{X}_\lambda \to X]$ in the cobordism $\Omega^*(X)$ of X. We write $[\widetilde{X}_\lambda]$ for these classes. The classes $([\widetilde{X}_\lambda])_{\lambda\in\mathcal{P}(k,n)}$ form a basis in any oriented cohomology theory and especially in cobordism:

$$\Omega^*(X) = \bigoplus_{\lambda \in \mathcal{P}(k,n)} \mathbb{L}[\widetilde{X}_\lambda],$$

where $\mathbb{L}$ is the Lazard ring; see [Hornbostel and Kiritchenko 2011].

2C. *Products in cobordism.* We want to understand the products with the classes $[X_{b^a}]$ in $\Omega^*(X)$. Note that the class $[X_{b^a}]$ is well defined without considering any resolution since $X^{b^a} \simeq \mathrm{Gr}(a, a+b)$ is smooth; hence its cobordism class is well defined.

2C1. *Sub-Grassmannians.* Let $Z = \mathrm{Gr}(a, a+b)$ be the Grassmannian of a-dimensional vector subspaces of k^{a+b}. Let $(f_i)_{i \in [1,a+b]}$ be the canonical basis of k^{a+b}. Define $F_i = \langle f_j \mid j \in [1,i] \rangle$ and $F^i = \langle f_j \mid j \in [a+b+1-i, a+b] \rangle$. For $\lambda \in \mathcal{P}(a, a+b)$ a partition contained in the $a \times b$ rectangle define the Schubert variety in Z (as above in X):

$$Z_\lambda = \{V_a \in Z \mid \dim(V_a \cap F_{i+\lambda_{a+1-i}}) \geq i \text{ for all } i \in [1,a]\},$$
$$Z^\lambda = \{V_a \in Z \mid \dim(V_a \cap F^{i+b-\lambda_i}) \geq i \text{ for all } i \in [1,a]\}.$$

If $w_Z : \mathsf{k}^{a+b} \to \mathsf{k}^{a+b}$ is the endomorphism defined by $f_i \mapsto f_{a+b+1-i}$, then $Z^\lambda = w_Z \cdot Z_{\lambda^{\vee_Z}}$ with $\mu = \lambda^{\vee_Z}$ defined by $\mu_i = b - \lambda_{a+1-i}$ for all $i \in [1,a]$.

Now define Bott–Samelson resolutions in Z. Define $W_{(i,j)} = F_{a+j-i}$ for all $(i,j) \notin \lambda$, where F_i is the zero space for $i \leq 0$ and $F_i = \mathsf{k}^{a+b} = F_{a+b}$ for $i \geq a+b$. Define $A_\lambda = \prod_{(i,j)\in\lambda} \mathrm{Gr}(a+j-i, a+b)$. Set

$$\widetilde{Z}_\lambda = \left\{ (W_{(i,j)})_{(i,j)\in\lambda} \in A_\lambda \;\middle|\; W_{(i+1,j)} \subset W_{(i,j)} \subset W_{(i,j+1)} \text{ for all } (i,j) \in \lambda \right\}.$$

The projection $\pi^Z_\lambda : \widetilde{Z}_\lambda \to Z$ defined by $\pi^Z_\lambda((W_{(i,j)})_{(i,j)\in\lambda}) = W_{1,1}$ induces a birational morphism onto Z_λ.

Embed Z in X with image $X_{(b^a)}$ as follows. Let $u : \mathsf{k}^{a+b} \to \mathsf{k}^n$ be the linear map defined by $u(f_i) = e_{k-a+i}$ for all $i \in [1, a+b]$. Note that $u(\mathsf{k}^{a+b}) = E_{[k-a+1,k+b]}$. Denote by $v : Z \to X$ the closed embedding defined by $W_a \mapsto E_{k-a} \oplus u(W_a)$.

Embed Z in X with image $X^{(b^a)^\vee}$ as follows. Let $u' : \mathsf{k}^{a+b} \to \mathsf{k}^n$ be the linear map defined by $u'(f_i) = e_{n-k-b+i}$ for all $i \in [1, a+b]$. Note that $u'(\mathsf{k}^{a+b}) = E_{[n-k-b+1,n-k+a]}$. Denote by $v' : Z \to X$ the closed embedding defined by $W_a \mapsto E^{k-a} \oplus u'(W_a)$.

2C2. *Intersection with Schubert varieties.* In this subsection we consider the classes of closed subvarieties $Y \subset X$ in Chow groups or in K-theory. To avoid introducing more notation we denote both theses classes by $[Y]$ and specify in which theory we are working. The product with the class $[X_{b^a}]$ in Chow groups or for K-theory is easy to compute.

Lemma 2.2. *Let $\lambda \in \mathcal{P}(k, n)$. We have*

$$v(Z) \cap X^\lambda = X_{b^a} \cap X^\lambda = \begin{cases} \varnothing & \text{for } \lambda \not\leq b^a, \\ v(Z^\lambda) & \text{for } \lambda \leq b^a. \end{cases}$$

Proof. Let $\mu = b^a$. As is well known, the intersection $X_\mu \cap X^\lambda$ is nonempty if and only if $\lambda \leq \mu$. Assume this holds. We also know that $X_\mu \cap X^\lambda$ is a Richardson variety thus reduced, irreducible of dimension $|\mu| - |\lambda|$. Since Z^λ has dimension $|\mu| - |\lambda|$ it is enough to prove the inclusion $v(Z^\lambda) \subset X_{b^a} \cap X^\lambda$. By construction, we have $v(Z) = X_{b^a}$ thus $v(Z^\lambda) \subset X_{b^a}$. We prove the inclusion $v(Z^\lambda) \subset X^\lambda$. Recall the definition

$$X^\lambda = \{V_k \in X \mid \dim(V_k \cap E^{i+n-k-\lambda_i}) \geq i \text{ for all } i \in [1, k]\}.$$

Since λ is contained in the $a \times b$ rectangle, we have $\ell(\lambda) \leq a$; thus the conditions $\dim(V_k \cap E^{i+n-k-\lambda_i}) \geq i$ for $i > a$ become $\dim(V_k \cap E^{i+n-k}) \geq i$ and are trivially satisfied. We need to check the conditions $\dim(V_k \cap E^{i+n-k-\lambda_i}) \geq i$ for $i \in [1, a]$ and $V_k = v(W_a)$ with $W_a \in Z^\lambda$. For all $i \in [1, a]$, we have $\dim(V_a \cap F^{i+b-\lambda_i}) \geq i$. Applying v we get the inequality

$$\dim\bigl(v(V_a) \cap v(F^{i+b-\lambda_i}) \cap E_{[k-a+1,k+b]}\bigr) \geq i.$$

But

$$v(F^{i+b-\lambda_i} \cap E_{[k-a+1,k+b]}) = E_{[k+1-\lambda_i-i,k+b]} \subset E_{[k+1-\lambda_i-i,n]} = E^{i+n-k-\lambda_i}.$$

In particular $\dim(v(V_a) \cap E^{i+n-k-\lambda_i}) \geq i$ for $i \in [1, a]$ proving the result. $\square$

Remark that $v(w_Z(F^i)) = E_{k-a} \oplus u(F_i) = E_i$; thus for $\lambda \in \mathcal{P}(a, a+b)$, we have $v(Z_\lambda) = X_\lambda$. In particular, we have $v(w_Z \cdot Z^\lambda) = v(Z_{\lambda^{\vee_Z}}) = X_{\lambda^{\vee_Z}}$. Consider k^{a+b} as a subspace of k^n via the embedding u and let w^Z be the endomorphism of k^n obtained by extending w_Z with the identity on the complement $\langle e_i \mid i \notin [k-a, k+b]\rangle$. We have $w^Z \circ v = v \circ w_Z$.

Corollary 2.3. *Let $\lambda \in \mathcal{P}(a, a+b)$. We have*

$$v(Z) \cap X^\lambda = X_{b^a} \cap X^\lambda = \begin{cases} \varnothing & \text{for } \lambda \not\leq b^a, \\ w^Z \cdot X_{\lambda^{\vee_Z}} & \text{for } \lambda \leq b^a. \end{cases}$$

Corollary 2.4. *Let $\lambda \in \mathcal{P}(a, a+b)$. We have*

$$X_\lambda \cap v'(Z) = X_\lambda \cap X^{b^{a\vee}} = \begin{cases} \varnothing & \text{for } \lambda \not\geq (b^a)^\vee, \\ w_X w^Z \cdot X_{(\lambda^\vee)^{\vee_Z}} & \text{for } \lambda \geq (b^a)^\vee. \end{cases}$$

Proof. Set $\mu = \lambda^\vee$, apply Corollary 2.3 to μ and multiply with w_X. $\square$

Corollary 2.5. *Let $\lambda \in \mathcal{P}(a, a+b)$. In $CH^*(X)$, we have*

$$[X_\lambda] \cup [X_{b^a}] = \begin{cases} [X_{(\lambda^\vee)^{\vee_Z}}] & \text{for } \lambda \geq (b^a)^\vee, \\ 0 & \text{for } \lambda \not\geq (b^a)^\vee. \end{cases}$$

Remark 2.6. The same result holds for K-theory; see [Buch 2002].

Our aim is to generalize the above results to Bott–Samelson resolutions and to cobordism. For this, the dual point of view of Corollary 2.4 is better suited.

2C3. *Fiber product.* Let μ be a partition in the $a \times b$ rectangle and let $\mu' = (\mu^{\vee z})^{\vee}$. We construct an embedding of $\widetilde{Z}_\mu \to \widetilde{X}_{\mu'}$. We denote by $v' : \mathrm{Gr}(i, a+b) \to \mathrm{Gr}(i+k-a, n)$ the embeddings induced by u' as follows: $v'(W_i) = u'(W_i) \oplus E^{k-a}$.

First remark that $\mu \le \mu'$ and that we get μ' from μ by adding $k-a$ lines (with $n-k$ boxes) and $n-k-b$ columns (with k boxes). In other words, $\mu'_i = n-k$ for $i \in [1, k-a]$ and $\mu'_i = \mu_i + n-k-b$ for $i \in [k-a+1, k]$.

Let $(W_{(i,j)})_{(i,j)\in\mu} \in \widetilde{Z}_\mu$. We define $(V_{(i,j)})_{(i,j)\in\mu'}$ as follows:

- For $i \in [1, k-a]$ and $j \in [1, n-k-b]$, set
$$V_{(i,j)} = (v'(W_{(1,1)}) \oplus E_{j-1}) \cap E_{n+1-i}.$$
- For $i \in [k-a+1, k]$ and $j \in [1, n-k-b]$, set
$$V_{(i,j)} = (v'(W_{(i-(k-a),1)}) \oplus E_{j-1}) \cap E_{n+a-k}.$$
- For $i \in [1, k-a]$ and $j \in [n-k-b+1, n-k]$, set
$$V_{(i,j)} = (v'(W_{(1,j-(n-k-b))}) \oplus E_{n-k-b}) \cap E_{n+1-i}.$$
- For $i \in [1, k-a]$ and $j \in [1, n-k-b]$, set
$$V_{(i,j)} = (v'(W_{(i-(k-a),j-(n-k-b))}) \oplus E_{n-k-b}) \cap E_{n+a-k}.$$
- For $(i, j) \notin \mu'$, set
$$V_{(i,j)} = (v'(W_{(i-(k-a),j-(n-k-b))}) \oplus E_{n-k-b}) \cap E_{n+a-k}.$$

Lemma 2.7. *We have $(V_{(i,j)})_{(i,j)\in\mu'} \in \widetilde{X}_{\mu'}$.*

Proof. Recall that $u'(\mathrm{k}^{a+b}) = E_{n-k-b,n-k+a}$, that $E^{k-a} \subset v'(W)$ and that $v'(W) \subset E^{k+b}$ for any subspace $W \subset \mathrm{k}^{a+b}$. In particular, in the above definition all sums are direct and all intersections are transverse. This implies $\dim V_{(i,j)} = k + j - i$; thus $(V_{(i,j)})_{(i,j)\in\lambda'} \in Y_{\mu'}$. For $(i, j) \notin \mu'$ we have
$$V_{(i,j)} = \big(v'(W_{(i-(k-a),j-(n-k-b))}) \oplus E_{n-k-b}\big) \cap E_{n+a-k} = E_{k+j-i}.$$
One easily proves that $V_{(i+1,j)} \subset V_{(i,j)} \subset V_{(i,j+1)}$. The result follows. □

Lemma 2.8. *The map $\varphi : \widetilde{Z}_\mu \to \widetilde{X}_{\mu'}$ is a closed embedding.*

Proof. We have $u'(W_{(i,j)}) = V_{(i+k-a,j+n-k-b)}) \cap E^{k+b}$. Since u is injective, the result follows. □

Lemma 2.9. *The map $\psi : \widetilde{Z}_\mu \to X$ defined by $(W_{(i,j)})_{(i,j)\in\mu} \mapsto V_{(1,1)}$ factors through $v'(Z)$.*

Proof. We have $V_{(1,1)} = v'(W_{(1,1)}) = u'(W_{(1,1)}) \oplus E^{k-a}$. In particular $E^{k-a} \subset V_{(1,1)} \subset E^{k+b}$. The result follows. □

Proposition 2.10. *Let $\mu \in \mathcal{P}(a, a+b)$ and consider $\widetilde{Z}_\mu$ as an X-scheme via ψ. We have $\widetilde{X}_{\mu'} \times_X v'(Z) = \widetilde{X}_{\mu'} \times_X X^{(b^a)^\vee} \simeq \widetilde{Z}_\mu$.*

Proof. We have morphisms $\varphi : \widetilde{Z}_\mu \to \widetilde{X}_{\mu'}$ and $\psi : \widetilde{Z}_\mu \to v'(Z)$ with φ a closed embedding. Furthermore the map $\pi_{\mu'} : \widetilde{X}_{\mu'} \to X$ is given by $(V_{(i,j)})_{(i,j)\in\mu'} \mapsto V_{(1,1)}$ so the composition $\pi_{\mu'} \circ \varphi$ is the map ψ. In particular we have a morphism $\varphi \times \psi : \widetilde{Z}_\mu \to \widetilde{X}_{\mu'} \times_X v'(Z)$. This is a closed embedding since φ is a closed embedding. To prove that this is an isomorphism, it is enough to prove that $\widetilde{X}_{\mu'} \times_X v'(Z)$ is irreducible and smooth of dimension $|\mu| = \dim \widetilde{Z}_\mu$. But $v'(Z) = X^{(b^a)^\vee}$ and $\widetilde{X}_{\mu'}$ are in general position. By Kleimann–Bertini [Kleiman 1974] any irreducible component is of dimension $|\mu| - \operatorname{codim}_X v'(Z) = |\mu|$. By Bertini again, the fiber product of $v'(Z)$ with the locus in $\widetilde{X}_{\mu'}$ where $\pi_{\mu'}$ is not an isomorphism, has dimension strictly less than $|\mu|$ and is therefore never an irreducible component. Now since $v'(Z) \cap X_{\mu'}$ is irreducible, the same holds for $\widetilde{X}_{\mu'} \times_X v'(Z)$. Furthermore by Bertini again this fiber product is smooth and therefore reduced. □

Corollary 2.11. *Let $\lambda \in \mathcal{P}(k, n)$. As X-schemes, we have*

$$\widetilde{X}_\lambda \times_X v'(Z) = \widetilde{X}_\lambda \times_X X^{b^a} \simeq \begin{cases} \varnothing & \text{for } \lambda \not\geq (b^a)^\vee, \\ \widetilde{Z}_\mu & \text{for } \lambda \geq (b^a)^\vee, \end{cases}$$

with $\mu = (\lambda^\vee)^{\vee_Z}$ for $\lambda \geq (b^a)^\vee$ and $\widetilde{Z}_\mu$ is considered as an X-scheme via ψ.

2C4. *Cobordism.* We construct another X-scheme isomorphism between $\widetilde{Z}_\mu$ and $w_X w^Z \cdot \widetilde{X}_\mu$. Here $\widetilde{Z}_\mu$ is an X-scheme via ψ, while $w_X w^Z \cdot \widetilde{X}_\mu$ is an X-scheme via $w_X w^Z \circ \pi_\mu$. The actions of w_X and w^Z on $\widetilde{X}_\mu$ being defined via the embedding of $\widetilde{X}_\mu$ in Y_μ and the actions on the later are given by the diagonal action on each factor (recall that Y_μ is a product of Grassmannians $\mathrm{Gr}(i, n)$ on which w_X and w^Z act).

Let $(W_{(i,j)})_{(i,j)\in\mu} \in \widetilde{Z}_\mu$. We define $(V_{(i,j)})_{(i,j)\in\mu}$ as follows. For $(i, j) \in \mu$, set $V_{(i,j)} = v'(W_{(i,j)})$. For $(i, j) \notin \mu$, set $V_{(i,j)} = w_X w^Z \cdot E_{k+j-i}$.

Lemma 2.12. *We have $(V_{(i,j)})_{(i,j)\in\mu} \in w_X w^Z \cdot \widetilde{X}_\mu$.*

Proof. For (i, j), $(i+1, j)$ and $(i, j+1)$ in μ, the conditions $V_{(i+1,j)} \subset V_{(i,j)} \subset V_{(i,j+1)}$ are clearly satisfied. We only need to check these conditions for $(i+1, j)$ or $(i, j+1)$ not in μ. But for $(i, j) \notin \mu$, we have $W_{(i,j)} = F_{a+j-i}$; thus

$$v'(W_{(i,j)}) = v'(F_{a+j-i}) = E^{k-a} \oplus E_{[n-k-b+1,n-k-b+a+j-i]} = w_X w^Z \cdot E_{k+j-i} = V_{(i,j)}$$

and the result follows. □

Proposition 2.13. *Let $\mu \in \mathcal{P}(a, a+b)$. The X-schemes $\widetilde{Z}_\mu$ (via ψ) and $w_X w^Z \cdot \widetilde{X}_\mu$ are isomorphic.*

Proof. The above morphism sending $(W_{(i,j)})_{(i,j)\in\mu} \in \widetilde{Z}_\mu$ to $(V_{(i,j)})_{(i,j)\in\mu} \in \widetilde{X}_\mu$ is a closed embedding. Since both schemes are smooth are irreducible of the same dimension, this map is an isomorphism. We need to check that the morphisms to X coincide. But the composition $\widetilde{Z}_\mu \to w_X w^Z \cdot \widetilde{X}_\mu \to X$ is given by $(W_{(i,j)})_{(i,j)\in\mu} \mapsto (V_{(i,j)})_{(i,j)\in\mu} \mapsto V_{(1,1)}$ and therefore maps $(W_{(i,j)})_{(i,j)\in\mu} \in \widetilde{Z}_\mu$ to $v'(W_{(1,1)}) = \psi(W_{(1,1)})$. It coincides with ψ. $\square$

Corollary 2.14. *Let* $\lambda \in \mathcal{P}(k,n)$. *As* X*-schemes, we have*

$$\widetilde{X}_\lambda \times_X v'(Z) = \widetilde{X}_\lambda \times_X X^{b^a} \simeq \begin{cases} \varnothing & \text{for } \lambda \not\geq (b^a)^\vee, \\ w_X w^Z \cdot \widetilde{X}_{(\lambda^\vee)^\vee z} & \text{for } \lambda \geq (b^a)^\vee. \end{cases}$$

Corollary 2.15. *Let* $\lambda \in \mathcal{P}(k,n)$. *Then in* $\Omega^*(X)$, *we have*

$$[X_{b^a}] \cdot [\widetilde{X}_\lambda] = \begin{cases} [\widetilde{X}_{(\lambda^\vee)^\vee z}] & \text{for } \lambda \geq (b^a)^\vee, \\ 0 & \text{for } \lambda \not\geq (b^a)^\vee. \end{cases}$$

Proof. The product $[X_{b^a}] \cdot [\widetilde{X}_\lambda]$ is given by pulling back the exterior product $X_{b^a} \times \widetilde{X}_\lambda \to X \times X$ along the diagonal map $\Delta : X \to X \times X$; see [Levine and Morel 2007, Remark 4.1.14]. We thus have $[X_{b^a}] \cdot [\widetilde{X}_\lambda] = \Delta^*[X_{b^a} \times \widetilde{X}_\lambda \to X \times X]$. Applying Corollary 6.5.5.1 of the same book, we get $\Delta^*[X_{b^a} \times \widetilde{X}_\lambda \to X \times X] = [X_{b^a} \times_X \widetilde{X}_\lambda]$ in $\Omega^*(X)$. $\square$

Remark 2.16. (1) These results were inspired by several similar results for other cohomology theories. In particular, the results explained in Corollary 2.5 are the classical part of Seidel symmetries [1997] in quantum cohomology. The results of Seidel are not explicit but were made explicit in [Chaput et al. 2007; 2009]. These results extend to quantum K-theory. This will be presented in a forthcoming work [Buch et al. $\geq$ 2018]. We expect the same results to be valid in quantum cobordism once the latter is defined.

(2) We expect more general results of the same type for other homogeneous spaces. These will be studied by the second author in forthcoming work.

3. Generalized Schubert polynomials and generalized Hecke algebras

Recall that classical Grothendieck polynomials are representatives of Schubert classes in Borel's presentation of K-theory. In this section, we discuss the difference between classical Grothendieck polynomials and the representatives in Borel's presentation of algebraic cobordism of Bott–Samelson resolutions of Schubert varieties. For K-theory (that is K_0), the computation of polynomial representatives for classes of Schubert varieties has been done in [Fomin and Kirillov 1996a; 1996b]. We establish a generalization of the main theorem of [Fomin and Kirillov 1996a]. Building on their work, Buch [2002] computed Littlewood–Richardson rules for K-theory.

3A. ***Divided difference operators.*** Recall that K-theory corresponds to the multiplicative formal group law. The methods of Buch and Fomin and Kirillov do not generalize to the universal formal group law, that is, to algebraic cobordism. However, we will show that they apply in a much weaker form to *hyperbolic formal group laws* (see Definition 3.6 below) since we need to impose one more relation in the Hecke algebra (see Definition 3.11 below). For $i \in [1, n-1]$, let s_i be the transposition of $[1, n]$ exchanging i and $i+1$.

Definition 3.1. Let F be a formal group law over R with inverse χ:

(1) For $i \in [1, n-1]$, define $\sigma_i \in \operatorname{End}(R[[x_1, \dots, x_n]])$ by

$$(\sigma_i f)(x_1, \dots, x_n) = f(x_{s_i(1)}, \dots, x_{s_i(n)}).$$

(2) For $i \in [1, n-1]$, define $C_i, \Delta_i \in \operatorname{End}(R[[x_1, \dots, x_n]])$ by

$$C_i = (\mathrm{Id} + \sigma_i)\frac{1}{F(x_i, \chi(x_{i+1}))} \quad \text{and} \quad \Delta_i = \frac{1}{F(x_{i+1}, \chi(x_i))}(\mathrm{Id} - \sigma_i).$$

Remark 3.2. Note that the above operators are well defined in $R[[x_1, \dots, x_n]]$ since $F(x, \chi(y))$ can be written $(x-y)g(x, y)$ with $g(x, y)$ invertible in $R[[x, y]]$.

This definition is taken from [Hornbostel and Kiritchenko 2011, p. 71] and [Calmès et al. 2013, Section 3]. When applying it to the additive formal group law, one recovers the usual definition as, e.g., in [Manivel 2001, Section 2.3.1] up to a sign (observe that $\sigma_i \circ F(x_{i+1}, \chi(x_i)) = F(x_i, \chi(x_{i+1}))$). For the multiplicative formal group law $F(x, y) = x + y + \beta xy$, the definition of C_i yields the β-DDO $\pi_i^{(\beta)}$ of [Fomin and Kirillov 1996a], which for $\beta = -1$ specializes to the isobaric DDO of [Buch 2002]. Moreover, still for the multiplicative formal group law $F(x, y) = x + y + \beta xy$, the operator Δ_i above, which equals the one of [Calmès et al. 2013, Section 3], coincides up to sign with the operator $\pi_i^{(\beta)} + \beta$ which appears in [Fomin and Kirillov 1996a, Lemma 2.5].

Recall [Bressler and Evens 1990] that the braid relations for the operators C_i only hold if the FGL is additive or multiplicative. We therefore need to keep track of reduced expressions to define generalized Schubert polynomials, which is not necessary in [Fomin and Kirillov 1996a, Definition 2.1].

3B. ***Generalized Schubert polynomials.*** The following definition generalizes both Schubert polynomials for Chow groups and Grothendieck polynomials for K-theory.

Definition 3.3. Let w be a permutation and $\underline{w}$ a reduced expression of w as product in the $(s_i)_{i \in [1, n-1]}$. Define the generalized Schubert polynomial $\mathfrak{L}_{\underline{w}}$ by induction:

(a) $\mathfrak{L}_1(x_1, \dots, x_n) = x_1^{n-1} x_2^{n-2} \cdots x_{n-1}$.

(b) $\mathfrak{L}_{\underline{w}s_i} := C_i \mathfrak{L}_{\underline{w}}$ if $\underline{w}s_i$ is a reduced word.

Note that this notation is different from the one used in [Fomin and Kirillov 1996a] and elsewhere: our $\mathfrak{L}_1$ corresponds to their $\mathfrak{L}_{w_0}$ and our $\mathfrak{L}_{\underline{w}}$ to their $\mathfrak{L}_{w_0 w}$. We decided to adopt this notation since there is a unique class for the point as well as a unique reduced expression for 1, but there is a Bott–Samelson resolution and a polynomial $\mathfrak{L}_{\underline{w}_0}$ for each reduced expression $\underline{w}_0$ of the element w_0.

For any permutation w, the Bott–Samelson resolutions $\widetilde{X}_{\underline{w}} \to X_w$ of the Schubert variety X_w are indexed by the reduced words $\underline{w}$ of w. It was proved in [Hornbostel and Kiritchenko 2011, Theorem 3.2] that the polynomial $\mathfrak{L}_{\underline{w}}$ represents the class of the resolution $\widetilde{X}_{\underline{w}} \to X_w$ in $\Omega^*(G/B)$.

Let S be the ideal in $R[[x_1, \dots, x_n]]$ generated by symmetric polynomials of positive degree. The polynomial $\mathfrak{L}_1$ corresponds to the cobordism class of a point. Modulo S, the polynomial $n!\,\mathfrak{L}_1$ has several equivalent descriptions; compare to, e.g., [Hornbostel and Kiritchenko 2011, Remark 2.7], where it differs by a scalar from D_n below.

Lemma 3.4. *Let $A^*(-)$ be an oriented cohomology theory and F its FGL.*

(a) *We have*

$$D_n := \prod_{1 \le i < j \le n} (x_i - x_j) \equiv n!\, x_1^{n-1} x_2^{n-2} \cdots x_{n-1} = n!\,\mathfrak{L}_1 \bmod S.$$

(b) *Setting $a -_F b = F(a, \chi(b))$, we have*

$$D_n \equiv D_n^F := \prod_{1 \le i < j \le n} (x_i -_F x_j) \bmod S.$$

Proof. To show (a), one first verifies that modulo S we have $\prod_{1<i\le n}(x_1 - x_i) \equiv n x_1^{n-1}$, deriving the equality $\prod_{1\le i\le n}(x - x_i) \equiv x^n$ and setting $x = x_1$. Then one shows $x_1^{n-1} p(x_2, \dots, x_n) \equiv 0$ for any symmetric nonconstant polynomial $p(x_2, \dots, x_n)$, writing $p(x_2, \dots, x_n) \equiv x_1 q(x_1, \dots, x_n)$ and using that $x_1^n \equiv 0$ modulo S. Now proceed by induction on n. The claim holds for $n = 1$. Using the factorization

$$\prod_{1\le i<j\le n} (x_i - x_j) = \prod_{1<i<j\le n} (x_i - x_j) \prod_{1<i\le n} (x_1 - x_i),$$

the claim for n follows using the induction hypothesis for $n-1$ and the above two equalities modulo S.

For (b), note that $x_i -_F x_j = 0$ if $x_j = x_i$, which implies that $x_i -_F x_j$ is divisible by $x_i - x_j$. Hence $x_i -_F x_j = (x_i - x_j) a(x_i, x_j)$ with $a(x_i, x_j) = 1 + b(x_i, x_j)$ and $b \in (x_i, x_j)$. Thus $D_n^F = D_n + D_n q(x_1, \dots, x_n)$ with $q(0, \dots, 0) = 0$. Now using part (a) and the equality $x_1^n \equiv 0 \bmod S$, we deduce that $D_n x_i \equiv 0 \bmod S$ for $i = 1$ and thus (use a suitable permutation) for all i. Hence $D_n q(x_1, \dots, x_n) \equiv 0 \bmod S$ as claimed. □

Remark 3.5. Some authors use $x_n^{n-1}x_{n-1}^{n-2}\cdots x_2$ in place of $x_1^{n-1}x_2^{n-2}\cdots x_{n-1}$. Modulo S these two classes only differ by the sign $(-1)^{n(n-1)/2}$.

3C. *Hyperbolic formal group laws.* We now define hyperbolic formal group laws, which generalize the additive and multiplicative ones.

Definition 3.6. The hyperbolic formal group law F over $R = \mathbb{Z}[\mu_1, \mu_2]$ and its inverse χ are given by

$$F(x, y) = \frac{x + y - \mu_1 xy}{1 + \mu_2 xy} \quad \text{and} \quad \chi(x) = -\frac{x}{1 - \mu_1 x}.$$

Recall that formal group laws are by definition power series in two variables, and all fractions here and below may be written as such. Note that any ring homomorphism $\mathbb{Z}[\mu_1, \mu_2] \to A$ induces a formal group law over A. Calling these induced formal group laws hyperbolic as well, we find that additive and multiplicative formal group laws are special cases of hyperbolic formal group laws. See, e.g., [Buchstaber and Bunkova 2010; Hoffnung et al. 2014, Example 2.2(d); Lenart and Zainoulline 2017, 2.2] for more on hyperbolic formal group laws. Combining their computations, we see that

$$F(x, y) = x + y - \mu_1 xy + \mu_2(x^2y + xy^2) + \mu_2\mu_1 x^2y^2 + O(5).$$

In Section 4B below, we explain how these FGLs lead to certain elliptic cohomology theories $E^*(-)$. If $\mu_2 = 0$, these cohomology theories specialize to Chow groups (if $\mu_1 = 0$), K_0 (if μ_1 is invertible, thus sometimes called periodic K-theory), connective K_0 and (if $\mu_1 = 0$ but $\mu_2 \neq 0$) theories associated with Lorentz FGLs.

Definition 3.7. Let F be a formal group law. Define

$$\kappa_i = \kappa_i^F = \frac{1}{F(x_i, \chi(x_{i+1}))} + \frac{1}{F(x_{i+1}, \chi(x_i))}.$$

Remark 3.8. In the above definition, κ_i is a formal series. Indeed, writing

$$F(x, \chi(y)) = (x - y)g(x, y)$$

with g a formal series with constant term equal to 1, we get

$$\kappa_i = \frac{g(y, x) - g(x, y)}{(x - y)g(x, y)g(y, x)}.$$

Since the numerator vanishes for $x = y$ there exists a formal series h such that $g(y, x) - g(x, y) = (x - y)h(x, y)$ and we get

$$\kappa_i = \frac{h(x, y)}{g(x, y)g(y, x)},$$

which can be written as a formal series.

Remark 3.9. An easy computation shows that $\Delta_i = \kappa_i - C_i$.

Example 3.10. The three formal group laws we have studied so far are F_a, F_m and F_e, namely the additive, the multiplicative and the elliptic (or hyperbolic) formal group laws:

$$F_a(x, y) = x + y, \quad F_m(x, y) = x + y - \mu_1 xy \quad \text{and} \quad F_e(x, y) = \frac{x + y - \mu_1 xy}{1 + \mu_2 xy}.$$

In these cases, we have $\kappa_i^{F_a} = 0$, $\kappa_i^{F_m} = \kappa_i^{F_e} = \mu_1$. So in all these examples, $\kappa := \kappa_i$ is independent of i.

We now define a variant of the Hecke algebra generalizing [Fomin and Kirillov 1996a, Definition 2.2] with respect to a fixed hyperbolic formal group law F. Setting $\mu_2 = 0$, we obtain the Hecke algebra of [Fomin and Kirillov 1996a], corresponding to (connective or periodic) K-theory.

Definition 3.11. For the hyperbolic formal group law F defined over $R = \mathbb{Z}[\mu_1, \mu_2]$ consider the commutative ring $\mathcal{R} := R[[x_1, \dots, x_n]]$. The *generalized Hecke algebra* $\mathcal{A}_n(\kappa)$ is the quotient of the associative $\mathcal{R}$-algebra $\mathcal{R}\langle u_1, \dots, u_{n-1}\rangle$ by the relations

- $u_i x_j = x_j u_i$ for all i, j,
- $u_i u_j = u_j u_i$ for $|i - j| > 1$,
- $u_i u_{i+1} u_i = u_{i+1} u_i u_{i+1}$ for all i,
- $u_i^2 = -\mu_1 u_i$ for all i,
- $\mu_2 x_i x_{i+1} u_i = 0$ for all i.

Although this algebra generalizes the ones of [Fomin and Kirillov 1996a; Buch 2002] and others, note that it is different from the formal *Demazure* algebras studied in [Calmès et al. 2013; Hoffnung et al. 2014]. See Remark 3.18 below for more details on this.

Remark 3.12. Note that the elements u_i satisfy the braid relations. Hence for any permutation w, we can define the element u_w as $u_w = u_{i_1} \cdots u_{i_r}$, where $w = s_{i_1} \cdots s_{s_{i_r}}$ is any reduced expression of w.

We now generalize [Fomin and Kirillov 1996a, Theorem 2.3] from multiplicative to hyperbolic formal group laws. Define

$$\mathfrak{S}(x_1, \dots, x_{n-1}) = \prod_{j=1}^{n-1} \prod_{i=n-1}^{j} (1 + x_j u_i),$$

where the interchanged bounds for i mean that the corresponding factors are multiplied in descending order, starting with $i = n - 1$.

Theorem 3.13. *For any hyperbolic FGL, in the generalized Hecke algebra $\mathcal{A}_n(\kappa)$ of Definition 3.11, we have*

$$\mathfrak{S}(x_1, \ldots, x_{n-1}) = \sum_{w \in \Sigma_n} \mathfrak{L}_{\underline{w}} u_{w_0 w},$$

where $\underline{w}$ is any reduced expression of w and $w_0(i) = n+1-i$ as usual.

Before proving this theorem, we compare the generalized Schubert polynomials $\mathfrak{L}_{\underline{w}}$ with the corresponding Grothendieck polynomials for K-theory.

Definition 3.14. Let w be a permutation and $w = s_{i_1} \cdots s_{s_{i_r}}$ any reduced expression:

(1) The support of w is the set $\mathrm{Supp}(w) = \{i_1, \ldots, i_r\}$. This is independent of the chosen reduced expression since it is preserved by the braid relations.

(2) Define $I(w)$ as the ideal in $\mathcal{R}$ generated by the polynomials $\mu_2 x_i x_{i+1}$ for $i \in \mathrm{Supp}(w_0 w)$.

(3) Let $\mathfrak{L}^K_w$ be the K-theoretic Grothendieck polynomial representing X_w.

Corollary 3.15. *Let $\underline{w} = s_{\alpha_{i_1}} \cdots s_{\alpha_{i_r}}$ be a reduced expression of w. Then for w a permutation and $\underline{w}$ any reduced expression for w, in $\mathcal{R}$ we have*

$$\mathfrak{L}_{\underline{w}} = \mathfrak{L}^K_w \mod I(w).$$

Some parts of the proof of [Fomin and Kirillov 1996a, Theorem 2.3] are formal and immediately generalize to arbitrary formal group laws. Lemma 2.5 of the same paper just rephrases Remark 3.9. Several other crucial parts of the proof do not generalize to arbitrary FGLs. However, they do generalize to hyperbolic FGLs when working with the generalized Hecke algebra $\mathcal{A}_n(\kappa)$. An important point in choosing hyperbolic FGL is the fact that the κ_i are independent of i, so we have an action of the symmetric group on $\mathcal{A}_n(\kappa)$ given by permutations on the variables x_i. From now on, we fix a hyperbolic formal group law F and a positive integer n.

Lemma 3.16. *Set $\alpha_i(x) = (1 + xu_{n-1}) \cdots (1 + xu_i)$. Then we have the following equalities in $\mathcal{A}_n(\kappa)$:*

(1) $\alpha_{i+1}(x_{i+1}) = \alpha_i(x_{i+1})(1 + \chi(x_{i+1})u_i)$.

(2) $1 + \chi(x_i)u_i = (1 + F(x_{i+1}, \chi(x_i))u_i)(1 + \chi(x_{i+1})u_i)$.

(3) $\Delta_i(1 + \chi(x_{i+1})u_i) = -(1 + \chi(x_{i+1})u_i)u_i$.

Proof. (1) The equality $\alpha_{i+1}(x_{i+1})(1 + x_{i+1}u_i) = \alpha_i(x_{i+1})$ implies

$$\alpha_{i+1}(x_{i+1})(1 + x_{i+1}u_i)(1 + \chi(x_{i+1})u_i) = \alpha_i(x_{i+1})(1 + \chi(x_{i+1})u_i).$$

A straightforward computation shows that $(1 + x_{i+1}u_i)(1 + \chi(x_{i+1})u_i) = 1$.

(2) To prove the claim, it suffices to prove that

$$\big(F(x_{i+1},\chi(x_i))+\chi(x_{i+1})-\chi(x_i)\big)u_i+\chi(x_{i+1})F(x_{i+1},\chi(x_i))u_i^2=0,$$

or equivalently that

$$\big(F(x_{i+1},\chi(x_i))+\chi(x_{i+1})-\chi(x_i)-\mu_1\chi(x_{i+1})F(x_{i+1},\chi(x_i))\big)u_i=0.$$

This holds by a computation using the explicit formulas for F and χ and the relation $\mu_2 x_i x_{i+1}(x_i-x_{i+1})u_i=0$. We use the stronger relation $\mu_2 x_i x_{i+1} u_i=0$ in the definition of our Hecke algebra since we need x_i-x_{i+1} to be a nonzero divisor for the next computation.

(3) We have

$$\begin{aligned}
-\Delta_i(1+\chi(x_{i+1})u_i) &= \frac{(1+\chi(x_i)u_i)-(1+\chi(x_{i+1})u_i)}{F(x_{i+1},\chi(x_i))}\\
&= \frac{1+F(x_{i+1},\chi(x_i))u_i-1}{F(x_{i+1},\chi(x_i))}(1+\chi(x_{i+1})u_i)\\
&= (1+\chi(x_{i+1})u_i)u_i.
\end{aligned}$$

The second equality follows from part (2). □

Proposition 3.17. *In the above notation, for all i we have the commutation*

$$\alpha_i(x_i)\alpha_i(x_{i+1})=\alpha_i(x_{i+1})\alpha_i(x_i).$$

Proof. Since we have the same relations for the u_i as in [Fomin and Kirillov 1996a], the proof of their Lemma 2.6 generalizes to our situation. More precisely, we may apply [Fomin and Kirillov 1996b, Corollary 5.4] as its assumptions (see Section 2 of that paper) are satisfied in our generalized Hecke algebra. □

Proof of Theorem 3.13. From $\mathfrak{S}(x_1,\dots,x_{n-1})=\alpha_1(x_1)\cdots\alpha_{n-1}(x_{n-1})$ we get

$$\mathfrak{S}(x_1,\dots,x_{n-1})=\alpha_1(x_1)\cdots\alpha_i(x_{i+1})(1+\chi(x_{i+1})u_i)\alpha_{i+2}(x_{i+2})\cdots\alpha_{n-1}(x_{n-1}).$$

Using Lemma 3.16(1), this implies $\Delta_i(\mathfrak{S}(x_1,\dots,x_{n-1}))$ is equal to the following formulas:

$$\begin{aligned}
&\alpha_1(x_1)\cdots\alpha_{i-1}(x_{i-1})\Delta_i\alpha_i(x_i)\alpha_i(x_{i+1})(1+\chi(x_{i+1})u_i)\alpha_{i+2}(x_{i+2})\cdots\alpha_{n-1}(x_{n-1})\\
&=\alpha_1(x_1)\cdots\alpha_{i-1}(x_{i-1})\alpha_i(x_i)\alpha_i(x_{i+1})\Delta_i(1+\chi(x_{i+1})u_i)\alpha_{i+2}(x_{i+2})\cdots\alpha_{n-1}(x_{n-1})\\
&=-\alpha_1(x_1)\cdots\alpha_i(x_i)\alpha_i(x_{i+1})(1+\chi(x_{i+1})u_i)u_i\alpha_{i+2}(x_{i+2})\cdots\alpha_{n-1}(x_{n-1})\\
&=-\alpha_1(x_1)\cdots\alpha_i(x_i)\alpha_i(x_{i+1})(1+\chi(x_{i+1})u_i)\alpha_{i+2}(x_{i+2})\cdots\alpha_{n-1}(x_{n-1})u_i.
\end{aligned}$$

Here the first equality follows from Proposition 3.17 and the fact that, as an operator, Δ_i commutes with the operator given by multiplication with a polynomial which is symmetric in x_i and x_{i+1}. The second equality follows from Lemma 3.16(3). We thus have shown

$$-\Delta_i(\mathfrak{S}(x_1,\dots,x_{n-1})) = (\mathfrak{S}(x_1,\dots,x_{n-1}))u_i,$$

which corresponds precisely to the induction step in Definition 3.3, using that $\Delta_i = \kappa - C_i$ and $u_i^2 = -\kappa u_i$. More precisely, write $\mathfrak{S} = \sum \widehat{\mathfrak{L}}_w u_{w_0 w}$, where the sum is taken over all $w \in \Sigma_n$. We wish to show that $\widehat{\mathfrak{L}}_w u_{w_0 w} = \mathfrak{L}_{\underline{w}} u_{w_0 w}$ by an ascending induction on the length of w. For $w = 1$ the claim is obviously true. Now fix $w \neq 1$ and choose i such that ws_i is reduced. Consider the coefficient of $u_{w_0 w}$ in

$$(C_i - \kappa_i)\mathfrak{S} = -\Delta_i \mathfrak{S} = \mathfrak{S} u_i.$$

Using that $u_i^2 = -\kappa_i u_i$ and the fact that $w_0 w s_i < w_0 w$, we deduce that

$$(C_i - \kappa_i)\widehat{\mathfrak{L}}_w u_{w_0 w} = (\widehat{\mathfrak{L}}_{ws_i} - \kappa_i \widehat{\mathfrak{L}}_w) u_{w_0 w};$$

hence $C_i \widehat{\mathfrak{L}}_w u_{w_0 w} = \widehat{\mathfrak{L}}_{ws_i} u_{w_0 w}$ as required. □

Remark 3.18. Note that the computations from [Fomin and Kirillov 1996a] cannot be done in the formal Demazure algebra of [Hoffnung et al. 2014]. E.g., the equality

$$(1 + x_{i+1}u_i)(1 + \chi(x_{i+1})u_i) = 1,$$

which was used to prove Lemma 3.16 above, does not hold, even for the additive FGL. This is related to the failure of $\kappa_i \Delta_i = \Delta_i \kappa_i$.

As for hyperbolic formal group laws, κ_i is independent of i (see Example 3.10); several other parts in [Buch 2002] on the Littlewood–Richardson rule for K_0 easily generalize to hyperbolic formal group laws when working with the generalized Hecke algebra $\mathcal{A}_n(\kappa)$ of Definition 3.11. For example, similar to [Buch 2002, p. 41], it is possible to introduce a stable generalized Schubert polynomial colim $\mathfrak{L}_{1^m \times \underline{w}}$ of $\mathfrak{L}_{\underline{w}}$ and to try to analyze its behavior along the lines of [Fomin and Kirillov 1996b, Section 6]. Also, there is a well-defined analog $\mathfrak{L}_{\nu/\lambda}$ of the polynomial $G_{\nu/\lambda}$ which is crucial for [Buch 2002, Theorem 3.1], as the construction on pages 41–42 of the same paper provides a reduced word $\underline{w}$ rather than just a permutation w. However, for hyperbolic formal group laws the operators C_i no longer satisfy the classical braid relation but a twisted version of it, namely $C_i C_{i+1} C_i + \mu_2 C_i = C_{i+1} C_i C_{i+1} + \mu_2 C_{i+1}$ [Hoffnung et al. 2014]. This will lead to additional difficulties when arguing inductively using these C_i and the corresponding geometric operators as, e.g., in [Buch 2002, Section 8]. This is also related to the discussion in [Lenart and Zainoulline 2017, Section 6]. On the other hand, Proposition 3.17 is wrong already for small values of n and i when replacing the classical braid relation for the u_i by its twisted analog in the definition of $\mathcal{A}_n(\kappa)$. We hope to return to these questions in future work.

4. Some examples

4A. *Polynomials representing some smooth Schubert varieties.* We first compute generalized Schubert polynomials for some of the smooth Schubert varieties considered in Section 2. Let $X = \mathrm{Gr}(k, n)$ be a Grassmannian and let λ be a partition of the form b^a. Denote by $\mathfrak{S}_\lambda$ the polynomial in $\Omega^*(G/B) \simeq \mathbb{L}[x_1, \dots, x_n]/S$ representing the pull-back along the canonical quotient map $\pi : G/B \to X$ of the cobordism class $[X_\lambda \to X]$. Recall [Heller and Malagón-López 2013, Section 3.2.4] that the induced map $\pi^* : \Omega^*(\mathrm{Gr}(k, n)) \to \Omega^*(G/B)$ is a ring monomorphism which identifies $\Omega^*(\mathrm{Gr}(k, n))$ with an explicit subring of $\mathbb{L}[x_1, \dots, x_n]/S$. The results in the sequel may thus be stated in either of these rings. (Recall that there is a standard map, see, e.g., [Buch 2002, p. 42], from partitions to permutations that corresponds to π^* and geometric operators for K-theory and Chow groups.)

Lemma 4.1. *In $\Omega^*(X)$, we have $[X_{(n-k)^{k-1}}]^a = [X_{(n-k)^{k-a}}]$ and $[X_{(n-k-1)^k}]^b = [X_{(n-k-b)^k}]$.*

Proof. We need to prove the formula $[X_{(n-k)^{k-1}}] \cdot [X_{(n-k)^{k-a}}] = [X_{(n-k)^{k-a-1}}]$. But the first class is represented by the sub-Grassmannian $X_{n-k} = \{V_k \in X \mid E_1 \subset V_k\}$, while the second class is represented by $X^{(n-k)^{k-a\vee}} = X^{(n-k)^a} = \{V_k \in X \mid E^a \subset V_k\}$. The product is represented by the intersection of these varieties and since E_1 and E^a do not meet we get

$$X_{n-k} \cap X^{(n-k)^a} = \{V_k \in X \mid E_1 \oplus E^a \subset V_k\}.$$

This last variety is a $\mathrm{GL}_n(\mathsf{k})$-translate of $X_{(n-k)^{k-a-1}} = \{V_k \in X \mid E_{a+1} \subset V_k\}$, proving the first formula. The second one is obtained along the same lines or deduced from the first one using the isomorphism $\mathrm{Gr}(k, n) \simeq \mathrm{Gr}(n-k, n)$. □

Proposition 4.2. *In $\Omega^*(G/B)$, we have the formulas*

$$\mathfrak{S}_{(n-k)^a} = (x_{k+1} \cdots x_n)^{k-a} \quad \textit{and} \quad \mathfrak{S}_{b^k} = (x_1 \cdots x_k)^{n-k-b}.$$

Proof. By the previous lemma, we only need to compute the class $[X_{(n-k)^{k-1}}]$ in $\Omega^*(X)$. Since $X_{(n-k)^{k-1}}$ is the zero locus of a section of the tautological quotient bundle whose Chern roots are $x_{k+1}, \dots, x_n$, the first equality of the proposition follows; see for example the proof of [Levine and Morel 2007, Lemma 6.6.7]. For the second formula, we just need to remark that $X_{(k-1)^k}$ is the zero locus of a global section of the dual of the tautological subbundle and apply the same method (or use the isomorphism $\mathrm{Gr}(k, n) \simeq \mathrm{Gr}(n-k, n)$ again). □

Corollary 4.3. *The classes of $[X_{(n-k)^a} \to X]$ and $[X_{b^k} \to X]$ are represented by the same polynomial in any oriented cohomology theory.*

Proof. Indeed we have $[X_{(n-k)^a} \to X] = \mathfrak{S}_{(n-k)^a}$ and $[X_{b^k} \to X] = \mathfrak{S}_{b^k}$, so this is independent of the FGL. □

Remark 4.4. We will see in the next subsection that this is no longer the case for the other classes of smooth Schubert varieties. Indeed, in Proposition 4.5, we prove that the class of the line in the elliptic cohomology of $\mathrm{Gr}(2,4)$ is given by $x_1x_2(x_1+x_2)-\mu_1x_1^2x_2^2$ and therefore depends on the FGL.

4B. ***Elliptic cohomology of* Gr(2, 4).** We now present explicit results concerning elliptic cohomology, i.e., for the hyperbolic FGL, of $\mathrm{Gr}(2,4)$. We compute the polynomial representatives for all Bott–Samelson classes as well as their products.

Let $X=\mathrm{Gr}(2,4)$ and let λ be a partition. Denote by $\mathfrak{L}_\lambda$ the polynomial in $\Omega^*(G/B)\simeq\mathbb{L}[x_1,x_2,x_3,x_4]/S$ representing the pull-back along the map $G/B\to X$ of the cobordism class $[\widetilde{X}_\lambda\to X]$, where $\widetilde{X}_\lambda$ is the Bott–Samelson resolution of X_λ.

Recall the hyperbolic FGL of [Buchstaber and Bunkova 2010, Example 63] as in Section 3C above. By the universal property of the formal group law of Ω^* established in [Levine and Morel 2007], we have a unique morphism of formal group laws, which yields in particular a ring morphism $\mathbb{L}\to\mathbb{Z}[\mu_1,\mu_2]$. This map is called the "Krichever genus" and is studied in detail in [loc. cit.]. In particular, μ_i has cohomological degree $-i$ for $i=1,2$. Note that (unlike in the bigraded case, see, e.g., [Levine et al. 2013]) this always yields an oriented cohomology theory, as there is no Landweber exactness condition to check. As the theory $E^*(-)$ is oriented in the sense of [Levine and Morel 2007], the analogs of the above theorems also hold for $E^*(G/B)$ and $E^*(\mathrm{Gr}(2,4))$, and the natural transformation $\Omega^*(-)\to E^*(-)$ commutes in particular with the ring monomorphisms π^*. Below, we use the notations $\widetilde{X}_\lambda$ and $\mathfrak{L}_\lambda$ for elements in $E^*(-)$ as well.

Proposition 4.5. *In* $E^*(\mathrm{Gr}(2,4))$, *we have the following formulas*:

$$\begin{aligned}
\mathfrak{L}_{(00)} &= x_1^2x_2^2,\\
\mathfrak{L}_{(10)} &= x_1x_2(x_1+x_2)-\mu_1x_1^2x_2^2,\\
\mathfrak{L}_{(20)} &= x_1^2+x_1x_2+x_2^2-\mu_1x_1x_2(x_1+x_2)-\mu_2x_1^2x_2^2,\\
\mathfrak{L}_{(11)} &= x_1x_2-\mu_2x_1^2x_2^2,\\
\mathfrak{L}_{(21)} &= x_1+x_2-\mu_1x_1x_2-\mu_2x_1x_2(x_1+x_2)-\mu_1\mu_2x_1^2x_2^2,\\
\mathfrak{L}_{(22)} &= 1-\mu_2(x_1+x_2)^2+\mu_1^2\mu_2x_1^2x_2^2.
\end{aligned}$$

Proof. Since the fiber of the map $\pi: G/B\to\mathrm{Gr}(2,4)$ is isomorphic to $\mathbb{P}^1\times\mathbb{P}^1$, the pull-back $\pi^*[\widetilde{X}_\lambda]\in E^*(G/B)$ of a Bott–Samelson class in $\mathrm{Gr}(2,4)$ is again a Bott–Samelson class $X_{\underline{w}}$. (Note that this is not true anymore in higher dimensions.) Moreover in this case, we can explicitly write down the reduced word $\underline{w}$ corresponding to λ under π^*. Now we wish to compute $\mathfrak{L}_\lambda\in E^*(\mathrm{Gr}(2,4))\subset E^*(G/B)$. The above, together with the results of [Hornbostel and Kiritchenko

2011], implies that both in $\Omega^*(G/B)$ and $E^*(G/B)$, we have

$$\begin{aligned} \pi^*[X_{(00)}] &= C_1C_3(\mathfrak{L}_1), & \pi^*[X_{(11)}] &= C_1C_3C_2C_1(\mathfrak{L}_1), \\ \pi^*[X_{(10)}] &= C_1C_3C_2(\mathfrak{L}_1), & \pi^*[X_{(21)}] &= C_1C_3C_2C_1C_3(\mathfrak{L}_1), \\ \pi^*[X_{(20)}] &= C_1C_3C_2C_3(\mathfrak{L}_1), & \pi^*[X_{(22)}] &= C_1C_3C_2C_1C_3C_2(\mathfrak{L}_1). \end{aligned}$$

Now the results follow from $\mathfrak{L}_1 = x_1^3x_2^2x_3$ and explicit computations with the C_i done with the help of a computer. □

We computed everything in elliptic cohomology for sake of simplicity, but a similar computation can be done in $\Omega^*(X)$.

Remark 4.6. In elliptic cohomology, the multiplication formula for the square of the hyperplane class in the Bott–Samelson basis is the same as the one in K-theory, namely $\mathfrak{L}_{(21)}^2 = \mathfrak{L}_{(20)} + \mathfrak{L}_{(11)} - \mu_1\mathfrak{L}_{(10)}$.

Acknowledgement

We thank the referee for careful reading and suggestions.

References

[Bott and Samelson 1958] R. Bott and H. Samelson, "Applications of the theory of Morse to symmetric spaces", *Amer. J. Math.* **80** (1958), 964–1029. MR Zbl

[Bressler and Evens 1990] P. Bressler and S. Evens, "The Schubert calculus, braid relations, and generalized cohomology", *Trans. Amer. Math. Soc.* **317**:2 (1990), 799–811. MR Zbl

[Brion and Polo 1999] M. Brion and P. Polo, "Generic singularities of certain Schubert varieties", *Math. Z.* **231**:2 (1999), 301–324. MR Zbl

[Buch 2002] A. S. Buch, "A Littlewood–Richardson rule for the K-theory of Grassmannians", *Acta Math.* **189**:1 (2002), 37–78. MR Zbl

[Buch et al. ≥ 2018] A. S. Buch, P. E. Chaput, L. C. Mihalcea, and N. Perrin, "Positivity of minuscule quantum K-theory", in preparation.

[Buchstaber and Bunkova 2010] V. M. Buchstaber and E. Y. Bunkova, "Elliptic formal group laws, integral Hirzebruch genera and Krichever genera", preprint, 2010. arXiv

[Calmès et al. 2013] B. Calmès, V. Petrov, and K. Zainoulline, "Invariants, torsion indices and oriented cohomology of complete flags", *Ann. Sci. Éc. Norm. Supér.* (4) **46**:3 (2013), 405–448. MR Zbl

[Chaput et al. 2007] P.-E. Chaput, L. Manivel, and N. Perrin, "Quantum cohomology of minuscule homogeneous spaces, II: Hidden symmetries", *Int. Math. Res. Not.* **2007**:22 (2007), art. id. rnm107. MR Zbl

[Chaput et al. 2009] P.-E. Chaput, L. Manivel, and N. Perrin, "Affine symmetries of the equivariant quantum cohomology ring of rational homogeneous spaces", *Math. Res. Lett.* **16**:1 (2009), 7–21. MR Zbl

[Demazure 1974] M. Demazure, "Désingularisation des variétés de Schubert généralisées", *Ann. Sci. École Norm. Sup.* (4) **7** (1974), 53–88. MR Zbl

[Fomin and Kirillov 1996a] S. Fomin and A. N. Kirillov, "Grothendieck polynomials and the Yang–Baxter equation", pp. 183–189 in *Formal power series and algebraic combinatorics* (Piscataway, NJ), edited by L. J. Billera et al., DIMACS Series in Discrete Mathematics and Theoretical Computer Science **24**, AMS, Providence, RI, 1996. MR

[Fomin and Kirillov 1996b] S. Fomin and A. N. Kirillov, "The Yang–Baxter equation, symmetric functions, and Schubert polynomials", *Discrete Math.* **153**:1-3 (1996), 123–143. MR Zbl

[Fulton 1998] W. Fulton, *Intersection theory*, 2nd ed., Ergebnisse der Mathematik (3) **2**, Springer, 1998. MR Zbl

[Gasharov and Reiner 2002] V. Gasharov and V. Reiner, "Cohomology of smooth Schubert varieties in partial flag manifolds", *J. London Math. Soc.* (2) **66**:3 (2002), 550–562. MR Zbl

[Hansen 1973] H. C. Hansen, "On cycles in flag manifolds", *Math. Scand.* **33** (1973), 269–274. MR Zbl

[Heller and Malagón-López 2013] J. Heller and J. Malagón-López, "Equivariant algebraic cobordism", *J. Reine Angew. Math.* **2013**:684 (2013), 87–112. MR Zbl

[Hoffnung et al. 2014] A. Hoffnung, J. Malagón-López, A. Savage, and K. Zainoulline, "Formal Hecke algebras and algebraic oriented cohomology theories", *Selecta Math.* (*N.S.*) **20**:4 (2014), 1213–1245. MR Zbl

[Hornbostel and Kiritchenko 2011] J. Hornbostel and V. Kiritchenko, "Schubert calculus for algebraic cobordism", *J. Reine Angew. Math.* **2011**:656 (2011), 59–85. MR Zbl

[Hudson and Matsumura 2016] T. Hudson and T. Matsumura, "Segre classes and Kempf–Laksov formula in algebraic cobordism", preprint, 2016. To appear in *IMPANGA 15: Schubert varieties, equivariant cohomology and characteristic classes*. arXiv

[Hudson and Matsumura 2017] T. Hudson and T. Matsumura, "Vexillary degeneracy loci classes in K-theory and algebraic cobordism", preprint, 2017. arXiv

[Kleiman 1974] S. L. Kleiman, "The transversality of a general translate", *Compositio Math.* **28**:3 (1974), 287–297. MR Zbl

[Lakshmibai and Brown 2015] V. Lakshmibai and J. Brown, *The Grassmannian variety: geometric and representation-theoretic aspects*, Developments in Mathematics **42**, Springer, 2015. MR Zbl

[Lazard 1955] M. Lazard, "Sur les groupes de Lie formels à un paramètre", *Bull. Soc. Math. France* **83** (1955), 251–274. MR Zbl

[Lenart and Zainoulline 2017] C. Lenart and K. Zainoulline, "Towards generalized cohmology Schubert calculus via formal root polynomials", *Math. Res. Lett.* **24**:3 (2017), 839–877. MR

[Levine and Morel 2007] M. Levine and F. Morel, *Algebraic cobordism*, Springer, 2007. MR Zbl

[Levine and Pandharipande 2009] M. Levine and R. Pandharipande, "Algebraic cobordism revisited", *Invent. Math.* **176**:1 (2009), 63–130. MR Zbl

[Levine et al. 2013] M. Levine, Y. Yang, and G. Zhao, "Algebraic elliptic cohomology theory and flops, I", preprint, 2013. To appear in *Math. Annalen*. arXiv

[Magyar 1998] P. Magyar, "Schubert polynomials and Bott–Samelson varieties", *Comment. Math. Helv.* **73**:4 (1998), 603–636. MR Zbl

[Manivel 2001] L. Manivel, *Symmetric functions, Schubert polynomials and degeneracy loci*, SMF/AMS Texts and Monographs **6**, American Mathematical Society, Providence, RI, 2001. MR Zbl

[Perrin 2007] N. Perrin, "Small resolutions of minuscule Schubert varieties", *Compos. Math.* **143**:5 (2007), 1255–1312. MR Zbl

[Perrin 2009] N. Perrin, "The Gorenstein locus of minuscule Schubert varieties", *Adv. Math.* **220**:2 (2009), 505–522. MR Zbl

[Seidel 1997] P. Seidel, "π_1 of symplectic automorphism groups and invertibles in quantum homology rings", *Geom. Funct. Anal.* **7**:6 (1997), 1046–1095. MR Zbl

[Zelevinskiĭ 1983] A. V. Zelevinskiĭ, "Small resolutions of singularities of Schubert varieties", *Funktsional. Anal. i Prilozhen.* **17**:2 (1983), 75–77. In Russian; translated in *Funct. Anal. Appl.*, **17**:2 (1983), 142–144. MR Zbl

Received March 21, 2017. Revised September 26, 2017.

JENS HORNBOSTEL
BERGISCHE UNIVERSITÄT WUPPERTAL
FACHGRUPPE MATHEMATIK UND INFORMATIK
WUPPERTAL
GERMANY

hornbostel@math.uni-wuppertal.de

NICOLAS PERRIN
LABORATOIRE DE MATHÉMATIQUES DE VERSAILLES, UVSQ, CNRS
UNIVERSITÉ PARIS-SACLAY
VERSAILLES
FRANCE

nicolas.perrin@uvsq.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 294, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.294.423

SOBOLEV INEQUALITIES ON A WEIGHTED RIEMANNIAN MANIFOLD OF POSITIVE BAKRY–ÉMERY CURVATURE AND CONVEX BOUNDARY

SAÏD ILIAS AND ABDOLHAKIM SHOUMAN

To the memory of our friend A. El Soufi

In this paper, we study some nonlinear elliptic equations on a compact n-dimensional weighted Riemannian manifold of positive m-Bakry–Émery–Ricci curvature and convex boundary. Our main purpose is to find conditions which imply that such elliptic equations admit only constant solutions. As an application, we obtain weighted Sobolev inequalities with explicit constants that extend the inequalities obtained by Ilias [1983; 1996] in the Riemannian setting. In a last part of the article, as applications we derive a new Onofri inequality, a logarithmic Sobolev inequality and estimates for the eigenvalues of a weighted Laplacian and for the trace of the weighted heat kernel.

1. Introduction and main result

Sobolev inequalities with sharp constants play an important role in Riemannian and conformal geometries. For example, on the unit sphere $\mathbb{S}^n$ endowed with its standard metric, we have (see [Aubin 1982]), for all $f \in H_1^2(\mathbb{S}^n)$,

$$\|f\|^2_{L^{2n/(n-2)}(dv)} \leq K(n,2)\|\nabla f\|^2_{L^2(dv)} + \mathrm{vol}(\mathbb{S}^n)^{-2/n}\|f\|^2_{L^2(dv)}, \tag{1-1}$$

where $K(n,2) := 4/(n(n-2))\,\mathrm{vol}(\mathbb{S}^n)^{-2/n}$, dv and $\mathrm{vol}(\mathbb{S}^n)$ are respectively the Riemannian measure and the Riemannian volume of $\mathbb{S}^n$. This inequality has been crucial in the study of the Yamabe problem on closed Riemannian manifolds. It corresponds to the limiting case in the Sobolev embedding

$$H_1^2 \hookrightarrow L^p \quad \left(2 < p \leq \hat{2} := \frac{2n}{n-2}\right)$$

MSC2010: 35J60, 53C21, 58J32, 58J50, 58J60.

Keywords: Sobolev inequality, Onofri inequality, weighted Riemannian manifold, convex boundary, manifold with density, weighted Laplacian, drifting Laplacian, Bakry–Émery–Ricci curvature, Neumann boundary condition, eigenvalues.

and the constants appearing in it are the best possible constants (see [Aubin 1982; Lee and Parker 1987; Ilias 1983]). Note that, using a stereographic projection and the conformal nature of this inequality, one can show its equivalence with the Euclidean Sobolev inequality (see for instance [Lee and Parker 1987]),

$$\forall f \in H_1^2(\mathbb{R}^n), \quad \|f\|^2_{L^{2n/(n-2)}(dx)} \le K(n,2)\|\nabla f\|^2_{L^2(dx)}, \tag{1-2}$$

where dx is the Lebesgue measure and $K(n,2)$ is the best constant in this Euclidean Sobolev inequality [Aubin 1982]. The conformal nature of inequality (1-1) can also be used to deduce the following Sobolev inequality on the hyperbolic space:

$$\|f\|^2_{L^{2n/(n-2)}(dx)} \le K(n,2)\|\nabla f\|^2_{L^2(dx)} - \operatorname{vol}(\mathbb{S}^n)^{-2/n}\|f\|^2_{L^2(dx)} \tag{1-3}$$

for all $f \in H_1^2(\mathbb{H}^n)$ (see [Hebey 1996] for another proof of this inequality).

Beckner [1993] extended the spherical inequality (1-1) to all the Sobolev exponents, proving that for all $p \in (2, \hat{2}]$,

$$\|f\|^2_{L^p(dv)} \le \operatorname{vol}(\mathbb{S}^n)^{-\frac{p-2}{p}}\left(\frac{p-2}{n}\|\nabla f\|^2_{L^2(dv)} + \|f\|^2_{L^2(dv)}\right). \tag{1-4}$$

This inequality is attributed in the literature to Beckner but it was proved in 1991 independently by Bidaut-Véron and Véron [1991].

In 1983, the first author generalized the spherical inequality (1-1) to any closed Riemannian manifold with positive Ricci curvature, see [Ilias 1983]. In fact, if (M, g) is a compact n-dimensional Riemannian manifold of Ricci curvature bounded-below by a positive constant k, then every function $f \in H_1^2(M)$ satisfies

$$\|f\|^2_{L^{2n/(n-2)}(dv_g)} \le \operatorname{vol}_g(M)^{-2/n}\left(\frac{4(n-1)}{n(n-2)k}\|\nabla f\|^2_{L^2(dv_g)} + \|f\|^2_{L^2(dv_g)}\right), \tag{1-5}$$

where dv_g and $\operatorname{vol}_g(M)$ are respectively the Riemannian measure and the Riemannian volume of (M, g).

This last inequality is derived from (1-1) using the Levy–Gromov isoperimetric inequality and an adapted symmetrization. Moreover, we observe that if we use the inequality (1-4) instead of (1-1), the same arguments of symmetrization extend (1-5) for all the Sobolev exponents $p \in (2, \hat{2}]$.

Bidaut-Véron and Véron [1991] were able to give another proof of the inequality (1-5). Their proof is based on the Bochner formula and a uniqueness result for some nonlinear elliptic equations strongly related to that Sobolev inequality. In fact, they improve a technique developed by Gidas and Spruck [1981]. The technique developed by Gidas and Spruck seems mysterious, but it is in fact inspired by that of Obata in his study of the unicity of an Einstein metric in a conformal class [Yano and Obata 1970]. And as we mentioned above, Bidaut-Véron and Véron [1991] obtained simultaneously the Beckner inequality (1-4). Note that, Bakry and Ledoux

[1996] obtained another important and different proof of the inequality (1-5), and this "probabilistic" proof, generalizes in fact, that inequality to the so-called Markov generators. Another recent generalization of the inequality (1-5) to metric measured spaces is due to Profeta [2015].

A natural question that can be asked is

"Is there a similar Sobolev inequality when the manifold has a boundary?"

For the hemisphere $\mathbb{S}^n_+$ endowed with its standard metric, using the exact value of its relative Yamabe infimum (see for instance [Escobar 1988; 1992]), we immediately get,

$$\|f\|^2_{L^{2n/(n-2)}(dv)} \leq \operatorname{vol}(\mathbb{S}^n_+)^{-2/n}\left(\frac{4}{n(n-2)}\|\nabla f\|^2_{L^2(dv)} + \|f\|^2_{L^2(dv)}\right) \tag{1-6}$$

for all $f \in H^2_1(\mathbb{S}^n_+)$.

In 1996, after a generalization of the method used by Bidaut-Véron and Véron, the first author [Ilias 1996] gave an answer to the above question by obtaining the same inequality as (1-5) for manifolds with convex boundary. More precisely, he proved that, for a compact Riemannian manifold (M, g) with convex boundary and of Ricci curvature bounded below by a constant $k > 0$, we have for any $p \in (2, \hat{2}]$ and any $f \in H^2_1(M)$,

$$\|f\|^2_{L^p(dv_g)} \leq \operatorname{vol}_g(M)^{-\frac{p-2}{p}}\left(\frac{(n-1)(p-2)}{nk}\|\nabla f\|^2_{L^2(dv_g)} + \|f\|^2_{L^2(dv_g)}\right). \tag{1-7}$$

The purpose of the present paper is to adapt the technique that has been used in [Ilias 1996] to the setting of weighted Riemannian manifolds with positive Bakry–Émery–Ricci curvature. More precisely, for a compact n-dimensional weighted Riemannian manifold (M^n, g, σ) of convex boundary and of m-Bakry–Émery–Ricci curvature (for some $m \in [n, \infty)$) bounded below by a positive constant k, we prove the analogue of (1-7) for any $p \in (2, 2^* := 2m/(m-2)]$, where the constant of the gradient term is $(m-1)(p-2)/(mk)$. In fact, we prove a stronger inequality where the constant of the gradient term depends on m, p, k, and the first nonzero Neumann eigenvalue λ^h_1 of the weighted Laplacian (see Theorem 3.6 for more details). Concerning the limiting case $p = 2^*$, our result shows that for any $f \in H^2_1(d\sigma)$,

$$\|f\|^2_{L^{2m/(m-2)}(d\sigma)} \leq \operatorname{vol}_h(M)^{-2/m}\left(\frac{4(m-1)}{m(m-2)k}\|\nabla f\|^2_{L^2(d\sigma)} + \|f\|^2_{L^2(d\sigma)}\right), \tag{1-8}$$

where $d\sigma$ is the weighted measure and $\operatorname{vol}_h(M)$ is the volume of M with respect to $d\sigma$ (see Corollary 3.7). These inequalities are a consequence of two uniqueness results for some nonlinear elliptic equations involving the weighted Laplacian (see Propositions 3.2 and 3.4) which are respectively generalizations to weighted manifolds with convex boundary of the result obtained by the first author [Ilias 1996] and that obtained by Licois and Véron [1995] (independently by Fontenas [1997]) for closed manifolds.

Using inequality (1-8) we can extend many Riemannian results to weighted Riemannian manifolds. Without being exhaustive, we will treat only applications which seems to us the most important. More precisely, we can obtain an upper bound with explicit constants (depending on the Sobolev constants) for the trace of the weighted heat kernel (see Section 4) and deduce therefrom a lower bound for the eigenvalues of the weighted Laplacian.

We also derive from our Sobolev inequalities, the analogue of the Onofri inequality. In fact, we prove in Corollary 4.1 that for any surface M of convex boundary and of Gaussian curvature bounded below by a positive constant k, it holds that

$$\log\left(\frac{1}{\mathrm{vol}_g(M)}\int_M e^{\varphi}\,dv_g\right) \le \frac{1}{\mathrm{vol}_g(M)}\left(\frac{1}{4k}\int_M |\nabla\varphi|^2\,dv_g + \int_M \varphi\,dv_g\right)$$

for all $\varphi \in H_1^2(M)$. In the case of the unit sphere and the unit hemisphere, we recover the classical Onofri inequalities (see [Onofri 1982; Chang and Yang 1988; Osgood et al. 1988]). As another consequence of our Sobolev inequalities, we obtain a logarithmic Sobolev inequality (Corollary 4.3) for weighted Riemannian manifolds with boundary.

This paper is organized as follows. In Section 2, we establish two elementary lemmas (Lemmas 2.1 and 2.2). The uniqueness results (Propositions 3.2 and 3.4) are discussed in Section 3 as well as Theorem 3.6 and Corollary 3.7. Finally, Section 4 is dedicated to some applications.

2. Preliminaries

Throughout the paper, we consider (M^n, g) as a smooth compact n-dimensional Riemannian manifold of boundary ∂M, endowed with a measure $d\sigma := \sigma\,dv_g$, where $\sigma = e^{-h}$ is a positive density (h is a smooth real-valued function on M) and dv_g is the Riemannian measure associated to the metric g. We denote by $\mathrm{vol}_g(M)$ and $\mathrm{vol}_h(M)$ respectively the volume of M with respect to dv_g and that with respect to $d\sigma$. Such a triplet (M, g, σ) is known in the literature as a weighted Riemannian manifold, a manifold with density, a Bakry–Émery manifold, or a Riemannian measure space. The associated weighted Laplacian Δ_h (also called drifted Laplacian, h-Laplacian or Bakry–Émery Laplacian) is given by

$$\Delta_h u = \Delta u - \frac{1}{\sigma}\langle\nabla\sigma, \nabla u\rangle = \Delta u + \langle\nabla h, \nabla u\rangle, \tag{2-1}$$

where Δ and ∇ are respectively the nonnegative Laplacian and the gradient with respect to g. It is self-adjoint on the space of square integrable functions on M

with respect to the weighted measure $d\sigma$, henceforth $L^2(d\sigma)$. We will denote by $H_1^2(d\sigma)$, the Sobolev space of $L^2(d\sigma)$ functions, such that the norm of their gradient is also in $L^2(d\sigma)$. Note that, since the manifold is compact and h is smooth, this Sobolev space coincides with the Sobolev space $H_1^2(M)$ of the Riemannian manifold (M, g), and these two spaces differ only in their norms.

The m-dimensional Bakry–Émery–Ricci curvature tensor (where $m \in [n, \infty)$) is a modified Ricci tensor more suitable to control the geometry of weighted manifolds and is defined by

$$\mathrm{Ric}_h^m := \mathrm{Ric} + D^2 h - \frac{1}{m-n}\, dh \otimes dh \tag{2-2}$$

where D^2 is the Hessian operator on M and Ric is the usual Ricci curvature of (M, g). The equation $\mathrm{Ric}_h^m = \kappa g$ correspond to the so-called *quasi-Einstein metric*, which has been studied by many authors (see for instance [Case et al. 2011]). When $m = \infty$, (2-2) gives the tensor $\mathrm{Ric}_h = \mathrm{Ric} + D^2 h$ introduced by Lichnerowicz [1970; 1971/72] and independently by Bakry and Émery [1985]. For $m = n$, (2-2) makes sense only when the function h is constant and so Ric_h^m is the usual Ricci tensor of M and Δ_h in this case is nothing but the Laplace–Beltrami operator Δ of M.

Let $\{e_1, \dots, e_n\}$ be a local orthonormal frame of M such that at $p \in \partial M$, the vectors $e_1, \dots, e_{n-1}$ are tangent to the boundary and the remaining vector $e_n := \nu$ is the outward unit normal vector to ∂M. The second fundamental form of ∂M at $p \in \partial M$ is defined as

$$\mathrm{II}(X, Y) := \langle \mathcal{A}X, Y\rangle = \langle \nabla_X \nu, Y\rangle$$

for any $X, Y \in T_p(\partial M)$, where $\mathcal{A}$ is the Weingarten endomorphism of $T_p M$. The mean curvature H of ∂M is defined as the trace of the second fundamental form II:

$$H = \sum_{i=1}^{n-1} \mathrm{II}(e_i, e_i).$$

In the sequel, we will need the following two lemmas. The first one is nothing but a little modification of the Bochner–Lichnerowicz–Weitzenböck formula for functions on weighted Riemannian manifolds which generalizes the Reilly identity ([Reilly 1977; Ma and Du 2010]). The version we present here is better suited to our purpose and its proof is a straightforward adaptation of that given by the first author [Ilias 1996] to the weighted setting.

Lemma 2.1 (generalized Reilly formula). *Let (M^n, g, σ) be a compact weighted Riemannian manifold with boundary ∂M. For any two smooth functions u and v on*

M, we have

$$\int_M v\big(|D^2u|^2-(\Delta_h u)^2\big)\,d\sigma$$
$$=-\int_M\big\{(\Delta_h u)\langle\nabla u,\nabla v\rangle+\tfrac12|\nabla u|^2\Delta_h v+\mathrm{Ric}_h(\nabla u,\nabla u)v\big\}\,d\sigma$$
$$+\int_{\partial M}\Big\{-\tfrac12|\nabla u|^2\frac{\partial v}{\partial\nu}+\Big\langle\nabla^\partial\Big(\frac{\partial u}{\partial\nu}\Big),\nabla^\partial u\Big\rangle v+(\Delta^\partial u)\Big(\frac{\partial u}{\partial\nu}\Big)v$$
$$-\mathrm{II}(\nabla^\partial u,\nabla^\partial u)v-Hv\Big(\frac{\partial u}{\partial\nu}\Big)^2+\langle\nabla u,\nabla h\rangle v\Big(\frac{\partial u}{\partial\nu}\Big)\Big\}\,d\sigma,$$

where ∇^∂ and Δ^∂ denote the gradient and the Laplacian of ∂M and for the sake of simplicity, we still denote by $d\sigma$ the induced weighted measure on ∂M.

Proof. From the classical Riemannian Bochner formula applied to u, one can easily deduce the following weighted one (see for instance [Setti 1998; Bakry and Émery 1985]):

$$\langle\nabla(\Delta_h u),\nabla u\rangle=|D^2u|^2+\tfrac12\Delta_h(|\nabla u|^2)+\mathrm{Ric}_h(\nabla u,\nabla u).\tag{2-3}$$

Multiplying (2-3) by v and integrating over M with respect to $d\sigma$, we get:

$$\int_M v\langle\nabla(\Delta_h u),\nabla u\rangle\,d\sigma$$
$$=\int_M v|D^2u|^2\,d\sigma+\tfrac12\int_M v\Delta_h(|\nabla u|^2)\,d\sigma+\int_M v\,\mathrm{Ric}_h(\nabla u,\nabla u)\,d\sigma.$$

Integration by parts in the left hand side and in the second term of the right hand side gives

$$\int_M v\big(|D^2u|^2-(\Delta_h u)^2\big)\,d\sigma\tag{2-4}$$
$$=-\int_M\big\{(\Delta_h u)\langle\nabla u,\nabla v\rangle+\tfrac12|\nabla u|^2\Delta_h v+\mathrm{Ric}_h(\nabla u,\nabla u)v\big\}\,d\sigma$$
$$+\int_{\partial M}\Big\{-\tfrac12|\nabla u|^2\frac{\partial v}{\partial\nu}+(\Delta_h u)\frac{\partial u}{\partial\nu}v+\tfrac12\frac{\partial(|\nabla u|^2)}{\partial\nu}v\Big\}\,d\sigma.$$

Now in the calculations that follow (at a point $x\in\partial M$), we will use an orthonormal local frame $\{e_1,\dots,e_n\}$ such that $e_1,\dots,e_{n-1}$ are tangent to the boundary and $e_n=\nu$ is the outward unit normal to ∂M. A direct calculation of the last two terms in (2-4) at a point $x\in\partial M$ yields

$$(\Delta_h u)\Big(\frac{\partial u}{\partial\nu}\Big)v+\tfrac12\frac{\partial(|\nabla u|^2)}{\partial\nu}v\tag{2-5}$$
$$=\sum_{i=1}^{n-1}\big(D^2u(e_n,e_i)e_i(u)-D^2u(e_i,e_i)e_n(u)\big)v+\langle\nabla h,\nabla u\rangle\Big(\frac{\partial u}{\partial\nu}\Big)v$$

and

$$\begin{aligned}\sum_{i=1}^{n-1} D^2u(e_n, e_i)e_i(u) &= \left\langle \nabla^\partial\left(\frac{\partial u}{\partial \nu}\right), \nabla^\partial u \right\rangle - \mathrm{II}(\nabla^\partial u, \nabla^\partial u),\\ \sum_{i=1}^{n-1} D^2u(e_i, e_i)e_n(u) &= -(\Delta^\partial u)\left(\frac{\partial u}{\partial \nu}\right) + H\left(\frac{\partial u}{\partial \nu}\right)^2.\end{aligned} \tag{2-6}$$

After incorporating the two identities of (2-6) in (2-5), and the obtained result in (2-4), we conclude the proof of Lemma 2.1. □

The second lemma, which has an elementary proof, arises naturally when we need to estimate the Hessian D^2u in terms of $\Delta_h u$ (see for example [Li 2005] for a proof):

Lemma 2.2. *Let u be a smooth function on M. For every $m \geq n$, we have*

$$|D^2u|^2 + \mathrm{Ric}_h(\nabla u, \nabla u) \geq \frac{1}{m}(\Delta_h u)^2 + \mathrm{Ric}_h^m(\nabla u, \nabla u). \tag{2-7}$$

Moreover, the equality in (2-7) holds if and only if

$$D^2u = -\frac{1}{n}(\Delta u)g \quad \textit{and} \quad \Delta_h u = \frac{m}{m-n}\langle \nabla h, \nabla u\rangle.$$

3. Weighted Sobolev inequalities

Let (M, g, σ) be a compact weighted Riemannian manifold of dimension $n \geq 2$. In this section, we seek conditions that guarantee the uniqueness of the positive solution of a nonlinear elliptic PDE (see (3-17)). This is an important step towards the Sobolev inequality.

We first start by giving the keystone of both uniqueness results (Propositions 3.2 and 3.4):

Proposition 3.1. *Let $q > 1$, $\lambda > 0$ and φ be a positive solution of the following system*

$$\begin{cases} \Delta_h \varphi + \lambda\varphi = \varphi^q & \text{in } M,\\ \dfrac{\partial \varphi}{\partial \nu} = 0 & \text{on } \partial M.\end{cases} \tag{3-1}$$

Put

$$J := |D^2u|^2 - \frac{1}{m}(\Delta_h u)^2 + \mathrm{Ric}_h(\nabla u, \nabla u). \tag{3-2}$$

(i) *For any two nonzero real numbers α and β, we have*

$$\begin{aligned}&\int_M u^\beta J\, d\sigma + \int_{\partial M} u^\beta \mathrm{II}(\nabla^\partial u, \nabla^\partial u)\, d\sigma\\ &\quad = A_1 \int_M u^{\beta+(q-1)/\alpha}|\nabla u|^2\, d\sigma + B_1 \int_M u^\beta |\nabla u|^2\, d\sigma + C_1 \int_M u^{\beta-2}|\nabla u|^4\, d\sigma,\end{aligned} \tag{3-3}$$

where

$$A_1 = \left\{\frac{m-1}{m}(\alpha\beta+q) - \tfrac{3}{2}\alpha\beta\right\},$$
$$B_1 = \left\{-\lambda\frac{m-1}{m}(\alpha\beta+1) + \tfrac{3}{2}\alpha\beta\lambda\right\},$$
$$C_1 = \left\{\frac{m-1}{m}\Big(\frac{\alpha-1}{\alpha}\Big)^2 + \tfrac{3}{2}\beta\Big(\frac{\alpha-1}{\alpha}\Big) + \frac{\beta(\beta-1)}{2}\right\}.$$

(ii) *For any two nonzero real numbers* α *and* $\beta \neq -2$, *we have the following identity*:

$$\begin{aligned}\int_M u^\beta J\,d\sigma + \int_{\partial M} u^\beta \mathrm{II}(\nabla^\partial u, \nabla^\partial u)\,d\sigma \\ = A_2\int_M \big(\Delta_h u^{(\beta+2)/2}\big)^2\,d\sigma + B_2\int_M u^\beta|\nabla u|^2\,d\sigma + C_2\int_M u^{\beta-2}|\nabla u|^4\,d\sigma,\end{aligned} \tag{3-4}$$

where

$$A_2 = -\frac{2}{m(\beta+2)^2}\left\{(m+2)\frac{\alpha\beta}{q} - 2(m-1)\right\},$$
$$B_2 = \frac{\alpha\beta}{q}\Big(\frac{m+2}{2m}\Big)\lambda(q-1),$$
$$C_2 = \frac{\alpha\beta}{q}\left[\frac{m+2}{2m}\left\{\Big(\frac{\beta}{4}+\frac{\alpha-1}{\alpha}\Big)\Big(\beta+\frac{q}{\alpha}\Big) + \Big(\frac{\alpha-1}{\alpha}\Big)^2\right\} + \frac{q(\beta-4)}{8\alpha}\right],$$

while if $\beta = -2$, *we have*

$$\begin{aligned}\int_M u^{-2} J\,d\sigma + \int_{\partial M} u^{-2} \mathrm{II}(\nabla^\partial u, \nabla^\partial u)\,d\sigma \\ = A_3\int_M (\Delta_h \ln u)^2\,d\sigma + B_3\int_M |\nabla \ln u|^2\,d\sigma + C_3\int_M |\nabla \ln u|^4\,d\sigma,\end{aligned} \tag{3-5}$$

where

$$A_3 = \frac{1}{m}\left\{(m+2)\frac{\alpha}{q} + (m-1)\right\},$$
$$B_3 = \frac{-2\alpha}{q}\Big(\frac{m+2}{2m}\Big)\lambda(q-1),$$
$$C_3 = -\frac{2\alpha}{q}\left[\frac{m+2}{2m}\left\{\Big(\frac{1}{2}-\frac{1}{\alpha}\Big)\Big(-2+\frac{q}{\alpha}\Big) + \Big(\frac{\alpha-1}{\alpha}\Big)^2\right\} - \frac{3q}{4\alpha}\right].$$

Proof. Let φ be a positive solution of (3-1). Let α and β be two nonzero real numbers to be determined later and take $u = \varphi^\alpha$ and $v = u^\beta$. Using (3-1), a direct calculation gives

$$
\begin{cases} \Delta_h u = \alpha u^{1+(q-1)/\alpha} - \alpha\lambda u - \left(\dfrac{\alpha-1}{\alpha}\right)\dfrac{|\nabla u|^2}{u} & \text{in } M, \\ \dfrac{\partial u}{\partial \nu} = 0 & \text{on } \partial M. \end{cases} \tag{3-6}
$$

Applying Lemma 2.1 with u and v as above, and using (3-6), we obtain

$$
\begin{aligned}
\int_M u^\beta\Big(|D^2u|^2 - \frac{1}{m}(\Delta_h u)^2 + \mathrm{Ric}_h(\nabla u,\nabla u)\Big)\,d\sigma
&= \frac{m-1}{m}\int_M u^\beta(\Delta_h u)^2\,d\sigma - \int_M (\Delta_h u)\langle\nabla u,\nabla u^\beta\rangle\,d\sigma \\
&\quad - \tfrac{1}{2}\int_M |\nabla u|^2\Delta_h u^\beta\,d\sigma - \int_{\partial M}\mathrm{II}(\nabla^\partial u,\nabla^\partial u)u^\beta\,d\sigma,
\end{aligned} \tag{3-7}
$$

which is equivalent to

$$
\begin{aligned}
\int_M u^\beta J\,d\sigma + \int_{\partial M}\mathrm{II}(\nabla^\partial u,\nabla^\partial u)u^\beta\,d\sigma
&= \frac{m-1}{m}\int_M u^\beta(\Delta_h u)^2\,d\sigma - \tfrac{3}{2}\beta\int_M u^{\beta-1}(\Delta_h u)|\nabla u|^2\,d\sigma \\
&\quad + \frac{\beta(\beta-1)}{2}\int_M u^{\beta-2}|\nabla u|^4\,d\sigma.
\end{aligned} \tag{3-8}
$$

Proof of (i). To prove identity (3-3), let us calculate the first two terms on the right hand side of (3-8). First:

$$
\begin{aligned}
I_1 &:= \frac{m-1}{m}\int_M u^\beta(\Delta_h u)^2\,d\sigma \\
&= \frac{m-1}{m}\int_M u^\beta\Big(\alpha u^{1+(q-1)/\alpha} - \alpha\lambda u - \Big(\frac{\alpha-1}{\alpha}\Big)\frac{|\nabla u|^2}{u}\Big)\Delta_h u\,d\sigma \\
&= \alpha\frac{m-1}{m}\Big(\beta+1+\frac{q-1}{\alpha}\Big)\int_M u^{\beta+(q-1)/\alpha}|\nabla u|^2\,d\sigma \\
&\qquad\qquad - \alpha\lambda\frac{m-1}{m}(\beta+1)\int_M u^\beta|\nabla u|^2\,d\sigma \\
&\quad - \frac{m-1}{m}\Big(\frac{\alpha-1}{\alpha}\Big)\int_M u^{\beta-1}|\nabla u|^2\Big(\alpha u^{1+(q-1)/\alpha} - \alpha\lambda u - \Big(\frac{\alpha-1}{\alpha}\Big)\frac{|\nabla u|^2}{u}\Big)\,d\sigma \\
&= \frac{m-1}{m}(\alpha\beta+q)\int_M u^{\beta+(q-1)/\alpha}|\nabla u|^2\,d\sigma - \lambda\frac{m-1}{m}(\alpha\beta+1)\int_M u^\beta|\nabla u|^2\,d\sigma \\
&\qquad\qquad + \frac{m-1}{m}\Big(\frac{\alpha-1}{\alpha}\Big)^2\int_M u^{\beta-2}|\nabla u|^4\,d\sigma,
\end{aligned}
$$

where we used integration by parts with respect to $d\sigma$. A similar type of calculation yields

$$\begin{aligned} I_2 &:= -\tfrac{3}{2}\beta \int_M u^{\beta-1}(\Delta_h u)|\nabla u|^2\,d\sigma \\ &= -\tfrac{3}{2}\beta \int_M u^{\beta-1}\Big(\alpha u^{1+(q-1)/\alpha} - \alpha\lambda u - \Big(\frac{\alpha-1}{\alpha}\Big)\frac{|\nabla u|^2}{u}\Big)|\nabla u|^2\,d\sigma \\ &= -\tfrac{3}{2}\alpha\beta \int_M u^{\beta+(q-1)/\alpha}|\nabla u|^2\,d\sigma + \tfrac{3}{2}\alpha\beta\lambda \int_M u^{\beta}|\nabla u|^2\,d\sigma \\ &\qquad\qquad + \tfrac{3}{2}\beta\Big(\frac{\alpha-1}{\alpha}\Big)\int_M u^{\beta-2}|\nabla u|^4\,d\sigma. \end{aligned}$$

Incorporating I_1 and I_2 in (3-8), we complete the proof of (3-3) by deducing that

$$\begin{aligned} &\int_M u^{\beta} J\,d\sigma + \int_{\partial M} \mathrm{II}(\nabla^{\partial} u, \nabla^{\partial} u)u^{\beta}\,d\sigma \\ &\quad = \Big\{\frac{m-1}{m}(\alpha\beta+q) - \tfrac{3}{2}\alpha\beta\Big\}\int_M u^{\beta+(q-1)/\alpha}|\nabla u|^2\,d\sigma \\ &\qquad + \Big\{-\lambda\frac{m-1}{m}(\alpha\beta+1) + \tfrac{3}{2}\alpha\beta\lambda\Big\}\int_M u^{\beta}|\nabla u|^2\,d\sigma \\ &\qquad + \Big\{\frac{m-1}{m}\Big(\frac{\alpha-1}{\alpha}\Big)^2 + \tfrac{3}{2}\beta\Big(\frac{\alpha-1}{\alpha}\Big) + \frac{\beta(\beta-1)}{2}\Big\}\int_M u^{\beta-2}|\nabla u|^4\,d\sigma. \end{aligned} \tag{3-9}$$

Proof of (ii). Let us now give the proof of (3-4). The main idea here is to replace the second term on the right hand side of (3-8) $\big($i.e., $\int_M u^{\beta-1}(\Delta_h u)|\nabla u|^2\,d\sigma\big)$ by an expression in which the sign is controllable. For that, we multiply the first equation of (3-6) by $u^{\beta-1}|\nabla u|^2$, and we obtain after integrating with respect to $d\sigma$,

$$\begin{aligned} &\int_M u^{\beta-1}(\Delta_h u)|\nabla u|^2\,d\sigma \\ &\qquad = \alpha \int_M (u^{\beta+(q-1)/\alpha} - \lambda u^{\beta})|\nabla u|^2\,d\sigma - \Big(\frac{\alpha-1}{\alpha}\Big)\int_M u^{\beta-2}|\nabla u|^4\,d\sigma. \end{aligned} \tag{3-10}$$

Similarly, multiplying the same equation of (3-6) by $u^{\beta}\Delta_h u$ and integrating by parts yields

$$\begin{aligned} &\int_M u^{\beta}(\Delta_h u)^2\,d\sigma \\ &\quad = -\frac{\alpha-1}{\alpha}\int_M u^{\beta-1}(\Delta_h u)|\nabla u|^2\,d\sigma + \alpha\int_M \big(u^{\beta+1+(q-1)/\alpha} - \lambda u^{\beta+1}\big)\Delta_h u\,d\sigma \\ &\quad = -\frac{\alpha-1}{\alpha}\int_M u^{\beta-1}(\Delta_h u)|\nabla u|^2\,d\sigma + \alpha\Big(\beta+1+\frac{q-1}{\alpha}\Big)\int_M u^{\beta+(q-1)/\alpha}|\nabla u|^2\,d\sigma \\ &\qquad\qquad - \alpha\lambda(\beta+1)\int_M u^{\beta}|\nabla u|^2\,d\sigma. \end{aligned} \tag{3-11}$$

In order to eliminate the term

$$\int_M u^{\beta+(q-1)/\alpha}|\nabla u|^2 d\sigma,$$

we multiply (3-10) by $(\beta+1+(q-1)/\alpha)$ and subtract it from (3-11) to get

$$\begin{aligned}\text{(3-12)}\quad \lambda(q-1)\int_M u^\beta|\nabla u|^2\,d\sigma &= -\Big(\beta+\frac{q}{\alpha}\Big)\int_M u^{\beta-1}(\Delta_h u)|\nabla u|^2\,d\sigma+\int_M u^\beta(\Delta_h u)^2\,d\sigma\\ &\quad-\Big(\frac{\alpha-1}{\alpha}\Big)\Big(\beta+1+\frac{q-1}{\alpha}\Big)\int_M u^{\beta-2}|\nabla u|^4\,d\sigma.\end{aligned}$$

On the other hand, by a straightforward calculation, we have for $\beta\neq -2$,

$$\begin{aligned}\text{(3-13)}\quad \int_M u^{\beta-1}(\Delta_h u)|\nabla u|^2\,d\sigma &= -\frac{4}{\beta(\beta+2)^2}\int_M(\Delta_h u^{(\beta+2)/2})^2\,d\sigma\\ &\quad+\frac{\beta}{4}\int_M u^{\beta-2}|\nabla u|^4\,d\sigma+\frac{1}{\beta}\int_M u^\beta(\Delta_h u)^2\,d\sigma.\end{aligned}$$

Now, we replace $\int_M u^{\beta-1}(\Delta_h u)|\nabla u|^2\,d\sigma$ by its expression given in (3-13) in the equations (3-8) and (3-12) respectively. So (3-8) gives

$$\begin{aligned}\text{(3-14)}\quad \int_M u^\beta J\,d\sigma+\int_{\partial M}u^\beta \mathrm{II}(\nabla^\partial u,\nabla^\partial u)\,d\sigma &= -\frac{m+2}{2m}\int_M u^\beta(\Delta_h u)^2\,d\sigma+\frac{6}{(\beta+2)^2}\int_M(\Delta_h u^{(\beta+2)/2})^2\,d\sigma\\ &\quad+\frac{\beta(\beta-4)}{8}\int_M u^{\beta-2}|\nabla u|^4\,d\sigma,\end{aligned}$$

and the Equation (3-12) gives

$$\begin{aligned}\text{(3-15)}\quad \lambda(q-1)\int_M u^\beta|\nabla u|^2\,d\sigma &= \frac{-q}{\alpha\beta}\int_M u^\beta(\Delta_h u)^2\,d\sigma+\frac{4(\beta+q/\alpha)}{\beta(\beta+2)^2}\int_M(\Delta_h u^{(\beta+2)/2})^2\,d\sigma\\ &\quad-\bigg(\frac{\beta}{4}\Big(\beta+\frac{q}{\alpha}\Big)+\Big(\frac{\alpha-1}{\alpha}\Big)\Big(\beta+1+\frac{q-1}{\alpha}\Big)\bigg)\int_M u^{\beta-2}|\nabla u|^4\,d\sigma.\end{aligned}$$

Thus in order to eliminate the term

$$\int_M u^\beta(\Delta_h u)^2\,d\sigma$$

from (3-14) and (3-15), we multiply (3-14) by $(q/\alpha\beta)$ and (3-15) by $(m+2)/2m$

and subtract them to obtain

$$
\begin{aligned}
\text{(3-16)}\quad & A\int_M u^{\beta-2}|\nabla u|^4\,d\sigma \\
&= -\frac{q}{\alpha\beta}\left(\int_M u^\beta J\,d\sigma + \int_{\partial M} u^\beta \mathrm{II}(\nabla^\partial u,\nabla^\partial u)\,d\sigma\right) - B\int_M (\Delta_h u^{(\beta+2)/2})^2\,d\sigma \\
&\qquad + \left(\frac{m+2}{2m}\right)\lambda(q-1)\int_M u^\beta|\nabla u|^2\,d\sigma,
\end{aligned}
$$

where the expressions of A and B are given by

$$
\begin{aligned}
A &= -\frac{m+2}{2m}\left\{\left(\frac{\beta}{4}+\frac{\alpha-1}{\alpha}\right)\left(\beta+\frac{q}{\alpha}\right)+\left(\frac{\alpha-1}{\alpha}\right)^2\right\} - \frac{q(\beta-4)}{8\alpha}, \\
B &= \frac{2}{m(\beta+2)^2}\left\{(m+2)-2(m-1)\frac{q}{\alpha\beta}\right\}
\end{aligned}
$$

This completes the proof of (3-4) in the case $\beta \neq -2$.

Concerning the case where $\beta = -2$, the second term on the right hand side of (3-16) can be written as

$$
\begin{aligned}
& B\int_M (\Delta_h u^{(\beta+2)/2})^2\,d\sigma \\
&\qquad = \frac{1}{2m}\left\{(m+2)-2(m-1)\frac{q}{\alpha\beta}\right\}\int_M\left(\Delta_h\left(\frac{u^{(\beta+2)/2}-u^0}{(\beta+2)/2}\right)\right)^2 d\sigma,
\end{aligned}
$$

and when β tends to -2 in (3-16), we obtain (3-5). Therefore the proof of Proposition 3.1 is completed. □

The first uniqueness result in this paper is the following:

Proposition 3.2. *Let (M^n, g, σ) be a compact weighted Riemannian manifold of dimension $n \geq 2$ with convex boundary ∂M. Assume that for some $m \in [n,\infty)$, the m-Bakry–Émery–Ricci curvature satisfies $\mathrm{Ric}_h^m \geq kg$, for a positive constant k. Let $q > 1$, $\lambda > 0$ and φ be a positive solution of the following system*:

$$
\text{(3-17)}\qquad \begin{cases} \Delta_h\varphi + \lambda\varphi = \varphi^q & \text{in } M, \\ \dfrac{\partial\varphi}{\partial\nu} = 0 & \text{on } \partial M. \end{cases}
$$

Suppose that

$$
q \leq \frac{m+2}{m-2}
$$

with no restriction on q if $m = 2$,

$$
\lambda \leq \frac{mk}{(m-1)(q-1)}
$$

and one of these two inequalities is strict, then φ is constant equal to $\lambda^{1/(q-1)}$.

In addition, if (M, g) is of constant scalar curvature R, nonisometric to the hemisphere $\mathbb{S}^n_+(\sqrt{n(n-1)/R})$, then

$$\lambda = \frac{mk}{(m-1)(q-1)} \quad \text{and} \quad q = \frac{m+2}{m-2}$$

ensure that φ is constant equal to $\lambda^{1/(q-1)}$.

Remark 3.3. • In the last Proposition, one can consider more generally an equation of the form

$$\Delta_h \varphi + \lambda \varphi = \mu \varphi^q,$$

where μ is a positive constant. In fact, if we take $\widetilde{\varphi} = \mu^{1/(q-1)} \varphi$, we can easily obtain $\Delta_h \widetilde{\varphi} + \lambda \widetilde{\varphi} = \widetilde{\varphi}^q$.

• *The solutions in the spherical case*: In the Riemannian case (i.e., the case $m = n$), the solutions are well known and they are related to the metrics conformal to the standard metric on the sphere, respectively the hemisphere (see for instance [Aubin 1982; Escobar 1990; 1992; Lee and Parker 1987]).

• *Unicity of an Einstein metric in a conformal class*: For (M^n, g) a compact Einstein manifold of totally geodesic boundary ∂M, Escobar [1990] proved that if g_1 is a metric conformally related to g with constant scalar curvature and for which ∂M is minimal, then g_1 is Einstein. Moreover, if (M^n, g) is not conformally equivalent to $\mathbb{S}^n_+$, then $g_1 = cg$ for some positive constant c. In fact, one can easily see that if $g_1 = u^{4/(n-2)} g$, then the scalar curvatures R_g and R_{g_1} satisfy

$$\begin{cases} \Delta u + \dfrac{n-2}{4(n-1)} R_g u = \dfrac{n-2}{4(n-1)} R_{g_1} u^{(n+2)/(n-2)} & \text{in } M, \\ \dfrac{\partial u}{\partial \nu} = 0 & \text{on } \partial M. \end{cases}$$

Now, when the scalar curvatures are positive (this is the difficult part in Escobar's result), using Proposition 3.2 one can easily deduce that u is constant. Note that Escobar's result is the generalization of the classical uniqueness result of Obata for Einstein compact manifolds without boundary (see for instance [Obata 1962]).

The situation is more complicated in the case of weighted Riemannian manifolds as one can see, for instance, in [Case et al. 2011; Case 2015; Chang et al. 2006; 2011]. Nevertheless, we expect that there is an analogue of Escobar's theorem for weighted Riemannian manifolds.

Proof of Proposition 3.2. One can easily infer from (3-3) the following:

$$\begin{aligned} \int_M u^\beta \big(J - \mathrm{Ric}^m_h(\nabla u, \nabla u)\big)\, d\sigma &+ \int_{\partial M} \mathrm{II}(\nabla^\partial u, \nabla^\partial u) u^\beta \, d\sigma \\ &= A_1 \int_M u^{\beta+(q-1)/\alpha} |\nabla u|^2 \, d\sigma + B_1 \int_M u^\beta |\nabla u|^2 \, d\sigma \\ &\qquad + C_1 \int_M u^{\beta-2} |\nabla u|^4 \, d\sigma - \int_M u^\beta \, \mathrm{Ric}^m_h(\nabla u, \nabla u). \end{aligned}$$

Using the hypothesis that the m-Bakry–Émery–Ricci curvature satisfies $\mathrm{Ric}_h^m \geq kg$, and the fact that ∂M is convex (i.e., $\mathrm{II} \geq 0$), we obtain:

$$\int_M u^\beta \big(J - \mathrm{Ric}_h^m(\nabla u, \nabla u)\big)\, d\sigma \leq A_1 \int_M u^{\beta+(q-1)/\alpha} |\nabla u|^2 \, d\sigma + (B_1 - k) \int_M u^\beta |\nabla u|^2 \, d\sigma + C_1 \int_M u^{\beta-2} |\nabla u|^4 \, d\sigma, \tag{3-18}$$

where A_1, B_1 and C_1 are as given in Proposition 3.1. Since Lemma 2.2 asserts that J is bounded below by $\mathrm{Ric}_h^m(\nabla u, \nabla u)$, it suffices to show the existence of α and β such that $A_1 \leq 0$, $(B_1 - k) \leq 0$, $C_1 \leq 0$ and at least one of these three inequalities is strict, in order to conclude from Equation (3-18) that u (and hence φ) is a constant.

By arguing as in [Ilias 1996], we see that if

$$q \leq \frac{m+2}{m-2} \quad \text{and} \quad \lambda \leq \frac{mk}{(m-1)(q-1)}$$

then there exist α, β such that A_1, $(B_1 - k)$, $C_1 \leq 0$. Moreover, if one of the above two inequalities is strict, we may choose α, β such that A_1, $(B_1 - k)$, $C_1 \leq 0$ and at least one of these inequalities is strict.

Now suppose that (M, g) is of constant scalar curvature R, nonisometric to the hemisphere $\mathbb{S}_+^n(\sqrt{n(n-1)/R})$. If $q = (m+2)/(m-2)$ and $\lambda = mk/(m-1)(q-1)$, then one can choose α and β such that $A_1 = B_1 - k = C_1 = 0$. From (3-18) we conclude that $J - \mathrm{Ric}_h^m(\nabla u, \nabla u) = 0$, which is equivalent by Lemma 2.2, to

$$D^2 u = -\frac{1}{n}(\Delta u) g \quad \text{and} \quad \Delta_h u = \frac{m}{m-n}\langle \nabla u, \nabla h\rangle. \tag{3-19}$$

Suppose that u is not constant and consider the vector field $Y = \nabla u$. First of all, Y is a conformal vector field because the first equality of (3-19) is nothing but $\mathcal{L}_Y g = (2/n)\rho g$, where $\rho = \operatorname{div} Y = -\Delta u$. Since Y is conformal and R is constant, we have (see equation (1.11) of [Yano and Obata 1970])

$$\mathcal{L}_Y(R) = Y(R) = \frac{2(n-1)}{n}\Delta\rho - \frac{2}{n}\rho R = 0,$$

and consequently

$$\Delta u - \frac{R}{n-1} u = \text{constant}. \tag{3-20}$$

We immediately deduce from (3-20) that $\frac{\partial}{\partial \nu}(\Delta u)\,|_{\partial M} = 0$.

Differentiating (3-20) two times we get

$$\begin{cases} D^2\rho + \dfrac{R}{n(n-1)}\rho g = 0 & \text{in } M, \\ \dfrac{\partial \rho}{\partial \nu} = 0 & \text{on } \partial M. \end{cases} \tag{3-21}$$

Since we have supposed that u is not constant, $\rho = -\Delta u$ is not identically zero on M and we can deduce from Escobar's theorem [1990, Theorem 4.2] that (M, g) is isometric to the upper hemisphere $\mathbb{S}^n_+(\sqrt{n(n-1)/R})$, which contradicts our hypothesis. Thus u (hence φ) is constant. $\square$

Now, we shall prove another kind of uniqueness result (under different conditions on λ) for the same nonlinear elliptic PDE (3-17) which generalizes a result of Licois and Véron [1995]. This result involves the first nonzero eigenvalue λ_1^h of the weighted Laplacian Δ_h under the Neumann boundary condition.

Proposition 3.4. *Let (M^n, g, σ) be a compact weighted Riemannian manifold of dimension $n \geq 2$ with convex boundary ∂M. Assume that for some $m \in [n, \infty)$, the m-Bakry–Émery–Ricci curvature satisfies $\mathrm{Ric}_h^m \geq kg$, for a positive constant k. Let $q > 1$, $\lambda > 0$ and φ be a positive solution of the following system*

$$\begin{cases} \Delta_h \varphi + \lambda\varphi = \varphi^q & \text{in } M, \\ \dfrac{\partial\varphi}{\partial\nu} = 0 & \text{on } \partial M. \end{cases} \tag{3-22}$$

Suppose that

$$q \leq \frac{m+2}{m-2} \tag{3-23}$$

with no restriction on q if $m = 2$,

$$(q-1)\lambda \leq \lambda_1^h + \frac{qm(m-1)}{q+m(m+2)}\left(k - \frac{m-1}{m}\lambda_1^h\right) \tag{3-24}$$

and one of these two inequalities is strict, then φ is constant equal to $\lambda^{1/(q-1)}$.

In addition, if (M, g) is of constant scalar curvature R nonisometric to the hemisphere $\mathbb{S}^n_+(\sqrt{n(n-1)/R})$, then

$$(q-1)\lambda = \lambda_1^h + \frac{qm(m-1)}{q+m(m+2)}\left(k - \frac{m-1}{m}\lambda_1^h\right) \quad \text{and} \quad q = \frac{m+2}{m-2}$$

ensure that φ is constant equal to $\lambda^{1/(q-1)}$.

Remark 3.5. We observe that the term $(k - ((m-1)/m)\lambda_1^h)$ in Equation (3-24) is always nonpositive and this is due to the Escobar–Lichnerowicz theorem (see [Escobar 1990, Theorem 4.3]) generalized to weighted Riemannian manifolds with convex boundary and satisfying $\mathrm{Ric}_h^m \geq kg > 0$ (see, for example, [Li and Wei 2015, Theorem 3] or [Ma and Du 2010, Theorem 2]). Moreover, (3-24) can be rewritten in the following form:

$$(q-1)\lambda \leq \left(1 - \frac{q(m-1)^2}{q+m(m+2)}\right)\lambda_1^h + \left(\frac{qm(m-1)}{q+m(m+2)}\right)k. \tag{3-25}$$

The coefficient of λ_1^h in (3-25) is positive if $q < (m+2)/(m-2)$ and equal to zero if $q = (m+2)/(m-2)$.

Proof of Proposition 3.4. From (3-4) of Proposition 3.1, we easily infer for $\beta \neq -2$,

$$\begin{aligned} A\int_M u^{\beta-2}|\nabla u|^4\,d\sigma &= -\frac{q}{\alpha\beta}\left(\int_M u^\beta J\,d\sigma + \int_{\partial M} u^\beta \mathrm{II}(\nabla^\partial u, \nabla^\partial u)\,d\sigma\right) - B\int_M (\Delta_h u^{(\beta+2)/2})^2\,d\sigma \\ &\qquad + \left(\frac{m+2}{2m}\right)\lambda(q-1)\int_M u^\beta|\nabla u|^2\,d\sigma, \end{aligned} \tag{3-26}$$

where the expressions of A and B are given by

$$\begin{aligned} A &= -\frac{m+2}{2m}\left\{\left(\frac{\beta}{4}+\frac{\alpha-1}{\alpha}\right)\left(\beta+\frac{q}{\alpha}\right)+\left(\frac{\alpha-1}{\alpha}\right)^2\right\} - \frac{q(\beta-4)}{8\alpha}, \\ B &= \frac{2}{m(\beta+2)^2}\left\{(m+2)-2(m-1)\frac{q}{\alpha\beta}\right\}. \end{aligned}$$

The idea is to replace the integral associated to B by one of the form $\int_M u^\beta|\nabla u|^2\,d\sigma$. Since $\frac{\partial u}{\partial \nu} = 0$ over the boundary ∂M, the variational characterization of λ_1^h yields

$$\begin{aligned} \int_M (\Delta_h u^{(\beta+2)/2})^2 d\sigma &\geq \frac{(\beta+2)^2}{4}\lambda_1^h \int_M u^\beta |\nabla u|^2\,d\sigma \quad &&\text{for } \beta \neq -2, \\ \int_M (\Delta_h \ln u)^2\,d\sigma &\geq \lambda_1^h \int_M |\nabla \ln u|^2\,d\sigma &&\text{for } \beta = -2. \end{aligned} \tag{3-27}$$

Therefore if we can find a couple of nonzero real numbers (α, β) such that

$$\alpha\beta > 0, \quad A \geqq 0 \quad \text{and} \quad B \geqq 0, \tag{3-28}$$

then by using the relation (3-27), the hypothesis on the m-Bakry–Émery–Ricci curvature and the convexity of ∂M, we deduce from (3-26),

$$\begin{aligned} 0 &\leq A\int_M u^{\beta-2}|\nabla u|^4\,d\sigma \\ &\leq \frac{-q}{\alpha\beta}\int_M u^\beta\big(J - \mathrm{Ric}_h^m(\nabla u, \nabla u)\big)\,d\sigma + C\int_M u^\beta|\nabla u|^2\,d\sigma, \end{aligned} \tag{3-29}$$

where

$$\begin{aligned} C &:= \frac{-q}{\alpha\beta}k - \frac{(\beta+2)^2}{4}\lambda_1^h B + \left(\frac{m+2}{2m}\right)(q-1)\lambda \\ &= \frac{m+2}{2m}(\lambda(q-1)-\lambda_1^h) - \frac{q}{\alpha\beta}\left(k - \frac{m-1}{m}\lambda_1^h\right), \end{aligned}$$

and using the hypothesis (3-23) as well as (3-24) concerning λ, we can find a couple among the (α, β) satisfying the conditions (3-28) such that C is nonpositive.

Moreover, when the equality is not achieved in (3-23) or in (3-24) we will be able to conclude that u is constant.

For $\beta = -2$, an immediate modification where we use identity (3-5) instead of (3-4) permits us to conclude.

In the rest of the proof we will show how to find such a couple (α, β) when $\beta \neq -2$. First we simplify the expressions of A, B and C, by setting

$$X = \frac{-1}{\alpha\beta}, \quad \delta = \frac{1}{\beta} + \frac{1}{2}, \quad \text{and} \quad \widetilde{A} = \frac{2m}{(m+2)\beta^2} A$$

to obtain

$$\begin{aligned}
\widetilde{A} &= -\delta^2 + 2\frac{q-(m+2)}{m+2} X\delta + (q-1)X^2 + \frac{(m-1)Xq}{2(m+2)}, \\
B &= \frac{2}{m(\beta+2)^2}\big((m+2) + 2(m-1)Xq\big), \\
C &= \frac{m+2}{2m}\big(\lambda(q-1) - \lambda_1^h\big) + Xq\Big(k - \frac{m-1}{m}\lambda_1^h\Big),
\end{aligned}$$

and then we maximize X in the interval

$$I := \Big[-\frac{m+2}{2(m-1)q}, 0\Big)$$

(which ensures that $B \geq 0$ and $\alpha\beta > 0$) such that

$$\widetilde{A} = -\delta^2 + 2\frac{q-(m+2)}{m+2} X\delta + (q-1)X^2 + \frac{(m-1)}{2(m+2)} Xq \geq 0.$$

The derivative of $\widetilde{A}$ with respect to δ is given by

$$\frac{d\widetilde{A}}{d\delta} = -2\Big(\delta - \frac{q-(m+2)}{m+2} X\Big).$$

Therefore the maximum of $\widetilde{A}$ with respect to δ is achieved for

$$\delta_0 := \frac{q-(m+2)}{m+2} X$$

and thus:

$$\begin{aligned}
\widetilde{A}(\delta_0, X) &= \delta_0^2 + (q-1)X^2 + \frac{(m-1)}{2(m+2)} Xq \\
&= \bigg(q - 1 + \Big(\frac{q-(m+2)}{m+2}\Big)^2\bigg) X^2 + \frac{(m-1)}{2(m+2)} Xq,
\end{aligned}$$

which admits a nontrivial negative solution

$$X_0 = -\frac{(m-1)(m+2)}{2\big(q + m(m+2)\big)}.$$

Using the hypothesis that $q \le (m+2)/(m-2)$, one has $X_0 \in I$ and therefore

$$\widetilde{A}(\delta_0, X) \ge 0 \quad \text{on } \Big[-\frac{m+2}{2(m-1)q}, X_0\Big] \subseteq I.$$

Moreover, a direct computation of C at the specific value $X = X_0$ gives

$$C(X_0) = \frac{m+2}{2m}\Big(\lambda(q-1) - \lambda_1^h - \frac{qm(m-1)}{q+m(m+2)}\Big(k - \frac{m-1}{m}\lambda_1^h\Big)\Big).$$

Thus, if we suppose that

$$q \le \frac{m+2}{m-2} \tag{3-30}$$

and

$$\lambda(q-1) \le \lambda_1^h + \frac{qm(m-1)}{q+m(m+2)}\Big(k - \frac{m-1}{m}\lambda_1^h\Big) \tag{3-31}$$

then we have two possibilities:

(1) The equality in (3-31) is not achieved (i.e., $C(X_0) < 0$). In this case, we obtain from (3-29) at $X = X_0$,

$$C(X_0)\int_M u^\beta |\nabla u|^2\, d\sigma = 0,$$

and since $C(X_0) < 0$, we deduce that u is constant.

(2) The equality in (3-31) is achieved (i.e., $C(X_0) = 0$) and the inequality (3-30) is strict. In this case, one can deduce that all the inequalities used to obtain (3-29) are in fact equalities. In particular, one has

$$B(X_0)\int_M (\Delta_h u^{(\beta+2)/2})^2 d\sigma = B(X_0)\frac{(\beta+2)^2}{4}\lambda_1^h \int_M u^\beta |\nabla u|^2\, d\sigma$$

and since (3-30) is strict, $B(X_0)$ is positive. Therefore

$$\int_M (\Delta_h u^{(\beta+2)/2})^2\, d\sigma = \lambda_1^h \int_M |\nabla u^{(\beta+2)/2}|^2\, d\sigma. \tag{3-32}$$

Thus, if u is not constant, then $u^{(\beta+2)/2}$ is an eigenfunction associated to λ_1^h and since $\frac{\partial u}{\partial \nu}|_{\partial M} = 0$, we have

$$\int_M u^{(\beta+2)/2}\, d\sigma = 0,$$

which contradicts the fact that u is positive. In conclusion u is constant.

To prove the last assertion in the proposition, suppose that (M, g) is of constant scalar curvature R, nonisometric to the hemisphere $\mathbb{S}^n_+(\sqrt{n(n-1)/R})$. If (3-30) and (3-31) are equalities, then we can conclude from (3-29) that $J - \mathrm{Ric}_h^m(\nabla u, \nabla u) = 0$.

Similar arguments as those used in the proof of Proposition 3.2 allow us to conclude that u is constant. □

Now we will deduce our Sobolev inequalities from the previous uniqueness results.

Theorem 3.6. *Let (M^n, g, σ) be a compact weighted Riemannian manifold of dimension $n \geq 2$ with convex boundary (i.e., $\mathrm{II} \geq 0$). Assume that for some $m \in [n, \infty)$, the m-Bakry–Émery–Ricci curvature satisfies $\mathrm{Ric}_h^m \geq kg$ for a positive constant k. Then every function $f \in H_1^2(d\sigma)$ satisfies*

$$\|f\|^2_{L^p(d\sigma)} \leq \mathrm{vol}_h(M)^{-(p-2)/p}\left(\frac{1}{\theta(m,p)}\|\nabla f\|^2_{L^2(d\sigma)} + \|f\|^2_{L^2(d\sigma)}\right), \tag{3-33}$$

for any $p \in (2, 2^*]$ with $2^* = 2m/(m-2)$ if $m > 2$ and for any $p \in (2, \infty)$ if $m = 2$ (i.e., $n = 2$ and h is constant) with $\theta(m, p) \in \{\theta_1(m, p), \theta_2(m, p)\}$ where

$$\theta_1(m,p) = \frac{mk}{(m-1)(p-2)},$$
$$\theta_2(m,p) = \frac{\lambda_1^h}{(p-2)} + \frac{m(m-1)(p-1)}{\big((p-1)+m(m+2)\big)(p-2)}\left(k - \frac{m-1}{m}\lambda_1^h\right).$$

Using the Escobar–Lichnerowicz theorem (see Section 4), we note that for $m > 2$

$$\theta_1(m,p) - \theta_2(m,p) = \frac{m(m+2) - m(m-2)(p-1)}{\big((p-1)+m(m+2)\big)(p-2)}\left(\frac{mk}{m-1} - \lambda_1^h\right) \leq 0$$

and for $m = 2$,

$$\theta_1(2,p) - \theta_2(2,p) = \frac{8}{((p-1)+8)(p-2)}(2k - \lambda_1^h) \leq 0$$

therefore, (3-33) is better with $\theta(m,p) = \theta_2(m,p)$ than with $\theta_1(m,p)$.

On the other hand, when $m > 2$ and p tends to the critical exponent $2^* = 2m/(m-2)$, we have

$$\theta_1(m, 2^*) = \theta_2(m, 2^*) = \frac{m(m-2)k}{4(m-1)}.$$

Therefore, one limiting case of Theorem 3.6 gives the following:

Corollary 3.7. *Let (M^n, g, σ) be a compact weighted Riemannian manifold of dimension $n \geq 2$ and of convex boundary (i.e., $\mathrm{II} \geq 0$). Assume that for some $m \in [n, \infty)$, the m-Bakry–Émery–Ricci curvature satisfies $\mathrm{Ric}_h^m \geq kg$, for a positive constant k. If $m > 2$, then every $f \in H_1^2(d\sigma)$ satisfies*

$$\|f\|^2_{L^{2m/(m-2)}(d\sigma)} \leq \mathrm{vol}_h(M)^{-2/m}\left(\frac{4(m-1)}{m(m-2)k}\|\nabla f\|^2_{L^2(d\sigma)} + \|f\|^2_{L^2(d\sigma)}\right). \tag{3-34}$$

Proof of Theorem 3.6. Suppose that $m > 2$, and consider the following family of functionals $\mathcal{J}_q$, defined by

$$\mathcal{J}_q(\varphi) = \int_M |\nabla\varphi|^2\, d\sigma + \Theta(m,q)\int_M \varphi^2\, d\sigma \quad \text{for } 1 < q < \frac{m+2}{m-2}$$

with $\Theta(m,q) \in \{\Theta_1(m,q), \Theta_2(m,q)\}$ where

$$\begin{aligned}\Theta_1(m,q) &= \frac{mk}{(m-1)(q-1)}\\ \Theta_2(m,q) &= \frac{\lambda_1^h}{(q-1)} + \frac{qm(m-1)}{(q-1)\big(q+m(m+2)\big)}\Big(k - \frac{m-1}{m}\lambda_1^h\Big)\end{aligned}$$

and consider $\mu_q := \inf\{\mathcal{J}_q(\varphi),\ \varphi \in \mathcal{H}_q\}$, where

$$\mathcal{H}_q := \Big\{\varphi \in H_1^2(d\sigma) : \int_M \varphi^{q+1}\, d\sigma = 1\Big\}.$$

Another crucial key here is the fact that the real-valued function

$$g : x \longmapsto \frac{x+2}{x-2}$$

is decreasing. So

$$\frac{m+2}{m-2} = g(m) < g(n) = \frac{n+2}{n-2}$$

as $n < m$.

Using the compactness of the inclusions

$$H_1^2(d\sigma) \hookrightarrow L^2(d\sigma) \quad \text{and} \quad H_1^2(d\sigma) \hookrightarrow L^{q+1}(d\sigma)$$

for any $q+1 < 2n/(n-2)$, we can prove that μ_q is achieved by a positive function $\psi_q \in \mathcal{H}_q$ and therefore, one can easily check that ψ_q verifies weakly the following system:

$$\text{(3-35)} \qquad \begin{cases} \Delta_h\psi_q + \Theta(m,q)\psi_q = \mu_q\psi_q^q & \text{in } M,\\ \dfrac{\partial\psi_q}{\partial\nu} = 0 & \text{on } \partial M.\end{cases}$$

Since h is smooth, the regularity result of Cherrier [1984, Theorem 1] shows that ψ_q is smooth, and hence by applying Proposition 3.2 if $\Theta(m,q) = \Theta_1(m,q)$ or Proposition 3.4 if $\Theta(m,q) = \Theta_2(m,q)$, we deduce that ψ_q is constant. Since $\psi_q \in \mathcal{H}_q$, we get

$$\psi_q = (\mathrm{vol}_h(M))^{-1/(q+1)} \quad \text{and} \quad \mu_q = \Theta(m,q)(\mathrm{vol}_h(M))^{(q-1)/(q+1)}.$$

Therefore, one can deduce from the definition of μ_q, that any $f \in H_1^2(d\sigma)$ satisfies

$$\left(\int_M |f|^{q+1}\, d\sigma\right)^{2/(q+1)} \le \mathrm{vol}_h(M)^{-(q-1)/(q+1)}\left(\frac{1}{\Theta(m,q)}\int_M |\nabla f|^2\, d\sigma + \int_M |f|^2\, d\sigma\right). \tag{3-36}$$

If we put $p = q+1$, then $\Theta(m,q) = \Theta(m, p-1) = \theta(m, p)$ and thus (3-36) completes the proof of Theorem 3.6 for $m > 2$. The case where $m = 2$, can be treated in a similar manner. □

4. Some applications

We can derive many interesting applications from the weighted Sobolev inequalities we obtain. The list is long but for brevity we will limit ourselves to a few and more significant examples.

(I) *The classical Escobar–Lichnerowicz lower bound*: As in [Bakry and Ledoux 1996], if we apply (3-34) to $f = 1 + t\phi$ $(t > 0)$, where $\int_M \phi\, d\sigma = 0$, and use the Taylor expansion $(1+x)^p \simeq_{x\to 0} 1 + px + \frac{1}{2}p(p-1)x^2$, then we obtain the analogue of the classical Escobar–Lichnerowicz theorem for measured spaces with convex boundary:

$$\lambda_1^h \ge \frac{mk}{m-1}. \tag{4-1}$$

We should also point out that many authors have obtained this estimate (see, for example, [Li and Wei 2015, Theorem 3; Ma and Du 2010, Theorem 2]).

Similarly, one can look for lower bounds depending on k for higher eigenvalues:

(II) *Lower bounds of higher eigenvalues with explicit constants*: Inspired by the work of Cheng and Li [1981] in the case of a Riemannian manifold and using Poincaré and Hölder inequalities, we can easily derive from (3-34) that any $f \in H_1^2(d\sigma)$ such that $\int_M f\, d\sigma = 0$ satisfies

$$\int_M |\nabla f|^2\, d\sigma \ge \mathcal{C}_1 \left(\int_M f^2\, d\sigma\right)^{\frac{m+2}{m}} \left(\int_M |f|\, d\sigma\right)^{-\frac{4}{m}}, \tag{4-2}$$

where $\mathcal{C}_1 = \lambda_1^h/(\lambda_1^h \mathcal{C}_0 + 1)(\mathrm{vol}_h(M))^{2/m}$ with $\mathcal{C}_0 = 4(m-1)/(m(m-2)k)$. Applying this last inequality (4-2) to the weighted heat kernel $H_h(t, x, y)$ with Neumann condition on the boundary and after using its semigroup property (see for instance the book of Grigor'yan [2009] for the properties of such heat kernel), we are able to obtain an explicit upper bound for its trace. In fact, we obtain

$$\int_M \left(H_h(t,x,x) - \frac{1}{\mathrm{vol}_h(M)} \right) d\sigma = \sum_{i=1}^{\infty} e^{-\lambda_i^h t} \le 4 \left(\frac{m}{2\mathcal{C}_1} \right)^{m/2} \mathrm{vol}_h(M)\, t^{-m/2}. \tag{4-3}$$

From this upper bound, taking $t = 1/\lambda_\ell^h$, where λ_ℓ^h is the ℓ-th Neumann eigenvalue of Δ_h on M, we can easily deduce a lower bound of λ_ℓ^h as follows:

$$\lambda_\ell^h \ge \left(\frac{1}{4e} \right)^{2/m} \left(\frac{2}{m} \right) \mathcal{C}_1 \left(\frac{\ell}{\mathrm{vol}_h(M)} \right)^{2/m}. \tag{4-4}$$

In the particular case, where the manifold is without boundary (respectively with convex boundary) and the density function is constant (hence $m = n$), λ_ℓ^h is nothing but λ_ℓ , the ℓ-th eigenvalue of the usual Laplacian (resp. the Neumann Laplacian), and thus we recover the lower bound obtained in [Cheng and Li 1981] but with explicit constants. It's worthwhile to point out that if we use the estimate (4-1) for λ_1^h obtained above, then the constant $\mathcal{C}_1$ in (4-2) can be replaced by the constant

$$\mathcal{C}_2 := \frac{m-2}{m+2} (\mathrm{vol}_h(M))^{2/m} \frac{mk}{m-1}$$

and thus (4-4) becomes

$$\lambda_\ell^h \ge 2 \left(\frac{1}{4e} \right)^{2/m} \frac{(m-2)k}{(m+2)(m-1)} \ell^{2/m}.$$

In the same spirit, but inspired this time by the work of Li and Yau [1983], one can consider the Neumann heat kernel of the operator Δ_h/q, where q is a positive potential on M. In this case, using (4-2), one can deduce that

$$\lambda_\ell^h \left(\int_M q^{\frac{m}{2}} \, d\sigma \right)^{\frac{2}{m}} \ge \left(\frac{\mathcal{C}_1}{e} \right) \ell^{2/m}$$

and as above $\mathcal{C}_1$ can be replaced by $\mathcal{C}_2$. Using this last inequality, the same arguments as in Corollary 2 of [Li and Yau 1983] gives an explicit estimate of the number of eigenvalues for a weighted Schrödinger operator $\Delta_h + V$ which are less than or equal to a given value.

(III) *Lower bound for the Yamabe invariant*: • *The case of a compact Riemannian manifold without boundary*: Let (M, g) be a compact Riemannian manifold without boundary of dimension $n > 2$, and as before denote by dv_g its Riemannian measure and by R_g its scalar curvature. Let $[g]$ be the class of conformal metrics to g. The Yamabe invariant of the conformal class $[g]$ (see for instance [Aubin 1982; Hebey and Vaugon 1996; Lee and Parker 1987]) is given by

$$\mu(M,[g]) = \inf_{u \in C^1(M)\setminus\{0\}} \frac{\left(4(n-1)/(n-2) \int_M |\nabla u|^2 \, dv_g + \int_M R_g u^2 \, dv_g \right)}{\left(\int_M u^{2n/(n-2)} \, dv_g \right)^{(n-2)/n}} \tag{4-5}$$

and under the condition $\mathrm{Ric}(M, g) \geq kg$ (with $k > 0$), using the Sobolev inequality of Corollary 3.7 and the fact that $R_g \geq nk$, we obtain the following lower bound of (4-5):

$$\mu(M, [g]) \geq nk(\mathrm{vol}(M))^{2/n}. \tag{4-6}$$

Note that since

$$\mu(M, [g]) \leq \mu(\mathbb{S}^n, [\mathrm{can}]) = n\,(n-1)\,(\mathrm{vol}(\mathbb{S}^n))^{2/n}$$

(see [Aubin 1982]), we obtain the Bishop inequality:

$$\mathrm{vol}_g(M) \leq \left(\frac{n-1}{k}\right)^{n/2} \mathrm{vol}(\mathbb{S}^n).$$

We also observe that Petean [2005] deduced from (4-6) that if g_0 is the Fubini–Study metric on $\mathbb{C}P^2$ and g is any other metric on $\mathbb{C}P^2$ with $\mathrm{Ric}_g \geq \mathrm{Ric}_{g_0}$ then $\mathrm{vol}_g(\mathbb{C}P^2) \leq \mathrm{vol}_{g_0}(\mathbb{C}P^2)$.

• *The case of a compact Riemannian manifold with boundary*: This case is more complicated than the first one (see [Cherrier 1984; Escobar 1992; Akutagawa 2001]) even if the strategy is the same. In this case the boundary Yamabe invariant is given by

$$\mu(M, [g]) = \inf_{u \in C^1(M)\setminus\{0\}} \frac{\left(4(n-1)/(n-2)\int_M |\nabla u|^2\, dv_g + \int_M R_g u^2\, dv_g + 2\int_{\partial M} H u^2\, dv_g\right)}{\left(\int_M u^{2n/(n-2)}\, dv_g\right)^{(n-2)/n}}, \tag{4-7}$$

where H is the mean curvature of ∂M and dv_g denotes the induced Riemannian measure on ∂M. As before, under the conditions $\mathrm{Ric}(M, g) \geq kg$ (with $k > 0$) and ∂M convex, we deduce since, in this case, $R_g \geq n\,k$ and $H \geq 0$, the following lower bound for the boundary Yamabe invariant, similar to that for the Yamabe invariant:

$$\mu(M, [g]) \geq nk(\mathrm{vol}(M))^{2/n}. \tag{4-8}$$

Note that since

$$\mu(M, [g]) \leq \mu(\mathbb{S}^n_+, [\mathrm{can}]) = n(n-1)(\mathrm{vol}(\mathbb{S}^n_+))^{2/n}$$

(see [Escobar 1992]), we obtain the equivalent of the Bishop inequality when the boundary of the manifold is convex:

$$\mathrm{vol}_g(M) \leq \left(\frac{n-1}{k}\right)^{n/2} \tfrac{1}{2}\,\mathrm{vol}(\mathbb{S}^n).$$

• *The case of a measured Riemannian space*: Here we consider a weighted Riemannian manifold of dimension $n > 2$, and when the manifold is with boundary, H denotes the mean curvature of its boundary. As we observed in Remark 3.3, the situation is more complicated. We don't know if there's an equivalent of the Yamabe invariant related to our Sobolev inequality of Corollary 3.7. If we consider the following infimum:

$$\text{(4-9)}\quad \mu(M,m,g,\sigma) := \inf_{u\in C^1(M)\setminus\{0\}} \frac{\left(4(m-1)/(m-2)\int_M |\nabla u|^2\,d\sigma + \int_M R_h^m u^2\,d\sigma + 2\int_{\partial M} Hu^2\,d\sigma\right)}{\left(\int_M u^{2m/(m-2)}\,d\sigma\right)^{(m-2)/m}},$$

where $R_h^m = \frac{m}{n}\operatorname{trace}(\mathrm{Ric}_h^m)$ and as before, if we suppose that $\mathrm{Ric}_h^m \geq kg > 0$ and the boundary is convex in the case where $\partial M \neq \varnothing$, we obtain

$$\mu(M,m,g,\sigma) \geq mk(\mathrm{vol}_h(M))^{2/m}.$$

For an extension of the Yamabe invariant in the case of weighted manifolds, one can consult [Chang et al. 2011; Case 2015].

(IV) *Onofri and logarithmic Sobolev inequalities*: Another interesting application is the Onofri inequality (see for example [Onofri 1982; Beckner 1993]), which appears as an endpoint of various families of interpolation inequalities in dimension two, exactly like Sobolev inequality in higher dimensions. The following corollary gives the analogue of Onofri's inequality on any 2-dimensional compact Riemannian manifold (M^2, g) with positive curvature and convex boundary (see also [Ilias 1983] for nonsharp Onofri inequalities in all dimensions). Since in this case we take $m = 2$, we must have h constant, and finally the measure $d\sigma$ is just a multiple of the Riemannian measure. The inequality being invariant by homothety on the measure, we can restrict ourselves to the Riemannian case.

Corollary 4.1. *Let (M, g) be a compact Riemannian surface of convex boundary and such that $\mathrm{Ric}_g \geq kg$ for a positive constant k. We have for any $\varphi \in H_1^2$,*

$$\log\left(\frac{1}{\mathrm{vol}_g(M)}\int_M e^{\varphi}\,dv_g\right) \leq \frac{1}{\mathrm{vol}_g(M)}\left(\frac{1}{4k}\int_M |\nabla\varphi|^2\,dv_g + \int_M \varphi\,dv_g\right).$$

Proof. For any $p \in (2,\infty)$ and $f \in H_1^2(M)$, (3-33) yields

$$\text{(4-10)}\quad \left(\int_M |f|^p\,dv_g\right)^{\frac{2}{p}} \leq \mathrm{vol}_g(M)^{-\frac{p-2}{p}}\left(\frac{1}{\theta(2,p)}\int_M |\nabla f|^2\,dv_g + \int_M |f|^2\,dv_g\right),$$

where $\theta(2,p) \in \{\theta_1(2,p), \theta_2(2,p)\}$ as defined in Theorem 3.6. Proceeding as in [Beckner 1993], if we choose $f = 1 + \varphi/p$, then (4-10) gives after applying the logarithm function to both sides

$$2\log\left(\int_M \left|1+\frac{\varphi}{p}\right|^p dv_g\right) \le 2\log(\mathrm{vol}_g(M)) + p\left\{\log\left(\frac{1}{\theta(2,p)}\frac{1}{p^2}\int_M |\nabla\varphi|^2\,dv_g + \mathrm{vol}_g(M) + \frac{2}{p}\int_M \varphi\,dv_g + \frac{1}{p^2}\int_M \varphi^2\,dv_g\right) - \log(\mathrm{vol}_g(M))\right\}.$$

Therefore when p tends to infinity, one can easily see that $1/\theta(2,p)p^2$ converges to zero, and thus the second term on the right hand side of the above equation converges to

$$\frac{1}{\mathrm{vol}_g(M)}\left(\frac{1}{2k}\int_M |\nabla\varphi|^2\,dv_g + 2\int_M \varphi\,dv_g\right). \qquad \square$$

Remark 4.2. • For any $\varphi \in H_1^2(\mathbb{S}^2)$ (respectively, $\varphi \in H_1^2(\mathbb{S}^2_+)$), Corollary 4.1 gives immediately

$$\log\left(\frac{1}{4\pi}\int_{\mathbb{S}^2} e^{\varphi}\,dv\right) \le \frac{1}{4\pi}\left(\frac{1}{4}\int_{\mathbb{S}^2} |\nabla\varphi|^2\,dv + \int_{\mathbb{S}^2}\varphi\,dv\right) \tag{4-11}$$

respectively,

$$\log\left(\frac{1}{2\pi}\int_{\mathbb{S}^2_+} e^{\varphi}\,dv\right) \le \frac{1}{2\pi}\left(\frac{1}{4}\int_{\mathbb{S}^2_+} |\nabla\varphi|^2\,dv + \int_{\mathbb{S}^2_+}\varphi\,dv\right), \tag{4-12}$$

where dv is the Riemannian measure of the unit 2-dimensional sphere $\mathbb{S}^2$ (respectively, hemisphere $\mathbb{S}^2_+$), see for instance [Chang and Yang 1988; Onofri 1982; Osgood et al. 1988] for different proofs. It is worth noting that our method (which is inspired by that of Beckner [1993] for the sphere) is the simplest among the existing ones concerning surfaces with boundary.

• We also observe that as in the proofs of [Ilias 1983] for surfaces without boundary, one can use the Levy–Gromov isoperimetric inequality and an adapted symmetrization to deduce the inequality of Corollary 4.1 from that of the 2-sphere.

In the last corollary, we give a logarithmic Sobolev inequality on a compact weighted Riemannian manifold (M, g, σ) of arbitrary dimension (for this kind of inequalities one can see for instance [Gross 1975]). In fact:

Corollary 4.3. *Let (M^n, g, σ) be a compact weighted Riemannian manifold of dimension $n \ge 2$ with convex boundary (i.e., $\mathrm{II} \ge 0$). Assume that for some $m \in [n, \infty)$, the m-Bakry–Émery–Ricci curvature satisfies $\mathrm{Ric}_h^m \ge kg$ for a positive constant k.*

Then for any $f \in H_1^2(d\sigma)$, we have:

$$\int_M |f|^2 \log |f|^2\, d\sigma - \int_M |f|^2 \log\left(\frac{\|f\|^2_{L^2(d\sigma)}}{\mathrm{vol}_h(M)}\right) d\sigma \\ \leq \frac{p}{p-2} \|f\|^2_{L^2(d\sigma)} \log\left(\frac{1}{\theta(m,p)} \frac{\int_M |\nabla f|^2\, d\sigma}{\|f\|^2_{L^2(d\sigma)}} + 1\right)$$

for any $p \in (2, 2^*]$ with $2^* = 2m/(m-2)$ if $m \geq 3$ and $p \in (2, \infty)$ if $m = 2$ (i.e., $n = 2$ and h is constant) with $\theta(m, p)$ as defined in Theorem 3.6.

Proof. It is equivalent to prove that

$$\int_M |f|^2 \log |f|^2\, d\sigma \leq \frac{p}{p-2} \log\left(\frac{1}{\theta(m,p)} \int_M |\nabla f|^2\, d\sigma + 1\right) - \log(\mathrm{vol}_h(M)),$$

for any $f \in H_1^2(d\sigma)$ with $\|f\|^2_{L^2(d\sigma)} = 1$. From Theorem 3.6, we have

$$\frac{2}{p} \log\left(\int_M |f|^p\, d\sigma\right) \\ \leq -\frac{p-2}{p} \log(\mathrm{vol}_h(M)) + \log\left(\frac{1}{\theta(m,p)} \int_M |\nabla f|^2\, d\sigma + 1\right). \tag{4-13}$$

Since the logarithmic function is concave, we use Jensen's inequality and the fact that $\int_M |f|^2\, d\sigma = 1$, to obtain

$$\log\left(\int_M |f|^p\, d\sigma\right) = \log\left(\int_M |f|^{p-2} |f|^2\, d\sigma\right) \geq \frac{p-2}{2} \int_M |f|^2 \log |f|^2\, d\sigma.$$

Replacing the above equation in (4-13), we finish the proof of Corollary 4.3. □

Remark 4.4. If we suppose that the weighted measure $d\sigma$ is a probability measure on M, that is $\mathrm{vol}_h(M) = 1$, then one can reformulate Theorem 3.6 as follows:

$$\eta(m,p)\, \frac{F(p) - F(2)}{p-2} \leq \|\nabla f\|^2_{L^2(d\sigma)}, \tag{4-14}$$

where $F(p) = \|f\|^2_{L^p(d\sigma)} = \left(\int_M |f|^p\, d\sigma\right)^{2/p}$ and $\eta(m,p) \in \{\eta_1(m,p), \eta_2(m,p)\}$ with

$$\eta_1(m,p) = \frac{mk}{(m-1)},$$
$$\eta_2(m,p) = \lambda_1^h + \frac{m(m-1)(p-1)}{\big((p-1) + m(m+2)\big)} \left(k - \frac{m-1}{m} \lambda_1^h\right).$$

By taking the limit $p \to 2$ in (4-14), we obtain

$$\eta(m,2)\, F'(2) \leq \int_M |\nabla f|^2\, d\sigma \tag{4-15}$$

but

$$F'(p) = \frac{2}{p} F(p)^{1-p/2} \int_M |f|^p \log |f| \, d\sigma - \frac{2}{p^2} F(p) \log\left(\int_M |f|^p \, d\sigma \right).$$

Substituting in (4-15), we obtain the following analogue for weighted Riemannian manifolds of the logarithmic Sobolev inequality:

$$\tfrac{1}{2}\eta(m,2)\left(\int_M |f|^2 \log |f|^2 \, d\sigma - \int_M |f|^2 \log\left(\int_M |f|^2 d\sigma \right) d\sigma \right) \le \int_M |\nabla f|^2 \, d\sigma.$$

In the case of a compact Riemannian manifold (without boundary), this last inequality was obtained by Fontenas [1997].

Acknowledgments

The authors are very grateful to the referee for his/her careful reading and valuable comments on the first version of this manuscript.

References

[Akutagawa 2001] K. Akutagawa, "Notes on the relative Yamabe invariant", pp. 105–113 in *Differential geometry* (Sakado, 2001), edited by Q.-M. Cheng, Josai Math. Monogr. **3**, Josai Univ., Sakado, 2001. MR Zbl

[Aubin 1982] T. Aubin, *Nonlinear analysis on manifolds: Monge–Ampère equations*, Grundlehren der Math. Wissenschaften **252**, Springer, 1982. MR Zbl

[Bakry and Émery 1985] D. Bakry and M. Émery, "Diffusions hypercontractives", pp. 177–206 in *Séminaire de probabilités, XIX, 1983/84*, Lecture Notes in Math. **1123**, Springer, 1985. MR Zbl

[Bakry and Ledoux 1996] D. Bakry and M. Ledoux, "Sobolev inequalities and Myers's diameter theorem for an abstract Markov generator", *Duke Math. J.* **85**:1 (1996), 253–270. MR Zbl

[Beckner 1993] W. Beckner, "Sharp Sobolev inequalities on the sphere and the Moser–Trudinger inequality", *Ann. of Math.* (2) **138**:1 (1993), 213–242. MR Zbl

[Bidaut-Véron and Véron 1991] M.-F. Bidaut-Véron and L. Véron, "Nonlinear elliptic equations on compact Riemannian manifolds and asymptotics of Emden equations", *Invent. Math.* **106**:3 (1991), 489–539. MR Zbl

[Case 2015] J. S. Case, "A Yamabe-type problem on smooth metric measure spaces", *J. Differential Geom.* **101**:3 (2015), 467–505. MR Zbl

[Case et al. 2011] J. S. Case, Y.-J. Shu, and G. Wei, "Rigidity of quasi-Einstein metrics", *Differential Geom. Appl.* **29**:1 (2011), 93–100. MR Zbl

[Chang and Yang 1988] S.-Y. A. Chang and P. C. Yang, "Conformal deformation of metrics on S^2", *J. Differential Geom.* **27**:2 (1988), 259–296. MR

[Chang et al. 2006] S.-Y. A. Chang, M. J. Gursky, and P. Yang, "Conformal invariants associated to a measure", *Proc. Natl. Acad. Sci. USA* **103**:8 (2006), 2535–2540. MR Zbl

[Chang et al. 2011] S.-Y. A. Chang, M. J. Gursky, and P. Yang, "Conformal invariants associated to a measure: conformally covariant operators", *Pacific J. Math.* **253**:1 (2011), 37–56. MR Zbl

[Cheng and Li 1981] S. Y. Cheng and P. Li, "Heat kernel estimates and lower bound of eigenvalues", *Comment. Math. Helv.* **56**:3 (1981), 327–338. MR Zbl

[Cherrier 1984] P. Cherrier, "Problèmes de Neumann non linéaires sur les variétés riemanniennes", *J. Funct. Anal.* **57**:2 (1984), 154–206. MR Zbl

[Escobar 1988] J. F. Escobar, "Sharp constant in a Sobolev trace inequality", *Indiana Univ. Math. J.* **37**:3 (1988), 687–698. MR Zbl

[Escobar 1990] J. F. Escobar, "Uniqueness theorems on conformal deformation of metrics, Sobolev inequalities, and an eigenvalue estimate", *Comm. Pure Appl. Math.* **43**:7 (1990), 857–883. MR Zbl

[Escobar 1992] J. F. Escobar, "The Yamabe problem on manifolds with boundary", *J. Differential Geom.* **35**:1 (1992), 21–84. MR Zbl

[Fontenas 1997] E. Fontenas, "Sur les constantes de Sobolev des variétés riemanniennes compactes et les fonctions extrémales des sphères", *Bull. Sci. Math.* **121**:2 (1997), 71–96. MR Zbl

[Gidas and Spruck 1981] B. Gidas and J. Spruck, "Global and local behavior of positive solutions of nonlinear elliptic equations", *Comm. Pure Appl. Math.* **34**:4 (1981), 525–598. MR Zbl

[Grigor'yan 2009] A. Grigor'yan, *Heat kernel and analysis on manifolds*, AMS/IP Studies in Advanced Mathematics **47**, American Mathematical Society, Providence, RI, 2009. MR Zbl

[Gross 1975] L. Gross, "Logarithmic Sobolev inequalities", *Amer. J. Math.* **97**:4 (1975), 1061–1083. MR Zbl

[Hebey 1996] E. Hebey, *Sobolev spaces on Riemannian manifolds*, Lecture Notes in Mathematics **1635**, Springer, 1996. MR Zbl

[Hebey and Vaugon 1996] E. Hebey and M. Vaugon, "Meilleures constantes dans le théorème d'inclusion de Sobolev", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **13**:1 (1996), 57–93. MR Zbl

[Ilias 1983] S. Ilias, "Constantes explicites pour les inégalités de Sobolev sur les variétés riemanniennes compactes", *Ann. Inst. Fourier* (*Grenoble*) **33**:2 (1983), 151–165. MR Zbl

[Ilias 1996] S. Ilias, "Inégalités de Sobolev et résultats d'isolement pour les applications harmoniques", *J. Funct. Anal.* **139**:1 (1996), 182–195. MR Zbl

[Lee and Parker 1987] J. M. Lee and T. H. Parker, "The Yamabe problem", *Bull. Amer. Math. Soc.* (*N.S.*) **17**:1 (1987), 37–91. MR Zbl

[Li 2005] X.-D. Li, "Liouville theorems for symmetric diffusion operators on complete Riemannian manifolds", *J. Math. Pures Appl.* (9) **84**:10 (2005), 1295–1361. MR Zbl

[Li and Wei 2015] H. Li and Y. Wei, "f-minimal surface and manifold with positive m-Bakry–Émery Ricci curvature", *J. Geom. Anal.* **25**:1 (2015), 421–435. MR Zbl

[Li and Yau 1983] P. Li and S. T. Yau, "On the Schrödinger equation and the eigenvalue problem", *Comm. Math. Phys.* **88**:3 (1983), 309–318. MR Zbl

[Lichnerowicz 1970] A. Lichnerowicz, "Variétés riemanniennes à tenseur C non négatif", *C. R. Acad. Sci. Paris Sér. A-B* **271** (1970), A650–A653. MR Zbl

[Lichnerowicz 1971/72] A. Lichnerowicz, "Variétés kählériennes à première classe de Chern non negative et variétés riemanniennes à courbure de Ricci généralisée non negative", *J. Differential Geometry* **6** (1971/72), 47–94. MR Zbl

[Licois and Véron 1995] J. R. Licois and L. Véron, "Un théorème d'annulation pour des équations elliptiques non linéaires sur des variétés riemanniennes compactes", *C. R. Acad. Sci. Paris Sér. I Math.* **320**:11 (1995), 1337–1342. MR Zbl

[Ma and Du 2010] L. Ma and S.-H. Du, "Extension of Reilly formula with applications to eigenvalue estimates for drifting Laplacians", *C. R. Math. Acad. Sci. Paris* **348**:21-22 (2010), 1203–1206. MR Zbl

[Obata 1962] M. Obata, "Certain conditions for a Riemannian manifold to be isometric with a sphere", *J. Math. Soc. Japan* **14** (1962), 333–340. MR Zbl

[Onofri 1982] E. Onofri, "On the positivity of the effective action in a theory of random surfaces", *Comm. Math. Phys.* **86**:3 (1982), 321–326. MR Zbl

[Osgood et al. 1988] B. Osgood, R. Phillips, and P. Sarnak, "Extremals of determinants of Laplacians", *J. Funct. Anal.* **80**:1 (1988), 148–211. MR Zbl

[Petean 2005] J. Petean, "Ricci curvature and Yamabe constants", preprint, 2005. arXiv

[Profeta 2015] A. Profeta, "The sharp Sobolev inequality on metric measure spaces with lower Ricci curvature bounds", *Potential Anal.* **43**:3 (2015), 513–529. MR Zbl

[Reilly 1977] R. C. Reilly, "Applications of the Hessian operator in a Riemannian manifold", *Indiana Univ. Math. J.* **26**:3 (1977), 459–472. MR Zbl

[Setti 1998] A. G. Setti, "Eigenvalue estimates for the weighted Laplacian on a Riemannian manifold", *Rend. Sem. Mat. Univ. Padova* **100** (1998), 27–55. MR Zbl

[Yano and Obata 1970] K. Yano and M. Obata, "Conformal changes of Riemannian metrics", *J. Differential Geometry* **4** (1970), 53–72. MR Zbl

Received April 17, 2017. Revised November 3, 2017.

SAÏD ILIAS
LABORATOIRE DE MATHÉMATIQUES ET PHYSIQUE THÉORIQUE, UMR CNRS 7350
UNIVERSITÉ FRANÇOIS RABELAIS DE TOURS
PARC DE GRANDMONT
37200 TOURS
FRANCE
said.ilias@lmpt.univ-tours.fr

ABDOLHAKIM SHOUMAN
LABORATOIRE DE MATHÉMATIQUES ET PHYSIQUE THÉORIQUE, UMR CNRS 7350
UNIVERSITÉ FRANÇOIS RABELAIS DE TOURS
PARC DE GRANDMONT
37200 TOURS
FRANCE
abdolhakim.shouman@lmpt.univ-tours.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 294, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.294.453

ON THE EXISTENCE OF CLOSED GEODESICS ON 2-ORBIFOLDS

CHRISTIAN LANGE

We show that on every compact Riemannian 2-orbifold there exist infinitely many closed geodesics of positive length.

1. Introduction

Existence and properties of closed geodesics on Riemannian manifolds have been the subject of intense research since Poincaré's work [1905]. A prominent result in the field is a theorem by Gromoll and Meyer [1969] that guarantees the existence of infinitely many closed geodesics on compact Riemannian manifolds on some cohomological assumption. This assumption is satisfied by a large class of manifolds but not by spheres. Generalizing ideas of Birkhoff [1927], Franks [1992] and Bangert [1993] together proved the existence of infinitely many closed geodesics on every Riemannian 2-sphere.

Guruprasad and Haefliger [2006] generalized the result by Gromoll and Meyer to the setting of Riemannian orbifolds. In this paper we generalize the result by Bangert and Franks in the following way.

Theorem 1. *On every compact Riemannian* 2*-orbifold there exist infinitely many geometrically distinct closed geodesics of positive length.*

For spindle orbifolds (see Figure 1), which also go by the name of football orbifolds, this statement was previously only known to hold in the rotational symmetric case [Borzellino et al. 2007]. In [Borzellino and Lorica 1996] the existence of a closed geodesic in the regular part of any 2-orbifold with isolated singularities is claimed. The alleged geodesic in the regular part is obtained as a limit of locally length-minimizing curves contained in the complement of shrinking δ-neighborhoods of the singular points. Note, however, that there are examples in which such curves have a limit that is not contained in the regular part. For instance, on a plane with two singular points of cone angle π which is otherwise flat, curves

The author is partially supported by the DFG funded project SFB/TRR 191.
MSC2010: 53C22, 57R18.
Keywords: closed geodesics, orbifolds.

of minimal length that enclose both δ-neighborhoods converge to a straight segment between the singular points.

To prove our result we first reduce its statement to the case of simply connected spindle orbifolds (see Section 2). Using the curve-shortening flow we are able to prove the existence of an embedded geodesic in the regular part of any simply connected spindle orbifold and to apply ideas from Bangert's proof and Frank's result in this case. Our proof relies on the observation that embedded loops in the regular part that evolve under the curve-shortening flow either stay in the regular part forever, or collapse into a singular point in finite time (see Section 4). The possibility of a limit curve being not entirely contained in the regular part is then excluded by topological arguments (see Proposition 4.5).

For spindle orbifolds with S^1-symmetry, more is known, namely the existence of infinitely many distinct (even modulo isometries) closed geodesics in the regular part [Borzellino et al. 2007]. For general spindle orbifolds (and also in other cases) this is not known to be true. For some 2-orbifolds, e.g., for a sphere with three singular points of order 2, even the existence of a single closed geodesic in the regular part does not seem to be rigorously proven, yet. In Section 6 we will add some more comments on the following questions.

Question 1. Does there always exist a closed geodesic in the regular part of a 2-orbifold with isolated singularities? When do there exist infinitely many?

Note that there are examples of surfaces with more general conical singularities that do not support infinitely many closed geodesics [Borzellino et al. 2007].

2. Preliminaries

2A. *Orbifolds.* Recall that a *length space* is a metric space in which the distance of any two points can be realized as the infimum of the lengths of all rectifiable paths connecting these points [Burago et al. 2001]. A Riemannian orbifold can be defined as follows [Lange 2018b].

Definition 2.1. An *n-dimensional Riemannian orbifold* $\mathcal{O}$ is a length space such that for each point $x \in \mathcal{O}$, there exists a neighborhood U of x in $\mathcal{O}$, an n-dimensional Riemannian manifold M and a finite group G acting by isometries on M such that U and M/G are isometric.

Behind the above definition lies the fact that an effective isometric action of a finite group on a simply connected Riemannian manifold can be recovered from the corresponding metric quotient. In the case of spheres this is proven in [Swartz 2002]; the general case, see [Lange 2018b, Corollary 3.10], can be deduced from it. In particular, the underlying topological space of a Riemannian orbifold in this sense admits a smooth orbifold structure and a compatible Riemannian structure in the usual sense (see [Borzellino 1992; Bridson and Haefliger 1999; Guruprasad and Haefliger 2006]) that in turn induces the metric. For a point x on a Riemannian orbifold the linearized isotropy group of a preimage of x in a Riemannian manifold chart is uniquely determined up to conjugation. Its conjugacy class in $\mathrm{O}(n)$ is denoted as G_x and is called the *local group* of $\mathcal{O}$ at x. The points with trivial local group form the *regular part* of $\mathcal{O}$, which is a Riemannian manifold. All other points are called *singular*. We will be particularly concerned with 2-orbifolds all of whose singular points are isolated. In this case all local groups are cyclic and we refer to their orders as the *orders of the singular points*.

We are interested in (orbifold) geodesics defined in the following way.

Definition 2.2. An (orbifold) *geodesic* on a Riemannian orbifold is a continuous path that can locally be lifted to a geodesic in a Riemannian manifold chart. A closed (orbifold) geodesic is a continuous loop that is an (orbifold) geodesic on each subinterval.

In the following, by a (closed) geodesic we always mean a (closed) orbifold geodesic. A geodesic that encounters an isolated singularity at an interior point is not locally length-minimizing [Borzellino 1992, Theorem 3]. On a 2-orbifold such a geodesic is either reflected or goes straight through the singular point depending on whether the order of the singular point is even or odd. We say that two geodesics are *geometrically distinct* if their geometric trajectories differ. Given a closed geodesic c, the iterations $c^m(t) := c(mt)$, where $m \in \mathbb{N}$, form a whole *tower* of geometrically equivalent closed geodesics. In the following, by infinitely many closed geodesics we always mean infinitely many geometrically distinct closed geodesics of positive length.

We need the following concept; see [Lange 2018b] for a metric definition.

Definition 2.3. A *covering orbifold* or *orbi-cover* of a Riemannian orbifold $\mathcal{O}$ is a Riemannian orbifold $\mathcal{O}'$ together with a surjective map $\varphi : \mathcal{O}' \to \mathcal{O}$ such that each point $x \in \mathcal{O}$ has a neighborhood U isometric to some M/G for which each connected component U_i of $\varphi^{-1}(U)$ is isometric to M/G_i for some subgroup $G_i < G$ such that the isometries respect the projections.

An orbifold is called *simply connected*, if it does not admit a nontrivial orbi-cover. An orbifold is called *good* (or developable) if it is covered by a manifold; otherwise it is called *bad* [Thurston 1980]. The only bad 2-orbifolds are depicted in Figure 1; see [Scott 1983, Theorem 2.3]. In fact, every compact good 2-orbifold is *very good*, meaning that it is finitely covered by a (necessarily compact) manifold [Scott 1983, Theorem 2.5]. Clearly, if an orbifold is finitely covered by an orbifold with infinitely many closed geodesics, then it has itself infinitely many closed geodesics. Since all Riemannian surfaces have infinitely many closed geodesics (see, e.g., [Berger 2010, XII.5] for a survey), in view of proving our main result it suffices to treat

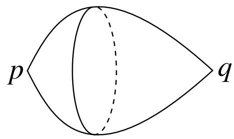


Figure 1. A (p,q)-*spindle* orbifold $S^2(p,q)$, i.e., a 2-orbifold with at most two isolated singularities of order p and q (with p or q perhaps being 1). Spindle orbifolds are also known as *footballs* and $(p,1)$-spindle orbifolds as *teardrops*. The orbifolds in the picture are bad if and only if $p \neq q$ and simply connected as orbifolds if and only if p and q are coprime.

simply connected spindle orbifolds, i.e., spindle orbifolds $S^2(p,q)$ with p and q coprime (see Figure 1).

2B. *Orbifold loop spaces.* We would like to apply Morse theory and homological methods to find closed geodesics on orbifolds. To this end a notion of a loop space is needed. Such a notion is defined in [Guruprasad and Haefliger 2006]. To any compact Riemannian orbifold $\mathcal{O}$ a free loop space $\Lambda\mathcal{O}$ is associated and endowed with a natural structure of a complete Riemannian Hilbert orbifold. We sketch this construction in Appendix A.

Here we give an alternative description of $\Lambda\mathcal{O}$ in the case in which $\mathcal{O}$ has only isolated singularities. So in this section $\mathcal{O}$ will always be a *Riemannian orbifold with only isolated singularities*. Let γ be a loop on $\mathcal{O}$. In the following we always assume that such a loop $\gamma : S^1 \to \mathcal{O}$ is of class H^1, i.e., that it locally lifts to absolutely continuous curves on manifold charts with square-integrable velocities.

Definition 2.4. A *development of* γ is a loop $\hat{\gamma}$ on a Riemannian manifold M together with a map $M \to \mathcal{O}$ which is locally an orbi-covering and which projects $\hat{\gamma}$ to γ (respecting the parametrizations). The development is called *geodesic* if $\hat{\gamma}$ is a geodesic on M.

Every loop on $\mathcal{O}$ can be locally lifted to Riemannian manifold charts. A development of a loop γ on $\mathcal{O}$ can be obtained by gluing together the Riemannian manifold charts that support the local lifts. In particular, this yields a loop after having carried out the identifications. Two developments $(M_1, \hat{\gamma}_1)$ and $(M_2, \hat{\gamma}_2)$ are said to be equivalent if there exist neighborhoods M_1' of $\hat{\gamma}_1$ in M_1 and M_2' of $\hat{\gamma}_2$ in M_2 and an isometry $M_1' \to M_2'$ that maps $\hat{\gamma}_1$ to $\hat{\gamma}_2$ (respecting the parametrizations).

Definition 2.5. An *orbifold loop* is a loop γ on $\mathcal{O}$ together with an equivalence class of developments of γ. The orbifold loop is called *geodesic* if the developments are geodesic.

The notion of a geodesic orbifold loop is equivalent to the notion of a closed orbifold geodesic. Every geodesic orbifold loop projects to a closed orbifold geodesic in the sense of Definition 2.2 and every closed orbifold geodesic gives rise to a unique equivalence class of geodesic developments. However, viewing a closed geodesic as a geodesic orbifold loop shows that it can be assigned local invariants like the index or the nullity as in the manifold case. Moreover, as a set the free loop space $\Lambda\mathcal{O}$ is the collection of all orbifold loops and we can even recover its metric structure using the concept of developments. Indeed, let $\mathcal{D} = (M, \hat{\gamma})$ be a development defining an orbifold loop γ. The free loop space ΛM has a natural structure of a Riemannian Hilbert manifold [Klingenberg 1978] and we have $\hat{\gamma} \in \Lambda M$. A loop $\hat{\gamma}' \in \Lambda M$ can be regarded as an orbifold loop represented by the development $(M, \hat{\gamma}')$. If γ is not a constant loop at a singular point of $\mathcal{O}$, then we can choose a neighborhood $U_{\mathcal{D}}$ of $\hat{\gamma}$ in ΛM such that any pair of distinct loops $\hat{\gamma}', \hat{\gamma}'' \in U_{\mathcal{D}}$ projects to distinct loops on $\mathcal{O}$ and hence corresponds to distinct orbifold loops. The $U_{\mathcal{D}}$ obtained in this way patch together to a Riemannian Hilbert manifold $\Lambda_{\text{reg}}\mathcal{O}$ by identifying elements that correspond to the same orbifold loop. Indeed, if distinct $\hat{\gamma}_1' \in U_{\mathcal{D}_1}$ and $\hat{\gamma}_2' \in U_{\mathcal{D}_2}$ are identified, then, by the definition of the equivalence relation on developments, a whole open neighborhood of $\hat{\gamma}_1'$ in $U_{\mathcal{D}_1}$ is isometrically identified with an open neighborhood of $\hat{\gamma}_2'$ in $U_{\mathcal{D}_2}$. The metric completion of $\Lambda_{\text{reg}}\mathcal{O}$ is the Riemannian Hilbert orbifold $\Lambda\mathcal{O}$ introduced in [Guruprasad and Haefliger 2006]; see Appendix A, and the singular set $\Lambda\mathcal{O} \setminus \Lambda_{\text{reg}}\mathcal{O}$ corresponds to the constant loops at the singular points of $\mathcal{O}$.

Given an atlas of $\mathcal{O}$, the free loop space $\Lambda\mathcal{O}$ can be written as a quotient of a Riemannian Hilbert manifold Ω_X (by a groupoid), where X is the disjoint union of the manifold charts of the atlas, and this description provides local manifold charts for $\Lambda\mathcal{O}$ [Guruprasad and Haefliger 2006]; see Appendix A. On $\Lambda\mathcal{O}$, the energy function E is defined and its critical points correspond to the closed geodesics. Since all singular points of $\Lambda\mathcal{O}$ have zero energy, an explicit knowledge of their structure will not be relevant for our argument (see the Appendix). For some $\kappa > 0$ we write $\Lambda^{\kappa} := \Lambda^{\kappa}\mathcal{O} := \Lambda\mathcal{O} \cap E^{-1}([0, \kappa))$ and for a geodesic loop c with $E(c) = \kappa$ we set $\Lambda(c) := \Lambda\mathcal{O}(c) := \Lambda^{\kappa}\mathcal{O}$. The spaces Ω_X and $\Lambda\mathcal{O}$ admit finite-dimensional approximations much as in the manifold case, see Proposition A.1.

From our description of $\Lambda\mathcal{O}$ it is clear that the index $\operatorname{ind}(c)$ and the nullity $\nu(c)$ of a nontrivial orbifold geodesic can be defined as in the manifold case. Moreover, it shows that the statements used in [Gromoll and Meyer 1969] on the index and the nullity of iterated geodesics and on local loop space homology remain valid in the following form since their proofs involve only local arguments. Note that there is a natural S^1-action on $\Lambda\mathcal{O}$ given by reparametrization.

Lemma 2.6. *For a Riemannian orbifold with isolated singularities and a nontrivial orbifold geodesic c on it, the following statements hold true*:

(i) *Either* $\operatorname{ind}(c^m) = 0$ *for all m or* $\operatorname{ind}(c^m)$ *grows linearly in m* [Gromoll and Meyer 1969, Lemma 1].

(ii) *There are positive integers* $k_1, \dots, k_s$ *and a sequence* $m^i_j \in \mathbb{N}$, *with* $i > 0$ *and* $j = 1, \dots, s$, *such that the numbers* $m^i_j k_j$ *are mutually distinct*, $m^1_j = 1$, $\{m^i_j k_j\} = \mathbb{N}$, *and* $\nu(c^{m^i_j k_j}) = \nu(c^{k_j})$ [Gromoll and Meyer 1969, Lemma 2].

(iii) *There exists some* k *such that* $H_p(\Lambda(c^m) \cup S^1 c^m, \Lambda(c^m)) = 0$ *except possibly for* $\operatorname{ind}(c^m) \le p \le \operatorname{ind}(c^m) + k$ [Gromoll and Meyer 1969, Corollary 1].

3. Homology generated by iterated geodesics

We will need the following slight generalizations of statements in [Bangert and Klingenberg 1983]. Our proofs are essentially the same as those contained in that paper. For convenience we summarize the arguments in the Appendix. Recall that a geodesic c is called *homologically invisible* if $H_*(\Lambda(c) \cup S^1 c, \Lambda(c)) = 0$.

Theorem 3.1 (cf. [Bangert and Klingenberg 1983, Theorem 3]). *Let* $\mathcal{O}$ *be a compact orbifold with isolated singularities and let* c *be a closed geodesic on* $\mathcal{O}$ *such that* $\operatorname{ind}(c^m) = 0$ *for all* $m \in \mathbb{N}$, *i.e.*, c *does not have conjugate points when defined on* $\mathbb{R}$. *Suppose* c *is neither homologically invisible nor an absolute minimum of* E *in its free homotopy class. Then there exist infinitely many closed geodesics on* $\mathcal{O}$.

Note that if $\mathcal{O}$ is simply connected, then a nontrivial geodesic c is never an absolute minimum in its free homotopy class.

Lemma 3.2 (cf. [Bangert and Klingenberg 1983, Lemma 2]). *Let* $\mathcal{O}$ *be a compact orbifold with isolated singularities and let* $\{S^1 c_i \mid i \in \mathbb{N}\}$ *be a sequence of pairwise disjoint critical orbits such that the* c_i *are not absolute minima of* E *in their free homotopy classes. Suppose there exists* $p \in \mathbb{N}$ *such that* $H_p(\Lambda(c_i) \cup S^1 c_i, \Lambda(c_i))) \neq 0$ *for all* $i \in \mathbb{N}$. *Then there exist infinitely many closed geodesics on* $\mathcal{O}$.

Note that the geodesics c_i in the lemma do not need to be geometrically distinct.

4. Existence of simple closed geodesics

The *curve-shortening flow* can be used to prove the existence of a simple closed geodesic on any Riemannian 2-sphere [Grayson 1989]. In this section we discuss properties of the curve-shortening flow on Riemannian 2-orbifolds that allow us to prove the existence of a *separating geodesic* on every simply connected Riemannian spindle orbifold. Here a loop embedded in the regular part of a spindle orbifold is called *separating* if each connected component of its complement contains at most one singular point. Let us first recall some well-known properties of the curve-shortening flow (see [Grayson 1989; Chou and Zhu 2001; Huisken and Polden 1999; Colding et al. 2015]). For a smoothly embedded curve $\gamma = \Gamma_0 : S^1 \to M$ in a closed Riemannian surface there exists a unique maximal smooth curve-shortening flow $\Gamma_t : S^1 \to M$ for $t \in [0, T)$, $T > 0$, satisfying $\partial\Gamma_t/\partial t = kN$ where k is the curvature of Γ_t and N is its normal vector, which moreover depends continuously on the initial condition γ. This flow can be considered as the negative gradient flow of the length functional. An important feature of the curve-shortening flow is that it is a *geometric flow* meaning that the evolution of the geometric image of γ does not depend on the initial parametrization. In the situation above, the curve Γ_t is embedded for each $t \in [0, T)$. Moreover, if T is finite, then Γ_t converges to a point. If T is infinite, then the curvature of Γ_t converges to zero in the C^∞-norm and a subsequence of Γ_t converges to a closed embedded geodesic on M. In particular, T is finite if the length of γ is sufficiently small [Grayson 1989, Lemma 7.1]. If M is not complete but the curvature is still bounded there is a single other alternative for finite T, namely that points on γ do not have a limit on M for $t \to T$. In particular, we see that the curve-shortening flow of an embedded curve in the regular part of a compact 2-orbifold is (a priori) defined until the flow hits a singular point or collapses to a point. In fact, if the 2-orbifold has only isolated singularities more is true as will be discussed below. One possibility to analyze the local behavior of a curve-shortening flow $\Gamma_t : S^1 \to M$ on a manifold at a space-time point (x, T), $T < \infty$, is to blow up the flow at (x, T) by a sequence of parabolic rescalings

$$((M, g), t) \to ((M, \lambda_1 g), \lambda_i^2(t - T) + T), \quad \lambda_i \to \infty,$$

where g denotes the Riemannian metric. Such a blow-up sequence subconverges to a self-shrinking tangent flow of an embedded curve on $T_x M$ (see [Colding et al. 2015]) which, according to a result of Abresch and Langer [1986], is either a self-shrinking circle or a static straight line through the origin. In the first case Γ_t converges to x as a "round point" and in the second case Γ_t is regular at (x, T). Another feature of the curve-shortening flow is that it satisfies the so-called *avoidance principle*, meaning that two initially disjoint curves remain disjoint under the flow. This is a consequence of the maximum principle. Moreover, by the strong maximum principle for parabolic PDEs [Evans 1998, Theorem 7.1.12], it is impossible for a closed curve to be disjoint from a (possibly noncompact) geodesic for $t < t_0$ and to touch it tangentially at $t = t_0$.

The following statement is proven in [Grayson 1989, Corollary 1.7] and says that the curve-shortening flow cannot spread out arbitrarily in finite time.

Lemma 4.1 [Grayson 1989]. *If $T < \infty$, then for every $\varepsilon > 0$ there exists $t_1 < T$, and an open set U in M such that U contains every $\Gamma_t(S^1)$, $t_1 < t < T$, and U is contained in the ε-neighborhood of each $\Gamma_t(S^1)$, $t_1 < t < T$.*

Now we analyze the evolution under the curve-shortening flow of a separating loop on a simply connected spindle orbifold.

Lemma 4.2. *Let $\mathcal{O} \cong S^2(p, q)$ be a Riemannian spindle orbifold, let γ be a separating loop on $\mathcal{O}$ and let $\Gamma : S^1 \times [0, T) \to \mathcal{O}_{\mathrm{reg}}$ be the evolution of γ under the* *curve-shortening flow in the regular part of $\mathcal{O}$. If Γ hits a singular point x of order $p > 2$ in finite time T, then the flow converges to this point.*

Proof. By Lemma 4.1 we can assume that there exists a point $y \neq x$ on $\mathcal{O}$ which is avoided by Γ such that each Γ_t separates x and y. If $\mathcal{O}$ has two singular points, we choose y to be a singular point. The open subset $\mathcal{O} \setminus \{y\}$ of $\mathcal{O}$ admits a p-fold manifold covering M with a cyclic group of deck transformations G of order p that acts by rotations around the preimage $\hat{x}$ of x in M. The preimages $\hat{\Gamma}_t$ of $\Gamma_t(S^1)$ in M are embedded G-invariant loops (or loops after choosing a parametrization, which does not make a difference for us since we are dealing with a geometric flow) and are solutions of the curve-shortening flow. By parabolically blowing up this flow at $(\hat{x}, T)$ as discussed above, we obtain a G-invariant tangent flow. Since $p = |G| > 2$, this tangent flow must be a circle and so $\hat{\Gamma}_t$ converges to $\hat{x}$, as, by transversality, we would otherwise obtain a contradiction to the embeddedness of the circles $\hat{\Gamma}_t$ for $t < T$. In particular, it follows that Γ_t converges to x in the limit $t \to T$. □

Lemma 4.3. *Let $\mathcal{O} \cong S^2(p, q)$ be a simply connected Riemannian spindle orbifold, let γ be a separating loop on $\mathcal{O}$ and let $\Gamma : S^1 \times [0, T) \to \mathcal{O}_{\mathrm{reg}}$ be the maximal evolution of γ under the curve-shortening flow in the regular part of $\mathcal{O}$. Then the curve γ either*

(i) *shrinks to a round point in the regular part of $\mathcal{O}$ in finite time T, or*

(ii) *collapses into a singular point of $\mathcal{O}$ in finite time T, or*

(iii) *$T = \infty$ and γ stays in the regular part of $\mathcal{O}$ forever.*

Note that the first case can only occur if $\mathcal{O}$ has at most one singular point.

Proof. We only have to exclude the case that Γ hits a singular point x of $\mathcal{O}$ in finite time $T < \infty$ without collapsing into this point. Suppose this were the case. By Lemma 4.2 we can assume that the order p of x is even and that the flow does not approach a singular point of odd order at time T. Since p and q are coprime by assumption, the order q must be odd. We can assume that there exists a point $y \neq x$ on $\mathcal{O}$ such that the flow avoids a whole neighborhood of y, such that each Γ_t separates x and y, and such that $\mathcal{O} \setminus \{x, y\}$ lies in the regular part of $\mathcal{O}$. In fact, for $q > 1$ we can choose y to be the singular point of odd order q, and otherwise we can apply Lemma 4.1. As in the proof of Lemma 4.2 the open subset $\mathcal{O} \setminus \{y\}$ of $\mathcal{O}$ admits a p-fold manifold covering M with a cyclic group of deck transformations G of order p that acts by rotations around the preimage $\hat{x}$ of x in M. Also, the preimages $\hat{\Gamma}_t$ of $\Gamma_t(S^1)$ in M are embedded G-invariant circles (or loops after choosing a parametrization) and are solutions of the curve-shortening flow. Since Γ avoids a neighborhood of y and since $\hat{\Gamma}_t$ does not converge to a point at time T, the flow $\hat{\Gamma}_t$ can be extended to a flow $\hat{\Gamma} : S^1 \times [0, T') \to M$ with $T < T'$. Now the fact that Γ hits x at time T implies by G-equivariance that the extended flow $\hat{\Gamma}$ develops a self-crossing at time T. This contradicts embeddedness and hence the claim follows. □

Using compactness and the fact that the curvature converges to zero in infinite time, the following lemma can be proven as in the manifold case [Grayson 1989, Section 7].

Lemma 4.4. *In the situation of Lemma 4.3*(iii), *the loop subconverges to a nontrivial* (*orbifold*) *geodesic.*

Note that the limit geodesic obtained in Lemma 4.4 is a priori not necessarily contained in the regular part of $\mathcal{O}$.

Proposition 4.5. *On every simply connected Riemannian spindle orbifold $\mathcal{O} \cong S^2(p, q)$ there exists a separating geodesic. In particular, there exists a closed geodesic in the regular part of any spindle orbifold.*

Remark 4.6. The first statement is optimal in the sense that there exist Riemannian metrics on $S^2(p, q)$ all of whose geodesics are closed but with only one embedded geodesic [Lange 2018a].

Proof. Choose distinct $x, y \in \mathcal{O}$ such that $\mathcal{O} \setminus \{x, y\}$ is contained in the regular part of $\mathcal{O}$. We smoothly foliate $\mathcal{O} \setminus \{x, y\}$ by circles separating x and y. Then, under the curve-shortening flow, small circles near x flow to a point in finite time, and so do small circles near y [Grayson 1989, Lemma 7.1]. However, the orientations of the limiting points will be opposite in both cases. Hence, by continuous dependence of the flow on the initial conditions there must be some circle γ in the middle that does not flow to a point in finite time, but instead stays in the regular part of $\mathcal{O}$ forever by Lemma 4.2. By Lemma 4.4 this circle subconverges to some nontrivial orbifold geodesic c.

It remains to show that c is contained in the regular part. In fact, in this case c is embedded and separating as a limit of embedded and separating loops. Let γ_i be a subsequence of the curve-shortening flow Γ_t that converges to c. Suppose that c hits a singular point, say x of order p. We can assume that the γ_i and c avoid a neighborhood of a point y' on $\mathcal{O}$. The open subset $\mathcal{O} \setminus \{y'\}$ of $\mathcal{O}$ admits a p-fold orbi-cover $\hat{\mathcal{O}}$ with a cyclic group of deck-transformations G of order p that acts by rotations around the preimage $\hat{x}$ of x in $\hat{\mathcal{O}}$. Let $s_i \in S^1$ be a sequence such that $\gamma_i(s_i)$ converges to some $c(s) \neq x,\ s \in S^1$. The restrictions of γ_i to $S^1 \setminus \{s_i\}$ and of c to $S^1 \setminus \{s\}$ can be lifted to embedded curves $\hat{\gamma}_i^j : S^1 \setminus \{s_i\} \to \hat{\mathcal{O}}_{\mathrm{reg}},\ j = 1, \ldots, p$, and to geodesics $\hat{c}^j : S^1 \setminus \{s\} \to \hat{\mathcal{O}},\ j = 1, \ldots, p$, that are permuted by the deck transformation group G. We can choose the $s_i,\ s$ and the j-numbering in such a way that $\hat{\gamma}_i^j$ converges to $\hat{c}^j$. Note that for fixed i the $\hat{\gamma}_i^j$ have disjoint images in $\hat{\mathcal{O}}$ since γ_i is embedded in $\mathcal{O}$. For $p > 2$ this yields a contradiction since the geodesics

$\hat{c}^j$ intersect transversally at $\hat{x}$ in this case. Suppose that $p = 2$. In this case $\mathcal{O}$ is rotated by π around $\hat{x}$ by the deck transformation group and the orbifold geodesic c is reflected at the singular point x on $\mathcal{O}$. The only way for c to reverse its direction is by being reflected at a singular point of even order. Hence, by periodicity it has to be reflected at singular points of even order twice during a single period. Since p and q are coprime by assumption, this second reflection also has to occur at x. Moreover, this second reflection has to occur from a different direction, because otherwise there would have to be an additional encounter with the singularity of even order in between. Therefore, the $\hat{c}_j$ have transverse self-intersections at $\hat{x}$. Now, in this case the transversality argument from the case $p \geq 3$ yields a contradiction since the $\hat{\gamma}_i^j$ are embedded and hence the claim follows. □

5. Proof of the main result

As seen in Section 2A, it is sufficient to prove our main result for simply connected spindle orbifolds. Given the results from the preceding section, in this case a proof can be given similarly as in [Bangert 1993] in the case of a 2-sphere:

5A. ***Outline of the proof.*** Let $\mathcal{O} \cong S^2(p, q)$ be a simply connected Riemannian spindle orbifold. By Proposition 4.5 there exists a separating geodesic,

$$c : S^1 = \mathbb{R}/\mathbb{Z} \to \mathcal{O}.$$

Suppose the following two conditions are satisfied:

(i) For every geodesic $d : [0, \infty) \to \mathcal{O}$ with initial point $d(0)$ on $c(S^1)$ there exists $t > 0$ with $d(t) \in c(S^1)$.

(ii) When we consider c as defined on $\mathbb{R}$ there exists a pair of conjugate points of c.

Note that the second statement is equivalent to the condition that for every $t_0 \in \mathbb{R}$ there exists $t_1 > t_0$ such that $c(t_0)$ and $c(t_1)$ are conjugate points of c [Lytchak 2009, Corollary 1.3]. In this case Birkhoff's annulus map B_c can be defined on the closed annulus $S^1 \times [0, \pi]$ as follows (see [Bangert 1993; Franks 1992; Birkhoff 1927, VI.10]). For $t \in S^1$ and $\alpha \in (0, \pi)$, consider a geodesic γ starting at $c(t)$ in a direction that forms an angle of α with $\dot{c}(t)$. Condition (i) guarantees the existence of some time t_2 at which γ returns to an encounter with c for the second time, say at $c(t')$. Then $\dot{\gamma}(t_2)$ and $c(t')$ enclose some angle $\alpha' \in (0, \pi)$ and one sets $B_c(t, \alpha) = (t', \alpha')$. Moreover, second conjugate points can be used to extend the map B_c to all of $S^1 \times [0, \pi]$ in a continuous way. The map B_c is isotopic to the identity and preserves a canonical area measure related to the Liouville measure on the unit tangent bundle of S^2 which is invariant under the geodesic flow. Moreover, the restrictions of B_c to the two boundary components are inverse to each other.

In this case the work of Franks [1992] implies that Birkhoff's annulus map has infinitely many periodic points (see [Franks 1992, Theorem 4.1]) and these points correspond to closed geodesics on $\mathcal{O}$.

In the following we show that the existence of infinitely many geodesics can still be shown if Birkhoff's annulus map cannot be defined.

5B. ***Simple closed geodesics without conjugate points.***

Proposition 5.1. *Suppose on a Riemannian spindle orbifold $\mathcal{O} \cong S^2(p,q)$ there exists a separating geodesic c without conjugate points. Then there exist infinitely many closed geodesics.*

Proof. The proof works similarly to the proof of [Bangert 1993, Theorem 1]. We choose Riemannian manifold charts $X_0 = \mathcal{O}_{\text{reg}},\ X_x,\ X_y$ of $\mathcal{O}$ where $\mathcal{O}\backslash\{x, y\} \subset \mathcal{O}_{\text{reg}}$ and where X_x and X_y cover $\mathcal{O} \setminus \{y\}$ and $\mathcal{O} \setminus \{x\}$, respectively. Moreover, we choose finite-dimensional approximations (see Proposition A.1) $\Omega = \Omega_X^{\kappa}(k)$ and $P = \Lambda\mathcal{O}^{\kappa}(k)$ of Ω_X^{κ} and $\Lambda\mathcal{O}^{\kappa}$ containing S^1c. Since c is separating, there exists a tubular neighborhood V of $c(S^1)$ in the regular part of $\mathcal{O}$ which is homeomorphic to an annulus. For every $\varepsilon > 0$ we let $U(V, \varepsilon)$ denote the set of $\gamma \in P$ which have energy $E(\gamma) < E(c) + \varepsilon$ and whose projections to $\mathcal{O}$ lie in V and are freely homotopic to c in V. Since we are looking for infinitely many closed geodesics we may assume that the only closed geodesics freely homotopic to c in V are those in S^1c. Moreover, we can choose V so small that every $\gamma \in U(V, \varepsilon)$ with $E(\gamma) < E(c)$ is disjoint from $c(S^1)$. This follows from the assumption that c does not have conjugate points and the Gauss lemma, see the proof of [Bangert 1993, Theorem 1]. If we choose arbitrarily small V and ε, the sets $U(V, \varepsilon)$ form a fundamental system of neighborhoods of S^1c in P. Therefore, we have that either (see [Bangert 1993, Theorem 1])

(i) c is a local minimum of E, or

(ii) c can be approximated by curves $\gamma \in P$ with $E(\gamma) < E(c)$ from both sides, or

(iii) c can be approximated by curves $\gamma \in P$ with $E(\gamma) < E(c)$ precisely from one side.

In the second case it follows as in [Bangert 1993, Theorem 1] that

$$H_1(\Lambda(c) \cup S^1c, \Lambda(c)) \neq 0$$

and this implies the existence of infinitely many closed geodesics by Theorem 3.1.

In the first or third cases, let D be a disk bounded by $c(S^1)$ such that c cannot be approximated by closed curves in D with $E(\gamma) < E(c)$ and suppose that $x \in D$ has order p. The disk D is p-foldly covered by a disk $\hat{D}$ in X_x. A parametrization $\hat{c}$ of the boundary of $\hat{D}$ by arclength is a geodesic that covers c a total of p times. For some sufficiently large κ_m we can choose finite-dimensional approximations

Ω_m of $\Omega_X^{\kappa_m}$ containing a homotopy in $\hat{D}$ from $\hat{c}^m$ to a point curve. As above we choose an annulus $V \supset c(S^1)$ so small that $S^1 c$ are the only closed geodesics freely homotopic to c in V. Moreover, we may assume that every $\gamma \in U(V, \varepsilon)$ which lies in the component V_0 of $V \setminus c(S^1)$ contained in D has energy $E(\gamma) \geq E(c)$. Hence we have $E(\gamma) > E(c)$ for all $\gamma \in U(V, \varepsilon) \setminus S^1 c$ which are contained in D, because otherwise such a γ with $E(\gamma) = E(c)$ would be a closed geodesic freely homotopic to c. The analogous statement also holds for c^m [Bangert 1993, page 5]. In particular, the analogous statement also holds for $\hat{c}^m \in \Omega_m$. In this very situation, min-max methods applied to homotopies in $\hat{D}$ from $\hat{c}^m$ to a point curve are used in the proof of [Bangert 1993, Theorem 1] to show the existence of a sequence of closed geodesics $\hat{d}_m$ in $\hat{D}$ such that $E(\hat{d}_m)$ tends to infinity and such that the local groups $H_1(\Omega_X(\hat{d}_m) \cup S^1 c, \Omega_X(\hat{d}_m))$ do not vanish. The gradient of the energy functional restricted to the finite-dimensional approximation is used to deform the homotopies, and so the fact that $\hat{D}$ is bounded by a geodesic guarantees that the construction remains in $\hat{D}$. The resulting geodesics project to (orbifold) geodesics d_m in D with $E(d_m)$ tending to infinity and with

$$H_1(\Lambda(d_m) \cup S^1 c, \Lambda(d_m)) = H_1(\Omega_X(\hat{d}_m) \cup S^1 c, \Omega_X(\hat{d}_m)) \neq 0.$$

Therefore there exist infinitely many closed geodesics on $\mathcal{O}$ by Lemma 3.2. □

5C. ***The non-Birkhoff case.*** In this section we study the case of a Riemannian spindle orbifold with a separating geodesic for which the corresponding Birkhoff map is not defined. We reduce this case to Proposition 5.1. Finally we summarize why this implies our main result.

The following lemma is a special case of [Bangert 1993, Lemma 2].

Lemma 5.2. *Suppose c is a separating geodesic with conjugate points on a Riemannian spindle orbifold. Then c can be approximated from either side by closed curves γ which are disjoint from c and satisfy $E(\gamma) < E(c)$. In particular, c can be approximated from either side by shorter, disjoint curves.*

Now we can prove the following proposition.

Proposition 5.3. *Suppose c is a separating geodesic with conjugate points on a simply connected Riemannian spindle orbifold $\mathcal{O} \cong S^2(p, q)$ and $d : (0, \infty) \to \mathcal{O}$ is a geodesic disjoint from c. Then there exist infinitely many closed geodesics.*

Proof. The proof is similar to the proof of [Bangert 1993, Theorem 2]. However, we use the curve-shortening flow instead of Birkhoff's curve shortening process and simplify the second part of the argument.

Let D be the component of $\mathcal{O} \setminus c(S^1)$ that contains $d(\mathbb{R})$. Since c is separating by assumption, there exists an open neighborhood V of the closure of D in $\mathcal{O}$ that admits a Riemannian manifold chart $\hat{V}$. The geodesic d lifts to a geodesic $\hat{d}$

on $\hat{V}$ and the geodesic c is covered p times by a geodesic $\hat{c}$ disjoint from $\hat{d}$. In [Bangert 1993, Theorem 2] it is proven that the closure of every limit geodesic $\bar{d}$ of $\hat{d}$, that is, every geodesic of the form $\bar{d} : \mathbb{R} \to \hat{V}$, $\bar{d}(t) = \exp_p(tv)$ where (p, v) is an accumulation point of $(\hat{d}, \hat{d}')$ in $T\hat{V}$, is disjoint from $\hat{c}(S^1)$. In particular, the closure of the image $\tilde{d}$ in V of such a limit geodesic of $\hat{d}$ is disjoint from $c(S^1)$. Let U_1 be the component of $D \setminus \mathrm{closure}(\tilde{d}(\mathbb{R}))$ that contains $c(S^1)$ in its closure. Since the closure of $\tilde{d}(\mathbb{R})$ is disjoint from $c(S^1)$, by Lemma 5.2 there is an embedded loop γ_1 in U_1 that is freely homotopic in the regular part of U_1 to c and shorter than c. We claim that the evolution Γ_t of γ_1 under the curve-shortening flow does not leave U_1. Otherwise there would exist some t_1 minimal with the property that $\Gamma_{t_1}(S^1)$ is not contained in U_1. Let $x \in \Gamma_{t_1}(S^1) \cap U_1^C$. By the avoidance principle applied to c and γ_1 the only possibility could be that x is contained in the closure of $\tilde{d}(\mathbb{R})$. Since γ_1 is noncontractible in U_1 and since $\tilde{d}(\mathbb{R})$ is not a point, the point x is regular by Lemma 4.2 and so is the loop Γ_{t_1} by Lemma 4.3. Let $d_0 : \mathbb{R} \to \mathcal{O}$ be the geodesic which is tangent to Γ_{t_1} at x. By minimality of t_1 the geodesic d_0 is contained in the closure of $\tilde{d}(\mathbb{R})$. In particular, the flow of γ_1 and the (static) flow of d_0 touch at (x, t_1) for the first time. This is impossible by the maximum principle (see Section 4) and hence the evolution of γ_1 stays in U_1 as claimed. Moreover, since γ_1 is noncontractible in U_1 by assumption, it evolves in the regular part forever, i.e., we are in case (iii) of Lemma 4.2. By Lemma 4.4 the flow subconverges to a simple closed geodesic $\tilde{d}_1$ contained in the closure of U_1. The proof of Proposition 4.5 shows that this limit geodesic actually lies in the regular part of $\mathcal{O}$. It is distinct from c since γ_1 is shorter than c and the curve-shortening flow does not increase the arclength. By the choice of γ_1 the geodesic $\tilde{d}_1$ is separating and so we are done in this case by Proposition 5.1 if $\tilde{d}_1$ does not have conjugate points. Otherwise we define U_2 to be the component of $D \setminus \tilde{d}_1(S^1)$ whose closure contains c. This component is bounded by two geodesics with conjugate points and contained in the regular part of $\mathcal{O}$ by the construction of $\tilde{d}_1$. Moreover, it contains a noncontractible embedded loop γ_2 which is shorter than $\tilde{d}_1$ by Lemma 5.2 and hence also shorter than c. Again, letting γ_2 evolve under the curve-shortening flow, the same argument as above yields a separating limit geodesic $\tilde{d}_2$ in $\overline{U}_2$ which is now distinct from both c and $\tilde{d}_1$ and hence contained in U_2. This process can be iterated. It either yields infinitely many (simple) closed geodesics on $\mathcal{O}$ with conjugate points or terminates at a separating geodesic without conjugate points which in turn implies the existence of infinitely many closed geodesics by Proposition 5.1. □

Together with Section 5A, we see Propositions 5.3 and 5.1 imply the following:

Proposition 5.4. *Let c be a separating geodesic on a Riemannian spindle orbifold $\mathcal{O} \cong S^2(p, q)$ for which Birkhoff's annulus map B_c is not defined. Then there exist infinitely many closed geodesics.*

Recall from Section 5A that Frank's work implies the existence of infinitely many closed geodesics in the case in which Birkhoff's annulus map B_c can be defined. Therefore, in any case there are infinitely many closed geodesics on a Riemannian spindle orbifold. By the remark at the end of Section 2A our main result, Theorem 1, follows.

6. Closed geodesics in the regular part

In this section we sketch some ideas and make some speculations on the question, posed in the introduction, of when there exists one, or even infinitely many closed geodesics in the regular part of a Riemannian 2-orbifold $\mathcal{O}$ with isolated singularities.

Let us begin with a general remark. In Section 4 we have shown that an embedded loop in the regular part of a simply connected spindle orbifold cannot flow into a singular point in finite time under the curve-shortening flow unless it collapses into this point entirely. The simply-connectedness assumption was used to handle the case of singular points of order 2 (see proof of Lemma 4.3). Using a local noncollapsing result of Brian White this *avoidance of singularities principle* can actually be shown to be true for embedded loops in the regular part of any Riemannian 2-orbifold. More precisely, if such a loop were to flow into a singular point in finite time, then we could locally lift the flow to a manifold chart and look at a tangent flow of a blow-up limit at the singular space-time point in question as in the proof of Lemma 4.2. Recall from that argument that such a tangent flow can only either be a self-shrinking circle or a static line, and that we used equivariance with respect to the deck transformation group in case of a singular point of order at least 3 in order to exclude the latter case. In case of a singular point of order 2 the blow-up sequence could in principle converge to a line, but this line would have to have multiplicity 2, i.e., two strands of the lifted flow that are permuted by the deck transformation group would converge to it in the blow-up limit. However, this possibility is ruled out by the noncollapsing result of White [2000, Theorem 9.1] (see also [White 2015, Section 7] for more details).

The discussed argument not only works for embedded loops, but also for loops that stay δ-embedded in the regular part under the curve-shortening flow for some $\delta > 0$ as long as it remains in the regular part. Here a loop is called *δ-embedded* if its restriction to each subinterval of length δ with respect to arclength parametrization is embedded. If the 2-orbifold $\mathcal{O}$ has at least 4 singular points, then one can find infinitely many loops in the regular part with this property that are pairwise homotopically different in the regular part. This is because the property of having a "minimal number of transverse self-intersections" in one's homotopy class is preserved under the curve-shortening flow. Each such loop subconverges to a closed (orbifold) geodesic on $\mathcal{O}$ and one is left to decide whether the limits lie in the regular part. If all singular points have orders at least 3, this is the case by the same argument as in the proof of Proposition 4.5, and so there exist infinitely many distinct closed geodesics in the regular part in this case. We believe that the same conclusion can be drawn in the presence of singular points of order 2 by arguments similar to the one in the proof of Proposition 4.5. For instance, if a limit geodesic hit a singular point of order 2, then, by the above arguments, it would have to oscillate between two singular points of order 2 and avoid all other singularities. By choosing the initial loops in a clever way, this limit behavior could possibly be ruled out in advance.

In the case that $\mathcal{O}$ has 3 singular points one should still be able to infer the existence of infinitely many closed geodesics if the orders of the singular points are sufficiently large by observing that the avoidance of singularities principle discussed above still holds for loops that "wind around a singular point of order $2n$ up to $n-1$ times". Otherwise, one might be able to use the existence of a finite manifold cover in this case to find at least one closed geodesic in the regular part.

In the case of (simply connected) spindle orbifolds the above arguments break down and we do not know how to find infinitely many closed geodesics in the regular part. There might also exist metrics without this property.

Appendix A: Orbifold loop spaces

We summarize the description of an orbifold loop space from [Guruprasad and Haefliger 2006]. For a Riemannian orbifold $\mathcal{O}$ let X be the disjoint union of Riemannian manifold charts covering $\mathcal{O}$ and let $\mathcal{G}$ be the small category with the set of objects X and arrows being the germs of change of charts of X. Then $(\mathcal{G}, X)$ with the usual topology of germs on $\mathcal{G}$ is an étale groupoid and $\mathcal{O}$ can be represented as a quotient $\mathcal{O} = \mathcal{G} \setminus X$ [Guruprasad and Haefliger 2006, Section 2.1.4]. A *$\mathcal{G}$-loop based at $x \in X$* over a subdivision $0 = t_0 < t_1 \cdots < t_k = 1$ of the interval $[0, 1]$ is a sequence $c = (g_0, c_1, g_1, \dots, c_k, g_k)$ where:

(i) $c_i : [t_{i-1}, t_i] \to X$ is of class H^1, i.e., c_i is absolutely continuous and the velocity functions $t \mapsto |\dot{c}_i(t)|$ are square integrable.

(ii) g_i is an element of $\mathcal{G}$ such that $\alpha(g_i) = c_{i+1}(t_i)$ for $i = 0, 1, \dots, k-1$, $\omega(g_i) = c_i(t_i)$ for $i = 1, \dots, k$ and $\omega(g_0) = \alpha(g_k) = x$. Here $\alpha(g)$ and $\omega(g)$ denote the source and the target of $g \in \mathcal{G}$; see [Guruprasad and Haefliger 2006, Section 2.1.4].

A $\mathcal{G}$-loop is called *geodesic* if all the c_i are geodesics and their velocities match up at the break points t_i via the g_i. A geodesic $\mathcal{G}$-loop gives rise to a closed geodesic on $\mathcal{O}$ in the sense of Definition 2.2. The space Ω_x is defined as the set of equivalence classes of $\mathcal{G}$-loops based at x under the equivalence relation generated by the following operations [Guruprasad and Haefliger 2006, Sections 2.3.2 and 3.3.2]:

(i) Given a $\mathcal{G}$-loop $c = (g_0, c_1, g_1, \dots, c_k, g_k)$ over the subdivision
$$0 = t_0 < \cdots < t_k = 1,$$
we can add a subdivision point $t' \in (t_{i-1}, t_i)$ together with the unit element $g' = 1_{c_i(t')}$ to get a new sequence, replacing c_i in c by $c_i',\ g', c_i''$, where c_i' and c_i'' are the restrictions of c_i to the intervals $[t_{i-1}, t']$ and $[t', t_i]$ and where 1_y, $y \in X$ is the germ of the identity map at y.

(ii) Replace a $\mathcal{G}$-loop c by a new loop $c' = (g_0', c_1', g_1', \dots, c_k', g_k')$ over the same subdivision as follows: for each $i = 1, \dots, k$ choose H^1-maps
$$h_i : [t_{i-1}, t_i] \to \mathcal{G}$$
such that $\alpha(h_i(t)) = c_i(t)$, and define $c_i' : t \mapsto \omega(h_i(t)),\ g_i' = h_i(t_i) g_i h_{i+1}(t_i)^{-1}$ for $i = 1, \dots, k-1,\ g_0' = g_0 h_1(0)^{-1}$ and $g_k' = h_k(1) g_k$.

The space Ω_X is defined to be $\Omega_X = \bigcup_{x \in X} \Omega_x$. It admits a natural structure of a Riemannian Hilbert manifold [Guruprasad and Haefliger 2006, Section 3.3.2]. A $\mathcal{G}$-loop c gives rise to a development $(M, \hat{\gamma})$ in the sense of Definition 2.4. A neighborhood of the equivalence class of c in Ω_X is isometric to a neighborhood of $\hat{\gamma}$ in ΛM. If $\mathcal{O}$ is compact, then Ω_X is complete with respect to the induced metric (see the proof of [Klingenberg 1978, Theorem 1.4.5]). An energy function can be defined on Ω_X whose critical points correspond to geodesic $\mathcal{G}$-loops [Guruprasad and Haefliger 2006, Section 3.4.1]. The groupoid $\mathcal{G}$ acts isometrically on the left on Ω_X. If $[c] \in \Omega_X$ is represented by the $\mathcal{G}$-loop $c = (g_0, c_1, g_1, \dots, c_k, g_k)$ based at x and g is an element of $\mathcal{G}$ with $\alpha(g) = x$ and $\omega(g) = y$, then $g[c]$ is represented by
$$gc := (gg_0, c_1, g_1, \dots, c_k, g_k g^{-1})$$
[Guruprasad and Haefliger 2006, Section 2.3.3]. The quotient $|\Lambda\mathcal{O}| = \mathcal{G} \setminus \Omega_X$ is called the *free loop space* of $\mathcal{O}$. The quotient map $\Omega_X \to \mathcal{G} \setminus \Omega_X$ induces a natural structure of a Riemannian Hilbert orbifold on $|\Lambda\mathcal{O}|$ denoted as $\Lambda\mathcal{O}$ [Guruprasad and Haefliger 2006, Section 3.3.4]. Since Ω_X is complete as a metric space, so is $\Lambda\mathcal{O}$. The space $\Lambda(\mathcal{O}_{\mathrm{reg}})$ is naturally a subset of $\Lambda\mathcal{O}$ and coincides with the ordinary loop space of $\mathcal{O}_{\mathrm{reg}}$ as a Riemannian manifold; see [Klingenberg 1978]. If $\mathcal{O}$ has only isolated singularities, then the only elements of Ω_X with nontrivial $\mathcal{G}$-isotropy are the loops that project to a singular point of $\mathcal{O}$. In this case $\mathcal{G}$-equivalence classes of $\mathcal{G}$-loops correspond to equivalence classes of developments discussed in Section 2B and the regular part of $\Lambda\mathcal{O}$ is isometric to the Riemannian Hilbert manifold $\Lambda_{\mathrm{reg}}\mathcal{O}$ constructed in Section 2B. In particular, $\Lambda\mathcal{O}$ is the metric completion of $\Lambda_{\mathrm{reg}}\mathcal{O}$, as $\Lambda_{\mathrm{reg}}\mathcal{O}$ is dense in the complete metric space $\Lambda\mathcal{O}$.

As in the manifold case, the spaces
$$\Omega_X^{\kappa} = \Omega_X \cap E^{-1}([0, \kappa)) \quad \text{and} \quad \Lambda\mathcal{O}^{\kappa} = \Lambda\mathcal{O}^{\kappa} \cap E^{-1}([0, \kappa))$$
admit finite-dimensional approximations [Guruprasad and Haefliger 2006, v.1].

Proposition A.1 (finite-dimensional approximation). *Let $\mathcal{O}$ be a compact orbifold and let $\kappa \geq 0$ be given.*

(i) *There exist $\varepsilon > 0$ and a big enough k such that every element of Ω_X^κ can be represented by a $\mathcal{G}$-path $c = (g_0, c_1, g_1, \dots, c_k, g_k)$ defined over the subdivision $0 = t_0 < t_1 < \cdots < t_k = 1$, where $t_i = i/k$, and each $c_i(t_{i-1})$ is the center of a convex open geodesic ball of radius ε containing the image of c_i.*

(ii) *The space Ω_X^κ retracts by energy-nonincreasing deformation onto the subspace $\Omega_X^\kappa(k)$ whose elements are represented by $\mathcal{G}$-loops as above for which each c_i is a geodesic segment.*

(iii) *The restrictions of the energy function to Ω_X^κ and to $\Omega_X^\kappa(k)$ have the same critical points, and at each such point the nullity and the index* (*which can still be defined in this case* [Guruprasad and Haefliger 2006]) *are the same.*

(iv) *The orbifold $\Lambda\mathcal{O}^\kappa := \mathcal{G} \setminus \Omega_X^\kappa$ retracts by energy-nonincreasing deformation onto the finite-dimensional orbifold $\Lambda\mathcal{O}^\kappa(k) = \mathcal{G} \setminus \Omega_X^\kappa(k)$.*

For a Riemannian spindle orbifold $\mathcal{O} = S^2(p, q)$ with a separating geodesic c we can choose the following convenient charts. Let $x, y \in \mathcal{O}$ be two points in different components of the complement of c such that $\mathcal{O} \setminus \{x, y\}$ lies in the regular part of $\mathcal{O}$. We choose X to be the disjoint union of $X_0 = \mathcal{O}_{\text{reg}}$ and Riemannian manifold charts X_x and X_y covering $\mathcal{O} \setminus \{y\}$ p-foldly and $\mathcal{O} \setminus \{x\}$ q-foldly, respectively. With respect to this choice any $\mathcal{G}$-loop based at $z \in X_0$ that projects to $\mathcal{O}_{\text{reg}}$ can be represented as $(1_z, c, 1_z)$ for an H^1-loop in X_0 with $c(0) = z = c(1)$ and can thus be identified with c. Any $\mathcal{G}$-equivalence class of $\mathcal{G}$-loops based at $z \in X_x$ which is represented by a $\mathcal{G}$-loop that projects to $\mathcal{O} \setminus \{y\}$ can be represented as $(1_z, c, g_z)$ for an H^1-loop in X_0 with $c(0) = z$ and $c(1) = gz$ and a deck transformation $g \in G_x$ of the covering $X_x \to \mathcal{O} \setminus \{y\}$.

Appendix B: Homology generated by iterated geodesics

In this section we explain why the proofs of [Bangert and Klingenberg 1983, Theorem 3; Bangert and Klingenberg 1983, Lemma 2] also work in the slightly more general situation of Theorem 3.1 and Lemma 3.2, which allow for isolated orbifold singularities. The first step is to obtain [Bangert and Klingenberg 1983, Theorem 1] in the following version. Here the map $\psi^m : \Lambda \to \Lambda$ sends a loop γ to its m-fold iteration γ^m.

Theorem B.1. *Let Λ be the free loop space of a compact Riemannian orbifold $\mathcal{O}$ with isolated singularities. For some $\kappa > 0$ let $[g]$ be a class in $\pi_p(|\Lambda|, |\Lambda^\kappa|)$. Then $[g^m] := [\psi^m \circ g] \in \pi_p(|\Lambda|, |\Lambda^{\kappa m^2}|)$ is trivial for almost all $m \in \mathbb{N}$.*

In the manifold case the idea of the proof is the following. Given a homotopy $h : [a, b] \to \Lambda$, then $h^m := \psi^m \circ h$ is a naturally associated homotopy from $h^m(a)$ to

$h^m(b)$. One can replace this homotopy, which pulls the m loops of $h^m(a)$ to $h^m(b)$ as a whole, by another homotopy $h_m : [a, b] \to \Lambda$ from $h_m(a) = h^m(a)$ to $h_m(b) = h^m(b)$ that pulls the m loops from $h^m(a)$ to $h^m(b)$ successively. The advantage of the new homotopy over the old one is that the energy of $h_m(t)$ depends only on $h(a)$ and $h(b)$ in the limit of large m and can thus be bounded appropriately. This construction can be applied fiberwise to a map $g : (I^p, \partial I^p) \to (|\Lambda|, |\Lambda^\kappa|)$ with respect to a splitting $I^p = I^{p-1} \times I$ and yields a homotopic map with image in Λ^κ. The same construction can be carried out in the case of an orbifold with isolated singularities. In fact, all regularity issues occurring in [Bangert and Klingenberg 1983] can be handled in the same way in this case since finite-dimensional approximations are also available for orbifold loop spaces (see Proposition A.1) and since the whole construction can be assumed to take place away from the energy zero level set, and hence in the manifold part of Λ.

The next step is to obtain a related result in homology as in [Bangert and Klingenberg 1983, Theorem 2].

Theorem B.2. *Let Λ be the free loop space of a compact Riemannian orbifold $\mathcal{O}$ with isolated singularities. Let H be an element of $H_*(\Lambda_\kappa, \Lambda^\kappa)$, where $\kappa > 0$ and Λ_κ is the union of all components of Λ intersecting Λ^κ. Let K be a finite set of integers $k \geq 2$. Then there exists $m \in \mathbb{N}$ such that no $k \in K$ divides m and such that $\psi_*^m(H)$ vanishes in $H_*(\Lambda_\kappa, \Lambda^{\kappa m^2})$.*

The proof of [Bangert and Klingenberg 1983, Theorem 2] can be taken verbatim as a proof for Theorem B.2. For sufficiently large m one can construct a homotopy from a representative of $\psi_*^m(H)$ to a representative in $\Lambda^{\kappa m^2}$ by using Theorem B.1 inductively.

Now we explain the proof of [Bangert and Klingenberg 1983, Theorem 3] and why it generalizes to the setting of Theorem 3.1. Suppose that there exist only finitely many towers of closed geodesics on $\mathcal{O}$. Then all critical S^1-orbits on Λ are isolated. Moreover, by Lemma 2.6(iii), and perhaps after choosing a different c, one can find $p \in \mathbb{N}$ such that $H_p(\Lambda(c) \cup S^1 c, \Lambda(c)) \neq 0$ and $H_q(\Lambda(d) \cup S^1 d, \Lambda(d)) = 0$ for every $q > p$ and every closed geodesic d with $\operatorname{ind}(d^m) = 0$ for all m. In the manifold case Lemma 2.6(ii) implies that there exist integers $\{k_1, \dots, k_s\}$, $k_i \geq 2$, such that

$$\psi_*^m : H_p(\Lambda(c) \cup S^1 c, \Lambda(c)) \to H_p(\Lambda(c^m) \cup S^1 c^m, \Lambda(c^m))$$

is an isomorphism whenever none of the k_i divides m (for details see the proof of [Bangert and Klingenberg 1983, Theorem 2]). The same conclusion holds in the present situation since its proof involves only local arguments. By Lemma 2.6(i) and the assumption of only finitely many towers of closed geodesics, there exists some $A > 0$ such that every closed geodesic d with $E(d) > A$ either satisfies $\operatorname{ind}(d) > p+1$ or $\operatorname{ind}(d^m) = 0$ for all m. Hence one has $H_{p+1}(\Lambda(d) \cup S^1 d, \Lambda(d)) = 0$ whenever

d is a closed geodesic with $E(d) > A$. Therefore, as in [Bangert and Klingenberg 1983], standard arguments from Morse theory imply that

$$i_* : H_p(\Lambda(c^m) \cup S^1 c^m, \Lambda(c^m)) \to H_p(\Lambda, \Lambda(c^m))$$

is one-to-one if $E(c^m) > A$. In particular, the composition

$$i_* \circ \psi_*^m : H_p(\Lambda(c) \cup S^1 c, \Lambda(c)) \to H_p(\Lambda, \Lambda(c^m))$$

is one-to-one. Since c is not an absolute minimum in its free homotopy class, this contradicts Theorem B.2.

Hence Theorem 3.1 holds. The proof of [Bangert and Klingenberg 1983, Lemma 2] and Lemma 3.2 works as follows. $H_p(\Lambda(c_i) \cup S^1 c_i, \Lambda(c_i)) \neq 0$ implies $\mathrm{ind}(c_i) \leq p$ by Lemma 2.6(iii). Since for every closed geodesic c either $\mathrm{ind}(c^m)$ grows linearly with m or $\mathrm{ind}(c^m) = 0$ for all m, the c_i can be iterates of a finite number of prime closed geodesics only if $\mathrm{ind}(c_i^m) = 0$ for some i and all m. In this case Theorem 3.1 proves the existence of infinitely many closed geodesics.

Acknowledgements

I would like to thank Nick Edelen and Alexander Lytchak for discussions and comments on a draft version of this paper. I also thank Victor Bangert for his remarks, and Stephan Wiesendorf for literature suggestions. I thank the referee for useful comments that helped to improve the exposition.

A partial support by the DFG funded project SFB/TRR 191 “Symplectic Structures in Geometry, Algebra and Dynamics” is gratefully acknowledged.

References

[Abresch and Langer 1986] U. Abresch and J. Langer, “The normalized curve shortening flow and homothetic solutions”, *J. Differential Geom.* **23**:2 (1986), 175–196. MR Zbl

[Bangert 1993] V. Bangert, “On the existence of closed geodesics on two-spheres”, *Internat. J. Math.* **4**:1 (1993), 1–10. MR Zbl

[Bangert and Klingenberg 1983] V. Bangert and W. Klingenberg, “Homology generated by iterated closed geodesics”, *Topology* **22**:4 (1983), 379–388. MR Zbl

[Berger 2010] M. Berger, *Geometry revealed*, Springer, 2010. MR Zbl

[Birkhoff 1927] G. D. Birkhoff, *Dynamical systems*, Colloquium Publ. **9**, American Mathematical Society, Providence, RI, 1927. Zbl

[Borzellino 1992] J. E. Borzellino, *Riemannian geometry of orbifolds*, Ph.D. thesis, University of California, Los Angeles, 1992, available at https://tinyurl.com/borzellinothes.

[Borzellino and Lorica 1996] J. E. Borzellino and B. G. Lorica, “The closed geodesic problem for compact Riemannian 2-orbifolds”, *Pacific J. Math.* **175**:1 (1996), 39–46. MR Zbl

[Borzellino et al. 2007] J. E. Borzellino, C. R. Jordan-Squire, G. C. Petrics, and D. M. Sullivan, “Closed geodesics on orbifolds of revolution”, *Houston J. Math.* **33**:4 (2007), 1011–1025. MR Zbl

[Bridson and Haefliger 1999] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Math. Wissenschaften **319**, Springer, 1999. MR Zbl

[Burago et al. 2001] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics **33**, American Mathematical Society, Providence, RI, 2001. MR Zbl

[Chou and Zhu 2001] K.-S. Chou and X.-P. Zhu, *The curve shortening problem*, Chapman Hall/CRC, Boca Raton, FL, 2001. MR Zbl

[Colding et al. 2015] T. H. Colding, W. P. Minicozzi, II, and E. K. Pedersen, "Mean curvature flow", *Bull. Amer. Math. Soc.* (*N.S.*) **52**:2 (2015), 297–333. MR Zbl

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

[Franks 1992] J. Franks, "Geodesics on S^2 and periodic points of annulus homeomorphisms", *Invent. Math.* **108**:2 (1992), 403–418. MR Zbl

[Grayson 1989] M. A. Grayson, "Shortening embedded curves", *Ann. of Math.* (2) **129**:1 (1989), 71–111. MR Zbl

[Gromoll and Meyer 1969] D. Gromoll and W. Meyer, "Periodic geodesics on compact Riemannian manifolds", *J. Differential Geometry* **3** (1969), 493–510. MR Zbl

[Guruprasad and Haefliger 2006] K. Guruprasad and A. Haefliger, "Closed geodesics on orbifolds", *Topology* **45**:3 (2006), 611–641. MR Zbl

[Huisken and Polden 1999] G. Huisken and A. Polden, "Geometric evolution equations for hypersurfaces", pp. 45–84 in *Calculus of variations and geometric evolution problems* (Cetraro, 1996), edited by S. Hildebrandt and M. Struwe, Lecture Notes in Math. **1713**, Springer, 1999. MR Zbl

[Klingenberg 1978] W. Klingenberg, *Lectures on closed geodesics*, Grundlehren der Math. Wissenschaften **230**, Springer, 1978. MR Zbl

[Lange 2018a] C. Lange, "On metrics on 2-orbifolds all of whose geodesics are closed", *J. Reine Angew. Math* (online publication January 2018).

[Lange 2018b] C. Lange, "Orbifolds from a metric viewpoint", preprint, 2018. arXiv

[Lytchak 2009] A. Lytchak, "Notes on the Jacobi equation", *Differential Geom. Appl.* **27**:2 (2009), 329–334. MR Zbl

[Poincaré 1905] H. Poincaré, "Sur les lignes géodésiques des surfaces convexes", *Trans. Amer. Math. Soc.* **6**:3 (1905), 237–274. MR JFM

[Scott 1983] P. Scott, "The geometries of 3-manifolds", *Bull. London Math. Soc.* **15**:5 (1983), 401–487. MR Zbl

[Swartz 2002] E. Swartz, "Matroids and quotients of spheres", *Math. Z.* **241**:2 (2002), 247–269. MR Zbl

[Thurston 1980] W. P. Thurston, "The geometry and topology of three-manifolds", lecture notes, Princeton University, 1980, available at http://library.msri.org/books/gt3m/.

[White 2000] B. White, "The size of the singular set in mean curvature flow of mean-convex sets", *J. Amer. Math. Soc.* **13**:3 (2000), 665–695. MR Zbl

[White 2015] B. White, "Mean curvature flow", Lecture notes, Stanford University, 2015, available at https://tinyurl.com/whitecurva.

Received March 20, 2017. Revised October 10, 2017.

Christian Lange
Mathematisches Institut der Universität zu Köln
Weyertal
Köln
Germany
clange@math.uni-koeln.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 294, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.294.473

A CASSELMAN–SHALIKA FORMULA FOR THE GENERALIZED SHALIKA MODEL OF SO_{4n}

MIYU SUZUKI

We compute the explicit formula (sometimes called the Casselman–Shalika formula) of the generalized Shalika model for unramified principal series of p-adic SO_{4n}. The method mainly used is the Casselman–Shalika method, modified by Y. Hironaka and applied by Y. Sakellaridis to the case of the Shalika model of GL_{2n}.

1. Introduction

Let $G = \mathrm{SO}_{4n}(F)$, the F-split $4n$-dimensional special orthogonal group, where F is a nonarchimedean local field of characteristic 0.

By P, we denote the Siegel parabolic subgroup of G and by N, the unipotent radical of P. Once we identify G with a subgroup of the isotropy group of the quadratic form defined by

$$\xi = \begin{pmatrix} & \mathbb{1}_{2n} \\ \mathbb{1}_{2n} & \end{pmatrix},$$

N is identified with the subgroup consisting of matrices of the form

$$\begin{pmatrix} \mathbb{1}_{2n} & X \\ & \mathbb{1}_{2n} \end{pmatrix} \quad \text{with } X + {}^tX = 0_{2n}.$$

Let M be the Levi component of P consisting of matrices of the form

$$\begin{pmatrix} A & \\ & {}^tA^{-1} \end{pmatrix} \quad \text{with } A \in \mathrm{GL}_{2n}(F).$$

Jiang and Qin [2007] introduced the notion of a generalized Shalika model for representations of G as follows. Take any nontrivial additive character ψ of F with conductor 0. The expression

$$\psi\left(\tfrac{1}{2}\,\mathrm{tr}(JX)\right)$$

defines a character Ψ on N, where

$$J = \begin{pmatrix} & \mathbb{1}_n \\ -\mathbb{1}_n & \end{pmatrix}.$$

MSC2010: primary 11F70, 22E50; secondary 11F85.
Keywords: Casselman–Shalika formula, Shalika models.

The stabilizer of this character in M is naturally isomorphic to $\mathrm{Sp}_{2n}(F)$, the symplectic group with respect to J,

$$\mathrm{Sp}_{2n}(F) = \{x \in \mathrm{GL}_{2n}(F) \mid {}^t x J x = J\}.$$

Define the subgroup (called the "generalized Shalika subgroup") H of P by

$$H := \mathrm{Stab}_M(\Psi)N \cong \mathrm{Sp}_{2n}(F) \ltimes N$$

and extend Ψ to a character of H, which will be again denoted by Ψ.

An admissible representation π of G is said to have a generalized Shalika model if there is a nonzero G-morphism from π to $\mathrm{Ind}_H^G(\Psi)$. Because of Frobenius reciprocity, this is equivalent to saying that there is a nonzero H-morphism from π to Ψ.

In this article, we will treat the case of unramified principal series $I(\chi)$ of G and determine a necessary and sufficient condition for $I(\chi)$ to have a generalized Shalika model. Moreover, we will give an explicit formula (a Casselman–Shalika formula) for the spherical vector in the generalized Shalika model of $I(\chi)$.

We will explain our results more precisely. Take any nonzero H-morphism Λ from $I(\chi)$ to Ψ. Let $K = \mathrm{SO}_{4n}(\mathfrak{o})$, the standard maximal compact subgroup of G, where $\mathfrak{o}$ is the ring of integers of F. There is a unique K-invariant vector ϕ_K in $I(\chi)$ which satisfies $\phi_K(1) = 1$. Let $\Omega(g) = \Lambda(R_g \phi_K)$. Our goal is to give an explicit formula for this function Ω.

The Weyl group of G is denoted by W. The main result involves the subgroup Γ of W. Let $\Sigma = \{e_i \pm e_j,\ 1 \le i,\ j \le 2n,\ i \ne j\}$ be the root system of G and $E_i = e_{2i-1} + e_{2i}$. Then, $\Phi = \{E_i - E_j,\ \pm E_k,\ 1 \le i,\ j,\ k \le n,\ i \ne j\}$ is a root system of type B_n and Γ is the Weyl group of Φ realized by the subgroup of W. For each root $\alpha \in \Sigma$, Casselman defined a certain constant $c_\alpha(\chi)$ (see [Casselman 1980, Section 3]). If $\beta \in \Phi$ is a short root, then β is in Σ and a_β is already defined. In this case, let $d_\beta(\chi) = \chi(a_\beta)$. If $\beta = E_i - E_j$ is a long root of Φ, define $a_\beta = a_{e_{2i-1} - e_{2j-1}}$. In this case, let

$$d_\beta(\chi) = \chi(a_\beta) \frac{1 - q^{-2}\chi(a_{-\beta})}{1 - q^{-2}\chi(a_\beta)}.$$

Our main result is as follows.

Theorem 1.1. *For every $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{Z}^n$ with $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge 0$,*

$$\Omega(g_\lambda) = \prod_{\alpha > 0} c_\alpha(\chi) \sum_{w \in \Gamma} (-1)^{l_\Gamma(w)} (w\chi)^{-1} \delta^{1/2}(h_\lambda) \prod_{\beta > 0,\ w\beta < 0} d_\beta(\chi),$$

where $g_\lambda = \mathrm{diag}(\varpi^{\lambda_1}, \dots, \varpi^{\lambda_n}, 1, \dots, 1)$, $h_\lambda = \mathrm{diag}(\varpi^{\lambda_1}, 1, \varpi^{\lambda_2}, 1, \dots, \varpi^{\lambda_n}, 1) \in M$ and l_Γ is the length function of Γ.

Note that Ω satisfies $\Omega(hgk) = \Psi(h)\Omega(g)$ for every $h \in H,\ k \in K,\ g \in G$ and hence we only need to compute the value of Ω for representatives $\{g_\lambda\}$ of $H \backslash G / K$.

The method we will use is based on works of Casselman and Shalika (see [Casselman 1980; 1980]) and the outline of this paper is essentially the same as that of [Sakellaridis 2006], where an explicit formula for the Shalika model is given.

2. Preliminaries

Notation. Let F be a nonarchimedean local field of characteristic 0. Let ϖ be a uniformizer, q the order of the residue field, $\mathfrak{o}$ the ring of integers, and $\mathfrak{p}$ the maximal ideal of F.

Let $G = \mathrm{SO}_{4n}(F)$, the F-split $4n$-dimensional special orthogonal group. The group G is identified with the subgroup of $\mathrm{SL}_{4n}(F)$ consisting of matrices satisfying

$$ {}^t g \xi g = \xi, \quad \xi = \begin{pmatrix} & \mathbb{1}_{2n} \\ \mathbb{1}_{2n} & \end{pmatrix}. $$

Denote by $\mathrm{Mat}_{2n}(F)$ the set of matrices of degree $2n$.

By P, we denote the Siegel parabolic subgroup of G, consisting of matrices of the form

$$ \begin{pmatrix} x & \\ & {}^t x^{-1} \end{pmatrix} \begin{pmatrix} \mathbb{1}_{2n} & X \\ & \mathbb{1}_{2n} \end{pmatrix} \tag{2-1} $$

with $x \in \mathrm{GL}_{2n}(F)$ and $X \in \mathrm{Mat}_{2n}(F)$, $X + {}^t X = 0$. Let N be the unipotent radical of P and M the Levi component with Levi decomposition $P = MN$ as (2-1). We will frequently identify M with $\mathrm{GL}_{2n}(F)$ without notice.

The Bruhat–Tits building of G is denoted by $\mathcal{B}(G)$. Each maximal F-split torus defines an apartment of $\mathcal{B}(G)$. We denote the split maximal torus consisting of diagonal matrices by T and corresponding apartment by $\mathcal{A}(T)$. Fix a special point $o \in \mathcal{A}(T)$ and identify $\mathcal{A}(T)$ with $2n$-dimensional Euclid space with origin o.

Let Σ be the set of roots of G with respect to T. By taking differentials, we identify elements of Σ with linear functions on $\mathfrak{t}$, the Lie algebra of T. We will naturally identify $\mathfrak{t}$ with an F-linear space of diagonal matrices:

$$ \mathfrak{t} = \{\mathrm{diag}(t_1, \ldots, t_{2n}, -t_1, \ldots, -t_{2n}) \mid t_1, \ldots, t_{2n} \in F\}. $$

For $1 \le i \le 2n$, the element e_i of the dual space of $\mathfrak{t}$ is defined by

$$ e_i : \mathrm{diag}(t_1, \ldots, t_{2n}, -t_1, \ldots, -t_{2n}) \mapsto t_i. $$

Then, under identifications mentioned above, $\Sigma = \{e_i \pm e_j \mid 1 \le i \ne j \le 2n\}$. Let $\Pi = \{\alpha_i := e_i - e_{i+1} \mid 1 \le i \le 2n-1,\ \alpha_{2n} := e_{2n-1} + e_{2n}\}$; this is a basis of the root system Σ. Elements of Σ are regarded as linear functions on $\mathcal{A}(T)$ and the set Σ_{aff} of affine roots of G as a subset of affine functions on $\mathcal{A}(T)$:

$$ \Sigma_{\mathrm{aff}} = \{\alpha + m \mid \alpha \in \Sigma, m \in \mathbb{Z}\}. $$

Let $C = \{x \in \mathcal{A}(T) \mid 0 < \alpha(x) < 1 \text{ for all } \alpha \in \Pi\}$ be an alcove of $\mathcal{B}(G)$. Let B be the Iwahori subgroup of G stabilizing C.

We denote the Weyl group of G by W. By $s_i \in W$, we denote the simple reflection attached to the simple root α_i.

Generalized Shalika model. Following [Jiang and Qin 2007], we define the generalized Shalika model for representations of G as follows. Let $\mathcal{A}$ be the set of nonsingular skew-symmetric matrices of degree $2n$. Take any nontrivial additive character ψ of F with conductor 0 and a skew-symmetric matrix $b \in \mathcal{A}$. The expression

$$\psi\left(\tfrac{1}{2}\operatorname{tr}(bX)\right)$$

defines a character Ψ^b on N. The stabilizer of this character in M is naturally isomorphic to $\mathrm{Sp}_{2n}^b(F)$, the symplectic group with respect to b,

$$\mathrm{Sp}_{2n}^b(F) = \{x \in \mathrm{GL}_{2n}(F) \mid {}^t x b x = b\}.$$

Form a group

$$H^b := \mathrm{Sp}_{2n}^b(F) \ltimes N$$

and extend Ψ^b to a character of H^b, which is again denoted by Ψ^b.

Let $J = \begin{pmatrix} & \mathbb{1}_n \\ -\mathbb{1}_n & \end{pmatrix} \in \mathcal{A}$. We will simply denote Ψ^b, H^b and $\mathrm{Sp}_{2n}^b(F)$ by Ψ (or sometimes by Ψ_H), H and $\mathrm{Sp}_{2n}(F)$ when $b = J$.

Definition. Let (π, V) be an irreducible admissible representation of G. We say that π has a generalized Shalika model if $\mathrm{Hom}_{H^b}(\pi, \Psi^b)$ is nonzero for some $b \in \mathcal{A}$.

Nien proved the uniqueness of generalized Shalika models:

Theorem 2.1 [Nien 2010]. *For any irreducible admissible representation π of $\mathrm{SO}_{4n}(F)$ and $b \in \mathcal{A}$,*

$$\dim \mathrm{Hom}_{H^b}(\pi, \Psi^b) \leq 1.$$

We will consider the generalized Shalika model for unramified principal series of G. The Borel subgroup of G consisting of matrices in the form of (2-1) with upper triangular $x \in \mathrm{GL}_{2n}(F)$ will be denoted by P_ϕ. Let $\chi = (|\cdot|^{z_1}, |\cdot|^{z_2}|, \dots, |\cdot|^{z_{2n}})$ be an unramified character of P_ϕ (i.e., $\chi : \mathrm{diag}(t_1, \dots, t_{2n}, t_1^{-1}, \dots, t_{2n}^{-1}) \mapsto |t_1|^{z_1} \cdots |t_{2n}|^{z_{2n}}$) and $I(\chi)$ the smooth unramified principal series of G. The representation space of $I(\chi)$ is realized by the space of locally constant functions on G which satisfy

$$f(pg) = \chi\delta^{1/2}(p) f(g)$$

for every $p \in P_\phi,\ g \in G$, where $\delta = (|\cdot|^{4n-2}, |\cdot|^{4n-4}, \dots, |\cdot|^2, |\cdot|^0)$ is the modular character of P_ϕ. Then G acts on this space by right translations R. There is a surjective map $\mathcal{P}_\chi$ to this space from $C_c^\infty(G)$ defined by

$$\mathcal{P}_\chi(f)(g) = \int_{P_\phi} \chi^{-1}\delta^{1/2}(p) f(pg)\, dp$$

for $f \in C_c^\infty(G)$ and $g \in G$. We will always assume that $P_\phi(\mathfrak{o})$ has total measure 1. Let $K = \mathrm{SO}_{4n}(\mathfrak{o})$ be the standard maximal compact subgroup of G and $\phi_K = \phi_{K,\chi}$ be the unique K-invariant element of $I(\chi)$ satisfying $\phi_K(1) = 1$. It is easy to see that ϕ_K is the image under $\mathcal{P}_\chi$ of the characteristic function of K.

Definition. Take a nontrivial element $\Lambda(= \Lambda_H = \Lambda_{H,\chi}) \in \mathrm{Hom}_H(I(\chi), \Psi))$. We define a generalized Shalika function

$$\Omega(g)(= \Omega_H(g) = \Omega_{H,\chi}(g)) = \Lambda(R_g\phi_K).$$

The aim of this paper is to give an explicit formula of this function.

The main results. We will briefly explain the statement of the main results in this subsection. At first, we need to introduce some more notation.

Since the function Ω satisfies

$$\Omega(hgk) = \Psi(h)\Omega(g)$$

for every $h \in H$, $g \in G$ and $k \in K$, it suffices to compute it for a set of double coset representatives in $H \setminus G/K$. By Iwasawa decomposition,

$$H \setminus G/K = H \setminus PK/K \cong \mathrm{Sp}_{2n}(F) \setminus \mathrm{GL}_{2n}(F)/\,\mathrm{GL}_{2n}(\mathfrak{o}).$$

Considering transitive right action of $\mathrm{GL}_{2n}(F)$ on $\mathcal{A}$ defined by $X * g := {}^t gXg$, we can naturally identify these double cosets with orbits in $\mathcal{A}$ under the action of $\mathrm{GL}_{2n}(\mathfrak{o})$.

Proposition 2.2. *We have the following double coset decomposition*:

$$\mathrm{GL}_{2n}(F) = \bigsqcup_{\lambda} \mathrm{Sp}_{2n}(F) g_\lambda \,\mathrm{GL}_{2n}(\mathfrak{o}),$$

where $g_\lambda := \mathrm{diag}(\varpi^{\lambda_1}, \varpi^{\lambda_2}, \dots \varpi^{\lambda_n}, 1, \dots, 1)$ *with* $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n) \in \mathbb{Z}^n$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.

Proof. The elementary divisor theorem shows that representatives of orbits in $\mathcal{A}$ under the action of $\mathrm{GL}_{2n}(\mathfrak{o})$ can be taken as follows:

$$X_\lambda = \left(\begin{array}{cccc|cccc} & & & & \varpi^{\lambda_1} & & & \\ & & & & & \varpi^{\lambda_2} & & \\ & & & & & & \ddots & \\ & & & & & & & \varpi^{\lambda_n} \\ \hline -\varpi^{\lambda_1} & & & & & & & \\ & -\varpi^{\lambda_2} & & & & & & \\ & & \ddots & & & & & \\ & & & -\varpi^{\lambda_n} & & & & \end{array}\right).$$

Since $X_\lambda = J * g_\lambda$, we obtain the double coset decomposition. $\square$

By an abuse of notation, we will write g_λ as $\mathrm{diag}(g_\lambda^{-1}, g_\lambda) \in G$. Then we only have to compute $\Omega(g_\lambda)$ for each λ.

Lemma 2.3. *If some λ_i is negative, then $\Omega(g_\lambda) = 0$.*

Proof. Assume that $\lambda_n < 0$ and let $X \in \mathrm{Mat}_{2n}(\mathfrak{o})$ be a matrix whose only nonzero entries are $X_{n,2n} = u$ and $X_{2n,n} = -u$, where $u \in \mathfrak{o}^\times$. Then

$$a = \begin{pmatrix} \mathbb{1}_{2n} & X \\ & \mathbb{1}_{2n} \end{pmatrix}$$

is an element of K and we have $\Omega(g_\lambda) = \Omega(g_\lambda a) = \psi^{-1}(u\varpi^{\lambda_n})\Omega(g_\lambda)$. Since the conductor of ψ is 0, we can choose u so that $\psi(u\varpi^{\lambda_n}) \neq 1$. □

Consequently, we only have to treat the case where λ is a dominant partition of some positive integer. Hereafter, we assume that λ denotes these partitions.

For each $w \in W$, there is an intertwining operator $T_w : I(\chi) \to I(w\chi)$ which satisfies the following relations (see [Casselman 1980]):

$$T_w(\phi_{K,\chi}) = c_w(\chi)\phi_{K,w\chi},$$

where

$$c_w(\chi) = \prod_{\alpha>0,\ w\alpha<0} c_\alpha(\chi), \qquad c_\alpha(\chi) = \frac{1 - q^{-1}\chi(a_\alpha)}{1 - \chi(a_\alpha)}.$$

Here α is a root of G and a_α is a diagonal matrix attached to α. For details, see [Casselman 1980]. Taking the adjoint, we get a G-morphism $T_w^* : I(w\chi)^* \to I(\chi)^*$, where * denotes the dual space of a complex linear space.

Denote the space of distributions on G by $\mathfrak{D}(G)$. By $\mathcal{P}_\chi : C_c^\infty(G) \to I(\chi)$, we obtain the adjoint G-morphism $\mathcal{P}_\chi^* : I(\chi)^* \to \mathfrak{D}(G)$. Let $\Delta (= \Delta_H = \Delta_{H,\chi}) := \mathcal{P}_\chi^*(\Lambda) \in \mathfrak{D}(G)$. Based on the work of Sakellaridis [2006] (also see [Casselman 1980] and [Hironaka 1999]), we get

$$\Omega(g) = Q^{-1} \sum_w \Bigl(\prod_{\alpha>0,\ w\alpha>0} c_\alpha(\chi) \Bigr) T_{w^{-1}}^* \Delta(R_g \,\mathrm{ch}_B), \tag{2-2}$$

where ch_B denotes the characteristic function of B, Q the volume of Bw_lB and w_l is the longest element of W. Hence the problem is reduced to computing $T_{w^{-1}}^* \Delta(R_g \,\mathrm{ch}_B)$ for $w \in W$ and $g = g_\lambda$.

The statement of our formula involves the subgroup Γ of W, which is isomorphic to the Weyl group of type B_n, and its root system. Therefore, let us fix some notation.

Let $E_i = e_{2i-1} + e_{2i}$, $\beta_i = E_i - E_{i+1}$ $(1 \le i < n)$ and $\beta_n = E_n$. Then, $\Phi := \{E_i - E_j, \pm E_k \mid 1 \le i, j, k \le n, i \neq j\}$ is a root system of type B_n and $\{\beta_i \mid 1 \le i \le n\}$ is a basis of Φ.

The subgroup Γ is generated by

$$w_i := \begin{array}{c} \begin{array}{c} i \\ \vee \end{array} \\ \begin{pmatrix} \mathbb{1}_2 & & \vdots & & & \\ & \ddots & \vdots & & & \\ & & 0_2 & \mathbb{1}_2 & & \\ & & \mathbb{1}_2 & 0_2 & & \\ & & & & \ddots & \\ & & & & & \mathbb{1}_2 \end{pmatrix} \end{array} \in M, \quad (1 \le i \le n-1)$$

and

$$w_n := \left(\begin{array}{cc|cc} \mathbb{1}_{2(n-1)} & & & \\ & 0_2 & & \varepsilon \\ \hline & & \mathbb{1}_{2(n-1)} & \\ & \varepsilon & & 0_2 \end{array}\right) \in G, \quad \text{where } \varepsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Note that Γ is naturally identified with the Weyl group of the root system Φ and under this identification, w_i is the simple reflection corresponding to β_i.

Definition. For each long root $\beta = E_i - E_j \in \Phi$, let $a_\beta = a_{e_{2i-1}-e_{2j-1}}$. For a short root $\beta \in \Phi$, a_β is already defined since $\beta \in \Sigma$.

We define $d_\beta(\chi)$ for each $\beta \in \Phi$ as follows: if β is a short root,

$$d_\beta(\chi) = \chi(a_\beta)$$

and if β is a long root,

$$d_\beta(\chi) = \chi(a_\beta)\frac{1-q^{-2}\chi(a_{-\beta})}{1-q^{-2}\chi(a_\beta)}.$$

Our main theorem is as follows.

Theorem 2.4. *Let $\chi = (|\cdot|^{z_1}, |\cdot|^{z_2}, \dots, |\cdot|^{z_{2n}})$ be an unramified character on P_ϕ and assume that this character satisfies $z_{2i-1} = 1 + z_{2i}$ for all $a \le i \le n$.*

(i) *If χ is not of the form as above (or its W-translate), then $I(\chi)$ does not have a generalized Shalika model.*

(ii) *For every $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{Z}^n$ with $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge 0$,*

$$\Omega(g_\lambda) = Q^{-1} \prod_{\alpha>0} c_\alpha(\chi) \sum_{w\in\Gamma} (-1)^{l_\Gamma(w)} (w\chi)^{-1} \delta^{1/2}(h_\lambda) \prod_{\substack{\beta>0 \\ w\beta<0}} d_\beta(\chi),$$

where $h_\lambda = \mathrm{diag}(\varpi^{\lambda_1}, 1, \varpi^{\lambda_2}, 1, \dots, \varpi^{\lambda_n}, 1) \in M$ and l_Γ is the length function of Γ.

3. The open coset

In this section, we will determine which double cosets in $P_\phi \backslash G/H$ are open (if they exist). We don't analyze this quotient space directly but consider $P_\phi \backslash G/P$, which is easily described by using Weyl groups. Since the unique open coset in $P_\phi \backslash G/P$ is $P_\phi \xi P$, the open cosets in $P_\phi \backslash G/H$ are in this coset (if they exist). So we will treat the following quotient space: $P_\phi \backslash P_\phi \xi P/H \cong (\xi^{-1} P_\phi \xi \cap P) \backslash P/H \cong P_0 \backslash G_0/H_0$, where $G_0 = \mathrm{GL}_{2n}(F)$, $H_0 = \mathrm{Sp}_{2n}(F)$ and P_0 is the Borel subgroup of G_0 consisting of lower triangular matrices.

The transitive left action of G_0 on $\mathcal{A}$ is defined by $g * X := gX{}^tg$. Then there is a natural surjective map θ from G_0 to $\mathcal{A}$ defined by $\theta(g) = g * J$. For $X \in G_0$ and each $1 \le i \le n$, X_i denotes the top left $2i \times 2i$-block and $d_i(X)$ its Pfaffian. Let $\mathcal{A}' = \{X \in \mathcal{A} \mid d_i(X) \neq 0\,(1 \le i \le n)\}$ be an open set in $\mathcal{A}$. We will show that the inverse image of this set under the map θ is a double coset in $P_0 \backslash G_0/H_0$. Identifying W_0, the Weyl group of G_0, with the symmetric group of degree $2n$, define the element w_0 of W_0 as a permutation such that

$$w_0(i) = \begin{cases} 2i-1 & (1 \le i \le n) \\ 2i-2n & (n+1 \le i \le 2n). \end{cases}$$

Let $\varepsilon = \begin{pmatrix} & 1 \\ -1 & \end{pmatrix}$. Then

$$\theta(w_0) = w_0 * J = \begin{pmatrix} \varepsilon & & & \\ & \varepsilon & & \\ & & \ddots & \\ & & & \varepsilon \end{pmatrix} \in \mathcal{A}'.$$

Lemma 3.1. $\mathcal{A}' = \theta(P_0 w_0 H_0)$. *In particular*, $P_0 w_0 H_0 = \theta^{-1}(\mathcal{A}')$ *is open.*

Proof. Since

$$p * X = \begin{pmatrix} p_i & \\ * & * \end{pmatrix} \begin{pmatrix} X_i & * \\ * & * \end{pmatrix} \begin{pmatrix} {}^tp_i & * \\ & * \end{pmatrix} = \begin{pmatrix} p_i * X_i & * \\ * & * \end{pmatrix},$$

we have $d_i(p * X) = \det(p_i)^2 d_i(X) \neq 0$ and this shows that $\mathcal{A}'$ is P_0-stable. Hence $\theta(P_0 w_0 H_0) \subset \mathcal{A}'$. By induction on i, we have to show that if X_i is of the form

$$\begin{pmatrix} \varepsilon & & \\ & \ddots & \\ & & \varepsilon \end{pmatrix},$$

there is a lower triangular matrix $p \in \mathrm{GL}_{2(i+1)}(F)$ such that

$$p * X_{i+1} = \begin{pmatrix} \varepsilon & & \\ & \ddots & \\ & & \varepsilon \end{pmatrix}.$$

Let A_j be 2×2-matrices and assume that X_{i+1} is expressed as

$$\begin{pmatrix} \varepsilon & & & A_1 \\ & \ddots & & \vdots \\ & & \varepsilon & A_i \\ -{}^tA_1 & \cdots & -{}^tA_i & A_{i+1} \end{pmatrix}.$$

Let $p_1 \in \mathrm{GL}_{2(i+1)}(F)$ denote a lower triangular matrix of the form

$$\begin{pmatrix} \mathbb{1}_2 & & & \\ & \ddots & & \\ & & \mathbb{1}_2 & \\ B_1 & \cdots & B_i & \mathbb{1}_2 \end{pmatrix},$$

where $B_j := {}^tA_j\varepsilon^{-1}$. Then there is a skew-symmetric matrix $C \in \mathrm{GL}_2(F)$ satisfying

$$p_1 * X_{i+1} = \begin{pmatrix} \varepsilon & & & \\ & \ddots & & \\ & & \varepsilon & \\ & & & C \end{pmatrix}.$$

It is clear that there is a diagonal matrix p_2 of degree $2(i+1)$ so that $(p_2p_1) * X_{i+1}$ becomes the desired form. $\square$

Remark. From the proof of Lemma 3.1, we easily obtain the following slight refinement. For any $X \in \mathcal{A}'$, there is a $p \in P_0$ with diagonal component $(c_1, \dots, c_{2n})$ which sends X to $\theta(w_0)$ and satisfies

$$c_{2i} = 1, \quad c_{2i+1} = \frac{d_{i+1}(X)}{d_i(X)}.$$

Let

$$B_0 = \begin{pmatrix} \mathfrak{o}^\times & & \mathfrak{p} \\ & \ddots & \\ \mathfrak{o} & & \mathfrak{o}^\times \end{pmatrix}$$

be the standard Iwahori subgroup corresponding to P_0 and

$$Y_\lambda = \begin{pmatrix} \varpi^{\lambda_1}\varepsilon & & 0 \\ & \ddots & \\ 0 & & \varpi^{\lambda_n}\varepsilon \end{pmatrix}$$

be an element of $\mathcal{A}'$.

Lemma 3.2. *For all $b \in B_0$ and λ, $1 \le i \le n$,*

$$|d_i(b * Y_{-\lambda})| = |d_i(Y_{-\lambda})|.$$

In particular, $\mathcal{A}' = P_0B_0 * Y_{-\lambda} = \theta(P_0B_0w_0H_0)$.

Proof. By Lemma 3.1, $\mathcal{A}' = P_0 * Y_{-\lambda} \subset P_0 B_0 * Y_{-\lambda} = \theta(P_0 B_0 w_0 H_0)$. The other inclusion follows once we prove the first equation. This is clear for elements in $P_0 \cap B_0$. Thus, by Iwahori decomposition, it suffices to prove this equation for elements in

$$N_0 := \begin{pmatrix} 1 & & \mathfrak{p} \\ & \ddots & \\ & & 1 \end{pmatrix}.$$

This will be proved by induction on the size of matrices. Let $n \in N_0$ and $X = n * Y_{-\lambda}$. Then for $1 \le i \le n-1$,

$$X_i \in n_i * (Y_{-\lambda})_i + \varpi^{-\lambda_{i+1}+2} \mathrm{M}_{2i}(\mathfrak{o}).$$

Since $-\lambda_i \le -\lambda_{i+1}$, any component of $n_i * (Y_{-\lambda})_i$ does not lie in $\varpi^{-\lambda_{i+1}+2}\mathfrak{o} = \mathfrak{p}^{-\lambda_{i+1}+2}$. Hence we see by induction hypothesis,

$$|d_i(X)| = |\det X_i|^{1/2} = |\det(n_i * (Y_{-\lambda})_i)|^{1/2} = |d_i(Y_{-\lambda})|. \qquad \square$$

From the two lemmas above, we have $\theta(P_0 w_0 H_0) = \mathcal{A}' = P_0 B_0 * Y_{-\lambda}$. Let $h_\lambda = \mathrm{diag}(\varpi^{\lambda_1}, 1, \varpi^{\lambda_2}, 1, \ldots, \varpi^{\lambda_n}, 1)$. Then $\theta(h_\lambda w_0) = Y_\lambda$ and $P_0 w_0 H_0 = P_0 B_0 h_{-\lambda} w_0 H_0$. In other words, $B_0 w_0 g_{-\lambda} \subset P_0 w_0 H_0$ since $g_{-\lambda} = w_0^{-1} h_{-\lambda} w_0$.

Lemma 3.3. *For all* λ, $B \xi w_0 g_{-\lambda} \subset P_\phi \xi w_0 H$.

Proof. Identifying G_0 with M by

$$g \mapsto \begin{pmatrix} g & \\ & {}^t g^{-1} \end{pmatrix},$$

we see that $\xi P_0 \xi^{-1} \subset P_\phi$, $H_0 \subset H$ and $\xi B_0 \xi^{-1} \subset B$. From the previous argument, $P_\phi \xi B_0 w_0 g_{-\lambda} H \subset P_\phi \xi P_0 w_0 H_0 H$.

Since ξ and w_0 are commutative, we obtain

$$P_\phi \xi P_0 w_0 H_0 H = P_\phi (\xi P_0 \xi^{-1}) \xi w_0 H = P_\phi \xi w_0 H.$$

On the other hand, $P_\phi \xi B_0 w_0 g_{-\lambda} H = P_\phi (\xi B_0 \xi^{-1}) \xi w_0 g_{-\lambda} H$. By Iwahori decomposition, $B = (B \cap P_\phi)(\xi B_0 \xi^{-1})(B \cap \xi N \xi^{-1})$ and since $w_0 g_{-\lambda} \in G_0$, $\xi N \xi^{-1} = (\xi w_0 g_{-\lambda}) N (\xi w_0 g_{-\lambda})^{-1}$. Therefore, $P_\phi (\xi B_0 \xi^{-1}) \xi w_0 g_{-\lambda} H = P_\phi B \xi w_0 g_{-\lambda} H$ and the desired inclusion follows. $\square$

Let $\eta = \xi w_0$, $S = \eta H \eta^{-1}$. Hereafter, we will treat S instead of H and so we need to translate all things defined above as follows:

$$\Psi_S(s) = \Psi_H(\eta^{-1} s \eta), \qquad \Lambda_S = \Lambda_H \circ R_{\eta^{-1}} \in \mathrm{Hom}_S(I(\chi), \Psi_S),$$
$$\Delta_S = \mathcal{P}_\chi^*(\Lambda_S) \in \mathcal{D}(G), \quad \Omega_S(g) = \Omega_H(\eta^{-1} g) = \Omega_H(\eta^{-1} g \eta).$$

We have to compute $\Omega_H(g_\lambda) = \Omega_S((\xi w_0) g_\lambda (\xi w_0)^{-1}) = \Omega_S(h_{-\lambda})$. Since by Lemma 3.3, we have $\operatorname{supp}(R_{h_{-\lambda}} \operatorname{ch}_B) = B h_\lambda \subset P_\phi S$, and taking (2-2) into consideration, we obtain the following result:

Proposition 3.4. *Let $\chi = (|\cdot|^{z_1}, |\cdot|^{z_2}, \dots, |\cdot|^{z_{2n}})$ be an unramified character on P_ϕ and assume that this character satisfies $z_{2i-1} = 1 + z_{2i}$ for all $1 \le i \le n$.*

(i) *If χ is not of the form above (or its W-translate), then $I(\chi)$ does not have a generalized Shalika model.*

(ii) *For $w \notin \Gamma$, we have $T^*_{w^{-1}} \Delta_S(R_{h_{-\lambda}} \operatorname{ch}_B) = 0$ for every λ, where Γ is the subgroup of W generated by*

$$w_i := \begin{array}{c} \begin{array}{cccccc} \phantom{\mathbb{1}_2} & & i & \phantom{\mathbb{1}_2} & & \phantom{\mathbb{1}_2} \\ & & \vee & & & \end{array} \\ \begin{pmatrix} \mathbb{1}_2 & & \vdots & & & \\ & \ddots & \vdots & & & \\ & & 0_2 & \mathbb{1}_2 & & \\ & & \mathbb{1}_2 & 0_2 & & \\ & & & & \ddots & \\ & & & & & \mathbb{1}_2 \end{pmatrix} \end{array} \in G_0, \ (1 \le i \le n-1) \ \textit{and}$$

$$w_n := \left(\begin{array}{cc|cc} \mathbb{1}_{2(n-1)} & & & \\ & 0_2 & & \varepsilon \\ \hline & & \mathbb{1}_{2(n-1)} & \\ & \varepsilon & & 0_2 \end{array}\right) \in G.$$

Proof. (essentially the same as [Sakellaridis 2006, Proposition 5.2])

(i) Let $I_S(w\chi)$ be the subspace of $I(w\chi)$ consisting of elements supported in $P_\phi S$. Then the restriction map induces an isomorphism $I_S(w\chi) \to \text{c-ind}^S_{P_\phi \cap S}(w\chi)\delta^{1/2}$. On the other hand, there is a surjective map $\mathcal{P}_r : C_c^\infty(S) \to \text{c-ind}^S_{P_\phi \cap S}(w\chi)\delta^{1/2}$ defined by

$$\mathcal{P}_r(f)(s) = \int_{P_\phi \cap S} (w\chi)\delta^{1/2}(p)^{-1} f(ps)\, d_r p$$

where $d_r p$ is a right Haar measure on $P_\phi \cap S$. Composed with these maps, $T^*_{w^{-1}} \Lambda_S$ can be taken as a distribution on S. Then $\Psi_S \cdot T^*_{w^{-1}} \Lambda_S$ is a right S-invariant distribution, which must be a Haar measure on S:

$$T^*_{w^{-1}} \Lambda_S = \Psi_S^{-1}\, ds.$$

For $x \in P_\phi \cap S$, $f \in C_c^\infty(S)$,

$$\begin{aligned}(w\chi)\delta^{1/2}\delta_{P_\phi\cap S}^{-1}(x)\int_S f(s)T^*_{w^{-1}}\Lambda_S(s)\,ds &= \int_S f(xs)T^*_{w^{-1}}\Lambda_S(s)\,ds\\ &= \int_S f(xs)\Psi_S^{-1}(s)\,ds\\ &= \Psi_S(x)\int_S f(s)\Psi_S^{-1}(s)\,ds\\ &= \Psi_S(x)\int_S f(s)T^*_{w^{-1}}\Lambda_S(s)\,ds,\end{aligned}$$

where $\delta_{P_\phi\cap S}$ is the modular character of $P_\phi \cap S$. Since $P_\phi \cap S$ consists of matrices of the form

$$p = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_n\end{pmatrix}, \quad A_i = \begin{pmatrix} a_i & b_i \\ & a_i^{-1}\end{pmatrix} \in \mathrm{SL}_2(F)$$

and is contained in $G_0 \cong M$, Ψ_S is trivial on $P_\phi \cap S$. So we have

$$\delta_{P_\phi\cap S}(x) = (w\chi)\delta^{1/2}(x) \tag{3-1}$$

for all $x \in P_\phi \cap S$.

An easy calculation shows that $\delta_{P_\phi\cap S}(p) = \delta(p) = \prod_i |a_i|^2$ and hence we get $(w\chi)(p) = \prod_i |a_i|$. If we put $w\chi = (|\cdot|^{z_1}, \dots, |\cdot|^{z_{2n}})$, then $(w\chi)(p) = \prod_i |a_i|^{z_{2i-1}-z_{2i}}$ and so it is necessary for the existence of a generalized Shalika model that $z_{2i-1} - z_{2i} = 1$ for every $1 \le i \le n$.

(ii) Note that Γ is isomorphic to the Weyl group of type B_n, in particular, it is a Coxeter group. It is easy to see that Γ consists of elements which preserves the condition (3-1) and the claim follows immediately. □

This proposition proves the first half of the main theorem. Throughout this paper, assume that χ satisfies the conditions stated in Proposition 3.4.

Since $\chi\delta^{1/2}\delta_{P_\phi\cap S}^{-1} = 1$ on $P_\phi \cap S$ and S is unimodular, there exists a nonzero right S-invariant linear functional $I : \text{c-ind}_{P_\phi\cap S}^S \chi\delta^{1/2} \to \mathbb{C}$ (where the action of S on $\mathbb{C}$ is trivial). We habitually use an integral expression

$$I(\varphi) = \int_{P_\phi\cap S\backslash S} \varphi(s)\,d\dot{s}$$

for $\varphi \in \text{c-ind}_{P_\phi\cap S}^S \chi\delta^{1/2}$. Note that this is not an integral in the usual sense since "integrands" are twisted by characters. This functional is uniquely determined by right S-invariance up to a positive constant factor (see [Bushnell and Henniart 2006, Proposition 3.4]). For an element φ of $\text{c-ind}_{P_\phi\cap S}^S \chi\delta^{1/2}$, $\varphi \cdot \Psi_S^{-1}$ is also an element of $\text{c-ind}_{P_\phi \cap S}^{S} \chi \delta^{1/2}$ and it follows that

$$\int_{P_\phi \cap S \backslash S} \varphi(s) \Psi_S^{-1}(s)\, d\dot{s}$$

is well defined. On the other hand, $\mathcal{P}_r^* I$ is a right S-invariant distribution on S, which is a Haar measure on S. Therefore, by the argument in the proof of Proposition 3.4,

$$\mathcal{P}_r^* \Lambda_S = \Psi_S^{-1} \mathcal{P}_r^* I = \mathcal{P}_r^* (\Psi_S^{-1} I).$$

In other words, the restriction of Λ_S to $I_S(\chi)$ has an integral expression:

Lemma 3.5. *For $\varphi \in I_S(\chi)$,*

$$\Lambda_S(\varphi) = \int_{P_\phi \cap S \backslash S} \varphi(s) \Psi_S^{-1}(s)\, d\dot{s}. \tag{3-2}$$

In a similar way, using uniqueness of invariant distributions and the linear functional $C_c^\infty(P_\phi \times S) \to C_c^\infty(P_\phi S)$ defined by

$$P_\phi S \ni ps \mapsto \int_{p_\phi \cap S} f(px^{-1}, xs)\, d_r x, \qquad f \in C_c^\infty(P_\phi \times S),$$

we obtain the following result:

Lemma 3.6. *The map $\Theta_\chi : P_\phi S \to \mathbb{C}$ defined by $\Theta_\chi(ps) = \chi^{-1}\delta^{1/2}(p)\Psi_S^{-1}(s)$ for $ps \in P_\phi S$ is well defined and for every $f \in C_c^\infty(P_\phi S)$ and*

$$\Delta_S(f) = \int_{P_\phi S} \Theta_\chi(x) f(x)\, dx, \tag{3-3}$$

where dx is a suitably normalized Haar measure on G.

Proposition 3.7. *Assume that* $\operatorname{Re} z_i > 0$ *for all i. Then* (3-2) *converges absolutely for every $\varphi \in I(\chi)$.*

Proof. (essentially the same as [Sakellaridis 2006, Proposition 7.1])

We will treat Λ_H in place of Λ_S. The equation (3-2) is equivalent to saying that for every $\varphi \in I(\chi)$ with support contained in $P_\phi \eta H$,

$$\Lambda_H(\varphi) = \int_{{}_H P_\phi \backslash H} \varphi(\eta h) \Psi_H^{-1}(h)\, d\dot{h}. \tag{3-4}$$

Here, ${}_H P_\phi = \eta^{-1} P_\phi \eta \cap H$. Hence we need to prove that (3-4) converges absolutely for every $\varphi \in I(\chi)$.

Since every element of $I(\chi)$ is dominated by some multiple of ϕ_K, it suffices to treat the case $\varphi = \phi_K$. By Iwasawa decomposition and K-invariance of ϕ_K, (3-4) is reduced to

$$\int \phi_K\left(\eta \begin{pmatrix} \mathbb{1}_{2n} & X \\ & \mathbb{1}_{2n} \end{pmatrix} \begin{pmatrix} m & \\ & {}^t m^{-1} \end{pmatrix}\right) \psi\left(\tfrac{1}{2}\operatorname{tr}(JX)\right) dX\, dm,$$

where X is a skew-symmetric matrix and

$$m = \begin{pmatrix} a & \\ & {}^t a^{-1} \end{pmatrix} \begin{pmatrix} \mathbb{1}_{2n} & Y \\ & \mathbb{1}_{2n} \end{pmatrix} \in \mathrm{Sp}_{2n}(F)$$

with an upper triangular unipotent matrix $a \in \mathrm{GL}_{2n}(F)$ and a symmetric matrix $Y \in \mathrm{Mat}_n(F)$. The integral over a is taken modulo matrices of the form

$$\begin{pmatrix} \mathbb{1}_n & b \\ & \mathbb{1}_n \end{pmatrix} \in \mathrm{GL}_{2n}(F),$$

where $b \in \mathrm{Mat}_n(F)$ is a diagonal matrix. Then

$$\begin{aligned}\int \phi_K\left(\eta \begin{pmatrix} \mathbb{1}_{2n} & X \\ & \mathbb{1}_{2n} \end{pmatrix} \begin{pmatrix} m & \\ & {}^t m^{-1} \end{pmatrix}\right) \psi\left(\tfrac{1}{2}\operatorname{tr}(JX)\right) dX\, d\dot{m} \\ = \int \phi_K\left(\eta \begin{pmatrix} m & \\ & {}^t m^{-1} \end{pmatrix} \begin{pmatrix} \mathbb{1}_{2n} & m^{-1}X{}^t m^{-1} \\ & \mathbb{1}_{2n} \end{pmatrix}\right) \psi\left(\tfrac{1}{2}\operatorname{tr}(JX)\right) dX\, d\dot{m}.\end{aligned}$$

Then $m^{-1}X{}^t m^{-1}$ can be replaced by X since H is unimodular and $m \in \mathrm{Sp}_{2n}(F)$.

Since $\phi_K \in I(\chi)$, m on the left factor can be assumed to be of the form

$$\begin{pmatrix} c & d \\ & \mathbb{1}_n \end{pmatrix},$$

with an upper triangular unipotent matrix $c \in \mathrm{GL}_n(F)$ and an upper triangular nilpotent matrix $d \in \mathrm{Mat}_n(F)$ (here, the integral is taken in the usual sense, not in that of Lemma 3.5). Therefore, the integral above is dominated absolutely by the integral representing the intertwining operator T_η (see [Casselman 1995, Lemma 6.4.2]), which converges absolutely when $\operatorname{Re} z_i > 0$ for all i by [Casselman 1980, Lemma 3.2]. □

Thanks to Proposition 3.7, exactly the same argument given in [Sakellaridis 2006, Section 7] suggests that for any $f \in C_c^\infty(G)$, $\Delta_{S,\chi}(f)$ is a rational function of χ.

4. End of calculations

Normalize the Haar measure on G so that $\mathrm{vol}(B) = 1$.

Lemma 4.1. *For any* λ, $\Delta_S(\mathrm{ch}_{Bh_\lambda}) = \chi^{-1}\delta^{1/2}(h_\lambda)$.

Proof. Since $Bh_\lambda \subset P_\phi S$ and (3-3), $\Delta_S(\mathrm{ch}_{Bh_\lambda}) = \int_{Bh_\lambda} \Theta_\chi(x)\, dx$. Using Iwahori decomposition of B and B_0, every $b \in B$ can be expressed in the form $b = pqr$, where

$$p=\begin{pmatrix}\mathbb{1}_{2n} & * \\ & \mathbb{1}_{2n}\end{pmatrix}\in B\cap N\subset P_\phi,\quad q=\begin{pmatrix}* & 0_{2n} \\ 0_{2n} & *\end{pmatrix}\in \xi B_0\xi^{-1},$$

$$r=\begin{pmatrix}\mathbb{1}_{2n} & \\ * & \mathbb{1}_{2n}\end{pmatrix}\in B\cap \xi N\xi^{-1}.$$

Then

$$bh_\lambda = pqr\eta g_{-\lambda}\eta^{-1} = p\xi\cdot\underbrace{\xi^{-1}q\xi}_{\in B_0}\cdot w_0 g_{-\lambda}\cdot\underbrace{(\eta g_{-\lambda})^{-1}r(\eta g_{-\lambda})}_{\in N}\cdot\eta^{-1}.$$

Since $B_0w_0g_{-\lambda}\subset P_0w_0H_0$, there are $p_0\in P_0$ and $h_0\in H_0$ satisfying $b_0w_0g_{-\lambda}=p_0w_0h_0$, where $b_0:=\xi^{-1}q\xi$. In other words,

$$b_0Y_{-\lambda}{}^tb_0=\theta(b_0w_0g_{-\lambda})=\theta(p_0w_0h_0)=p_0Y_0{}^tp_0=:X\in\mathscr{A}.$$

By this and Lemma 3.2,

$$|d_i(X)|=|d_i(b_0*Y_{-\lambda})|=|d_i(Y_{-\lambda})|=|d_i(p_0*Y_0)|=|\det(p_0)_i|.$$

Denote the diagonal component of p_0 by $(c_1,\dots,c_{2n})$. Then we have $|d_i(X)|=\prod_{j=1}^{2i}|c_j|=q^{\lambda_1+\cdots+\lambda_i}$ and therefore the remark on page 481 shows that p_0 can be chosen so that $c_{2i}=1,\ \ |c_{2i-1}|=q^{\lambda_i}$ for each i.

Let $n_0=(\eta g_{-\lambda})^{-1}r(\eta g_{-\lambda})$. Then

$$bh_\lambda=p\xi p_0w_0h_0n_0\eta^{-1}=\underbrace{p\cdot\xi p_0\xi^{-1}}_{\in P_\phi}\cdot\underbrace{\eta h_0n_0\eta^{-1}}_{\in S}$$

Hence,

$$\Theta_\chi(bh_\lambda)=\chi^{-1}\delta^{1/2}(p\xi p_0\xi^{-1})\Psi_H(h_0n_0)=\prod_{i=1}^{n}q^{-(2n-2i+1-z_{2i-1})\lambda_i}\Psi_H(n_0)$$

Express r in the form

$$\begin{pmatrix}\mathbb{1}_{2n} & \\ X & \mathbb{1}_{2n}\end{pmatrix},$$

where X is an element of $\mathrm{Mat}_{2n}(\mathfrak{p})$. Since

$$n_0=(w_0g_{-\lambda})^{-1}\begin{pmatrix}\mathbb{1}_{2n} & X \\ & \mathbb{1}_{2n}\end{pmatrix}(w_0g_{-\lambda})=\begin{pmatrix}\mathbb{1}_{2n} & g_\lambda{}^tw_0Xw_0g_\lambda \\ & \mathbb{1}_{2n}\end{pmatrix}$$

and the conductor of ψ is assumed to be 0,

$$\Psi_H(n_0)=\psi\left(\tfrac12\operatorname{tr}(Jg_\lambda{}^tw_0Xw_0g_\lambda)\right)=\psi\left(\tfrac12\operatorname{tr}(X\cdot Y_\lambda)\right)=1.$$

Some additional simple computations show that $\Delta_S(\mathrm{ch}_{Bh_\lambda})=\chi^{-1}\delta^{1/2}(h_\lambda)$. $\square$

Proposition 4.2.

$$\Omega_S(h_{-\lambda}) = Q^{-1}\sum_{w\in\Gamma}\Bigg(\prod_{\substack{\alpha>0\\ w\alpha>0}} c_\alpha(\chi)\Bigg)(w\chi)^{-1}\delta^{1/2}(h_\lambda)T^*_{w^{-1}}\Delta_{S,\chi}(\mathrm{ch}_B).$$

Proof. By the uniqueness of the generalized Shalika model, $T^*_{w^{-1}}\Lambda_{S,\chi}$ is a scalar multiple of $\Lambda_{S,w\chi}$. Hence,

$$\begin{aligned}\frac{T^*_{w^{-1}}\Delta_{S,\chi}(R_{h_{-\lambda}}\,\mathrm{ch}_B)}{T^*_{w^{-1}}\Delta_{S,\chi}(\mathrm{ch}_B)} = \frac{T^*_{w^{-1}}\Lambda_{S,\chi}(R_{h_{-\lambda}}\phi_B)}{T^*_{w^{-1}}\Lambda_{S,\chi}(\phi_B)} &= \frac{\Lambda_{S,w\chi}(R_{h_{-\lambda}}\phi_B)}{\Lambda_{S,w\chi}(\phi_B)}\\ &= \frac{\Delta_{S,w\chi}(R_{h_{-\lambda}}\,\mathrm{ch}_B)}{\Delta_{S,w\chi}(\mathrm{ch}_B)}\\ &= (w\chi)^{-1}\delta^{1/2}(h_\lambda).\end{aligned}$$

Applying this to (2-2), the desired result follows. □

We denote the length function of W by l and that of Γ by l_Γ. The following lemma suggests that we only have to treat the case $w = w_i$ in the notation of Proposition 3.4.

Lemma 4.3. *For $w, w' \in \Gamma$, $l_\Gamma(ww') = l_\Gamma(w) + l_\Gamma(w')$ implies that $l(ww') = l(w) + l(w')$.*

Notice that a reduced expression of each w_i is given as follows:

$$w_i = s_{2i}s_{2i-1}s_{2i+1}s_{2i},\ (1 \le i \le n-1), \quad w_n = s_{2n}.$$

Following [Casselman 1980], we denote $\mathcal{P}_\chi(\mathrm{ch}_{BwB})$ by $\phi_{w,\chi}$ for each $w \in W$.

Let N_ϕ be the unipotent radical of P_ϕ and N_ϕ^- be that of the opposite of P_ϕ. For $\alpha \in \Sigma$, N_ϕ^α (resp. $N_\phi^{-,\alpha}$) will denote the image of standard embedding $F \to N_\phi$ (resp. $F \to N_\phi^-$) corresponding to α. We will use $N_\phi^{\hat\alpha}$ (resp. $N_\phi^{-,\widehat{-\alpha}}$) to denote the product (in any order) of all N_ϕ^β (resp. $N_\phi^{-,-\beta}$), $(0 < \beta \ne \alpha)$. Similarly, for a subset $\Sigma' \subset \Sigma$, we define $N_\phi^{\Sigma'}$, $N_\phi^{\widehat{\Sigma'}}$, etc. Let $P_\phi^\alpha = T \cdot N_\phi^\alpha$ and so on.

We use the following fundamental equation of intertwining operators T_w and functions ϕ (see [Casselman 1980, Theorem 3.4]): for each simple reflection s_k and $w \in W$ with $l(s_k w) = l(w) + 1$, we have

$$T_{s_k}(\phi_{w,s_k\chi}) = (c_{\alpha_k}(s_k\chi) - 1)\phi_{w,\chi} + q^{-1}\phi_{s_k w,\chi}, \tag{4-1}$$

$$T_{s_k}(\phi_{w,s_k\chi}) = \phi_{w,\chi} + (c_{\alpha_k}(s_k\chi) - q^{-1})\phi_{s_k w,\chi}. \tag{4-2}$$

Lemma 4.4. *Let $w = w_n$ and $\beta = \beta_n$. Then, $T^*_{w^{-1}}\Delta_{S,\chi}(\mathrm{ch}_B) = -c_\beta(\chi)\chi(a_\beta)$.*

Proof. Since $w = s_{2n}$ is a simple reflection, we can apply (4-1) and obtain $T_{w^{-1}}(\phi_{B,w\chi}) = (c_\beta(w\chi) - 1)\phi_{1,\chi} + q^{-1}\phi_{w,\chi}$. Using the integral expression (3-3), it follows that $\Lambda_S(\phi_{1,\chi}) = 1$ (with Haar measure normalized so that the volume of B is 1). Therefore, it remains to compute $\Lambda_{S,\chi}(\phi_{w,\chi})$.

Assume $\mathrm{Re}\, z_i > 0$ for all i so that Δ_S is given by (3-3). In order to use the integral expression (3-3) again, we need to express elements of BwB in the form $P_\phi S$. Note that BwB need not be contained in $P_\phi S$, but almost all (i.e., except elements in certain set of measure 0) elements must be contained.

We use the following measure-preserving decomposition where all compact groups which appear are assumed to be total measure 1:

$$BwB = P_\phi(\mathfrak{o}) w N^\beta(\mathfrak{o}) N^{-,\widehat{-\beta}}(\mathfrak{p}).$$

An easy calculation shows that $\mathrm{Lie}(N^{-,\widehat{-\beta}})(\mathfrak{p}) \subset \mathrm{Lie}(P_\phi^{\hat{\beta}})(\mathfrak{p}) + \mathrm{Lie}(S)(\mathfrak{p})$, and by an argument similar to the proof of [Sakellaridis 2006, Lemma 5.1], we have $N^{-,\widehat{-\beta}}(\mathfrak{p}) \subset P_\phi^{\hat{\beta}}(\mathfrak{o}) S(\mathfrak{o})$. Consequently,

$$\Lambda_S(\phi_{w,\chi}) = \Delta_S(\mathrm{ch}_{BwB}) = \int_{BwB} \Theta_\chi(x)\, dx = q \int_{wN_\phi^\beta(\mathfrak{o})} \Theta_\chi(x)\, dx.$$

The domain of the integral $wN_\phi^\beta(\mathfrak{o})$ consists of elements of the form

$$\left(\begin{array}{cc|cc} \mathbb{1}_{2(n-1)} & & & \\ & 0_2 & & \varepsilon \\ \hline & & \mathbb{1}_{2(n-1)} & \\ & \varepsilon & & -x\cdot\mathbb{1}_2 \end{array}\right) =: A(x),$$

with $x \in \mathfrak{o}$. If $x \neq 0$,

$$A(x) = \left(\begin{array}{cc|cc} \mathbb{1}_{2(n-1)} & & & \\ & x^{-1}\cdot\mathbb{1}_2 & & -\varepsilon \\ \hline & & \mathbb{1}_{2(n-1)} & \\ & & & x\cdot\mathbb{1}_2 \end{array}\right) \left(\begin{array}{cc|cc} \mathbb{1}_{2(n-1)} & & & \\ & -\mathbb{1}_2 & & \\ \hline & & \mathbb{1}_{2(n-1)} & \\ & x^{-1}\varepsilon & & -\mathbb{1}_2 \end{array}\right) \in P_\phi S.$$

Therefore,

$$\Theta_\chi(A(x)) = |x|^{z_{2n-1}+z_{2n}-1} \psi\big(\tfrac{1}{2}\,\mathrm{tr}(x^{-1}\varepsilon^2)\big) = |x|^{z_{2n-1}+z_{2n}-1} \psi^{-1}(x^{-1})$$

and

$$\Lambda_S(\phi_w,\chi) = q\int_{\mathfrak{o}} |x|^{z_{2n-1}+z_{2n}-1} \psi^{-1}(x^{-1})\, dx = q\sum_{i=0}^{\infty} (q\chi(a_\alpha))^i \int_{\mathfrak{p}^i - \mathfrak{p}^{i+1}} \psi^{-1}(x^{-1})\, dx.$$

Substituting

$$\int_{\mathfrak{p}^i-\mathfrak{p}^{i+1}} \psi^{-1}(x^{-1})\, dx = \begin{cases} 1-q^{-1} & (i=0) \\ -q^{-2} & (i=1) \\ 0 & (i \geq 2) \end{cases}$$

for the above equation, it follows that

$$\Lambda_S(\phi_{w,\chi}) = q(1-q^{-1}-q^{-1}\chi(a_\beta)).$$

By putting all this together and after some simple algebraic manipulation, the desired equation follows. By rationality, we can drop the assumption of $\operatorname{Re} z_i > 0$ and the result follows for all χ. $\square$

Lemma 4.5. *Let $w = w_i$ for fixed $1 \le i \le n-1$. Then*

$$T^*_{w^{-1}}\Delta_S(\mathrm{ch}_B) = \frac{q^{-1}(q^{-2}x-1)(x-q^{-1})}{(q^{-1}x-1)(x-1)},$$

where $x = \chi(a_{\alpha_{2i}})$.

Proof. Applying (4-1), we obtain $T_{s_j^{-1}}\phi_{1,s_j\chi} = (c_{\alpha_j}(s_j\chi)-1)\phi_{1,\chi} + q^{-1}\phi_{s_j\chi}$ for each $2i-1 \le j \le 2i+1$. Since $s_j \notin \Gamma$, $T^*_{s_j^{-1}}\Lambda_{S,\chi}(\phi_{1,s_j\chi}) = 0$ and we get

$$q^{-1}\Lambda_{S,\chi}(\phi_{s_j,\chi}) = -(c_{\alpha_j}(s_j\chi)-1). \tag{4-3}$$

Repeating the same argument gives us the following equations: for every distinct $j,k,l \in \{2i-1, 2i, 2i+1\}$,

$$q^{-2}\Lambda_{S,\chi}(\phi_{s_k s_j,\chi}) = (c_{\alpha_k}(s_k\chi)-1)(c_{\alpha_j}(s_j\chi)-1), \tag{4-4}$$

$$q^{-3}\Lambda_{S,\chi}(\phi_{s_l s_k s_j,\chi}) = -(c_{\alpha_l}(s_l\chi)-1)(c_{\alpha_k}(s_k\chi)-1)(c_{\alpha_j}(s_j\chi)-1). \tag{4-5}$$

For $j \in \{2i-1, 2i+1\}$, we also obtain

$$\begin{aligned} q^{-3}&\Lambda_{S,\chi}(\phi_{s_{2i}s_j s_{2i}},\chi) \\ &= (c_{\alpha_{2i+1}}(s_{2i+1}s_{2i}\chi)-1)(c_{\alpha_{2i}}(s_{2i}\chi)-q^{-1})(c_{\alpha_{2i}}(s_{2i}\chi)-1) \\ &\qquad - q^{-1}(c_{\alpha_{2i+1}}(s_{2i+1}s_2\chi)-1) - (c_{\alpha_{2i}}(s_{2i}\chi)-1)^2(c_{\alpha_{2i+1}}(s_{2i+1}\chi)-1). \end{aligned} \tag{4-6}$$

Using (4-1) and (4-2) repeatedly, we can express $T_{w^{-1}}(\phi_{1,w\chi})$ as a linear combination of functions ϕ. Substituting (4-3), (4-4) and (4-5), we obtain

$$\begin{aligned} \Lambda_S(T_{w^{-1}}(\phi_{1,w\chi})) &= (c_{\alpha_{2i}}(s_{2i}\chi)-1)^2(c_{\alpha_{2i-1}}(s_{2i-1}\chi)-1)(c_{\alpha_{2i+1}}(s_{2i+1}s_{2i}\chi)-1) \\ &\quad + (c_{\alpha_{2i-1}}(s_{2i-1}s_{2i}\chi)-1)(c_{\alpha_{2i}}(s_{2i}\chi)-q^{-1}) \\ &\qquad\qquad (c_{\alpha_{2i}}(s_{2i}\chi)-1)(c_{\alpha_{2i+1}}(s_{2i+1}s_{2i}\chi)-1) \\ &\quad - q^{-1}(c_{\alpha_{2i-1}}(s_{2i-1}s_{2i}\chi)-1)(c_{\alpha_{2i+1}}(s_{2i+1}s_{2i}\chi)-1) \\ &\quad - (c_{\alpha_{2i}}(s_{2i}\chi)-1)^2(c_{\alpha_{2i-1}}(s_{2i-1}\chi)-1)(c_{\alpha_{2i+1}}(s_{2i+1}s_{2i}\chi)-1) \\ &\quad - (c_{\alpha_{2i}}(s_{2i}\chi)-1)^2(c_{\alpha_{2i-1}}(s_{2i-1}\chi)-1)(c_{\alpha_{2i+1}}(s_{2i+1}\chi)-1) \\ &\quad + q^{-4}\Lambda_S(\phi_{w,\chi}). \end{aligned}$$

Simple computations using

$$
\begin{aligned}
c_{\alpha_{2i+1}}(s_{2i+1}\chi) = c_{\alpha_{2i-1}}(s_{2i-1}\chi) &= 0, \\
c_{\alpha_{2i}}(s_{2i}\chi) - 1 &= \frac{1-q^{-1}}{\chi(a_{\alpha_{2i}})-1}, \\
c_{\alpha_{2i+1}}(s_{2i+1}s_{2i}\chi) - 1 = c_{\alpha_{2i-1}}(s_{2i-1}s_{2i+1}s_{2i}\chi) - 1 &= \frac{1-q^{-1}}{q^{-1}\chi(a_{\alpha_{2i}})-1}, \\
c_{\alpha_{2i}}(s_{2i}\chi) - q^{-1} &= \frac{1-q^{-1}}{\chi(a_{\alpha_{2i}})-1}\chi(a_{\alpha_{2i}})
\end{aligned}
$$

show that

$$
\begin{aligned}
\Lambda_S(T_{w^{-1}}(\phi_{1,w\chi})) &= \frac{(1-q^{-1})^4}{(q^{-1}x-1)^2(x-1)^2}x - q^{-1}\frac{(1-q^{-1})^2}{(q^{-1}x-1)^2} \\
&\qquad - \frac{(1-q^{-1})^2}{(x-1)^2} + q^{-4}\Lambda_S(\phi_{w,\chi}) \\
&= -\frac{(1+q^{-1})(1-q^{-1})^2}{(q^{-1}x-1)(x-1)} + q^{-4}\Lambda_S(\phi_{w,\chi}),
\end{aligned}
$$

where $x = \chi(a_{\alpha_{2i}})$.

It remains to compute $\Lambda_S(\phi_{w,\chi})$. This can be done by essentially the same method as the proof of Lemma 4.4.

Assume $\operatorname{Re} z_1 > \operatorname{Re} z_2 > \cdots > \operatorname{Re} z_{2n} > 0$ so that Δ_S is given by (3-3). For later use, we make a stronger assumption. Let

$$
\Sigma_i = \{\alpha \in \Sigma \mid \alpha > 0,\ w\alpha < 0\} = \{e_{2i-1}-e_{2i+1}, e_{2i-1}-e_{2i+2}, e_{2i}-e_{2i+1}, e_{2i}-e_{2i+2}\}.
$$

Then $BwB = P_\phi(\mathfrak{o})wN_\phi^{\Sigma_i}(\mathfrak{o})N_\phi^{-,\widehat{-\Sigma_i}}(\mathfrak{p})$. An easy calculation shows that

$$
\mathrm{Lie}(N_\phi^{-,\widehat{-\Sigma_i}})(\mathfrak{p}) \subset \mathrm{Lie}(S)(\mathfrak{p}) + \mathrm{Lie}(P_\phi^{\widehat{\Sigma_i}})(\mathfrak{p}),
$$

and by an argument similar to the proof of Lemma 5.1 of [Sakellaridis 2006], we have $N_\phi^{-,\widehat{-\Sigma_i}}(\mathfrak{p}) \subset P_\phi^{\widehat{\Sigma_i}}(\mathfrak{o})S(\mathfrak{o})$. Therefore,

$$
\Lambda_S(\phi_{w,\chi}) = \Delta_S(\mathrm{ch}_{BwB}) = \int_{BwB}\Theta_\chi(x)\,dx = q^4\int_{wN_\phi^{\Sigma_i}(\mathfrak{o})}\Theta_\chi(x)\,dx
$$

Then $wN_\phi^{\Sigma_i}(\mathfrak{o})$ consists of elements of the form

$$
B(a) = \begin{pmatrix} \mathbb{1}_{2(i-1)} & & & \\ & 0_2 & \mathbb{1}_2 & \\ & \mathbb{1}_2 & a & \\ & & & \mathbb{1}_{2(n-i-1)} \end{pmatrix} \in G_0
$$

with $a \in \mathrm{Mat}_2(\mathfrak{o})$. If $\det a \neq -1$, let

$$b = \begin{pmatrix} 1+\det a & 0 \\ 0 & 1 \end{pmatrix}.$$

Then

$$\begin{pmatrix} \mathbb{1}_{2(i-1)} & & & \\ & b & \varepsilon b^{-1}{}^t a \varepsilon & \\ & & b^{-1} & \\ & & & \mathbb{1}_{2(n-i-1)} \end{pmatrix} B(a) \in S \cap G_0.$$

Thus,

$$\Theta_\chi(B(a)) = |1+\det a|^{z_{2i-1}-z_{2i+1}-2}$$

and

$$\begin{aligned} \Lambda_S(\phi_{w,\chi}) &= q^4 \int_{\left(\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right) \in \mathrm{Mat}_2(o)} |1+xw-yz|^{z_{2i-1}-z_{2i+1}-2}\, dx\, dy\, dz\, dw \\ &= q^4 \cdot \mathrm{vol}(\mathrm{Mat}_2(\mathfrak{o}) - \mathrm{GL}_2(\mathfrak{o})) \\ &\qquad + q^4 \int_{\left(\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathfrak{o})} |1+xw-yz|^{z_{2i-1}-z_{2i+1}-2}\, dx\, dy\, dz\, dw. \end{aligned}$$

The first term can be computed as follows. Since the restriction of a Haar measure on $\mathrm{Mat}_2(\mathfrak{o})$ to $\mathrm{GL}_2(\mathfrak{o})$ is equal to the restriction of a Haar measure on $\mathrm{GL}_2(F)$,

$$\mathrm{vol}(\mathrm{GL}_2(\mathfrak{o})) = (q+1) \cdot \mathrm{vol} \begin{pmatrix} \mathfrak{o}^\times & \mathfrak{o} \\ \mathfrak{p} & \mathfrak{o}^\times \end{pmatrix} = q^{-3}(q-1)^2(q+1),$$

and hence $\mathrm{vol}(\mathrm{Mat}_2(\mathfrak{o}) - \mathrm{GL}_2(\mathfrak{o})) = 1 - q^{-3}(q-1)^2(q+1)$.

Next, we need to compute the second term. There is a diffeomorphism f between $\mathrm{GL}_2(\mathfrak{o})$ and $\mathfrak{o}^\times \times \mathrm{SL}_2(\mathfrak{o})$ given by

$$\mathfrak{o}^\times \times \mathrm{SL}_2(\mathfrak{o}) \ni \left(t, \begin{pmatrix} x & y \\ z & w \end{pmatrix}\right) \mapsto \begin{pmatrix} tx & ty \\ z & w \end{pmatrix} \in \mathrm{GL}_2(\mathfrak{o}).$$

The Jacobian of f on the region $\mathfrak{o}^\times \times \{w \neq 0\} \subset \mathfrak{o}^\times \times \mathrm{SL}_2(\mathfrak{o})$ is $Jf = t/w$. Since the complement of this region is a set of measure 0, we can transform the second term into an integral on $\mathfrak{o}^\times \times \mathrm{SL}_2(\mathfrak{o})$:

$$\begin{aligned} &\int_{\left(\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathfrak{o})} |1+xw-yz|^{z_{2i-1}-z_{2i+1}-2}\, dx\, dy\, dz\, dw \\ &\qquad = \int_{\mathfrak{o}^\times} \int_{\{w \neq 0\}} |1+t|^{z_{2i-1}-z_{2i+1}-2} |tw^{-1}|\, d^\times t\, dy\, dz\, dw \\ &\qquad = \int_{\mathfrak{o}^\times} |1+t|^{z_{2i-1}-z_{2i+1}-2}\, dt \cdot \int_{\{w \neq 0\}} |w|^{-1}\, dy\, dz\, dw. \end{aligned}$$

First, we consider the integral $\int_{\mathfrak{o}^\times} |1+t|^{z_{2i-1}-z_{2i+1}-2}\,dt$. Split the integral into $1+t \in \mathfrak{o}^\times$ and $1+t \in \mathfrak{p}$. The former contributes $1-2q^{-1}$ and the latter meshing

$$\begin{aligned}\sum_{j=1}^{\infty} |\varpi^j|^{z_{2i-1}-z_{2i+1}-2} \cdot (1-q^{-1})q^{-j} &= (1-q^{-1}) \sum_{j=1}^{\infty} \chi(a_{\alpha_{2i}})^j \\ &= (1-q^{-1})\chi(a_{\alpha_{2i}})(1-\chi(a_{\alpha_{2i}})^{-1}.\end{aligned}$$

Here, we used the assumption $\operatorname{Re} z_1 > \operatorname{Re} z_2 > \cdots > \operatorname{Re} z_{2n} > 0$. This implies $|\chi(a_{\alpha_{2i}})| < 1$, which is necessary for convergence of the above power series.

Therefore, we have

$$\int_{\mathfrak{o}^\times} |1+t|^{z_{2i-1}-z_{2i+1}-2}\,dt = -q^{-1} + (1-q^{-1})(1-\chi(a_{\alpha_{2i}}))^{-1}.$$

Second, we compute the integral $\int_{\{w\neq 0\}} |w|^{-1}\,dy\,dz\,dw$. Splitting the integral into $w \in \mathfrak{o}^\times$ and $w \in \varpi^j\mathfrak{o}^\times$, we get

$$\begin{aligned}&\int_{\{w\neq 0\}} |w|^{-1}\,dy\,dz\,dw \\ &\qquad = \int_{w\in\mathfrak{o}^\times}\int_{y,z\in\mathfrak{o}} dy\,dz\,dw + \sum_{j=1}^{\infty}\int_{w\in\varpi^j\mathfrak{o}^\times}\int_{yz\in -1+\mathfrak{p}^j} |\varpi^j|^{-1}\,dy\,dz\,dw \\ &\qquad = 1-q^{-1} + \sum_{j=1}^{\infty}(1-q^{-1})^2 q^{-j} \\ &\qquad = 1-q^{-2}.\end{aligned}$$

Consequently, we obtain

$$\Lambda_S(\phi_{w,\chi}) = q^2 - q(q-1)^2(q+1)(\chi(a_{\alpha_{2i}})-1)^{-1}.$$

Putting all this together, the desired equation follows. By rationality, we can drop the assumption of $\operatorname{Re} z_1 > \operatorname{Re} z_2 > \cdots > \operatorname{Re} z_{2n} > 0$, and the result follows for all χ. $\square$

Some more computation enables us to rewrite these results.

Corollary. *For $w = w_n$ we have*:

$$T^*_{w^{-1}}\Lambda_{S,\chi} = -\chi(a_\beta)c_\beta(\chi)\Lambda_{S,w\chi},$$

where $\beta = \beta_n$.

For $w = w_i$ $(1 \le i < n)$ we have:

$$T^*_{w^{-1}}\Lambda_{S,\chi} = -\chi(a_\beta)\frac{1-q^{-2}\chi(a_{-\beta})}{1-q^{-2}\chi(a_\beta)}\prod_{\alpha\in\Sigma_i} c_\alpha(\chi)\Lambda_{S,w\chi},$$

where $\beta = \beta_i$.

More compactly, for every $w \in \Gamma$,

$$T^*_{w^{-1}} \Lambda_{S,\chi} = (-1)^{l_\Gamma(w)} \prod_{\substack{\alpha>0 \\ w\alpha<0}} c_\alpha(\chi) \prod_{\substack{\beta>0 \\ w\beta<0}} d_\beta(\chi) \Lambda_{S,w\chi},$$

where $\alpha \in \Sigma$, $\beta \in \Phi$.

These complete the proof of Theorem 2.4.

Acknowledgements

I would like to thank Tamotsu Ikeda, my advisor, for directing my attention to this question and for guiding and encouraging my research.

References

[Bushnell and Henniart 2006] C. J. Bushnell and G. Henniart, *The local Langlands conjecture for* GL(2), Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] **335**, Springer, Berlin, 2006. MR Zbl

[Casselman 1980] W. Casselman, "The unramified principal series of $\mathfrak{p}$-adic groups. I. The spherical function", *Compositio Math.* **40**:3 (1980), 387–406. MR Zbl

[Casselman 1995] W. Casselman, "Introduction to the theory of admissible repressentaitons of p-adic reductive groups", book draft, 1995, Available at https://www.math.ubc.ca/∼cass/research/pdf/p-adic-book.pdf.

[Casselman and Shalika 1980] W. Casselman and J. Shalika, "The unramified principal series of p-adic groups. II. The Whittaker function", *Compositio Math.* **41**:2 (1980), 207–231. MR Zbl

[Hironaka 1999] Y. Hironaka, "Spherical functions and local densities on Hermitian forms", *J. Math. Soc. Japan* **51**:3 (1999), 553–581. MR Zbl

[Jiang and Qin 2007] D. Jiang and Y. Qin, "Residues of Eisenstein series and generalized Shalika models for SO_{4n}", *J. Ramanujan Math. Soc.* **22**:2 (2007), 101–133. MR Zbl

[Nien 2010] C. Nien, "Local uniqueness of generalized Shalika models for SO_{4n}", *J. Algebra* **323**:2 (2010), 437–457. MR Zbl

[Sakellaridis 2006] Y. Sakellaridis, "A Casselman–Shalika formula for the Shalika model of GL_n", *Canad. J. Math.* **58**:5 (2006), 1095–1120. MR Zbl

Received February 20, 2017.

MIYU SUZUKI
DEPARTMENT OF MATHEMATICS
KYOTO UNIVERSITY
KYOTO
JAPAN
msuzuki@math.kyoto-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 294, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.294.495

NONTAUTOLOGICAL BIELLIPTIC CYCLES

JASON VAN ZELM

Let $[\overline{\mathcal{B}}_{2,0,20}]$ and $[\mathcal{B}_{2,0,20}]$ respectively be the classes of the loci of stable and of smooth bielliptic curves with 20 marked points where the bielliptic involution acts on the marked points as the permutation $(1\ 2)\cdots(19\ 20)$. Graber and Pandharipande proved that these classes are nontautological. In this note we show that their result can be extended to prove that $[\overline{\mathcal{B}}_g]$ is nontautological for $g \geq 12$ and that $[\mathcal{B}_{12}]$ is nontautological.

1. Introduction

The system of *tautological rings* $\{R^\bullet(\overline{\mathcal{M}}_{g,n})\}$ is defined (see [Faber and Pandharipande 2005]) to be the minimal system of $\mathbb{Q}$-subalgebras of the Chow rings $A^\bullet(\overline{\mathcal{M}}_{g,n})$ closed under pushforward along the natural gluing and forgetful morphisms

$$\begin{aligned}\overline{\mathcal{M}}_{g_1,n_1+1} \times \overline{\mathcal{M}}_{g_2,n_2+1} &\to \overline{\mathcal{M}}_{g_1+g_2,n_1+n_2},\\ \overline{\mathcal{M}}_{g,n+2} &\to \overline{\mathcal{M}}_{g+1,n},\\ \overline{\mathcal{M}}_{g,n+1} &\to \overline{\mathcal{M}}_{g,n}.\end{aligned}$$

The tautological ring $R^\bullet(\mathcal{M}_{g,n})$ of the moduli space of smooth curves is the image of $R^\bullet(\overline{\mathcal{M}}_{g,n})$ under the localization morphism $A^\bullet(\overline{\mathcal{M}}_{g,n}) \to A^\bullet(\mathcal{M}_{g,n})$. We denote the image of $R^\bullet(\overline{\mathcal{M}}_{g,n})$ under the cycle map $A^\bullet(\overline{\mathcal{M}}_{g,n}) \to H^{2\bullet}(\overline{\mathcal{M}}_{g,n})$ by $RH^{2\bullet}(\overline{\mathcal{M}}_{g,n})$ and define $RH^{2\bullet}(\mathcal{M}_{g,n})$ accordingly. We say a cohomology class is *tautological* if it lies in the tautological subring of its cohomology ring; otherwise we say it is *nontautological*. In this note we work over $\mathbb{C}$ and all Chow and cohomology rings are assumed to be taken with rational coefficients.

The tautological rings are relatively well understood. An additive set of generators for the vector spaces $R^\bullet(\overline{\mathcal{M}}_{g,n})$ is given by decorated boundary strata and there exists an algorithm for computing the intersection product; see [Graber and Pandharipande 2003]. The class of many "geometrically defined" loci can be shown to be tautological. For example, this is the case for the class of the locus $\overline{\mathcal{H}}_g$ of hyperelliptic curves in $\overline{\mathcal{M}}_g$; see [Faber and Pandharipande 2005, Theorem 1].

MSC2010: primary 14H10; secondary 14H30, 14H37.
Keywords: nontautological, bielliptic.

Any odd cohomology class of $\overline{\mathcal{M}}_{g,n}$ is nontautological by definition. Deligne proved that $H^{11}(\overline{\mathcal{M}}_{1,11}) \neq 0$, thus providing a first example of the existence of nontautological classes. In fact, it is known that $H^\bullet(\overline{\mathcal{M}}_{0,n}) = RH^\bullet(\overline{\mathcal{M}}_{0,n})$ [Keel 1992] and that $H^{2\bullet}(\overline{\mathcal{M}}_{1,n}) = RH^{2\bullet}(\overline{\mathcal{M}}_{1,n})$ [Petersen 2014, Corollary 1.2].

Examples of geometrically defined loci which can be proven to be nontautological are still relatively scarce. In [Graber and Pandharipande 2003], Graber and Pandharipande hunt for algebraic classes in $H^{2\bullet}(\overline{\mathcal{M}}_{g,n})$ and in $H^{2\bullet}(\mathcal{M}_{g,n})$ which are nontautological. In particular, they show that the classes of the loci $\overline{\mathcal{B}}_{g,n,2m}$ and $\mathcal{B}_{g,n,2m}$ of, respectively, stable and smooth bielliptic curves of genus g, with n marked points fixed by the bielliptic involution and $2m$ marked points pairwise switched by the bielliptic involution, are nontautological when $g = 2$, $n = 0$ and $2m = 20$ (i.e., $[\overline{\mathcal{B}}_{2,0,20}] \notin RH^\bullet(\overline{\mathcal{M}}_{2,20})$ and $[\mathcal{B}_{2,0,20}] \notin RH^\bullet(\mathcal{M}_{2,20})$). They also show that for sufficiently high odd genus h, the class of the locus of stable curves of genus $2h$ admitting a map to a curve of genus h is nontautological in $\overline{\mathcal{M}}_{2h}$. Their result relies on the existence of odd cohomology in $H^\bullet(\overline{\mathcal{M}}_{h,1})$, which was proven in [Pikaart 1995] for all $h \geq 8069$. See [Faber and Pandharipande 2013] for a recent survey of different methods of detecting nontautological classes.

In [Petersen and Tommasi 2014; Petersen 2016], Petersen and Tommasi proved that $H^{2\bullet}(\overline{\mathcal{M}}_{2,n})$ is tautological for all $n < 20$ and that $H^{2\bullet}(\overline{\mathcal{M}}_{2,20})$ is additively generated by tautological classes, by the class $[\overline{\mathcal{B}}_{2,0,20}]$, and by its conjugates under the action of the symmetric group on 20 elements. In this sense the result of Graber and Pandharipande for the bielliptic locus is sharp.

In this note we prove the following two new results.

Theorem 1. *The cohomology class $[\overline{\mathcal{B}}_{g,n,2m}]$ is nontautological for all $g + m \geq 12$, $0 \leq n \leq 2g - 2$ and $g \geq 2$.*

Theorem 2. *The cohomology class $[\mathcal{B}_{g,0,2m}]$ is nontautological when $g + m = 12$ and $g \geq 2$.*

Theorem 1 reduces the genus for which algebraic nontautological classes on $\overline{\mathcal{M}}_g$ are known to exist from 16138 to 12. As far as the author is aware, Theorem 2 provides the first example of a nontautological algebraic class on $\mathcal{M}_g$.

2. Preliminaries

Let $\mathcal{B}_{g,n,2m} \subset \mathcal{M}_{g,n+2m}$ be the locus of smooth bielliptic curves for which the bielliptic involution acts on the last $2m$ markings as the involution $(1\ 2) \cdots (2m-1\ 2m)$ and fixes the remaining points, and let $\overline{\mathcal{B}}_{g,n,2m} \subset \overline{\mathcal{M}}_{g,n+2m}$ be its closure. A modular interpretation of $\overline{\mathcal{B}}_{g,n,2m}$ can be given by *admissible double covers* [Abramovich et al. 2003].

Definition 3. We define $\overline{\mathrm{Adm}}(g, h)_{2m}$ to be the stack parameterizing *admissible double covers* from curves of genus g to curves of genus h with $2m$ points switched by the involution. Specifically, $\overline{\mathrm{Adm}}(g, h)_{2m}$ parameterizes tuples

$$(C, D, f, y_1, \dots, y_{2m})$$

together with a *total ordering* of the smooth ramification points of f such that

- $f : C \to D$ is a double cover of connected nodal curves of arithmetic genus g and h, respectively,
- $y_1, \dots, y_{2m}$ are points in the smooth locus of C such that the covering involution swaps the points y_{2k-1} and y_{2k} pairwise,
- the image of each node of C under f is a node,
- the curve C, equipped with the markings given by the set of all ramification points and the points $y_1, \dots, y_{2m}$, is stable, and so is the curve D, equipped with the markings given by the ordered set of all smooth branch points and the images of the points $y_1, \dots, y_{2m}$.

There is a natural map $\phi_n : \overline{\mathrm{Adm}}(g, h)_{2m} \to \overline{\mathcal{M}}_{g,n+2m}$ which assigns to an admissible double cover $(C, D, f, y_1, \dots, y_{2m})$ the stabilization of the curve

$$(C, x_1, \dots, x_n, y_1, \dots, y_{2m}),$$

where the $(x_i)_{i=1}^n$ are the first n smooth ramification points of f. The space $\overline{\mathcal{B}}_{g,n,2m}$ equals the image of $\overline{\mathrm{Adm}}(g, 1)_{2m}$ under ϕ_n.

By using the Riemann–Hurwitz formula inductively on the number of nodes of D, we see that the map f must have $2g + 2 - 4h$ ramification points. The map $\overline{\mathrm{Adm}}(g, h)_{2m} \to \overline{\mathcal{M}}_{h,2g+2-4h+m}$, mapping each admissible cover to its target curve together with its marked points, is finite. In the bielliptic case it follows that the dimension of $\overline{\mathcal{B}}_{g,n,2m}$ is $2g - 2 + 2m$. The classes of these loci are denoted by

$$[\overline{\mathcal{B}}_{g,n,2m}] \in A^{g-1+n+2m}(\overline{\mathcal{M}}_{g,2m+n}) \quad \text{and} \quad [\mathcal{B}_{g,n,2m}] \in A^{g-1+n+2m}(\mathcal{M}_{g,2m+n}).$$

Similarly, we let $\mathcal{H}_{g,n,2m}$ be the locus of smooth hyperelliptic curves with n marked points fixed and $2m$ points pairwise permuted by the hyperelliptic involution. We denote its closure inside $\overline{\mathcal{M}}_{g,n+2m}$ by $\overline{\mathcal{H}}_{g,n,2m}$. This closure equals the image of $\overline{\mathrm{Adm}}(g, 0)_{2m}$ under ϕ_n.

Our proof of Theorem 1 relies on the following result for pullbacks along gluing morphisms.

Proposition 4 [Graber and Pandharipande 2003, Proposition 1]. *Let*

$$\xi : \overline{\mathcal{M}}_{g_1,n_1+1} \times \overline{\mathcal{M}}_{g_2,n_2+1} \to \overline{\mathcal{M}}_{g_1+g_2,n_1+n_2}$$

be the gluing morphism and $\gamma \in RH^\bullet(\overline{\mathcal{M}}_{g_1+g_2,n_1+n_2})$. *Then*

$$\xi^*(\gamma) \in RH^\bullet(\overline{\mathcal{M}}_{g_1,n_1+1}) \otimes RH^\bullet(\overline{\mathcal{M}}_{g_2,n_2+1}).$$

We say that a cycle $\lambda \in H^\bullet(\overline{\mathcal{M}}_{g_1,n_1}) \otimes H^\bullet(\overline{\mathcal{M}}_{g_2,n_2})$ *admits a tautological Künneth decomposition* if $\lambda \in RH^\bullet(\overline{\mathcal{M}}_{g_1,n_1}) \otimes RH^\bullet(\overline{\mathcal{M}}_{g_2,n_2})$. Proposition 4 says that the pullback of a tautological class admits a tautological Künneth decomposition. It can be shown that the pullback of a tautological class under the gluing morphism $\overline{\mathcal{M}}_{g,n+2} \to \overline{\mathcal{M}}_{g+1,n}$ and the forgetful morphism $\overline{\mathcal{M}}_{g,n+1} \to \overline{\mathcal{M}}_{g,n}$ is tautological. In this sense the tautological ring is also closed under pullbacks along the gluing and the forgetful morphisms.

3. Proof of Theorems 1 and 2

We are now ready to prove Theorem 1. We start by proving the following weaker result.

Proposition 5. *We have*

$$[\overline{\mathcal{B}}_{g,0,2m}] \notin RH^\bullet(\overline{\mathcal{M}}_{g,2m})$$

for $g+m=12$ and $g \geq 2$.

Proof. Let

$$i : \overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11} \to \overline{\mathcal{M}}_{g,2m} \tag{$\dagger$}$$

be the gluing morphism that pairwise identifies the first $g-1$ points on the first curve with the first $g-1$ points on the second curve. In Lemma 6 we will prove that the restriction of $i^*[\overline{\mathcal{B}}_{g,0,2m}]$ to the interior $\mathcal{M}_{1,11} \times \mathcal{M}_{1,11}$ is a positive scalar multiple α of the class $[\Delta]$ of the diagonal. Let $\partial(\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11})$ denote the normalization of $(\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11}) \backslash (\mathcal{M}_{1,11} \times \mathcal{M}_{1,11})$. It follows from the localization sequence

$$A^{10}(\partial(\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11})) \to A^{11}(\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11}) \to A^{11}(\mathcal{M}_{1,11} \times \mathcal{M}_{1,11}) \to 0$$

that $i^*[\overline{\mathcal{B}}_{g,0,2m}] = \alpha \cdot \Delta + B$, with B supported on the image of $\partial(\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11})$.

The class B admits a tautological Künneth decomposition by Lemma 7(i). Given a homogeneous basis $\{e_i\}_{i\in I}$ for $H^\bullet(\overline{\mathcal{M}}_{1,11})$ with dual basis $\{\hat{e}_i\}_{i\in I}$, the cohomology class of the diagonal can be written as

$$[\Delta] = \sum_{i\in I} (-1)^{\deg e_i} e_i \otimes \hat{e}_i.$$

In particular, since $H^{11}(\overline{\mathcal{M}}_{1,11}) \neq 0$, the diagonal $[\Delta]$ does not admit a tautological Künneth decomposition. Since the pullback of a tautological class along a (composition of) gluing morphisms admits a tautological Künneth decomposition by repeated application of Proposition 4, this shows that $[\overline{\mathcal{B}}_{g,0,2m}]$ is nontautological. □

Lemma 6. *Let $g+m=12$ and $g \geq 2$. The pullback of $[\overline{\mathcal{B}}_{g,0,2m}]$ to $\mathcal{M}_{1,11} \times \mathcal{M}_{1,11}$, under the restriction j of the gluing map i defined in ($\dagger$), is a scalar multiple α of the class of the diagonal Δ.*

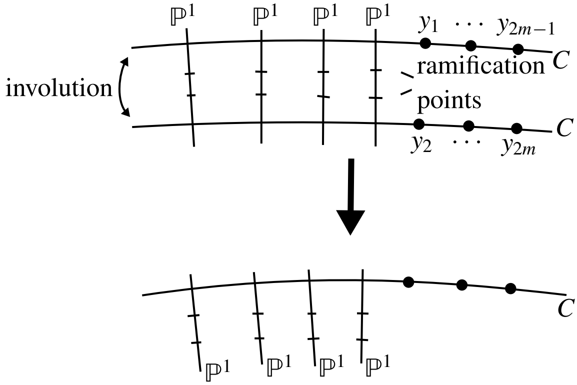


Figure 1. The image of C under η.

Proof. Let η be the map $\mathcal{M}_{1,11} \to \overline{\mathrm{Adm}}(g,1)_{2m}$ which maps a curve $(C, x_1, \ldots, x_{11})$ to the admissible cover which has as a source curve two copies of C glued together by rational bridges attached to the first $g-1$ points of each copy of C, as covering involution the bielliptic involution which switches around the two copies of C and has two fixed points on each of the rational bridges, and as target curve a single copy of C with a rational component attached to the first $g-1$ points (see Figure 1). Let $\delta : \mathcal{M}_{1,11} \to \mathcal{M}_{1,11} \times \mathcal{M}_{1,11}$ be the diagonal morphism. Consider the diagram

$$(\ddagger)\qquad \begin{tikzcd} \mathcal{M}_{1,11} \arrow[rrd, bend left, "\eta"] \arrow[rd, dashed, "\zeta"] \arrow[rdd, bend right, "\delta"'] & & \\ & F \arrow[r] \arrow[d, "\tilde{\phi}_0"] \arrow[rd, phantom, "\square"] & \overline{\mathrm{Adm}}(g,1)_{2m} \arrow[d, "\phi_0"] \\ & \mathcal{M}_{1,11} \times \mathcal{M}_{1,11} \arrow[r, "j"'] & \overline{\mathcal{M}}_{g,2m} \end{tikzcd}$$

By unwrapping definitions one verifies that $j \circ \delta = \phi_0 \circ \eta$. By the universal property of fiber products this defines a unique map $\zeta : \mathcal{M}_{1,11} \to F$, making diagram (‡) commute.

Claim: The morphism ζ is surjective on closed points.

Assuming the claim, it follows that $\tilde{\phi}_{0*}[F]$ is a positive scalar multiple of $\delta_*[\mathcal{M}_{1,11}] = [\Delta]$. Since

$$\operatorname{codim}_{\mathcal{M}_{1,11}\times\mathcal{M}_{1,11}} \Delta = 11 = \operatorname{codim}_{\overline{\mathcal{M}}_{g,2m}} \overline{\mathcal{B}}_{g,0,2m},$$

it follows that there is no excess of intersection between $\mathcal{M}_{1,11} \times \mathcal{M}_{1,11}$ and $\overline{\mathcal{B}}_{g,0,2m} = \phi_0(\overline{\mathrm{Adm}}(g,1)_{2m})$ in diagram (‡). We deduce that $j^*[\overline{\mathcal{B}}_{g,0,2m}] = \alpha[\Delta]$ for some $\alpha \in \mathbb{Q}_{>0}$.

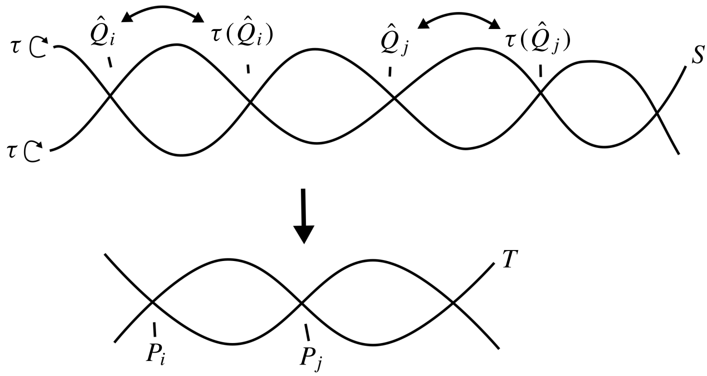


Figure 2. The admissible cover $S \to T$ when τ fixes C_1 and C_2.

Proof of the claim. By definition, an object of $F(\mathbb{C})$ consists of a curve $\widetilde{C} := (\tilde{C}_1, \widetilde{C}_2)$ in $\mathcal{M}_{1,11} \times \mathcal{M}_{1,11}(\mathbb{C})$, an object $(S \to T) \in \overline{\text{Adm}}(g, 1)_{2m}(\mathbb{C})$ and an isomorphism $\gamma : j(\widetilde{C}) \xrightarrow{\sim} \phi_0(S \to T)$. To prove the claim, we show that $(\widetilde{C}, (S \to T), \gamma)$ is isomorphic to an object in the image of ζ. Let $f : \widetilde{C}_1 \cup \widetilde{C}_2 \to j(\widetilde{C})$ be the map of curves induced by j, set $C := j(\widetilde{C})$, $C_1 := f(\widetilde{C}_1)$ and $C_2 := f(\widetilde{C}_2)$, let τ be the involution on C induced by the bielliptic involution of $S \to T$ and let Q_i be the node of C corresponding to the i-th marking of $\widetilde{C}_1$ and $\widetilde{C}_2$ via the morphism f.

Since C_1 and C_2 are smooth, there are two possibilities for the action of τ on C: either it fixes C_1 and C_2 or it switches the whole of C_1 with the whole of C_2.

Suppose τ fixes C_1 and C_2. By construction the involution τ maps marked points lying on C_1 to marked points lying on C_2 so this is only possible if C has no marked points at all. In this case τ must fix the different branches of C at each Q_i. If the preimage of Q_i in S were to be a genus 0 curve R_i, contracted by the stabilization map, then R_i would have 2 marked ramification points which are not nodes. But this would imply that τ switches the nodes on R_i and it would therefore also switch the branches of C at Q_i. It follows that the preimage of each Q_i in S is a single node $\hat{Q}_i$. Since C_1 and C_2 are smooth, τ induces an involution on the set of nodes $\{\hat{Q}_1, \ldots, \hat{Q}_{11}\}$. We can thus find distinct $\hat{Q}_i$, $\hat{Q}_j \neq \tau(\hat{Q}_i)$ such that $S - \{\hat{Q}_i, \tau(\hat{Q}_i), \hat{Q}_j, \tau(\hat{Q}_j)\}$ is connected. If P_i and P_j are the images of Q_i and Q_j, respectively, under the admissible cover $S \to T$ then this means that $T - \{P_i, P_j\}$ is connected (see Figure 2). This implies that the arithmetic genus of T is at least 2, which is a contradiction.

We can therefore assume τ maps C_1 to C_2. Let us first suppose that τ does not fix all nodes, so there exist some distinct i, j such that $\tau(Q_i) = Q_j$ (see Figure 3). If the preimage of Q_i in S is a component of S contracted by the stabilization map, then this component must contain a ramification point. This would be a fixed point of the involution, contradicting the assumption that $\tau(Q_i) = Q_j$. So the preimage

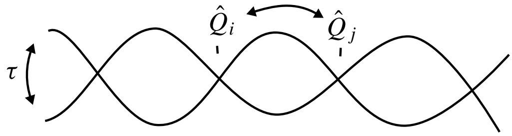


Figure 3. Nodes in S not fixed by τ.

of Q_i and Q_j in S are nodes $\hat{Q}_i$ and $\hat{Q}_j$. Let P be the image of $\{\hat{Q}_i, \hat{Q}_j\}$ under the bielliptic map. Arguing as at the end of the last paragraph, we see that $T\backslash\{P\}$ is connected. Therefore, since T has arithmetic genus 1, it has geometric genus 0. However, if S_1 is the irreducible component of S which surjects onto C_1 under the stabilization map, then S_1 is a smooth curve of geometric genus 1. This is a contradiction because $S_1 \to T_1$ is a birational map.

We have thus proven that τ switches the components C_1 and C_2 and fixes the nodes Q_i, which implies that $((\widetilde{C}_1, \widetilde{C}_2), (S \to T), \gamma)$ is isomorphic to an object in the image of $\mathcal{M}_{1,11}(\mathbb{C})$. This concludes the proof that the map $\mathcal{M}_{1,11} \to F$ is surjective on closed points. $\square$

Lemma 7. (i) *Every algebraic class of codimension* 11 *in* $\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11}$ *supported on* $\partial(\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11})$ *admits a tautological Künneth decomposition.*

(ii) *Every algebraic class on* $\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11}$ *of codimension less than* 11 *admits a tautological Künneth decomposition.*

Proof. This is a slightly weaker version of [Graber and Pandharipande 2003, Lemma 3]; the proof given there requires that $RH^{2\bullet}(\overline{\mathcal{M}}_{1,n}) = H^{2\bullet}(\overline{\mathcal{M}}_{1,n})$ and $H^{\mathrm{odd}}(\overline{\mathcal{M}}_{1,n}) = 0$ for $n < 11$, for which there was no reference at the time of their paper. The first equation is [Petersen 2014, Corollary 1.2]. The second condition follows from the computations for $n < 11$ in [Getzler 1998]. $\square$

We have now concluded the proof of Proposition 5. To prove Theorem 1 it remains to show that $[\overline{\mathcal{B}}_{g,n,2m}]$ is nontautological for all g, n, m with $0 \le n \le 2g-2$ and $g+m > 12$.

Proof of Theorem 1. We will show in Lemma 8 and 9 that if $[\overline{\mathcal{B}}_{g,n,2m}]$ is nontautological then so are $[\overline{\mathcal{B}}_{g,n+1,2m}]$ for $n \le 2g-3$, and $[\overline{\mathcal{B}}_{g,n,2m+2}]$. In Lemma 10 we will show that if $[\overline{\mathcal{B}}_{g,1,0}]$ is nontautological then so is $[\overline{\mathcal{B}}_{g+1}]$. Using these statements, and by induction with the statement of Proposition 5 as the base case, we conclude that $[\overline{\mathcal{B}}_{g,n,2m}]$ is nontautological for all $g+m \ge 12$. $\square$

Lemma 8. *If* $[\overline{\mathcal{B}}_{g,n,2m}]$ *is nontautological and* $n \le 2g-3$, *then so is* $[\overline{\mathcal{B}}_{g,n+1,2m}]$.

Proof. Let $\pi : \overline{\mathcal{M}}_{g,n+1+2m} \to \overline{\mathcal{M}}_{g,n+2m}$ be the morphism that forgets the first point and stabilizes. By definition $\pi_*([\overline{\mathcal{B}}_{g,n+1,2m}])$ is a positive scalar multiple of $[\overline{\mathcal{B}}_{g,n,2m}]$. Because the pushforward of a tautological class by the forgetful morphism is tautological by definition, the result follows. $\square$

Lemma 9. *If $[\overline{\mathcal{B}}_{g,n,2m}]$ is nontautological, then so is $[\overline{\mathcal{B}}_{g,n,2m+2}]$.*

Proof. If $n \le 2g-3$ then $[\overline{\mathcal{B}}_{g,n+1,2m}]$ is nontautological by Lemma 8. Consider the gluing morphism

$$\sigma : \overline{\mathcal{M}}_{g,n+2m+1} \times \overline{\mathcal{M}}_{0,3} \to \overline{\mathcal{M}}_{g,n+2m+2}$$

which glues the first points of both curves together; then $\sigma^{-1}(\overline{\mathcal{B}}_{g,n,2m+2}) = \overline{\mathcal{B}}_{g,n+1,2m}$.

Since

$$\operatorname{codim}_{\overline{\mathcal{M}}_{g,n+2m+2}} \overline{\mathcal{B}}_{g,n,2m+2} = \operatorname{codim}_{\overline{\mathcal{M}}_{g,n+2m+1}} \overline{\mathcal{B}}_{g,n+1,2m},$$

it follows that $\sigma^*[\overline{\mathcal{B}}_{g,n,2m+2}] = \alpha[\overline{\mathcal{B}}_{g,n+1,2m}]$ for some $\alpha \in \mathbb{Q}_{>0}$. Since σ is a gluing morphism and the pullback of a tautological class along σ admits tautological Künneth decomposition, $[\overline{\mathcal{B}}_{g,n,2m+2}]$ is nontautological.

If $n = 2g-2$ we first prove that $[\overline{\mathcal{B}}_{g,n-1,2m+2}]$ is nontautological as above by pulling back along the map $\overline{\mathcal{M}}_{g,n+2m} \times \overline{\mathcal{M}}_{0,3} \to \overline{\mathcal{M}}_{g,n+2m+1}$ and then applying Lemma 8. □

Lemma 10. *If $[\overline{\mathcal{B}}_{g,1,0}]$ is nontautological, then so is $[\overline{\mathcal{B}}_{g+1}]$.*

Proof. Let $\epsilon : \overline{\mathcal{M}}_{g,1} \times \overline{\mathcal{M}}_{1,1} \to \overline{\mathcal{M}}_{g+1}$ be the gluing morphism. From the description of the boundary divisors of $\overline{\mathcal{B}}^{\mathrm{Adm}}_{g+1}$ [Pagani 2016, pp. 1275–1276], it follows that there exist $\alpha, \beta \in \mathbb{Q}_{>0}$ such that

$$\epsilon^*[\overline{\mathcal{B}}_{g+1}] = \alpha[\overline{\mathcal{B}}_{g,1,0} \times \overline{\mathcal{M}}_{1,1}] + \beta[(\overline{\mathcal{H}}_{g-1,0,2}, \overline{\mathcal{M}}_{1,1})] \in H^\bullet(\overline{\mathcal{M}}_{g,1} \times \overline{\mathcal{M}}_{1,1}),$$

where $(\overline{\mathcal{H}}_{g-1,0,2}, \overline{\mathcal{M}}_{1,1})$ denotes the locus of pairs $(C, E) \in \overline{\mathcal{M}}_{g,1} \times \overline{\mathcal{M}}_{1,1}$, where C consists of a genus $g-1$ hyperelliptic curve C' glued to an elliptic curve E' isomorphic to E, with the hyperelliptic involution switching the marked point of C' with the point of intersection with E'. The class $[(\overline{\mathcal{H}}_{g-1,0,2}, \overline{\mathcal{M}}_{1,1})]$ admits a tautological Künneth decomposition because the diagonal inside $\overline{\mathcal{M}}_{1,1} \times \overline{\mathcal{M}}_{1,1}$ does, the class of the hyperelliptic locus is tautological by [Faber and Pandharipande 2005, Theorem 1], and the pushforward of tautological classes under a gluing morphism is tautological by definition. The class $[\overline{\mathcal{B}}_{g,1} \times \overline{\mathcal{M}}_{1,1}]$ does not admit a tautological Künneth decomposition because $[\overline{\mathcal{B}}_{g,1}]$ is nontautological. It follows by Proposition 4 that $[\overline{\mathcal{B}}_{g+1}]$ is nontautological. □

We now complete the proof of Theorem 2.

Proof of Theorem 2. The case $g = 2$ is treated in [Graber and Pandharipande 2003, Section 3]. We use a similar argument to prove the remaining cases. The proof runs by contradiction.

Suppose $[\mathcal{B}_{g,0,2m}] \in RH^\bullet(\mathcal{M}_{g,2m})$; then there is a collection of cycles Z_k in $\overline{\mathcal{M}}_{g,2m}$, of codimension 11 and supported on $\partial\overline{\mathcal{M}}_{g,2m}$, such that $\sum[Z_k] + [\overline{\mathcal{B}}_{g,0,2m}]$ is tautological. Consider again the gluing morphism $i : \overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11} \to \overline{\mathcal{M}}_{g,2m}$ of (†). By assumption, the pullback of $\sum[Z_k] + [\overline{\mathcal{B}}_{g,0,2m}]$ to $\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11}$

admits a tautological Künneth decomposition whereby the pullback of $\sum[Z_k]$ to $\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11}$ must be nontautological (by Proposition 4 and since the pullback of $[\overline{\mathcal{B}}_{g,0,2m}]$ is nontautological, as we have shown in the proof of Theorem 1).

We denote by Δ_h the locus of curves in $\overline{\mathcal{M}}_{g,2m}$ consisting of two curves, one of which has genus h, glued together in a single node, and by Δ_{irr} the locus that generically parameterizes irreducible singular curves. So $\partial\overline{\mathcal{M}}_{g,2m} = \Delta_{\mathrm{irr}} \cup \bigcup_h \Delta_h$.

Suppose Z_k is supported on Δ_h for some h. Since $i(\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11})$ does not have a separating node, we see that $i(\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11}) \not\subset \Delta_h$. The intersection

$$\Delta_h \cap (\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11})$$

therefore lies in the image of $\partial(\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11})$. It follows by Lemma 7(i) that $i^*[Z_k]$ admits a tautological Künneth decomposition.

Suppose now that Z_k is supported on Δ_{irr}. We decompose the map i as

$$\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11} \xrightarrow{i_1} \overline{\mathcal{M}}_{g-1,2m+2} \xrightarrow{i_2} \overline{\mathcal{M}}_{g,2m}.$$

Then there exist cycles Y_k in $\overline{\mathcal{M}}_{g-1,2m+2}$ such that $i_{2*}[Y_k] = [Z_k]$. Now

$$i^*[Z_k] = i_1^* i_2^*[Z_k] = i_1^*(c_1(N_{\overline{\mathcal{M}}_{g-1,2m+2}}\overline{\mathcal{M}}_{g,2m}) \cap [Y_k]).$$

We see that $i^*[Z_k]$ decomposes as a product of algebraic classes of codimension less than 11, all of which admit tautological Künneth decomposition by Lemma 7(ii).

We conclude that all the cycles $[Z_k]$ have a tautological Künneth decomposition when pulled back to $\overline{\mathcal{M}}_{1,11} \times \overline{\mathcal{M}}_{1,11}$, which is a contradiction. □

Acknowledgements

The author would like to thank his Ph.D. supervisor Nicola Pagani for the many helpful discussions leading up to this paper. The author would also like to thank the referees for the many helpful comments. The author is supported by a GTA Ph.D. fellowship at the University of Liverpool.

References

[Abramovich et al. 2003] D. Abramovich, A. Corti, and A. Vistoli, "Twisted bundles and admissible covers", *Comm. Algebra* **31**:8 (2003), 3547–3618. MR Zbl

[Faber and Pandharipande 2005] C. Faber and R. Pandharipande, "Relative maps and tautological classes", *J. Eur. Math. Soc.* (*JEMS*) **7**:1 (2005), 13–49. MR Zbl

[Faber and Pandharipande 2013] C. Faber and R. Pandharipande, "Tautological and non-tautological cohomology of the moduli space of curves", pp. 293–330 in *Handbook of moduli*, vol. I, edited by G. Farkas and I. Morrison, Advanced Lectures in Math. **24**, International Press, Somerville, MA, 2013. MR Zbl

[Getzler 1998] E. Getzler, "The semi-classical approximation for modular operads", *Comm. Math. Phys.* **194**:2 (1998), 481–492. MR Zbl

[Graber and Pandharipande 2003] T. Graber and R. Pandharipande, "Constructions of nontautological classes on moduli spaces of curves", *Michigan Math. J.* **51**:1 (2003), 93–109. MR Zbl

[Keel 1992] S. Keel, "Intersection theory of moduli space of stable n-pointed curves of genus zero", *Trans. Amer. Math. Soc.* **330**:2 (1992), 545–574. MR Zbl

[Pagani 2016] N. Pagani, "Moduli of abelian covers of elliptic curves", *J. Pure Appl. Algebra* **220**:3 (2016), 1258–1279. MR Zbl

[Petersen 2014] D. Petersen, "The structure of the tautological ring in genus one", *Duke Math. J.* **163**:4 (2014), 777–793. MR Zbl

[Petersen 2016] D. Petersen, "Tautological rings of spaces of pointed genus two curves of compact type", *Compos. Math.* **152**:7 (2016), 1398–1420. MR Zbl

[Petersen and Tommasi 2014] D. Petersen and O. Tommasi, "The Gorenstein conjecture fails for the tautological ring of $\overline{\mathcal{M}}_{2,n}$", *Invent. Math.* **196**:1 (2014), 139–161. MR Zbl

[Pikaart 1995] M. Pikaart, "An orbifold partition of $\overline{M}_g^n$", pp. 467–482 in *The moduli space of curves* (Texel Island, Netherlands, 1994), edited by R. Dijkgraaf et al., Progress in Math. **129**, Birkhäuser, Boston, 1995. MR Zbl

Received December 7, 2016. Revised October 18, 2017.

JASON VAN ZELM
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF LIVERPOOL
LIVERPOOL
UNITED KINGDOM
jasonvanzelm@outlook.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 294, No. 2, 2018
dx.doi.org/10.2140/pjm.2018.294.505

ADDENDUM: SINGULARITIES OF FLAT FRONTS IN HYPERBOLIC SPACE

MASATOSHI KOKUBU, WAYNE ROSSMAN, KENTARO SAJI,
MASAAKI UMEHARA AND KOTARO YAMADA

Volume **221**:2 (2005), 303–351

This is an addendum to the authors' previous paper in which criteria for cuspidal edges and swallowtails on surfaces are given by applying the so-called Zakalyukin's lemma. The original statement in Zakalyukin's paper assumed the properness of the mappings. However, the lemma in the appendix of our paper did not assume properness. Recently, we noticed that the proof given in the appendix was implicitly relying on properness. In this addendum, we prove that mappings satisfying the criteria of cuspidal edges and swallowtails have properness. Consequently, the criteria are clarified.

In [Kokubu et al. 2005], to which this note is an addendum, we found an omitted condition in the statement of Lemma 2.2, which was explained there as a lemma given by Zakalyukin. The original statement in [Zakalyukin 1976] assumed the properness of the mappings f_1 and f_2 in the lemma. We have discovered that the proof given in the appendix of [Kokubu et al. 2005] was implicitly using the properness of the mappings f_i, $(i = 1, 2)$.

In [Kokubu et al. 2005, Proposition 1.3], this lemma was applied to prove criteria for cuspidal edges and swallowtails. In this paper, we show that these criteria still remain valid. In fact, we prepare the following new lemma to replace Lemma 2.2 in [Kokubu et al. 2005].

Lemma. *Let $U(\subset \mathbb{R}^n)$ be a neighborhood of the origin, and let the mappings $f_i : (U, o) \to (\mathbb{R}^{n+1}, \mathbf{0})$, with $i = 1, 2$, be wave fronts, where o and $\mathbf{0}$ are the origins of $\mathbb{R}^n$ and $\mathbb{R}^{n+1}$, respectively. Suppose that o is a singular point of f_i and the set of regular points of f_i is dense in U for each $i = 1, 2$. Moreover, suppose that $f_i^{-1}(\mathbf{0})$ is a finite set. Then the following two statements are equivalent*:

(1) *There exist neighborhoods $V_1, V_2(\subset \mathbb{R}^n)$ of the origin o and a local diffeomorphism on $\mathbb{R}^{n+1}$ which maps the image $f_1(V_1)$ to $f_2(V_2)$, namely the image of f_1 is locally diffeomorphic to that of f_2.*

MSC2010: primary 53C42, 57R45; secondary 53A05, 53A99.
Keywords: wave front, cuspidal edge, swallowtail.

(2) *There exists a local diffeomorphism h on $\mathbb{R}^{n+1}$ and a local contact diffeomorphism Φ on the unit cotangent bundle $T_1^*\mathbb{R}^{n+1}$ of $\mathbb{R}^{n+1}$ with respect to the Euclidean metric of $\mathbb{R}^{n+1}$ which sends fibers to fibers such that $\Phi \circ L_{f_1} = L_{f_2} \circ h$, namely the lift L_{f_1} is **Legendrian equivalent** to the lift L_{f_2}.*

Remark 1. Instead of the properness in the original Zakalyukin lemma, we assume the finiteness of the inverse images $f_i^{-1}(\mathbf{0})$, $i = 1, 2$, which was dropped in Lemma 2.2 in [Kokubu et al. 2005]. The condition that $f_i^{-1}(\mathbf{0})$ is a finite set relates to the $\mathcal{K}$-finiteness of the map f_i (cf. [Wall 1981]), which plays an important role in singularity theory.

We prepare the following assertion:

Proposition. *Let $U(\subset \mathbb{R}^n)$ be a neighborhood of a point $p \in \mathbb{R}^n$, and $B_{\mathbf{0}}(r)$ be an open ball of radius $r(>0)$ centered at the origin in $\mathbb{R}^N$, and let $f : (U, p) \to (\mathbb{R}^N, \mathbf{0})$ $(N \geq n)$ be a continuous map such that $f^{-1}(\mathbf{0})$ is a finite set. Then for sufficiently small $r > 0$, the connected component V of $f^{-1}(B_{\mathbf{0}}(r))$ containing p satisfies $\overline{V} \subset U$. Moreover, the restriction of the map f to V with image inside $B_{\mathbf{0}}(r)$ is a proper mapping.*

Proof. Take a ball $W := D_p(\epsilon)$ of radius ϵ centered at p such that $\overline{W}$ is contained in U. Since $f^{-1}(\mathbf{0})$ is a finite set, we may choose the radius ϵ so that

$$f^{-1}(\mathbf{0}) \cap W = \{p\} \tag{1}$$

holds. We denote by $V(r)$ the connected component of $f^{-1}(B_{\mathbf{0}}(r))$ containing p. It is sufficient to show that $\overline{V(1/k)} \subset W$ for any sufficiently large integers $k > 0$. If not, there exists a point $q_k \notin W$ lying in $\overline{V(1/k)}$. If $q_k \notin \partial W (:= \overline{W} \setminus W)$, then q_k is an exterior point of W. Then we can find a point $q_k' \in V(1/k)$ such that q_k' is also an exterior point of W. Since $V(1/k)$ is connected, there exists a continuous curve on $V(1/k)$ joining p and q_k'. By the intermediate value theorem, for each positive integer k, there exists a point p_k satisfying

$$p_k \in \overline{V(1/k)} \cap \partial W. \tag{2}$$

On the other hand, if $q_k \in \partial W$, then (2) trivially holds by setting $p_k := q_k$.

We then take a sequence $\{q_{j,k}\}_{j=1}^{\infty}$ lying in $V(1/k)$ converging to p_k. By definition, we have $f(q_{j,k}) \in B_{\mathbf{0}}(1/k)$. By the continuity of f,

$$f(p_k) \in \overline{B_{\mathbf{0}}(1/k)} \qquad (k = 1, 2, 3, \ldots) \tag{3}$$

holds, where $\overline{B_{\mathbf{0}}(1/k)}$ is the closure of the open ball $B_{\mathbf{0}}(1/k)$. Since ∂W is compact, we can take a subsequence $\{p_{k_m}\}_{m=1}^{\infty}$ of $\{p_k\}$ which converges to a point $p_\infty \in \partial W$. Letting $m \to \infty$, equation (3) yields that $f(p_\infty) = \mathbf{0}$, which contradicts (1).

We next prove the final assertion: Suppose that K is a compact subset of $B_{\mathbf{0}}(r)$ and $f^{-1}(K)$ is not compact. Then we can take a sequence $\{x_k\}$ in $f^{-1}(K)$ not accumulating to any point of V. Since $\overline{V}$ is compact, we may assume that $\{x_k\}$ converges to a point x_∞ on $\partial V := \overline{V} \setminus V$. Since f is a continuous map on U and $\overline{V} \subset U$, there exists a connected open neighborhood O of x_∞ such that $f(O) \subset B_{\mathbf{0}}(r)$. Then $V' := V \cup O$ is a connected open subset such that $f(V') \subset B_{\mathbf{0}}(r)$, which contradicts the definition of V, since $V \subsetneq V \cup O$. □

Proof of Lemma. (1) follows from (2) immediately, so it is sufficient to show (1) implies (2). By Fact A.3 in the appendix of [Kokubu et al. 2005], we may assume $f_1(V_1) = f_2(V_2)$. By the above proposition, we can take $r > 0$ such that $V(r) := f^{-1}(B_{\mathbf{0}}(r))$ satisfies $\overline{V(r)} \subset V_1 \cap V_2$. Then we have that

$$f_1(V(r)) = f_2(V(r)).$$

By [Kokubu et al. 2005, Fact A.1], we may assume that the associated Legendrian immersion $L_{f_i} : \tilde{U}_i \to T_1^*\mathbb{R}^{n+1}$ $(i = 1, 2)$ is an embedding. Since $\overline{V(r)}$ is compact, we have

$$f_1(\overline{V(r)}) = \overline{f_1(V(r))} = \overline{f_2(V(r))} = f_2(\overline{V(r)}).$$

Thus by [Kokubu et al. 2005, Proposition A.4], we have $L_{f_1}(\overline{V_1}) = L_{f_2}(\overline{V_2})$. In particular, we have

$$L_{f_1}(V_1) \subset L_{f_2}(U_2),$$

and by [Kokubu et al. 2005, Fact A.2], there exists a local diffeomorphism φ on $\mathbb{R}^n$ such that $L_{f_2} = L_{f_1} \circ \varphi$, which proves the assertion. □

We next show the following claim, that is, that wave fronts satisfying our criteria for cuspidal edges or swallowtails also satisfy the assumption of the above lemma. Consequently, the statement of [Kokubu et al. 2005, Proposition 1.3] is clarified.

Claim 1. *Let U be a domain in $\mathbb{R}^2$, and let $f : (U, p) \to (\mathbb{R}^3, \mathbf{0})$ be a wave front such that p is a nondegenerate singular point. Take a regular curve $\gamma(t)$ parametrizing the singular set such that $\gamma(0) = p$. If f satisfies one of the two conditions*

(1) *the null vector $\eta(0)$ is linearly independent of $\dot{\gamma}(0)$, or*

(2) *$\eta(0)$ is proportional to $\dot{\gamma}(0)$, and*

$$\left.\frac{d}{dt}\right|_{t=0} \det(\dot{\gamma}(t), \eta(t)) \neq 0,$$

then the inverse image $f^{-1}(\mathbf{0})$ is a finite set.

Proof. Let ν_0 be the unit normal vector of f at p, and T the plane passing through $\mathbf{0}$ perpendicular to ν_0. We denote by $\pi : \mathbb{R}^3 \to T$ the orthogonal projection. Then $\varphi := \pi \circ f : U \to \mathbb{R}^2$ is a smooth map having a singular point at p. Then the condition (1) (respectively, (2)) turns out to be a well-known criterion for a fold singularity (respectively, Whitney cusp singularity), see [Whitney 1955] or Theorem A1 in the appendix of [Saji et al. 2009] (A_1-Morin singularity means fold singularity, and A_2-Morin singularity means Whitney cusp singularity). So φ is right-left equivalent to the map germ $(u, v) \mapsto (u^2, v)$ (respectively, $(u, v) \mapsto (u^3 - 3uv, v)$). Thus, $f^{-1}(\mathbf{0})$ is a finite set. $\square$

In [Saji et al. 2009], we gave a criterion for an A_{k+1}-singular point of a wave front for $k \geq 1$, as a generalization of the case of cuspidal edges and swallowtails. Then the same problem has arisen in that case as well, that is, to clarify the criterion, we must show that the map satisfies the condition that the inverse image of the singular point is a finite set. However, by the following claim, this is actually true:

Claim 2. *Let U be a domain in $\mathbb{R}^n$, and let $f : (U, p) \to (\mathbb{R}^{n+1}, \mathbf{0})$ be a wave front such that p is a nondegenerate singular point. If f satisfies the criterion given in [Saji et al. 2009, Theorem 2.4], then the inverse image $f^{-1}(\mathbf{0})$ consists of finitely many points.*

The proof is the same as for Claim 1: Taking the unit normal vector ν_0 of f at p, we define the orthogonal projection $\pi : \mathbb{R}^{n+1} \to H$, where H is the hyperplane passing through p orthogonal to the vector ν_0. Then $\varphi := \pi \circ f : U \to H (= \mathbb{R}^n)$ is a smooth map having a singular point at p. Then the criterion for an A_{k+1} singularity on the wave front f corresponds to the criterion for an A_k-Morin singularity of φ given in [Saji et al. 2009, Theorem A1]. Hence, the inverse image of the origin is a finite set. $\square$

Remark 2. In [Izumiya and Saji 2010; Izumiya et al. 2010; Saji 2011], criteria for cuspidal lips, cuspidal beaks, cuspidal butterflies and D_4 singularities are given. In these cases as well, one can similarly show the finiteness of the inverse image of the singular point, assuming that the criteria given in those papers hold. However, the arguments are longer and will be given in a separate work by those authors.

Acknowledgement

The authors thank Professor Shyuichi Izumiya for valuable comments.

References

[Izumiya and Saji 2010] S. Izumiya and K. Saji, "The mandala of Legendrian dualities for pseudo-spheres in Lorentz–Minkowski space and 'flat' spacelike surfaces", *J. Singul.* **2** (2010), 92–127. MR Zbl

[Izumiya et al. 2010] S. Izumiya, K. Saji, and M. Takahashi, "Horospherical flat surfaces in hyperbolic 3-space", *J. Math. Soc. Japan* **62**:3 (2010), 789–849. MR Zbl

[Kokubu et al. 2005] M. Kokubu, W. Rossman, K. Saji, M. Umehara, and K. Yamada, "Singularities of flat fronts in hyperbolic space", *Pacific J. Math.* **221**:2 (2005), 303–351. MR Zbl

[Saji 2011] K. Saji, "Criteria for D_4 singularities of wave fronts", *Tohoku Math. J.* (2) **63**:1 (2011), 137–147. MR Zbl

[Saji et al. 2009] K. Saji, M. Umehara, and K. Yamada, "A_k singularities of wave fronts", *Math. Proc. Cambridge Philos. Soc.* **146**:3 (2009), 731–746. MR Zbl

[Wall 1981] C. T. C. Wall, "Finite determinacy of smooth map-germs", *Bull. London Math. Soc.* **13**:6 (1981), 481–539. MR Zbl

[Whitney 1955] H. Whitney, "On singularities of mappings of Euclidean spaces, I : Mappings of the plane into the plane", *Ann. of Math.* (2) **62**:3 (1955), 374–410. MR Zbl

[Zakalyukin 1976] V. M. Zakalyukin, "Lagrangian and Legendre singularities", *Funkcional. Anal. i Priložen.* **10**:1 (1976), 26–36. In Russian; translated in *Func. Anal. Appl.* **10**:1 (1976), 23-31. MR Zbl

Received November 16, 2016.

Masatoshi Kokubu
Department of Mathematics, School of Engineering
Tokyo Denki University
Tokyo
Japan

kokubu@cck.dendai.ac.jp

Wayne Rossman
Department of Mathematics, Faculty of Science
Kobe University
Kobe
Japan

wayne@math.kobe-u.ac.jp

Kentaro Saji
Department of Mathematics, Graduate School of Science
Kobe University
Kobe
Japan

saji@math.kobe-u.ac.jp

Masaaki Umehara
Department of Mathematical and Computing Sciences
Tokyo Institute of Technology
Tokyo
Japan

umehara@is.titech.ac.jp

Kotaro Yamada
Department of Mathematics
Tokyo Institute of Technology
Tokyo
Japan

kotaro@math.titech.ac.jp

CONTENTS

Volume 294, 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 294 No. 2 June 2018