Pacific Journal of Mathematics

Volume 251 No. 1 May 2011

PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

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

See inside back cover or www.pjmath.org for submission instructions.

The subscription price for 2011 is US $420/year for the electronic version, and $485/year for print and electronic.
Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. Prior back issues are obtainable from Periodicals Service Company, 11 Main Street, Germantown, NY 12526-5635. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 969 Evans Hall, Berkeley, CA 94720-3840, is published monthly except July and August. 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 PACIFIC JOURNAL OF MATHEMATICS
at the University of California, Berkeley 94720-3840
A NON-PROFIT CORPORATION
Typeset in LaTeX

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 1, 2011

AN ANALOGUE OF THE CARTAN DECOMPOSITION FOR p-ADIC SYMMETRIC SPACES OF SPLIT p-ADIC REDUCTIVE GROUPS

PATRICK DELORME AND VINCENT SÉCHERRE

Let k be a nonarchimedean locally compact field of residue characteristic p, let G be a connected reductive group defined over k, let σ be an involutive k-automorphism of G, and H an open k-subgroup of the fixed points group of σ. We denote by G_k and H_k the groups of k-points of G and H. We obtain an analogue of the Cartan decomposition for the reductive symmetric space $\mathrm{H}_k\backslash\mathrm{G}_k$ in the case where G is k-split and p is odd. More precisely, we obtain a decomposition of G_k as a union of $(\mathrm{H}_k, \mathrm{K})$-double cosets, where K is the stabilizer of a special point in the Bruhat–Tits building of G over k. This decomposition is related to the H_k-conjugacy classes of maximal σ-antiinvariant k-split tori in G. In a more general context, Benoist and Oh obtained a polar decomposition for any p-adic reductive symmetric space. In the case where G is k-split and p is odd, our decomposition makes more precise that of Benoist and Oh, and generalizes results of Offen for GL_n.

1. Introduction

Let k be a nonarchimedean locally compact field of odd residue characteristic. Let G be a connected reductive group defined over k, let σ be an involutive k-automorphism of G and let H be an open k-subgroup of the fixed points group of σ. We denote by G_k and H_k the groups of k-points of G and H. Harmonic analysis on the reductive symmetric space $\mathrm{H}_k\backslash\mathrm{G}_k$ is the study of the action of G_k on the space of complex square integrable functions on $\mathrm{H}_k\backslash\mathrm{G}_k$. This study is related to the classification of H_k-distinguished representations of G_k, that is representations having a nonzero space of H_k-invariant linear forms. Offen [2004] has investigated the harmonic analysis of spherical functions in some cases related to GL_n. Hironaka [1988] has described a Cartan decomposition for the pair (GL_n, O_n). Blanc and Delorme [2008] have studied H_k-distinguishedness for families of parabolically induced representations of G_k. Lagier [2008], and independently Kato and Takano

MSC2000: 22E35.
Keywords: p-adic reductive group, building, Cartan decomposition, symmetric space.

[2008], have introduced the notion of relative cuspidality for irreducible H_k-distinguished representations of G_k and constructed "Jacquet maps" at the level of invariant linear forms. In this paper, we investigate the geometry of the reductive symmetric space $\mathrm{H}_k\backslash\mathrm{G}_k$.

Connected reductive groups can be considered as reductive symmetric spaces. Indeed, if G' is such a group, the map

$$\sigma : (x, y) \mapsto (y, x)$$

defines a k-involution of $\mathrm{G} = \mathrm{G}' \times \mathrm{G}'$ whose fixed points group H is the diagonal image of G' in G, and the reductive symmetric space $\mathrm{H}_k\backslash\mathrm{G}_k$ naturally identifies with G'_k via the map $(x, y) \mapsto x^{-1}y$. Moreover, if K' is a subgroup of G'_k, and if we set $\mathrm{K} = \mathrm{K}' \times \mathrm{K}'$, then this map induces a bijective correspondence:

$$\{(\mathrm{H}_k, \mathrm{K})\text{-double cosets of } \mathrm{G}_k\} \leftrightarrow \{\mathrm{K}'\text{-double cosets of } \mathrm{G}'_k\}.$$

In particular, if K' is the G'_k-stabilizer of a special point in the Bruhat–Tits building of G' over k, the decomposition of $\mathrm{H}_k\backslash\mathrm{G}_k$ into K-orbits corresponds to the Cartan decomposition of G'_k relative to K' [Bruhat and Tits 1972, Proposition 4.4.3].

In this paper, we obtain an analogue of the Cartan decomposition for $\mathrm{H}_k\backslash\mathrm{G}_k$ when the group G is k-split. In a more general context (k any nonarchimedean locally compact field of odd characteristic and G any connected reductive group over k), Benoist and Oh [2007] have obtained a polar decomposition for $\mathrm{H}_k\backslash\mathrm{G}_k$. In the case where k has odd residue characteristic and G is k-split, our decomposition is a refinement of Benoist–Oh's polar decomposition (see 4.14). This decomposition can be seen as a p-adic analogue of the Cartan decomposition for real reductive symmetric spaces [Flensted-Jensen 1978, Theorem 4.1]. It generalizes the decompositions obtained by Offen [2004, Proposition 3.1] for $\mathrm{G} = \mathrm{GL}_{2n}$ in what he called Cases 1 and 3.

Let $\{\mathrm{A}^j \mid j \in \mathrm{J}\}$ be a set of representatives of the H_k-conjugacy classes of maximal σ-antiinvariant k-split tori of G (called maximal (σ, k)-*split* tori in [Helminck 1994]; see also Definition 4.2). These tori, as well as related entities, have been studied in [Helminck 1994; Helminck and Helminck 1998; Helminck and Wang 1993]. In particular, the set J is finite and the A^j, $j \in \mathrm{J}$, are all conjugate under G_k. Let S be a σ-stable maximal k-split torus of G containing a maximal (σ, k)-split torus A. For each $j \in \mathrm{J}$, we choose $y_j \in \mathrm{G}_k$ such that $y_j\mathrm{A}y_j^{-1} = \mathrm{A}^j$. Our main result is this:

Theorem 1.1 (see Theorem 4.13). *Assume* G *is k-split. Let* K *be the stabilizer in* G_k *of a special point in the apartment attached to* S. *Then*

$$\mathrm{G}_k = \bigcup_{j\in\mathrm{J}} \mathrm{H}_k y_j \mathrm{S}_k \mathrm{K}. \tag{1-1}$$

If one compares with Offen's decompositions [2004, Proposition 3.1], one sees that in each of his Cases 1 and 3 (where $G = GL_{2n}$ for $n \geqslant 1$), the set J reduces to a single element and y_j can be chosen to be trivial. In general however, one cannot avoid having several non-H_k-conjugate maximal σ-antiinvariant k-split tori of G appearing in (1-1).

To prove Theorem 1.1, we make generous use of Bruhat–Tits theory [1972; 1984a]. First, let G be any connected reductive group over k, and let $\mathcal{B}$ be its Bruhat–Tits building. It is endowed with an action of σ. Then:

Proposition 1.2 (see Proposition 3.8). *$\mathcal{B}$ is the union of its σ-stable apartments.*

Note that in the case where $G = G' \times G'$ and $\sigma(x, y) = (y, x)$ as above, the building $\mathcal{B}$ identifies with the product of two copies of the building of G' over k and the proposition simply says that two arbitrary points in the building of G' are always contained in a common apartment.

When G is k-split, we obtain the following refinement of the proposition above:

Proposition 1.3 (see Proposition 4.8). *Assume* G *is k-split, and let x be a special point of $\mathcal{B}$. There is a σ-stable maximal k-split torus* S *of* G *such that the apartment corresponding to* S *contains x and the maximal σ-antiinvariant subtorus of* S *is a maximal (σ, k)-split torus of* G.

As we will see in 5.13, this is no longer true for nonsplit groups.

Summary. In Section 2, we recall the main properties of the Bruhat–Tits building attached to a connected reductive group defined over k. In Section 3, we study the set of all apartments containing a given σ-stable subset of the building, and we prove Proposition 1.2. In Section 4, we prove our main theorem for G a k-split group. In Section 5, we study in more detail the case of $G_k = GL_n(k)$ and $\sigma(g) =$ transpose of g^{-1}, and the case of $G_k = GL_n(k')$ with k' quadratic over k and $\mathrm{id} \neq \sigma \in \mathrm{Gal}(k'/k)$. When $n = 2$ and k' is totally ramified over k, the second case provides an example of a nonsplit group for which Proposition 1.3 is not satisfied.

2. The Bruhat–Tits building

Let k be a nonarchimedean nondiscrete locally compact field, and let ω be its normalized valuation. In this section, we recall the main properties of the Bruhat–Tits building attached to a connected reductive group defined over k. The reader may refer to [Bruhat and Tits 1972; 1984a] or to the more concise presentations [Landvogt 1995; Schneider and Stuhler 1997; Tits 1979].

If G is a linear algebraic group defined over k, the group of its k-points will be denoted by G_k or $G(k)$, and its neutral component will be denoted by G°. If X is a subset of G, then $N_G(X)$ and $Z_G(X)$ denote respectively the normalizer and centralizer of X in G, and, given $g \in G$, we write ${}^gX$ for gXg^{-1}.

2.1. Let G be a connected reductive group defined over k, and let S be a maximal k-split torus of G. We denote by $\mathrm{X}^*(\mathrm{S}) = \mathrm{Hom}(\mathrm{S}, \mathrm{GL}_1)$ the group of algebraic characters, and by $\mathrm{X}_*(\mathrm{S}) = \mathrm{Hom}(\mathrm{GL}_1, \mathrm{S})$ the group of cocharacters, of S. We define a map

$$\mathrm{X}_*(\mathrm{S}) \times \mathrm{X}^*(\mathrm{S}) \to \mathbb{Z} \tag{2-1}$$

as follows. If $\lambda \in \mathrm{X}_*(\mathrm{S})$ and $\chi \in \mathrm{X}^*(\mathrm{S})$, then $\chi \circ \lambda$ is an endomorphism of the multiplicative group GL_1, which corresponds to an endomorphism of the ring $\mathbb{Z}[t, t^{-1}]$. It is of the form $t \mapsto t^n$ for some $n \in \mathbb{Z}$. This integer n is denoted by $\langle \lambda, \chi \rangle$. The map (2-1) defines a perfect duality [Borel 1991, § 8.6].

2.2. Let N and Z denote the normalizer and centralizer of S in G. If we extend the map (2-1) by $\mathbb{R}$-linearity, there exists a unique group homomorphism

$$\nu : \mathrm{Z}_k \to \mathrm{X}_*(\mathrm{S}) \otimes_{\mathbb{Z}} \mathbb{R} \tag{2-2}$$

such that the condition

$$\langle \nu(z), \chi \rangle = -\omega(\chi(z))$$

holds for any $z \in \mathrm{Z}_k$ and any k-rational character $\chi \in \mathrm{X}^*(\mathrm{Z})_k$ [Tits 1979, § 1.2]. According to [Landvogt 1995, Proposition 1.2], the kernel of (2-2) is the maximal compact subgroup of Z_k.

2.3. Let C denote the connected center of G and let $\mathrm{X}_*(\mathrm{C})$ be the group of its algebraic cocharacters. It is a subgroup of the free abelian group $\mathrm{X}_*(\mathrm{S})$. We denote by $\mathscr{A}$ the space

$$\mathrm{V} = (\mathrm{X}_*(\mathrm{S}) \otimes_{\mathbb{Z}} \mathbb{R})/(\mathrm{X}_*(\mathrm{C}) \otimes_{\mathbb{Z}} \mathbb{R}),$$

considered as an affine space on itself and by $\mathrm{Aff}(\mathscr{A})$ the group of its affine automorphisms. By making V act on $\mathscr{A}$ by translations, we can think of V as a subgroup of $\mathrm{Aff}(\mathscr{A})$. It is the kernel of the natural group homomorphism $\mathrm{Aff}(\mathscr{A}) \to \mathrm{GL}(\mathrm{V})$ which associates to any affine automorphism its linear part.

2.4. The map (2-2) induces a homomorphism

$$\mathrm{Z}_k \to \mathrm{Aff}(\mathscr{A}), \tag{2-3}$$

which we still denote by ν. Its image is contained in V. An important property of this homomorphism is that it extends to a homomorphism $\mathrm{N}_k \to \mathrm{Aff}(\mathscr{A})$ [Tits 1979, § 1.2]. It does not extend in a unique way, but two homomorphisms extending (2-3) to N_k are conjugated by a *unique* element of $\mathrm{Aff}(\mathscr{A})$ [Landvogt 1995, Proposition 1.8].

2.5. The affine space $\mathscr{A}$ endowed with an action of N_k defined by a group homomorphism $\nu : \mathrm{N}_k \to \mathrm{Aff}(\mathscr{A})$ extending the homomorphism (2-3) is called the (reduced) *apartment* attached to S. It satisfies these conditions:

A1. $\mathscr{A}$ is an affine space on V;

A2. ν is a group homomorphism $\mathrm{N}_k \to \mathrm{Aff}(\mathscr{A})$ extending the canonical homomorphism $\mathrm{Z}_k \to \mathrm{V}$.

It has the following uniqueness property: if $(\mathscr{A}', \nu')$ satisfies A1 and A2, there is a unique affine and N_k-equivariant isomorphism from $\mathscr{A}'$ to $\mathscr{A}$.

Remark 2.6. As in [Tits 1979], one obtains the *nonreduced* apartment $\mathscr{A}_{\mathrm{nr}}$ by replacing V by $\mathrm{X}_*(\mathrm{S}) \otimes_{\mathbb{Z}} \mathbb{R}$. It is not as canonical as the reduced one: two homomorphisms extending the map $\nu_{\mathrm{nr}} : \mathrm{Z}_k \to \mathrm{Aff}(\mathscr{A}_{\mathrm{nr}})$ to N_k are conjugated by an element of $\mathrm{Aff}(\mathscr{A}_{\mathrm{nr}})$ which is not necessarily unique [Landvogt 1995, Chapter 1, § 1; Tits 1979, § 1.2].

2.7. Let $\Phi = \Phi(\mathrm{G}, \mathrm{S})$ denote the set of roots of G relative to S. It is a subset of $\mathrm{X}^*(\mathrm{S})$. Therefore, any root $a \in \Phi$ can be seen as a linear form on $\mathrm{X}_*(\mathrm{S}) \otimes_{\mathbb{Z}} \mathbb{R}$ which is trivial on the subspace $\mathrm{X}_*(\mathrm{C}) \otimes_{\mathbb{Z}} \mathbb{R}$, hence as a linear form on V [Landvogt 1995, Chapter 1, § 1].

For $a \in \Phi$, we denote by U_a the root subgroup associated to a, which is a unipotent subgroup of G normalized by Z [Borel 1991, Proposition 21.9], and by s_a the reflection corresponding to a, considered as an element of GL(V) — or, more precisely, of the quotient of $\nu(\mathrm{N}_k)$ by $\nu(\mathrm{Z}_k)$.

2.8. Let $a \in \Phi$ and $u \in \mathrm{U}_a(k) - \{1\}$. The intersection

$$\mathrm{U}_{-a}(k) u \mathrm{U}_{-a}(k) \cap \mathrm{N}_k \tag{2-4}$$

consists of a single element, called $m(u)$, whose image by ν is an affine reflection the linear part of which is s_a [Borel and Tits 1965, § 5]. The set $\mathscr{H}_{a,u}$ of fixed points of $\nu(m(u))$ is an affine hyperplane of $\mathscr{A}$, which is called a *wall* of $\mathscr{A}$.

A *chamber* of $\mathscr{A}$ is a connected component of the complementary in $\mathscr{A}$ of the union of its walls. Note that a chamber is open in $\mathscr{A}$.

A point $x \in \mathscr{A}$ is said to be *special* if, for all root $a \in \Phi$, there is a root $b \in \Phi \cap \mathbb{R}_+ a$ and an element $u \in \mathrm{U}_b(k) - \{1\}$ such that $x \in \mathscr{H}_{b,u}$ [Landvogt 2000, § 1.2.3; Tits 1979, § 1.9].

2.9. Let $\theta(a, u)$ denote the affine function $\mathscr{A} \to \mathbb{R}$ whose linear part is a and whose vanishing hyperplane is the wall $\mathscr{H}_{a,u}$ of fixed points of $\nu(m(u))$. We fix a base point in $\mathscr{A}$, so that $\mathscr{A}$ can be identified with the vector space V. For $r \in \mathbb{R}$, we set

$$\mathrm{U}_a(k)_r = \left\{ u \in \mathrm{U}_a(k) - \{1\} \mid \theta(a, u)(x) \geqslant a(x) + r \text{ for all } x \in \mathscr{A} \right\} \cup \{1\}.$$

Thus we obtain a filtration of $\mathrm{U}_a(k)$ by subgroups. If we change the base point in $\mathscr{A}$, this filtration is only modified by a translation of the indexation.

2.10. Let Ω be a nonempty subset of $\mathscr{A}$. We set

$$\mathrm{N}_\Omega = \{n \in \mathrm{N}_k \mid \nu(n)(x) = x \text{ for all } x \in \Omega\},$$

and we denote by U_Ω the subgroup of G_k generated by all the $\mathrm{U}_a(k)_r$ such that the affine function $x \mapsto a(x) + r$ is nonnegative on Ω. According to [Landvogt 1995, § 12], this subgroup is compact in G_k, and we have $n\mathrm{U}_\Omega n^{-1} = \mathrm{U}_{\nu(n)(\Omega)}$ for $n \in \mathrm{N}_k$. In particular, N_Ω normalizes U_Ω. The subgroup $\mathrm{P}_\Omega = \mathrm{N}_\Omega\mathrm{U}_\Omega$ is open in G_k [Landvogt 1995, Corollary 12.12].

2.11. Let $\Phi = \Phi^- \cup \Phi^+$ be a decomposition of Φ into positive and negative roots. We denote by U^+ (U^-) the subgroup of G_k generated by the U_a for all $a \in \Phi^+$ ($a \in \Phi^-$). Then the group P_Ω has the following Iwahori decomposition [Landvogt 1995, Corollary 12.6; Bruhat and Tits 1972, § 7.1.4]:

$$\mathrm{P}_\Omega = (\mathrm{U}_\Omega \cap \mathrm{U}^-) \cdot (\mathrm{U}_\Omega \cap \mathrm{U}^+) \cdot \mathrm{N}_\Omega. \tag{2-5}$$

2.12. Bruhat and Tits [1972; 1984a] associate to the apartment $(\mathscr{A}, \nu)$ a G_k-set $\mathscr{B} = \mathscr{B}(\mathrm{G}, k)$ containing $\mathscr{A}$, called the (reduced) *building* of G over k and satisfying the following conditions:

B1. The set $\mathscr{B}$ is the union of the $g \cdot \mathscr{A}$ for $g \in \mathrm{G}_k$.

B2. The subgroup N_k is the stabilizer of $\mathscr{A}$ in G_k, and $n \cdot x = \nu(n)(x)$ for all $x \in \mathscr{A}$ and $n \in \mathrm{N}_k$.

B3. For all $a \in \Phi$ and $r \in \mathbb{R}$, the subgroup $\mathrm{U}_a(k)_r$ defined in 2.9 fixes the subset $\{x \in \mathscr{A} \mid a(x) + r \geqslant 0\}$ pointwise.

The building has the following uniqueness property: if $\mathscr{B}'$ is a G_k-set containing $\mathscr{A}$ and satisfying B1–B3, there is a unique G_k-equivariant bijection from $\mathscr{B}'$ to $\mathscr{B}$ [Tits 1979, § 2.1; Prasad and Yu 2002, § 1.9].

2.13. The subsets of $\mathscr{B}$ of the form $g \cdot \mathscr{A}$ with $g \in \mathrm{G}_k$ are called *apartments*. According to B1, the building is the union of its apartments. For $g \in \mathrm{G}_k$, the apartment $g \cdot \mathscr{A}$ can be naturally endowed with a structure of affine space and an action of ${}^g\mathrm{N}_k$ by affine isomorphisms. Up to unique isomorphism, it is the apartment attached to the maximal k-split torus ${}^g\mathrm{S}$ (see 2.5). This defines a unique G_k-equivariant map

$$\mathrm{S}' \mapsto \mathscr{A}(\mathrm{S}') \subseteq \mathscr{B} \tag{2-6}$$

between maximal k-split tori of G and apartments of $\mathscr{B}$, such that S maps to $\mathscr{A}$.

Note that the building $\mathscr{B}$ does not depend on the maximal k-split torus S. Indeed, let S′ be a maximal k-split torus of G, let $(\mathscr{A}', \nu')$ be the apartment attached to S′ and $\mathscr{B}'$ be the building of G over k relative to this apartment (see 2.12). If we identify $\mathscr{A}'$ with the unique apartment of $\mathscr{B}$ corresponding to S′ via (2-6), then $\mathscr{B}' = \mathscr{B}$.

2.14. The building has the following important properties [Bruhat and Tits 1972, § 7.4; Landvogt 1995, Chapter 4, § 13]:

(1) Let Ω be a nonempty subset of $\mathscr{A}$. Then P_Ω is the subgroup of G_k made of those elements fixing Ω pointwise.

(2) Let $g \in \mathrm{G}_k$. There is $n \in \mathrm{N}_k$ such that $g \cdot x = n \cdot x$ for any $x \in \mathscr{A} \cap g^{-1} \cdot \mathscr{A}$.

In particular, (1) together with B2 imply that $\mathrm{N}_\Omega = \mathrm{N}_k \cap \mathrm{P}_\Omega$.

2.15. Let σ be a k-automorphism of G. There is a unique bijective map from $\mathscr{B}$ to itself, still denoted σ, such that

(1) the condition
$$\sigma(g \cdot x) = \sigma(g) \cdot \sigma(x)$$
holds for any $g \in \mathrm{G}_k$ and $x \in \mathscr{B}$; and

(2) the map σ permutes the apartments and, for any apartment $\mathscr{A}$, the restriction of σ to $\mathscr{A}$ is an affine isomorphism from $\mathscr{A}$ to $\sigma(\mathscr{A})$.

This makes (2-6) into a σ-equivariant map. In particular, an apartment is σ-stable if and only if its corresponding maximal k-split torus of G is σ-stable [Bruhat and Tits 1984a, § 4.2.12].

3. Existence of σ-stable apartments

From now on, k will be a nonarchimedean locally compact field of odd residue characteristic. Let G be connected reductive group defined over k and let σ be a k-involution on G. According to 2.15, the building $\mathscr{B}$ of G over k is endowed with an action of σ. In this section, we prove that, given $x \in \mathscr{B}$, there exists a σ-stable apartment containing x. We keep using notation of Section 2.

3.1. Let Ω be a nonempty σ-stable subset of $\mathscr{B}$ contained in some apartment, and let $\mathrm{Ap}(\Omega)$ be the set of all apartments of $\mathscr{B}$ containing Ω. It is a nonempty set on which the group P_Ω acts transitively [Landvogt 1995, Corollary 13.7]. Because Ω is σ-stable, both P_Ω and $\mathrm{Ap}(\Omega)$ are σ-stable. Note that the σ-stable apartments containing Ω are exactly the σ-fixed points in $\mathrm{Ap}(\Omega)$.

3.2. Let us fix an apartment $\mathscr{A} \in \mathrm{Ap}(\Omega)$ and an element $u \in \mathrm{P}_\Omega$ such that $\sigma(\mathscr{A}) = u \cdot \mathscr{A}$. Let N denote the normalizer in G of the maximal k-split torus of G corresponding to $\mathscr{A}$. As σ is involutive, we have

$$\sigma(u)u \in \mathrm{P}_\Omega \cap \mathrm{N}_k = \mathrm{N}_\Omega. \tag{3-1}$$

The map $\rho : g \mapsto g \cdot \mathscr{A}$ induces a P_Ω-equivariant bijection between the homogeneous spaces $\mathrm{P}_\Omega/\mathrm{N}_\Omega$ and $\mathrm{Ap}(\Omega)$. The automorphism

$$\theta : x \mapsto u^{-1}\sigma(x)u$$

of the group G_k stabilizes P_Ω and N_Ω. Indeed $\sigma(\mathrm{N}_k) = u\mathrm{N}_k u^{-1}$, and

$$\theta(\mathrm{N}_\Omega) = u^{-1}\sigma(\mathrm{P}_\Omega \cap \mathrm{N}_k)u = \mathrm{P}_\Omega \cap u^{-1}\sigma(\mathrm{N}_k)u = \mathrm{N}_\Omega.$$

Note that the condition (3-1) implies that $\theta \circ \theta$ is conjugation by some element of N_Ω. As N_Ω is θ-stable, the map

$$(\sigma, g\mathrm{N}_\Omega) \mapsto u\theta(g\mathrm{N}_\Omega), \quad g \in \mathrm{P}_\Omega,$$

defines an action of σ on $\mathrm{P}_\Omega/\mathrm{N}_\Omega$, making ρ into a σ-equivariant bijection. Note that this action differs from the natural action of σ on $\mathrm{P}_\Omega/\mathrm{N}_\Omega$ (which obviously has fixed points).

3.3. Let Ω be a nonempty σ-stable subset of $\mathscr{B}$ contained in some apartment.

Proposition 3.4. *Assume that Ω contains a point of a chamber of $\mathscr{B}$. Then Ω is contained in some σ-stable apartment.*

Proof. We describe the quotient $\mathrm{P}_\Omega/\mathrm{N}_\Omega$ as a projective limit of finite σ-sets. According to [Cartier 1979, § 1.2], Example (f), the group G_k is locally compact and totally disconnected. Therefore we can choose a decreasing filtration $(\mathrm{Q}_i)_{i \geqslant 0}$ of the open subgroup P_Ω of G_k satisfying the following properties:

(A) The intersection of the Q_i is reduced to $\{1\}$.

(B) For any $i \geqslant 0$, the subgroup Q_i is compact open and normal in P_Ω.

Lemma 3.5. *Consider the decreasing filtration of P_Ω formed by the subgroups $\mathrm{P}_{\Omega,i} = \mathrm{N}_\Omega\mathrm{Q}_i \cap \theta(\mathrm{N}_\Omega\mathrm{Q}_i)$, for $i \geqslant 0$.*

(1) *The intersection of the $\mathrm{P}_{\Omega,i}$ is reduced to N_Ω.*

(2) *For any $i \geqslant 0$, the subgroup $\mathrm{P}_{\Omega,i}$ is θ-stable and of finite index in P_Ω.*

Proof. As N_Ω is θ-stable, it is contained in the intersection of the $\mathrm{P}_{\Omega,i}$. Let g be in this intersection. For any $i \geqslant 0$, there exist $n_i \in \mathrm{N}_\Omega$ and $q_i \in \mathrm{Q}_i$ such that $g = n_i q_i$. Because of (A) above, q_i converges to 1. Therefore n_i converges to a limit contained in the closed subgroup N_Ω, and this limit is g. This proves (1).

Now recall that $\theta\circ\theta$ is conjugation by some element of N_Ω. This implies that $\mathrm{P}_{\Omega,i}$ is θ-stable. As $\mathrm{P}_{\Omega,i}$ is open in P_Ω and contains N_Ω, the quotient $\mathrm{P}_\Omega/\mathrm{P}_{\Omega,i}$ can be identified with the quotient of U_Ω, which is compact, by some open subgroup. This gives (2). ☐

Because of Lemma 3.5(2), the map

$$(\sigma, g\mathrm{P}_{\Omega,i}) \mapsto u\theta(g\mathrm{P}_{\Omega,i}), \quad g\in\mathrm{P}_\Omega,$$

defines an action of σ on the finite quotient $\mathrm{P}_\Omega/\mathrm{P}_{\Omega,i}$, which gives us a projective system $(\mathrm{P}_\Omega/\mathrm{P}_{\Omega,i})_{i\geqslant 0}$ of finite σ-sets. Since P_Ω is complete, and thanks to Lemma 3.5(1), the natural σ-equivariant map from $\mathrm{P}_\Omega/\mathrm{N}_\Omega$ to the projective limit of the $\mathrm{P}_\Omega/\mathrm{P}_{\Omega,i}$ is bijective.

Lemma 3.6. *Let $(\mathrm{X}_i)_{i\geqslant 0}$ be a projective system of finite σ-sets. For all $i\geqslant 0$, assume the transition maps $\varphi_i : \mathrm{X}_{i+1}\to\mathrm{X}_i$ to be surjective and X_i to have odd cardinality. Then the projective limit X has a σ-fixed point.*

Proof. For each $i\geqslant 0$, the set X_i^σ of σ-fixed points of X_i is nonempty, since X_i has odd cardinality. This defines a projective system $(\mathrm{X}_i^\sigma)_{i\geqslant 0}$ whose transition maps may not be surjective. For each $i\geqslant 0$, let Y_i denote the intersection in X_i of the images of the X_{i+n}^σ, for $n\geqslant 0$. Then Y_i is nonempty, and the transition maps $\varphi_i : \mathrm{Y}_{i+1}\to\mathrm{Y}_i$ are surjective. Therefore, the projective limit $\mathrm{Y}=\mathrm{X}^\sigma\subseteq\mathrm{X}$ of the system $(\mathrm{Y}_i)_{i\geqslant 0}$ is nonempty. ☐

Let p denote the residue characteristic of k.

Lemma 3.7. *Let K be a normal subgroup of finite index in P_Ω containing N_Ω. Then the index of K in P_Ω is a power of p.*

Proof. Let S be the maximal k-split torus associated to $\mathscr{A}$, let Φ be the set of roots of G relative to S and let $\Phi=\Phi^-\cup\Phi^+$ be a decomposition of Φ into positive and negative roots. According to (2-5), the group P_Ω has the Iwahori decomposition

$$\mathrm{P}_\Omega = (\mathrm{U}_\Omega\cap\mathrm{U}^-)\cdot(\mathrm{U}_\Omega\cap\mathrm{U}^+)\cdot\mathrm{N}_\Omega.$$

That Ω contains a point of a chamber of $\mathscr{B}$ implies that the group N_Ω is reduced to $\mathrm{Ker}(\nu)$, hence normalizes the groups $\mathrm{V}^+=\mathrm{U}_\Omega\cap\mathrm{U}^+$ and $\mathrm{V}^-=\mathrm{U}_\Omega\cap\mathrm{U}^-$. The index of K in P_Ω can be decomposed as

$$(\mathrm{P}_\Omega:\mathrm{K}) = (\mathrm{P}_\Omega:\mathrm{V}^+\mathrm{K})\cdot(\mathrm{V}^+\mathrm{K}:\mathrm{K}).$$

On the one hand, the index

$$(\mathrm{V}^+\mathrm{K}:\mathrm{K}) = (\mathrm{V}^+:\mathrm{V}^+\cap\mathrm{K})$$

is a power of p, since V^+ is a pro-p-group. On the other hand, the index

$$(\mathrm{P}_\Omega:\mathrm{V}^+\mathrm{K}) = (\mathrm{V}^-:\mathrm{V}^-\cap\mathrm{V}^+\mathrm{K})$$

is a power of p, since V^- is a pro-p-group. The result follows. □

According to Lemma 3.7, the cardinality of each $\mathrm{P}_\Omega/\mathrm{P}_{\Omega,i}$, with $i \geqslant 0$, is odd (recall that p is different from 2). Proposition 3.4 follows from Lemma 3.6. □

We now prove the first main result of this section.

Proposition 3.8. *For any $x \in \mathscr{B}$, there exists a σ-stable apartment containing x.*

Proof. Let x be a point in $\mathscr{B}$, and let y be a point of a chamber of $\mathscr{B}$ whose closure contains x. The set $\Omega = \{y, \sigma(y)\}$ is a σ-stable subset of $\mathscr{B}$ satisfying the conditions of Proposition 3.4. Hence we get a σ-stable apartment of $\mathscr{B}$ containing y. Such an apartment contains the closure of the chamber of y. In particular, it contains x. □

3.9. Let S be a σ-stable maximal k-split torus, and let N and Z denote the normalizer and centralizer of S in G. Let $\mathrm{X} = \mathrm{X}(\mathrm{S})$ denote the set of all $g \in \mathrm{G}_k$ such that $g^{-1}\sigma(g) \in \mathrm{N}_k$, let $\mathscr{A}$ denote the σ-stable apartment corresponding to S and, given $x \in \mathscr{A}$, let P_x denote the subgroup P_Ω (see 2.11) with $\Omega = \{x\}$.

Proposition 3.10. X *is a finite union of* $(\mathrm{H}_k, \mathrm{Z}_k)$*-double cosets and* $\mathrm{G}_k = \mathrm{XP}_x$.

Proof. Let us fix a minimal parabolic k-subgroup P of G containing the torus S. According to Helminck and Wang [1993, Proposition 6.8], the map $g \mapsto \mathrm{H}_k g \mathrm{P}_k$ induces a bijection between the $(\mathrm{H}_k, \mathrm{Z}_k)$-double cosets in X and the $(\mathrm{H}_k, \mathrm{P}_k)$-double cosets in G_k. The first part of the proposition then follows from [Helminck and Wang 1993, Corollary 6.16].

Note that we have $g \in \mathrm{X}$ if and only if $g \cdot \mathscr{A}$ is σ-stable. For $g \in \mathrm{G}_k$, we set $x' = g \cdot x$. According to Proposition 3.8, there is a σ-stable apartment $\mathscr{A}'$ containing x'. Let $g' \in \mathrm{X}$ be such that $\mathscr{A}' = g' \cdot \mathscr{A}$. According to Property (2) in 2.14, there is $n \in \mathrm{N}_k$ such that we have $g'^{-1} g \cdot x = n \cdot x$. Hence we get $g \in \mathrm{XN}_k\mathrm{P}_x$. As $\mathrm{XN}_k = \mathrm{X}$, we obtain the expected result. □

4. Decomposition of $\mathrm{H}_k\backslash\mathrm{G}_k$

In all this section, we assume that G is k-split. Let H be an open k-subgroup of the fixed points group G^σ. Equivalently, H is a k-subgroup of G^σ containing $(\mathrm{G}^\sigma)^\circ$.

4.1. If T is a σ-stable torus in G, we write T^+ for the neutral component of $\mathrm{T} \cap \mathrm{H}$ and T^- for the neutral component of the subgroup $\{t \in \mathrm{T} \mid \sigma(t) = t^{-1}\}$. The torus T is the almost direct product of T^+ and T^-, that is $\mathrm{T} = \mathrm{T}^+\mathrm{T}^-$ and the intersection $\mathrm{T}^+ \cap \mathrm{T}^-$ is finite [Borel 1991, xi].

Definition 4.2 [Helminck and Wang 1993, § 4.4]. A σ-stable torus T of G is said to be (σ, k)*-split* if it is k-split and if $\mathrm{T} = \mathrm{T}^-$.

By Proposition 10.3 of the same reference, two arbitrary maximal (σ, k)-split tori of G are G_k-conjugated.

4.3. Let $\mathscr{D}$G denote the derived subgroup of G, and recall that C denotes the connected center of G. This latter subgroup is a k-split torus of G.

Lemma 4.4. *Let* T *be a k-split torus of* G.

(1) *There is a k-subtorus* T' *of* C *such that the groups* $\mathrm{T}\cdot\mathscr{D}\mathrm{G}$ *and* $\mathrm{T}'\cdot\mathscr{D}\mathrm{G}$ *are equal.*

(2) *If* T *is* (σ, k)*-split, any* T' *satisfying* (1) *is* (σ, k)*-split.*

(3) *Assume that* $\mathscr{D}$G *is contained in* H *and* T *is* (σ, k)*-split. Then any* T' *satisfying* (1) *is* (σ, k)*-split and has the same dimension as* T.

Proof. We set $\tilde{\mathrm{G}} = \mathrm{G}/\mathscr{D}\mathrm{G}$ and, for any k-subgroup K of G, we write $\tilde{\mathrm{K}}$ for the image of K in $\tilde{\mathrm{G}}$. According to [Borel 1991, Proposition 14.2], the group G is the almost direct product of C and $\mathscr{D}$G, which means that G is equal to the product $\mathrm{C}\cdot\mathscr{D}\mathrm{G}$ and that the intersection $\mathrm{C}\cap\mathscr{D}\mathrm{G}$ is finite. This implies that $\tilde{\mathrm{C}} = \tilde{\mathrm{G}}$. Let f denote the k-rational map $\mathrm{C} \to \tilde{\mathrm{C}}$. It is surjective with finite kernel. Hence $\tilde{\mathrm{G}}$ is a k-split torus, and we denote by $\tilde{\sigma}$ the involutive k-automorphism of $\tilde{\mathrm{G}}$ induced by σ. We now prove each conclusion claim in the lemma.

(1) By [Borel 1991, Proposition 8.2(c)], the neutral component of the inverse image $f^{-1}(\tilde{\mathrm{T}})$ is a k-split subtorus of C which we denote by T'. It has finite index in $f^{-1}(\tilde{\mathrm{T}})$. The image $f(\mathrm{T}')$ is then a subtorus of finite index in the connected group $\tilde{\mathrm{T}}$, so that $\tilde{\mathrm{T}}' = \tilde{\mathrm{T}}$.

(2) Assume that T is (σ, k)-split, and let T' satisfy (1). Let us consider the map $t \mapsto t\sigma(t)$ from T' to itself. As $\tilde{\mathrm{T}}' = \tilde{\mathrm{T}}$ is a $(\tilde{\sigma}, k)$-split torus, the image of this map is a connected k-subgroup contained in the kernel of f, which is finite.

(3) Assume that $\mathscr{D}$G is contained in H and T is (σ, k)-split. Then the map $\mathrm{T} \to \tilde{\mathrm{T}}$ has finite kernel, which implies that T and $\tilde{\mathrm{T}}$ have the same dimension. Now let T' satisfy (1). According to (2), such a torus is (σ, k)-split, and it has the same dimension as $\tilde{\mathrm{T}}' = \tilde{\mathrm{T}}$. □

4.5. Let S be a σ-stable maximal (k-split) torus of G, let $\mathscr{A}$ be the apartment corresponding to S and let Φ be the set of roots of G relative to S. Let $x \in \mathscr{A}$ be a special point (see 2.8), and write U_x for U_Ω (see 2.11) with $\Omega = \{x\}$. Let $a \in \Phi$ be a σ-invariant root, which means that $a \circ \sigma = a$.

Lemma 4.6. *Assume that* $\mathrm{U}_{-a}(k)$ *is contained in* $\{g \in \mathrm{G}_k \mid \sigma(g) = g^{-1}\}$. *Then there are* $n \in \mathrm{N}_k$ *and* $c \in \mathrm{U}_x$ *such that* $n = c^{-1}\sigma(c)$ *and* $\nu(n)$ *is the affine reflection of* $\mathscr{A}$ *which let* x *invariant and whose linear part is* s_a.

Proof. We fix a base point in the apartment $\mathscr{A}$, so that it can be identified with the vector space V. For any $b \in \Phi$, this defines a filtration of the group $\mathrm{U}_b(k)$ (see 2.9). For $u \in \mathrm{U}_b(k) - \{1\}$, we denote by $\varphi_b(u)$ the greatest real number $r \in \mathbb{R}$ such that $u \in \mathrm{U}_b(k)_r$. Let us choose $w \in \mathrm{U}_{-a}(k) - \{1\}$ such that x is contained in the wall $\mathscr{H}_{-a,w}$. Thus $\nu(m(w))$ is the affine reflection of $\mathscr{A}$ which fixes x and whose linear part is s_a, and we can set

$$n = m(w) \in \mathrm{N}_k.$$

Moreover $\theta(-a, w)$, which is the unique affine function from $\mathscr{A}$ to $\mathbb{R}$ whose linear part is $-a$ and whose vanishing hyperplane is $\mathscr{H}_{-a,w}$, vanishes on x. Therefore it is equal to

$$y \mapsto -a(y) + a(x),$$

which implies that $\varphi_{-a}(w) = a(x)$. According to B3 (see 2.12), it follows that w fixes x.

The group $\mathrm{U}_{-a}(k)$ is isomorphic to the additive group of k. Thus, for $r \in \mathbb{R}$, the subgroup $\mathrm{U}_{-a}(k)_r$ corresponds through this isomorphism to a nontrivial sub-$\mathcal{O}$-module of k, where $\mathcal{O}$ denotes the ring of integers of k [Landvogt 1995, Proposition 7.7]. Therefore, there is a unique element $v \in \mathrm{U}_{-a}(k)$ such that $w = v^2$ and $\varphi_{-a}(v) = \varphi_{-a}(w)$, hence $v \in \mathrm{U}_x$.

The map $\mathrm{U}_a(k) \times \mathrm{U}_a(k) \to \mathrm{G}_k$ defined by $(u, u') \mapsto uwu'$ is injective and the intersection given by (2-4) consists of a single element, which is n. If we choose $u, u' \in \mathrm{U}_a(k)$ such that $uwu' = n$, then the element

$$\sigma(u')^{-1} w \sigma(u)^{-1} = \sigma(n)^{-1}$$

is contained in the intersection (2-4). Hence $\sigma(n)^{-1}$ is equal to n, and the uniqueness property implies that $u' = \sigma(u)^{-1}$. Moreover, according to [Landvogt 1995, Lemma 7.4(ii)], the real numbers $\varphi_a(u)$ and $\varphi_a(\sigma(u))$ are both equal to $-\varphi_{-a}(w)$. This implies that u and $\sigma(u)$ are contained in U_x. Since v is σ-antiinvariant and $w = v^2$, we get the expected result by choosing $c = (uv)^{-1}$. □

Remark 4.7. Note that $\sigma(c) \in \mathrm{U}_x$. Indeed we have $\sigma(v) = v^{-1} \in \mathrm{U}_x$ and $\sigma(u) \in \mathrm{U}_x$. Hence $n = c^{-1}\sigma(c) \in \mathrm{N}_k \cap \mathrm{U}_\Omega$, which is contained in N_Ω with $\Omega = \{x, \sigma(x)\}$.

Let $\mathscr{B}$ denote the building of G over k.

Proposition 4.8. *Let x be a special point of $\mathscr{B}$. There is a σ-stable maximal k-split torus* S *of* G *such that the apartment corresponding to* S *contains x and such that* S^- *is a maximal (σ, k)-split torus of* G.

Remark 4.9. In 5.13, we give an example of a *nonsplit* k-group G such that Proposition 4.8 does not hold.

Proof. Let $\mathscr{A}$ be a σ-stable apartment containing x (see Proposition 3.8) and let S be the corresponding maximal k-split torus of G. Assume that $\mathscr{A}$ has been chosen such that the dimension of the (σ, k)-split torus S^- is maximal. If it is a maximal (σ, k)-split torus of G, then Proposition 4.8 is proved. Assume that this is not the case, and let A be a maximal (σ, k)-split torus of G containing S^-. The dimension of A is greater than $\dim \mathrm{S}^-$ (if not, the containment $\mathrm{S}^- \subseteq \mathrm{A}$ would imply that $\mathrm{S}^- = \mathrm{A}$). Let G' be the neutral component of the centralizer of S^- in G. It is a k-split connected reductive subgroup of G containing S and A, which is naturally endowed with a nontrivial action of σ. Let C' denote the connected center of G'.

Lemma 4.10. *There is* $a \in \Phi(\mathrm{G}', \mathrm{S})$ *such that the corresponding root subgroup* U'_a *is not contained in* H, *and such a root is* σ*-invariant.*

Proof. Assume that $\mathrm{U}'_a \subseteq \mathrm{H}$ for each root $a \in \Phi(\mathrm{G}', \mathrm{S})$. Then the derived subgroup $\mathscr{D}\mathrm{G}'$, which is generated by the U'_a for $a \in \Phi(\mathrm{G}', \mathrm{S})$, is contained in H [Humphreys 1975, Theorem 27.5(e)]. According to Lemma 4.4(iii), there exists a (σ, k)-subtorus A' of C' such that $\mathrm{A} \cdot \mathscr{D}\mathrm{G}' = \mathrm{A}' \cdot \mathscr{D}\mathrm{G}'$ and $\dim(\mathrm{A}) = \dim(\mathrm{A}')$. The subgroup generated by C' and S is a k-torus of G'. As G' is k-split, S is a maximal torus of G', hence it contains C'. Therefore S^- contains A' which has the same dimension as A, and this dimension is greater than $\dim \mathrm{S}^-$. This gives us a contradiction.

Now let a be a root in $\Phi(\mathrm{G}', \mathrm{S})$ such that U'_a is not contained in H. The root a and its conjugate $a \circ \sigma$ coincide on S^+ and are both trivial on S^-. As S is the almost direct product of S^+ and S^- (see 4.1), they are equal. Therefore a is σ-invariant. This ends the proof of Lemma 4.10. □

Let $a \in \Phi(\mathrm{G}', \mathrm{S})$ as in Lemma 4.10. If we think of a as a root in $\Phi(\mathrm{G}, \mathrm{S})$, then U_a is σ-stable and is not contained in H. Moreover:

Lemma 4.11. $\mathrm{U}_a(k)$ *is contained in* $\{g \in \mathrm{G}_k \mid \sigma(g) = g^{-1}\}$.

Proof. As G is k-split, U_a is k-isomorphic to the additive group. Thus the action of σ on $\mathrm{U}_a(k)$ corresponds to an involutive automorphism of the k-algebra $k[t]$. It has the form $t \mapsto \lambda t$ for some $\lambda \in k^\times$ with $\lambda^2 = 1$. As U_a is not contained in H, we have $\lambda = -1$. This gives us the expected result. □

According to Lemma 4.6, there are $n \in \mathrm{N}_k$ and $c \in \mathrm{U}_x$ such that $n = c^{-1}\sigma(c)$ and $\nu(n)$ is the affine reflection of $\mathscr{A}$ which let x invariant and whose linear part is s_a. For any $t \in \mathrm{S}$, note that

$$\sigma(ctc^{-1}) = cn\sigma(t)n^{-1}c^{-1} = cs_a(\sigma(t))c^{-1}.$$

Let $\mathscr{A}'$ denote the apartment $c \cdot \mathscr{A}$ and let $\mathrm{S}' = {}^c\mathrm{S}$ be the corresponding maximal k-split torus of G. Then $\mathscr{A}'$ contains x and is σ-stable. Moreover, since the root a is trivial on S^- and s_a fixes the kernel of a pointwise, the conjugate ${}^c(\mathrm{S}^-)$ is a (σ, k)-split subtorus of S'. Thus S'^- has dimension not smaller than $\dim \mathrm{S}^-$.

Now let S_a denote the maximal k-split torus in the set of all $t \in \mathrm{S}$ such that $s_a(t) = t^{-1}$. Since a is σ-invariant, such a torus is σ-stable. It is also one-dimensional and its intersection with $\mathrm{Ker}(a)$ is finite. Therefore ${}^c\mathrm{S}_a$ is a nontrivial (σ, k)-split subtorus of S' which is not contained in ${}^c(\mathrm{S}^-)$. Thus the dimension of S'^-, which contains ${}^c(\mathrm{S}_a\mathrm{S}^-)$, is greater than $\dim \mathrm{S}^-$, which contradicts the maximality property of $\mathscr{A}$. This ends the proof of Proposition 4.8. □

4.12. Let A be a maximal (σ, k)-split torus of G, let S be a σ-stable maximal k-split torus of G containing A and let $\mathscr{A}$ denote the apartment corresponding to S. Let $\{\mathrm{A}^j \mid j \in \mathrm{J}\}$ be a set of representatives of the H_k-conjugacy classes of maximal (σ, k)-split tori in G. According to [Helminck and Wang 1993], the set J is finite. Let $x \in \mathscr{A}$ be a special point and write K for its stabilizer in G_k.

Theorem 4.13. *For $j \in \mathrm{J}$, let $y_j \in \mathrm{G}_k$ such that ${}^{y_j}\mathrm{A} = \mathrm{A}^j$. We have*

$$\mathrm{G}_k = \bigcup_{j \in \mathrm{J}} \mathrm{H}_k y_j \mathrm{S}_k \mathrm{K}.$$

Proof. By Proposition 4.8, for any $g \in \mathrm{G}_k$, there is a σ-stable maximal k-split torus S' of G such that the apartment corresponding to it contains $g \cdot x$ and such that S'^- is a maximal (σ, k)-split torus of G. Let $j \in \mathrm{J}$ be such that S'^- is H_k-conjugate to A^j. According to Helminck and Helminck [1998, Lemma 2.2], there is $h \in \mathrm{H}_k$ such that $\mathrm{S}' = {}^{hy_j}\mathrm{S}$. Hence $g \cdot x$ is contained in $hy_j \cdot \mathscr{A}$. According to Property (2) in 2.14, there exists $n \in \mathrm{N}_k$ such that $g \cdot x = hy_jn \cdot x$. Therefore G_k is the union of the $\mathrm{H}_k y_j \mathrm{N}_k \mathrm{K}$ for $j \in \mathrm{J}$. As x is special, we have $\mathrm{N}_k\mathrm{K} = \mathrm{S}_k\mathrm{K}$ and we get the expected result. □

4.14. In the case where G is not necessarily k-split, we have the following result. For each j, let $\mathrm{W}_{\mathrm{G}_k}(\mathrm{A}^j)$ be the quotient of the normalizer of A^j in G_k by its centralizer, and likewise with G_k replaced by H_k. According to [Helminck and Wang 1993], the group $\mathrm{W}_{\mathrm{G}_k}(\mathrm{A}^j)$ is the Weyl group of a root system. For $j \in \mathrm{J}$, let $\mathcal{N}_j \subseteq \mathrm{N}_{\mathrm{G}_k}(\mathrm{A}^j)$ be a set of representatives of

$$\mathrm{W}_{\mathrm{H}_k}(\mathrm{A}^j)\backslash \mathrm{W}_{\mathrm{G}_k}(\mathrm{A}^j),$$

and let $y_j \in \mathrm{G}_k$ be such that ${}^{y_j}\mathrm{A} = \mathrm{A}^j$. Let P be a minimal parabolic k-subgroup of G containing S and such that $\mathrm{P} \cap \sigma(\mathrm{P})$ is a Levi component of P [Helminck and Wang 1993, § 4]. Let ϖ be a uniformizer of k, and write Λ for the lattice made of the images of ϖ by the various algebraic cocharacters of A and Λ^- for the subset of antidominant elements of Λ relative to P. Then one can derive from Proposition 3.10 the existence of a compact subset Q of G_k such that

$$G_k = \bigcup_{j \in J} \bigcup_{n \in \mathcal{N}_j} H_k n y_j \Lambda^- Q. \tag{4-1}$$

Benoist and Oh [2007] have obtained a similar decomposition of G_k, with a weaker condition on the base field k (they assume k to have odd characteristic).

Remark 4.15. In the split case, starting from Theorem 4.13, one can obtain a sharper result than the decomposition (4-1).

Let us mention that the question of the disjointness of the various components appearing in the decomposition (4-1) has been investigated in [Lagier 2008].

5. Examples

Let k be a nonarchimedean locally compact field of odd residue characteristic. Let $\mathcal{O}$ be its ring of integers and $\mathfrak{p}$ be the maximal ideal of $\mathcal{O}$.

5.1. We now consider the k-split reductive group $G = GL_n$, $n \geqslant 1$, endowed with the k-involution $\sigma : g \mapsto {}^t g^{-1}$, where ${}^t g$ denotes the transpose of g. We set $K = GL_n(\mathcal{O})$ and $H = G^\sigma$, and write S for the diagonal torus of G. This case has been explicitly investigated by Hironaka [1988] from a different point of view.

We start with the following lemma.

Lemma 5.2. *Let* V *be a finite dimensional k-vector space and* B *a symmetric bilinear form on* V. *Then any free $\mathcal{O}$-submodule of finite rank of* V *has a basis which is orthogonal relative to* B.

Proof. Let Λ be a free $\mathcal{O}$-submodule of finite rank of V. The proof goes by induction on the rank of Λ. If B is null, then the result is trivial. If not, we denote by B_Λ the restriction of B to $\Lambda \times \Lambda$. Its image is of the form $\mathfrak{p}^m$ for some integer $m \in \mathbb{Z}$. If ϖ is a uniformizer of k, then the form $B^0_\Lambda = \varpi^{-m} B_\Lambda$ has image $\mathcal{O}$ on $\Lambda \times \Lambda$. Therefore, it defines a nontrivial bilinear form

$$\bar{B}^0_\Lambda : \Lambda/\mathfrak{p}\Lambda \times \Lambda/\mathfrak{p}\Lambda \to \mathcal{O}/\mathfrak{p}.$$

Let $e \in \Lambda$ be a vector whose reduction modulo $\mathfrak{p}$ is not isotropic relative to $\bar{B}^0_\Lambda$, which means that $B^0_\Lambda(e, e)$ is a unit of $\mathcal{O}$. Then Λ is the direct sum of $\mathcal{O}e$ and $\Lambda \cap ke^\perp$, where $ke^\perp$ denotes the orthogonal of ke in V. Indeed, it follows from the decomposition

$$x = \frac{B(e, x)}{B(e, e)} e + \Big(x - \frac{B(e, x)}{B(e, e)} e \Big), \quad \text{for any } x \in \Lambda.$$

As $\Lambda \cap ke^{\perp}$ is a free $\mathcal{O}$-submodule of finite rank of V whose rank is smaller than the rank of Λ, we conclude by induction. □

We introduce the set Y of all $g \in \mathrm{G}_k$ such that ${}^t gg \in \mathrm{S}_k$. Using Lemma 5.2, we get the following decomposition of G_k.

Proposition 5.3. *We have* $\mathrm{G}_k = \mathrm{YK}$.

Proof. We make G_k act on the quotient G_k/K, which can be identified to the set of all $\mathcal{O}$-lattices (that is, cocompact free $\mathcal{O}$-submodules) of the k-vector space $\mathrm{V} = k^n$. Let B denote the symmetric bilinear form on V making the canonical basis of V into an orthonormal basis. According to Lemma 5.2, for any $g \in \mathrm{G}_k$, the $\mathcal{O}$-lattice Λ corresponding to the class $g\mathrm{K}$ has a basis which is orthogonal relative to B. This means that there exists $u \in \mathrm{K}$ such that the element $g' = gu^{-1} \in g\mathrm{K}$ maps the canonical basis of V to an orthogonal basis of Λ. Therefore we have $g' \in \mathrm{Y}$; thus $g \in \mathrm{YK}$. □

We now investigate the maximal (σ, k)-split tori of G. Note that S is a maximal (σ, k)-split torus of G.

Proposition 5.4. *The map* $g \mapsto {}^g\mathrm{S}$ *induces a bijection between* $(\mathrm{H}_k, \mathrm{N}_k)$*-double cosets of* Y *and* H_k*-conjugacy classes of maximal* (σ, k)*-split tori of* G.

Proof. One easily checks that this map is well defined and injective. For $g \in \mathrm{G}_k$, the conjugate ${}^g\mathrm{S}$ is a maximal (σ, k)-split torus of G if and only if $g^{-1}\sigma(g) \in \mathrm{S}_k$, which amounts to saying that $g \in \mathrm{Y}$ and proves surjectivity. □

Let $\mathcal{Q}$ denote the set of all equivalence classes of nondegenerate quadratic forms on k^n. For $a = \mathrm{diag}(a_1, \dots, a_n) \in \mathrm{S}_k$ we denote by Q_a the diagonal quadratic form $a_1\mathrm{X}_1^2 + \cdots + a_n\mathrm{X}_n^2$. Note that the map $a \mapsto \mathrm{Q}_a$ induces a surjective map from S_k to $\mathcal{Q}$.

We write H^0 and H^1 for the set of σ-fixed points and the first set of nonabelian cohomology of σ, respectively.

Proposition 5.5. (1) *The map* $g \mapsto {}^t gg$ *induces an injection* ι *from the set of* $(\mathrm{H}_k, \mathrm{N}_k)$*-double cosets of* Y *to* $\mathrm{H}^1(\mathrm{N}_k)$.

(2) *Given* $a \in \mathrm{S}_k$*, the class of* a *in* $\mathrm{H}^1(\mathrm{N}_k)$ *is in the image of* ι *if and only if* $\mathrm{Q}_a \sim \mathrm{X}_1^2 + \cdots + \mathrm{X}_n^2$.

Proof. We have an exact sequence

$$\mathrm{H}_k \to \mathrm{H}^0(\mathrm{G}_k/\mathrm{N}_k) \to \mathrm{H}^1(\mathrm{N}_k) \to \mathrm{H}^1(\mathrm{G}_k),$$

where the map from $\mathrm{H}^0(\mathrm{G}_k/\mathrm{N}_k)$ to $\mathrm{H}^1(\mathrm{N}_k)$ is induced by $g \mapsto {}^t gg$. As the set of $(\mathrm{H}_k, \mathrm{N}_k)$-double cosets of Y is a subset of $\mathrm{H}_k\backslash\mathrm{H}^0(\mathrm{G}_k/\mathrm{N}_k)$, we get the first assertion. To obtain the second one, it is enough to remark that $\mathrm{H}^1(\mathrm{G}_k)$ canonically identifies with $\mathcal{Q}$. □

Remark 5.6. Recall from [Serre 1970, IV.2.3] that for $a, b \in \mathrm{S}_k$, the quadratic forms Q_a, Q_b are equivalent if and only if they have the same discriminant and the same Hasse invariant.

Proposition 5.7. *Let* $\{a^j \mid j \in \mathrm{J}\} \subseteq \mathrm{S}_k$ *form a set of representatives of* $\mathrm{Im}(\iota)$ *in* $\mathrm{H}^1(\mathrm{N}_k)$. *For* $j \in \mathrm{J}$, *we choose* $y_j \in \mathrm{Y}$ *such that* ${}^t y_j y_j = a^j$. *Then,*

$$\mathrm{G}_k = \bigcup_{j \in \mathrm{J}} \mathrm{H}_k y_j \mathrm{S}_k \mathrm{K}.$$

Proof. Propositions 5.3 and 5.4 imply that G_k is the union of the components $\mathrm{H}_k y_j \mathrm{N}_k \mathrm{K}$ for $j \in \mathrm{J}$. As $\mathrm{N}_k\mathrm{K} = \mathrm{S}_k\mathrm{K}$, we get the expected result. □

Example 5.8. In the case where $n = 2$, we give an explicit description of $\mathrm{Im}(\iota)$. Let ϖ denote a uniformizer of $\mathcal{O}$ and $\xi \in \mathcal{O}^\times$ a nonsquare unit of $\mathcal{O}$, so that $\{1, \xi, \varpi, \xi\varpi\}$ is a set of representatives of $k^\times$ modulo $k^{\times 2}$. The set of elements of $k^\times$ which are represented by the quadratic form $\mathrm{Q}_1 = \mathrm{X}^2 + \mathrm{Y}^2$ depends on the image of p in $\mathbb{Z}/4\mathbb{Z}$. If $p \equiv 1 \bmod 4$, all elements of $k^\times$ are represented by Q_1. If $p \equiv 3 \bmod 4$, an element of $k^\times$ is represented by Q_1 if and only if its normalized valuation if even. We set

$$\mathrm{J} = \begin{cases} \{1, \xi, \varpi, \xi\varpi\} & \text{if } p \equiv 1 \bmod 4, \\ \{1, \xi\} & \text{if } p \equiv 3 \bmod 4. \end{cases}$$

For each $j \in \mathrm{J}$, set $a^j = \mathrm{diag}(j, j)$. Then the elements a^j form a set of representatives of $\mathrm{Im}(\iota)$ in $\mathrm{H}^1(\mathrm{N}_k)$.

5.9. We now consider the connected reductive k-group $\mathrm{G} = \mathrm{Res}_{k'/k}\mathrm{GL}_n$, where k' is a quadratic extension of k, endowed with the involutive k-automorphism σ of G induced by the nontrivial element of $\mathrm{Gal}(k'/k)$. This case has been explicitly investigated by Offen [2004] when k'/k is unramified.

We set $\mathrm{H} = \mathrm{G}^\sigma$, so that we have $\mathrm{G}_k = \mathrm{GL}_n(k')$ and $\mathrm{H}_k = \mathrm{GL}_n(k)$. We denote by S the diagonal torus of G and by K the maximal compact subgroup $\mathrm{GL}_n(\mathcal{O}')$ of G_k, where $\mathcal{O}'$ denotes the ring of integers of k'. Note that S is σ-invariant.

As usual, N and Z denote the normalizer and centralizer of S in G. Let $\mathfrak{S}_n$ denote the group of permutation matrices in G_k, so that N_k is the semidirect product of $\mathfrak{S}_n$ by Z_k. Note that S_k (resp. Z_k) is the subgroup of all diagonal matrices of G_k with entries in k (resp. in k').

Lemma 5.10. $\mathrm{H}^1(\mathrm{N}_k)$ *can be identified with the set of conjugacy classes of elements of* $\mathfrak{S}_n$ *of order* 1 *or* 2.

Proof. According to Hilbert's Theorem 90, the group $\mathrm{H}^1(\mathrm{Z}_k)$ is trivial. Therefore we have an exact sequence

$$1 \to \mathrm{H}^1(\mathrm{N}_k) \to \mathrm{H}^1(\mathrm{N}_k/\mathrm{Z}_k). \tag{5-1}$$

As σ acts trivially on $\mathrm{N}_k/\mathrm{Z}_k \simeq \mathfrak{S}_n$, the set $\mathrm{H}^1(\mathrm{N}_k/\mathrm{Z}_k)$ can be identified to the set of $\mathfrak{S}_n$-conjugacy classes of $\mathrm{Hom}(\mathbb{Z}/2\mathbb{Z}, \mathfrak{S}_n)$, that is, to the set of conjugacy classes of elements of $\mathfrak{S}_n$ of order 1 or 2. This proves that $\mathrm{H}^1(\mathrm{N}_k)$ can be naturally embedded in the set of conjugacy classes of elements of $\mathfrak{S}_n$ of order $\leqslant 2$.

Now two elements $w, w' \in \mathfrak{S}_n$ define the same class in $\mathrm{H}^1(\mathrm{N}_k)$ if and only if they are conjugate in $\mathfrak{S}_n$, thus if and only if $w\mathrm{Z}_k$ and $w'\mathrm{Z}_k$ define the same class in $\mathrm{H}^1(\mathrm{N}_k/\mathrm{Z}_k)$. Therefore (5-1) is a bijection. □

Proposition 5.11. (1) *The number of* H_k*-conjugacy classes of* σ*-stable maximal* k*-split tori in* G_k *is* $[n/2]+1$.

(2) *There is a unique* H_k*-conjugacy class of maximal* (σ, k)*-split tori in* G_k.

Proof. (1) Let X denote the set of all $g \in \mathrm{G}_k$ such that $g^{-1}\sigma(g) \in \mathrm{N}_k$. Then the map $g \mapsto {}^g\mathrm{S}$ defines an injective map from the set of $(\mathrm{H}_k, \mathrm{N}_k)$-double cosets of X to $\mathrm{H}^1(\mathrm{N}_k)$. Therefore we are reduced to proving that this map is surjective, and the first assertion will follow from Lemma 5.10. For $n=2$, let τ denote the nontrivial element of $\mathfrak{S}_2$ and choose an element $a \in k'$ which is not in k. Then the element

$$u = \begin{pmatrix} a & \sigma(a) \\ 1 & 1 \end{pmatrix} \in \mathrm{GL}_2(k') \tag{5-2}$$

satisfies the relation $u^{-1}\sigma(u) = \tau$. For an arbitrary integer $n \geqslant 2$, let $w \in \mathfrak{S}_n$ have order $\leqslant 2$. Then there is an integer $0 \leqslant i \leqslant [n/2]$ such that w is conjugate to the element

$$\tau_i = \mathrm{diag}(\tau, \dots, \tau, 1, \dots, 1) \in \mathrm{GL}_n(k'),$$

where $\tau \in \mathrm{GL}_2(k')$ appears i times and $1 \in \mathrm{GL}_1(k')$ appears $n-2i$ times. Thus

$$u_i = \mathrm{diag}(u, \dots, u, 1, \dots, 1) \in \mathrm{GL}_n(k') \tag{5-3}$$

satisfies the relation $u_i^{-1}\sigma(u_i) = \tau_i$. Therefore any 1-cocycle in N_k is G_k-cohomologous to the neutral element $1 \in \mathrm{G}_k$, which proves the first assertion.

(2) For any $0 \leqslant i \leqslant [n/2]$, the dimension of the (σ, k)-split torus $({}^{u_i}\mathrm{S})^-$ is equal to i. According to (1), the map

$$\mathrm{H}_k g \mathrm{N}_k \mapsto \text{class of } g^{-1}\sigma(g) \text{ in } \mathrm{H}^1(\mathrm{N}_k)$$

is a bijection from the set of $(\mathrm{H}_k, \mathrm{N}_k)$-double cosets of X to $\mathrm{H}^1(\mathrm{N}_k)$, and the elements of this latter set are the classes of the τ_i for $0 \leqslant i \leqslant [n/2]$. This gives us the expected result. □

Proposition 5.12. *For* $0 \leqslant i \leqslant [n/2]$, *let* u_i *denote the element defined by* (5-2) *and* (5-3). *Then*

$$\mathrm{G}_k = \bigcup_{i=0}^{[n/2]} \mathrm{H}_k u_i \mathrm{Z}_k \mathrm{K}.$$

Proof. According to the proof of Proposition 5.11, the set X is the union of the double cosets $\mathrm{H}_k u_i \mathrm{N}_k$ with $0 \leqslant i \leqslant [n/2]$. The result then follows from Proposition 3.10 and from the fact that $\mathrm{N}_k\mathrm{K} = \mathrm{Z}_k\mathrm{K}$. □

5.13. We now give an example (due to Bertrand Lemaire) of a nonsplit k-group such that Proposition 4.8 does not hold. We set $\mathrm{G} = \mathrm{Res}_{k'/k}\mathrm{GL}_2$, where k' is now a *ramified* quadratic extension of k. The k-involution σ is still induced by the nontrivial element of $\mathrm{Gal}(k'/k)$ and we set $\mathrm{H} = \mathrm{GL}_2$. Let $\mathcal{B}'$ (resp. $\mathcal{B}$) denote the building of G (resp. H) over k.

Bruhat and Tits [1984b] give a description of the faces of $\mathcal{B}$ in terms of hereditary $\mathcal{O}$-orders of $\mathrm{M}_2(k)$. More precisely, there is a bijective correspondence

$$\mathrm{F} \mapsto \mathcal{M}_\mathrm{F}$$

between the faces of $\mathcal{B}$ and the hereditary $\mathcal{O}$-orders of $\mathrm{M}_2(k)$, such that the stabilizer of F in $\mathrm{GL}_2(k)$ in the normalizer of $\mathcal{M}_\mathrm{F}$ in $\mathrm{GL}_2(k)$. For $x \in \mathcal{B}$, we will denote by $\mathcal{M}_x$ the hereditary order corresponding to the face of $\mathcal{B}$ which contains x. We have a similar correspondence between faces of $\mathcal{B}'$ and hereditary $\mathcal{O}'$-orders of $\mathrm{M}_2(k')$. Moreover, since k' is tamely ramified over k, there is a bijective correspondence j from the set $\mathcal{B}'^\sigma$ of σ-fixed points of $\mathcal{B}'$ to $\mathcal{B}$ such that, for any $x \in \mathcal{B}'^\sigma$, we have

$$\mathcal{M}_{j(x)} = \mathcal{M}_x \cap \mathrm{M}_2(k).$$

Let q denote the cardinality of the residue field of k. As k' is totally ramified over k, any vertex of $\mathcal{B}$ has exactly $q+1$ neighbors in $\mathcal{B}$, and likewise for $\mathcal{B}'$. Let x be a σ-invariant point of $\mathcal{B}'$. Recall that, according to Proposition 3.8, it is contained in a σ-stable apartment.

- If $j(x)$ is in a chamber of $\mathcal{B}$, then x has $q+1$ neighbors in $\mathcal{B}'$ but only two σ-fixed ones. Thus x has non-σ-fixed neighbors.
- If $j(x)$ is a vertex of $\mathcal{B}$, then x has $q+1$ neighbors in $\mathcal{B}'$ as in $\mathcal{B}$. Therefore any neighbor of x in $\mathcal{B}'$ is σ-invariant, which implies that any σ-stable apartment containing x is σ-invariant. For instance, this is the case of the vertex x corresponding to the $\mathcal{O}'$-order $\mathrm{M}_2(\mathcal{O}')$, as its image $j(x)$ corresponds to the maximal $\mathcal{O}$-order $\mathrm{M}_2(\mathcal{O}') \cap \mathrm{M}_2(k) = \mathrm{M}_2(\mathcal{O})$. For such a special point, Proposition 4.8 does not hold.

Acknowledgements

We thank F. Courtès, B. Lemaire, G. Rousseau, and S. Stevens for stimulating discussions. Particular thanks go to J. Bernstein for having suggested to the first author the use of the Bruhat–Tits building, and to J.-P. Labesse for answering numerous questions.

References

[Benoist and Oh 2007] Y. Benoist and H. Oh, "Polar decomposition for p-adic symmetric spaces", *Int. Math. Res. Not.* **2007**:24 (2007), article ID rnm121. MR 2009a:22006

[Blanc and Delorme 2008] P. Blanc and P. Delorme, "Vecteurs distributions H-invariants de représentations induites, pour un espace symétrique réductif p-adique G/H", *Ann. Inst. Fourier* (*Grenoble*) **58**:1 (2008), 213–261. MR 2009e:22015 Zbl 1151.22012

[Borel 1991] A. Borel, *Linear algebraic groups*, 2nd ed., Grad. Texts in Math. **126**, Springer, New York, 1991. MR 92d:20001 Zbl 0726.20030

[Borel and Tits 1965] A. Borel and J. Tits, "Groupes réductifs", *Inst. Hautes Études Sci. Publ. Math.* **27** (1965), 55–150. MR 34 #7527 Zbl 0145.17402

[Bruhat and Tits 1972] F. Bruhat and J. Tits, "Groupes réductifs sur un corps local", *Inst. Hautes Études Sci. Publ. Math.* **41** (1972), 5–251. MR 48 #6265 Zbl 0254.14017

[Bruhat and Tits 1984a] F. Bruhat and J. Tits, "Groupes réductifs sur un corps local. II. Schémas en groupes. Existence d'une donnée radicielle valuée", *Inst. Hautes Études Sci. Publ. Math.* **60** (1984), 197–376.

[Bruhat and Tits 1984b] F. Bruhat and J. Tits, "Schémas en groupes et immeubles des groupes classiques sur un corps local", *Bull. Soc. Math. France* **112**:2 (1984), 259–301. MR 86i:20064 Zbl 0565.14028

[Cartier 1979] P. Cartier, "Representations of p-adic groups: a survey", pp. 111–155 in *Automorphic forms, representations and L-functions, I* (Corvallis, OR, 1977), edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, Amer. Math. Soc., Providence, R.I., 1979. MR 81e:22029 Zbl 0421.22010

[Flensted-Jensen 1978] M. Flensted-Jensen, "Spherical functions of a real semisimple Lie group. A method of reduction to the complex case", *J. Funct. Anal.* **30**:1 (1978), 106–146. MR 80f:43022 Zbl 0419.22019

[Helminck 1994] A. G. Helminck, "Symmetric k-varieties", pp. 233–279 in *Algebraic groups and their generalizations: classical methods* (University Park, PA, 1991), edited by W. J. Haboush and B. J. Parshall, Proc. Sympos. Pure Math. **56**, Amer. Math. Soc., Providence, RI, 1994. MR 1278710 Zbl 0819.20048

[Helminck and Helminck 1998] A. G. Helminck and G. F. Helminck, "A class of parabolic k-subgroups associated with symmetric k-varieties", *Trans. Amer. Math. Soc.* **350**:11 (1998), 4669–4691. MR 99g:20082 Zbl 0912.20041

[Helminck and Wang 1993] A. G. Helminck and S. P. Wang, "On rationality properties of involutions of reductive groups", *Adv. Math.* **99**:1 (1993), 26–96. MR 94d:20051 Zbl 0788.22022

[Hironaka 1988] Y. Hironaka, "Spherical functions of Hermitian and symmetric forms, I", *Japan. J. Math.* (*N.S.*) **14**:1 (1988), 203–223. MR 90c:11027 Zbl 0674.43006

[Humphreys 1975] J. E. Humphreys, *Linear algebraic groups*, Grad. Texts in Math. **21**, Springer, New York, 1975. MR 53 #633 Zbl 0325.20039

[Kato and Takano 2008] S.-i. Kato and K. Takano, "Subrepresentation theorem for p-adic symmetric spaces", *Int. Math. Res. Not.* **2008**:11 (2008), Art. ID rnn028. MR 2009i:22021

[Lagier 2008] N. Lagier, "Terme constant de fonctions sur un espace symétrique réductif p-adique", *J. Funct. Anal.* **254**:4 (2008), 1088–1145. MR 2009d:22013 Zbl 1194.22010

[Landvogt 1995] E. Landvogt, *A compactification of the Bruhat–Tits building*, Lecture Notes in Math. **1619**, Springer, Berlin, 1995. MR 98h:20081 Zbl 0935.20034

[Landvogt 2000] E. Landvogt, "Some functorial properties of the Bruhat–Tits building", *J. Reine Angew. Math.* **518** (2000), 213–241. MR 2001g:20029 Zbl 0937.20026

[Offen 2004] O. Offen, "Relative spherical functions on $\wp$-adic symmetric spaces (three cases)", *Pacific J. Math.* **215**:1 (2004), 97–149. MR 2005f:11103

[Prasad and Yu 2002] G. Prasad and J.-K. Yu, "On finite group actions on reductive groups and buildings", *Invent. Math.* **147**:3 (2002), 545–560. MR 2003e:20036 Zbl 1020.22003

[Schneider and Stuhler 1997] P. Schneider and U. Stuhler, "Representation theory and sheaves on the Bruhat–Tits building", *Inst. Hautes Études Sci. Publ. Math.* **85** (1997), 97–191. MR 98m:22023 Zbl 0892.22012

[Serre 1970] J.-P. Serre, *Cours d'arithmétique*, Collection SUP: "Le Mathématicien" **2**, Presses Universitaires de France, Paris, 1970. MR 41 #138 Zbl 0225.12002

[Tits 1979] J. Tits, "Reductive groups over local fields", pp. 29–69 in *Automorphic forms, representations and L-functions, I* (Corvallis, OR, 1977), edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, Amer. Math. Soc., Providence, R.I., 1979. MR 80h:20064 Zbl 0415.20035

Received January 18, 2010.

PATRICK DELORME
UNIVERSITÉ DE LA MEDITERRANÉE
INSTITUT DE MATHÉMATIQUES DE LUMINY, UMR 6206
CAMPUS DE LUMINY, CASE 907
13288 MARSEILLE, CEDEX 9
FRANCE

delorme@iml.univ-mrs.fr
http://iml.univ-mrs.fr/fiche/Patrick_Delorme.html

VINCENT SÉCHERRE
UNIVERSITÉ DE VERSAILLES SAINT-QUENTIN
LABORATOIRE DE MATHÉMATIQUES DE VERSAILLES
45, AVENUE DES ÉTATS-UNIS
78035 VERSAILLES CEDEX
FRANCE

secherre@math.uvsq.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 1, 2011

UNITAL QUADRATIC QUASI-JORDAN ALGEBRAS

RAÚL FELIPE

Forty-six years ago, McCrimmon defined the notion of a unital quadratic Jordan algebra. Here we introduce and study the notion of a unital quadratic quasi-Jordan algebra, following earlier work by Loday, Velasquez and the author.

1. Introduction

In the past century, among nonassociative systems, Jordan algebras and unital quadratic Jordan algebras have occupied a very special place. For instance, Jordan algebras occur in quantum mechanics in connection with the representation of physical observables from an algebraic point of view.

It is well known that an associative algebra A gives rise to a Jordan algebra A^+ via the Jordan product $x \circ y = \frac{1}{2}(xy + yx)$; it also gives rise to a Lie algebra by means of the product $[x, y] = xy - yx$. A Jordan algebra is called *special* if it is isomorphic to a subalgebra of a Jordan algebra A^+ for some associative algebra A; otherwise it is *exceptional*. A major problem in the theory of Jordan algebra has been, from the beginning, the classification of simple Jordan algebras. Its solution began with the works of Jordan, von Neumann, Wigner and Albert around 1934 for finite-dimensional algebras and was concluded with Zelmanov's outstanding work in the general case [Albert 1934; Jordan et al. 1934; Zelmanov 1979; 1983].

Jordan algebras also play an important role in others areas of mathematics, such as differential geometry (exceptional algebras; see for instance [Bertram 2000]), and the analysis of nonconvex optimization problems over symmetric cones (specifically, Euclidean Jordan algebras; see [Faybusovich 1997] for more details).

Unital quadratic Jordan algebras were introduced by McCrimmon [1966; 1978] in order to understanding Jordan structures where there is no scalar $\frac{1}{2}$, which necessitate a quadratic approach based in the product xyx instead of $x \circ y = \frac{1}{2}(xy + yx)$. McCrimmon developed this concept to introduce uniform methods in the study

---

This research was supported in part under CONACYT grant no. 106923 and in part under the Cuba National Research Project *Theory and efficient algorithms for the solution of problems in operator algebra and geometric modeling*.
MSC2000: primary 17C50; secondary 16W10.
Keywords: Dialgebras, Leibniz algebras, quasi-Jordan algebras.

of Jordan algebras over characteristic 2. In a strict sense, unital quadratic Jordan algebras are not algebras, because they do not have a bilinear product; however, their connection to Jordan algebras motivated this terminology.

More recently, Loday [1993; 2001] discovered interesting generalizations of associative and Lie algebras, which are now well known as *dialgebras* and *Leibniz algebras*. All this leads in a natural way to the question of finding a similar analogue for Jordan algebras, and study the unital quadratic Jordan algebras associated to these new structures. With this purpose, we introduced in [Velásquez and Felipe 2008] the notion of quasi-Jordan algebras.

More specifically, a Leibniz algebra is a generalization of a Lie algebra where the skew-symmetry of the bracket is dropped and the Jacobi identity is changed by the Leibniz identity. Loday observed that the relationship between Lie algebras and associative algebras translate into an analogous relationship between Leibniz algebras and so-called dialgebras, which are a generalization of associative algebras possessing two products: Namely, a *dialgebra* over a field K is a K-vector space D equipped with two associative products

$$\dashv\, : D \times D \to D, \quad \vdash\, : D \times D \to D$$

satisfying the identities

$$x \dashv (y \dashv z) = x \dashv (y \vdash z), \tag{1}$$

$$(x \vdash y) \dashv z = x \vdash (y \dashv z), \tag{2}$$

$$(x \vdash y) \vdash z = (x \dashv y) \vdash z. \tag{3}$$

We say that $e \in D$ is a bar unit of D if for all $x \in D$ we have $e \vdash x = x = x \dashv e$.

Loday showed that any dialgebra $(D, \vdash, \dashv)$ becomes a Leibniz algebra under the Leibniz bracket $[x, y] = x \dashv y - y \vdash x$.

Our notion of quasi-Jordan algebra bears to Leibniz algebras a relationship similar to the one between Jordan algebras and Lie algebras. More precisely, in [Velásquez and Felipe 2008] we attached a quasi-Jordan algebra QJ_x to any Q-Jordan element x in a Leibniz algebra. Soon, Kolesnikov [2008] and Bremner [2010] (see also [Bremner and Peresi 2010]) found independently an interesting particular case of quasi-Jordan algebras, in which the analysis of its derivations has a promising future (see [Felipe 2009]). We observe that in a dialgebra over a field of characteristic other than 2 the Jordan quasiproduct takes the form

$$x \lhd y := \tfrac{1}{2}(x \dashv y + y \vdash x). \tag{4}$$

In other words, any dialgebra over a field of characteristic other than 2 leads to a quasi-Jordan algebra.

In this paper we generalize the notion of unital quadratic Jordan algebras, beginning with dialgebras. As we will see, one arrives to a new structure (the unital quadratic quasi-Jordan algebra) which include the notion introduced by McCrimmon in 1966.

2. Definitions and basic examples

Definition 1 [McCrimmon 2004, page 83]. A unital quadratic Jordan algebra J consists of a Φ-module on which a product $U_x y$ is defined which is linear in y and quadratic in x (i.e., $U : x \mapsto U_x$ is a mapping of J into $\mathrm{End}_\Phi(J)$, homogeneous of degree 2), together with a choice of a unit element e, such that the following operator identities hold, where we have defined

$$V_{x,y}z = (U_{x+z} - U_x - U_z)y \tag{5}$$

for all $x, y, z \in J$:

(a) $U_e = \mathrm{Id}$.

(b) $V_{x,y}U_x = U_x V_{y,x}$.

(c) $U_{U_x y} = U_x U_y U_x$.

Any associative algebra A determines a quadratic Jordan algebra QA^+ with the product $U_x y = xyx$.

In his original paper, McCrimmon [1966] included in the definition of unital quadratic Jordan algebras the condition that the identities (b) and (c) remain valid under extensions of the ring of scalars, and pointed out that this condition is equivalent to the assumption that the linearizations of the identities hold. He subsequently eliminated this requirement [1978; 2004]. We return to this point in Section 3.

Definition 2. A *unital quadratic quasi-Jordan algebra* over a field K is a quadruple $(\mathfrak{J}, U, W, e)$, where $\mathfrak{J}$ is a K-vector space, e is a distinguished element of $\mathfrak{J}$, and U and W are maps $a \mapsto U_a$ and $a \mapsto W_a$ of $\mathfrak{J}$ into $\mathrm{End}_K(\mathfrak{J})$ satisfying the following axioms:

(QQJ1) $U_e = \mathrm{Id}$ and $W_e e = e$.

(QQJ2) $W_z U_x V_{y,x} = W_z V_{x,y} U_x$ for all $x, y, z \subset \mathfrak{J}$, in the notation of (5).

(QQJ3) $U_{U_x y} = U_x U_y U_x$, for every $x, y \subset \mathfrak{J}$.

(QQJ4) $U_{\lambda x} e = \lambda^2 U_x e$ for any $x \in \mathfrak{J}$.

We say that e is the *unit* of the unital quadratic quasi-Jordan algebra.

The need for a second operator W arises as follows. We wish to include split quasi-Jordan algebras (where the product $\triangleleft$ is right commutative) among unital quadratic quasi-Jordan algebras. But in general, it is not true that $U_x V_{y,x} = V_{x,y} U_x$

for unital quadratic quasi-Jordan algebras, as will become clear after Lemma 4. The operator W is responsible, so to speak, for ensuring that $U_{(\,.\,)}$ and $V_{(\,.\,,\,.\,)}$ "commute" (QQJ2). Moreover, we want to be able to construct quasi-Jordan algebras from unital quadratic quasi-Jordan algebras (Section 4).

Lemma 3. *Any unital quadratic Jordan algebra is a unital quadratic quasi-Jordan algebra in which $W_a = U_a$ for all $a \in \Im$. In this case U_x is K-quadratic with respect to x.*

Proof. This is immediately checked from the definitions. □

The real motivation for Definition 2 is the following lemma.

Lemma 4. *Let $(D, \vdash, \dashv, e)$ be a unital K-dialgebra. We need not suppose that the field K is of characteristic other than 2. Define*

$$U_x y = (x \vdash y) \dashv x = x \vdash (y \dashv x), \qquad W_x y = (x \dashv y) \dashv x = x \dashv (y \dashv x).$$

Then (D, U, W, e) is a unital quadratic quasi-Jordan algebra, for which U and W are homogeneous of degree 2 (as maps $D \to \mathrm{End}_K(D)$).

The unital quadratic quasi-Jordan algebra built from a unital dialgebra D will be denoted by $(QQ(D), e)$.

Proof. It is clear that $U_e x = x$ for all $x \in D$. Next, $W_e e = (e \dashv e) \dashv e = e$. The homogeneity condition — that is, $U_{\lambda x} y = \lambda^2 U_x y$ and $W_{\lambda x} y = \lambda^2 W_x y$ for any $x, y \in \Im$ and any scalar λ — is also easy to check.

To show that QQJ3 holds, we write

$$\begin{aligned}
U_{U_x y} z = U_{(x \vdash y) \dashv x} z &= \big(((x \vdash y) \dashv x) \vdash z\big) \dashv ((x \vdash y) \dashv x) \\
&= ((x \vdash y) \dashv x) \vdash \big(z \dashv ((x \vdash y) \dashv x)\big) \\
&= ((x \vdash y) \dashv x) \vdash \big(z \dashv (x \vdash (y \dashv x))\big) \\
&= ((x \vdash y) \dashv x) \vdash \big(z \dashv (x \dashv (y \dashv x))\big) \\
&= ((x \vdash y) \vdash x) \vdash \big(z \dashv (x \dashv (y \dashv x))\big) \\
&= (x \vdash y) \vdash \big(x \vdash (z \dashv (x \dashv (y \dashv x)))\big) \\
&= (x \vdash y) \vdash \big((U_x z) \dashv (y \dashv x)\big) \\
&= U_x U_y U_x z.
\end{aligned}$$

To simplify the rest of the proof we introduce some notation. If $a_1, a_2, \ldots, a_n$ are elements of D and $1 \le k \le n$, we set

$$\begin{aligned}
&a_1 a_2 \ldots a_{k-1} \widehat{a_k} a_{k+1} \ldots a_{n-1} a_n \\
&\qquad = (a_1 \vdash a_2 \vdash \cdots \vdash a_{k-2} \vdash a_{k-1}) \vdash a_k \dashv (a_{k+1} \dashv a_{k+2} \dashv \cdots \dashv a_{n-1} \dashv a_n),
\end{aligned}$$

where the right-hand side is well defined by associativity.

Next we verify the axiom QQJ2. We have

$$\begin{aligned}
W_cU_xV_{y,x}z &= W_cU_x[(y\vdash x)\dashv z+(z\vdash x)\dashv y]\\
&= W_c\big[\big(x\vdash((y\vdash x)\dashv z)\big)\dashv x+\big(x\vdash((z\vdash x)\dashv y)\big)\dashv x\big]\\
&= \big(c\dashv((x\vdash((y\vdash x)\dashv z))\dashv x)\big)\dashv c+\big(c\dashv((x\vdash((z\vdash x)\dashv y))\dashv x)\big)\dashv c\\
&= \widehat{c}xyxzxc+\widehat{c}xzxyxc;
\end{aligned}$$

on the other hand

$$\begin{aligned}
W_cV_{x,y}U_xz &= W_cV_{x,y}((x\vdash z)\dashv x)\\
&= W_c\big[\big((x\vdash y)\dashv((x\vdash z)\dashv x)\big)+\big(((((x\vdash z)\dashv x)\vdash y)\dashv x\big)\big]\\
&= (c\dashv((x\vdash y)\dashv((x\vdash z)\dashv x)))\dashv c+\big(c\dashv((((x\vdash z)\dashv x)\vdash y)\dashv x)\big)\dashv c\\
&= \widehat{c}xyxzxc+\widehat{c}xzxyxc.
\end{aligned}$$

Thus, QQJ2 follows. Finally that U_x and W_x belong to $\mathrm{End}_K(D)$ for any $x\in D$ is evident. □

It is not hard to see that $U_xV_{y,x}$ and $V_{x,y}U_x$ need not coincide for unital quadratic quasi-Jordan algebras. In fact, from the proof of Lemma 4 it follows that

$$U_xV_{y,x}z=\big(x\vdash((y\vdash x)\dashv z)\big)\dashv x+\big(x\vdash((z\vdash x)\dashv y)\big)\dashv x, \tag{6}$$

$$V_{x,y}U_xz=\big((x\vdash y)\dashv((x\vdash z)\dashv x)\big)+\big((((x\vdash z)\dashv x)\vdash y)\dashv x\big). \tag{7}$$

Taking $x=e$, one obtains from (6) that $U_eV_{y,e}z=y\vdash e\dashv z+z\vdash e\dashv y$, and from (7) that $V_{e,y}U_ez=y\dashv z+z\vdash y$. Thus, for nonzero $y\in Z_B(D)$ we have $U_eV_{y,e}e=0$, but $V_{e,y}U_ee=2y$, which is nonzero if the characteristic is not 2.

3. Linearization

We now turn to the "linearization interpretation" of the axioms in Definition 2. We restrict ourselves to the case of unital quadratic quasi-Jordan algebras $(QQ(D),e)$.

Recall that in the proof of Lemma 4 we used the equality

$$V_{x,y}z=(x\vdash y)\dashv z+(z\vdash y)\dashv x.$$

Recall also that $U_xy=(x\vdash y)\dashv x$. If we replace x by $x+\alpha z$ in this latter equality, we obtain

$$U_{x+\alpha z}y=U_xy+(V_{x,y}z)\alpha+(U_zy)\alpha^2;$$

that is, we can consider to $V_{x,y}$ as the "linearization" of U_x, which justifies its presence in axiom QQJ2.

One can see, after a cumbersome calculation, that if the field of scalars over which a unital quadratic quasi-Jordan algebra $(QQ(D),e)$ is defined has at least four elements, the linearization of QQJ2 is

$$W_v(U_x V_{y,w} z + V_{w,V_{y,x}z} x) = W_v(V_{w,y} U_x z + V_{x,y} V_{w,z} x) \quad \text{for } v, x, y, z, w \in D. \tag{8}$$

If the field of scalars has at least five elements, linearizing QQJ3 we obtain

$$U_x U_y V_{x,w} z + V_{w,U_y U_x z} x = V_{V_{w,y}x,z} U_x y, \tag{9}$$

for all $x, y, z, w \in D$. Thus, if D is a dialgebra with a bar unit defined over a field with at least five elements, the axioms QQJ2 and QQJ3 for $(QQ(D), e)$ can be linearized in the form (8) and (9) respectively.

4. Relation to quasi-Jordan algebras

Let $(D, \vdash, \dashv, e)$ be a unital dialgebra. The unital quadratic quasi-Jordan algebra $(QQ(D), e)$ is *restrictive*, that is, it satisfies the condition

$$V_{z,(V_{(V_{y,y}e),x}e)}e - V_{(V_{z,(V_{y,y}e)}e),x}e = 2V_{y,(V_{z,(V_{y,x}e)}e - V_{(V_{z,y}e),x}e)}e, \tag{10}$$

for all $x, y, z \in \Im$. Indeed, (10) is the Bremner–Kolesnikov identity for the quasi-Jordan product defined from dialgebras [Bremner 2010; Felipe 2009; Kolesnikov 2008].

It is well known that any unital Jordan algebra $(J, \bullet, e)$ over a field of characteristic other than 2 gives rise to a unital quadratic Jordan algebra (and so also a unital quadratic quasi-Jordan algebra) for which, if $R_x y$ denotes the product of y by x,

$$U_x y = (2R_x^2 - R_{x^2})y \quad \text{and} \quad x \bullet y = \tfrac{1}{2}(U_{x+y} - U_x - U_y)e = K_{x,y}e.$$

At the same time, Bremner [2010] has shown that the Bremner–Kolesnikov identity holds in Jordan algebras. Hence, we have

$$K_{(K_{a,(K_{b,b}e)}e),c}e - K_{a,(K_{(K_{b,b}e),c}e)}e = 2K_{(K_{(K_{a,b}e),c}e - K_{a,(K_{b,c}e)}e),b}e,$$

for all $a, b, c \in J$. Since $V_{x,y}$ and $K_{x,y}$ act differently on a element, this last equality is distinct from (10). This is not surprising, because in general the quasi-Jordan algebra arising from a dialgebra is not a Jordan algebra.

We know that by means of the right and left products of a K-dialgebra over a field K of characteristic other than 2, we can build a new product on the same underlying vector space (see below after the next definition) with respect to which it becomes a *quasi-Jordan algebra* (in fact, this new product is right commutative). See [Velásquez and Felipe 2008; 2009] for details.

Definition 5. A *quasi-Jordan algebra* is a vector space $\Im$ over a field K of a characteristic other than 2 equipped with a bilinear product $\lhd : \Im \times \Im \to \Im$ such that

$$x \lhd (y \lhd z) = x \lhd (z \lhd y) \quad \text{(right commutativity)} \tag{11}$$

and

$$(y \lhd x) \lhd x^2 = (y \lhd x^2) \lhd x \quad \text{(right Jordan identity)} \tag{12}$$

for all $x, y, z \in \Im$, where $x^2 = x \lhd x$. A *unit* of a quasi-Jordan algebra $\Im$ is an element $e \in \Im$ such that $x \lhd e = x$ for all $x \in \Im$.

Example 6. As noted earlier, quasi-Jordan algebras appear in the study of the product

$$x \lhd y := \tfrac{1}{2}(x \dashv y + y \vdash x), \tag{13}$$

where x and y are elements in a dialgebra $(D, \vdash, \dashv)$ over a field K of characteristic other than 2. The quasi-Jordan algebra defined over D with the product (13) is denoted by $(\Im(D), \lhd)$.

From the results above we see that if D has a bar unit e, our construction defines over D a unital quadratic quasi-Jordan algebra $(QQ(D), e)$. In this case we have:

Lemma 7. *For any $x \in QQ(D)$, the linear transformation U_x can be recovered as*

$$U_x y = (2R_x^2 - R_{x^2})y,$$

where R_x is right multiplication by x (that is, the element of $\operatorname{End}(\Im(D))$ *defined by $R_x y = y \lhd x$). The product $\lhd$ in $((\Im(D), \lhd), e)$ is recovered as $y \lhd x = \frac{1}{2}V_{x,y}e$.*

Proof. We prove the first statement; the proof of the equality $y \lhd x = \frac{1}{2}V_{x,y}e$ is similar. In fact,

$$\begin{aligned}
&(2R_x^2 - R_{x^2})y \\
&\quad = 2(y \lhd x) \lhd x - y \lhd (x \lhd x) \\
&\quad = \tfrac{1}{2}\big((y \dashv x + x \vdash y) \dashv x + x \vdash (y \dashv x + x \vdash y)\big) - \tfrac{1}{4}\big(y \dashv (x \dashv x + x \vdash x) + (x \dashv x + x \vdash x) \vdash y\big) \\
&\quad = (x \vdash y) \dashv x = U_x y. \qquad \square
\end{aligned}$$

For a quasi-Jordan algebra $\Im$ we introduce

$$Z^r(\Im) = \{z \in \Im : x \lhd z = 0 \text{ for all } x \in \Im\}.$$

We denote by $\Im^{\text{ann}}$ the subspace of $\Im$ spanned by elements of the form $x \lhd y - y \lhd x$, with $x, y \in \Im$, and call it *the annihilator ideal* of the quasi-Jordan algebra $\Im$. Then $\Im$ is a Jordan algebra if and only if $\Im^{\text{ann}} = \{0\}$. It follows from right commutativity (11) that in any quasi-Jordan algebra

$$x \lhd (y \lhd z - z \lhd y) = 0.$$

The last identity implies that $\Im^{\text{ann}} \subset Z^r(\Im)$. One can prove that both $\Im^{\text{ann}}$ and $Z^r(\Im)$ are two-sided ideals of $\Im$. Now recall from [Velásquez and Felipe 2008] that if $\Im$ is a unital quasi-Jordan algebra, with a specific unit e, then

$$\Im^{\text{ann}} = Z^r(\Im), \quad \Im^{\text{ann}} = \{x \in \Im : e \lhd x = 0\}. \tag{14}$$

It is now clear that units in quasi-Jordan algebras are not unique; indeed, the set of units $U_r(\Im)$ of $\Im$ is given by

$$U_r(\Im) = \{x + e : x \in \Im^{\text{ann}}\}.$$

Definition 8. Let $\Im$ be a quasi-Jordan algebra and let I be an ideal in $\Im$ such that $\Im^{\text{ann}} \subset I \subset Z^r(\Im)$. We say that $\Im$ is *split* over I if there is a subalgebra J of $\Im$ such that $\Im = I \oplus J$ as a direct sum of subspaces.

Clearly, if $\Im$ is split over an ideal I with complement J, then J is a Jordan algebra with respect to the product $\lhd$ restricted to J. This is equivalent to saying that $(J, \lhd|_J)$ is a Jordan algebra. In fact, for $x, y \in J$, then $x \lhd y,\ y \lhd x \in J$ and $x \lhd y - y \lhd x \in I \cap J = \{0\}$; that is, $\lhd|_J$ is commutative and therefore the right Jordan identity over $\Im$ implies that $(J, \lhd|_J)$ is a Jordan algebra.

Additionally, for $a, b \in I$ and $x, y \in J$ we have

$$(a + x) \lhd (b + y) = a \lhd y + x \lhd y,$$

because $I \subset Z^r(\Im)$.

Reciprocally, let $(J, \bullet)$ be a Jordan algebra and let M be a Jordan bimodule over J. We consider the direct sum $\Im := M \oplus J$ and we define the product $\lhd$ over $\Im$ by

$$(a + x) \lhd (b + y) = ay + x \bullet y,$$

for all $a, b \in M$ and $x, y \in J$. Then $(\Im, \lhd)$ is a quasi-Jordan algebra, called the *demisemidirect product* of M with J.

It is possible to see that $\Im^{\text{ann}} \cong MJ$ and

$$Z^r(\Im) = M \oplus \{y \in Z(J) : uy = 0 \text{ for all } u \in M\},$$

where $Z(J) = \{y \in J : x \bullet y = 0 \text{ for all } x \in J\}$. Finally, $M \cong M \oplus \{0\}$ is an ideal of $\Im$ such that $\Im^{\text{ann}} \subset M \subset Z^r(\Im)$. In addition, $\Im/M \cong J$ and $\Im$ is split over M with complement J.

Let $(\Im, \bullet)$ be an algebra. Assume that $\Im = I \oplus J$, where $(J, \bullet)$ is a Jordan algebra and I is an ideal of $\Im$. In general I is not a Jordan bimodule over J with respect to the product $\bullet$. However, we can define a new product on $\Im$ by

$$(a + x) \lhd (b + y) = a \bullet y + x \bullet y, \tag{15}$$

for all $a, b \in I$ and $x, y \in J$.

Lemma 9. *Let $(\Im, \bullet)$ be an algebra such that $\Im = I \oplus J$, where $(J, \bullet)$ is a Jordan algebra and I is an ideal of $\Im$. Suppose that $(a \bullet x^2) \bullet x = (a \bullet x) \bullet x^2$ for all $a \in I$*

and $x \in J$, *where* $x^2 = x \bullet x$. *Then* $(\mathfrak{I}, \lhd)$ *is a quasi-Jordan algebra, where* $\lhd$ *is the product defined by* (15). *Moreover* $\mathfrak{I}^{\text{ann}} \subset I \subset Z^r(\mathfrak{I})$.

We refer to $(\mathfrak{I}, \lhd)$ *as the demisemidirect product of* I *with* J.

Proof. The product (15) is right commutative; in fact, if $a, b, c \in I$ and $x, y, z \in J$,

$$\begin{aligned}(a+x) \lhd ((b+y) \lhd (c+z)) &= a \bullet (y \bullet z) + x \bullet (y \bullet z)\\ &= a \bullet (z \bullet y) + x \bullet (z \bullet y)\\ &= (a+x) \lhd ((c+z) \lhd (b+y)).\end{aligned}$$

Observe that $(a+x) \lhd (a+x) = a \bullet x + x^2$. Now

$$\begin{aligned}((b+y) \lhd (a+x)) \lhd (a \bullet x + x^2) &= (b \bullet x) \bullet x^2 + (y \bullet x) \bullet x^2\\ &= (b \bullet x^2) \bullet x + (y \bullet x^2) \bullet x\\ &= ((b+y) \lhd (a \bullet x + x^2)) \lhd (a+x).\end{aligned}$$

Thus, the right Jordan identity holds. On the other hand,

$$(a+x) \lhd (b+y) - (b+y) \lhd (a+x) = a \bullet y - b \bullet x \in I.$$

It shows that $\mathfrak{I}^{\text{ann}} \subset I$. Finally, we have

$$(a+x) \lhd b = (a+x) \lhd (b+0) = a \bullet 0 + x \bullet 0 = 0,$$

which implies that $I \subset Z^r(\mathfrak{I})$. □

Theorem 10. *Let* $\mathfrak{I}$ *be a quasi-Jordan algebra and let* I *be an ideal of* $\mathfrak{I}$ *such that* $\mathfrak{I}^{\text{ann}} \subset I \subset Z^r(\mathfrak{I})$. *Then* $\mathfrak{I}$ *is split over* I *if and only if* $\mathfrak{I}$ *is the demisemidirect product of* I *with a Jordan algebra* J.

Proof. This follows from Lemma 9 and the discussion preceding that lemma. □

The property of being a split quasi-Jordan algebra is important for us, among other reasons because every quasi-Jordan algebra is isomorphic to a subalgebra of a split quasi-Jordan algebra.

Now suppose that $\mathfrak{I}$ is a split quasi-Jordan algebra with a specific unit e. Since, by (14), $\mathfrak{I}^{\text{ann}}$ and $Z^r(\mathfrak{I})$ coincide, there is a Jordan algebra J such that $\mathfrak{I} = \mathfrak{I}^{\text{ann}} \oplus J$.

Because $e \in \mathfrak{I}$ is a unit in $\mathfrak{I}$, there are elements $a \in \mathfrak{I}^{\text{ann}}$ and $\epsilon \in J$ such that $e = a + \epsilon$. If $b + y \in \mathfrak{I}$, with $b \in \mathfrak{I}^{\text{ann}}$ and $y \in J$, we have

$$b + y = (b+y) \lhd e = (b+y) \lhd (a+\epsilon) = b \lhd \epsilon + y \lhd \epsilon = (b+y) \lhd \epsilon.$$

The last equality implies that ϵ is a unit in $\mathfrak{I}$ and a unit in the Jordan algebra J. Also, ϵ is the only element in J such that $a + \epsilon$ is a unit in $\mathfrak{I}$ for all $a \in \mathfrak{I}^{\text{ann}}$. This shows that the units in a split quasi-Jordan algebra are of the form $a + \epsilon$, where $a \in \mathfrak{I}^{\text{ann}}$ and ϵ is the unique unit of a unital Jordan algebra; hence $U_r(\mathfrak{I}) = \mathfrak{I}^{\text{ann}} \oplus \{\epsilon\}$.

Theorem 11. *Let $\Im = \Im^{\mathrm{ann}} \oplus J$ be a unital split quasi-Jordan algebra and $\epsilon \in J$ a unit of $\Im$ which is also the unique unit of the Jordan algebra J. Then $(\Im, U, W, \epsilon)$ is a unital quadratic quasi-Jordan algebra in which U and W are defined as follows (if $x, y \in J$, we denote the product of x with y by xy instead of $x \lhd y$):*

$$U_{a+x}(b+y) = b + U_x y, \quad W_{a+x}(b+y) = -a \lhd y + (xy), \tag{16}$$

where $a, b \in \Im^{\mathrm{ann}}$, $x, y \in J$ and $U_x y = (2R_x^2 - R_{x^2})y$. Here $R_x y = yx = xy$.

As the reader probably has noticed, where no misunderstanding can arise, we will use the letter U to denote simultaneously the map U_{a+x} for any $a+x \in \Im$ and the map U_z for every $z \in J$.

Proof. Keep in mind that J is a Jordan algebra. We have $U_\epsilon(b+y) = b + U_\epsilon y = b + y$; thus $U_\epsilon = I_d$. At the same time, $W_\epsilon \epsilon = \epsilon$.

Obviously $U_{a+x}(b+y)$ and $W_{a+x}(b+y)$ are linear with respect to $(b+y)$ and $U_{\lambda(a+x)}\epsilon = U_{\lambda x}\epsilon = \lambda^2 U_x \epsilon = \lambda^2 U_{a+x}\epsilon$.

Next,

$$U_{U_{a+x}(b+y)}(c+z) = U_{b+U_x y}(c+z) = c + U_{U_x y} z. \tag{17}$$

On the other hand,

$$\begin{aligned} U_{a+x}U_{b+y}U_{a+x}(c+z) &= U_{a+x}U_{b+y}(c+U_x z) \\ &= U_{a+x}(c+U_y U_x z) = c + U_x U_y U_x z; \end{aligned} \tag{18}$$

since $U_{U_x y} z = U_x U_y U_x z$. From (17) and (18) we have

$$U_{U_{a+x}(b+y)} = U_{a+x}U_{b+y}U_{a+x}.$$

Next we check condition QQJ2. First we obtain

$$\begin{aligned} V_{(b+y),(a+x)}(c+z) &= (U_{((b+c)+(y+z))} - U_{(b+y)} - U_{(c+z)})(a+x) \\ &= (a + U_{y+z}x) - (a + U_y x) - (a + U_z x) \\ &= -a + (U_{y+z}x - U_y x - U_z x) = -a + V_{y,x} z. \end{aligned} \tag{19}$$

Similarly, $V_{(a+x),(b+y)}(c+z) = -b + V_{x,y} z$. Hence

$$U_{a+x}V_{(b+y),(a+x)}(c+z) = U_{a+x}(-a + V_{y,x}z) = -a + U_x V_{y,x} z,$$

which implies that

$$\begin{aligned} W_{d+w}U_{a+x}V_{(b+y),(a+x)}(c+z) &= W_{d+w}(-a + U_x V_{y,x} z) \\ &= -d \lhd (U_x V_{y,x} z) + w(U_x V_{y,x} z). \end{aligned} \tag{20}$$

Observe also that

$$V_{(a+x),(b+y)}U_{a+x}(c+z) = V_{(a+x),(b+y)}(c + U_x z) = -b + V_{x,y} U_x z,$$

and from this we conclude that

$$W_{d+w}V_{(a+x),(b+y)}U_{a+x}(c+z) = W_{d+w}(-b+V_{x,y}U_xz) \\ = -d \triangleleft (V_{x,y}U_xz) + w(V_{x,y}U_xz). \tag{21}$$

Using the commutativity property $U_xV_{y,x} = V_{x,y}U_x$ of Jordan algebras, it follows from (20) and (21) that $W_{d+w}U_{a+x}V_{(b+y),(a+x)} = W_{d+w}V_{(a+x),(b+y)}U_{a+x}$ for all $(a+x)$, $(b+y)$, $(d+w) \in \Im$. This concludes the proof of the theorem. □

Let $\Im = \Im^{\text{ann}} \oplus J$ be a unital split quasi-Jordan algebra with $\epsilon \in J$ as unit, then we denote $\wp(\Im)$ for the unital quadratic quasi-Jordan algebra $(\Im, U, W, \epsilon)$ corresponding to the previous theorem.

5. Split unital quadratic quasi-Jordan algebras

For a unital quadratic quasi-Jordan algebra $(\Im, U, W, e)$ we put

$$Z^r(\Im) = \{z \in \Im : W_xz = 0 \text{ for all } x \in \Im\}.$$

We denote by $\Im^{\text{ann}}$ the subspace of $\Im$ spanned by elements of the form

$$(U_xV_{y,x} - V_{x,y}U_x)z, \quad \text{with } x, y, z \in \Im.$$

$\Im$ is a unital quadratic Jordan algebra if and only if $\Im^{\text{ann}} = \{0\}$ and U_x is K-quadratic with respect to all $x \in \Im$. From QQJ2 follows that $\Im^{\text{ann}} \subset Z^r(\Im)$.

Proposition 12. *If $(\Im, U, W, e)$ is a unital quadratic quasi-Jordan algebra, the unit e does not belong to $\Im^{\text{ann}}$.*

Proof. Otherwise, one can write $e = \sum(U_{x_i}V_{y_i,x_i} - V_{x_i,y_i}U_{x_i})z_i$, where the sum is finite. Applying W_e to this equality and taking into account QQJ1 and QQJ2 we obtain $e = 0$, which is impossible. □

In fact a more general statement holds: e does not belong to $Z^r(\Im)$.

Definition 13. We say that a unital quadratic quasi-Jordan algebra $(\Im, U, W, e)$ is *split* if there exists a subspace QJ such that $\Im = \Im^{\text{ann}} \oplus QJ$ as a direct sum of subspaces and $U_xQJ \subset QJ$ for all $x \in QJ$.

Lemma 14. *Let $(\Im, U, W, e)$ be a split unital quadratic quasi-Jordan algebra such that $\Im = \Im^{\text{ann}} \oplus QJ$. Then, if U is K-quadratic, QJ is a unital quadratic Jordan algebra.*

Proof. Take $x, y, z \in QJ$. We have $(U_xV_{y,x} - V_{x,y}U_x)z \in \Im^{\text{ann}} \cap QJ$; therefore

$$(U_xV_{y,x} - V_{x,y}U_x)z = 0,$$

so $U_xV_{y,x} = V_{x,y}U_x$ for all $x, y \in QJ$. This shows that $(QJ, U_{|QJ}, W_{|QJ}, e)$ is a unital quadratic Jordan algebra. □

Now suppose that $(D, \vdash, \dashv, e)$ is a unital split dialgebra such that $D = D^{\text{ann}} \oplus A$, where A is an associative algebra (so $\vdash = \dashv$ on A) and e is a bar unit of D which is the unique unit of A. (D^{ann}, the *annihilator ideal* of D, is the subspace of D spanned by elements of the form $x \dashv y - x \vdash y$; see [Velásquez and Felipe 2009] for details). Then

$$(a+i) \dashv (b+j) = (a \dashv j) + ij \quad \text{and} \quad (a+i) \vdash (b+j) = (i \vdash b) + ij,$$

where $a, b \in D^{\text{ann}}$ and $i, j \in A$, moreover D^{ann} is spanned by elements of the form $a \dashv i$ and $k \vdash b$.

Theorem 15. *If $D = D^{ann} \oplus A$ is a unital split dialgebra as above, the unital quadratic quasi-Jordan algebra $(QQ(D), e)$ is split.*

Proof. Since $U_x y = (x \vdash y) \dashv x = xyx \in A$ if $x, y \in A$, it is sufficient to check that $D^{\text{ann}} = (QQ(D))^{\text{ann}}$. Now, it is easy to show through calculation that the term in the expression

$$U_{(a+i)} V_{(b+j),(a+i)}(c+k) - V_{(a+i),(b+j)} U_{(a+i)}(c+k) \tag{22}$$

that belongs to A is $(i((ji)k))\,i + (i((ki)\,j))\,i - ((ij)((ik)i)) - (((ik)\,i)\,j) = T$; but since A is associative we conclude that $T = 0$. The remaining four terms are of the form $d \dashv l$ and $m \vdash f$. It follows that $(QQ(D))^{\text{ann}} \subset D^{\text{ann}}$. On the other hand, taking $i = j = e$ in (22), this expression will be equal to $a \dashv k + k \vdash a - b \dashv k - k \vdash b$. Setting $b = 0$ we conclude that the elements of the form $d \dashv l$ and $m \vdash f$ (which span D^{ann}) can be obtained by means of (22). Thus $D^{\text{ann}} \subset (QQ(D))^{\text{ann}}$. This completes the proof of the theorem. $\square$

Proposition 16. *Let $\wp(\Im) = (\Im, U, W, \epsilon)$ be the unital quadratic quasi-Jordan algebra associated to a unital split quasi-Jordan algebra $\Im = \Im^{\text{ann}} \oplus J$ with $\epsilon \in J$ as a unit. Then $\wp(\Im)$ is split.*

Proof. It follows from (16) that $U_x y \in J$ for any $x, y \in J$. At the same time,

$$\begin{aligned}(U_{a+x} V_{(b+y),(a+x)} - V_{(a+x),(b+y)} U_{a+x})(c+z) &= (-a + U_x V_{y,x} z) - (-b + V_{x,y} U_x z) \\ &= b - a,\end{aligned}$$

where $a, b, c \in \Im^{\text{ann}}$ and $x, y, z \in J$. We obtain $b = U_x V_{(b+y),x} - V_{x,(b+y)} U_x)(c+z)$ by setting $a = 0$. Since $b \in \Im^{\text{ann}}$ is arbitrary, this implies that $\wp(\Im)^{\text{ann}} = \Im^{\text{ann}}$. $\square$

6. Concluding remarks

We propose a few possible directions of work:

(i) Inner ideals play a role in the theory of quadratic Jordan algebras analogous to that played by the one-sided ideals in the theory of associative algebras. It is therefore important to develop a corresponding ideal theory for quadratic quasi-Jordan algebras.

(ii) Although representations do not play as much of a role in the theory of Jordan algebras as they do in the associative or Lie theories, we propose to develop a representation theory for unital quadratic quasi-Jordan algebras. There exists some previous work of McCrimmon about this subject for quadratic Jordan algebras.

(iii) One of the most controversial concepts about dialgebras and quasi-Jordan algebras, one which is still under study, is that of a regular or invertible element. We think the reason for this is the nonuniqueness of the unit in these algebraic structures. Hence, an interesting subject of study could be the notion of a regular element on a unital quadratic quasi-Jordan algebra. Maybe this could help unify views and opinions in the near future.

(iv) There are some techniques for establishing identities in Jordan algebras and quadratic Jordan algebras, among which the best known are Macdonald's principle, Kocher's principle and McCrimmon's principle. It would be useful to find corresponding principles for unital quadratic quasi-Jordan algebras with the help of which we may know, for instance, whether (8) and (9) hold for any unital quadratic quasi-Jordan algebra.

Acknowledgment

The author thanks the anonymous referee for his/her careful reading and remarks which helped to improve the manuscript.

References

[Albert 1934] A. A. Albert, "On a certain algebra of quantum mechanics", *Ann. of Math.* **35**:1 (1934), 65–73. MR 1503142 Zbl 0008.42104

[Bertram 2000] W. Bertram, *The geometry of Jordan and Lie structures*, Lecture Notes in Math. **1754**, Springer, Berlin, 2000. MR 2002e:17041 Zbl 1014.17024

[Bremner 2010] M. R. Bremner, "On the definition of quasi-Jordan algebra", *Comm. Algebra* **38**:12 (2010), 4695–4704. Zbl 05859516

[Bremner and Peresi 2010] M. R. Bremner and L. A. Peresi, "Special identities for quasi-Jordan algebras", preprint, 2010. arXiv 1008.2723

[Faybusovich 1997] L. Faybusovich, "Euclidean Jordan algebras and interior-point algorithms", *Positivity* **1**:4 (1997), 331–357. MR 99m:90108 Zbl 0973.90095

[Felipe 2009] R. Felipe, "Restrictive split and unital quasi-Jordan algebras", preprint 1-09-09, Centro de Investigación en Matemáticas, Guanajuato, 2009, Available at http://www.cimat.mx/reportes/enlinea/I-09-09.pdf.

[Jordan et al. 1934] P. Jordan, J. von Neumann, and E. Wigner, "On an algebraic generalization of the quantum mechanical formalism", *Ann. of Math.* **35**:1 (1934), 29–64. MR 1503141 Zbl 0008.42103

[Kolesnikov 2008] P. S. Kolesnikov, "Varieties of dialgebras, and conformal algebras", *Sib. Math. J.* **49**:2 (2008), 257–272. MR 2009b:17002 Zbl 1164.17002

[Loday 1993] J.-L. Loday, "Une version non commutative des algèbres de Lie: les algèbres de Leibniz", *Enseign. Math.* (2) **39**:3-4 (1993), 269–293. MR 95a:19004 Zbl 0806.55009

[Loday 2001] J.-L. Loday, "Dialgebras", pp. 7–66 in *Dialgebras and related operads*, edited by J.-L. Loday et al., Lecture Notes in Math. **1763**, Springer, Berlin, 2001. MR 2002i:17004 Zbl 0999.17002

[McCrimmon 1966] K. McCrimmon, "A general theory of Jordan rings", *Proc. Nat. Acad. Sci. U.S.A.* **56** (1966), 1072–1079. MR 34 #2643 Zbl 0139.25502

[McCrimmon 1978] K. McCrimmon, "Jordan algebras and their applications", *Bull. Amer. Math. Soc.* **84**:4 (1978), 612–627. MR 57 #6115 Zbl 0421.17010

[McCrimmon 2004] K. McCrimmon, *A taste of Jordan algebras*, Springer, New York, 2004. MR 2004i:17001 Zbl 1044.17001

[Velásquez and Felipe 2008] R. Velásquez and R. Felipe, "Quasi-Jordan algebras", *Comm. Algebra* **36**:4 (2008), 1580–1602. MR 2009b:17004 Zbl 1188.17021

[Velásquez and Felipe 2009] R. Velásquez and R. Felipe, "Split dialgebras, split quasi-Jordan algebras and regular elements", *J. Algebra Appl.* **8**:2 (2009), 191–218. MR 2010m:17024 Zbl 1188.17022

[Zelmanov 1979] E. I. Zelmanov, "Jordan division algebras", *Algebra i Logika* **18**:3 (1979), 286–310. MR 81m:17021

[Zelmanov 1983] E. I. Zelmanov, "Prime Jordan algebras, II", *Sibirsk. Mat. Zh.* **24**:1 (1983), 89–104. MR 85d:17011

Received February 19, 2010. Revised May 10, 2010.

RAÚL FELIPE
raulf@cimat.mx
CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS (CIMAT)
JALISCO S/N, COL. VALENCIANA
36240 GUANAJUATO, GTO.
MEXICO

and

INSTITUTO DE CIBERNÉTICA, MATEMÁTICA Y FÍSICA (ICIMAF)
CALLE F ESQUINA A 15, NO 309
VEDADO
10400 CIUDAD DE LA HABANA, CUBA

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 1, 2011

THE DIRICHLET PROBLEM FOR CONSTANT MEAN CURVATURE GRAPHS IN $\mathbb{H} \times \mathbb{R}$ OVER UNBOUNDED DOMAINS

ABIGAIL FOLHA AND SOFIA MELO

We study graphs of constant mean curvature H in $\mathbb{H} \times \mathbb{R}$, where $\mathbb{H}$ is the hyperbolic plane. When $0 < H < \frac{1}{2}$, we find necessary and sufficient conditions for the existence of these graphs over unbounded domains in $\mathbb{H}$, having prescribed, possibly infinite, boundary data.

1. Introduction

This work deals with graphs in $\mathbb{H} \times \mathbb{R}$, where $\mathbb{H}$ is the hyperbolic plane, having constant mean curvature H defined over unbounded domains in $\mathbb{H}$. In the Euclidean space $\mathbb{R}^3$, Finn [1963; 1965] and Jenkins and Serrin [1966] studied the existence of a function whose graph over a bounded domain $\mathfrak{D} \subset \mathbb{R}^2$ is minimal and has prescribed boundary data. Finn studied the behavior of graphs in $\mathbb{R}^3$ over bounded convex domains in $\mathbb{R}^2$ having constant mean curvature $H = 0$ and established criteria to determine when a graph tends to infinity over a boundary arc of the domain. Jenkins and Serrin showed that necessary conditions for the existence of graphs over a domain $D \subset \mathbb{R}^2$ having unbounded boundary values given by the flux (see Section 5 for precise definition) on D are also sufficient.

The work of Jenkins and Serrin inspired many extensions to other ambient spaces and some of their ideas are present in these extensions. In $\mathbb{H} \times \mathbb{R}$ the existence theorem was proved by Nelli and Rosenberg [2002]. Collin and Rosenberg [2010] treated the case in which the domain $\mathfrak{D}$ in $\mathbb{H}$ is unbounded and Mazet, Rodríguez and Rosenberg [2008] dealt with a more general setting. Spruck [1972] extended the theorem of Jenkins and Serrin to constant mean curvature graphs in $\mathbb{R}^3$ over bounded domains of $\mathbb{R}^2$. Spruck's work introduced an important idea for the case $H \neq 0$: the reflection of the curves in order to get values $-\infty$ over boundary arcs. The case of graphs of constant mean curvature over bounded domains in $\mathbb{H}$

Folha was supported by FAPERJ. Melo was supported by CAPES.
MSC2000: 53A10, 53C42.
Keywords: constant mean curvature, unbounded domain, graph, Dirichlet problem.

was considered by Hauswirth, Rosenberg and Spruck [2009]. There are other articles about this theory; see, for example, [Rosenberg 2002; Pinheiro 2009; Gálvez and Rosenberg 2010].

It is a well known fact that there is no entire graph for H greater than $1/2$ in $\mathbb{H} \times \mathbb{R}$; moreover, Hauswirth, Rosenberg and Spruck [2008] prove that a complete graph with $H = 1/2$ in $\mathbb{H} \times \mathbb{R}$ is an entire graph. Hence, we consider in this work values of $H > 0$ less than $1/2$. We take a convex domain $\mathcal{D}$ whose boundary $\partial\mathcal{D}$ is composed of ideal arcs $\{A_i\}$, $\{B_j\}$ and $\{C_k\}$ such that the curvatures of the arcs with respect to the domain are $\kappa(A_i) = 2H$, $\kappa(B_j) = -2H$ and $\kappa(C_k) \geq 2H$. We give necessary and sufficient conditions on the geometry of the domain $\mathcal{D}$ which assure the existence of a function u defined in $\mathcal{D}$, whose graph has constant mean curvature and u assumes the value $+\infty$ on each A_i, $-\infty$ on each B_j and prescribed continuous data on each C_k. The conditions, as in Jenkins and Serrin's work [1966], will be considered in terms of the lengths and the areas of inscribed polygons. Since these quantities are infinite in general, the formulation of the conditions is somewhat delicate. For an example, the reader may look at Section 8. In order to control lengths we do the same as Collin and Rosenberg [2010]; however, the new and key idea appears when we consider the area and we split it in two parts, one finite and the other infinite (see Section 3).

This paper is organized as follows. In Section 2, we introduce notation. In Section 3, we state the main theorems, which will be proved in Section 7. Sections 4 and 5 contain general maximum principles and the flux formulas, which are useful tools to prove preliminary results and the necessary conditions of the main theorems. In Section 6, we state results about divergence lines, which are essential to prove the sufficient conditions of the main theorems. Finally, in Section 8, we construct an example.

2. Notation

Let $\mathbb{H}$ be the hyperbolic plane, and $\mathbb{H} \times \mathbb{R}$ be given the product metric. Let $u : D \subset \mathbb{H} \to \mathbb{R}$ be a function in $C^2(D)$, where D is a simply connected domain. Denote the graph of u by $S = \mathrm{Graph}(u) = \{\, (p, u(p)) \mid p \in D \,\}$. Since S is a graph, there are two choices for the unit normal vector $N(P)$ to S at a point $P = (p, u(p))$, $p \in D$. We choose

$$N(P) = \frac{-\nabla u + \partial_t}{\sqrt{1 + \|\nabla u\|^2}},$$

that is, the normal vector pointing up.

Let $\overrightarrow{H}(P)$ be the mean curvature vector of S at P. The mean curvature function of S at a point P is defined by $H(P) = \langle N, \overrightarrow{H} \rangle(P)$. Consider graphs with $0 < H(P) < \frac{1}{2}$ for all $P \in S$; in particular, $\overrightarrow{H}$ points up.

The graph S has constant mean curvature H if $H(P) = H$ for all $P \in S$. This means u satisfies the equation

$$Mu := \operatorname{div}\left(\frac{\nabla u}{\sqrt{1+|\nabla u|^2}}\right) = 2H, \tag{1}$$

where the divergence and gradient are taken with respect to the metric on $\mathbb{H}$. A function that satisfies this equation in D is called a *solution in D*. We will use the notation $X_u = \nabla u/W_u$, where $W_u = \sqrt{1+|\nabla u|^2}$.

Let $E \subset \mathbb{H}$ be a smooth curve. Denote by $\kappa(p)$ the (nonnegative) curvature of E at a point $p \in E$ and when $\kappa(p) = K$ for all $p \in E$, we will say $\kappa(E) = K$. When E is a boundary arc of a domain D, we will often let $\kappa(p)$, $p \in E$, denote the algebraic curvature of E at p with respect to D, that is, $\kappa(p) \geq 0$ if E is convex with respect to D, and $\kappa(p) < 0$ otherwise.

We will consider ideal domains in $\mathbb{H}$ whose asymptotic boundary is composed only of a finite number of isolated points. Domains mean a connected, simply connected open set. The boundary of an ideal domain will be called *ideal polygon*.

3. Main theorems

In this section, we state the theorems that give necessary and sufficient conditions for the existence of constant mean curvature graphs which take the boundary values $+\infty$ on certain arcs A_i, $-\infty$ on arcs B_i and continuous data on arcs C_i.

Definition 3.1 (admissible domain). We say that an unbounded domain $\mathcal{D}$ in $\mathbb{H}$ is admissible if it is simply connected and $\partial\mathcal{D}$ is an ideal polygon with sides $\{A_i\}$, $\{B_i\}$ and $\{C_i\}$ satisfying $\kappa(A_i) = 2H$, $\kappa(B_i) = -2H$ and $\kappa(C_i) \geq 2H$, respectively (with respect to the interior of $\mathcal{D}$). Suppose that no two of the arcs A_i and no two of the arcs B_i have a common endpoint. Moreover, all the sides of $\partial\mathcal{D}$ are contained in $\mathbb{H}$ and all the vertices of $\partial\mathcal{D}$ are in the asymptotic boundary of $\mathbb{H}$.

Definition 3.2 (Dirichlet problem). Let $\mathcal{D}$ be an admissible domain and fix $0 < H < \frac{1}{2}$. The generalized Dirichlet problem is to find a solution of (1) in $\mathcal{D}$ of mean curvature H, which assumes the value $+\infty$ on each A_i, $-\infty$ on each B_i and prescribed continuous data on each C_i.

Definition 3.3 (admissible inscribed polygon). Let $\mathcal{D}$ be an admissible domain. We say that $\mathcal{P}$ is an admissible inscribed polygon if $\mathcal{P} \subset \mathcal{D} \cup \partial\mathcal{D}$, its sides have curvature $\pm 2H$ and all the vertices of $\mathcal{P}$ are vertices of $\mathcal{D}$.

In [Hauswirth et al. 2009], the Dirichlet problem was solved for bounded admissible domains. The necessary and sufficient conditions in this case are in terms of the lengths and areas of inscribed polygons. When the domain is unbounded, these quantities can be infinite. Using the ideas in [Collin and Rosenberg 2010], we control the lengths as follows.

Let $\mathcal{P}$ be an inscribed polygon in $\mathcal{D}$ and let $\{d_i\}$ be the vertices of $\mathcal{P}$. Consider the set

$$\Theta = \{\mathcal{H}_i \mid \mathcal{H}_i \text{ is a horocycle at } d_i, \mathcal{H}_i \cap \mathcal{H}_j = \varnothing,\ i \neq j, \\ \text{and these horocycles satisfy condition (5)}\}.$$

Remark 3.1. We define condition (5) in Section 7. This is a technical condition which is always satisfied for sufficiently "small" horocycles at the vertices d_i. Throughout we only consider horocycles $\mathcal{H}_i$ contained in this set Θ.

Let F_i be the convex horodisk with boundary $\mathcal{H}_i$. Each A_i meets exactly two horodisks. Denote by $\widetilde{A}_i$ the compact arc of A_i which is the part of A_i outside the two horodisks; we define $|A_i|$ as the length of $\widetilde{A}_i$. For each arc $\eta_j \in \mathcal{P}$ we define $\widetilde{\eta}_j$ and $|\eta_j|$ in the same way.

We define

$$\alpha(\mathcal{P}) = \sum_{A_i \in \mathcal{P}} |A_i|, \quad \beta(\mathcal{P}) = \sum_{B_i \in \mathcal{P}} |B_i| \quad \text{and} \quad l(\mathcal{P}) = \sum_j |\eta_j|$$

where $\mathcal{P} = \bigcup_j \eta_j$.

Now, let $\gamma_i = \mathcal{H}_i \cap (\mathcal{D} \cup \partial\mathcal{D})$. Consider γ_i^* the geodesic reflection of γ_i about the geodesic joining the endpoints of γ_i.

Denote by Ω the domain bounded by $\mathcal{P}$ and $\widetilde{\Omega} = \bigcup_j (\Omega \cap F_j)$, where the area $\mathcal{A}(\Omega \cap F_j)$ is finite.

Let $\mathcal{H} = \{\mathcal{H}_i\}_{i=1,\dots,n}$ be a family of horocycles.

For each family $\mathcal{H}$, we define

$$\widetilde{\mathcal{A}}(\Omega) := \mathcal{A}(\Omega_{\mathcal{H}}) + \mathcal{A}(\widetilde{\Omega}),$$

where

$$\mathcal{A}(\Omega_{\mathcal{H}}) = \mathcal{A}\big(\Omega - \big(\textstyle\bigcup_i (\Omega \cap F_i)\big)\big)$$

for all i. This definition plays an important role in this work – actually, this is the key idea which we need to extend previous results of [Collin and Rosenberg 2010; Hauswirth et al. 2009] to our setting. In Section 7, we will point out where this definition is used.

Notice that the definitions of $\alpha(\mathcal{P})$, $\beta(\mathcal{P})$ and $l(\mathcal{P})$ can be extended to the boundary of $\mathcal{D}$ and $\widetilde{\mathcal{A}}(\Omega)$ to $\mathcal{D}$.

Remark 3.2. When $\partial\mathcal{D}$ only has sides of type A_i and B_i, we have that $\widetilde{\mathcal{A}}(\mathcal{D}) = \mathcal{A}(\mathcal{D})$, because $\mathcal{A}(\mathcal{D} \cap F_i)$ is finite for all i (this may be infinite when there are arcs C_i present). Also, in this case, for all admissible polygons $\mathcal{P}$ in $\mathcal{D}$ we have $\widetilde{\mathcal{A}}(\Omega) = \mathcal{A}(\Omega)$.

With these definitions we can state the main theorems.

Theorem 3.1. *Consider the Dirichlet problem in an admissible domain $\mathscr{D}$ and suppose the family $\{C_i\}$ is empty. Then, there exists a solution to the Dirichlet problem if and only if for some choice of the horocycles* (*in* Θ) *at the vertices*,

$$\alpha(\partial\mathscr{D}) = \beta(\partial\mathscr{D}) + 2H\widetilde{\mathscr{A}}(\mathscr{D}) \tag{2}$$

and for all admissible polygons $\mathscr{P}$,

$$2\alpha(\mathscr{P}) < l(\mathscr{P}) + 2H\widetilde{\mathscr{A}}(\Omega) \quad \textit{and} \quad 2\beta(\mathscr{P}) < l(\mathscr{P}) - 2H\widetilde{\mathscr{A}}(\Omega). \tag{3}$$

Now we remove the hypothesis that $\{C_i\}$ is empty from Theorem 3.1.

Theorem 3.2. *Consider the Dirichlet problem in an admissible domain $\mathscr{D}$ and suppose the family $\{C_i\}$ is nonempty. Then there exists a solution to the Dirichlet problem if and only if for some choice of the horocycles* (*in* Θ) *at the vertices*,

$$2\alpha(\mathscr{P}) < l(\mathscr{P}) + 2H\widetilde{\mathscr{A}}(\Omega) \quad \textit{and} \quad 2\beta(\mathscr{P}) < l(\mathscr{P}) - 2H\widetilde{\mathscr{A}}(\Omega) \tag{4}$$

for all admissible polygons $\mathscr{P}$.

4. Maximum principles

The next results are general maximum principles for sub- and supersolutions of the constant mean curvature operator for boundary data having a finite number of discontinuities. The first one is in a bounded domain and the second one is in an unbounded domain. First we state a local lemma whose proof is in [Hauswirth et al. 2009].

Lemma 4.1. *Let u^1 and u^2 be functions in $C^2(D)$, $D \subset \mathbb{H}$. Then*

$$\left\langle \nabla u^1 - \nabla u^2, \frac{\nabla u^1}{W_1} - \frac{\nabla u^2}{W_2} \right\rangle \geq 0,$$

with equality at a point if and only if $\nabla u^1 = \nabla u^2$. Here $W_i = W(\nabla u^i)$, $W(p) = \sqrt{1+|p|^2}$, $i = 1, 2$.

Theorem 4.1 (general maximum principle 1). *Let u^1 and u^2 satisfy $Mu^1 \geq 2H \geq Mu^2$ in a bounded domain $D \subset \mathbb{H}$. Suppose that* $\liminf(u^2 - u^1) \geq 0$ *for any approach to ∂D with the possible exception of a finite number of points of ∂D. Then $u^2 \geq u^1$ with strict inequality unless $u^2 \equiv u^1$.*

Theorem 4.2 (general maximum principle 2). *Let D be a domain with ∂D an ideal polygon. Let $W \subset D$ be a domain and let $u^1, u^2 \in C^0(\overline{W})$ be two solutions of* (1) *in W with $u^1 \leq u^2$ on ∂W. Suppose that for each vertex p of ∂D,* $\liminf \mathrm{dist}_{\mathbb{H}}(\Gamma_1, \Gamma_2) \to 0$ *as one converges to p, where Γ_1, Γ_2 are the curves on ∂D with p as vertex. Then $u^1 \leq u^2$ in W.*

The proof of Theorem 4.1 is given in [Hauswirth et al. 2009]. The proof of Theorem 4.2 is analogous to the one of Theorem 2 in [Collin and Rosenberg 2010] using Lemma 4.1.

We will see examples of barriers which will enable us to control convergence of solutions on ∂D, when we know they converge in D. Then the limit of the sequence on the boundary is the limit of the boundary values and the limit solution extends continuously to the boundary. The following examples can be found in [Hauswirth et al. 2009].

Example 4.1. Let $B \subset \mathbb{H}$ be a ball of radius δ centered at p. Let p_1 and p_2 be "antipodal" points on ∂B. We choose points d_1, d_2 on ∂B symmetric with respect to the geodesic through $p_1 p p_2$. Now let B_1 be an arc of curvature $-2H$ (as seen from p) joining d_1, d_2 and set $A_1 = B_1^*$, where B_1^* is the geodesic reflection of B_1. Let B_2 be the reflection of B_1 with respect to the geodesic orthogonal to $p_1 p p_2$ through p, and set $A_2 = B_2^*$. For δ small compared with H, there is a solution u^+ in B^+, the connected domain bounded by A_1, A_2 and arcs of ∂B such that u^+ is $+\infty$ on A_1 and A_2 and a constant $M > 0$ on the rest of ∂B^+. Similarly, there is a solution u^- in B^-, the domain bounded by B_1, B_2 and parts of ∂B such that u^- is $-\infty$ on B_1 and B_2 and a constant $-M$, $M > 0$ on the rest of ∂B^-.

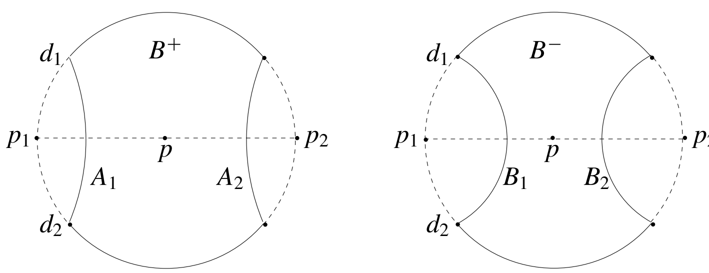


Figure 1. Domains of the solutions u^+ and u^- in Example 4.1.

5. Flux formulas

In this section, we state some results about the flux of a solution. As in [Jenkins and Serrin 1966], the flux will give us the necessary conditions, which also will be sufficient, to the existence of solutions having infinite boundary values. Finn [1963] proved that if a minimal solution in Euclidean space tends to $+\infty$ or $-\infty$ over a boundary arc Γ, then Γ is a line. The flux formula gives the requirement on the curvature of the boundary arcs of an admissible domain.

Let $u \in C^2(D) \cap C^1(\bar{D})$ be a solution in the bounded domain D. Then integrating (1) over D, we have

$$2H\mathscr{A}(D) = \int_{\partial D} \left\langle \frac{\nabla u}{W}, \nu \right\rangle ds,$$

where $\mathscr{A}(D)$ is the area of D and ν is the outer normal to ∂D. This integral is called *the flux of u across ∂D*. Let η be a subarc of ∂D (homeomorphic to [0, 1]). Even if u is not differentiable on η we can define the flux of u across η as follows; see [Hauswirth et al. 2009].

Definition 5.1. Choose Υ to be an embedded smooth curve in D so that $\eta \cup \Upsilon$ bounds a simply connected domain Δ_Υ. We then define the flux of u across η as

$$F_u(\eta) = 2H\mathscr{A}(\Delta_\Upsilon) - \int_{\Upsilon} \left\langle \frac{\nabla u}{W}, \nu \right\rangle ds.$$

The last integral is well defined, and $F_u(\eta)$ does not depend in the choice of Υ.

With this definition we can remove the condition $u \in C^2(D) \cap C^1(\bar{D})$ and state important flux formulas, whose proofs are in [Hauswirth et al. 2009].

Theorem 5.1. *Let u be a solution in D.*

(i) *If ∂D is a compact cycle, we have $F_u(\partial D) = 2H\mathscr{A}(D)$.*

(ii) *If D is bounded in part by a C^1 arc η, then:*

(a) *If u tends to $+\infty$ on η, we have $\kappa(\eta) = 2H$ and*

$$\int_{\eta} \left\langle \frac{\nabla u}{W}, \nu \right\rangle ds = |\eta|.$$

(b) *If u tends to $-\infty$ on η, we have $\kappa(\eta) = -2H$ and*

$$\int_{\eta} \left\langle \frac{\nabla u}{W}, \nu \right\rangle ds = -|\eta|.$$

(c) *If η is C^2, $\kappa(\eta) \geq 2H$ and u is continuous on η, we have*

$$\left| \int_{\eta} \left\langle \frac{\nabla u}{W}, \nu \right\rangle ds \right| < |\eta|.$$

Lemma 5.1. *Let D be a domain bounded in part by an arc η with $\kappa(\eta) = 2H$. We take a sequence of solutions $\{u_n\}$ in D with each u_n continuous on η. Then if the sequence diverges to $-\infty$ uniformly on compact subsets of D while remaining uniformly bounded on compact subsets of η, we have*

$$\lim_{n\to\infty} \int_{\Upsilon} \left\langle \frac{\nabla u}{W}, \nu \right\rangle ds = |\eta|.$$

The next lemma is almost a converse of the above Theorem 5.1. We follow the ideas in [Mazet et al. 2008].

Lemma 5.2. *Let u be a solution in D. Let $\tilde{\eta} \subset \partial D$ be an arc with $\kappa(\tilde{\eta}) = 2H$ ($\kappa(\tilde{\eta}) = -2H$) such that $F_u(\eta) = |\eta|$ ($F_u(\eta) = -|\eta|$), for every compact arc $\eta \subset \tilde{\eta}$. Then u takes boundary value $+\infty$ ($-\infty$) on $\tilde{\eta}$.*

Proof. Suppose that $\kappa(\tilde{\eta}) = 2H$. Let η be a compact arc as in the lemma, small enough so that the domain Δ bounded by η and η^* (the geodesic reflection of η) is contained in D. Consider the solution v which takes values $+\infty$ on η and $v = u$ on η^*; this solution exists by [Hauswirth et al. 2009, Theorem 7.11]. We need to show that $u = v$. If this is not the case, the set $O = \{u - v < \epsilon\}$ is nonempty, where $\epsilon > 0$ is a regular value of $u - v$. Let D' be the connected component of the complement of O in Δ which has $\partial\Delta - \eta$ in its boundary and let O' be the complement of D' in Δ, so $O \subset O'$ and $\partial O' \subset \partial O$. Let q be a point in $\partial O' - \eta$. For $\mu > 0$, let $O'(\mu)$ be the set defined by $O'(\mu) = \{p \in O' \mid \mathrm{dist}_{\mathbb{H}}(p, \eta) > \mu\}$. Let q_1, q_2 be the endpoints of the connected component of $\partial O' \cap \partial O'(\mu)$ which contains q. Let p_i be the projection of q_i on η. Let $\widetilde{O}(\mu)$ be the domain bounded by the segments $[p_1, q_1]$, $[p_2, q_2]$, the arc $[p_1, p_2] \subset \eta$ and the boundary component of $O'(\mu)$ between q_1, q_2, which is denoted by $\Gamma(\mu)$. On $\Gamma(\mu)$ the vector $X_u - X_v$ points outside $\widetilde{O}(\mu)$. Calculating the flux of $u - v$ across $\partial O'$ gives

$$0 = F_{u-v} = \int_{\Gamma(\mu)} \langle X_u - X_v, \nu\rangle + \int_{[p_1,q_1]\cup[p_2,q_2]} \langle X_u - X_v, \nu\rangle + \int_{[p_1,p_2]} \langle X_u - X_v, \nu\rangle\,.$$

So applying the flux formula, we have

$$\begin{aligned} 0 < \int_{\Gamma(\mu)} \langle X_u - X_v, \nu\rangle &= -\int_{[p_1,q_1]\cup[p_2,q_2]} \langle X_u - X_v, \nu\rangle - \int_{[p_1,p_2]} \langle X_u - X_v, \nu\rangle \\ &\le 4\mu, \end{aligned}$$

since the last term in the first line vanishes by the hypothesis on u and Theorem 5.1 applied to v. Note that the integral on $\Gamma(\mu)$ increases when $\mu \to 0$. So this inequality cannot occur.

If $\kappa(\tilde{\eta}) = -2H$, we consider the domain Δ which is bounded by η and an arc η' of curvature greater than $2H$ (with respect to the domain Δ) contained in D having the same endpoints as η. Then we consider v the solution on Δ with values $-\infty$ on η and $v = u$ on η'; this solution exists by [Hauswirth et al. 2009, Theorem 7.11]. Then the same argument made in the case $\kappa(\eta) = 2H$ can be applied. □

6. Divergence lines

In this section, we will study some characteristics of the sets where a sequence of solutions in a domain D converges or diverges. Jenkins and Serrin [1966] studied the convergence of a sequence (monotone) using a maximum principle. They also presented the structure of the divergence set of this sequence. Here, we study the convergence of a sequence defined over bounded or unbounded domains (not necessarily monotone) without the aid of a maximum principle. Nevertheless, the structure of the set where such a sequence converges is the same one found by Jenkins and Serrin. Many ideas found here were inspired by [Mazet et al. 2008].

Definition 6.1. Let D be a domain with piecewise smooth boundary, and u_n a sequence of solutions in D. We define the convergence set as

$$\mathcal{U} = \left\{\, p \in D \mid \{\|\nabla u_n(p)\|\} \text{ is bounded independent of } n \,\right\}$$

and the divergence set as

$$\mathcal{V} = D - \mathcal{U}.$$

In this section, D denotes a domain in $\mathbb{H}$ with piecewise smooth boundary.

Lemma 6.1. *Let $p \in D$ and u_n be a sequence of solutions in the domain D. If $p \in \mathcal{U}$, there is a subsequence of $\{v_n\}$ with $v_n = u_n - u_n(p)$ converging uniformly to a solution in a neighborhood of p in D. If $p \in \mathcal{V}$, there is a compact arc $L_p(\tilde{\delta})$ of curvature $2H$ containing p such that, after passing to a subsequence, $\{N_{v_n}(p)\}$ converges to a horizontal vector which is orthogonal to $L_p(\tilde{\delta})$ having the same direction as the curvature vector $\vec{\kappa}$ of $L_p(\tilde{\delta})$, where $N_{v_n}(p)$ is the upward unit normal vector to the graph of v_n at $(p, 0)$.*

Remark 6.1. All the vectors $\{N_{u_n}(p)\}$ can be thought as vectors at $(p, 0)$ by vertical translation, with the identification $N_{u_n}(p) = N_{v_n}(p)$.

Proof of Lemma 6.1. Denote by $G(v_n)$ the graph of v_n over D. Note that $N_{u_n}(q) = N_{v_n}(q)$, and the convergence and divergence sets are the same for $\{u_n\}$ and $\{v_n\}$.

The curvature estimates (see [Zhang 2005]) give us a $\delta > 0$ independent of n (in fact δ depends only on the distance from p to ∂D) such that a neighborhood of $P = (p, v_n(p)) = (p, 0)$ in $G(v_n)$ is a graph, in geodesic coordinates, with height and slope uniformly bounded over the disk $\mathbb{D}^n_\delta(P)$ of radius δ centered at the origin of $T_P G(v_n)$. We call this graph $G_P(v_n, \delta)$.

If $p \in \mathcal{U}$ the sequence $\{\|\nabla u_n\|\}$ is bounded, so there is a subsequence of $\{N_{v_n}(p)\}$, still called $\{N_{v_n}(p)\}$, which converges to a nonhorizontal vector and consequently the tangent planes associated to this subsequence converge to a nonvertical plane Π. Then, since the graphs $G_P(v_n, \delta)$ have height and slope uniformly bounded, there is a subsequence of $\{v_n\}$ such that these graphs converge to a graph $G_P(\delta)$ with constant mean curvature H over a disk of radius δ centered at the origin of Π. Since this plane Π is a nonvertical plane, there is $\tilde{\delta}$, $0 < \tilde{\delta} \le \delta$ such that $G_P(\delta)$ is a graph over a geodesic ball in D centered at p of radius $\tilde{\delta}$. We conclude that there is a neighborhood of $p \in D$ such that a subsequence of $\{v_n\}$ converges to a solution in this neighborhood.

Now, suppose that $p \in \mathcal{V}$. Since $\{\|\nabla u_n\|\}$ is unbounded, there is a subsequence of $\{N_{v_n}(p)\}$ that converges to a horizontal vector N_P, so (for this subsequence) the tangent planes $T_P G(v_n)$ converge to a vertical plane Π and the graphs $G_P(v_n, \delta)$ converge to a constant mean curvature H graph $G_P(\delta')$ over a disk of radius $\delta' \leq \delta$ centered at the origin of Π. By the choice of the direction of the normal vector and the choice of $H > 0$, the limit of the curvature vectors of $G_P(v_n, \delta)$ has the same direction as the normal limit.

Take the curve $L_p \subset D$ passing through p orthogonal to N_P, with curvature $2H$ and the curvature vector at p having the same direction as N_P. We want to prove that $G_P(\delta') \subset (L_p \times \mathbb{R})$.

Since $G_P(\delta')$ is tangent to $L_p \times \mathbb{R}$ at P, if $G_P(\delta')$ is on one side of $L_p \times \mathbb{R}$, by the maximum principle, we have that $G_P(\delta') \subset (L_p \times \mathbb{R})$. If this is not the case, $G_P(\delta') \cap (L_p \times \mathbb{R})$ is composed of $k \geq 2$ curves passing through p, meeting transversely at p. So in a neighborhood of p these curves separate $G_P(\delta')$ in $2k$ components and the adjacent components lie in alternate sides of $L_p \times \mathbb{R}$. Moreover, the curvature vector alternates from pointing down to pointing up when one goes from one component to the other. This implies that the normal vector to $G_P(\delta)$ points down and up. So, for n large enough, the normal vector to $G_P(v_n, \delta)$ would point down and up, which does not occur.

Let $L_p(\tilde{\delta}) \subset D$, $\delta' \geq \tilde{\delta}$, be the curve contained in $G_P(\delta') \cap (L_p \times \{0\})$ which contains p and has length $2\tilde{\delta}$. Since $G_P(\delta') \subset (L_p \times \mathbb{R})$, we have that for all $q \in L_p(\tilde{\delta})$ the normal vector to $G_P(\delta')$ at q is a horizontal vector normal to $L_p(\tilde{\delta})$ having the same direction as the curvature vector of $L_p(\tilde{\delta})$ at q. □

Remark 6.2. Lemma 6.1 shows that the convergence set is a domain.

Lemma 6.2. *Let $\{u_n\}$ be a sequence of solutions in D. Given $p \in \mathcal{V}$, there is a curve $L \subset D$ of curvature $2H$ which passes through p and such that, after passing to a subsequence, the sequence of normal vectors $\{N_{u_n}|_L\}$ converges to a horizontal vector normal to L having the same direction as the curvature vector of L. This curve L contains the compact arc $L_p(\tilde{\delta})$ given in Lemma 6.1.*

Proof. Let L be the curve of constant curvature $2H$ in D which contains $L_p(\tilde{\delta})$ joining the points of ∂D ($L_p(\tilde{\delta})$ is given in Lemma 6.1). Given $p, q \in D$, denote by $\overline{pq}$ the compact arc in L between p, q. We define

$$\Lambda = \{\, q \in L \mid \text{there is a subsequence of } \{u_n\} \text{ such that } \{N_{u_n}|_{\overline{pq}}\} \text{ becomes horizontal, orthogonal to } L \text{ having the same direction as the curvature vector of } L\}.$$

We want to prove that $\Lambda = L$. Since $p \in \Lambda$, Λ is nonempty. We will prove that Λ is open and closed. First, we will prove that Λ is open. Let q be a point in Λ. Denote $\{u_{\Lambda(n)}\}$ the subsequence associated to Λ. Since $\Lambda \subset \mathcal{V}$, Lemma 6.1 gives us a curve

$L_q(\delta)$ through q such that, after passing to a subsequence, $\{N_{u_{\Lambda(n)}}|_{L_q(\delta)}\}$ becomes horizontal and having the same direction as the curvature vector of $L_q(\delta)$. Note that this subsequence of $\{N_{u_{\Lambda(n)}}|_{L_q(\delta)}\}$ converges to a horizontal vector normal to $L_q(\delta)$ and to L simultaneously, so $L_q(\delta) \subset L$, then Λ is open.

Now we will prove that Λ is closed. We take a convergent sequence $\{q_n\}$ in Λ, $q_n \to q \in L$. We will show that $q \in \Lambda$. For each m, there is a subsequence of $\{u_{\Lambda(n)}\}$ such that $\{N_{u_{\Lambda(n)}}|_{\overline{pq}_m}\}$ becomes horizontal with the same direction as the curvature vector in $\overline{pq}_m$. By the diagonal process we obtain a subsequence of $\{u_{\Lambda(n)}\}$ such that $\{N_{u_{\Lambda(n)}}|_{\overline{pq}_m}\}$ converges to a horizontal vector having the same direction as the curvature vector of L in $\overline{pq}_m$ for all m. Then by Lemma 6.1, we can find a curve $L_{q_m}(\delta)$ having constant curvature $2H$ through q_m, (for m large, δ depends only on the distance from q to ∂D) such that $\{N_{u_{\Lambda(n)}}|_{\overline{pq}_m}\}$ converges to a horizontal vector having the same direction as the curvature vector to $L_{q_m}(\delta)$. So $L_{q_m}(\delta) \subset L$ and since $q_m \to q$, we have that, for all m large enough, $q \in L_{q_m}(\delta)$. Consequently, $q \in \Lambda$. □

An important conclusion of this lemma is that the divergence set is given by $\mathcal{V} = \bigcup_{i\in I} L_i$, where L_i is a curve, called a *divergence line*, having curvature $2H$.

Lemma 6.3. *Let $\{u_n\}$ be a sequence of solutions in D. Suppose that the divergence set $\mathcal{V}$ of $\{u_n\}$ is composed of a countable number of divergence lines. Then there is a subsequence of $\{u_n\}$, again denoted by $\{u_n\}$, such that*

(1) *the divergence set of $\{u_n\}$ is composed of a countable number of pairwise disjoint divergence lines*;

(2) *for any connected component $\mathcal{U}'$ of $\mathcal{U} = D - \mathcal{V}$ and for any $p \in \mathcal{U}'$, the sequence $\{u_n - u_n(p)\}$ converges uniformly on compact subsets of $\mathcal{U}'$ to a solution in $\mathcal{U}'$.*

Proof. Suppose that $\mathcal{V} \neq \varnothing$ and let L_1 be a divergence line of $\{u_n\}$. Lemma 6.1 guarantees that, after passing to a subsequence, $\{N_{u_n}(q)\}$ converges to a horizontal vector orthogonal to L_1 at q for all q in L_1. The divergence set of this subsequence is contained in the divergence set of the original sequence, so the divergence set associated to this subsequence has only a countable number of lines. This subsequence is still denoted by $\{u_n\}$ and its divergence set by $\mathcal{V}$. If there is a divergence line $L_2 \neq L_1$ in $\mathcal{V}$, we can find a subsequence such that $\{N_{u_n}(q)\}$ converges to a horizontal vector orthogonal to L_2 at q for each $q \in L_2$. This implies that $L_1 \cap L_2 = \varnothing$. In fact, if this does not occur, we take a point $q \in L_1 \cap L_2$ so the sequence $\{N_{u_n}(q)\}$ converges to a horizontal vector orthogonal to L_1 and L_2 at q having the same direction as the curvature vector of L_1 and L_2. Then the uniqueness of a curve through q having curvature $2H$ with a given tangent vector shows that $L_1 = L_2$. We continue this process to get a subsequence of $\{u_n\}$, still denoted by $\{u_n\}$, whose divergence set is composed of a countable number of pairwise disjoint divergence lines.

Lemma 6.1 shows that there is a subsequence of $\{u_n\}$ and a neighborhood of each point $p \in \mathcal{U}$ such that the sequence $\{u_n - u_n(p)\}$ converges to a constant mean curvature graph H, and this convergence is uniform on compact subsets of this neighborhood. Then taking a countable dense sequence $\{p_i\}$ in $\mathcal{U}'$, by the diagonal process we obtain a subsequence of $\{u_n\}$ such that $\{u_n - u_n(p)\}$ converges uniformly on compact subsets of $\mathcal{U}'$ for all $p \in \mathcal{U}'$. $\square$

Lemma 6.4. *Let $\{u_n\}$ be a sequence of solutions in D such that its divergence set is composed of a countable number of pairwise disjoint divergence lines. Suppose that $\{u_n\}$ converges to a solution u in a connected set $\mathcal{U}' \subset D$. Let γ be a compact arc in $\partial\mathcal{U}'$ included in a divergence line of $\{u_n\}$ such that $X_{u_n} \to \nu$ along γ, where ν is the outer conormal to γ with respect to $\mathcal{U}'$. Then if $p \in \mathcal{U}'$ and $q \in \gamma$, we have*

$$\lim_{n\to\infty} (u_n(q) - u_n(p)) = +\infty.$$

Proof. We choose p, q as in the hypothesis of the lemma. Since $X_{u_n} \to \nu$ we have $F_{u_n}(\gamma) \to |\gamma|$, where $F_{u_n}(\gamma)$ is the flux of u_n across γ. So Lemma 5.2 ensures that $u|_\gamma = +\infty$.

Claim 6.1. *There is an $\epsilon > 0$ such that $\partial u_n/\partial t \geq 0$ on $\{\Upsilon(t) \mid -\epsilon < t \leq 0\}$, where $\Upsilon(t)$ $(-\theta < t \leq 0,\ \theta \geq \epsilon)$ is the geodesic in $\mathcal{U}'$ such that $\Upsilon(0) = (q, 0)$ and $\Upsilon'(0) = \nu$. The inequality is strict on $\{\Upsilon(t) \mid -\epsilon < t < 0\}$.*

Using Lemma 6.1 and the fact that $u|_\gamma = +\infty$, we obtain a $\epsilon > 0$ such that $\partial u/\partial t \geq 1$ in $\{\Upsilon(t) \mid -\epsilon < t < 0\}$. The convergence $u_n \to u$ implies that $\partial u_n/\partial t > 0$ in $\{\Upsilon(t) \mid -\epsilon < t < -\eta\}$, for every $0 < \eta < \epsilon$ and $n \geq n_0(\eta)$.

If the claim is not true, considering a subsequence if necessary, there is a sequence $\{q_n\}$ in $\{\Upsilon(t) \mid -\eta \leq t \leq 0\}$ such that $q_n \to q$ and $(\partial u_n/\partial t)(q_n) = 0$.

If the sequence $\{\|\nabla u_n(q_n)\|\}$ is bounded, we have from the curvature estimates that $\{\|\nabla u_n\|\}$ is uniformly bounded on a disk D_n of radius independent of n, centered at q_n. Since $q_n \to q$, the sequence $\{\|\nabla u_n(q)\|\}$ is bounded, because for n large enough, $q \in D_n$. This contradicts that q is contained in the divergence set.

If the sequence $\{\|\nabla u_n(q_n)\|\}$ is unbounded, consider the sequence $\{u_n - u_n(q_n)\}$ and $\mathbb{D}_n^1$ the disk of radius δ in the graph of $\{u_n - u_n(q_n)\}$ centered at $(q_n, 0)$ given by the curvature estimates, δ independent of n. Since $(\partial u_n/\partial t)(q_n) = 0$, the disks $\mathbb{D}_n^1$ converge to a δ vertical disk centered at $(q, 0)$ in $\widetilde{\Upsilon} \times \mathbb{R}$, where $\widetilde{\Upsilon}$ is a curve having constant curvature $2H$ through q orthogonal to γ. Let $\mathbb{D}_n^2$ be the disk of radius δ centered at $(q, 0)$ in the graph of $\{u_n - u_n(q)\}$. Since γ is contained in a divergence line, $\{\mathbb{D}_n^2\}$ converges to a vertical disk centered at $(q, 0)$ in $\gamma \times \mathbb{R}$. Then, for n large enough, these disks $\mathbb{D}_n^1$ and $\mathbb{D}_n^2$ intersect transversally, but this is impossible because the normal vectors to $\mathbb{D}_n^1$ and $\mathbb{D}_n^2$ only depend on the gradient of u_n, so they are the same vector (on domains where both sequences are defined) for the two sequences. This proves Claim 6.1.

Let $q_t \in \mathcal{U}'$ be the point $q_t = \Upsilon(t)$, $t < 0$, for t small enough. Claim 6.1 ensures that for n large,

$$\begin{aligned} u_n(q) - u_n(p) &\geq u_n(q_t) - u_n(p) \\ &\geq u(q_t) - u(p) - 1. \end{aligned}$$

The second inequality comes from the convergence of $\{u_n\}$ to u. The third term is as large as we want, because $u|_\gamma = +\infty$. □

Lemma 6.5. *Let $E \subset \partial D$ be a smooth arc having $\kappa(E) \geq 2H$. Consider a sequence of solutions $\{u_n\}$ in $\mathfrak{D}$ such that $\lim_{n\to\infty} u_n|_E = f$ for f a continuous function. Then a divergence line cannot finish at an interior point of E.*

Proof. Let $p \in E$ be an interior point. If $\kappa(E) > 2H$ at p, Lemma 4.9 in [Hauswirth et al. 2009] (see also the lemma on page 139 of [Finn 1965]) shows that $\{u_n\}$ is uniformly bounded in a neighborhood of p in D. Then, a divergence line cannot end at p.

If $\kappa(E) = 2H$ at p, by [Hauswirth et al. 2009, Lemma 4.9], we have that the sequence $\{u_n\}$ does not diverge to $+\infty$ in a neighborhood of p. Suppose there is one divergence line L leaving p. Then there is a subset $V \subset D$ which contains a subarc (containing p) of E in its boundary, and the sequence diverges to $-\infty$ on V. Consider a point $q \in E \cap \partial V$, and denote by $\overline{pq}$ the arc contained in E joining the points p and q. Let s be a point in L and $\overline{ps}$ the arc in L joining p and s. Denote by $\overline{sq}$ the geodesic joining s and q, suppose that q is as close to s as necessary, in order to guarantee $\overline{sq} \subset V$. We choose this "triangle" T so that the sequence $\{u_n\}$ diverges to $-\infty$ in the domain $\Delta_T \subset V$ bounded by T. By the flux formulas,

$$2H\mathcal{A}(\Delta_T) = F_{u_n}(\overline{ps}) + F_{u_n}(\overline{pq}) + F_{u_n}(\overline{sq}).$$

We have

$$\lim_{n\to+\infty} F_{u_n}(\overline{pq}) = |\overline{pq}|.$$

Since $\overline{ps} \subset L$, either

$$\lim_{n\to+\infty} F_{u_n}(\overline{ps}) = |\overline{ps}| \quad \text{or} \quad \lim_{n\to+\infty} F_{u_n}(\overline{ps}) = -|\overline{ps}|.$$

First, suppose that

$$\lim_{n\to+\infty} F_{u_n}(\overline{ps}) = |\overline{ps}|.$$

Then,

$$\begin{aligned} \lim_{n\to+\infty} 2H\mathcal{A}(\Delta_T) &= \lim_{n\to+\infty} F_{u_n}(\overline{ps}) + \lim_{n\to+\infty} F_{u_n}(\overline{pq}) + \lim_{n\to+\infty} F_{u_n}(\overline{sq}) \\ &\geq |\overline{ps}| + |\overline{pq}| - |\overline{sq}| \end{aligned}$$

which implies

$$\frac{2H\mathscr{A}(\Delta_T)}{|\overline{sq}|} \geq \frac{|\overline{ps}|+|\overline{pq}|}{|\overline{sq}|} - 1.$$

We move q to q' and s to s' so that $|\overline{pq}'| = \lambda|\overline{pq}|$ and $|\overline{ps}'| = \lambda|\overline{ps}|$. When $\lambda \to 0$, the inequality

$$\frac{2H\mathscr{A}(\Delta_T)}{|\overline{sq}|} \geq \frac{|\overline{ps}|+|\overline{pq}|}{|\overline{sq}|} - 1$$

tends to zero on the left side, but is bounded from zero in the right side; a contradiction.

Now we consider the case where

$$\lim_{n\to+\infty} F_{u_n}(\overline{ps}) = -|\overline{ps}|.$$

By Lemma 6.4 we have that $\{u_n\}$ diverges to $-\infty$ on a subset of $D - V$ which has L and a subarc of E in its boundary. Then applying the same argument as above, we get a contradiction.

Now, suppose that there are two or more divergence lines leaving from p. We fix two divergence lines, L_1, L_2. The point $p \in E$ divides E in two curves E_1, E_2, and we orient L_1, L_2, E_1, E_2 such that W_1 is the domain bounded in part by $L_1 \cup E_1$ and not containing L_2, W_2 is the domain bounded in part by $E_2 \cup L_2$ and not containing L_1 and finally W_3 is the domain bounded in part by $L_1 \cup L_2$ and not containing $E_1 \cup E_2$. Let $q \in L_1$, $s \in L_2$, $p_1 \in E$, $p_2 \in E$ be points. Denote by $\overline{pq}$ the segment in L_1 joining p and q, by $\overline{ps}$ the segment in L_2 joining p and s, by $\overline{sq} \subset W_3$ the segment of the geodesic joining q to s, by $\overline{qp_1} \subset W_1$ the segment of the geodesic joining q and p_1, and by $\overline{sp_2} \subset W_2$ the segment of the geodesic joining s and p_2. In some of these subsets W_i, $i = 1, 2, 3$, the sequence $\{u_n\}$ diverges to $-\infty$. Suppose that in W_3 the sequence diverges to $-\infty$, and that $\overline{sq} \subset W_3$.

If either

$$\lim_{n\to+\infty} F_{u_n}(\overline{ps}) = |\overline{ps}| \quad \text{or} \quad \lim_{n\to+\infty} F_{u_n}(\overline{pq}) = |\overline{pq}|,$$

with respect to W_3, applying the flux formulas to the triangle formed by $\overline{ps}$, $\overline{pq}$ and $\overline{sq}$, we obtain a contradiction as before.

If, with respect to W_3, either

$$\lim_{n\to+\infty} F_{u_n}(\overline{ps}) = -|\overline{ps}| \quad \text{or} \quad \lim_{n\to+\infty} F_{u_n}(\overline{pq}) = -|\overline{pq}|,$$

then doing as we have done before to the triangle formed by $\overline{qp_1}$, $\overline{pq}$ and $\overline{p_1p}$, if $\lim_{n\to+\infty} F_{u_n}(\overline{pq}) = -|\overline{pq}|$, or to the triangle formed by $\overline{ps}$, $\overline{pp_2}$ and $\overline{sp_2}$ if $\lim_{n\to+\infty} F_{u_n}(\overline{ps}) = -|\overline{ps}|$, we obtain a contradiction. $\square$

7. Proof of the main theorems

Before the proof of the theorems we need to show that the conditions of the hypothesis make sense, that is, we have to show that they are preserved for smaller horocycles.

Let $\mathcal{H}_i$ be an horocycle at d_i. Suppose that the conditions of Theorems 3.1 and 3.2 are satisfied for a family of horocycles $\mathcal{H} = \{\mathcal{H}_i\}_{i=1,\dots,n}$. These conditions are

(i) $\alpha(\partial\mathcal{D}) - \beta(\partial\mathcal{D}) = 2H\widetilde{\mathcal{A}}(\mathcal{D})$,

and for all admissible polygons $\mathcal{P} \neq \partial\mathcal{D}$,

(ii) $2\alpha(\mathcal{P}) < l(\mathcal{P}) + 2H\widetilde{\mathcal{A}}(\Omega)$,

(iii) $2\beta(\mathcal{P}) < l(\mathcal{P}) - 2H\widetilde{\mathcal{A}}(\Omega)$.

Fixing $s \in \{1, \dots, n\}$, we will show that these conditions are also true for a family $\mathcal{H}' = \{\mathcal{H}_i\}_{i\neq s} \cup \{\mathcal{H}'_s\}$, where $\mathcal{H}'_s$ is contained in the horodisk F_s bounded by $\mathcal{H}_s$. We are interested in "smaller" horocycles because in this way we have an exhaustion of $\mathcal{P}$. To prove this we will use subindices T and T' to clarify the dependence of $\alpha(\mathcal{P})$, $\beta(\mathcal{P})$ and $l(\mathcal{P})$ with respect to $\mathcal{H}$ and $\mathcal{H}'$ respectively.

First, consider condition (i). We observe that when we change the family of horocycles, the left side of (i) does not change. So our definition for $\widetilde{\mathcal{A}}$ should not change. This is the first reason for the definition of $\widetilde{\mathcal{A}}$.

Note that

$$\alpha(\partial\mathcal{D}_{T'}) - \beta(\partial\mathcal{D}_{T'}) = \alpha(\partial\mathcal{D}_T) - \beta(\partial\mathcal{D}_T) = \text{constant}.$$

Thus, if (i) is true for $\mathcal{H}$, then it is also true for $\mathcal{H}'$.

Condition (ii) is equivalent to

$$2\alpha(\mathcal{P}) - l(\mathcal{P}) < 2H\widetilde{\mathcal{A}}(\Omega).$$

When we change from family $\mathcal{H}$ to family $\mathcal{H}'$ the left side of the above inequality is nonincreasing and the right side is nondecreasing, so the inequality is preserved.

Finally, we handle the inequality of condition (iii).

There are two distinct situations. The first one is when the horocycle $\mathcal{H}_s$ meets sides E_1, E_2 where $\kappa(E_1) = -2H$, $\kappa(E_2) = 2H$. The second one is when $\mathcal{H}_s$ meets sides E_1, E_2 with $\kappa(E_1) = 2H$, $\kappa(E_2) = 2H$.

In the first case, the area $\widetilde{\mathcal{A}}(\Omega)$ does not change when we change from the family $\mathcal{H}$ to $\mathcal{H}'$, and $2\beta(\mathcal{P}) - l(\mathcal{P})$ is nonincreasing, so the inequality is preserved.

The second case is the most delicate one. Here, it will be necessary to have horocycles small enough.

More precisely, we consider the half-space model of $\mathbb{H}$. We can suppose that the vertices of $\mathcal{P}$ are $d_j = (x_j, 0)$ for all $j \neq l$ and $d_l \in \{\partial\mathbb{H} - \{y = 0\}\}$. We choose the family $\{\mathcal{H}_i\}$ of horocycles at the vertices d_i. We define

$$(0, M_l) = \mathcal{H}_l \cap \{x = 0\}.$$

The necessary condition is

$$M_l > \frac{2H(|x_{l-1}| + |x_{l+1}|)}{2\sqrt{1-4H^2}} \quad \text{for all } l = 1, \dots, n. \tag{5}$$

Remark 7.1. This is always the case for sufficiently small horocycles.

With this hypothesis on the horocycles, we can finish that the inequality in (iii) is preserved for the family $\mathcal{H}'$.

Suppose that $\mathcal{H}_s$ meets sides E_1 and E_2, where $\kappa(E_1) = \kappa(E_2) = 2H$. We point out that this is the case where we use (5) and also the definition of $\widetilde{\mathcal{A}}$, since $\widetilde{\mathcal{A}}$ should have the right behavior as the area is infinite.

Note that

$$2\beta(\mathcal{P}_{T'}) = 2\beta(\mathcal{P}_T) < l(\mathcal{P}_T) - 2H\widetilde{\mathcal{A}}(\Omega_T).$$

We will show

$$l(\mathcal{P}_T) - 2H\widetilde{\mathcal{A}}(\Omega_T) < l(\mathcal{P}_{T'}) - 2H\widetilde{\mathcal{A}}(\Omega_{T'}),$$

that is,

$$\big(l(\mathcal{P}_{T'}) - l(\mathcal{P}_T)\big) - \big(2H\widetilde{\mathcal{A}}(\Omega_{T'}) - 2H\widetilde{\mathcal{A}}(\Omega_T)\big) > 0. \tag{6}$$

In fact, we show that $l(\mathcal{P}_T) - 2H\widetilde{\mathcal{A}}(\Omega_T)$ increases when $\mathcal{H}$ decreases.

Consider the half-space model of $\mathbb{H}$. We can assume that $d_s = (0, 0) \in \partial_\infty\mathbb{H}$. Using an inversion I with respect to the geodesic centered at $(0, 0)$ of radius 1, we have that $\mathcal{H}_s$ and $\mathcal{H}'_s$ are taken to the horizontal straight lines through $(0, M)$ and $(0, y_0)$, respectively, and the sides A and E are taken to tilted lines leaving the points $(-x_0, 0)$ and $(x_1, 0)$ and making an angle θ with the vertical, where $\sin\theta = 2H$, $x_0 > 0$ and $x_1 > 0$; see Figure 2.

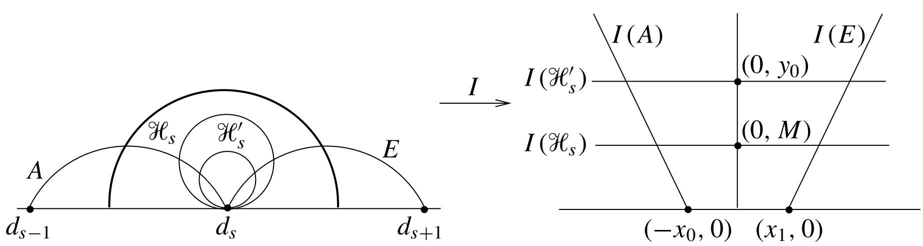


Figure 2. Using the inversion I.

Now, we calculate the length of the arcs of $I(A)$ and $I(E)$ bounded by $I(\mathcal{H}'_s)$ and $I(\mathcal{H}_s)$, denoted by $l(A_{\mathcal{H},\mathcal{H}'})$ and $l(E_{\mathcal{H},\mathcal{H}'})$, and the area limited by $I(A)$, $I(E)$, $I(\mathcal{H}_s)$ and $I(\mathcal{H}'_s)$, denoted by $\mathcal{A}(\Omega_{\mathcal{H},\mathcal{H}'})$.

Then,

$$l(A_{\mathcal{H},\mathcal{H}'}) = l(E_{\mathcal{H},\mathcal{H}'}) = \int_M^{y_0} \frac{\sec\theta}{y}\,dy = \sec\theta \ln y\Big|_M^{y_0}$$

and the area satisfies

$$\begin{aligned}\mathcal{A}(\Omega_{\mathcal{H},\mathcal{H}'}) &= \int_M^{y_0}\int_{-x_0-y\tan\theta}^{x_1+y\tan\theta} \frac{dxdy}{y^2}\\ &= \int_M^{y_0}\Big(\frac{2\tan\theta}{y} + \frac{(x_1+x_0)}{y^2}\Big)\,dy\\ &= 2\tan\theta\ln y\Big|_M^{y_0} - (x_1+x_0)\frac{1}{y}\Big|_M^{y_0}.\end{aligned}$$

Therefore,

$$\begin{aligned}&l(A_{\mathcal{H},\mathcal{H}'}) + l(E_{\mathcal{H},\mathcal{H}'}) - 2H\mathcal{A}(\Omega_{\mathcal{H},\mathcal{H}'})\\ &\qquad = 2(\sec\theta - 2H\tan\theta)\ln y\Big|_M^{y_0} + 2H\frac{(x_1+x_0)}{y}\Big|_M^{y_0}\\ &\qquad = 2\Big(\frac{1-\sin^2\theta}{\cos\theta}\Big)\ln y\Big|_M^{y_0} + \frac{2H(x_1+x_0)}{y}\Big|_M^{y_0}\\ &\qquad = 2\cos\theta\ln y_0 + \frac{2H(x_1+x_0)}{y_0} - 2\cos\theta\ln M - \frac{2H(x_1+x_0)}{M}.\end{aligned}$$

Then, to prove the inequality (6) it suffices to show that the function of y_0 above is increasing, because when $y_0 = M$, it is zero. We show that its derivative is greater than zero.

Differentiating we have

$$\frac{2\cos\theta}{y_0} - \frac{2H(x_1+x_0)}{y_0^2}.$$

So

$$\frac{2\cos\theta}{y_0} - \frac{2H(x_1+x_0)}{y_0^2} > 0 \iff 2y_0\cos\theta - 2H(x_1+x_0) > 0,$$

that is,

$$y_0 > \frac{2H(x_1+x_0)}{2\cos\theta}.$$

But our family $\mathcal{H}$ satisfies

$$M > \frac{2H(x_1+x_0)}{2\cos\theta}.$$

Thus, we have the inequality (6) as desired, and consequently the inequality in (iii) is satisfied.

We fix some notation which will be useful in the proof of the theorems. Let $\{d_i = (x_i, y_i)\}$ be the set of vertices of $\partial\mathcal{D}$. For each i, let $\mathcal{H}_i(n)$ be a horocycle asymptotic to d_i such that $\mathcal{H}_i(n)$ belongs to Θ for all i, n. We choose $\mathcal{H}_i(n)$ such that $\mathcal{H}_i(n+1) \subset F_i(n)$, where $F_i(n)$ is the convex horodisk bounded by $\mathcal{H}_i(n)$. Let $\mathcal{D}(n) \subset \mathcal{D}$ be the domain bounded by

$$\partial\mathcal{D}(n) = \big(\partial\mathcal{D} - \big(\textstyle\bigcup_i F_i(n)\big)\big) \cup \big(\textstyle\bigcup_i \gamma_i(n)\big),$$

where $\gamma_i(n) = \mathcal{H}_i(n) \cap (\partial\mathcal{D} \cup \mathcal{D})$. Let $\mathcal{D}^*(n) \subset \mathcal{D}$ be the domain bounded by

$$\partial\mathcal{D}^*(n) = \big(\partial\mathcal{D} - \big(\textstyle\bigcup_i F_i(n)\big)\big) \cup \big(\textstyle\bigcup_i \gamma_i^*(n)\big),$$

where $\gamma_i^*(n)$ is the geodesic reflection of $\gamma_i(n)$. Similarly, we define $\Omega(n)$ as the domain whose boundary is

$$\mathcal{P}(n) = \big(\mathcal{P} - \big(\textstyle\bigcup_i F_i(n)\big)\big) \cup \big(\textstyle\bigcup_i \gamma_i(n) \cap \Omega(n)\big)$$

and $\Omega^*(n)$ as the domain bounded by

$$\partial\Omega^*(n) = \big(\mathcal{P} - \big(\textstyle\bigcup_i F_i(n)\big)\big) \cup \big(\textstyle\bigcup_i (\gamma_i^*(n) \cap \Omega^*(n))\big).$$

Finally, given an arc $\eta \subset \mathcal{P}$, we define $\eta(n) = \eta \cap \mathcal{P}(n)$.

Proof of Theorem 3.1. Suppose that the conditions (2) and (3) are true for all polygons in $\mathcal{D}$.

Claim 7.1. *There is a solution in $\mathcal{D}$ which boundary values*

$$u_n = \begin{cases} n & \text{on } \bigcup_k A_k, \\ -n & \text{on } \bigcup_l B_l^*. \end{cases}$$

Assume this Claim is true and take $\{u_n\}$ a sequence of solutions in $\mathcal{D}$, where u_n is defined as in the Claim. Then, this sequence has, or does not have, a divergence line.

First, we assume that there is some divergence line, and we will obtain a contradiction. By Lemma 6.5, the endpoints of these lines are among vertices of $\mathcal{D}$. Since $\partial\mathcal{D}$ has only a finite number of vertices, we can suppose that the divergence set is composed of a finite number of disjoint divergence lines. These lines separate the domain $\mathcal{D}$ in at least two connected components, and the interior of these components belongs to the convergence domain. By Lemma 6.4, in some connected components of the convergence set, the sequence $\{u_n\},\ p \in \mathcal{D}$, diverges to $+\infty$ or $-\infty$. Suppose that in some connected component of the convergent set $\mathcal{U}'$, the sequence diverges to $+\infty$ (the case $-\infty$ is similar).

Since $\mathcal{U}' \subset \mathcal{U}$, where $\mathcal{U}$ is the convergence domain, we have that the sequence $\{u_n - u_n(p)\},\ p \in \mathcal{U}'$, converges uniformly on compact subsets of $\mathcal{U}'$ to a solution u in $\mathcal{U}'$. On the other hand, by the choice of $\mathcal{U}'$ we have $u_n(p) \to +\infty,\ p \in \mathcal{U}'$. Moreover, we note that $\partial\mathcal{U}' = \mathcal{P}$ is an admissible polygon, we can choose $\mathcal{P}$ satisfying the next Claim.

Claim 7.2. *One can choose $\mathcal{P}$ so that*

$$F_u\Big(\mathcal{P}(n) - \Big(\big(\bigcup_i A_i(n)\big) \cup \big(\bigcup_i (\gamma_i(n) \cap \mathcal{U}')\big)\Big)\Big) = -l\Big(\mathcal{P}(n) - \Big(\big(\bigcup_i A_i(n)\big) \cup \big(\bigcup_i (\gamma_i(n) \cap \mathcal{U}')\big)\Big)\Big),$$

where $\partial\mathcal{U}' = \mathcal{P}$.

See [Mazet et al. 2008] for a proof.

We are supposing that there is a divergence line, so $\mathcal{P} \neq \partial\mathcal{D}$. By Claim 7.2 and the flux formulas

$$\begin{aligned}
F_u(\mathcal{P}(n)) &= 2H\mathcal{A}(\mathcal{U}'(n)) \\
&= F_u\Big(\mathcal{P}(n) - \Big(\big(\bigcup_i A_i(n)\big) \cup \big(\bigcup_i (\gamma_i(n) \cap \mathcal{U}')\big)\Big)\Big) \\
&\qquad + F_u\Big(\big(\bigcup_i A_i(n)\big) \cup \big(\bigcup_i (\gamma_i(n) \cap \mathcal{U}')\big)\Big) \\
&\le -l\Big(\mathcal{P}(n) - \Big(\big(\bigcup_i A_i(n)\big) \cup \big(\bigcup_i (\gamma_i(n) \cap \mathcal{U}')\big)\Big)\Big) \\
&\qquad + l\Big(\big(\bigcup_i A_i(n)\big) \cup \big(\bigcup_i (\gamma_i(n) \cap \mathcal{U}')\big)\Big) \\
&= 2\alpha(\mathcal{P}) - l(\mathcal{P}) + l\big(\bigcup_i (\gamma_i(n) \cap \mathcal{U}')\big).
\end{aligned}$$

When $n \to \infty$, the area $\mathcal{A}\big(\mathcal{D} \cap \big(\bigcup_j F_j\big)\big)$ tends to zero, so

$$2H\mathcal{A}(\mathcal{U}') \le 2\alpha(\mathcal{P}) - l(\mathcal{P}),$$

contradicting the hypothesis. So the sequence $\{u_n\}$ has no divergence lines.

Since the sequence $\{u_n\}$ does not have any divergence lines, $\mathcal{D}$ is the convergence domain, so there is a subsequence of $\{u_n - u_n(p)\}$, $p \in \mathcal{D}$ which converges to a solution u on $\mathcal{D}$. If the sequence $\{u_n\}$ is bounded at the point $p \in \mathcal{D}$, u has the boundary values as desired, that is, $u|_{A_k} = +\infty$ and $u|_{B_l} = -\infty$. We will show that even if the sequence $\{u_n\}$ is unbounded, the solution u has the boundary values as prescribed.

Suppose the sequence $\{u_n(p)\}$ tends to $-\infty$. By the flux formulas,

$$\begin{aligned}
\lim_{n\to\infty} F_{u_n}(\mathcal{P}(m)) &= 2H\mathcal{A}(\mathcal{D}(m)) \\
&= 2H\widetilde{\mathcal{A}}(\mathcal{D}) - 2H\mathcal{A}\big(\mathcal{D} \cap \big(\bigcup_i F_i(m)\big)\big) \\
&= \sum \lim_{n\to\infty} F_{u_n}(A_i(m)) + \sum \lim_{n\to\infty} F_{u_n}(B_i(m)) \\
&\qquad + \sum \lim_{n\to\infty} F_{u_n}(\gamma_i(m)) \\
&\ge \alpha(\mathcal{P}) - \beta(\mathcal{P}) - \sum |\gamma_i(m)|
\end{aligned}$$

which implies

$$2\beta(\mathcal{P}) \ge l(\mathcal{P}) - 2H\mathcal{A}(\Omega).$$

The hypothesis does not allow $2\beta(\mathcal{P}) > l(\mathcal{P}) - 2H\mathcal{A}(\Omega)$. Then equality holds: $2\beta(\mathcal{P}) = l(\mathcal{P}) - 2H\mathcal{A}(\Omega)$. This implies that $\lim_{n\to\infty} F_{u_n}(B_l(m)) = |B_l(m)|$. So $\{u_n - u_n(p)\}$ tends to $-\infty$ on B_l for all l.

Suppose the sequence $\{u_n(p)\}$ tends to $+\infty$. By the flux formulas,

$$\begin{aligned}\lim_{n\to\infty} F_{u_n}(\mathcal{P}(m)) &= 2H\mathcal{A}(\mathcal{D}(m)) = 2H\widetilde{\mathcal{A}}(\mathcal{D}) - 2H\mathcal{A}\big(\mathcal{D}\cap\big(\textstyle\bigcup_i F_i(m)\big)\big)\\ &= \sum \lim_{n\to\infty} F_{u_n}(A_i(m)) + \sum \lim_{n\to\infty} F_{u_n}(B_i(m)) \lim_{n\to\infty} F_{u_n}(\gamma_i(m))\\ &\le \alpha(\mathcal{P}) - \beta(\mathcal{P}) + \sum |\gamma_i(m)|,\end{aligned}$$

which implies

$$2\alpha(\mathcal{P}) \ge l(\mathcal{P}) + 2H\mathcal{A}(\Omega).$$

Since we cannot have $2\alpha(\mathcal{P}) > l(\mathcal{P}) + 2H\mathcal{A}(\Omega)$, we have $2\alpha(\mathcal{P}) = l(\mathcal{P}) + 2H\mathcal{A}(\Omega)$, which implies $\lim_{n\to\infty} F_{u_n}(A_k(m)) = |A_k(m)|$. Then $\{u_n - u_n(p)\}$ tends to $+\infty$ on A_k for all k.

Proof of Claim 7.1. By the existence theorem for continuous boundary values and bounded domains [Hauswirth et al. 2009], for each m in $\mathcal{D}^*(m)$ there is a solution with boundary values

$$u_m = \begin{cases} n & \text{on } \bigcup_k A_k(m),\\ -n & \text{on } \bigcup_l B_l^*(m),\\ 0 & \text{on } \bigcup_i \gamma_i^*(m).\end{cases}$$

Fix m_0. For all $m > m_0$, we have that $\{u_m|_{\mathcal{D}^*(m_0)}\}$ is a sequence of solutions in $\mathcal{D}^*(m_0)$. If there were any divergence lines, we would find a divergence set which would contradict the hypothesis, as in the proof of Theorem 3.1. Moreover, as there are no divergence lines, either this sequence is bounded or it is not bounded. If this sequence is not bounded, say $u_m(p) \to +\infty,\ p \in \mathcal{D}^*(m_0)$, a subsequence $\{u_m|_{\mathcal{D}^*(m_0)} - u_m(p)\}$ converges to a solution in $\mathcal{D}^*(m_0)$ and tends to $-\infty$ on each arc $A_i(m_0)$, which cannot occur. If $\{u_m(p)\} \to -\infty,\ p \in \mathcal{D}^*(m_0)$, some subsequence of $\{u_m|_{\mathcal{D}^*(m_0)} - u_m(p)\}$ converges to a solution in $\mathcal{D}^*(m_0)$ and tends to $+\infty$ on each arc $A_i(m_0)$, $B_l^*(m_0)$. Taking $m_0 \to \infty$ we again get a contradiction, since two arcs with the same vertex point have values $+\infty$. So this sequence is bounded and some subsequence is convergent, by the boundary values of the $\{u_m\}$, we have $u_m|_{A_k(m_0)} = n$ and $u_m|_{B_l(m_0)} = -n$. By the diagonal process, we have in $\mathcal{D}$ a solution u_n given by

$$u_n = \begin{cases} n & \text{on } \bigcup_k A_k,\\ -n & \text{on } \bigcup_l B_l^*,\end{cases}$$

which completes the proof. □

We return to the proof of Theorem 3.1 and prove the necessary conditions. Suppose there is a solution u in $\mathcal{D}$ of the Dirichlet problem. Applying the flux formulas to $\mathcal{P}(n)=\partial\mathcal{D}(n)$, and remembering that, in this case, $\widetilde{\mathcal{A}}=\mathcal{A}$, we have

$$\begin{aligned}F_u(\mathcal{P}(n))=2H\mathcal{A}(\mathcal{D}(n))&=2H\widetilde{\mathcal{A}}(\mathcal{D})-2H\mathcal{A}\big(\mathcal{D}\cap\big(\textstyle\bigcup_i F_i(n)\big)\big)\\&=\textstyle\sum F_u(A_i(n))+\sum F_u(B_i(n))+\sum F_u(\gamma_i(n)).\end{aligned}$$

Since $\widetilde{\mathcal{D}}=\mathcal{D}\cap\big(\bigcup_i F_i(n)\big)$,

$$\begin{aligned}\textstyle\sum|A_i(n)|-\sum|B_i(n)|-\sum|\gamma_i(n)|&\le 2H\widetilde{\mathcal{A}}(\mathcal{D})-2H\mathcal{A}(\widetilde{\mathcal{D}})\\&\le\textstyle\sum|A_i(n)|-\sum|B_i(n)|+\sum|\gamma_i(n)|.\end{aligned}$$

It follows that

$$\alpha(\mathcal{D})-\beta(\mathcal{D})-\textstyle\sum|\gamma_i(n)|\le 2H\widetilde{\mathcal{A}}(\mathcal{D})-2H\mathcal{A}(\widetilde{\mathcal{D}})\le\alpha(\mathcal{D})-\beta(\mathcal{D})+\sum|\gamma_i(n)|.$$

When $n\to\infty$, we have $|\gamma_i|\to 0$ and $\mathcal{A}(\widetilde{\mathcal{D}})\to 0$, so $\alpha(\mathcal{D})-\beta(\mathcal{D})=2H\widetilde{\mathcal{A}}(\mathcal{D})$. Now, we prove the inequalities (3). Applying the flux formulas to the polygon $\mathcal{P}(n)$, and denoting its interior arcs by E_m, we have

$$\begin{aligned}F_u(\mathcal{P}(n))&=2H\mathcal{A}(\Omega(n))\\&=\sum_k F_u(A_k(n))+\sum_l F_u(B_l(n))+\sum_m F_u(E_m(n))+\sum_j F_u(\gamma_j(n)\cap\Omega(n))\\&\ge\sum_k|A_k(n)|-\sum_l|B_l(n)|+\delta-\sum|E_m(n)|-\sum_j|\gamma_j(n)\cap\Omega(n)|\\&=2\alpha(\mathcal{P})-l(\mathcal{P})+\delta-\sum_j|\gamma_j(n)\cap\Omega(n)|.\end{aligned}$$

We see that $\widetilde{\mathcal{A}}(\Omega)>\mathcal{A}(\Omega(n))$ and $\sum_j|\gamma_j(n)\cap\Omega(n)|-\delta<0$ for n large enough, so

$$2\alpha(\mathcal{P})<l(\mathcal{P})+2H\widetilde{\mathcal{A}}(\Omega).$$

Similarly,

$$\begin{aligned}F_u(\mathcal{P}(n))&=2H\mathcal{A}(\Omega(n))\\&=\sum_k F_u(A_k(n))+\sum_l F_u(B_l(n))+\sum_m F_u(E_m(n))+\sum_j F_u(\gamma_j(n)\cap\Omega)\\&\le\sum_k|A_k(n)|-\sum_l|B_l(n)|-\delta+\sum_m|E_m(n)|+\sum_j|\gamma_j(n)\cap\Omega|\\&=-2\beta(\mathcal{P})+l(\mathcal{P})-\delta;\end{aligned}$$

that is, for n sufficiently large,

$$\begin{aligned}2\beta(\mathcal{P})&\le l(\mathcal{P})-2H\mathcal{A}(\Omega(n))-\delta\\&<l(\mathcal{P})-2H\widetilde{\mathcal{A}}(\Omega).\end{aligned}$$

□

Proof of Theorem 3.2. This is similar to the proof of Theorem 3.1.

Claim 7.3. *There is a solution on $\mathfrak{D}$ having boundary values*

$$u_n = \begin{cases} n & \text{on } A_k, \\ -n & \text{on } B_l^*, \\ f_n & \text{on } C_m, \end{cases}$$

where $f_n = \varphi \circ f$ for $\varphi : \mathbb{R} \to \mathbb{R}$ defined by

$$\varphi(x) = \begin{cases} x & \text{if } -n \le x \le n, \\ -n & \text{if } x < -n, \\ n & \text{if } x > n. \end{cases}$$

Assume that Claim 7.3 is true and take a sequence $\{u_n\}$ on $\mathfrak{D}$ given by this claim.

Suppose that $\{u_n\}$ has a divergence line. By Lemma 6.5, we can suppose that the divergence set is composed of a finite number of disjoint divergence lines. These lines separate the domain $\mathfrak{D}$ in at least two connected components, and the interior of these components belongs to the convergence domain. By Lemma 6.4, in connected components of the convergence set the sequence $\{u_n\}$, $p \in \mathfrak{D}$, diverges to $+\infty$ or $-\infty$. We observe that if there is some arc $C \subset \partial\mathfrak{D}$ having $\kappa(C) > 2H$, Lemma 4.9 in [Hauswirth et al. 2009] ensures that in a neighborhood of this arc the sequence $\{u_n\}$ is bounded.

As in the proof of Theorem 3.1 we will work on subdomains of $\mathfrak{D}$ where the sequence diverges to $+\infty$ or $-\infty$, so the boundary of these domains only has arcs of curvature $2H$. This means that the boundary of these domains are admissible polygons. From now on, the proof is similar to the proof of Theorem 3.1.

Proof of Claim 7.3. The only difference between Claim 7.3 and Claim 7.1 is found in the construction of solutions over bounded domains. Let $\{d_i\}$ be the vertices points of $\mathfrak{D}$, after some isometry of the hyperbolic plane, we can assume that each d_i belongs to $\{(x, y) \in \mathbb{R}^2 \mid y = 0\}$. Let $\sigma_i[m]$ be geodesics which are semicircles centered at d_i with radius $1/m$. The hypothesis on the curvature of the arcs C_i enables us to conclude that, if m is big enough, $\sigma_i[m]$ divides $\mathfrak{D}$ in exactly two components, one of them having d_i in its asymptotic boundary. Let $\varrho_i[m]$ be the arc of the equidistant curve to $\sigma_i[m]$ having curvature $2H$ joining points of the boundary of $\mathfrak{D}$. Then $\varrho_i[m]$ divides $\mathfrak{D}$ in exactly two components, one having d_i in its asymptotic boundary. We chose the curvature vector of $\varrho_i[m]$ pointing to the component of $\mathfrak{D}$ which does not have d_i on its boundary. Now we can find a solution with prescribed boundary values using the existence theorem of [Hauswirth et al. 2009]. Let $A_i[m]$ be the compact arcs contained in A_i bounded by the endpoints of $\{\varrho_i[m]\}$, $B_i[m]$ be the compact arcs contained in B_i bounded by the endpoints of $\{\varrho_i[m]\}$ and $C_i[m]$ be the compact arcs contained in C_i bounded by the endpoints of $\{\varrho_i[m]\}$. So there exists

$$u_n = \begin{cases} n & \text{on } A_i[m], \\ -n & \text{on } B_i^*[m], \\ f_n & \text{on } C_i[m], \\ 0 & \text{on } \varrho_i[m], \end{cases}$$

where $f_n = \varphi \circ f$, for $\varphi : \mathbb{R} \to \mathbb{R}$ given by

$$\varphi(x) = \begin{cases} x & -n \le x \le n, \\ -n & x < -n, \\ n & x > n. \end{cases}$$

From now on, the same procedure as in Claim 7.1 enables us to conclude the existence of a solution over $\mathcal{D}$ as desired in Claim 7.3. $\square$

Now, we go back to the proof of Theorem 3.2. Suppose that there is a solution u for the Dirichlet problem. Let Ω be the domain bounded by the admissible polygon $\mathcal{P}$ and $\Omega(n), \mathcal{P}(n)$ as found in the notation at the beginning of this section. Applying the flux formulas,

$$\begin{aligned} &F_u(\mathcal{P}(n)) \\ &= 2H\mathcal{A}(\Omega(n)) \\ &= \sum_k F_u(A_k(n)) + \sum_l F_u(B_l(n)) + \sum_p F_u(C_p(n)) + \sum_m F_u(E_m(n)) + \sum_j F_u(\gamma_j(n)\cap\Omega) \\ &\ge \sum_k |A_k(n)| - \sum_l |B_l(n)| - \sum_p |C_p(n)| + \delta - \sum |E_m(n)| - \sum_j |\gamma_j(n)\cap\Omega| \\ &= 2\alpha(\mathcal{P}) - l(\mathcal{P}) + \delta - \sum_j |\gamma_j(n)\cap\Omega|. \end{aligned}$$

Either $\mathcal{A}(\mathcal{D}) < \infty$, or $\mathcal{A}(\mathcal{D}) = \infty$. If $\mathcal{A}(\mathcal{D}) < \infty$, since $\widetilde{\mathcal{A}}(\Omega) > \mathcal{A}(\Omega_{\mathcal{H}})$ and $|\gamma_j(n)| \to 0$ for all j, we have

$$2\alpha(\mathcal{P}) < l(\mathcal{P}) + 2H\mathcal{A}(\Omega(n)) < l(\mathcal{P}) + 2H\widetilde{\mathcal{A}}(\Omega).$$

If $\mathcal{A}(\mathcal{D}) = \infty$, we have

$$2H\widetilde{\mathcal{A}}(\Omega) \ge 2H\mathcal{A}(\Omega_{\mathcal{H}}) > 2\alpha(\mathcal{P}) - l(\mathcal{P}) - \sum_j |\gamma_j(n)\cap\Omega|,$$

Then,

$$2H\widetilde{\mathcal{A}}(\Omega) + l(\mathcal{P}) - 2\alpha(\mathcal{P}) > -\sum_j |\gamma_j(n)\cap\Omega|.$$

Remembering that $l(\mathcal{P}) - 2\alpha(\mathcal{P})$ is nondecreasing, we have that the left side of this inequality is increasing and tends to $+\infty$, when the horocycles tend to the vertices.

Therefore, we can suppose

$$2H\widetilde{\mathcal{A}}(\Omega)+l(\mathcal{P})-2\alpha(\mathcal{P})>0.$$

Similarly,

$$\begin{aligned}
&F_u(\mathcal{P}(n))\\
&=2H\mathcal{A}(\Omega_{\mathcal{H}})\\
&=\sum_k F_u(A_k(n))+\sum_l F_u(B_l(n))+\sum_p F_u(C_p(n))+\sum_m F_u(E_m(n))+\sum_j F_u(\gamma_j(n)\cap\Omega)\\
&\quad\le \sum_k |A_k(n)|-\sum_l |B_l(n)|+\sum_p |C_p(n)|-\delta+\sum_m |E_m(n)|+\sum_j |\gamma_j(n)\cap\Omega|\\
&=-2\beta(\mathcal{P})+l(\mathcal{P})-\delta+\sum_j |\gamma_j(n)\cap\Omega|.
\end{aligned}$$

Then, if $\mathcal{A}(\mathcal{D})<\infty$,

$$2\beta(\mathcal{P})<l(\mathcal{P})-2H\mathcal{A}(\Omega_{\mathcal{H}})-\frac{\delta}{2}\le l(\mathcal{P})-2H\widetilde{\mathcal{A}}(\Omega),$$

since we can choose $2H\mathcal{A}(\Omega\cap(\cup_i F_i))\le\frac{\delta}{2}$ and $\sum_j|\gamma_j(n)\cap\Omega|<\frac{\delta}{2}$.

If $\mathcal{A}(\mathcal{D})=\infty$,

$$2H\widetilde{\mathcal{A}}(\Omega)+2\beta(\mathcal{P})-l(\mathcal{P})<2H\mathcal{A}(\Omega(n))+2\beta(\mathcal{P})-l(\mathcal{P})\le\sum_j|\gamma_j(n)|,$$

because we can choose $2H\mathcal{A}(\widetilde{\Omega})<\delta$. Since $2H\widetilde{\mathcal{A}}(\Omega)+2\beta(\mathcal{P})-l(\mathcal{P})$ tends to $-\infty$ when the horocycles converge to vertices, we can suppose

$$2H\widetilde{\mathcal{A}}(\Omega)+2\beta(\mathcal{P})-l(\mathcal{P})<0. \qquad \square$$

8. Example

Consider a domain $\mathcal{D}$ whose boundary has sides A_1, B_1, A_2 and B_2 and vertices $d_1=(x_{d_1},0)$, $d_2=(x_{d_2},0)$, $d_3=(x_{d_3},0)$ and $d_4\in\{\partial_\infty\mathbb{H}-y=0\}$ with $x_{d_1}<x_{d_2}<x_{d_3}$. Suppose that the vertices of A_1 are d_4 and d_1, the vertices of B_1 are d_1 and d_2, the vertices of A_2 are d_2 and d_3 and the vertices of B_2 are d_3 and d_4. So A_1, B_2 are tilted lines and B_1 and A_2 are contained in Euclidean circles; see Figure 3.

Denote by $2\mu=x_{d_2}-x_{d_1}$, $2\omega=x_{d_3}-x_{d_2}$, and $0<\theta<\frac{\pi}{2}$ the angle such that $2H=\sin\theta$. This domain is not defined for all values of μ,ω,θ. We have to suppose that $B_1\cap B_2=\varnothing$.

Claim 8.1. *With the notation above, for*

$$2H<\sqrt{\frac{\omega}{\omega+\mu}},$$

the domain $\mathcal{D}$ is well defined.

Proof. Since B_2 is a tilted line making angle θ with vertical, we can write

$$B_2(y) = (x_{d_3} - y\tan(\theta),\, y).$$

The curve B_1 satisfies $(x-(x_{d_1}+\mu))^2 + (y-\mu\tan\theta)^2 = \left(\frac{\mu}{\cos\theta}\right)^2$ for $y>0$. Since $x_{d_3} = x_{d_1} + 2\mu + 2\omega$, we have $B_1 \cap B_2 \neq \varnothing$ if

$$y > 0 \quad \text{and} \quad (\mu + 2\omega - y\tan\theta)^2 + (y-\mu\tan\theta)^2 = \left(\frac{\mu}{\cos\theta}\right)^2.$$

Then $B_1 \cap B_2 = \varnothing$ if

$$2H = \sin(\theta) < \sqrt{\frac{\omega}{\omega+\mu}}\,. \qquad \square$$

We will assume that the domain $\mathcal{D}$ is well defined. We will show that the conditions of Theorem 3.1 are true for some choice of the horocycles at the vertices of $\mathcal{D}$, provided that $2H < \sqrt{2}/2$.

Suppose that B_1 and A_2 are contained in Euclidean circles centered at $(x_{d_1}+\mu,\, h)$ and $(x_{d_2}+\omega,\, R_A)$, respectively, where $R_A = \omega/\cos\theta$, $R_B = \mu/\cos\theta$ are the Euclidean radii of these circles and $l = \omega\tan\theta$, $h = \mu\tan\theta$; see Figure 3.

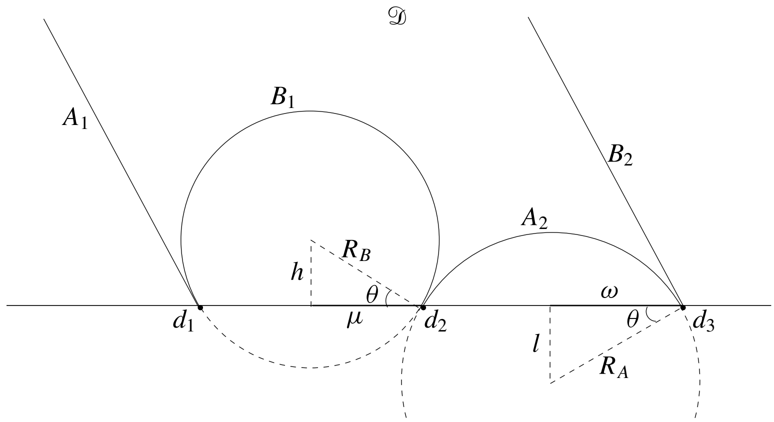


Figure 3. The domain $\mathcal{D}$.

On each vertex d_i we put horocycles $\mathcal{H}_i$, $\mathcal{H}_i \cap \mathcal{H}_j = \varnothing$, $i \neq j$. Since this domain does not have inscribed polygons we will verify only condition (3) of Theorem 3.1. When $\mu = \omega$ and $2H < \sqrt{2}/2$ we have, for this choice of horocycles, that $\alpha(\partial\mathcal{D}) = \beta(\partial\mathcal{D})$, so condition (2) of Theorem 3.1 can't occur. The next proposition shows that there is a choice of ω such that this condition is satisfied for $2H < \sqrt{2}/2$.

Proposition 8.1. *With the notation above, given $\mu \geq 3$ and $2H < \sqrt{2}/2$, there is $\omega_0 \geq \mu$ such that the condition $\alpha(\partial\mathcal{D}) - \beta(\partial\mathcal{D}) = 2H\mathcal{A}(\mathcal{D})$ is satisfied.*

Proof. First, we calculate the area $\mathcal{A}(\mathcal{D})$. Since the arc B_1 satisfies the equation $(x-(x_{d_1}+\mu))^2+(y-h)^2=R_B^2$ and the arc A_2 satisfies the equation $(x-(x_{d_2}+\omega))^2+(y+l)^2=R_A^2$, we have

$$\mathcal{A}(\mathcal{D})=\lim_{a\to 0}\frac{2(\mu+\omega)}{a}-2\lim_{a\to 0^+}\int_a^{R_B+h}\int_{x_{d_1}+\mu}^{\sqrt{R_B^2-(y-h)^2}+x_{d_1}+\mu}\frac{dxdy}{y^2}$$
$$-2\lim_{a\to 0^+}\int_a^{R_A-l}\int_{x_{d_2}+\omega}^{\sqrt{R_A^2-(y+l)^2}+x_{d_2}+\omega}\frac{dxdy}{y^2},$$

where the first term is the area between the arcs A_1, B_2 and straight line segment joining d_1, d_2, d_3.

Then

$$\mathcal{A}(\mathcal{D})=2\pi+2\tan\theta\ln\frac{2\omega^2(R_A-l)}{R_A^2-lR_A}+2\tan\theta\ln\frac{R_B^2+hR_B}{2\mu^2(R_B+h)}=2\Big(\pi+\ln\frac{\omega}{\mu}\Big).$$

Now, we are interested in the difference $\alpha(\partial\mathcal{D})-\beta(\partial\mathcal{D})$. We can suppose the horocycles $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$ are the same, that is, they differ by a horizontal translation. With this choice of the horocycles, we have $\alpha(\partial\mathcal{D})-\beta(\partial\mathcal{D})=|A_2|-|B_1|$, where $|A_2|$ and $|B_1|$ are the lengths of the compact arcs of A_2, B_1, respectively, which are outside of the horodisks bounded by $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$. Moreover, we will suppose that $\omega\geq\mu$ and that $\mathcal{H}_i\cap\Upsilon_i=(x_{d_i},\mu/2)$, where Υ_i is the vertical geodesic through x_{d_i}. It is possible to show that the intersection of B_1 and $\mathcal{H}_1$ occurs at $(x_{d_1},0)$ and at

$$(x_0,y_0)=\Big(-\sqrt{R_B^2-(y_0-h)^2}+x_{d_1}+\mu,\ \frac{8\mu^3}{17\mu^2+16h^2-8h\mu}\Big),$$

where B_1 and $\mathcal{H}_1$ satisfy the equations $(x-(x_{d_1}+\mu))^2+(y-h)^2=R_B^2$ and $(x-x_{d_1})^2+(y-\mu/4)^2=\mu^2/16$ respectively.

Similarly, the intersection of A_2 and $\mathcal{H}_2$ occurs at $(x_{d_2},0)$ and at

$$(x_1,y_1)=\Big(-\sqrt{R_A^2-(y_1+l)^2}+x_{d_2}+\omega,\ \frac{8\omega^2\mu}{16\omega^2+\mu^2+16l^2+8\mu l}\Big),\tag{7}$$

where A_2 and $\mathcal{H}_2$ satisfy the equations $(x-(x_{d_2}+\omega))^2+(y+l)^2=R_A^2$ and $(x-x_{d_2})^2+(y-\mu/4)^2=\mu^2/16$, respectively.

Then, the length of B_1 with respect to the horocycles $\mathcal{H}_1, \mathcal{H}_2$ is

$$|B_1|=2\int_{y_0}^{R_B+h}\frac{R_B}{y\sqrt{R_B^2-(y-h)^2}}\,dy$$
$$=\frac{2}{\cos\theta}\Big(-\ln R_B-\ln y_0+\ln\big(\mu\sqrt{R_B^2-(y_0-h)^2}+\mu^2+hy_0\big)\Big).$$

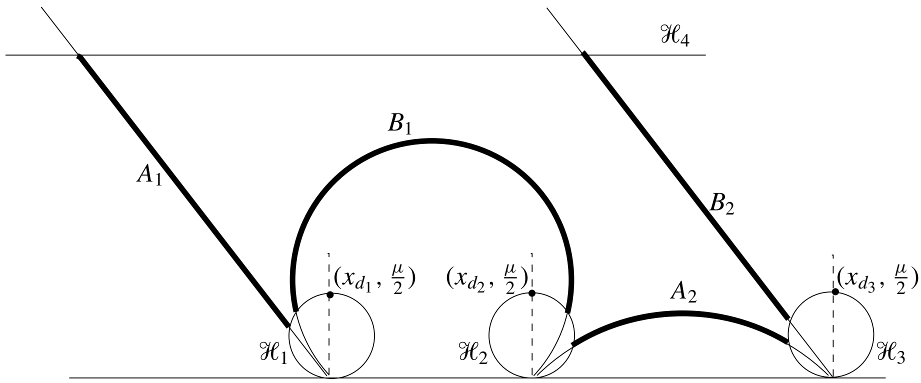


Figure 4. The domain $\mathcal{D}$ with the horocycles.

Analogously, the length of A_2 with respect to the horocycles $\mathcal{H}_2$, $\mathcal{H}_3$ is

$$\begin{aligned}|A_2| &= 2\int_{y_1}^{R_A-l} \frac{R_A}{y\sqrt{R_A^2-(y+l)^2}}\,dy \\ &= \frac{2}{\cos\theta}\Big(-\ln R_A - \ln y_1 + \ln\big(\omega\sqrt{R_A^2-(y_1+l)^2}+\omega^2-ly_1\big)\Big).\end{aligned}$$

So $\alpha(\partial\mathcal{D})-\beta(\partial\mathcal{D})-2H\mathcal{A}(\mathcal{D})$ only depends on μ and ω, because θ also depends on μ or ω. Thus consider, for each $\mu\in\mathbb{R}$, $\mu\geq 3$ fixed, the function

$$F(\omega)=\alpha(\partial\mathcal{D})-\beta(\partial\mathcal{D})-2H\mathcal{A}(\mathcal{D}).$$

We will show that at any moment this function is zero. We know for $\mu=\omega$ that $F(\omega)=-2H\mathcal{A}(\mathcal{D})<0$; thus we must show that for ω large enough, $F(\omega)>0$, so there exists a ω_0 such that $F(\omega_0)=0$ for each $\mu\geq 3$ fixed. We have

$$\begin{aligned}F(\omega) &= \frac{2}{\cos\theta}\Big(-\ln R_A - \ln y_1 + \ln\big(\omega\sqrt{R_A^2-(y_1+l)^2}+\omega^2-ly_1\big). \\ &\qquad + \ln R_B + \ln y_0 - \ln\big(\mu\sqrt{R_B^2-(y_0-h)^2}+\mu^2+hy_0\big)\Big) \\ &\qquad - 4H\Big(\pi+\ln\frac{\omega}{\mu}\Big) \\ &= \frac{2}{\cos\theta}\left(\ln\Big(\frac{1}{R_Ay_1}\big(\omega\sqrt{R_A^2-(y_1+l)^2}+\omega^2-ly_1\big)\Big)\right. \\ &\qquad \left. + \ln\frac{R_By_0}{\mu\sqrt{R_B^2-(y_0-h)^2}+\mu^2+hy_0}\right) \\ &\qquad - 4H\pi - 2\sin\theta\ln\frac{\omega}{\mu}.\end{aligned}$$

The second logarithmic term in the big parentheses is constant, because we are supposing μ fixed. As for the remaining terms, we substitute the value of y_1 from (7) and find that the difference

$$\frac{2}{\cos\theta}\ln\Big(\frac{1}{R_A y_1}\big(\omega\sqrt{R_A^2-(y_1+l)^2}+\omega^2-ly_1\big)\Big)-2\sin\theta\ln\frac{\omega}{\mu}$$

is strictly positive and increasing, so the function F is increasing and unbounded. Thus there is a ω_0 such that $F(\omega_0)=0$. □

References

[Collin and Rosenberg 2010] P. Collin and H. Rosenberg, "Construction of harmonic diffeomorphisms and minimal graphs", *Ann. of Math.* (2) **172** (2010), 1879–1906. MR 2726102 Zbl 05850189

[Finn 1963] R. Finn, "New estimates for equations of minimal surface type", *Arch. Rational Mech. Anal.* **14** (1963), 337–375. MR 28 #336 Zbl 0133.04601

[Finn 1965] R. Finn, "Remarks relevant to minimal surfaces, and to surfaces of prescribed mean curvature", *J. Analyse Math.* **14** (1965), 139–160. MR 32 #6337 Zbl 0163.34604

[Gálvez and Rosenberg 2010] J. A. Gálvez and H. Rosenberg, "Minimal surfaces and harmonic diffeomorphisms from the complex plane onto certain Hadamard surfaces", *Amer. J. Math.* **132**:5 (2010), 1249–1273. MR 2732346 Zbl 05816759

[Hauswirth et al. 2008] L. Hauswirth, H. Rosenberg, and J. Spruck, "On complete mean curvature $\frac{1}{2}$ surfaces in $\mathbb{H}^2\times\mathbb{R}$", *Comm. Anal. Geom.* **16**:5 (2008), 989–1005. MR 2010d:53009 Zbl 1166.53041

[Hauswirth et al. 2009] L. Hauswirth, H. Rosenberg, and J. Spruck, "Infinite boundary value problems for constant mean curvature graphs in $\mathbb{H}^2\times\mathbb{R}$ and $\mathbb{S}^2\times\mathbb{R}$", *Amer. J. Math.* **131**:1 (2009), 195–226. MR 2010e:53008 Zbl 1178.53062

[Jenkins and Serrin 1966] H. Jenkins and J. Serrin, "Variational problems of minimal surface type, II: Boundary value problems for the minimal surface equation", *Arch. Rational Mech. Anal.* **21** (1966), 321–342. MR 32 #8221 Zbl 0171.08301

[Mazet et al. 2008] L. Mazet, M. Rodríguez, and H. Rosenberg, "The Dirichlet problem for the minimal surface equation — with possible infinite boundary data — over domains in a Riemannian surface", preprint, 2008. to appear in *J. London Math. Soc.* arXiv 0806.0498

[Nelli and Rosenberg 2002] B. Nelli and H. Rosenberg, "Minimal surfaces in $\mathbb{H}^2\times\mathbb{R}$", *Bull. Braz. Math. Soc.* (*N.S.*) **33**:2 (2002), 263–292. MR 2004d:53014 Zbl 1038.53011

[Pinheiro 2009] A. L. Pinheiro, "A Jenkins–Serrin theorem in $M^2\times\mathbb{R}$", *Bull. Braz. Math. Soc.* (*N.S.*) **40**:1 (2009), 117–148. MR 2010a:53017 Zbl 1176.53019

[Rosenberg 2002] H. Rosenberg, "Minimal surfaces in $\mathbb{M}^2\times\mathbb{R}$", *Illinois J. Math.* **46**:4 (2002), 1177–1195. MR 2004d:53015 Zbl 1036.53008

[Spruck 1972] J. Spruck, "Infinite boundary value problems for surfaces of constant mean curvature", *Arch. Rational Mech. Anal.* **49** (1972), 1–31. MR 48 #12329 Zbl 0263.53008

[Zhang 2005] S. Zhang, "Curvature estimates for CMC surfaces in three dimensional manifolds", *Math. Z.* **249**:3 (2005), 613–624. MR 2005i:53011 Zbl 1132.53303

Received March 16, 2010.

ABIGAIL FOLHA
INSTITUTO DE MATEMÁTICA - DEPARTAMENTO DE GEOMETRIA
UNIVERSIDADE FEDERAL FLUMINENSE
R. MÁRIO SANTOS BRAGA, S/N
CAMPUS DO VALONGUINHO
24020-140 NITERÓI RJ
BRAZIL
abigailfolha@vm.uff.br

SOFIA MELO
DEPARTAMENTO DE MATEMÁTICA – ICE
UNIVERSIDADE FEDERAL DE JUIZ DE FORA
CIDADE UNIVERSITÁRIA
36036-330 JUIZ DE FORA, MG
BRAZIL
sofia.melo@ufjf.edu.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 1, 2011

OSGOOD–HARTOGS-TYPE PROPERTIES OF POWER SERIES AND SMOOTH FUNCTIONS

BUMA L. FRIDMAN AND DAOWEI MA

We study the convergence of a formal power series of two variables if its restrictions on curves belonging to a certain family are convergent. Also analyticity of a given C^∞ function f is proved when the restriction of f on analytic curves belonging to some family is analytic. Our results generalize two known statements: a theorem of P. Lelong and the Bochnak–Siciak theorem. The questions we study can be regarded as problems of Osgood–Hartogs type.

Introduction

Hartogs' theorem is a fundamental result in complex analysis: *A function f in $\mathbb{C}^n$, where $n > 1$, is holomorphic if it is holomorphic in each variable separately.* That is, f is holomorphic in $\mathbb{C}^n$ if for each axis it is holomorphic on every complex line parallel to this axis. In the last interpretation this statement leads to a number of questions described in an article by K. Spallek, P. Tworzewski, T. Winiarski [Spallek et al. 1990] in the following way: "Osgood–Hartogs-type problems ask for properties of 'objects' whose restrictions to certain 'test-sets' are well known". The article has a number of examples of such problems. Here are two classical examples: a theorem of P. Lelong and one proved independently by J. Bochnak and J. Siciak.

Theorem [Lelong 1951]. *A formal power series $g(x, y)$ converges in some neighborhood of the origin if there exists a set $E \subset \mathbb{C}$ of positive capacity such that, for each $s \in E$, the formal power series $g(x, sx)$ converges in some neighborhood of the origin* (*of a size possibly depending on s*).

Theorem [Bochnak 1970; Siciak 1970]. *Let $f \in C^\infty(D)$, where D is a domain in $\mathbb{R}^n$ containing* 0. *Suppose f is analytic on every line segment through* 0. *Then f is analytic in a neighborhood of* 0 (*as a function of n variables*).

In many articles the same two "objects" are usually considered: power series and functions of several variables. The test sets in many cases form a family of linear

MSC2000: 26E05, 30C85, 40A05.
Keywords: formal power series, analytic functions, capacity.

subspaces of lower dimension. For example, articles by S. S. Abhyankar, T. T. Moh [1970], N. Levenberg and R. E. Molzon, [1988], R. Ree [1949], A. Sathaye [1976], M. A. Zorn [1947] and others consider the convergence of formal power series of several variables provided the restriction of such a series on each element of a sufficiently large family of linear subspaces is convergent. T. S. Neelon [2009; 2006] proved that a formal power series is convergent if its restrictions to certain families of curves or surfaces parametrized by polynomial maps are convergent. The articles [Bochnak 1970; Neelon 2004; 2009; Siciak 1970], among others, prove that a function of several variables is highly smooth (or even analytic) if it is smooth enough on each of a sufficiently large set of linear or algebraic curves (or surfaces of lower dimension). The publication by E. Bierstone, P. D. Milman, A. Parusiński [Bierstone et al. 1991] provides an interesting example of a noncontinuous function in $\mathbb{R}^2$ that is analytic on every analytic curve.

In this article we also consider both: power series with complex coefficients and functions in a neighborhood of the origin in $\mathbb{R}^2$. As test sets we consider separately two families. They are derived the following way. First consider a nonlinear analytic curve $\Gamma = \{x, \gamma(x)\}$, with $\gamma(0) = 0$. One family, $\Im_1$, is a set of dilations of Γ: $\Im_1 = \{sx, s\gamma(x)\}, s \in \Lambda_1\}$, where $\Lambda_1 \subset \mathbb{R}$ is a closed subset of $\mathbb{C}$ of positive capacity. The other family, $\Im_2$, consists of curves Γ_θ, each of which is a rotation of Γ about the origin by an angle $\theta \in \Lambda_2$, where Λ_2 is a subset of $[0, 2\pi]$ of positive capacity. If f is C^∞ and its restriction on every curve of $\Im_1$ can be extended as an analytic function in a neighborhood of that curve, then f is real analytic in a neighborhood of the origin in the region covered by the curves of $\Im_1$. The same is true regarding $\Im_2$. (For precise statements see Theorems 2.1 and 2.2).

We start however with two results related to power series. First we prove a generalization of P. Lelong's theorem. Namely, if $g(x, y)$ is a formal power series and $h(x)$, $h(0) = 0$, is a convergent power series such that the inhomogeneous dilations $g(s^\sigma x, s^\tau h(x))$ are convergent for sufficiently many s (σ, τ are fixed), then $g(x, y)$ is convergent (for the precise statement see Theorem 1.1). Theorem 1.2 is devoted to a reverse claim: if $h(x)$ is a formal power series and $g(s^\sigma x, s^\tau h(x))$ converges for sufficiently many s, then $h(x)$ is convergent.

The results in this paper do not carry over in a routine way to dimensions greater than two. We intend to study corresponding problems for higher dimensions in future work.

1. On the convergence of a power series in two variables

Let $\mathbb{C}[[x_1, x_2, \dots, x_n]]$ denote the set of (formal) power series

$$g(x_1, \dots, x_n) = \sum_{k_1, \dots, k_n \geq 0} a_{k_1 \dots k_n} x_1^{k_1} \cdots x_n^{k_n}$$

in n variables with complex coefficients. Let $g(0) = g(0, \dots, 0)$ denote the coefficient $a_{0,\dots,0}$. A power series equals 0 if all of its coefficients $a_{k_1 \dots k_n}$ are equal to 0. A power series $g \in \mathbb{C}[[x_1, x_2, \dots, x_n]]$ is said to be convergent if there is a constant $C = C_g$ such that $|a_{k_1 \dots k_n}| \le C^{k_1 + \cdots + k_n}$ for all $(k_1, \dots, k_n) \ne (0, \dots, 0)$. If g is convergent, then it represents a holomorphic function in some neighborhood of 0 in $\mathbb{C}^n$. If $g \in \mathbb{C}[[x_1, x_2, \dots, x_n]]$ and $s \in \mathbb{C}^n$, then $g_s(t) := g(s_1 t, \dots, s_n t)$ is well defined and belongs to $\mathbb{C}[[t]]$. By [Zorn 1947], g is convergent if and only if $g_s(t)$ is convergent for each $s \in \mathbb{C}^n$. The partial derivatives of a power series are well defined even when it is divergent (not convergent). For example, if $g \in \mathbb{C}[[x, y]]$ and if $g = \sum a_{ij} x^i y^j$, then

$$g'_y = \frac{\partial g}{\partial y} = \sum j a_{ij} x^i y^{j-1}.$$

Thus $g'_y \ne 0$ simply means that $g \notin \mathbb{C}[[x]]$. If $g \in \mathbb{C}[[x, y]]$, and if $h \in \mathbb{C}[[x]]$ with $h(0) = 0$, then $g(x, h(x))$ is a well-defined element of $\mathbb{C}[[x]]$.

As mentioned above, a lot of work has been done on the convergence of a power series with the assumption that the series is convergent after restriction to sufficiently many subspaces; see [Abhyankar and Moh 1970; Levenberg and Molzon 1988; Lelong 1951; Siciak 1970; 1982].

We consider substitution of a power series $y = h(x)$ into an inhomogeneous dilation $g(s^\sigma x, s^\tau y)$ of a series $g(x, y)$, where σ, τ are integers.

Let

$$Q := \{(\sigma, \tau) : \sigma, \tau \in \mathbb{Z}, (\sigma, \tau) \ne (0, 0)\}.$$

Let $\operatorname{cap} E$ denote the (logarithmic) capacity of a closed set E in the complex plane.

We now present our two main theorems.

Theorem 1.1. *Let $g \in \mathbb{C}[[x, y]]$ be a power series of two variables x, y, let $h \in \mathbb{C}[[x]]$ be a nonzero convergent power series with $h(0) = 0$, let E be a closed subset of $\mathbb{C} \setminus \{0\}$ with* $\operatorname{cap} E > 0$, *and let (σ, τ) be a pair in the set Q. Assume, in case $\sigma\tau > 0$, that $h(x)$ is not a monomial of the form $b_k x^k$ with $\sigma k - \tau = 0$. Suppose that $g(s^\sigma x, s^\tau h(x))$ is convergent for each $s \in E$. Then g is convergent.*

Theorem 1.2. *Let $g \in \mathbb{C}[[x, y]]$ be a power series with $g'_y \ne 0$, let $h \in \mathbb{C}[[x]]$ be a nonzero power series with $h(0) = 0$, let E be a closed subset of $\mathbb{C} \setminus \{0\}$ with* $\operatorname{cap} E > 0$, *and let (σ, τ) be a pair in the set Q with $\sigma\tau > 0$. Suppose that $g(s^\sigma x, s^\tau h(x))$ is convergent for each $s \in E$. Then h is convergent.*

The examples in Section 3 show that if any condition in these two theorems is dispensed with, the resulting statement is false. We now prove some auxiliary results.

The following theorem is a consequence of a result by B. Malgrange [1966]. We present an independent short proof.

Theorem 1.3. *Let $g \in \mathbb{C}[[x_1, \dots, x_n, y]]$ with $g'_y \neq 0$, and let $h \in \mathbb{C}[[x_1, \dots, x_n]]$ with $h(0) = 0$. Suppose that g and $g(x_1, \dots, x_n, h(x_1, \dots, x_n))$ are convergent. Then h must be convergent.*

Proof. Let $f \in \mathbb{C}[[x_1, \dots, x_n, y]]$ be defined by

$$f(x_1, \dots, x_n, y) = g(x_1, \dots, x_n, y) - g(x_1, \dots, x_n, h(x_1, \dots, x_n)).$$

Then f is convergent and $f(x_1, \dots, x_n, h(x_1, \dots, x_n)) = 0$. Fix $s = (s_1, \dots, s_n) \in \mathbb{C}^n$. Let $f_s(t, y) \in \mathbb{C}[[t, y]]$ be defined by $f_s(t, y) = f(s_1 t, \dots, s_n t, y)$. Then $f_s(t, y)$ is convergent and $f_s(t, h_s(t)) = 0$. By the Weierstrass preparation theorem (see [Griffiths and Harris 1978, p. 8], for example), there is a nonnegative integer k such that $f_s(t, y) = t^k P(t, y) Q(t, y)$, where $P(t, y) = y^m + a_1(t) y^{m-1} + \cdots + a_m(t)$ is a polynomial in y with coefficients being convergent power series in t, and $Q(t, y)$ is a convergent power series with $Q(0, 0) \neq 0$. Hence $P(t, h_s(t)) = 0$. By [Fuks 1963, Theorem 4.12, p. 73] there is a positive integer r such that $P(t^r, y)$ splits into linear factors in y:

$$P(t^r, y) = (y - u_1(t)) \cdots (y - u_m(t)),$$

where the $u_j(t)$ are convergent power series. Thus

$$0 = P(t^r, h_s(t^r)) = (h_s(t^r) - u_1(t)) \cdots (h_s(t^r) - u_m(t)).$$

It follows that $h_s(t^r) = u_j(t)$ for some j. Therefore $h_s(t)$ is convergent. Since $h_s(t)$ is convergent for each $s \in \mathbb{C}^n$, the series $h(x_1, \dots, x_n)$ must be convergent. □

Let E be a closed bounded set in the complex plane. The *transfinite diameter* of E is defined as

$$d_\infty(E) = \lim_n \Big(\max \Big\{ \prod_{i<j} |z_i - z_j|^{2/n(n-1)} : z_1, \dots, z_n \in E \Big\} \Big).$$

For a probability measure μ on the compact set E, the *logarithmic potential* of μ is

$$p_\mu(z) = \lim_{N \to \infty} \int \min\Big(N, \log \frac{1}{|z - \zeta|}\Big)\, d\mu(\zeta),$$

and the *capacity* of E is defined by

$$\operatorname{cap} E = \exp(- \min_{\mu(E)=1} \sup_{z \in \mathbb{C}} p_\mu(z)).$$

It turns out that $d_\infty(E) = \operatorname{cap} E$ [Ahlfors 1973, pp. 23–28]. It follows from the definition of the transfinite diameter that, for $E_1 \supset E_2 \supset \cdots$,

$$E = \bigcap E_n \implies \operatorname{cap} E = \lim(\operatorname{cap} E_n),$$

and from the definition of the capacity that, if $E_1 \subset E_2 \subset \cdots$,

$$E = \bigcup E_n \implies \operatorname{cap} E = \lim(\operatorname{cap} E_n). \tag{1}$$

If E is a closed set, its capacity can be defined by

$$\operatorname{cap} E = \lim_n \operatorname{cap}(E \cap \{|x| \le n\}).$$

Lemma 1.4 (Bernstein inequality). *Let E be a compact set in the complex plane with* $\operatorname{cap} E > 0$. *Then there exists a positive constant $C = C_E$, depending only on E, such that for each positive integer n and each polynomial $P(z) = \sum a_k z^k \in \mathbb{C}[z]$ of degree n, each coefficient a_k, $0 \le k \le n$, of $P(z)$ satisfies*

$$|a_k| \le C^n \max_{z \in E} |P(z)|.$$

Proposition 4.6 in [Neelon 2009] can be used to prove this statement. Also (we thank Nessim Sibony for pointing this out to us) this lemma follows from considerations in [Sibony 1985]. We present here an independent short proof.

Proof. Without loss of generality we assume that $\max_{z \in E} |P(z)| = 1$. Let Ω be the unbounded component of the complement of E in $\mathbb{C}$. It is known that Ω has a Green's function with a pole at ∞ [Ahlfors 1966; 1973, pp. 25–27]. The Green's function is harmonic in Ω, 0 on $\partial\Omega$, and its asymptotic behavior at ∞ is

$$u(z) = \log|z| - \log\alpha + o(1),$$

where $\alpha := \operatorname{cap} E$. On applying the maximum principle to the subharmonic function $\log|P(z)| - (n+\epsilon)u(z)$, we obtain $|P(z)| \le e^{nu(z)}$ for $z \in \Omega$. Choose an $R > 1$ so that $E \subset \{z : |z| < R\}$. Set $C = \max_{|z|=R} e^{u(z)}$. Then $|P(z)| \le C^n$ if $|z| = R$, and

$$|a_k| = \left| \frac{1}{2\pi i} \int_{|z|=R} \frac{P(z)}{z^{k+1}} \, dz \right| \le R^{-k} \max_{|z|=R} |P(z)| \le C^n.$$

This proves the lemma. □

Proof of Theorem 1.1. We assume that $a_{00} = g(0,0) = 0$, that E is bounded, that $\gcd(\sigma, \tau) = 1$, that $\sigma \ge 0$, and, in case $\sigma = 0$, that $\tau = -1$. This causes no loss of generality. Indeed, if E is unbounded, set $E_n = \{s \in E : n \ge |s| \ge 1/n\}$. Since $\lim \operatorname{cap} E_n = \operatorname{cap} E > 0$, the set E_n has positive capacity when n is sufficiently large. On replacing E by E_n, we obtain that $0 \notin E$ and E is bounded. If $d := \gcd(\sigma, \tau) > 1$, we can replace (σ, τ) by $(\sigma/d, \tau/d)$, and E by the set $\{s \in \mathbb{C} : s^d \in E\}$. Finally, if $\sigma < 0$, or if $(\sigma, \tau) = (0, 1)$, we can replace (σ, τ) by $(-\sigma, -\tau)$, and E by $\{s \in \mathbb{C} : s^{-1} \in E\}$.

Let

$$h(x) = \sum_{i=1}^{\infty} b_i x^i.$$

Then

$$h(x)^j = \sum_{k=j}^{\infty} c_{jk} x^k,$$

where

$$c_{jk} = \sum_{l_1+\cdots+l_j=k} b_{l_1} \cdots b_{l_j}.$$

Note that $c_{jk} = 0$ for $k < j$. Hence

$$g(s^\sigma x, s^\tau h(x)) = \sum_{i,j,k} a_{ij} c_{jk} s^{\sigma i+\tau j} x^{i+k} = \sum_{p=1}^{\infty} \Bigg(\sum_{q=-\tau^- p}^{(\sigma+\tau^+)p} d_{pq} s^q \Bigg) x^p,$$

where $\tau^+ = \max(0, \tau)$, $\tau^- = -\min(0, \tau)$, and

$$d_{pq} = \sum_{\sigma i+\tau j=q} a_{ij} c_{j,p-i}. \tag{2}$$

For each $p \geq 1$ and each $q \in \mathbb{Z}$, the sum (2) contains only a finite number of nonzero terms. Let $u_p(s) = \sum_q d_{pq} s^q$. Then $s^{\tau^- p} u_p(s)$ is a polynomial in s of degree at most $(\sigma + |\tau|)p$, and $g(s^\sigma x, s^\tau h(x)) = \sum u_p(s) x^p$. For $s \in E$, since $g(s^\sigma x, s^\tau h(x))$ is convergent, its coefficients $u_p(s)$ satisfy $|u_p(s)| \leq C_s^p$ for some positive constant C_s, possibly depending on s, and $p = 1, 2, \ldots$. Set, for $n = 1, 2, \ldots,$

$$E_n = \{s \in E : |u_p(s)| \leq n^p \text{ for all } p > 0\}.$$

The sequence (E_n) is an increasing sequence of closed sets. Since $\lim \operatorname{cap} E_n = \operatorname{cap} E > 0$, the set E_n has positive capacity for some n. On replacing E by E_n, we obtain $|u_p(s)| \leq n^p$ for $s \in E$ and $p = 1, 2, \ldots$. The polynomial $s^{\tau^- p} u_p(s)$ is of degree at most $(\sigma + |\tau|)p$, and satisfies

$$|s^{\tau^- p} u_p(s)| \leq M^{\tau^- p} n^p, \quad s \in E,$$

where $M = \max_E |s|$. By Lemma 1.4, the coefficients of the above mentioned polynomial satisfy $|d_{pq}| \leq C_E^{(\sigma+|\tau|)p} M^{\tau^- p} n^p$, where C_E is the constant in Lemma 1.4, depending only on E. Set $C = C_E^{\sigma+|\tau|} M^{\tau^-} n$. Then

$$|d_{pq}| \leq C^p. \tag{3}$$

Let

$$g_q(x, y) = \sum_{\sigma i+\tau j=q} a_{ij} x^i y^j, \tag{4}$$

and let $\phi_q(x) = g_q(x, h(x))$, for $q \in \mathbb{Z}$. Then $g_q \in \mathbb{C}[[x, y]]$ in general, and it is a polynomial when $\sigma, \tau > 0$. It is straightforward to verify that

$$\phi_q(x) = g_q(x, h(x)) = \sum_{p=1}^{\infty} d_{pq} x^p. \tag{5}$$

The series $\phi_q(x)$ is convergent because of (3). Choose a positive number $r < 1/C$, where C is the constant in (3), so that $h(x)$ converges in a neighborhood of the closed ball $\{x \in \mathbb{C} : |x| \le r\}$ and $h(x) \neq 0$ when $0 < |x| \le r$. Let $m = \min_{|x|=r} |h(x)|$. Then $m > 0$. For $x \in \mathbb{C}$, $|x| \le r$,

$$|\phi_q(x)| \le \sum |d_{pq}||x|^p \le \sum (Cr)^p = \frac{1}{1 - Cr}.$$

We now consider two cases, depending on whether $\sigma\tau$ is positive.

Case (i): $\sigma > 0$, $\tau > 0$. Let

$$\Omega_q = \{(i, j) : i, j \in \mathbb{Z},\ i, j \ge 0,\ \sigma i + \tau j = q\}. \tag{6}$$

Let ω_q be the cardinality of Ω_q. It is clear that $\omega_q \le q + 1$. Fix a $q \ge 1$ so that $\omega_q > 0$. Let (λ, μ) be the element of Ω_q so that μ is the minimum. Then

$$\Omega_q = \{(\lambda - k\tau,\ \mu + k\sigma) : k = 0, 1, \dots, \omega_q - 1\},$$

and

$$g_q(x, y) = x^\lambda y^\mu \sum_{k=0}^{\omega_q - 1} a_{\lambda - k\tau,\, \mu + k\sigma} (x^{-\tau} y^\sigma)^k.$$

Let

$$\psi_q(t) = \sum_{k=0}^{\omega_q - 1} a_{\lambda - k\tau,\, \mu + k\sigma} t^k,$$

so that $g_q(x, y) = x^\lambda y^\mu \psi_q(x^{-\tau} y^\sigma)$, and

$$\psi_q(x^{-\tau} h(x)^\sigma) = x^{-\lambda} h(x)^{-\mu} \phi_q(x). \tag{7}$$

Let $u(x) = x^{-\tau} h(x)^\sigma$, $S = \{x \in \mathbb{C} : |x| = r\}$, and $F = u(S)$. Since $h(x)$ is not a monomial of the form $b_k x^k$ with $\sigma k - \tau = 0$, the function $u(x)$ is a nonconstant meromorphic function, hence F has positive capacity. For $t = x^{-\tau} h(x)^\sigma \in F$, we obtain, by (7), that

$$|\psi_q(t)| \le \frac{r^{-\lambda} m^{-\mu}}{1 - Cr} \le \frac{(1 + r^{-1} + m^{-1})^{\lambda + \mu}}{1 - Cr}. \tag{8}$$

The summand 1 in the right-hand side of the above inequality is included to ensure that the numerator is greater than 1 as needed later. Hence $|\psi_q(t)| \le L^q$ on F,

where

$$L = \frac{1+r^{-1}+m^{-1}}{1-Cr},$$

for $\lambda+\mu \le q$. By Lemma 1.4, the coefficients of ψ_q are bounded by $L^q C_F^{\omega_q-1}$. Thus for $(i, j) \in \Omega_q$,

$$|a_{ij}| \le L^q C_F^{\omega_q-1} \le (L+C_F)^{2q} \le (L+C_F)^{2(\sigma+\tau)(i+j)},$$

or $|a_{ij}| \le K^{i+j}$, where $K = (L+C_F)^{2(\sigma+\tau)}$. The number K does not depend on q. It follows that

$$|a_{ij}| \le K^{i+j}, \text{ if } \sigma i + \tau j \ge 1.$$

This proves that g is convergent.

Case (ii): $\sigma \ge 0$, $\tau \le 0$. In this case the set Ω_q in (6) can be written as

$$\Omega_q = \{(\lambda - k\tau,\ \mu + k\sigma) : k = 0, 1, 2, \dots\},$$

where (λ, μ) is the element in Ω_q with least value of μ when $\sigma > 0$, and $(\lambda, \mu) = (0, -q)$ when $(\sigma, \tau) = (0, -1)$. Let

$$\psi_q(t) = \sum_{k=0}^{\infty} a_{\lambda+k|\tau|,\, \mu+k\sigma} t^k.$$

Then $g_q(x, y) = x^\lambda y^\mu \psi_q(x^{|\tau|} y^\sigma)$. The formal power series $\psi_q(t)$ satisfies $\phi_q(x) = x^\lambda h(x)^\mu \psi_q(x^{|\tau|} h(x)^\sigma)$. Since $x^\lambda h(x)^\mu$ and $\phi_q(x)$ are convergent, the series

$$\alpha(x) := \psi_q(x^{|\tau|} h(x)^\sigma)$$

has to be convergent. Write $x^{|\tau|} h(x)^\sigma = cx^\nu + \cdots$, $c \ne 0$. There is a power series $\beta(x)$, also convergent in a neighborhood of $\{|x| \le r\}$, such that $x^{|\tau|} h(x)^\sigma = \beta(x)^\nu$. Reducing r if necessary, we assume that $\beta(x)$ is univalent in a neighborhood of $\{|x| \le r\}$. Note that the reduction in the value of r is independent of q. The set $\{\beta(x) : |x| < r\}$ contains an open disc $\{z \in \mathbb{C} : |z| < \delta\}$. The series $\beta(x)$ has an inverse $\gamma(z)$, convergent in $\{z \in \mathbb{C} : |z| < \delta\}$, such that $\gamma(\beta(x)) = x$ and $\beta(\gamma(z)) = z$. Now $\psi_q(z^\nu)$ is converge nt in $\{|z| < \delta\}$, so $\psi_q(t)$ is convergent in $\{|t| < \delta^\nu\}$. Let $t \in \mathbb{C}$ with $|t| < \delta^\nu$. Then $t = z^\nu$ for some z with $|z| < \delta$, and $z = \beta(x)$ for some x with $|x| < r$. Hence

$$|\psi_q(t)| = |\psi_q(\beta(x)^\nu)| = |\alpha(x)| \le \max_{|x|=r} |\alpha(x)|.$$

Thus

$$\sup_{|t|<\delta^\nu} |\psi_q(t)| \le \max_{|x|=r} \left| \frac{\phi_q(x)}{x^\lambda h(x)^\mu} \right| \le \frac{r^{-\lambda} m^{-\mu}}{1-Cr}.$$

By the Cauchy estimates, the coefficients of ψ_q satisfy

$$|a_{\lambda+k|\tau|,\,\mu+k\sigma}| \leq \frac{r^{-\lambda}m^{-\mu}}{1-Cr}\delta^{-k\nu} \leq \frac{(1+r^{-1}+m^{-1}+\delta^{-\nu})^{\lambda+\mu+k}}{1-Cr}.$$

The summand 1 in the right-hand side of the above inequality is included to ensure that the numerator is greater than 1 as needed later. It follows that, for $(i, j) \in \Omega_q$,

$$|a_{ij}| \leq \left(\frac{1+r^{-1}+m^{-1}+\delta^{-\nu}}{1-Cr}\right)^{i+j}.$$

The number $K := (1+r^{-1}+m^{-1}+\delta^{-\nu})/(1-Cr)$ does not depend on q. Therefore, $|a_{ij}| \leq K^{i+j}$ for all (i, j). This proves that g is convergent. □

Proof of Theorem 1.2. This proof and the proof of Theorem 1.1 share the discussion through Equation (5). Note that the convergence of h has not been used in the derivation of (5). We define polynomials $g_q(x, y)$ by (4). Then $g_q(x, h(x))$ are convergent by (3) and (5). Since $g'_y(x, y) \neq 0$, $\partial g_q/\partial y \neq 0$ for some q. It follows from Theorem 1.3 that $h(x)$ is convergent. □

For $h \in \mathbb{C}[[x]]$ with $h(0) = 0$, let $h_s(x) = s^{-1}h(sx)$.

Corollary 1.5. *Let $g \in \mathbb{C}[[x, y]]$ be a power series, let $h \in \mathbb{C}[[x]]$ be a nonzero and nonlinear power series with $h(0) = 0$, and let E be a closed subset of $\mathbb{R} \setminus \{0\}$ with* cap $E > 0$. *Suppose that $g(x, h_s(x))$ is convergent for each $s \in E$. Then g is convergent.*

Proof. If $g'_y = 0$ then the statement holds. Suppose $g'_y \neq 0$. For $s \neq 0$, $g(x, h_s(x))$ is convergent if and only if $g(s^{-1}x, h_s(s^{-1}x)) = g(s^{-1}x, s^{-1}h(x))$ is convergent. By Theorem 1.2, h is convergent. Then g is convergent by Theorem 1.1. □

For $f \in \mathbb{C}[[x, y]]$ and $\theta \in [0, 2\pi]$, write

$$f_\theta(x, y) = f(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta).$$

Theorem 1.6. *Let $f \in \mathbb{C}[[x, y]]$ be a power series, let $h \in \mathbb{C}[[x]]$ be a convergent power series with $h(0) = 0$, and let E be a closed subset of $[0, 2\pi]$ with* cap $E > 0$. *Suppose that $f_\theta(x, h(x))$ is convergent for each $\theta \in E$. Then f is convergent.*

Proof. Let $g(x, y) = f((x+y)/2, -i(x-y)/2)$. Then $f(x, y) = g(x+iy, x-iy)$ and $f_\theta(x, y) = g(e^{i\theta}(x+iy), e^{-i\theta}(x-iy))$. Let $\phi_\theta(x) = f_\theta(x, h(x)) = g(e^{i\theta}(x+ih(x)), e^{-i\theta}(x-ih(x)))$. Then $\phi_\theta(x)$ is convergent for $\theta \in E$. The x terms of the two series $x \pm ih(x)$ cannot both be zero. Say, the x term of $x+ih(x)$ is nonzero. So $x+ih(x)$ has an inverse $\psi(x)$ which is a convergent power series such that $\psi(x)+ih(\psi(x)) = x$. Set $\psi(x)-ih(\psi(x)) = \omega(x)$. Then $\phi_\theta(\psi(x)) = g(e^{i\theta}x, e^{-i\theta}\omega(x))$ is convergent for $\theta \in E$. It follows that $g(sx, s^{-1}\omega(x))$ is convergent for each s in the set $\{e^{i\theta} : \theta \in E\}$, which has positive capacity. By Theorem 1.1, g is convergent. Therefore f is convergent. □

2. Analytic functions in $\mathbb{R}^2$

Suppose that $f(x, y), \phi(x), q(x)$ are C^∞ functions defined near the origin with $\phi(0) = 0$. Let $\hat{f}, \hat{\phi}, \hat{q}$ denote the Taylor series at 0 of those functions. Then $\hat{f}$ lies in $\mathbb{C}[[x, y]]$ and $\hat{\phi}, \hat{q}$ lie in $\mathbb{C}[[x]]$. By the chain rule, $f(x, \phi(x)) = q(x)$ implies $\hat{f}(x, \hat{\phi}(x)) = \hat{q}(x)$. We consider here complex-valued analytic functions of real variables. If I is an interval and if $\Gamma = \{(t, \gamma(t)) : t \in I\}$ is a curve, the dilation by s of Γ is

$$\Gamma_s = \{(st, s\gamma(t))\} = \{(t, \gamma_{1/s}(t))\}, \quad \gamma_s(t) = s^{-1}\gamma(st).$$

Theorem 2.1. *Let f be a C^∞ function defined in an open set $\Omega \subset \mathbb{R}^2$ containing the origin, let $\Gamma = \{(t, \phi(t))\}$ be a nonlinear analytic curve in $\mathbb{R}^2$ passing through or ending at the origin, and let E be a closed subset of $\mathbb{R} \setminus \{0\}$ of positive capacity. Suppose that for each $s \in E$, there is a real analytic function F_s defined in a neighborhood Q_s of $\Gamma_s \cap \Omega$ in $\mathbb{R}^2$ such that f and F_s coincide on $\Gamma_s \cap \Omega$. Then there is a neighborhood U of the origin, and a real analytic function F defined on U that coincides with f on $U \cap \Lambda$, where $\Lambda := \bigcup_{s \in E} \Gamma_s$.*

Proof. Without loss of generality we assume that $\phi(0) = 0$. Since f and F_s coincide on Γ_s, we have

$$f(x, \phi_{1/s}(x)) = F_s(x, \phi_{1/s}(x)). \tag{9}$$

Let g, h denote the Taylor series of f, ϕ respectively. Then (9) implies

$$g(x, h_{1/s}(x)) = F_s(x, h_{1/s}(x)).$$

Hence $g(x, h_{1/s}(x))$ is convergent for $s \in E$. By Corollary 1.5, g is convergent. Thus g represents a real analytic function F in some neighborhood U of the origin that satisfies $F(x, h_{1/s}(x)) = F_s(x, h_{1/s}(x))$. It follows that the real analytic function F coincides with f on $U \cap \Lambda$. □

Note that f does not need to be analytic in a neighborhood of the origin.

If $\Gamma = \{(t, \phi(t) : t \in I\}$ is a curve, its rotation by θ is

$$\Gamma_\theta = \{(t\cos\theta + \phi(t)\sin\theta, -t\sin\theta + \phi(t)\cos\theta) : t \in I\}.$$

Theorem 2.2. *Let f be a C^∞ function defined in an open set $\Omega \subset \mathbb{R}^2$ containing the origin, let $\Gamma = \{(t, \phi(t))\}$ be an analytic curve in $\mathbb{R}^2$ passing through or ending at the origin, and let E be a closed subset of $[0, 2\pi]$ of positive capacity. Suppose that for each $\theta \in E$, there is a real analytic function F_θ defined in a neighborhood Q_θ of $\Gamma_\theta \cap \Omega$ in $\mathbb{R}^2$ such that f and F_θ coincide on $\Gamma_\theta \cap \Omega$. Then there is an analytic function F defined in some neighborhood U of the origin that coincides with f on $U \cap \Lambda$, where $\Lambda := \bigcup_{\theta \in E} \Gamma_\theta$.*

Proof. The proof is similar to that of Theorem 2.1. Let

$$g_\theta(x, y) := g(x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta).$$

Then $g_\theta(x, h(x))$ is convergent for each $\theta \in E$. By Theorem 1.6, g is convergent. □

Corollary 2.3. *Let f be a C^∞ function defined in a neighborhood of 0 in $\mathbb{R}^2$, and let $\Gamma = \{(t, \phi(t))\}$ be an analytic curve passing through or ending at the origin in $\mathbb{R}^2$. Suppose that for each $\theta \in [0, 2\pi]$, the restriction of f to Γ_θ extends to a real analytic function F_θ in a neighborhood Q_θ of the origin. Then f is analytic in a neighborhood of the origin.*

Remark 2.4. We can see from the proofs that in Theorem 2.1, Theorem 2.2, and Corollary 2.3 the hypothesis on f can be weakened to f having a Taylor series at the origin in the sense that there are numbers a_{ij} such that for each positive integer n,

$$f(x, y) - \sum_{i+j\le n} a_{ij}x^i y^j = o((x^2 + y^2)^{n/2}).$$

3. Examples

Here we show that the restrictions in our main theorems are necessary.

Example 3.1. P. Lelong [1951] proved that if E is a set with $\operatorname{cap} E = 0$ then one can find a divergent power series $g(x, y)$ such that for all $s \in E$, $g(x, sx)$ is convergent. For completeness we present here a construction of such an example. Since $\operatorname{cap} E = 0$, there is a sequence of positive numbers (δ_n) with $\lim \delta_n = 0$, and a sequence of polynomials $(P_n(x))$ with $\max_{x\in E} |P(x)| \le \delta_n^n$, where

$$P_n(x) = \sum_{j=0}^{n} b_{nj}x^{n-j},$$

with $b_{n0} = 1$. Let

$$a_{ij} = \delta_{i+j}^{-(i+j)} b_{i+j,i} \quad \text{and} \quad g(x, y) = \sum a_{ij}x^i y^j.$$

Then

$$g(x, sx) = \sum \delta_n^{-n} P_n(s)x^n.$$

For $s \in E$ we have $|\delta_n^{-n} P_n(s)| \le 1$, so $g(x, sx)$ is convergent. Note that $a_{0j} = \delta_j^{-j}$, which obviously implies that g is divergent, since $\lim \delta_j = 0$.

Example 3.2. We show that the condition in Theorem 1.1 that $h(x)$ is not a monomial of the form $b_k x^k$ with $\sigma k - \tau = 0$ cannot be dispensed with. Let σ, k be positive integers, and $\phi \in \mathbb{C}[[x]]$ a divergent series with $\phi(0) = 0$. Let $g(x, y) = \phi(x^k) - \phi(y)$ and $h(x) = x^k$. Then g is divergent; but $g(s^\sigma x, s^{\sigma k} h(x)) = 0$ for each $s \in \mathbb{C}$.

Example 3.3. We show that the hypothesis in Theorem 1.1 that $h(x)$ is convergent cannot be dispensed with when $\sigma\tau \le 0$. (By Theorem 1.2 that hypothesis can be dispensed with when $\sigma\tau > 0$.) The example also shows that Theorem 1.2 fails for $\sigma\tau \le 0$.

Suppose that $\tau \le 0$, $\sigma > 0$. Let $u(x) = x + \cdots$ be a divergent series. Let $h(x), \phi(x)$ be the series satisfying $\phi(u(x)) = x$ and $x^{|\tau|}h(x)^\sigma = u(x^{\sigma+|\tau|})$. Then ϕ, h are divergent. Let $f(x, y) = \phi(x^{|\tau|}y^\sigma)$. Then f is divergent; but

$$f(s^\sigma x, s^\tau h(x)) = x^{\sigma+|\tau|} \quad \text{for each } s \in \mathbb{C} \setminus \{0\}.$$

Now we consider the case where $\sigma = 0$, $\tau = 1$. Let $h(x) = x + \cdots$ be a divergent series, and let $\phi(x)$ be the series satisfying $h(x)\phi(x) = x^2$. Then ϕ is divergent. Let $f(x, y) = \phi(x)y$. Then f is divergent; but $f(x, sh(x)) = sx^2$ for each $s \in \mathbb{C}$.

Acknowledgements

We thank T. S. Neelon for informing us of the result of B. Malgrange [1966] and pointing out several misprints. We are also grateful to the referee for his exhaustive and highly professional report. The changes implemented due to the suggestions in the report have significantly improved the paper.

References

[Abhyankar and Moh 1970] S. S. Abhyankar and T. T. Moh, "A reduction theorem for divergent power series", *J. Reine Angew. Math.* **241** (1970), 27–33. MR 41 #3800 Zbl 0191.04403

[Ahlfors 1966] L. V. Ahlfors, *Complex analysis: An introduction to the theory of analytic functions of one complex variable*, 2nd ed., McGraw-Hill, New York, 1966. MR 32 #5844 Zbl 0154.31904

[Ahlfors 1973] L. V. Ahlfors, *Conformal invariants: topics in geometric function theory*, McGraw-Hill, New York, 1973. MR 50 #10211 Zbl 0272.30012

[Bierstone et al. 1991] E. Bierstone, P. D. Milman, and A. Parusiński, "A function which is arc-analytic but not continuous", *Proc. Amer. Math. Soc.* **113**:2 (1991), 419–423. MR 91m:32008 Zbl 0739.32009

[Bochnak 1970] J. Bochnak, "Analytic functions in Banach spaces", *Studia Math.* **35** (1970), 273–292. MR 42 #8275 Zbl 0199.18402

[Fuks 1963] B. A. Fuks, *Theory of analytic functions of several complex variables*, American Mathematical Society, Providence, R.I., 1963. MR 29 #6049 Zbl 0138.30902

[Griffiths and Harris 1978] P. Griffiths and J. Harris, *Principles of algebraic geometry*, Wiley, New York, 1978. MR 80b:14001 Zbl 0408.14001

[Lelong 1951] P. Lelong, "On a problem of M. A. Zorn", *Proc. Amer. Math. Soc.* **2** (1951), 11–19. MR 12,694a

[Levenberg and Molzon 1988] N. Levenberg and R. E. Molzon, "Convergence sets of a formal power series", *Math. Z.* **197**:3 (1988), 411–420. MR 89b:32002 Zbl 0617.32001

[Malgrange 1966] B. Malgrange, *Ideals of differentiable functions*, Tata Studies in Mathematics **3**, Oxford Univ. Press, London, 1966. MR 35 #3446 Zbl 0177.17902

[Neelon 2004] T. S. Neelon, "Ultradifferentiable functions on smooth plane curves", *J. Math. Anal. Appl.* **299**:1 (2004), 61–71. MR 2005h:26005 Zbl 1092.26018

[Neelon 2006] T. Neelon, "A Bernstein-Walsh type inequality and applications", *Canad. Math. Bull.* **49**:2 (2006), 256–264. MR 2007b:26066

[Neelon 2009] T. Neelon, "Restrictions of power series and functions to algebraic surfaces", *Analysis* (*Munich*) **29**:1 (2009), 1–15. MR 2010d:32001 Zbl 1179.26088

[Ree 1949] R. Ree, "On a problem of Max A. Zorn", *Bull. Amer. Math. Soc.* **55** (1949), 575–576. MR 11,25b Zbl 0032.40403

[Sathaye 1976] A. Sathaye, "Convergence sets of divergent power series", *J. Reine Angew. Math.* **283/284** (1976), 86–98. MR 53 #5553 Zbl 0334.13010

[Sibony 1985] N. Sibony, "Sur la frontière de Shilov des domaines de $\mathbb{C}^n$", *Math. Ann.* **273**:1 (1985), 115–121. MR 87d:32029 Zbl 0573.32017

[Siciak 1970] J. Siciak, "A characterization of analytic functions of n real variables", *Studia Math.* **35** (1970), 293–297. MR 43 #4986 Zbl 0197.05801

[Siciak 1982] J. Siciak, "Extremal plurisubharmonic functions and capacities in $\mathbb{C}^n$", Kokyuroku Math. 14, Sophia University, Tokyo, 1982.

[Spallek et al. 1990] K. Spallek, P. Tworzewski, and T. Winiarski, "Osgood-Hartogs-theorems of mixed type", *Math. Ann.* **288**:1 (1990), 75–88. MR 92b:32019 Zbl 0712.32002

[Zorn 1947] M. A. Zorn, "Note on power series", *Bull. Amer. Math. Soc.* **53** (1947), 791–792. MR 9,139d Zbl 0031.29601

Received May 5, 2010. Revised August 24, 2010.

BUMA L. FRIDMAN
DEPARTMENT OF MATHEMATICS
WICHITA STATE UNIVERSITY
WICHITA, KS 67260-0033
UNITED STATES
buma.fridman@wichita.edu

DAOWEI MA
DEPARTMENT OF MATHEMATICS
WICHITA STATE UNIVERSITY
WICHITA, KS 67260-0033
UNITED STATES
dma@math.wichita.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 1, 2011

TWISTED CAPPELL–MILLER HOLOMORPHIC AND ANALYTIC TORSIONS

RUNG-TZUNG HUANG

Recently, Cappell and Miller extended the classical construction of the analytic torsion for de Rham complexes to coupling with an arbitrary flat bundle and the holomorphic torsion for $\bar{\partial}$-complexes to coupling with an arbitrary holomorphic bundle with compatible connection of type $(1, 1)$. Cappell and Miller also studied the behavior of these torsions under metric deformations. On the other hand, Mathai and Wu generalized the classical construction of the analytic torsion to the twisted de Rham complexes with an odd degree closed form as a flux and later, more generally, to the $\mathbb{Z}_2$-graded elliptic complexes. Mathai and Wu also studied the properties of analytic torsions for the $\mathbb{Z}_2$-graded elliptic complexes, including the behavior under metric and flux deformations. In this paper we define the Cappell–Miller holomorphic torsion for the twisted Dolbeault-type complexes and the Cappell–Miller analytic torsion for the twisted de Rham complexes. We obtain variation formulas for the twisted Cappell–Miller holomorphic and analytic torsions under metric and flux deformations.

1. Introduction

Ray and Singer, in the celebrated works [1971; 1973], defined the analytic torsion for de Rham complexes and the holomorphic torsion for $\bar{\partial}$-complexes of complex manifolds. They studied properties of these torsions, including the behavior under metric deformations and coupled the Riemannian Laplacian and the $\bar{\partial}$-Laplacian with unitary flat vector bundles and obtained self-adjoint operators. Hence, the analytic torsion and holomorphic torsion are real numbers in the acyclic cases considered by Ray and Singer and are expressed as elements of real determinant line in the nonacyclic case.

Recently, Cappell and Miller [2010] extended the classical construction of the analytic torsion to coupling with an arbitrary flat bundle and the holomorphic torsion to coupling with an arbitrary holomorphic bundle with compatible connection

MSC2000: 58J52.
Keywords: determinant, analytic torsion.

of type $(1, 1)$; see Definition 3.1. This includes both unitary and flat (not necessarily unitary) bundles as special cases. However, in this general setting, the associated operators are not necessarily self-adjoint and the torsions are complex-valued. Cappell and Miller also studied the behavior of these torsions under metric deformations.

Mathai and Wu [2008; 2010b] generalized the classical construction of the Ray–Singer torsion for de Rham complexes to the twisted de Rham complex with an odd degree closed differential form H as a flux. Later, in [Mathai and Wu 2010a], they extended this to $\mathbb{Z}_2$-graded elliptic complexes. The definitions use pseudo-differential operators and residue traces. Mathai and Wu also studied the properties of analytic torsion for $\mathbb{Z}_2$-graded elliptic complexes, including the behavior under the variation of metric and flux.

Let E be a holomorphic bundle with a compatible type-$(1, 1)$ connection D (see Definition 3.1) over a complex manifold W of complex dimension n and $H \in A^{0,\bar{1}}(W, \mathbb{C})$ be a $\bar{\partial}$-closed differential form of type $(0, \text{odd})$. In Definition 3.5, for each p, $1 \leq p \leq n$, we define the twisted Cappell–Miller holomorphic torsion $\tau_{\text{holo},p}(W, E, H)$ as a nonvanishing element of the determinant line:

$$\tau_{\text{holo},p}(W, E, H) \in \operatorname{Det} H^{p,\bullet}_{\bar{\partial}_E}(W, E, H) \otimes \big(\operatorname{Det} H^{\bullet,n-p}_{D^{1,0}}(W, E, H)\big)^{(-1)^{n+1}}.$$

We show that the variation of the twisted Cappell–Miller holomorphic torsion $\tau_{\text{holo},p}(W, E, H)$ under the deformation of the metric is given by a local formula; see Theorem 3.8. We also show that along any deformation of H that fixes the cohomology class $[H]$ and the natural identification of determinant lines, the variation of the twisted Cappell–Miller holomorphic torsion $\tau_{\text{holo},p}(W, E, H)$ under the deformation of the flux is given by a local formula; see Theorem 3.12.

Let $\mathscr{E}$ be a complex flat vector bundle over a closed manifold M endowed with a flat connection ∇ and let $\mathscr{H}$ be an odd degree flux form. Then the Cappell–Miller analytic torsion $\tau(\nabla, \mathscr{H})$ (see Definition 4.2) for the twisted de Rham complexes is an element of $\operatorname{Det} H^\bullet(M, \mathscr{E} \oplus \mathscr{E}', \mathscr{H})$, where $\mathscr{E}'$ is the dual of the vector bundle $\mathscr{E}$. We show that the variation of the twisted Cappell–Miller analytic torsion $\tau(\nabla, \mathscr{H})$ under the deformation of the metric is given by a local formula; see Theorem 4.3. We also show that along any deformation of $\mathscr{H}$ that fixes the cohomology class $[\mathscr{H}]$ and the natural identification of determinant lines, the variation of the twisted Cappell–Miller analytic torsion $\tau(\nabla, \mathscr{H})$ under the deformation of the flux is given by a local formula; see Theorem 4.4. In particular, we show that if the manifold M is an odd-dimensional closed oriented manifold, then the twisted Cappell–Miller analytic torsion is independent of the Riemannian metric and the representative $\mathscr{H}$ in the cohomology class $[\mathscr{H}]$. See also [Su 2011, Section 6]. We also compare the twisted Cappell–Miller analytic torsion with the twisted refined analytic torsion [Huang 2010]; see Theorem 4.5.

In the paper just cited we defined and studied the refined analytic torsion of Braverman and Kappeler [2007; 2008b] for the twisted de Rham complexes. Later, Su [2011] defined and studied the Burghelea–Haller [2007; 2008; 2010] analytic torsion for the twisted de Rham complexes and compared the twisted Burghelea–Haller torsion with the twisted refined analytic torsion. Su [2011] also briefly discussed the twisted Cappell–Miller analytic torsion when the dimension of the manifold is odd.

The rest of the paper is organized as follows. In Section 2, we define and calculate the Cappell–Miller torsion for the $\mathbb{Z}_2$-graded finite-dimensional bigraded complex. In Section 3, we first define the Dolbeault-type bigraded complexes twisted by a flux form and its (co)homology groups. We then define the Cappell–Miller holomorphic torsion for the twisted Dolbeault-type bigraded complexes. We prove variation theorems for the twisted Cappell–Miller holomorphic torsion under metric and flux deformations. In Section 4, we first define the de Rham bigraded complex twisted by a flux form and its (co)homology groups. Then we define the Cappell–Miller analytic torsion for the twisted de Rham bigraded complex. We prove variation theorems for the twisted Cappell–Miller analytic torsion under metric and flux deformations.

Throughout this paper, a bar over an integer means taking the value modulo 2.

2. The Cappell–Miller torsion for a $\mathbb{Z}_2$-graded finite-dimensional bigraded complex

In this section we define and calculate the Cappell–Miller torsion for the $\mathbb{Z}_2$-graded finite-dimensional bigraded complex. For the $\mathbb{Z}$-graded case, see [Cappell and Miller 2010, Section 6]. Throughout this section $\boldsymbol{k}$ is a field of characteristic zero.

Determinant lines of a $\mathbb{Z}_2$-graded finite-dimensional bigraded complex. Given a $\boldsymbol{k}$-vector space V of dimension n, the determinant line of V is the line $\mathrm{Det}(V) := \bigwedge^n V$, where $\bigwedge^n V$ denotes the n-th exterior power of V. By definition, we set $\mathrm{Det}(0) := \boldsymbol{k}$. Further, we denote by $\mathrm{Det}(V)^{-1}$ the dual line of $\mathrm{Det}(V)$. Let

$$C^{\bar{0}} = C^{\mathrm{even}} = \bigoplus_{i=0}^{[m/2]} C^{2i},$$
$$C^{\bar{1}} = C^{\mathrm{odd}} = \bigoplus_{i=0}^{[(m-1)/2]} C^{2i+1},$$

where C^i, $i = 0, \ldots, m$, are finite-dimensional $\boldsymbol{k}$-vector spaces. Let

$$(C^\bullet, d) : \quad \cdots \xrightarrow{d} C^{\bar{0}} \xrightarrow{d} C^{\bar{1}} \xrightarrow{d} C^{\bar{0}} \xrightarrow{d} \cdots \tag{2-1}$$

be a $\mathbb{Z}_2$-graded cochain complex of finite dimensional $\boldsymbol{k}$-vector spaces. Denote by $H^\bullet(d) = H^{\bar{0}}(d) \oplus H^{\bar{1}}(d)$ its cohomology. Set

$$\begin{aligned} \mathrm{Det}(C^\bullet) &:= \mathrm{Det}(C^{\bar{0}}) \otimes \mathrm{Det}(C^{\bar{1}})^{-1}, \\ \mathrm{Det}(H^\bullet(d)) &:= \mathrm{Det}(H^{\bar{0}}(d)) \otimes \mathrm{Det}(H^{\bar{1}}(d))^{-1}. \end{aligned} \tag{2-2}$$

Assume that $C^\bullet$ has another differential $d^* : C^{\bar{k}} \to C^{\overline{k-1}}$ giving the complex

$$(C^\bullet, d^*): \quad \cdots \xleftarrow{d^*} C^{\bar{0}} \xleftarrow{d^*} C^{\bar{1}} \xleftarrow{d^*} C^{\bar{0}} \xleftarrow{d^*} \cdots .$$

Denote its homology by $H_\bullet(d^*) = H_{\bar{0}}(d^*) \oplus H_{\bar{1}}(d^*)$. Set

$$\mathrm{Det}(H_\bullet(d^*)) := \mathrm{Det}(H_{\bar{0}}(d^*)) \otimes \mathrm{Det}(H_{\bar{1}}(d^*))^{-1}.$$

The fusion isomorphisms. (See [Braverman and Kappeler 2007, Section 2.3].) For two finite-dimensional $\boldsymbol{k}$-vector spaces V and W, we denote by $\mu_{V,W}$ the canonical fusion isomorphism

$$\mu_{V,W} : \mathrm{Det}(V) \otimes \mathrm{Det}(W) \to \mathrm{Det}(V \oplus W). \tag{2-3}$$

For $v \in \mathrm{Det}(V)$, $w \in \mathrm{Det}(W)$, we have

$$\mu_{V,W}(v \otimes w) = (-1)^{\dim V \cdot \dim W} \mu_{W,V}(w \otimes v). \tag{2-4}$$

By a slight abuse of notation, denote by $\mu_{V,W}^{-1}$ the transpose of the inverse of $\mu_{V,W}$.

Similarly, if $V_1, \dots, V_r$ are finite-dimensional $\boldsymbol{k}$-vector spaces, we define an isomorphism

$$\mu_{V_1,\dots,V_r} : \mathrm{Det}(V_1) \otimes \cdots \otimes \mathrm{Det}(V_r) \to \mathrm{Det}(V_1 \oplus \cdots \oplus V_r). \tag{2-5}$$

The isomorphism between determinant lines. For $k = 0, 1$, fix a direct sum decomposition

$$C^{\bar{k}} = B^{\bar{k}} \oplus H^{\bar{k}} \oplus A^{\bar{k}}, \tag{2-6}$$

such that $B^{\bar{k}} \oplus H^{\bar{k}} = (\mathrm{Ker}\, d) \cap C^{\bar{k}}$ and $B^{\bar{k}} = d(C^{\overline{k-1}}) = d(A^{\overline{k-1}})$. Then $H^{\bar{k}}$ is naturally isomorphic to the cohomology $H^{\bar{k}}(d)$ and d defines an isomorphism $d : A^{\bar{k}} \to B^{\overline{k+1}}$.

Fix $c_{\bar{k}} \in \mathrm{Det}(C^{\bar{k}})$ and $x_{\bar{k}} \in \mathrm{Det}(A^{\bar{k}})$. Let $d(x_{\bar{k}}) \in \mathrm{Det}(B^{\overline{k+1}})$ be the image of $x_{\bar{k}}$ under the map $\mathrm{Det}(A^{\bar{k}}) \to \mathrm{Det}(B^{\overline{k+1}})$ induced by the isomorphism $d : A^{\bar{k}} \to B^{\overline{k+1}}$. Then there is a unique element $h_{\bar{k}} \in \mathrm{Det}(H^{\bar{k}})$ such that

$$c_{\bar{k}} = \mu_{B^{\bar{k}}, H^{\bar{k}}, A^{\bar{k}}}\big(d(x_{\overline{k-1}}) \otimes h_{\bar{k}} \otimes x_{\bar{k}}\big), \tag{2-7}$$

where $\mu_{B^{\bar{k}}, H^{\bar{k}}, A^{\bar{k}}}$ is the fusion isomorphism; see (2-5) and [Braverman and Kappeler 2007, Section 2.3].

Define the canonical isomorphism

$$\phi_{C^\bullet} = \phi_{(C^\bullet,d)} : \mathrm{Det}(C^\bullet) \longrightarrow \mathrm{Det}(H^\bullet(d)) \tag{2-8}$$

by the formula

$$\phi_{C^\bullet} : c_{\bar 0} \otimes c_{\bar 1}^{-1} \mapsto h_{\bar 0} \otimes h_{\bar 1}^{-1}. \tag{2-9}$$

Following the sign convention of [Braverman and Kappeler 2007, (2-14)], Equation (2.10) of [Huang 2010] introduced a sign-refined version of the canonical isomorphism (2-8). Here we follow the sign convention of [Cappell and Miller 2010, Section 6].

Similarly, for $k = 0, 1$, fix a direct sum decomposition

$$C^{\bar k} = B_{\bar k} \oplus H_{\bar k} \oplus A_{\bar k}, \tag{2-10}$$

such that $B_{\bar k} \oplus H_{\bar k} = (\mathrm{Ker}\, d^*) \cap C^{\bar k}$ and $B_{\bar k} = d^*(C^{\overline{k+1}}) = d^*(A_{\overline{k+1}})$. Then $H_{\bar k}$ is naturally isomorphic to the homology $H_{\bar k}(d^*)$ and d^* defines an isomorphism $d^* : A_{\bar k} \to B_{\overline{k-1}}$.

Similarly, fix $c_{\bar k} \in \mathrm{Det}(C^{\bar k})$ and $y_{\bar k} \in \mathrm{Det}(A_{\bar k})$. Let $d^*(y_{\bar k}) \in \mathrm{Det}(B_{\overline{k-1}})$ denote the image of $y_{\bar k}$ under the map $\mathrm{Det}(A_{\bar k}) \to \mathrm{Det}(B_{\overline{k-1}})$ induced by the isomorphism $d^* : A_{\bar k} \to B_{\overline{k-1}}$. Then there is a unique element $h'_{\bar k} \in \mathrm{Det}(H^{\bar k})$ such that

$$c_{\bar k} = \mu_{B_{\bar k}, H_{\bar k}, A_{\bar k}}\big(d^*(y_{\overline{k+1}}) \otimes h'_{\bar k} \otimes y_{\bar k}\big), \tag{2-11}$$

where $\mu_{B_{\bar k}, H_{\bar k}, A_{\bar k}}$ is the fusion isomorphism; see (2-5) and [Braverman and Kappeler 2007, Section 2.3].

Define the canonical isomorphism

$$\phi'_{C^\bullet} = \phi'_{(C^\bullet,d^*)} : \mathrm{Det}(C^\bullet) \longrightarrow \mathrm{Det}(H_\bullet(d^*)) \tag{2-12}$$

by the formula

$$\phi'_{C^\bullet} : c_{\bar 0} \otimes c_{\bar 1}^{-1} \mapsto h'_{\bar 0} \otimes {h'_{\bar 1}}^{-1}. \tag{2-13}$$

The Cappell–Miller torsion for a $\mathbb{Z}_2$-graded finite-dimensional bigraded complex. Let $C^\bullet = C^{\bar 0} \oplus C^{\bar 1}$ and $\widetilde{C}^\bullet = \widetilde{C}^{\bar 0} \oplus \widetilde{C}^{\bar 1}$ be finite-dimensional $\mathbb{Z}_2$-graded $\boldsymbol{k}$-vector spaces. The fusion isomorphism

$$\mu_{C^\bullet, \widetilde{C}^\bullet} : \mathrm{Det}(C^\bullet) \otimes \mathrm{Det}(\widetilde{C}^\bullet) \to \mathrm{Det}(C^\bullet \oplus \widetilde{C}^\bullet)$$

is defined by the formula

$$\mu_{C^\bullet, \widetilde{C}^\bullet} := (-1)^{\mathcal{M}(C^\bullet, \widetilde{C}^\bullet)} \mu_{C^{\bar 0}, \widetilde{C}^{\bar 0}} \otimes \mu^{-1}_{C^{\bar 1}, \widetilde{C}^{\bar 1}}, \tag{2-14}$$

where

$$\mathcal{M}(C^\bullet, \widetilde{C}^\bullet) := \dim C^{\bar 1} \cdot \dim \widetilde{C}^{\bar 0}. \tag{2-15}$$

Consider the element $c := c_{\bar{0}} \otimes c_{\bar{1}}^{-1}$ of $\mathrm{Det}(C^\bullet)$. Then, for the bigraded complex $(C^\bullet, d, d^*)$, the Cappell–Miller torsion is the algebraic torsion invariant

$$\tau(C^\bullet, d, d^*) := (-1)^{S(C^\bullet)} \phi_{C^\bullet}(c)(\phi'_{C^\bullet}(c))^{-1} \in \mathrm{Det}(H^\bullet(d)) \otimes \mathrm{Det}(H_\bullet(d^*))^{-1}, \tag{2-16}$$

where $(-1)^{S(C^\bullet)}$ is defined by the formula

$$S(C^\bullet) := \sum_{k=0,1} \big(\dim B_{\overline{k-1}} \cdot \dim B^{\overline{k+1}} + \dim B^{\overline{k+1}} \cdot \dim H_{\bar{k}} + \dim B_{\overline{k-1}} \cdot \dim H^{\bar{k}}\big). \tag{2-17}$$

Calculation of the $\mathbb{Z}_2$-graded Cappell–Miller torsion. We first compute the torsion in the case that the combinatorial Laplacian $\Delta := d^*d + dd^*$ is bijective.

For $k = 0, 1$, define

$$C_+^{\bar{k}} := \mathrm{Ker}\, d^* \cap C^{\bar{k}}, \quad C_-^{\bar{k}} := \mathrm{Ker}\, d \cap C^{\bar{k}}. \tag{2-18}$$

The proof of the following proposition is similar to the proof of the $\mathbb{Z}$-graded case [Cappell and Miller 2010, Section 6.2, Claim B].

Proposition 2.1. *Suppose that the combinatorial Laplacian Δ has no zero eigenvalue. Then the cohomology group $H^\bullet(d) = 0$ and the homology group $H_\bullet(d^*) = 0$. Moreover*,

$$\tau(C^\bullet, d, d^*) = \mathrm{Det}(d^*d|_{C_+^{\bar{0}}}) \cdot \mathrm{Det}(d^*d|_{C_+^{\bar{1}}})^{-1}, \tag{2-19}$$

Proof. The proof of the first assertion that $H^\bullet(d) = 0$ and $H_\bullet(d^*) = 0$ is standard, so we skip the proof.

To compute $\tau(C^\bullet, d, d^*)$ (see (2-16)), we first compute $\phi'_{C^\bullet}(c)$. For each $k = 0, 1$, we now have the direct sum decomposition

$$C^{\bar{k}} = d^* C^{\overline{k+1}} \oplus d C^{\overline{k-1}}. \tag{2-20}$$

We also have the isomorphisms

$$d : d^* C^{\overline{k+1}} \cong d C^{\bar{k}}, \quad d^* : d C^{\bar{k}} \cong d^* C^{\overline{k+1}}. \tag{2-21}$$

By (2-18), (2-20) and (2-21), we know that

$$C_+^{\bar{k}} = d^* C^{\overline{k+1}}, \quad C_-^{\bar{k}} = d C^{\overline{k-1}}. \tag{2-22}$$

By (2-6), (2-10), (2-21), (2-22) and the first assertion we know that

$$C_+^{\bar{k}} = B_{\bar{k}} \cong A^{\bar{k}}, \quad C_-^{\bar{k}} = B^{\bar{k}} \cong A_{\bar{k}}. \tag{2-23}$$

Let $\{ d^* y_{\overline{k+1},i} \mid 1 \le i \le \dim B_{\bar{k}} \}$ be a basis for $B_{\bar{k}} = d^* C^{\overline{k+1}} \cong A^{\bar{k}}$. Since

$$d^*d : d^* C^{\overline{k+1}} \to d^* C^{\overline{k+1}}$$

is an isomorphism, there is a unique vector

$$x_{\bar{k},i} \in B_{\bar{k}} = d^* C^{\overline{k+1}}$$

such that

$$d^* d x_{\bar{k},i} = d^* y_{\overline{k+1},i}. \tag{2-24}$$

Then $\{ x_{\bar{k},i} \mid 1 \le i \le \dim B_{\bar{k}} \}$ is also a basis for $B_{\bar{k}} \cong A^{\bar{k}}$. Since $d : d^* C^{\overline{k+1}} \to dC^{\bar{k}}$ is an isomorphism, it follows that $\{ dx_{\bar{k},i} \mid 1 \le i \le \dim B_{\bar{k}} \}$ is a basis for $B^{\overline{k+1}} = dC^{\bar{k}} \cong A_{\overline{k+1}}$. Hence, in view of the decomposition (2-20), we conclude that

$$\{ d^* y_{\overline{k+1},i} \mid 1 \le i \le \dim B_{\bar{k}} \} \cup \{ dx_{\overline{k-1},i} \mid 1 \le i \le \dim B_{\overline{k-1}} \}$$

forms a basis for $C^{\bar{k}}$. In particular, by the first assertion and (2-6), we have

$$\dim B^{\bar{k}} = \dim B_{\overline{k-1}}. \tag{2-25}$$

With this particular choice of basis, we set

$$\begin{aligned} y_{\overline{k+1}} &:= y_{\overline{k+1},1} \wedge \cdots \wedge y_{\overline{k+1},\dim B_{\bar{k}}} \in \operatorname{Det}(A_{\overline{k+1}}) \\ x_{\overline{k-1}} &:= x_{\overline{k-1},1} \wedge \cdots \wedge x_{\overline{k-1},\dim B_{\overline{k-1}}} \in \operatorname{Det}(B_{\overline{k-1}}). \end{aligned}$$

Let $d^* y_{\overline{k+1}}$ and $dx_{\overline{k-1}}$ be the induced elements in $\operatorname{Det}(B_{\bar{k}})$ and $\operatorname{Det}(A_{\bar{k}})$. Set

$$c_{\bar{k}} = \mu_{B_{\bar{k}},A_{\bar{k}}}(d^* y_{\overline{k+1}} \otimes dx_{\overline{k-1}}). \tag{2-26}$$

To compute $\phi'_{C^\bullet}(c)$ (see (2-13)), we need to compute $h'_{\bar{k}} \in \operatorname{Det}(H_{\bar{k}}(d^*)) \cong \boldsymbol{k}$.

If L is a complex line and $x, y \in L$ with $y \ne 0$, we denote by $[x : y] \in \boldsymbol{k}$ the unique number such that $x = [x : y]y$. Then

$$\begin{aligned} h'_{\bar{k}} &= [c_{\bar{k}} : \mu_{B_{\bar{k}},A_{\bar{k}}}(d^* y_{\overline{k+1}} \otimes dx_{\overline{k-1}})] && \text{by (2-24)} \\ &= [\mu_{B_{\bar{k}},A_{\bar{k}}}(d^* y_{\overline{k+1}} \otimes dx_{\overline{k-1}}) : \mu_{B_{\bar{k}},A_{\bar{k}}}(d^* y_{\overline{k+1}} \otimes dx_{\overline{k-1}})] && \text{by (2-26)} \\ &= 1. \end{aligned} \tag{2-27}$$

We next compute $\phi_{C^\bullet}(c)$. By (2-9), we need to compute $h_{\bar{k}}$. By our choice of basis, we have

$$\begin{aligned} h_{\bar{k}} &= [c_{\bar{k}} : \mu_{B^{\bar{k}},A^{\bar{k}}}(dx_{\overline{k-1}}) \otimes x_{\bar{k}}] \\ &= [\mu_{B_{\bar{k}},A_{\bar{k}}}(d^* y_{\overline{k+1}} \otimes dx_{\overline{k-1}}) : \mu_{B^{\bar{k}},A^{\bar{k}}}(dx_{\overline{k-1}}) \otimes x_{\bar{k}}] && \text{by (2-26)} \\ &= [\mu_{B_{\bar{k}},A_{\bar{k}}}(d^* dx_{\bar{k}} \otimes dx_{\overline{k-1}}) : \mu_{A_{\bar{k}},B_{\bar{k}}}(dx_{\overline{k-1}}) \otimes x_{\bar{k}}] && \text{by (2-23), (2-24)} \\ &= (-1)^{\dim B_{\bar{k}} \dim A_{\bar{k}}} [\mu_{B_{\bar{k}},A_{\bar{k}}}(d^* dx_{\bar{k}} \otimes dx_{\overline{k-1}}) : \mu_{B_{\bar{k}},A_{\bar{k}}}(x_{\bar{k}} \otimes dx_{\overline{k-1}})] \\ &= (-1)^{\dim B_{\overline{k-1}} \dim B^{\overline{k+1}}} \operatorname{Det}(d^* d|_{C^{\bar{k}}_+}), && \text{by (2-23), (2-25).} \end{aligned} \tag{2-28}$$

Combining (2-16), (2-17), (2-27), (2-28) with the first assertion gives (2-19). □

We now compute the torsion in the case that the combinatorial Laplacian $\Delta := d^*d + dd^*$ is *not* bijective. For simplicity, we restrict to the case $\boldsymbol{k} = \mathbb{C}$ for the rest of discussion in this section. The operator Δ maps $C^{\bar{k}}$ into itself. For an arbitrary interval $\mathcal{I} \subset [0, \infty)$, let $C^{\bar{k}}_{\mathcal{I}} \subset C^{\bar{k}}$ denote the linear span of the generalized eigenvectors of the restriction of Δ to $C^{\bar{k}}$, corresponding to eigenvalue λ with $|\lambda| \in \mathcal{I}$. Since both d and d^* commute with Δ, we have $d(C^{\bar{k}}_{\mathcal{I}}) \subset C^{\overline{k+1}}_{\mathcal{I}}$ and $d^*(C^{\bar{k}}_{\mathcal{I}}) \subset C^{\overline{k-1}}_{\mathcal{I}}$. Hence, we obtain a subcomplex $C^{\bullet}_{\mathcal{I}}$ of $C^{\bullet}$. We denote by $H^{\bullet}_{\mathcal{I}}(d)$ the cohomology of the complex $(C^{\bullet}_{\mathcal{I}}, d_{\mathcal{I}})$ and $H_{\bullet,\mathcal{I}}(d^*)$ the homology of the complex $(C^{\bullet}_{\mathcal{I}}, d^*_{\mathcal{I}})$. Denote by $d_{\mathcal{I}}$ and $d^*_{\mathcal{I}}$ the restrictions of d and d^* to $C^{\bar{k}}_{\mathcal{I}}$ and denote by $\Delta_{\mathcal{I}}$ the restriction of Δ to $C^{\bar{k}}_{\mathcal{I}}$. Then $\Delta_{\mathcal{I}} = d^*_{\mathcal{I}} d_{\mathcal{I}} + d_{\mathcal{I}} d^*_{\mathcal{I}}$. For $k = 0, 1$, we also denote by $C^{\bar{k}}_{\pm,\mathcal{I}}$ the restrictions of $C^{\bar{k}}_{\pm}$ to $C^{\bar{k}}_{\mathcal{I}}$.

For each $\lambda \geq 0$, we have $C^{\bullet} = C^{\bullet}_{[0,\lambda]} \oplus C^{\bullet}_{(\lambda,\infty)}$. Then $H^{\bullet}_{(\lambda,\infty)}(d) = 0$ whereas $H^{\bullet}_{[0,\lambda]}(d) \cong H^{\bullet}(d)$ and $H_{\bullet,(\lambda,\infty)}(d^*) = 0$ whereas $H_{\bullet,[0,\lambda]}(d^*) \cong H_{\bullet}(d^*)$. Hence there are canonical isomorphisms

$$\Phi_\lambda : \mathrm{Det}(H^{\bullet}_{(\lambda,\infty)}(d)) \to \mathbb{C}, \qquad \Psi_\lambda : \mathrm{Det}(H^{\bullet}_{[0,\lambda]}(d)) \to \mathrm{Det}(H^{\bullet}(d))$$
$$\Phi'_\lambda : \mathrm{Det}(H_{\bullet,(\lambda,\infty)}(d^*)) \to \mathbb{C}, \quad \Psi'^{*}_\lambda : \mathrm{Det}(H_{\bullet,[0,\lambda]}(d^*))^{-1} \to \mathrm{Det}(H_{\bullet}(d^*))^{-1}.$$

In the sequel, we will write t for $\Phi_\lambda(t) \in \mathbb{C}$ and t' for $\Phi'_\lambda(t') \in \mathbb{C}$.

Proposition 2.2. *Let $(C^{\bullet}, d, d^*)$ be a $\mathbb{Z}_2$-graded bigraded complex of finite-dimensional $\boldsymbol{k}$-vector spaces. Then, for each $\lambda \geq 0$,*

$$\tau(C^{\bullet}, d, d^*) = \mathrm{Det}(d^*d|_{C^{\bar{0}}_{+,(\lambda,\infty)}}) \cdot \mathrm{Det}(d^*d|_{C^{\bar{1}}_{+,(\lambda,\infty)}})^{-1} \cdot \tau(C^{\bullet}_{[0,\lambda]}, d, d^*), \tag{2-29}$$

where we view $\tau(C^{\bullet}_{[0,\lambda]}, d, d^*)$ as an element of $\mathrm{Det}(H^{\bullet}(d)) \otimes \mathrm{Det}(H_{\bullet}(d^*))^{-1}$ *via the canonical isomorphism $\Psi_\lambda \otimes \Psi'^{*}_\lambda : \mathrm{Det}(H^{\bullet}_{[0,\lambda]}(d)) \otimes \mathrm{Det}(H_{\bullet,[0,\lambda]}(d^*))^{-1} \to$* $\mathrm{Det}(H^{\bullet}(d)) \otimes \mathrm{Det}(H_{\bullet}(d^*))^{-1}$.

In particular, the right side of (2-29) *is independent of* $\lambda \geq 0$.

Proof. Recall the natural isomorphisms

$$\begin{aligned}\mathrm{Det}(H^{\bar{k}}_{[0,\lambda]}(d) \otimes H^{\bar{k}}_{(\lambda,\infty)}(d)) &\cong \mathrm{Det}(H^{\bar{k}}_{[0,\lambda]}(d) \oplus H^{\bar{k}}_{(\lambda,\infty)}(d)) \\ &= \mathrm{Det}(H^{\bar{k}}(d)),\end{aligned} \tag{2-30}$$

$$\begin{aligned}\mathrm{Det}(H_{\bar{k},[0,\lambda]}(d^*) \otimes H_{\bar{k},(\lambda,\infty)}(d^*)) &\cong \mathrm{Det}(H_{\bar{k},[0,\lambda]}(d^*) \oplus H_{\bar{k},(\lambda,\infty)}(d^*)) \\ &= \mathrm{Det}(H_{\bar{k}}(d^*)).\end{aligned} \tag{2-31}$$

From (2-16), Proposition 2.1, (2-30) and (2-31) we obtain the result. □

3. Twisted Cappell–Miller holomorphic torsion

In this section we first review the $\bar{\partial}$-Laplacian for a holomorphic bundle with compatible type (1,1) connection introduced in [Cappell and Miller 2010]. Then we define the Dolbeault-type bigraded complexes twisted by a flux form and its cohomology and homology groups. We define the Cappell–Miller holomorphic torsion for the twisted Dolbeault-type bigraded complexes. We also prove variation theorems for the twisted Cappell–Miller holomorphic torsion under metric and flux deformations.

The $\bar{\partial}$-Laplacian for a holomorphic bundle with compatible type $(1,1)$ connection. In this section we review some materials from [Cappell and Miller 2010]; see also [Liu and Yu 2010].

Let (W, J) be a complex manifold of complex dimension n with the complex structure J and let g^W be any Hermitian metric on TW. Let $E \to W$ be a holomorphic bundle over W endowed with a linear connection D and let h^E be a Hermitian metric on E.

The complex structure J induces a splitting $TW \otimes_{\mathbb{R}} \mathbb{C} = T^{(1,0)}W \oplus T^{(0,1)}W$, where $T^{(1,0)}W$ and $T^{(0,1)}W$ are eigenbundles of J corresponding to eigenvalues i and $-i$, respectively. Let $T^{*(1,0)}W$ and $T^{*(0,1)}W$ be the corresponding dual bundles. For $0 \le p, q \le n$, let

$$A^{p,q}(W, E) = \Gamma\big(W, \textstyle\bigwedge^p(T^{*(1,0)}W) \otimes \bigwedge^q(T^{*(0,1)}W) \otimes E\big)$$

be the space of smooth (p,q)-forms on W with values in E. Set

$$A^{\bullet,\bullet}(W, E) = \bigoplus_{p,q=0}^{n} A^{p,q}(W, E).$$

Let $\bar{\partial} : A^{p,q}(W, \mathbb{C}) \to W^{p,q+1}(W, \mathbb{C})$ and $\partial : A^{p,q}(W, \mathbb{C}) \to A^{p+1,q}(W, \mathbb{C})$ be the standard operators obtained by decomposing by type the exterior derivative

$$d = \bar{\partial} + \partial$$

acting on complex-valued smooth forms of type (p,q). From $d^2 = 0$, we have $\bar{\partial}^2 = 0$, $\partial^2 = 0$.

Since E is holomorphic, the operator $\bar{\partial}$ on $A^{\bullet,\bullet}(W, \mathbb{C})$ has a unique natural extension to $A^{\bullet,\bullet}(W, E)$ (see [Cappell and Miller 2010, page 139])

$$\bar{\partial}_E : A^{p,q}(W, E) \to W^{p,q+1}(W, E).$$

Under the splitting $\Gamma(W, (T^*W \otimes_{\mathbb{R}} \mathbb{C}) \otimes_{\mathbb{C}} E) = A^{1,0}(W, E) \oplus A^{0,1}(W, E)$, the connection D decomposes as a sum $D = D^{1,0} \oplus D^{0,1}$ with

$$D^{1,0} : \Gamma(W, E) \to A^{1,0}(W, E), \quad D^{0,1} : \Gamma(W, E) \to A^{0,1}(W, E).$$

Extend the connection D on $\Gamma(W, E)$ in a unique way to $A^{\bullet,\bullet}(W, E)$ by the Leibniz formula [Berline et al. 2004, page 21]. The extended D again decomposes as a sum $D = D^{1,0} + D^{0,1}$ also satisfying the Leibniz formula [Berline et al. 2004, page 131].

Recall the following definition from [Cappell and Miller 2010, pages 139–140] or [Liu and Yu 2010, Definition 2.1].

Definition 3.1. The connection D is said to be *compatible* with the holomorphic structure on E if $D^{0,1} = \bar{\partial}_E$. The connection D is said to be of *type* $(1, 1)$ if the curvature D^2 is of type $(1, 1)$, that is, $(D^{1,0})^2 = 0$ and $(D^{0,1})^2 = 0$.

The complex Hodge star operator $\star$ acting on forms is a complex conjugate linear mapping

$$\star : A^{p,q}(W, \mathbb{C}) \to A^{n-p,n-q}(W, \mathbb{C})$$

induced by a conjugate linear bundle isomorphism; see [Cappell and Miller 2010, page 141] for this and other statements on this page.

The natural conjugate mapping

$$\mathrm{conj} : A^{p,q}(W, \mathbb{C}) \to A^{q,p}(W, \mathbb{C})$$

is a complex linear mapping induced by the bundle automorphism

$$T^*W \otimes_{\mathbb{R}} \mathbb{C} \to T^*W \otimes_{\mathbb{R}} \mathbb{C},\ v \otimes \lambda \mapsto v \otimes \bar{\lambda}, \qquad v \in T^*W,\ \lambda \in \mathbb{C},$$

of the complexified cotangent bundle. Define $\hat{\star} := \mathrm{conj}\,\star$. Then

$$\hat{\star} = \mathrm{conj}\,\star : A^{p,q}(W, \mathbb{C}) \to A^{n-q,n-p}(W, \mathbb{C})$$

is a complex linear mapping. Clearly, $\hat{\star} = \mathrm{conj}\,\star = \star\,\mathrm{conj}$.

As pointed out by Cappell and Miller, since $\hat{\star}$ is complex linear, it may be coupled to a complex linear bundle mapping, for example, the identity mapping. We also denote by $\hat{\star}$ the complex linear mapping

$$\hat{\star} : A^{p,q}(W, E) \to A^{n-q,n-p}(W, E).$$

Recall that the adjoint $\bar{\partial}^*$ of $\bar{\partial}$ with respect to the chosen Hermitian inner product on TW is given by

$$\bar{\partial}^* = -\star\,\bar{\partial}\,\star.$$

In particular,

$$\bar{\partial}^* = -\hat{\star}\ \mathrm{conj}\ \bar{\partial}\ \mathrm{conj}\ \hat{\star} = -\hat{\star}\,\partial\,\hat{\star}.$$

Let D be a compatible $(1, 1)$ connection. Following Cappell and Miller, we define

$$\bar{\partial}^*_{E,D^{1,0}} = -\hat{\star}\,D^{1,0}\,\hat{\star}$$

and the $\bar{\partial}$-Laplacian for the holomorphic bundle E with compatible type-$(1, 1)$ connection D by

$$\square_{E,\bar{\partial}} = \bar{\partial}_E \bar{\partial}^*_{E,D^{1,0}} + \bar{\partial}^*_{E,D^{1,0}} \bar{\partial}_E.$$

Note that $(\bar{\partial}^*_{E,D^{1,0}})^2 = 0$, since $(D^{1,0})^2 = 0$ and $\hat{\star}^2 = \star^2 = \pm 1$.

Denote by δ_E the adjoint of the $\bar\partial$-operator $\bar\partial_E$ with respect to the inner product $\langle\cdot,\cdot\rangle_E$ on $A^{\bullet,\bullet}(W,E)$ induced by the Hermitian metrics g^W and h^E. Then the associated self-adjoint $\bar\partial$-Laplacian is defined as

$$\square_E=(\bar\partial_E+\delta_E)^2=\bar\partial_E\delta_E+\delta_E\bar\partial_E.$$

Recall that, in general, the operator $\square_{E,\bar\partial}$ is not self-adjoint with respect to the inner product $\langle\cdot,\cdot\rangle_E$ on $A^{\bullet,\bullet}(W,E)$, but has the same leading symbol as the operator $\square_E$; see [Cappell and Miller 2010, Section 3]. When the connection on E is compatible with the Hermitian inner product $\langle\cdot,\cdot\rangle_E$ on $A^{\bullet,\bullet}(W,E)$, the operator $\square_{E,\bar\partial}$ recovers the self-adjoint operators considered in [Bismut 1993; Bismut et al. 1988a; 1988b; 1988c; 1990; Bismut and Lebeau 1989; 1991]. When the bundle E is unitary flat, the operator $\square_{E,\bar\partial}$ recovers the self-adjoint operators of [Ray and Singer 1973]. For more details about the operator $\square_{E,\bar\partial}$, see [Cappell and Miller 2010].

Twisted Dolbeault-type cohomology and homology groups. For each $0\le p\le n$, denote by $A^{p,\bar 0}(W,E):=A^{p,\mathrm{even}}(W,E)$ and $A^{p,\bar 1}(W,E):=A^{p,\mathrm{odd}}(W,E)$. Let $H\in A^{0,\bar 1}(W,\mathbb{C})$ and $\bar\partial_E^H:=\bar\partial_E+H\wedge\cdot$. We assume that $\bar\partial H=0$. Then, as in the de Rham case, $(\bar\partial_E^H)^2=0$. Hence, we can consider the twisted complex

$$(A^{p,\bullet}(W,E),\bar\partial_E^H):\ \cdots\xrightarrow{\bar\partial_E^H}A^{p,\bar 0}(W,E)\xrightarrow{\bar\partial_E^H}A^{p,\bar 1}(W,E)\xrightarrow{\bar\partial_E^H}A^{p,\bar 0}(W,E)\xrightarrow{\bar\partial_E^H}\cdots.$$

Define the twisted Dolbeault-type cohomology groups of $(A^{p,\bullet}(W,E),\bar\partial_E^H)$ as

$$H^{p,\bar k}_{\bar\partial_E}(W,E,H):=\frac{\mathrm{Ker}(\bar\partial_E^H:A^{p,\bar k}(W,E)\to A^{p,\overline{k+1}}(W,E))}{\mathrm{Im}(\bar\partial_E^H:A^{p,\overline{k-1}}(W,E)\to A^{p,\bar k}(W,E))},\quad k=0,1.$$

Define $\overline{H}:=\mathrm{conj}\,H$. Let $D^{1,0}_{\overline{H}}:=D^{1,0}+\overline{H}\wedge\cdot$. Then $(D^{1,0}_{\overline{H}})^2=0$. Hence, we can also consider the twisted complex

$$(A^{\bullet,p}(W,E),D^{1,0}_{\overline{H}}):\cdots\xrightarrow{D^{1,0}_{\overline{H}}}A^{\bar 0,p}(W,E)\xrightarrow{D^{1,0}_{\overline{H}}}A^{\bar 1,p}(W,E)\xrightarrow{D^{1,0}_{\overline{H}}}A^{\bar 0,p}(W,E)\xrightarrow{D^{1,0}_{\overline{H}}}\cdots.$$

Define the twisted Dolbeault-type cohomology groups of $(A^{\bullet,p}(W,E),D^{1,0}_{\overline{H}})$ as

$$H^{\bar k,p}_{D^{1,0}}(W,E,H):=\frac{\mathrm{Ker}(D^{1,0}_{\overline{H}}:A^{\bar k,p}(W,E)\to A^{\overline{k+1},p}(W,E))}{\mathrm{Im}(D^{1,0}_{\overline{H}}:A^{\overline{k-1},p}(W,E)\to A^{\bar k,p}(W,E))},\quad k=0,1.$$

Define $\bar\partial^{*,H}_{E,D^{1,0}}:=-\hat\star\,(D^{1,0}+\mathrm{conj}\,H\wedge\cdot)\,\hat\star=-\hat\star\,D^{1,0}_{\overline{H}}\,\hat\star$. Then $(\bar\partial^{*,H}_{E,D^{1,0}})^2=0$. Again we can consider the twisted complex

$$\begin{aligned}(A^{p,\bullet}(W,E),\bar\partial^{*,H}_{E,D^{1,0}}):\ \cdots\xleftarrow{\bar\partial^{*,H}_{E,D^{1,0}}}A^{p,\bar 0}(W,E)\xleftarrow{\bar\partial^{*,H}_{E,D^{1,0}}}A^{p,\bar 1}(W,E)&\\ \xleftarrow{\bar\partial^{*,H}_{E,D^{1,0}}}A^{p,\bar 0}(W,E)\xleftarrow{\bar\partial^{*,H}_{E,D^{1,0}}}\cdots&.\end{aligned}$$

Define the twisted Dolbeault-type homology groups of $(A^{p,\bullet}(W,E), \bar\partial^{*,H}_{E,D^{1,0}})$ as

$$H_{\bar k}(A^{p,\bullet}(W,E), \bar\partial^{*,H}_{E,D^{1,0}}) := \frac{\mathrm{Ker}(\bar\partial^{*,H}_{E,D^{1,0}} : A^{p,\bar k}(W,E) \to A^{p,\overline{k-1}}(W,E))}{\mathrm{Im}(\bar\partial^{*,H}_{E,D^{1,0}} : A^{p,\overline{k+1}}(W,E) \to A^{p,\bar k}(W,E))}, \; k=0,1.$$

The operator $\hat\star$ induces a $\mathbb{C}$-linear isomorphism from $(A^{p,\bullet}(W,E), \bar\partial^{*,H}_{E,D^{1,0}})$ to $(A^{\overline{n-\bullet},n-p}(W,E), \pm D^{1,0}_{\bar H})$. Hence, as in the $\mathbb{Z}$-graded case (see [Cappell and Miller 2010, page 151] or [Liu and Yu 2010, (2.19)]), we have the isomorphism

$$H^{\overline{n-k},n-p}_{D^{1,0}}(W,E,H) \cong H_{\bar k}(A^{p,\bullet}(W,E), \bar\partial^{*,H}_{E,D^{1,0}}), \quad k=0,1. \tag{3-1}$$

ζ-function and ζ-regularized determinant. In this section we briefly recall some definitions of ζ-regularized determinants of non-self-adjoint elliptic operators. See [Braverman and Kappeler 2007, Section 6] for more details. Let F be a complex (respectively, holomorphic) vector bundle over a closed smooth (respectively, complex) manifold N. Let $D : C^\infty(N,F) \to C^\infty(N,F)$ be an elliptic differential operator of order $m \ge 1$. Assume that θ is an Agmon angle; see, for example, [Braverman and Kappeler 2007, Definition 6.3]. Let $\Pi : L^2(N,F) \to L^2(N,F)$ denote the spectral projection of D corresponding to all nonzero eigenvalues of D. The ζ-function $\zeta_\theta(s,D)$ of D is defined as

$$\zeta_\theta(s,D) = \mathrm{Tr}\,\Pi D_\theta^{-s}, \quad \mathrm{Re}\, s > \frac{\dim N}{m}. \tag{3-2}$$

Seeley [1967] (see also [Shubin 2001]) showed that $\zeta_\theta(s,D)$ has a meromorphic extension to the whole complex plane and that 0 is a regular value of $\zeta_\theta(s,D)$.

Definition 3.2. The ζ-regularized determinant of D is defined by the formula

$$\mathrm{Det}'_\theta(D) := \exp\Bigl(-\frac{d}{ds}\Big|_{s=0} \zeta_\theta(s,D)\Bigr).$$

Define

$$\mathrm{LDet}'_\theta(D) = -\frac{d}{ds}\Big|_{s=0} \zeta_\theta(s,D).$$

Let Q be a 0-th order pseudo-differential projection, that is, a 0-th order pseudo-differential operator satisfying $Q^2 = Q$. Set

$$\zeta_\theta(s,Q,D) = \mathrm{Tr}\, Q\Pi D_\theta^{-s}, \quad \mathrm{Re}\, s > \frac{\dim M}{m}. \tag{3-3}$$

The function $\zeta_\theta(s,Q,D)$ also has a meromorphic extension to the whole complex plane and by [Wodzicki 1984, Section 7], it is regular at 0.

Definition 3.3. Suppose that Q is a 0-th order pseudo-differential projection commuting with D. Then $V := \mathrm{Im}\, Q$ is D invariant subspace of $C^\infty(M,E)$. The ζ-regularized determinant of the restriction $D|_V$ of D to V is defined by the formula

$$\mathrm{Det}'_\theta(D|_V) := e^{\mathrm{LDet}'_\theta(D|_V)},$$

where

$$\mathrm{LDet}'_\theta(D|_V) = -\frac{d}{ds}\Big|_{s=0}\zeta_\theta(s, Q, D). \tag{3-4}$$

Remark 3.4. The prime in Det'_θ and LDet'_θ indicates that we ignore the zero eigenvalues of the operator in the definition of the regularized determinant. If the operator is invertible we usually omit the prime and write Det_θ and LDet_θ instead.

Twisted Cappell–Miller holomorphic torsion. For each $0 \le p \le n$, the twisted flat $\bar\partial$-Laplacian, defined as

$$\square^H_{E,\bar\partial} := (\bar\partial^H_E + \bar\partial^{*,H}_{E,D^{1,0}})^2,$$

maps $A^{p,\bar k}(W, E)$, $k = 0, 1$, into itself. Suppose that $\mathcal{I}$ is an interval of the form $[0, \lambda]$, $(\lambda, \mu]$ or $(\lambda, \infty)(\mu > \lambda \ge 0)$. Denote by $\Pi^{E,\mathcal{I}}$ the spectral projection of $\square^H_{E,\bar\partial}$ corresponding to the set of generalized eigenvalues, whose absolute values lie in $\mathcal{I}$. Set

$$A^{p,\bar k}_{\mathcal{I}}(W, E) := \Pi^{E,\mathcal{I}}(A^{p,\bar k}(W, E)) \subset A^{p,\bar k}(W, E), \quad k = 0, 1.$$

If the interval $\mathcal{I}$ is bounded, then for each $0 \le p \le n$, the space $A^{p,\bar k}_{\mathcal{I}}(W, E)$, $k = 0, 1$, is finite-dimensional. The differentials $\bar\partial^H_E$ and $\bar\partial^{*,H}_{E,D^{1,0}}$ commute with $\square^H_{E,\bar\partial}$, so the subspace $A^{p,\bar k}_{\mathcal{I}}(W, E)$ is a subcomplex of the twisted bigraded complex $(A^{p,\bullet}(W, E), \bar\partial^H_E, \bar\partial^{*,H}_{E,D^{1,0}})$. Clearly, for each $\lambda \ge 0$, the complex $A^{p,\bar k}_{(\lambda,\infty)}(W, E)$ is doubly acyclic, that is,

$$H^{\bar k}(A^{p,\bullet}_{(\lambda,\infty)}(W, E), \bar\partial^H_E) = 0 \quad \text{and} \quad H_{\bar k}(A^{p,\bullet}_{(\lambda,\infty)}(W, E), \bar\partial^{*,H}_{E,D^{1,0}}) = 0.$$

Since

$$A^{p,\bar k}(W, E) = A^{p,\bar k}_{[0,\lambda]}(W, E) \oplus A^{p,\bar k}_{(\lambda,\infty)}(W, E),$$

we have the isomorphisms

$$H^{\bar k}(A^{p,\bullet}_{[0,\lambda]}(W, E), \bar\partial^H_E) \cong H^{p,\bar k}_{\bar\partial_E}(W, E, H)$$

and, by (3-1),

$$H_{\bar k}(A^{p,\bullet}_{[0,\lambda]}(W, E), \bar\partial^{*,H}_{E,D^{1,0}}) \cong H^{\overline{n-k}}(A^{\bullet,n-p}_{[0,\lambda]}(W, E), \pm D^{1,0}_H) \cong H^{\overline{n-k},n-p}_{D^{1,0}}(W, E, H).$$

In particular, we have the isomorphisms

$$\mathrm{Det}\, H^\bullet(A^{p,\bullet}_{[0,\lambda]}(W, E), \bar\partial^H_E) \cong \mathrm{Det}\, H^{p,\bullet}_{\bar\partial_E}(W, E, H), \tag{3-5}$$

$$\mathrm{Det}\, H_\bullet(A^{p,\bullet}_{[0,\lambda]}(W, E), \bar\partial^{*,H}_{E,D^{1,0}}) \cong \mathrm{Det}\, H^{\overline{n-\bullet},n-p}_{D^{1,0}}(W, E, H). \tag{3-6}$$

For any $\lambda \ge 0$, $0 \le p \le n$, let $\tau_{p,[0,\lambda]}$ denote the Cappell–Miller torsion of the twisted bigraded complex $(A^{p,\bar k}_{[0,\lambda]}(W, E), \bar\partial^H_E, \bar\partial^{*,H}_{E,D^{1,0}})$; see (2-16). Then, by (3-5) and (3-6), we can view $\tau_{p,[0,\lambda]}$ as an element of the determinant line

$$\begin{aligned}(3\text{-}7)\qquad \tau_{p,[0,\lambda]} &\in \operatorname{Det} H^{p,\bullet}_{\bar\partial_E}(W,E,H)\otimes\big(\operatorname{Det} H^{\overline{n-\bullet},n-p}_{D^{1,0}}(W,E,H)\big)^{-1}\\ &\cong \operatorname{Det} H^{p,\bullet}_{\bar\partial_E}(W,E,H)\otimes\big(\operatorname{Det} H^{\bullet,n-p}_{D^{1,0}}(W,E,H)\big)^{(-1)^{n+1}}.\end{aligned}$$

For each $k=0,1$ and each $0\le p\le n$, set

$$\begin{aligned}A^{p,\bar k}_{+,\mathcal{I}}(W,E) &:= \operatorname{Ker}(\bar\partial^H_E\bar\partial^{*,H}_{E,D^{1,0}})\cap A^{p,\bar k}_{\mathcal{I}}(W,E),\\ A^{p,\bar k}_{-,\mathcal{I}}(W,E) &:= \operatorname{Ker}(\bar\partial^{*,H}_{E,D^{1,0}}\bar\partial^H_E)\cap A^{p,\bar k}_{\mathcal{I}}(W,E).\end{aligned}$$

Clearly,

$$A^{p,\bar k}_{\mathcal{I}}(W,E) = A^{p,\bar k}_{+,\mathcal{I}}(W,E)\oplus A^{p,\bar k}_{-,\mathcal{I}}(W,E),\ \text{ if } 0\notin\mathcal{I}.$$

Let $\theta\in(0,2\pi)$ be an Agmon angle of the operator $\Box^H_{E,\bar\partial}$; see, for example, [Braverman and Kappeler 2007, Section 6]. Since the leading symbol of the operator $\Box^H_{E,\bar\partial}$ is positive definite, the ζ-regularized determinant

$$\operatorname{Det}_\theta(\bar\partial^{*,H}_{E,D^{1,0}}\bar\partial^H_E)|_{A^{p,\bar k}_{+,\mathcal{I}}(W,E)}$$

is independent of the choice of the Agmon angle θ of the operator $\Box^H_{E,\bar\partial}$.

For any $0\le\lambda\le\mu\le\infty$, one easily sees that

$$\begin{aligned}(3\text{-}8)\quad &\prod_{k=0,1}\big(\operatorname{Det}_\theta(\bar\partial^{*,H}_{E,D^{1,0}}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\big)^{(-1)^k}\\ &= \Big(\prod_{k=0,1}\big(\operatorname{Det}_\theta(\bar\partial^{*,H}_{E,D^{1,0}}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\mu)}(W,E)}\big)^{(-1)^k}\Big)\\ &\qquad\cdot\Big(\prod_{k=0,1}\big(\operatorname{Det}_\theta(\bar\partial^{*,H}_{E,D^{1,0}}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\mu,\infty)}(W,E)}\big)^{(-1)^k}\Big)\end{aligned}$$

By Proposition 2.2 and (3-8), we know that the element

$$(3\text{-}9)\qquad \tau_{\mathrm{holo},p}(W,E,H) := \tau_{p,[0,\lambda]}\cdot\prod_{k=0,1}\big(\operatorname{Det}_\theta(\bar\partial^{*,H}_{E,D^{1,0}}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\big)^{(-1)^k}$$

is independent of the choice of λ. It is also independent of the choice of the Agmon angle $\theta\in(0,2\pi)$ of the operator $\Box^H_{E,\bar\partial}$.

Definition 3.5. The nonvanishing element of the determinant

$$\tau_{\mathrm{holo},p}(W,E,H)\in \operatorname{Det} H^{p,\bullet}_{\bar\partial_E}(W,E,H)\otimes\big(\operatorname{Det} H^{\bullet,n-p}_{D^{1,0}}(W,E,H)\big)^{(-1)^{n+1}}$$

defined in (3-9) is called the twisted Cappell–Miller holomorphic torsion.

Twisted Cappell–Miller holomorphic torsion under metric deformation. Let g_u^W, $u \in \mathbb{R}$, be a smooth family of Hermitian metrics on the complex manifold W. Denote by $\star_u$ the Hodge star operators associated to the metrics g_u^W and denote by

$$\bar\partial^{*,H}_{E,D^{1,0},u} := -\hat\star_u\,(D^{1,0} + \operatorname{conj} H \wedge \cdot\,)\,\hat\star_u.$$

Let $\square^H_{E,\bar\partial,u} = (\bar\partial^H_E + \bar\partial^{*,H}_{E,D^{1,0},u})^2$ be the flat Laplacian operators associated to the metrics g_u^W.

Fix $u_0 \in \mathbb{R}$ and choose $\lambda \ge 0$ so that there are no eigenvalues of $\square^H_{E,\bar\partial,u}$ whose absolute values are equal to λ. Then there exists $\delta > 0$ such that the same is true for all $u \in (u_0-\delta, u_0+\delta)$. In particular, if we denote by $A^{p,\bullet}_{[0,\lambda],u}(W, E)$ the span of the generalized eigenvectors of $\square^H_{E,\bar\partial,u}$ corresponding to eigenvalues with absolute value $\le \lambda$, then $\dim A^{p,\bullet}_{[0,\lambda],u}(W, E)$ is independent of $u \in (u_0-\delta, u_0+\delta)$.

For any $\lambda \ge 0$, $0 \le p \le n$, let $\tau_{p,[0,\lambda],u}$ denote the Cappell–Miller torsion of the twisted bigraded complex $(A^{p,\bullet}_{[0,\lambda]}(W, E), \bar\partial^H_E, \bar\partial^{*,H}_{E,D^{1,0},u})$. Set

$$\alpha_u = \star_u^{-1} \cdot \frac{d}{du}\star_u = \hat\star_u^{-1} \cdot \frac{d}{du}\hat\star_u.$$

Let $Q_{p,\bar k}$ be the spectral projection onto $A^{p,\bar k}_{[0,\lambda]}(W, E)$. The proof of the following lemma is similar to the proof of [Cappell and Miller 2010, Lemma 7.1], where the untwisted case was treated.

Lemma 3.6. *Under the assumptions above*, *we have*

$$\frac{d}{du}\tau_{p,[0,\lambda],u} = -\sum_{k=0,1}(-1)^k \operatorname{Tr}(\alpha_u Q_{p,\bar k}) \cdot \tau_{p,[0,\lambda],u}.$$

Lemma 3.7. *Under the assumptions above*, *we have*

$$\frac{d}{du}\Bigl(\sum_{k=0,1}(-1)^k \operatorname{LDet}_\theta(\bar\partial^{*,H}_{E,D^{1,0},u}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\Bigr) = \sum_{k=0,1}(-1)^k \operatorname{Tr}(\alpha_u Q_{p,\bar k}) + \sum_{k=0,1}(-1)^k \int_W b_{n,p,\bar k,u},$$

where $b_{n,p,\bar k,u}$ is given by a local formula.

Proof. Set

$$\begin{aligned}(3\text{-}10)\quad f(s,u) &= \sum_{k=0,1}(-1)^k \int_0^\infty t^{s-1}\operatorname{Tr}\Bigl(\exp\bigl(-t(\bar\partial^{*,H}_{E,D^{1,0},u}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\bigr)\Bigr)dt \\ &= \Gamma(s)\sum_{k=0,1}(-1)^k\zeta\bigl(s, (\bar\partial^{*,H}_{E,D^{1,0},u}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\bigr).\end{aligned}$$

The equality

$$(3\text{-}11)\qquad \frac{d}{du}\bar\partial^{*,H}_{E,D^{1,0},u}|_{A^{p,\overline{k+1}}_{-,(\lambda,\infty)}(W,E)} = -\bigl[\alpha_u, \bar\partial^{*,H}_{E,D^{1,0},u}|_{A^{p,\overline{k+1}}_{-,(\lambda,\infty)}(W,E)}\bigr],$$

follows easily from $\bar\partial^{*,H}_{E,D^{1,0},u} := -\hat\star_u\,(D^{1,0} + \operatorname{conj} H \wedge \cdot\,)\,\hat\star_u$ and the equality

$$\star_u^{-1} \cdot \frac{d}{du}\star_u = -\frac{d}{du}\,\star_u \cdot \star_u^{-1}\,.$$

If A is of trace class and B is a bounded operator, it is well known that $\operatorname{Tr}(AB) = \operatorname{Tr}(BA)$. By this and the semigroup property of the heat operator, we have

(3-12)

$$\begin{aligned}
&\operatorname{Tr}\Big(\bar\partial^{*,H}_{E,D^{1,0},u}|_{A^{p,\overline{k+1}}_{-,(\lambda,\infty)}(W,E)}\alpha_u\bar\partial^H_E|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\exp\big(-t\,(\bar\partial^{*,H}_{E,D^{1,0},u}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\big)\Big)\\
&\quad= \operatorname{Tr}\Big(\exp\Big(-\frac{t}{2}(\bar\partial^{*,H}_{E,D^{1,0},u}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\Big)\bar\partial^{*,H}_{E,D^{1,0},u}|_{A^{p,\overline{k+1}}_{-,(\lambda,\infty)}(W,E)}\\
&\qquad\qquad\cdot\alpha_u\bar\partial^H_E|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\exp\Big(-\frac{t}{2}(\bar\partial^{*,H}_{E,D^{1,0},u}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\Big)\Big)\\
&\quad= \operatorname{Tr}\Big(\alpha_u\bar\partial^H_E|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\exp\Big(-\frac{t}{2}(\bar\partial^{*,H}_{E,D^{1,0},u}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\Big)\\
&\qquad\qquad\cdot\exp\big(-\frac{t}{2}(\bar\partial^{*,H}_{E,D^{1,0},u}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\big)\bar\partial^{*,H}_{E,D^{1,0},u}|_{A^{p,\overline{k+1}}_{-,(\lambda,\infty)}(W,E)}\Big)\\
&\quad= \operatorname{Tr}\Big(\alpha_u(\bar\partial^H_E\bar\partial^{*,H}_{E,D^{1,0},u})|_{A^{p,\overline{k+1}}_{-,(\lambda,\infty)}(W,E)}\exp\big(-t\,(\bar\partial^H_E\bar\partial^{*,H}_{E,D^{1,0},u})|_{A^{p,\overline{k+1}}_{-,(\lambda,\infty)}(W,E)}\big)\Big)
\end{aligned}$$

Now, by (3-10), (3-11) and (3-12), we have

(3-13)

$$\begin{aligned}
\frac{d}{du}f(s,u) &= \sum_{k=0,1}(-1)^k\int_0^\infty t^{s-1}\operatorname{Tr}\Big(t\big[\alpha_u,\,\bar\partial^{*,H}_{E,D^{1,0}}|_{A^{p,\overline{k+1}}_{-,(\lambda,\infty)}(W,E)}\big]\\
&\qquad\qquad\times\exp\big(-t\,(\bar\partial^{*,H}_{E,D^{1,0}}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\big)\Big)\,dt\\
&= \sum_{k=0,1}(-1)^k\int_0^\infty t^{s-1}\\
&\qquad\times\operatorname{Tr}\Big(t\alpha_u\Big((\bar\partial^{*,H}_{E,D^{1,0}}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\exp\big(-t\,(\bar\partial^{*,H}_{E,D^{1,0}}\bar\partial^H_E)|_{A^{p,\bar k}_{+,(\lambda,\infty)}(W,E)}\big)\\
&\qquad-(\bar\partial^H_E\bar\partial^{*,H}_{E,D^{1,0},u})|_{A^{p,\overline{k+1}}_{-,(\lambda,\infty)}(W,E)}\exp\big(-t\,(\bar\partial^H_E\bar\partial^{*,H}_{E,D^{1,0},u})|_{A^{p,\overline{k+1}}_{-,(\lambda,\infty)}(W,E)}\big)\Big)\Big)\,dt\\
&= \sum_{k=0,1}(-1)^k\int_0^\infty t^s\operatorname{Tr}\Big(\alpha_u(\Box^H_{E,\bar\partial,u})|_{A^{p,\bar k}_{(\lambda,\infty)}(W,E)}\\
&\qquad\qquad\times\exp\big(-t\,(\Box^H_{E,\bar\partial,u})|_{A^{p,\bar k}_{(\lambda,\infty)}(W,E)}\big)\Big)\,dt\\
&= -\sum_{k=0,1}(-1)^k\int_0^\infty t^s\frac{d}{dt}\operatorname{Tr}\big(\alpha_u\exp\big(-t\,(\Box^H_{E,\bar\partial,u})|_{A^{p,\bar k}_{(\lambda,\infty)}(W,E)}\big)\big)\,dt\\
&= s\sum_{k=0,1}(-1)^k\int_0^\infty t^{s-1}\operatorname{Tr}\big(\alpha_u\exp\big(-t\,(\Box^H_{E,\bar\partial,u})|_{A^{p,\bar k}_{(\lambda,\infty)}(W,E)}\big)\big)\,dt,
\end{aligned}$$

where the second equality holds by (3-12) and we used integration by parts for the last equality. Since $\square^H_{E,\bar\partial,u}$ is an elliptic operator, the dimension of $A^{p,\bullet}_{[0,\lambda]}(W,E)$ is finite. Then we can rewrite (3-13) as

$$\begin{aligned}\text{(3-14)}\quad \frac{d}{du}f(s,u) = s\sum_{k=0,1}(-1)^k\int_0^1 t^{s-1}\,\mathrm{Tr}\big(\alpha_u\exp\big(-t(\square^H_{E,\bar\partial,u})|_{A^{p,\bar k}(W,E)}\big)\big)\,dt\\ +s\sum_{k=0,1}(-1)^k\int_1^\infty t^{s-1}\,\mathrm{Tr}\big(\alpha_u\exp\big(-t(\square^H_{E,\bar\partial,u})|_{A^{p,\bar k}(W,E)}\big)\big)\,dt\\ -s\sum_{k=0,1}(-1)^k\int_0^1 t^{s-1}\,\mathrm{Tr}\big(\alpha_u\exp\big(-t(\square^H_{E,\bar\partial,u})|_{A^{p,\bar k}_{[0,\lambda]}(W,E)}\big)\big)\,dt\\ -s\sum_{k=0,1}(-1)^k\int_1^\infty t^{s-1}\,\mathrm{Tr}\big(\alpha_u\exp\big(-t(\square^H_{E,\bar\partial,u})|_{A^{p,\bar k}_{[0,\lambda]}(W,E)}\big)\big)\,dt.\end{aligned}$$

Now $\dim W = 2n$ is even, so for small time asymptotic expansion for

$$\mathrm{Tr}\big(\alpha_u\exp\big(-t(\square^H_{E,\bar\partial,u})|_{A^{p,\bar k}(W,E)}\big)\big)$$

has a term $a_{n,p,\bar k,u}t^0$ in its expansion about $t=0$. That means

$$\mathrm{Tr}\big(\alpha_u\exp\big(-t(\square^H_{E,\bar\partial,u})|_{A^{p,\bar k}(W,E)}\big)\big) - a_{n,p,\bar k,u}t^0$$

does not contain a constant term as $t\downarrow 0$. Hence, the integrals

$$\sum_{k=0,1}(-1)^k\int_0^1 t^{s-1}\,\mathrm{Tr}\big(\alpha_u\exp\big(-t(\square^H_{E,\bar\partial,u})|_{A^{p,\bar k}(W,E)}\big)\big) - a_{n,p,\bar k,u}t^0\,dt$$

do not have poles at $s=0$. But the integrals

$$\sum_{k=0,1}(-1)^k\int_0^1 t^{s-1}a_{n,p,\bar k,u}t^0\,dt$$

have poles of order 1 with residue $a_{n,p,\bar k,u}$, $k=0,1$. And, because of the exponential decay of $\mathrm{Tr}\big(\alpha_u\exp\big(-t(\square^H_{E,\bar\partial,u})|_{A^{p,\bar k}(W,E)}\big)\big)$ and $\mathrm{Tr}\big(\alpha_u\exp\big(-t(\square^H_{E,\bar\partial,u})|_{A^{p,\bar k}_{[0,\lambda]}(W,E)}\big)\big)$ for large t, the integrals of the second and fourth terms on the right-hand side of (3-14) are entire functions in s. Hence we have

$$\begin{aligned}\text{(3-15)}\quad \frac{d}{du}\Big|_{s=0}f(s,u) &= -s\bigg(\sum_{k=0,1}(-1)^k\int_0^1 t^{s-1}\big(\mathrm{Tr}[\alpha_u Q_{p,\bar k}]-a_{n,p,\bar k,u}\big)\,dt\bigg)\bigg|_{s=0}\\ &= -\sum_{k=0,1}(-1)^k\mathrm{Tr}[\alpha_u Q_{p,\bar k}]+\sum_{k=0,1}(-1)^k a_{n,p,\bar k,u}.\end{aligned}$$

Hence, the result follows. □

By combining Lemma 3.6 with Lemma 3.7, we obtain the main theorem of this section. For the untwisted case, see [Cappell and Miller 2010, Theorem 4.4].

Theorem 3.8. *Let W be a complex manifold of complex dimension n and let E be a holomorphic bundle with connection D that is compatible and of type* $(1, 1)$ *over W. Suppose that $H \in A^{0,\bar{1}}(W, \mathbb{C})$ and $\bar{\partial} H = 0$. Let g_u^W, $u \in (u_0 - \delta, u_0 + \delta)$, be a smooth family of Riemannian metrics on the complex manifold W. Then the corresponding twisted Cappell–Miller holomorphic torsion $\tau_{\text{holo},p,u}(W, E, H)$ varies smoothly and the variation of $\tau_{\text{holo},p,u}(W, E, H)$ is given by a local formula*

$$\frac{d}{du}\tau_{\text{holo},p,u}(W, E, H) = \left(\sum_{k=0,1}(-1)^k \int_W b_{n,p,\bar{k},u}\right)\cdot \tau_{\text{holo},p,u}(W, E, H).$$

We have the following corollary. See also [Mathai and Wu 2010a, Theorem 5.3, Corollary 7.1] for the case of analytic torsion on $\mathbb{Z}_2$-graded elliptic complexes.

Corollary 3.9. *Let W be a complex manifold of complex dimension n and let E be a holomorphic bundle with connection D that is compatible and of type* $(1, 1)$ *over W. Suppose that $H \in A^{0,\bar{1}}(W, \mathbb{C})$ and $\bar{\partial} H = 0$. Let F_1, F_2 be two flat complex bundles over W of the same dimension. Then*

$$\tau_{\text{holo},p}(W, E \otimes F_1, H) \otimes \big(\tau_{\text{holo},p}(W, E \otimes F_2, H)]^{-1}\big)$$

in the tensor product of determinant lines

$$\Big(\operatorname{Det} H^{p,\bullet}_{\bar{\partial}_E}(W, E \otimes F_1, H) \otimes \big(\operatorname{Det} H^{\bullet,n-p}_{D^{1,0}}(W, E \otimes F_1, H)\big)^{(-1)^{n+1}}\Big)$$
$$\otimes \Big(\operatorname{Det} H^{p,\bullet}_{\bar{\partial}_E}(W, E \otimes F_2, H) \otimes \big(\operatorname{Det} H^{\bullet,n-p}_{D^{1,0}}(W, E \otimes F_2, H)\big)^{(-1)^{n+1}}\Big)^{-1}$$

is independent of the Hermitian metric g^W chosen.

This follows from the fact that the two bundles $E \otimes F_1$ and $E \otimes F_2$ are locally identical as bundles. For the untwisted case, see [Cappell and Miller 2010, Corollary 4.5].

Twisted Cappell–Miller holomorphic torsion under flux deformation. Suppose that the flux form H is deformed smoothly along a one-parameter family with parameter $v \in \mathbb{R}$ in such a way that the cohomology class $[H] \in H^{0,\bar{1}}(W, \mathbb{C})$ is fixed. Then $(d/dv)H = -\bar{\partial} B$ for some form $B \in A^{0,\bar{0}}(W, \mathbb{C})$ that depends smoothly on v. Let $\beta = B \wedge \cdot$. Fix $v_0 \in \mathbb{R}$ and choose $\lambda > 0$ such that there are no eigenvalues of $\square^H_{E,\bar{\partial},v_0}$ of absolute value λ. Then there exists $\delta > 0$ small enough that the same holds for the spectrum of $\square^H_{E,\bar{\partial},v}|_{A^{p,\bar{k}}_{(\lambda,\infty)}(W,E)}$ for $v \in (v_0 - \delta, v_0 + \delta)$. For simplicity, we omit the parameter v in the notation in the following discussion. Recall that $Q_{p,\bar{k}}$ is the spectral projection onto $A^{p,\bar{k}}_{[0,\lambda]}(W, E)$.

The proof of the following lemma is similar to the proof of [Mathai and Wu 2008, Lemma 3.7]; see also [Huang 2010, Lemma 4.7].

Lemma 3.10. *Under the assumptions above, we have*

$$\frac{d}{dv}\tau_{p,[0,\lambda]} = -\sum_{k=0,1}(-1)^k \operatorname{Tr}[\beta Q_{p,\bar{k}}]\cdot \tau_{p,[0,\lambda]},$$

upon identification of determinant lines under the deformation.

Lemma 3.11. *Under the assumptions above, we have*

$$\frac{d}{dv}\biggl(\sum_{k=0,1}(-1)^k \operatorname{LDet}_\theta(\bar\partial^{*,H}_{E,\nabla^{1,0}}\bar\partial^H_E)|_{A^{p,\bar{k}}_{+,(\lambda,\infty)}(W,E)}\biggr) = \sum_{k=0,1}(-1)^k \operatorname{Tr}[\beta Q_{p,\bar{k}}] + \sum_{k=0,1}(-1)^k\int_W c_{n,p,\bar{k}},$$

where $c_{n,p,\bar{k}}$ is given by a local formula.

Proof. Under the deformation, we have

$$\frac{d}{dv}\bar\partial^H_E = [\beta, \bar\partial^H_E], \qquad \frac{d}{dv}\bar\partial^{*,H}_{E,D^{1,0}} = -[\beta, \bar\partial^{*,H}_{E,D^{1,0}}].$$

Following the proof of [Mathai and Wu 2008, Lemma 3.5], we obtain the desired variation formula. □

By combining Lemma 3.10 with Lemma 3.11, we obtain the main theorem of this section.

Theorem 3.12. *Let W be a complex manifold of complex dimension n and let E be a holomorphic bundle with connection D that is compatible and of type* $(1, 1)$ *over W. Along any one parameter deformation of H that fixes the cohomology class $[H]$ and the natural identification of determinant lines, we have the variation formula*

$$\frac{d}{dv}\tau_{\mathrm{holo},p}(W, E, H) = \biggl(\sum_{k=0,1}(-1)^k\int_W c_{n,p,\bar{k}}\biggr)\cdot \tau_{\mathrm{holo},p}(W, E, H).$$

As with Corollary 3.9, we have the following corollary. See also [Mathai and Wu 2010a, Corollary 7.1] for the case of analytic torsion on $\mathbb{Z}_2$-graded elliptic complexes.

Corollary 3.13. *Let W be a complex manifold of complex dimension n and let E be a holomorphic bundle with connection D that is compatible and of type* $(1, 1)$ *over W. Suppose that $H \in A^{0,\bar{1}}(W, \mathbb{C})$ and $\bar\partial H = 0$. Let F_1, F_2 be two flat complex bundles over W of the same dimension. Then*

$$\tau_{\mathrm{holo},p}(W, E\otimes F_1, H)\otimes\bigl(\tau_{\mathrm{holo},p}(W, E\otimes F_2, H)\bigr)^{-1}$$

is invariant under any deformation of H by an $\bar\partial$-exact form, up to natural identification of the determinant lines.

4. Twisted Cappell–Miller analytic torsion

In this section we first define the de Rham bigraded complex twisted by a flux form H and its (co)homology groups. Then we define the Cappell–Miller analytic torsion for the twisted de Rham bigraded complex. We obtain the variation theorems of the twisted Cappell–Miller analytic torsion under metric and flux deformations. Su, in a recent preprint [2011], also briefly discussed the twisted Cappell–Miller analytic torsion when dimension of the manifold M is odd.

The twisted de Rham complexes. Suppose M is a closed oriented m-dimensional smooth manifold and let $\mathcal{E}$ be a complex vector bundle over M endowed with a flat connection ∇. We denote by $\Omega^p(M,\mathcal{E})$ the space of p-forms with values in the flat bundle $\mathcal{E}$, that is, $\Omega^p(M,\mathcal{E}) = \Gamma(M, \bigwedge^p(T^*M)_{\mathbb{R}} \otimes \mathcal{E})$ and by

$$\nabla : \Omega^\bullet(M,\mathcal{E}) \to \Omega^{\bullet+1}(M,\mathcal{E})$$

the covariant differential induced by the flat connection on $\mathcal{E}$. Fix a Riemannian metric g^M on M and let $\star : \Omega^\bullet(M,\mathcal{E}) \to \Omega^{m-\bullet}(M,\mathcal{E})$ denote the Hodge $\star$ operator. We choose a Hermitian metric $h^{\mathcal{E}}$ so that together with the Riemannian metric g^M we can define a scalar product $\langle \cdot\,,\cdot \rangle_M$ on $\Omega^\bullet(M,\mathcal{E})$. Define the chirality operator $\Gamma = \Gamma(g^M) : \Omega^\bullet(M,\mathcal{E}) \to \Omega^\bullet(M,\mathcal{E})$ by [Braverman and Kappeler 2007, (7-1)]

$$\Gamma\omega := i^r(-1)^{q(q+1)/2} \star \omega, \quad \omega \in \Omega^q(M,\mathcal{E}), \tag{4-1}$$

where $r = (m+1)/2$ if m is odd and $r = m/2$ if m is even. The numerical factor in (4-1) has been chosen so that $\Gamma^2 = \mathrm{Id}$; see [Berline et al. 2004, Proposition 3.58].

Assume $\mathcal{H}$ is an odd degree closed differential form on M. Let $\Omega^{\bar{0}/\bar{1}}(M,\mathcal{E}) := \Omega^{\text{even}/\text{odd}}(M,\mathcal{E})$ and $\nabla^{\mathcal{H}} := \nabla + \mathcal{H} \wedge \cdot\,$. Assume that $\mathcal{H}$ does not contain a 1-form component, which can be absorbed in the flat connection ∇.

It is not difficult to check that $(\nabla^{\mathcal{H}})^2 = 0$. Clearly, for each $k = 0, 1$, we have $\nabla^{\mathcal{H}} : \Omega^{\bar{k}}(M,\mathcal{E}) \to \Omega^{\overline{k+1}}(M,\mathcal{E})$. Hence we can consider the twisted de Rham complex

$$(\Omega^\bullet(M,\mathcal{E}), \nabla^{\mathcal{H}}) : \quad \cdots \xrightarrow{\nabla^{\mathcal{H}}} \Omega^{\bar{0}}(M,\mathcal{E}) \xrightarrow{\nabla^{\mathcal{H}}} \Omega^{\bar{1}}(M,\mathcal{E}) \xrightarrow{\nabla^{\mathcal{H}}} \Omega^{\bar{0}}(M,\mathcal{E}) \xrightarrow{\nabla^{\mathcal{H}}} \cdots . \tag{4-2}$$

We define the *twisted de Rham cohomology group of* $(\Omega^\bullet(M,\mathcal{E}), \nabla^{\mathcal{H}})$ as

$$H^{\bar{k}}(M,\mathcal{E},\mathcal{H}) := \frac{\mathrm{Ker}(\nabla^{\mathcal{H}} : \Omega^{\bar{k}}(M,\mathcal{E}) \to \Omega^{\overline{k+1}}(M,\mathcal{E}))}{\mathrm{Im}(\nabla^{\mathcal{H}} : \Omega^{\overline{k-1}}(M,\mathcal{E}) \to \Omega^{\bar{k}}(M,\mathcal{E}))}, \quad k = 0, 1.$$

The groups $H^{\bar{k}}(M,\mathcal{E},\mathcal{H})$, $k = 0, 1$, are independent of the choice of the Riemannian metric on M or the Hermitian metric on $\mathcal{E}$. Replacing $\mathcal{H}$ by $\mathcal{H}' = \mathcal{H} - d\mathcal{B}$ for some $\mathcal{B} \in \Omega^{\bar{0}}(M)$ gives an isomorphism $\varepsilon_{\mathcal{B}} := e^{\mathcal{B}} \wedge \cdot\, : \Omega^\bullet(M,\mathcal{E}) \to \Omega^\bullet(M,\mathcal{E})$

satisfying

$$\varepsilon_{\mathcal{B}} \circ \nabla^{\mathcal{H}} = \nabla^{\mathcal{H}'} \circ \varepsilon_{\mathcal{B}}.$$

Therefore $\varepsilon_{\mathcal{B}}$ induces an isomorphism on the twisted de Rham cohomology. Also denote by $\varepsilon_{\mathcal{B}}$ the map

$$\varepsilon_{\mathcal{B}} : H^{\bullet}(M, \mathcal{E}, \mathcal{H}) \to H^{\bullet}(M, \mathcal{E}, \mathcal{H}'). \tag{4-3}$$

Denote by $\nabla^{\mathcal{H},*}$ the adjoint of $\nabla^{\mathcal{H}}$ with respect to the scalar product $\langle \cdot, \cdot \rangle_M$. Then the Laplacian

$$\Delta^{\mathcal{H}} := \nabla^{\mathcal{H},*}\nabla^{\mathcal{H}} + \nabla^{\mathcal{H}}\nabla^{\mathcal{H},*}$$

is an elliptic operator and therefore the complex (4-2) is elliptic. By Hodge theory, we have the isomorphism $\operatorname{Ker} \Delta^{\mathcal{H}} \cong H^{\bullet}(M, \mathcal{E}, \mathcal{H})$. For more details of the twisted de Rham cohomology, see, for example, [Mathai and Wu 2008].

Now denote by ∇' the connection on $\mathcal{E}$ dual to the connection ∇ with respect to the Hermitian metric $h^{\mathcal{E}}$ [Braverman and Kappeler 2007, Section 10.1]. Denote by $\mathcal{E}'$ the flat bundle $(\mathcal{E}, \nabla')$, referring to $\mathcal{E}'$ as the dual of the flat vector bundle $\mathcal{E}$. We emphasize that, similar to the untwisted case [Braverman and Kappeler 2007, (10-8); Cappell and Miller 2010, (8.4)],

$$\nabla^{\mathcal{H},*} = \Gamma \nabla'^{\mathcal{H}} \Gamma,$$

where $\nabla'^{\mathcal{H}} = \nabla' + \mathcal{H} \wedge \cdot$.

Let $\nabla^{\mathcal{H},\sharp} := \Gamma \nabla^{\mathcal{H}} \Gamma$. Then $(\nabla^{\mathcal{H},\sharp})^2 = 0$. Clearly, $\nabla^{\mathcal{H},\sharp} : \Omega^{\bar{k}}(M, \mathcal{E}) \to \Omega^{\overline{k-1}}(M, \mathcal{E})$. Hence we can consider the twisted de Rham complex

$$(\Omega^{\bullet}(M, \mathcal{E}), \nabla^{\mathcal{H},\sharp}) : \quad \cdots \xleftarrow{\nabla^{\mathcal{H},\sharp}} \Omega^{\bar{0}}(M, \mathcal{E}) \xleftarrow{\nabla^{\mathcal{H},\sharp}} \Omega^{\bar{1}}(M, \mathcal{E}) \xleftarrow{\nabla^{\mathcal{H},\sharp}} \Omega^{\bar{0}}(M, \mathcal{E}) \xleftarrow{\nabla^{\mathcal{H},\sharp}} \cdots . \tag{4-4}$$

We also define the homology group of the complex $(\Omega^{\bullet}(M, \mathcal{E}), \nabla^{\mathcal{H},\sharp})$ as

$$H_{\bar{k}}(\Omega^{\bullet}(M, \mathcal{E}), \nabla^{\mathcal{H},\sharp}) := \frac{\operatorname{Ker}(\nabla^{\mathcal{H},\sharp} : \Omega^{\bar{k}}(M, \mathcal{E}) \to \Omega^{\overline{k-1}}(M, \mathcal{E}))}{\operatorname{Im}(\nabla^{\mathcal{H},\sharp} : \Omega^{\overline{k+1}}(M, \mathcal{E}) \to \Omega^{\bar{k}}(M, \mathcal{E}))}, \quad k = 0, 1.$$

Similarly, the groups $H_{\bar{k}}(\Omega^{\bullet}(M, \mathcal{E}), \nabla^{\mathcal{H},\sharp})$, $k = 0, 1$, are independent of the choice of the Riemannian metric on M or the Hermitian metric on $\mathcal{E}$. Suppose that $\mathcal{H}$ is replaced by $\mathcal{H}'' = \mathcal{H} - \delta\mathcal{B}'$ for some $\mathcal{B}' \in \Omega^{\bar{0}}(M)$ and δ the adjoint of d with respect to the scalar product induced by the Riemannian metric g^M. Then there is an isomorphism $\varepsilon_{\mathcal{B}'} := e^{\mathcal{B}'} \wedge \cdot : \Omega^{\bullet}(M, \mathcal{E}) \to \Omega^{\bullet}(M, \mathcal{E})$ satisfying

$$\varepsilon_{\mathcal{B}'} \circ \nabla^{\mathcal{H},\sharp} = \nabla^{\mathcal{H}'',\sharp} \circ \varepsilon_{\mathcal{B}'}.$$

Therefore $\varepsilon_{\mathcal{B}'}$ induces an isomorphism on the twisted de Rham homology. Also denote by $\varepsilon_{\mathcal{B}'}$ the map

$$\varepsilon_{\mathcal{B}'} : H_\bullet(\Omega^\bullet(M,\mathcal{E}), \nabla^{\mathcal{H},\sharp}) \to H_\bullet(\Omega^\bullet(M,\mathcal{E}), \nabla^{\mathcal{H}'',\sharp}). \tag{4-5}$$

Denote by $\nabla^{\mathcal{H},\sharp,*}$ the adjoint of $\nabla^{\mathcal{H},\sharp}$ with respect to the scalar product $\langle\,\cdot\,,\cdot\,\rangle_M$. Then we have the equalities

$$\nabla^{\mathcal{H},\sharp,*} = \nabla'^{\mathcal{H}}, \quad \Delta'^{\mathcal{H}} := \nabla'^{\mathcal{H},*}\nabla'^{\mathcal{H}} + \nabla'^{\mathcal{H}}\nabla'^{\mathcal{H},*} = \nabla^{\mathcal{H},\sharp}\nabla^{\mathcal{H},\sharp,*} + \nabla^{\mathcal{H},\sharp,*}\nabla^{\mathcal{H},\sharp}.$$

Again the Laplacian $\Delta'^{\mathcal{H}}$ is an elliptic operator and thus the complex (4-4) is elliptic. By Hodge theory, we have the isomorphism $\operatorname{Ker}\Delta'^{\mathcal{H}} \cong H_\bullet(\Omega^\bullet(M,\mathcal{E}), \nabla^{\mathcal{H},\sharp})$. In particular, for $k = 0, 1$,

$$H_{\bar{k}}(\Omega^\bullet(M,\mathcal{E}), \nabla^{\mathcal{H},\sharp}) \cong H^{\bar{k}}(M,\mathcal{E}',\mathcal{H}). \tag{4-6}$$

Definition of twisted Cappell–Miller analytic torsion. The flat Laplacian

$$\Delta^{\mathcal{H},\sharp} := (\nabla^{\mathcal{H}} + \nabla^{\mathcal{H},\sharp})^2$$

maps $\Omega^{\bar{k}}(M,\mathcal{E})$ into itself. Suppose $\mathcal{I}$ is an interval of the form $[0,\lambda]$, $(\lambda,\mu]$, or (λ,∞) $(\mu > \lambda \geq 0)$. Denote by $\Pi^{\sharp,\mathcal{I}}$ the spectral projection of $\Delta^{\mathcal{H},\sharp}$ corresponding to the set of generalized eigenvalues, whose absolute values lie in $\mathcal{I}$. Set

$$\Omega^{\bar{k}}_{\mathcal{I}}(M,\mathcal{E}) := \Pi^{\sharp,\mathcal{I}}(\Omega^{\bar{k}}(M,\mathcal{E})) \subset \Omega^{\bar{k}}(M,\mathcal{E}).$$

If the interval $\mathcal{I}$ is bounded, then the space $\Omega^{\bar{k}}_{\mathcal{I}}(M,\mathcal{E})$ is finite dimensional. Since $\nabla^{\mathcal{H}}$ and $\nabla^{\mathcal{H},\sharp}$ commute with $\Delta^{\sharp}_{\mathcal{H}}$, the subspace $\Omega^\bullet_{\mathcal{I}}(M,\mathcal{E})$ is a subcomplex of the twisted de Rham bi-complex $(\Omega^\bullet(M,\mathcal{E}), \nabla^{\mathcal{H}}, \nabla^{\mathcal{H},\sharp})$. Clearly, for each $\lambda \geq 0$, the complex $\Omega^\bullet_{(\lambda,\infty)}(M,\mathcal{E})$ is doubly acyclic, that is, $H^{\bar{k}}(\Omega^\bullet_{(\lambda,\infty)}(M,\mathcal{E}), \nabla^{\mathcal{H}}) = 0$ and $H_{\bar{k}}(\Omega^\bullet_{(\lambda,\infty)}(M,\mathcal{E}), \nabla^{\mathcal{H},\sharp}) = 0$. Since

$$\Omega^{\bar{k}}(M,\mathcal{E}) = \Omega^{\bar{k}}_{[0,\lambda]}(M,\mathcal{E}) \oplus \Omega^{\bar{k}}_{(\lambda,\infty)}(M,\mathcal{E}), \tag{4-7}$$

the homology $H_{\bar{k}}(\Omega^\bullet_{[0,\lambda]}(M,\mathcal{E}), \nabla^{\mathcal{H},\sharp})$ of the complex $(\Omega^\bullet_{[0,\lambda]}(M,\mathcal{E}), \nabla^{\mathcal{H},\sharp})$ is naturally isomorphic to the homology $H_{\bar{k}}(\Omega^\bullet(M,\mathcal{E}), \nabla^{\mathcal{H},\sharp}) \cong H^{\bar{k}}(M,\mathcal{E}',\mathcal{H})$ (see (4-6)), and the cohomology $H^{\bar{k}}(\Omega^\bullet_{[0,\lambda]}(M,\mathcal{E}), \nabla^{\mathcal{H}})$ of $(\Omega^\bullet_{[0,\lambda]}(M,\mathcal{E}), \nabla^{\mathcal{H}})$ is naturally isomorphic to the cohomology $H^{\bar{k}}(M,\mathcal{E},\mathcal{H})$.

Similar to the $\mathbb{Z}$-graded case [Cappell and Miller 2010, Section 8], the chirality operator Γ establishes a complex linear isomorphism of the homology groups with cohomology groups

$$H_{\bar{k}}(\Omega^\bullet_{[0,\lambda]}(M,\mathcal{E}), \nabla^{\mathcal{H},\sharp}) \cong H^{\overline{m-k}}(\Omega^\bullet_{[0,\lambda]}(M,\mathcal{E}), \nabla^{\mathcal{H}}) \cong H^{\overline{m-k}}(M,\mathcal{E},\mathcal{H}).$$

In particular, we have the isomorphism

$$\begin{aligned}\operatorname{Det} H_\bullet(\Omega^\bullet(M,\mathscr{E}),\nabla^{\mathscr{H},\sharp}) &\cong \operatorname{Det} H_\bullet(\Omega^\bullet_{[0,\lambda]}(M,\mathscr{E}),\nabla^{\mathscr{H},\sharp})\\ &\cong (\operatorname{Det} H^\bullet(M,\mathscr{E},\mathscr{H}))^{(-1)^m}.\end{aligned}\tag{4-8}$$

Using Poincaré duality, we also have the isomorphism

$$\operatorname{Det} H^{\overline{m-k}}(M,\mathscr{E},\mathscr{H})^{-1} \cong \operatorname{Det} H^{\bar{k}}(M,\mathscr{E}',\mathscr{H}),\tag{4-9}$$

where $\mathscr{E}'$ is the dual vector bundle of the vector bundle E. Therefore, we have

$$\begin{aligned}&\operatorname{Det} H^\bullet(M,\mathscr{E},\mathscr{H}) \otimes \operatorname{Det} H^{\overline{m-\bullet}}(M,\mathscr{E},\mathscr{H})^{-1}\\ &\cong \operatorname{Det} H^\bullet(M,\mathscr{E},\mathscr{H}) \otimes \operatorname{Det} H^\bullet(M,\mathscr{E}',\mathscr{H}) \quad \text{by (4-9)}\\ &\cong \operatorname{Det} H^\bullet(M,\mathscr{E}\oplus\mathscr{E}',\mathscr{H}).\end{aligned}\tag{4-10}$$

For $k=0,1$, set

$$\begin{aligned}\Omega^{\bar{k}}_{+,\mathscr{I}}(M,\mathscr{E}) &:= \operatorname{Ker}(\nabla^{\mathscr{H}}\nabla^{\mathscr{H},\sharp}) \cap \Omega^{\bar{k}}_{\mathscr{I}}(M,\mathscr{E}),\\ \Omega^{\bar{k}}_{-,\mathscr{I}}(M,\mathscr{E}) &:= \operatorname{Ker}(\nabla^{\mathscr{H},\sharp}\nabla^{\mathscr{H}}) \cap \Omega^{\bar{k}}_{\mathscr{I}}(M,\mathscr{E}).\end{aligned}\tag{4-11}$$

Clearly,

$$\Omega^{\bar{k}}_{\mathscr{I}}(M,\mathscr{E}) = \Omega^{\bar{k}}_{+,\mathscr{I}}(M,\mathscr{E}) \oplus \Omega^{\bar{k}}_{-,\mathscr{I}}(M,\mathscr{E}) \quad \text{if } 0 \notin \mathscr{I}.$$

Let $\theta \in (0,2\pi)$ be an Agmon angle; see [Shubin 2001]. Since the leading symbol of $\nabla^{\mathscr{H},\sharp}\nabla^{\mathscr{H}}$ is positive definite, the ζ-regularized determinant $\operatorname{Det}_\theta(\nabla^{\mathscr{H},\sharp}\nabla^{\mathscr{H}})|_{\Omega^{\bar{k}}_{+,\mathscr{I}}(M,\mathscr{E})}$ is independent of the choice of θ.

For any $0 \le \lambda \le \mu \le \infty$, one easily sees that

$$\begin{aligned}\prod_{k=0,1}&\bigl(\operatorname{Det}_\theta(\nabla^{\mathscr{H},\sharp}\nabla^{\mathscr{H}})|_{\Omega^{\bar{k}}_{+,(\lambda,\infty)}(M,\mathscr{E})}\bigr)^{(-1)^k}\\ &= \biggl(\prod_{k=0,1}\bigl(\operatorname{Det}_\theta(\nabla^{\mathscr{H},\sharp}\nabla^{\mathscr{H}})|_{\Omega^{\bar{k}}_{+,(\lambda,\mu)}(M,\mathscr{E})}\bigr)^{(-1)^k}\biggr)\\ &\qquad\cdot\biggl(\prod_{k=0,1}\bigl(\operatorname{Det}_\theta(\nabla^{\mathscr{H},\sharp}\nabla^{\mathscr{H}})|_{\Omega^{\bar{k}}_{+,(\mu,\infty)}(M,\mathscr{E})}\bigr)^{(-1)^k}\biggr).\end{aligned}$$

For any $\lambda \ge 0$, denote by $\tau_{[0,\lambda]}$ the Cappell–Miller torsion of the twisted de Rham bigraded complex $\bigl(\Omega^\bullet_{[0,\lambda]}(M,\mathscr{E}),\nabla^{\mathscr{H}},\nabla^{\mathscr{H},\sharp}\bigr)$. Via the isomorphisms

$$\begin{aligned}H_\bullet(\Omega^\bullet_{[0,\lambda]}(M,\mathscr{E}),\nabla^{\mathscr{H},\sharp}) &\cong H_\bullet(\Omega^\bullet(M,\mathscr{E}),\nabla^{\mathscr{H},\sharp}),\\ H^\bullet(\Omega^\bullet_{[0,\lambda]}(M,\mathscr{E}),\nabla^{\mathscr{H}}) &\cong H^\bullet(M,\mathscr{E},\mathscr{H}),\end{aligned}$$

and (4-10), we can view $\tau_{[0,\lambda]}$ as an element of $\operatorname{Det} H^\bullet(M,\mathscr{E}\oplus\mathscr{E}',\mathscr{H})$. In particular, if m is odd, then, up to an isomorphism,

$$\tau_{[0,\lambda]} \in \operatorname{Det} H^\bullet(M,\mathscr{E},\mathscr{H}) \otimes \operatorname{Det} H^\bullet(M,\mathscr{E},\mathscr{H}) \cong \operatorname{Det} H^\bullet(M,\mathscr{E}\oplus\mathscr{E}',\mathscr{H}).\tag{4-12}$$

The proof of the following lemma is similar to the proof of [Cappell and Miller 2010, Theorem 8.3].

Lemma 4.1. *The element*

$$\tau_{[0,\lambda]} \cdot \prod_{k=0,1} \left(\mathrm{Det}_\theta(\nabla^{\mathcal{H},\sharp}\nabla^{\mathcal{H}})|_{\Omega^{\bar{k}}_{+,(\lambda,\infty)}(M,\mathcal{E})}\right)^{(-k)}$$

is independent of the choice of λ.

We now define the Cappell–Miller analytic torsion for the de Rham complex twisted by a flux.

Definition 4.2. Let $(\mathcal{E}, \nabla)$ be a complex vector bundle over a connected oriented m-dimensional closed Riemannian manifold M and $\mathcal{H}$ be a closed odd degree form (not a 1-form). Further, let

$$\nabla^{\mathcal{H}} = \nabla + \mathcal{H} \wedge \cdot \quad \text{and} \quad \nabla^{\mathcal{H},\sharp} = \Gamma\nabla^{\mathcal{H}}\Gamma.$$

Let $\theta \in (0, 2\pi)$ be an Agmon angle for the operator $\Delta^{\mathcal{H},\sharp} := (\nabla^{\mathcal{H}} + \nabla^{\mathcal{H},\sharp})^2$. The Cappell–Miller torsion $\tau(\nabla, \mathcal{H})$ for the twisted de Rham bigraded complex $(\Omega^\bullet(M,\mathcal{E}), \nabla^{\mathcal{H}}, \nabla^{\mathcal{H},\sharp})$ is an element of $\mathrm{Det}\, H^\bullet(M,\mathcal{E},\mathcal{H}) \otimes \left(\mathrm{Det}\, H^\bullet(M,\mathcal{E},\mathcal{H})\right)^{(-1)^{m+1}}$ defined as

$$\tau(\nabla, \mathcal{H}) := \tau_{[0,\lambda]} \cdot \prod_{k=0,1} \left(\mathrm{Det}_\theta(\nabla^{\mathcal{H},\sharp}\nabla^{\mathcal{H}})|_{\Omega^{\bar{k}}_{+,(\lambda,\infty)}(M,\mathcal{E})}\right)^{(-k)}. \tag{4-13}$$

Twisted Cappell–Miller analytic torsion under metric and flux deformations. In this section we obtain the variation formulas for the twisted Cappell–Miller analytic torsion $\tau(\nabla, \mathcal{H})$ under the metric and flux deformations. In particular, we show that if the manifold M is an odd-dimensional closed oriented manifold, then the twisted Cappell–Miller analytic torsion is independent of the Riemannian metric and the representative $\mathcal{H}$ in the cohomology class $[\mathcal{H}]$. See also [Su 2011].

The proof of the following theorem is similar to the proof of Theorem 3.8.

Theorem 4.3. *Let* $(\mathcal{E}, \nabla)$ *be a complex vector bundle over a* m*-dimensional connected oriented closed Riemannian manifold* M *and* $\mathcal{H}$ *be a closed odd degree form* (*not a* 1*-form*)*. Let* g^M_v, $a \le v \le b$, *be a smooth family of Riemannian metrics on* M. *Then the corresponding twisted Cappell–Miller analytic torsion* $\tau_v(\nabla, \mathcal{H})$ *varies smoothly and the variation of* $\tau_v(\nabla, \mathcal{H})$ *is given by a local formula*

$$\frac{d}{dv}\tau_v(\nabla, \mathcal{H}) = \left(\sum_{k=0,1} (-1)^k \int_M b_{m/2,\bar{k},v}\right) \cdot \tau_v(\nabla, \mathcal{H}).$$

In particular, *if the dimension of the manifold* M *is odd*, *then twisted Cappell–Miller analytic torsion* $\tau(\nabla, \mathcal{H})$ *is independent of the Riemannian metric* g^M.

For the untwisted case considered in [Bismut and Zhang 1992], the variation of the torsion can be integrated to an anomaly formula.

The proof of the following is similar to that of [Mathai and Wu 2010a, Theorem 6.1]. See also [Mathai and Wu 2008, Theorem 3.8].

Theorem 4.4. *Let* $(\mathscr{E}, \nabla)$ *be a complex vector bundle over a* m*-dimensional connected oriented closed Riemannian manifold* M *and* $\mathscr{H}$ *be a closed odd degree form (not a* 1*-form). Under the natural identification of determinant lines and along any one parameter deformation* $\mathscr{H}_v$ *of* $\mathscr{H}$ *that fixes the cohomology class* $[\mathscr{H}]$, *we have the variation formula*

$$\frac{d}{dv}\tau(\nabla, \mathscr{H}_v) = \left(\sum_{k=0,1}(-1)^k\int_M c_{m/2,\bar{k},v}\right)\cdot\tau(\nabla, \mathscr{H}_v).$$

In particular, if the dimension of the manifold M *is odd, then, under the natural identification of determinant lines, the twisted Cappell–Miller analytic torsion* $\tau(\nabla, \mathscr{H})$ *is independent of any deformation of* $\mathscr{H}$ *that fixes the cohomology class* $[\mathscr{H}]$.

Relationship with the twisted refined analytic torsion. In this section we assume that M is a closed compact oriented manifold of odd dimension. Recall that in [Huang 2010, (3.13)], for each $\lambda > 0$, we define the twisted refined torsion $\rho_{\Gamma_{[0,\lambda]}}$ of the twisted finite-dimensional complex $(\Omega^{\bullet}_{[0,\lambda]}(M, \mathscr{E}), \nabla^{\mathscr{H}})$ corresponding to the chirality operator $\Gamma_{[0,\lambda]}$. In our setting, as in the $\mathbb{Z}$-graded case [Braverman and Kappeler 2008a, (5.1)], the twisted Cappell–Miller torsion can be described as (see (4-12))

$$\tau_{[0,\lambda]} := \rho_{\Gamma_{[0,\lambda]}} \otimes \rho_{\Gamma_{[0,\lambda]}} \in \operatorname{Det} H^{\bullet}(M, \mathscr{E}, \mathscr{H}) \otimes \operatorname{Det} H^{\bullet}(M, \mathscr{E}, \mathscr{H}). \tag{4-14}$$

By combining (3.14), (3.20), (5.28) and Definition 4.5 of [Huang 2010], the twisted refined analytic torsion can be written as

$$\begin{aligned}\rho_{\mathrm{an}}(\nabla^{\mathscr{H}}) = \pm\rho_{\Gamma_{[0,\lambda]}}\cdot\prod_{k=0,1}\big(\operatorname{Det}_{\theta}(\nabla^{\mathscr{H},\sharp}\nabla^{\mathscr{H}})|_{\Omega^{\bar{k}}_{+,(\lambda,\infty)}(M,\mathscr{E})}\big)^{-k/2}&\\ \cdot\exp\big(-i\pi(\eta(\mathscr{B}_0^{\mathscr{H}}(\nabla^{\mathscr{H}})) - \operatorname{rank}\mathscr{E}\cdot\eta_{\mathrm{trivial}})\big),&\end{aligned} \tag{4-15}$$

where $\eta(\mathscr{B}_0^{\mathscr{H}}(\nabla^{\mathscr{H}})) - \operatorname{rank}\mathscr{E}\cdot\eta_{\mathrm{trivial}}$ is the ρ-invariant of the twisted odd signature operator $\mathscr{B}_0^{\mathscr{H}}(\nabla^{\mathscr{H}})$ defined in [Huang 2010, (3.2)].

By combining (4-13), (4-14) with (4-15), we have the following comparison theorem of the twisted Cappell–Miller analytic torsion and twisted refined analytic torsion.

Theorem 4.5. *Let* $(\mathscr{E}, \nabla)$ *be a complex vector bundle over a connected oriented odd-dimensional closed Riemannian manifold* M *and* $\mathscr{H}$ *be a closed odd degree* *form* (*not a* 1*-form*). *Further*, *let* $\nabla^{\mathscr{H}} = \nabla + \mathscr{H} \wedge \cdot$. *Then*

$$\tau(\nabla, \mathscr{H}) \cdot \exp\bigl(-2i\pi(\eta(\mathscr{B}_0^{\mathscr{H}}(\nabla^{\mathscr{H}})) - \operatorname{rank}\mathscr{E} \cdot \eta_{\text{trivial}})\bigr) = \rho_{\text{an}}(\nabla^{\mathscr{H}}) \otimes \rho_{\text{an}}(\nabla^{\mathscr{H}}).$$

Su in [2011, Theorem 5.1] compared the twisted Burghelea–Haller analytic torsion which he introduced with the twisted refined analytic torsion. By combining [Su 2011, Theorem 5.1] with Theorem 4.5, we can also obtain the comparison theorem of the twisted Burghelea–Haller torsion and the twisted Cappell–Miller analytic torsion. We skip the details.

References

[Berline et al. 2004] N. Berline, E. Getzler, and M. Vergne, *Heat kernels and Dirac operators*, Springer, Berlin, 2004. MR 2007m:58033 Zbl 1037.58015

[Bismut 1993] J.-M. Bismut, "From Quillen metrics to Reidemeister metrics: some aspects of the Ray–Singer analytic torsion", pp. 273–324 in *Topological methods in modern mathematics* (Stony Brook, NY, 1991), edited by L. R. Goldberg and A. V. Phillips, Publish or Perish, Houston, 1993. MR 94e:58144 Zbl 0799.58076

[Bismut and Lebeau 1989] J.-M. Bismut and G. Lebeau, "Immersions complexes et métriques de Quillen", *C. R. Acad. Sci. Paris Sér. I Math.* **309** (1989), 487–491. MR 91k:58138 Zbl 0681.53034

[Bismut and Lebeau 1991] J.-M. Bismut and G. Lebeau, "Complex immersions and Quillen metrics", *Inst. Hautes Études Sci. Publ. Math.* **74** (1991), 1–298. MR 94a:58205 Zbl 0784.32010

[Bismut and Zhang 1992] J.-M. Bismut and W. Zhang, *An extension of a theorem by Cheeger and Müller*, Astérisque **205**, Société Math. de France, Paris, 1992. MR 93j:58138 Zbl 0781.58039

[Bismut et al. 1988a] J.-M. Bismut, H. Gillet, and C. Soulé, "Analytic torsion and holomorphic determinant bundles, I, Bott–Chern forms and analytic torsion", *Comm. Math. Phys.* **115**:1 (1988), 49–78. MR 89g:58192a Zbl 0651.32017

[Bismut et al. 1988b] J.-M. Bismut, H. Gillet, and C. Soulé, "Analytic torsion and holomorphic determinant bundles, II, Direct images and Bott–Chern forms", *Comm. Math. Phys.* **115**:1 (1988), 79–126. MR 89g:58192b Zbl 0651.32017

[Bismut et al. 1988c] J.-M. Bismut, H. Gillet, and C. Soulé, "Analytic torsion and holomorphic determinant bundles, III, Quillen metrics on holomorphic determinants", *Comm. Math. Phys.* **115**:2 (1988), 301–351. MR 89g:58192c Zbl 0651.32017

[Bismut et al. 1990] J.-M. Bismut, H. Gillet, and C. Soulé, "Complex immersions and Arakelov geometry", pp. 249–331 in *The Grothendieck Festschrift, Vol. I*, edited by P. Cartier et al., Progr. Math. **86**, Birkhäuser, Boston, 1990. MR 92a:14019 Zbl 0744.14015

[Braverman and Kappeler 2007] M. Braverman and T. Kappeler, "Refined analytic torsion as an element of the determinant line", *Geom. Topol.* **11** (2007), 139–213. MR 2008a:58031 Zbl 1135.58014

[Braverman and Kappeler 2008a] M. Braverman and T. Kappeler, "A canonical quadratic form on the determinant line of a flat vector bundle", *Int. Math. Res. Not.* **2008**:11 (2008), Art. ID rnn030. MR 2009j:58048

[Braverman and Kappeler 2008b] M. Braverman and T. Kappeler, "Refined analytic torsion", *J. Differential Geom.* **78**:2 (2008), 193–267. MR 2009a:58041 Zbl 1147.58034

[Burghelea and Haller 2007] D. Burghelea and S. Haller, "Complex-valued Ray–Singer torsion", *J. Funct. Anal.* **248**:1 (2007), 27–78. MR 2008b:58035 Zbl 1131.58020

[Burghelea and Haller 2008] D. Burghelea and S. Haller, "Torsion, as a function on the space of representations", pp. 41–66 in *C^*-algebras and elliptic theory II*, edited by D. Burghelea et al., Birkhäuser, Basel, 2008. MR 2009i:58045 Zbl 1147.58035

[Burghelea and Haller 2010] D. Burghelea and S. Haller, "Complex valued Ray–Singer torsion II", *Math. Nachr.* **283**:10 (2010), 1372–1402. MR 2744135 Zbl 1208.58029

[Cappell and Miller 2010] S. E. Cappell and E. Y. Miller, "Complex-valued analytic torsion for flat bundles and for holomorphic bundles with (1, 1) connections", *Comm. Pure Appl. Math.* **63**:2 (2010), 133–202. MR 2011a:58061 Zbl 1201.58021

[Huang 2010] R.-T. Huang, "Refined analytic torsion for twisted de Rham complexes", preprint, 2010. arXiv 1001.0654v1

[Liu and Yu 2010] B. Liu and J. Yu, "On the anomaly formula for the Cappell–Miller holomorphic torsion", *Sci. China Math.* **53**:12 (2010), 3225–3241. MR 2746319 Zbl 05857741

[Mathai and Wu 2008] V. Mathai and S. Wu, "Analytic torsion for twisted de Rham complexes", preprint, 2008. arXiv 0810.4204v3

[Mathai and Wu 2010a] V. Mathai and S. Wu, "Analytic torsion for $\mathbb{Z}_2$-graded elliptic complexes", preprint, 2010. arXiv 1001.3212v1

[Mathai and Wu 2010b] V. Mathai and S. Wu, "Twisted analytic torsion", *Sci. China Math.* **53**:3 (2010), 555–563. MR 2608312 Zbl 1202.58019

[Ray and Singer 1971] D. B. Ray and I. M. Singer, "R-torsion and the Laplacian on Riemannian manifolds", *Advances in Math.* **7** (1971), 145–210. MR 45 #4447 Zbl 0239.58014

[Ray and Singer 1973] D. B. Ray and I. M. Singer, "Analytic torsion for complex manifolds", *Ann. of Math.* (2) **98** (1973), 154–177. MR 52 #4344 Zbl 0267.32014

[Seeley 1967] R. T. Seeley, "Complex powers of an elliptic operator", pp. 288–307 in *Singular Integrals* (Chicago, Ill., 1966), edited by A. P. Calderón, Proc. Sympos. Pure Math. **10**, Amer. Math. Soc., Providence, R.I., 1967. MR 38 #6220 Zbl 0159.15504

[Shubin 2001] M. A. Shubin, *Pseudodifferential operators and spectral theory*, 2nd ed., Springer, Berlin, 2001. MR 2002d:47073 Zbl 0980.35180

[Su 2011] G. Su, "Burghelea–Haller analytic torsion for twisted de Rham complexes", *Pac. J. Math.* **250**:2 (2011), 421–437. Zbl 05879446

[Wodzicki 1984] M. Wodzicki, "Local invariants of spectral asymmetry", *Invent. Math.* **75**:1 (1984), 143–177. MR 85g:58089 Zbl 0538.58038

Received April 1, 2010.

RUNG-TZUNG HUANG
INSTITUTE OF MATHEMATICS, ACADEMIA SINICA
6TH FLOOR, ASTRONOMY-MATHEMATICS BUILDING
NO. 1 SEC. 4, ROOSEVELT ROAD
TAIPEI 106-17
TAIWAN

rthuang@math.sinica.edu.tw

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 1, 2011

GENERALIZATIONS OF AGOL'S INEQUALITY AND NONEXISTENCE OF TIGHT LAMINATIONS

THILO KUESSNER

We give a general lower bound for the normal Gromov norm of genuine laminations in terms of the topology of the complementary regions.

In the special case of 3-manifolds, this yields a generalization of Agol's inequality from incompressible surfaces to tight laminations. In particular, the inequality excludes the existence of tight laminations with nonempty guts on 3-manifolds of small simplicial volume.

1. Results

Agol's inequality [1999, Theorem 2.1] is the following (see Section 7 for notation):

Let M be a hyperbolic 3-manifold containing an incompressible, properly embedded surface F. Then

$$\mathrm{Vol}(M) \geq -2V_3\chi(\mathrm{Guts}(\overline{M-F})),$$

where V_3 is the volume of a regular ideal tetrahedron in hyperbolic 3-space.

In [Agol et al. 2007], this inequality was improved to

$$\mathrm{Vol}(M) \geq \mathrm{Vol}(\mathrm{Guts}(\overline{M-F})) \geq -V_{\mathrm{oct}}\chi(\mathrm{Guts}(\overline{M-F})),$$

where V_{oct} is the volume of a regular ideal octahedron in hyperbolic 3-space.

In this paper we will, building on ideas from [Agol 1999], prove a general inequality for the (transversal) Gromov norm $\|M\|_{\mathcal{F}}$ and the normal Gromov norm $\|M\|_{\mathcal{F}}^{\mathrm{norm}}$ of laminations.

To state the result in its general form we need two definitions.

Definition (pared acylindrical). Let Q be a manifold with a given decomposition

$$\partial Q = \partial_0 Q \cup \partial_1 Q.$$

The pair $(Q, \partial_1 Q)$ is called a pared acylindrical manifold if any continuous map of pairs $f : (\mathbb{S}^1 \times [0, 1], \mathbb{S}^1 \times \{0, 1\}) \to (Q, \partial_1 Q)$ that is π_1-injective as a map of pairs

MSC2000: primary 57R30; secondary 53C23, 57M27, 57N10, 57M50.

Keywords: lamination, hyperbolic, 3-manifold, volume, Gromov norm, tight, Weeks manifold, simplicial, simplex, straightening.

is necessarily homotopic, as a map of pairs $(\mathbb{S}^1 \times [0, 1], \mathbb{S}^1 \times \{0, 1\}) \to (Q, \partial_1 Q)$, into ∂Q.

Definition (essential decomposition). Let $(N, \partial N)$ be a pair of topological spaces such that $N = Q \cup R$ for two subspaces Q, R. Let

$\partial_0 Q = Q \cap R$, $\partial_1 Q = Q \cap \partial N$, $\partial_1 R = R \cap \partial N$, $\partial Q = \partial_0 Q \cup \partial_1 Q$, $\partial R = \partial_0 Q \cup \partial_1 R$.

We say that the decomposition $N = Q \cup R$ is an essential decomposition of $(N, \partial N)$ if the inclusions

$$\partial_1 Q \to Q \to N,\ \partial_1 R \to R \to N,\ \partial N \to N,\ \partial_0 Q \to Q,\ \partial_0 Q \to R$$

are each π_1-injective (for each path component).

Theorem 1.1. *Let M be a compact, orientable, connected n-manifold and $\mathcal{F}$ a lamination (of codimension one) of M. Assume that $N := \overline{M - \mathcal{F}}$ has a decomposition $N = Q \cup R$ into orientable n-manifolds (with boundary) Q, R such that the following assumptions are satisfied for $\partial_0 Q = Q \cap R$, $\partial_1 Q = Q \cap \partial N$, $\partial_1 R = R \cap \partial N$:*

(i) *Each path component of $\partial_0 Q$ has amenable fundamental group.*

(ii) *$(Q, \partial_1 Q)$ is pared acylindrical and $\partial_1 Q$ is acylindrical.*

(iii) *$Q, \partial N, \partial_1 Q, \partial_1 R, \partial_0 Q$ are aspherical.*

(iv) *The decomposition $N = Q \cup R$ is an essential decomposition of $(N, \partial N)$.*

Then

$$\|M, \partial M\|_{\mathcal{F}}^{\text{norm}} \geq \frac{1}{n+1} \|\partial Q\|.$$

In the case of 3-manifolds M carrying an essential lamination $\mathcal{F}$, considering $Q = \operatorname{Guts}(\overline{M - \mathcal{F}})$ yields a special case:

Theorem 1.2. *Let M be a compact 3-manifold with (possibly empty) boundary consisting of incompressible tori, and let $\mathcal{F}$ be an essential lamination of M. Then*

$$\|M, \partial M\|_{\mathcal{F}}^{\text{norm}} \geq -\chi(\operatorname{Guts}(\overline{M - \mathcal{F}})).$$

More generally, if P is a polyhedron with f faces, then

$$\|M, \partial M\|_{\mathcal{F}, P}^{\text{norm}} \geq -\frac{2}{f-2} \chi(\operatorname{Guts}(\overline{M - \mathcal{F}})).$$

The following corollary applies, for example, to all hyperbolic manifolds M obtained by Dehn-filling the complement of the figure-eight knot in $\mathbb{S}^3$. (It is known that each of these M contains tight laminations. By the following corollary, all these tight laminations have empty guts.)

Corollary 1.3. *If M is a finite-volume hyperbolic* 3*-manifold with* $\mathrm{Vol}(M) < 2V_3 = 2.02\ldots$, *then M carries no essential lamination $\mathcal{F}$ with $\|M\|^{\mathrm{norm}}_{\mathcal{F},P} = \|M\|_P$ for all polyhedra P, and nonempty guts. In particular, there is no tight essential lamination with nonempty guts.*

Calegari and Dunfield [2003] observed that their own results about tight laminations with empty guts would imply the following corollary, in the presence of a generalization of Agol's inequality to the case of tight laminations.

Corollary 1.4 [Calegari and Dunfield 2003, Conjecture 9.7]. *The Weeks manifold* (*the closed hyperbolic manifold of smallest volume*) *admits no tight lamination $\mathcal{F}$.*

Taking into account the main result of [Li 2006], this can be strengthened:

Corollary 1.5. *The Weeks manifold admits no transversely orientable essential lamination.*

We also have an application of Theorem 1.1 to higher-dimensional manifolds.

Corollary 1.6. *Let M be a compact Riemannian n-manifold of negative sectional curvature and finite volume. Let $F \subset M$ be a geodesic* $(n-1)$*-dimensional hypersurface of finite volume. Then* $\|F\| \leq \frac{1}{2}(n+1)\|M\|$.

The basic idea of Theorem 1.1, say for simplicity in the special situation of Corollary 1.6, is the following: a simplex which contributes to a normalized fundamental cycle of M should intersect $\partial Q = 2F$ in at most $n+1$ codimension-one simplices. This is of course not true in general: simplices can wrap around M many times and intersect F arbitrarily often, and even a homotopy rel vertices will not change this. As an obvious example, look at the following situation: Let γ be a closed geodesic transverse to F, and for some large N let σ be a straight simplex contained in a small neighborhood of γ^N. Then σ and F intersect N times and, since σ is already straight, this number of intersections can of course not be reduced by straightening. This shows that some more involved straightening must take place, and that the acylindricity of F is an essential condition. The way to use acylindricity will be to find a normalization such that many subsets of simplices are mapped to cylinders, which degenerate and thus can be removed without changing the homology class.

We remark that many technical points, including the use of multicomplexes, can be omitted if (in the setting of Theorem 1.2) one does not consider incompressible surfaces or essential laminations, but just geodesic surfaces in hyperbolic manifolds. In this case, all essential parts of the proof of Theorem 1.1 enter without the notational complications caused by the use of multicomplexes. Therefore we have given a fairly detailed outline of the proof for this special case in the beginning of Section 6. This should help to motivate the general proof in the second half of that section (156). (We mention that Theorem 1.1 is not true without assuming amenability of $\pi_1 \partial_0 Q$. This indicates that the proof of multicomplexes in the proof of Theorem 1.1 seems unavoidable.)

2. Preliminaries

2A. ***Laminations.*** Let M be an n-manifold, possibly with boundary. In this paper all manifolds will be smooth and orientable. (Hence they are triangulable by Whitehead's theorem and possess a locally finite fundamental class.) A (codimension 1) lamination $\mathcal{F}$ of M is a foliation of a closed subset $\mathcal{F}$ of M, i.e., a decomposition of a closed subset $\mathcal{F} \subset M$ into immersed codimension 1 submanifolds (leaves) so that M is covered by charts $\phi_j : \mathbb{R}^{n-1} \times \mathbb{R} \to M$, the intersection of any leaf with the image of any chart ϕ_j being a union of plaques of the form $\phi_j(\mathbb{R}^{n-1} \times \{*\})$. (We will denote by $\mathcal{F}$ both the lamination and the laminated subset of M, i.e., the union of leaves.) If M has boundary, we will always assume without further mentioning that $\mathcal{F}$ is *either transverse to* ∂M (that is, every leaf is transverse to $\mathcal{F}$) *or tangential to* ∂M (that is, ∂M is a leaf of $\mathcal{F}$). If neither of these two conditions were true, then the transverse and normal Gromov norm would be infinite, therefore all lower bounds will be trivially true.

To construct the leaf space T of $\mathcal{F}$, one considers the pull-back lamination $\widetilde{\mathcal{F}}$ on the universal covering $\widetilde{M}$. The space of leaves T is defined as the quotient of $\widetilde{M}$ under the following equivalence relation $\sim$. Two points $x, y \in \widetilde{M}$ are equivalent if either they belong to the same leaf of $\widetilde{\mathcal{F}}$, or they belong to the same connected component of the metric completion $\overline{\widetilde{M} - \widetilde{\mathcal{F}}}$ (for the path metric inherited by $\widetilde{M} - \widetilde{\mathcal{F}}$ from an arbitrary Riemannian metric on $\widetilde{M}$).

2B. ***Laminations of 3-manifolds.*** A lamination $\mathcal{F}$ of a 3-manifold M is called *essential* if no leaf is a sphere or a torus bounding a solid torus, $\overline{M - \mathcal{F}}$ is irreducible, and $\partial(\overline{M - \mathcal{F}})$ is incompressible and end-incompressible in $\overline{M - \mathcal{F}}$, where again the metric completion $\overline{M - \mathcal{F}}$ of $M - \mathcal{F}$ is taken with respect to the path metric inherited from any Riemannian metric on M; see [Gabai and Oertel 1989, Chapter 1]. (Note that $\overline{M - \mathcal{F}}$ is immersed in M, the leaves of $\mathcal{F}$ in the image of the immersion are called boundary leaves.)

Examples of essential laminations are taut foliations or compact, incompressible, boundary-incompressible surfaces in compact 3-manifolds. (We always consider laminations without isolated leaves. If a lamination has isolated leaves, then it can be converted into a lamination without isolated leaves by replacing each two-sided isolated leaf S_i with the trivially foliated product $S_i \times [0, 1]$, resp. each one-sided isolated leaf with the canonically foliated normal I-bundle, without changing the topological type of M.)

If $\mathcal{F}$ is an essential lamination, then the leaf space T is an order tree, with segments corresponding to directed, transverse, efficient arcs. (An order tree T is a set T with a collection of linearly ordered subsets, called segments, such that the axioms of [Gabai and Oertel 1989, Definition 6.9], are satisfied.) Moreover, T is an $\mathbb{R}$-order tree, that is, it is a countable union of segments and each segment is order isomorphic to a closed interval in $\mathbb{R}$. T can be topologized by the order topology on segments (and declaring that a set is closed if the intersection with each segment is closed). For this topology, $\pi_0 T$ and $\pi_1 T$ are trivial (see, for example, [Roberts et al. 2003], Chapter 5, and its references).

The order tree T comes with a fixed-point free action of $\pi_1 M$. Fenley [2007] has exhibited hyperbolic 3-manifolds whose fundamental groups do not admit any fixed-point free action on $\mathbb{R}$-order trees. Thus there are hyperbolic 3-manifolds not carrying any essential lamination.

If M is hyperbolic and $\mathcal{F}$ an essential lamination, then $\overline{M - \mathcal{F}}$ has a characteristic submanifold which is the maximal submanifold that can be decomposed into I-bundles and solid tori, respecting boundary patterns (see [Jaco and Shalen 1979], [Johannson 1979] for precise definitions). The complement of this characteristic submanifold is denoted by Guts($\mathcal{F}$). It admits a hyperbolic metric with geodesic boundary and cusps. (Be aware that some authors, like [Calegari and Dunfield 2003], include the solid tori into the guts.) If $\mathcal{F} = F$ is a properly embedded, incompressible, boundary-incompressible surface, then Agol's inequality states that $\mathrm{Vol}(M) \geq -2V_3\chi(\mathrm{Guts}(F))$. This implies, for example, that a hyperbolic manifold of volume $< 2V_3$ can not contain any geodesic surface of finite area. Agol, Storm, and Thurston [Agol et al. 2007], using estimates coming from Perelman's work on the Ricci flow, have improved this inequality to

$$\mathrm{Vol}(M) \geq \mathrm{Vol}(\mathrm{Guts}(F)) \geq -V_{\mathrm{oct}}\chi(\mathrm{Guts}(F)).$$

Assume that $\mathcal{F}$ is a codimension one lamination of an n-manifold M such that its leaf space T is an $\mathbb{R}$-order tree. (For example this is the case if $n = 3$ and $\mathcal{F}$ is essential.) An essential lamination is called *tight* if T is Hausdorff. It is called unbranched if T is homeomorphic to $\mathbb{R}$. It is said to have *two-sided branching* [Calegari 2000, Definition 2.5.2] if there are leaves $\lambda, \lambda_1, \lambda_2, \mu, \mu_1, \mu_2$ such that the corresponding points in the T satisfy $\lambda < \lambda_1$, $\lambda < \lambda_2$, $\mu > \mu_1$, $\mu > \mu_2$, but λ_1, λ_2 are incomparable and μ_1, μ_2 are incomparable. It is said to have one-sided branching if it is neither unbranched nor has two-sided branching.

If M is a hyperbolic 3-manifold and carries a tight lamination with empty guts, we know from [2003, Theorem 3.2] that $\pi_1 M$ acts effectively on the circle, i.e., there is an injective homomorphism $\pi_1 M \to \mathrm{Homeo}(\mathbb{S}^1)$. This implies that the Weeks manifold cannot carry a tight lamination with empty guts [Calegari and Dunfield 2003, Corollary 9.4]. The aim of this paper is to find obstructions to the existence of laminations with nonempty guts.

2C. *Simplicial volume and refinements.* Let M be a compact, orientable, connected n-manifold, possibly with boundary. Its top integer (singular) homology group $H_n(M, \partial M; \mathbb{Z})$ is cyclic. The image of a generator under the change-of-coefficients homomorphism $H_n(M, \partial M; \mathbb{Z}) \to H_n(M, \partial M; \mathbb{R})$ is called a fundamental class and is denoted $[M, \partial M]$. If M is not connected, we define $[M, \partial M]$ to be the formal sum of the fundamental classes of its connected components.

The simplicial volume $\|M, \partial M\|$ is defined as

$$\|M, \partial M\| = \inf\left\{\sum_{i=1}^{r} |a_i|\right\},$$

where the infimum is taken over all singular chains $\sum_{i=1}^{r} a_i \sigma_i$ (with real coefficients) representing the fundamental class in $H_n(M, \partial M; \mathbb{R})$.

If $M - \partial M$ carries a complete hyperbolic metric of finite volume $\mathrm{Vol}(M)$, then

$$\|M, \partial M\| = \frac{1}{V_n} \mathrm{Vol}(M),$$

with $V_n = \sup\{\mathrm{Vol}(\Delta) : \Delta \subset \mathbb{H}^n \text{ geodesic simplex}\}$; see [Gromov 1982; Thurston 1980; Benedetti and Petronio 1992; Francaviglia 2004].

More generally, let P be any polyhedron. Then the invariant $\|M, \partial M\|_P$ is defined in [Agol 1999] as follows: denoting by $C_*(M, \partial M; P; \mathbb{R})$ the complex of P-chains with real coefficients and by $H_*(M, \partial M; P; \mathbb{R})$ its homology, there is a canonical chain homomorphism $\psi : C_*(M, \partial M; P; \mathbb{R}) \to C_*(M, \partial M; \mathbb{R})$, given by certain triangulations of P which are to be chosen so that all possible cancellations of boundary faces are preserved. Then $\|M, \partial M\|_P$ is defined as the infimum of $\sum_{i=1}^{r} |a_i|$ over all P-chains $\sum_{i=1}^{r} a_i P_i$ such that $\psi(\sum_{i=1}^{r} a_i P_i)$ represents the fundamental class $[M, \partial M]$. Set $V_P := \sup\{\mathrm{Vol}(\Delta)\}$, where the supremum is taken over all straight P-polyhedra $\Delta \subset \mathbb{H}^3$.

Proposition 2.1 [Agol 1999, Lemma 4.1]. *If $M - \partial M$ admits a hyperbolic metric of finite volume* $\mathrm{Vol}(M)$, *then*

$$\|M, \partial M\|_P = \frac{1}{V_P} \mathrm{Vol}(M).$$

(The proof in [Agol 1999] is quite short, and it does not give details for the cusped case. However, the proof in the cusped case can be completed using the arguments in [Francaviglia 2004, Sections 5 and 6].)

Let M be a manifold and $\mathcal{F}$ a codimension-one lamination of M. Let Δ^n be the standard simplex in $\mathbb{R}^{n+1}$, and $\sigma : \Delta^n \to M$ some continuous singular simplex. The lamination $\mathcal{F}$ induces an equivalence relation on Δ^n, whereby $x \sim y$ if and only if $\sigma(x)$ and $\sigma(y)$ belong to the same connected component of $L \cap \sigma(\Delta^n)$ for some leaf L of $\mathcal{F}$. We say that a singular simplex $\sigma : \Delta^n \to M$ is laminated if the equivalence relation $\sim$ is induced by a lamination $\mathcal{F}|_\sigma$ of Δ^n. We call a lamination $\mathcal{F}$ of Δ^n *affine* if there is an affine mapping $f : \Delta^n \to \mathbb{R}$ such that $x, y \in \Delta^n$ belong to the same leaf if and only if $f(x) = f(y)$. We say that a lamination $\mathcal{G}$ of Δ^n is conjugate to an affine lamination if there is a simplicial homeomorphism $H : \Delta^n \to \Delta^n$ such that $H^*\mathcal{G}$ is an affine lamination.

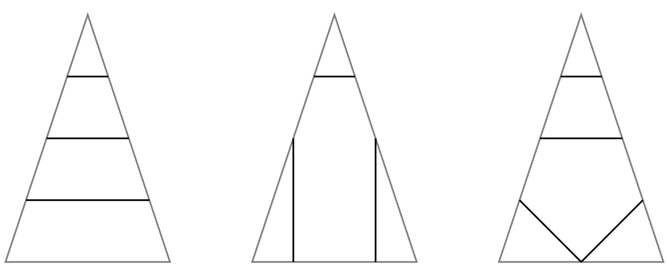
We say that a singular n-simplex $\sigma : \Delta^n \to M$, $n \geq 2$, is *transverse* to $\mathcal{F}$ if it is laminated and it is either contained in a leaf, or $\mathcal{F}|_\sigma$ is conjugate to an affine lamination $\mathcal{G}$ of Δ^n.

For $n = 1$, we say that a singular 1-simplex $\sigma : \Delta^1 \to M$ is transverse to $\mathcal{F}$ if it is either contained in a leaf, or for each lamination chart $\phi : U \to \mathbb{R}^{m-1} \times \mathbb{R}^1$ (with m-th coordinate map $\phi_m : U \to \mathbb{R}^1$) one has that $\phi_m \circ \sigma|_{\sigma^{-1}(U)} : \sigma^{-1}(U) \to \mathbb{R}^1$ is locally surjective at all points of $\operatorname{int}(\Delta^1)$, i.e., for all $p \in \operatorname{int}(\Delta^1) \cap \sigma^{-1}(U)$, the image of $\phi_m \circ \sigma|_{\sigma^{-1}(U)}$ contains a neighborhood of $\phi_m \circ \sigma(p)$.

We say that the simplex $\sigma : \Delta^n \to M$ is *normal* to $\mathcal{F}$ if, for each leaf F, $\sigma^{-1}(F)$ consists of normal disks, i.e., disks meeting each edge of Δ^n at most once. (If $F = \partial M$ is a leaf of $\mathcal{F}$ we also allow that $\sigma^{-1}(F)$ can be a face of Δ^n). In particular, any transverse simplex is normal.

In the special case of foliations, $\mathcal{F}$ one has that the transversality of a singular simplex σ is implied by (hence equivalent to) the normality of σ, as can be shown along the lines of [Kuessner 2004, Section 1.3].

More generally, let P be any polyhedron. Then we say that a singular polyhedron $\sigma : P \to M$ is normal to $\mathcal{F}$ if, for each leaf F, $\sigma^{-1}(F)$ consists of normal disks, i.e., disks meeting each edge of P at most once (or being equal to a face of P, if F is a boundary leaf).



Definition 2.2. Let M be a compact, oriented, connected n-manifold, possibly with boundary, and let $\mathcal{F}$ be a foliation or lamination on M. Let Δ^n be the standard simplex and P any polyhedron. Let Σ be the set of singular simplices $\Delta^n \to M$ transverse to $\mathcal{F}$. We define

$$\|M, \partial M\|_{\mathcal{F}} := \inf\left\{ \sum_{i=1}^{r} |a_i| : \psi\left(\sum_{i=1}^{r} a_i\sigma_i\right) \text{ represents } [M, \partial M] \text{ for some } \sigma_i \in \Sigma \right\}$$

and

$$\|M, \partial M\|^{\text{norm}}_{\mathcal{F},P} := \inf\Big\{\sum_{i=1}^{r} |a_i| : \psi\Big(\sum_{i=1}^{r} a_i\sigma_i\Big) \text{ represents } [M, \partial M] \text{ for some } \sigma_i \in \Sigma\Big\}.$$

In particular, we define $\|M, \partial M\|^{\text{norm}}_{\mathcal{F}} = \|M, \partial M\|^{\text{norm}}_{\mathcal{F},\Delta^n}$.

All these norms are finite, under the assumption that $\mathcal{F}$ is transverse or tangential to ∂M. There are obvious inequalities

$$\|M, \partial M\| \leq \|M, \partial M\|^{\text{norm}}_{\mathcal{F}} \leq \|M, \partial M\|_{\mathcal{F}}.$$

In the case of foliations, this last inequalities becomes an equality.

(We remark that all definitions extend in an obvious way to disconnected manifolds by summing over the connected components.)

The next proposition and lemma are straightforward generalizations of [Calegari 2000, Theorem 2.5.9] and of arguments in [Agol 1999].

Proposition 2.3. *Let M be a compact, oriented 3-manifold.*

(a) *If $\mathcal{F}$ is an essential lamination which is either unbranched or has one-sided branching such that the induced lamination of ∂M is unbranched, then*

$$\|M, \partial M\|^{\text{norm}}_{\mathcal{F},P} = \|M, \partial M\|_P$$

for each polyhedron P.

(b) *If $\mathcal{F}$ is a tight essential lamination, then*

$$\|M, \partial M\|^{\text{norm}}_{\mathcal{F},P} = \|M, \partial M\|_P$$

for each polyhedron P.

Proof. Since $\mathcal{F}$ is an essential lamination, we know from [Gabai and Oertel 1989, Theorem 6.1] that the leaves are π_1-injective, the universal covering $\widetilde{M}$ is homeomorphic to $\mathbb{R}^3$ and that the leaves of the pull-back lamination are planes, in particular aspherical. Therefore Proposition 2.3 is a special case of the next result. □

Lemma 2.4. *Let M be a compact, oriented, aspherical manifold, and $\mathcal{F}$ a lamination of codimension one. Assume that the leaves are π_1-injective and aspherical, and that the leaf space T is an $\mathbb{R}$-order tree.*

(a) *If the leaf space T is either $\mathbb{R}$ or branches in only one direction, so that the induced lamination of ∂M has leaf space $\mathbb{R}$, then $\|M, \partial M\|^{\text{norm}}_{\mathcal{F},P} = \|M, \partial M\|_P$ for each polyhedron P.*

(b) *If the leaf space is a Hausdorff tree, then $\|M, \partial M\|^{\text{norm}}_{\mathcal{F},P} = \|M, \partial M\|_P$ for each polyhedron P.*

Proof. To prove the wanted equalities, it suffices in each case to show that any (relative) cycle can be homotoped to a cycle consisting of normal polyhedra. We denote by $\widetilde{\mathcal{F}}$ the pull-back lamination of $\widetilde{M}$ and by $p : \widetilde{M} \to T = \widetilde{M}/\widetilde{\mathcal{F}}$ the projection to the leaf space.

(a) First we consider the case that P is a simplex [Calegari 2000, Section 4.1] and $\mathcal{F}$ is unbranched. For this case, we can repeat the argument in [Calegari 2000, Lemma 2.2.8]. Namely, given a (relative) cycle $\sum_{i=1}^{r} a_i \sigma_i$, lift it to a $\pi_1 M$-equivariant (relative) cycle on $\widetilde{M}$ and then perform an (equivariant) straightening, by induction on the dimension of subsimplices of the lifts $\tilde{\sigma}_i$ as follows: for each edge $\tilde{e}$ of any lift $\tilde{\sigma}_i$, its projection $p(\tilde{e})$ to the leaf space T is homotopic to a unique straight arc $\mathrm{str}(p(\tilde{e}))$ in $T \simeq \mathbb{R}$. It is easy to see (covering the arc by foliation charts and then extending the lifted arc stepwise) that $\mathrm{str}(p(\tilde{e}))$ can be lifted to an arc $\mathrm{str}(\tilde{e})$ with the same endpoints as $\tilde{e}$, and that the homotopy between $\mathrm{str}(p(\tilde{e}))$ and $p(\tilde{e})$ can be lifted to a homotopy between $\mathrm{str}(\tilde{e})$ and $\tilde{e}$. $\mathrm{str}(\tilde{e})$ is transverse to $\mathcal{F}$, because its projection is a straight arc in T. These homotopies of edges can be extended to a homotopy of the whole (relative) cycle. Thus we have straightened the 1-skeleton of the given (relative) cycle.

Now let us be given a 2-simplex $\tilde{f} : \Delta^2 \to \widetilde{M}$ with transverse edges. There is an obvious straightening $\mathrm{str}(p(\tilde{f}))$ of $p(\tilde{f}) : \Delta^2 \to T$ as follows: if, for $t \in T$, $(p\tilde{f})^{-1}(t)$ has two preimages x_1, x_2 on edges of Δ^2 (which are necessarily unique), then $\mathrm{str}(p(\tilde{f}))$ maps the line which connects x_1 and x_2 in Δ^2 constantly to t. It is clear that this defines a continuous map $\mathrm{str}(p(\tilde{f})) : \Delta^2 \to T$.

Since the leaves $\widetilde{F}$ of $\widetilde{\mathcal{F}}$ are connected ($\pi_0 \widetilde{F} = 0$), $\mathrm{str}(p(\tilde{f}))$ can be lifted to a map $\mathrm{str}(\tilde{f}) : \Delta^2 \to \widetilde{M}$ with $p(\mathrm{str}(\tilde{f})) = \mathrm{str}(p(\tilde{f}))$. The 2-simplex $\mathrm{str}(\tilde{f})$ is transverse to $\mathcal{F}$, because its projection is a straight simplex in T.

There is an obvious homotopy between $p(\tilde{f})$ and $\mathrm{str}(p(\tilde{f}))$. For each $t \in T$, the restriction of the homotopy to $(p\tilde{f})^{-1}(t)$ can be lifted to a homotopy in $\widetilde{M}$, because $\pi_1 \widetilde{M} = 0$. Since $\pi_2 \widetilde{M} = 0$, these homotopies for various $t \in T$ fit together continuously to give a homotopy between $\tilde{f}$ and $\mathrm{str}(\tilde{f})$.

These homotopies of 2-simplices leave the (already transverse) boundaries pointwise fixed; thus they can be extended to a homotopy of the whole (relative) cycle. Hence we have straightened the 2-skeleton of the given (relative) cycle.

Assume that we have already straightened the k-skeleton, for some $k \in \mathbb{N}$. The analogous procedure, using $\pi_{k-1} \widetilde{F} = 0$ for all leaves, and $\pi_k \widetilde{M} = 0$, $\pi_{k+1} \widetilde{M} = 0$, allows to straighten the $(k+1)$-skeleton of the (relative) cycle. This finishes the proof in the case that $\mathcal{F}$ is unbranched.

The generalization to the case that $\mathcal{F}$ has one-sided branching and the induced lamination of ∂M is unbranched works as in [Calegari 2000, Theorem 2.6.6].

We remark that in the case that P is a simplex we get not only a normal cycle, but even a transverse cycle.

Now we consider the case of arbitrary polyhedra P. Let $\sum_{i=1}^{r} a_i \sigma_i$ be a P-cycle. It can be subtriangulated to a simplicial cycle $\sum_{i=1}^{r} a_i \sum_{j=1}^{s} \tau_{i,j}$. Again the argument in [Calegari 2000, Lemma 2.2.8], and the corresponding argument for manifolds with boundary, shows that this simplicial cycle can be homotoped such that each $\tau_{i,j}$ is transverse (and such that boundary cancellations are preserved). But transversality of each $\tau_{i,j}$ implies by definition that $\sigma_i = \sum_{j=1}^{s} \tau_{i,j}$ is normal (though in general not transverse) to $\mathcal{F}$.

(b) By assumption $\widetilde{M}/\widetilde{\mathcal{F}}$ is a Hausdorff tree. Its branching points are the projections of complementary regions: Indeed, if F is a leaf of $\mathcal{F}$, then $\widetilde{F}$ is a submanifold of the contractible manifold $\widetilde{M}$. By asphericity and π_1-injectivity of F, $\widetilde{F}$ must be contractible. By Alexander duality it follows that $\widetilde{M} - \widetilde{F}$ has two connected components. Therefore the complement of the point $p(\widetilde{F})$ in the leaf space has (at most) two connected components, so $p(\widetilde{F})$ cannot be a branch point.

Again, to define a straightening of P-chains it suffices to define a canonical straightening of singular polyhedra P such that straightenings of common boundary faces will agree. Let $\tilde{v}_0, \dots, \tilde{v}_n$ be the vertices of the image of P. For each pair $\{\tilde{v}_i, \tilde{v}_j\}$ there exists at most one edge $\tilde{e}_{ij}$ with vertices $\tilde{v}_i, \tilde{v}_j$ in the image of P. Since the leaf space is a tree, we have a unique straight arc $\mathrm{str}(p(\tilde{e}_{ij}))$ connecting the points $p(\tilde{v}_i)$ and $p(\tilde{v}_j)$ in the leaf space. As in (a), one can lift this straight arc $\mathrm{str}(p(\tilde{e}_{ij}))$ to an arc $\mathrm{str}(\tilde{e}_{ij})$ in $\widetilde{M}$, connecting $\tilde{v}_i$ and $\tilde{v}_j$, which is transverse to $\mathcal{F}$. We define this arc $\mathrm{str}(\tilde{e}_{ij})$ to be the straightening of $\tilde{e}_{ij}$. As in (a), we have homotopies of 1-simplices, which extend to a homotopy of the whole (relative) cycle. Thus we have straightened the 1-skeleton.

Now let us be given the 3 vertices $\tilde{v}_0, \tilde{v}_1, \tilde{v}_2$ of a 2-simplex $\tilde{f}$ with straight edges. If the projections $p(\tilde{v}_0), p(\tilde{v}_1), p(\tilde{v}_2)$ belong to a subtree isomorphic to a connected subset of $\mathbb{R}$, then we can straighten $\tilde{f}$ as in (a). If not, the projection of the 1-skeleton of this simplex has exactly one branch point, which corresponds to a complementary region. (The projection may of course meet many branch points of the tree, but the image of the projection, considered as a subtree, can have at most one branch point. In general, a subtree with n vertices can have at most $n-2$ branch points.) The preimage of the complement of this complementary region consists of three connected subsets of the 2-simplex (the "corners" around the vertices). We can straighten each of these subsets and do not need to care about the complementary region corresponding to the branch point. Thus we have straightened the 2-skeleton.

Assume that we have already straightened the k-skeleton, for some $k \in \mathbb{N}$. Given the $k+2$ vertices $\tilde{v}_0, \tilde{v}_1, \dots, \tilde{v}_{k+1}$ of a $(k+1)$-simplex with straight faces, we have (at most k) branch points in the projection of the simplex, which correspond to complementary regions. Again we can straighten the parts of the simplex which do not belong to these complementary regions as in (a), since they are projected to linearly ordered subsets of the tree. Thus we have straightened the $(k+1)$-skeleton.

Since, by the recursive construction, we have defined straightenings of simplices with common faces by first defining (the same) straightenings of their common faces, the straightening of a (relative) cycle will be again a (relative) cycle, in the same (relative) homology class. □

Remark. For $\|M\|_{\mathcal{F}}$ instead of $\|M\|_{\mathcal{F}}^{\text{norm}}$, equality (b) is in general wrong, and equality (a) is unknown (but presumably wrong).

If $\mathcal{F}$ is essential but not tight, one may still try to homotope cycles to be transverse, by possibly changing the lamination. In the special case that the cycle is coming from a triangulation, this has been done by Brittenham [1995] and Gabai [1999]. It is not obvious how to generalize their arguments to cycles with overlapping simplices.

3. Retracting chains to codimension zero submanifolds

3A. ***Definitions.*** The results of this section are essentially all due to Gromov, but we follow mainly our exposition in [Kuessner 2010]. We start with some recollections about multicomplexes; for details, see [Gromov 1982, Section 3; Kuessner 2010, Section 1].

A multicomplex K is a topological space $|K|$ with a decomposition into simplices, where each n-simplex is attached to the $(n-1)$-skeleton K_{n-1} by a simplicial homeomorphism $f:\partial\Delta^n \to K_{n-1}$. (In particular, each n-simplex has $n+1$ distinct vertices.) In contrast with simplicial complexes, in a multicomplex there may be n-simplices with the same $(n-1)$-skeleton.

We call a multicomplex minimally complete if the following condition holds: Let $\sigma : \Delta^n \to |K|$ be a singular n-simplex such that $\partial_0\sigma, \dots, \partial_n\sigma$ are distinct simplices of K. Then σ is homotopic relative $\partial\Delta^n$ to a *unique* simplex in K.

We call a minimally complete multicomplex K aspherical if all simplices $\sigma \neq \tau$ in K satisfy $\sigma_1 \neq \tau_1$. That means that simplices are uniquely determined by their 1-skeleton.

Orientations of multicomplexes are defined as usual in simplicial theory. If σ is a simplex, $\bar{\sigma}$ will denote the simplex with the opposite orientation.

A submulticomplex L of a multicomplex K is a subset of the set of simplices closed under face maps. (K, L) is a pair of multicomplexes if K is a multicomplex and L is a submulticomplex of K.

A group G acts simplicially on a pair of multicomplexes (K, L) if it acts on the set of simplices of K, mapping simplices in L to simplices in L, so that the action commutes with all face maps. For $g \in G$ and σ a simplex in K, we denote by $g\sigma$ the simplex obtained by this action.

3B. ***Construction of $K(X)$.*** We recall the construction from [Kuessner 2010, Section 1.3] (originally found in [Gromov 1982, pp. 45–46]).

For a topological space X, we denote by $S_*(X)$ the simplicial set of all singular simplices in X and by $|S_*(X)|$ its geometric realization.

For a topological space X, a multicomplex $\widehat{K}(X) \subset |S_*(X)|$ is constructed as follows. The 0-skeleton $\widehat{K}_0(X)$ equals $S_0(X)$. The 1-skeleton $\widehat{K}_1(X)$ contains one element in each homotopy class (rel $\{0, 1\}$) of singular 1-simplices $f : [0, 1] \to X$ with $f(0) \neq f(1)$. For $n \geq 2$, assuming by recursion that the $(n-1)$-skeleton is defined, the n-skeleton $\widehat{K}_n(X)$ contains *one* singular n-simplex in each homotopy class (rel boundary) of singular n-simplices $f : \Delta^n \to X$ with $\partial f \in \widehat{K}_{n-1}(X)$. We can choose $\widehat{K}(X)$ with the property that $\sigma \in \widehat{K}(X) \iff \bar{\sigma} \in \widehat{K}(X)$ (recall that the bar denotes orientation reversal). We will henceforth assume that $\widehat{K}(X)$ is constructed according to this condition.

According to [Gromov 1982], $|\widehat{K}(X)|$ is weakly homotopy equivalent to X.

The multicomplex $K(X)$ is defined as the quotient

$$K(X) := \widehat{K}(X)/\sim$$

where simplices in $\widehat{K}(X)$ are identified if and only if they have the same 1-skeleton. Let p be the canonical projection $p : \widehat{K}(X) \to K(X)$.

$K(X)$ is minimally complete and aspherical.

If $X' \subset X$ is a subspace, we have (not necessarily injective) simplicial mappings $\hat{j} : \widehat{K}(X') \to \widehat{K}(X)$ and $j : K(X') \to K(X)$.

If $\pi_1 X' \to \pi_1 X$ is injective (for each path-connected component of X'), then j is injective ([Kuessner 2010], Section 1.3) and we can (and will) consider $K(X')$ as a submulticomplex of $K(X)$. (Since simplices in $\widehat{K}(X')$ have image in X', this means that we assume we have constructed $\widehat{K}(X)$ so that simplices in $\widehat{K}(X)$ have image in X' whenever this is possible.) If moreover $\pi_n X' \to \pi_n X$ is injective for all $n \geq 2$ (say, if X' is aspherical), then $\hat{j}$ is also injective and $\widehat{K}(X')$ can be considered as a submulticomplex of $\widehat{K}(X)$.

In particular, *if X and X' are aspherical and $\pi_1 X' \to \pi_1 X$ is injective, there is an inclusion*

$$i_* : C_*^{\text{simp}}(K(X), K(X')) = C_*^{\text{simp}}(\widehat{K}(X), \widehat{K}(X')) \to C_*^{\text{sing}}(X, X')$$

into the relative singular chain complex of (X, X').

3C. ***Infinite and locally finite chains.*** In this paper we will also work with infinite chains, and in particular with locally finite chains on noncompact manifolds, as introduced in [Gromov 1982, Section 0.2].

For a topological space X, a formal sum $\sum_{i \in I} a_i \sigma_i$ of singular k-simplices with real coefficients (with a possibly infinite index set I, and the convention $a_i \neq 0$

for $i \in I$) is an infinite singular k-chain. It is said to be a locally finite chain if each point of X is contained in the image of at most finitely many σ_i. Infinite k-chains form a real vector space denoted by $C_k^{\mathrm{inf}}(X)$, and locally finite k-chains one denoted by $C_k^{\mathrm{lf}}(X)$. The boundary operator maps locally finite k-chains to locally finite $(k-1)$-chains, hence, for a pair of spaces (X, X') the homology $H_*^{\mathrm{lf}}(X, X')$ of the complex of locally finite chains can be defined.

For a noncompact, orientable n-manifold X with (possibly noncompact) boundary ∂X, one has a fundamental class $[X, \partial X] \in H_n^{\mathrm{lf}}(X, \partial X)$. We will say that an infinite chain $\sum_{i\in I} a_i\sigma_i$ represents $[X, \partial X]$ if it is homologous to a locally finite chain representing $[X, \partial X] \in H_n^{\mathrm{lf}}(X, \partial X)$.

For a simplicial complex K, we denote by $C_k^{\mathrm{simp,inf}}(K)$ the $\mathbb{R}$-vector space of (possibly infinite) formal sums $\sum_{i\in I} a_i\sigma_i$ with $a_i \in \mathbb{R}$ and σ_i k-simplices in K. If $\pi_n X' \to \pi_n X$ is injective for $n \geq 1$, we have again the obvious inclusion $i_* : C_*^{\mathrm{simp,inf}}(\widehat{K}(X), \widehat{K}(X')) \to C_*^{\mathrm{inf}}(X, X')$.

The following observation is of course a well-known application of the homotopy extension property, but we will use it so often that we state it here for reference.

Observation 3.1. *Let X be a topological space and $\sigma_0 : \Delta^n \to X$ a singular simplex. Let $H : \partial\Delta^n \times I \to X$ be a homotopy with $H(x, 0) = \sigma_0(x)$ for all $x \in \partial\Delta^n$. Then there exists a homotopy $\overline{H} : \Delta^n \times I \to X$ with $\overline{H}|_{\partial\Delta^n\times I} = H$ and $\overline{H}|_{\Delta^n\times\{0\}} = \sigma_0$.*

If $X' \subset X$ is a subspace and the images of σ_0 and H belong to X', we can choose $\overline{H}$ so that its image belongs to X'.

Lemma 3.2. *Let (X, X') be a pair of topological spaces. Assume $\pi_n X' \to \pi_n X$ is injective for each path component of X' and each $n \geq 1$.*

(a) *Let $\sum_{i\in I} a_i\tau_i \in C_n^{\mathrm{inf}}(X, X')$ be a (possibly infinite) singular n-chain. Assume that I is countable, and that each path component of X and each nonempty path component of X' contain uncountably many points. Then $\sum_{i\in I} a_i\tau_i$ is homotopic to a (possibly infinite) simplicial chain*

$$\sum_{i\in I} a_i\tau_i' \in C_n^{\mathrm{simp,inf}}(\widehat{K}(X), \widehat{K}(X')) \subset C_*^{\mathrm{inf}}(X, X').$$

In particular, $\sum_{i\in I} a_i\tau_i'$ is homologous to $\sum_{i\in I} a_i\tau_i$.

(b) *Let $\sigma_0 \in \widehat{K}(X)$ and $H : \Delta^n \times [0, 1] \to X$ a homotopy with $H(\cdot, 0) = \sigma_0$. Consider a minimal triangulation $\Delta^n \times [0, 1] = \Delta_0 \cup \ldots \Delta_n$ of $\Delta^n \times [0, 1]$ into $n+1$ $(n+1)$-simplices. Assume that $H(\partial\Delta^n \times [0, 1])$ consists of simplices in $\widehat{K}(X)$. Then H is homotopic (rel $\Delta^n \times \{0\} \cup \partial\Delta^n \times [0, 1]$) to a map $\overline{H} : \Delta^n \times [0, 1] \to X$ such that $\overline{H}|_{\Delta_i} \in \widehat{K}(X)$; in particular $\sigma_1 := \overline{H}(\cdot, 1) \in \widehat{K}(X)$.*

Proof. (a) From the assumptions it follows that there exists a homotopy of the 0-skeleton such that each vertex is moved into a distinct point of X, and such that vertices in X' remain in X' during the homotopy. By Observation 3.1, this homotopy can by induction be extended to a homotopy of the whole chain.

Now we prove the claim by induction on k ($0 \le k < n$). We assume that the k-skeleton of $\sum_{i\in I} a_i\tau_i$ consists of simplices in $\widehat{K}(X)$ and we want to homotope $\sum_{i\in I} a_i\tau_i$ such that the homotoped $(k+1)$-skeleton consists of simplices in $\widehat{K}(X)$.

By construction, each singular $(k+1)$-simplex σ in X with boundary a simplex in $\widehat{K}(X)$ is homotopic (rel boundary) to a unique $(k+1)$-simplex in $\widehat{K}(X)$. Since the homotopy keeps the boundary fixed, the homotopies of different $(k+1)$-simplices are compatible. By Observation 3.1, the homotopy of the $(k+1)$-skeleton can by induction be extended to a homotopy of the whole chain.

If the image of the $(k+1)$-simplex σ is contained in X', then it is homotopic rel boundary to a simplex in $\widehat{K}(X')$, for a homotopy with image in X'. Thus we can realize the homotopy in such a way that all simplices with image in X' are homotoped inside X'.

(b) follows by the same argument as (a), successively applied to $\Delta_0, \dots, \Delta_n$. $\square$

We remark that there exists a canonical simplicial map

$$p : C_*^{\text{simp,inf}}(\widehat{K}(X), \widehat{K}(X')) \to C_*^{\text{simp,inf}}(K(X), K(X')),$$

defined by induction. It is defined to be the identity on the 1-skeleton. If it is defined on the $(n-1)$-skeleton, for $n \ge 2$, then, for an n-simplex τ, $p(\tau) \in K(X)$ is the unique simplex with $\partial_i p(\tau) = p(\partial_i\tau)$ for $i = 0, \dots, n$.

3D. ***Action of $G = \Pi(A)$.*** We repeat the definitions from [Kuessner 2010, Section 1.5] (originally due to Gromov), as they will be frequently used in the remainder of the paper.

Let (P, A) be a pair of minimally complete multicomplexes. We define its space of nontrivial loops $\Omega^* A$ as the set of homotopy classes (rel $\{0, 1\}$) of continuous maps $\gamma : [0, 1] \to |A|$ with $\gamma(0) = \gamma(1)$ and not homotopic (rel $\{0, 1\}$) to a constant map.

We define

$$\Pi(A) := \big\{\{\gamma_1, \dots, \gamma_n\} : n \in \mathbb{N},\ \gamma_1, \dots, \gamma_n \in A_1 \cup \Omega^* A,\ \gamma_i(0) = \gamma_i(1) \text{ for all } i, \\ \gamma_i(0) \ne \gamma_j(0), \gamma_i(1) \ne \gamma_j(1) \text{ for } i \ne j\big\}.$$

If γ, γ' are elements of A_1 with $\gamma' \ne \overline{\gamma}$ and $\gamma(0) = \gamma'(1)$, we denote by $\gamma * \gamma' \in A_1$ the unique edge of A in the homotopy class of the concatenation.[1] If $\gamma \in A_1$ and $\gamma' \in \Omega^* A$ (or vice versa), with $\gamma(1) \ne \gamma(0) = \gamma'(1) = \gamma'(0)$, we also denote

[1] We follow the usual convention of defining the concatenation of paths by $\gamma * \gamma'(t) = \gamma(2t)$ if $t \le \frac{1}{2}$ and $\gamma * \gamma'(t) = \gamma'(2t-1)$ if $t \ge \frac{1}{2}$. Unfortunately this implies that, in order for $\Pi(A)$ to act on P, we will need the multiplication in $\Pi(A)$ to satisfy, for example, $\{\gamma\}\{\gamma'\} = \{\gamma' * \gamma\}$. We hope that this does not lead to confusion.

by $\gamma * \gamma' \in A_1$ the unique edge in the homotopy class of the concatenation. If $\gamma, \gamma' \in \Omega^* A$ with $\gamma(1) = \gamma(0) = \gamma'(1) = \gamma'(0)$, we denote by $\gamma * \gamma' \in \Omega^* A$ the concatenation of homotopy classes of loops.

We then define a multiplication on $\Pi(A)$ as follows: Given $\{\gamma_1, \dots, \gamma_m\}$ and $\{\gamma'_1, \dots, \gamma'_n\}$, we reindex the unordered sets $\{\gamma_1, \dots, \gamma_m\}$ and $\{\gamma'_1, \dots, \gamma'_n\}$ so that $\gamma_j(1) = \gamma'_j(0)$ for $1 \le j \le i$ and $\gamma_j(1) \ne \gamma'_k(0)$ for $j \ge i+1$ and $k \ge i+1$. (Since we are assuming that all $\gamma_j(1)$ are pairwise distinct, and also all $\gamma'_j(0)$ are pairwise distinct, such a reindexing exists for some $i \ge 0$, and it is unique up to permuting the indices $\le i$ and permuting separately the indices of the γ_j and γ'_k with $j \ge i+1$ and $k \ge i+1$.) Moreover we permute the indices $\{1, \dots, i\}$ so that there exists some h with $0 \le h \le i$ satisfying the following conditions:

- For $1 \le j \le h$ we have either $\gamma'_j \ne \overline{\gamma}_j \in A_1$ or $\gamma'_j \ne \gamma_j^{-1} \in \Omega^* A$.
- For $h < j \le i$ we have either $\gamma'_j = \overline{\gamma}_j \in A_1$ or $\gamma'_j = \gamma_j^{-1} \in \Omega^* A$.

With this fixed reindexing we define

$$\{\gamma_1, \dots, \gamma_m\}\{\gamma'_1, \dots, \gamma'_n\} := \{\gamma'_1 * \gamma_1, \dots, \gamma'_h * \gamma_h, \gamma_{i+1}, \dots, \gamma_m, \gamma'_{i+1}, \dots, \gamma'_n\}.$$

(Note that we have omitted all $\gamma'_j * \gamma_j$ with $j > h$. The choice of $\gamma'_j * \gamma_j$ rather than $\gamma_j * \gamma'_j$ is just because we want to define a left action on (P, A).)

We have shown in [Kuessner 2010] (footnote to Section 1.5.1) that the product belongs to $\Pi(A)$. Moreover, the multiplication so defined is independent of the chosen reindexing. It is clearly associative. A neutral element is given by the empty set. The inverse to $\{\gamma_1, \dots, \gamma_n\}$ is given by $\{\gamma'_1, \dots, \gamma'_n\}$, with $\gamma'_i = \overline{\gamma}_i$ if $\gamma_i \in A_1$ and $\gamma'_i = \gamma_i^{-1}$ if $\gamma_i \in \Omega^* A$. (Indeed, in this case $h = 0$; thus $\{\gamma_1, \dots, \gamma_n\}\{\gamma'_1, \dots, \gamma'_n\}$ is the empty set.) Thus we have defined a group law on $\Pi(A)$.

Remark. There is an inclusion

$$\Pi(A) \subset \mathrm{map}_0\big(A_0, \big[[0,1], |A|\big]_{|P|}\big),$$

where $\big[[0,1], |A|\big]_{|P|}$ is the set of homotopy classes (in $|P|$) rel $\{0,1\}$ of maps from $[0,1]$ to $|A|$, and $\mathrm{map}_0\big(A_0, \big[[0,1], |A|\big]_{|P|}\big)$ is the set of maps $f : A_0 \to \big[[0,1], |A|\big]_{|P|}$ with

- $f(y)(0) = y$ for all $y \in A_0$, and
- $f(\cdot)(1) : A_0 \to A_0$ is a bijection.

This inclusion is given by sending $\{\gamma_1, \dots, \gamma_n\}$ to the map f defined by $f(\gamma_i(0)) = [\gamma_i]$ for $i = 1, \dots, n$, and $f(y) = [c_y]$ (the constant path) for $y \notin \{\gamma_1(0), \dots, \gamma_n(0)\}$. The inclusion is a homomorphism with respect to the group law defined by

$$[gf(y)] := [f(y)] * [g(f(y)(1))]$$

on $\mathrm{map}_0(A_0, \big[[0,1], |A|\big]_{|P|})$.

3E. ***Action of*** $\boldsymbol{\Pi(A)}$ ***on*** $\boldsymbol{P}$**.** From now on we assume that P *is aspherical.* We define an action of $\mathrm{map}_0\big(A_0, \big[[0,1], |A|\big]_{|P|}\big)$ on P. This gives an action of $\Pi(A)$ on P.

Let $g \in \mathrm{map}_0(A_0, [(0,1), |A|]_{|P|})$. Define $gy = g(y)(1)$ for $y \in A_0$ and $gx = x$ for $x \in P_0 - A_0$. This defines the action on the 0-skeleton of P.

We extend this to an action on the 1-skeleton of P. Recall that, by minimal completeness of P, 1-simplices σ are in one-to-one correspondence with homotopy classes (rel $\{0,1\}$) of (nonclosed) singular 1-simplices in $|P|$ with vertices in P_0. Using this correspondence, define

$$g\sigma := [\overline{g(\sigma(0))}] * [\sigma] * [g(\sigma(1))],$$

where $*$ denotes concatenation of (homotopy classes of) paths.

In [Kuessner 2010, Section 1.5.1] we proved that this defines an action on P_1 and that there is an extension of ths action to an action on P. (The extension is unique because P is aspherical.)

We remark, because this will be one of the assumptions to apply Lemma 3.7, that the action of any element $g \in \Pi(A)$ is homotopic to the identity. The homotopy between the action of the identity and the action of $\{\gamma_1, \dots, \gamma_r\}$ given by the action of $\{\gamma_1^t, \dots, \gamma_r^t\}$, $0 \le t \le 1$, with $\gamma_i^t(s) = \gamma_i(st)$.

The next lemma follows directly from the construction, but we will use it so often that we want to explicitly state it.

Lemma 3.3. *Let (P, A) be a pair of aspherical, minimally complete multicomplexes, with the action of $G = \Pi(A)$. If $\sigma \in P$ is a simplex all of whose vertices are not in A, then $g\sigma = \sigma$ for all $g \in G$.*

For a topological space and a subset $P \subset S_*(X)$ closed under face maps, the (antisymmetric) bounded cohomology $H_b^*(P)$ and its pseudonorm are defined literally like for multicomplexes in [Gromov 1982, Section 3.2]. The following well-known fact will be needed for applications of Lemma 3.7 (to the setting of Theorem 1.1) with $P = K^{\mathrm{str}}(\partial Q)$, $G = \Pi(K(\partial_0 Q))$.

Lemma 3.4. (a) *Let (P, A) be a pair of minimally complete multicomplexes. If each connected component of $|A|$ has amenable fundamental group, then $\Pi(A)$ is amenable.*

(b) *Let X be a topological space, $P \subset S_*(X)$ a subset closed under face maps, and G an amenable group acting on P. Then the canonical homomorphism*

$$\mathrm{id} \otimes 1 : C_*^{\mathrm{simp}}(P) \to C_*^{\mathrm{simp}}(P) \otimes_{\mathbb{Z}G} \mathbb{Z}$$

induces an isometric monomorphism in bounded cohomology.

The proof of (a) is an obvious adaptation of that of [Kuessner 2010, Lemma 4]. Part (b) is proved by averaging bounded cochains; see [Gromov 1982].

3F. ***Retraction to central simplices.***

Lemma 3.5. *Let $(N, \partial N)$ be a pair of topological spaces with $N = Q \cup R$ for two subspaces Q, R. Let*

$$\partial_0 Q = Q \cap R,\ \partial_1 Q = Q \cap \partial N,\ \partial_1 R = R \cap \partial N,\ \partial Q = \partial_0 Q \cup \partial_1 Q,\ \partial R = \partial_0 Q \cup \partial_1 R.$$

Assume that $\partial_1 Q \to Q \to N$, $\partial_1 R \to R \to N$, $\partial N \to N$, $\partial_0 Q \to Q$, $\partial_0 Q \to R$ are π_1-injective, and that ∂N, $\partial_1 Q$, $\partial_1 R$, $\partial_0 Q$ are aspherical (so the corresponding $K(\cdot)$ can be considered as submulticomplexes of $K(N)$).

In connection with the simplicial action of $G = \Pi(K(\partial_0 Q))$ on $K(N)$, there is a chain homomorphism

$$r : C_*^{\text{simp,inf}}(K(N)) \otimes_{\mathbb{Z}G} \mathbb{Z} \to C_*^{\text{simp,inf}}(K(Q)) \otimes_{\mathbb{Z}G} \mathbb{Z}$$

in degrees $* \geq 2$, mapping $C_*^{\text{simp,inf}}(GK(\partial N)) \otimes_{\mathbb{Z}G} \mathbb{Z}$ to $C_*^{\text{simp,inf}}(GK(\partial_1 Q)) \otimes_{\mathbb{Z}G} \mathbb{Z}$, and such that

- *if σ is a simplex in $K(N)$, then $r(\sigma \otimes 1) = \kappa \otimes 1$, where either κ is a simplex in $K(Q)$ or $\kappa = 0$;*
- *if σ is a simplex in $K(Q)$, then $r(\sigma \otimes 1) = \sigma \otimes 1$;*
- *if σ is a simplex in $K(R)$, then $r(\sigma \otimes 1) = 0$.*

Proof. This is [Kuessner 2010, Proposition 6]. (We have replaced the assumption $\ker(\pi_1 \partial_0 Q \to \pi_1 Q) = \ker(\pi_1 \partial_0 Q \to \pi_1 R)$ from that reference by the stronger assumption of π_1-injectivity, since this will be true in all our applications and we have no need for the more general assumption.) The conclusion is stated in [Kuessner 2010] for locally finite chains, but of course r extends linearly to infinite chains. □

Remark. If some edge of σ is contained in $K(\partial_0 Q) = K(Q) \cap K(R)$, then

$$\sigma \otimes 1 = 0 \in C_*^{\text{simp,inf}}(K(N)) \otimes_{\mathbb{Z}G} \mathbb{Z};$$

see [Kuessner 2010, Section 1.5.2]. (The proof is essentially the same as that of Lemma 5.17 below.) In particular, if σ is contained in both $K(Q)$ and $K(R)$, then $r(\sigma \otimes 1) = r(0) = 0$.

3G. ***Fundamental cycles in $K(N)$ and $K(Q)$.*** Let N be a (possibly noncompact) connected, orientable n-manifold with (possibly noncompact) boundary ∂N. Then $H_n^{\text{lf}}(N, \partial N) \simeq \mathbb{Z}$ by Whitehead's theorem and a generator is called $[N, \partial N]$. (It is only defined up to sign, but this will not concern our arguments.) Recall that an infinite chain is said to represent $[N, \partial N]$ if it is homologous to a locally finite chain representing $[N, \partial N]$.

If $\partial N \to N$ is π_1-injective and ∂N is aspherical, we know from Section 3B that

$$C_*^{\mathrm{simp,inf}}(\widehat{K}(N), \widehat{K}(\partial N)) \subset C_*^{\mathrm{sing,inf}}(N, \partial N).$$

Thus it makes sense to say that some chain $z \in C_*^{\mathrm{simp,inf}}(\widehat{K}(N), \widehat{K}(\partial N))$ represents the fundamental class $[N, \partial N]$.

If $\partial_1 Q \to Q$ is π_1-injective and Q and $\partial_1 Q$ are aspherical, and if we set $G := \Pi(K(\partial_0 Q))$, then $C_*^{\mathrm{simp,inf}}(GK(\partial_1 Q)) = C_*^{\mathrm{simp,inf}}(G\widehat{K}(\partial_1 Q)) \subset C_*^{\mathrm{sing,inf}}(\partial Q)$, as G maps simplices in $\mathrm{im}(K(\partial Q) \to K(Q))$ to simplices in $\mathrm{im}(K(\partial Q) \to K(Q))$. Thus it makes sense to say that some chain $z \in C_*^{\mathrm{simp,inf}}(K(Q), GK(\partial_1 Q))$ represents the fundamental class $[Q, \partial Q]$.

The projection $p : \widehat{K}(N) \to K(N)$ is defined at the end of Section 3B.

Lemma 3.6. *Let N be an orientable n-manifold with boundary* (*where $n \geq 2$*), *and let $Q, R \subset N$ be orientable n-manifolds with boundary such that $N = Q \cup R$ satisfies the assumptions of Lemma 3.5 and that $\partial_0 Q, \partial_1 Q \subset \partial Q$ and $\partial_1 R \subset \partial R$ are* $(n-1)$*-dimensional submanifolds* (*with boundary*) *of ∂Q or ∂R. Assume also that Q is aspherical. Let $\sum_i a_i \sigma_i \in C_n^{\mathrm{simp,inf}}(\widehat{K}(N), \widehat{K}(\partial N))$ represent $[N, \partial N]$.*

(a) $\sum_i a_i r(p(\sigma_i)) \otimes 1 \in C_n^{\mathrm{simp,inf}}(K(Q), GK(\partial_1 Q)) \otimes_{\mathbb{Z}G} \mathbb{Z}$ *represents* $[Q, \partial Q] \otimes 1$.

(b) $\partial \sum_i a_i r(p(\sigma_i)) \otimes 1 \in C_n^{\mathrm{simp,inf}}(GK(\partial Q)) \otimes_{\mathbb{Z}G} \mathbb{Z}$ *represents* $[\partial Q] \otimes 1$.

Remark. Explicitly, statement (a) means that the element on the left represents the image of $h \otimes 1$ under the canonical homomorphism $H_n^{\mathrm{sing,inf}}(Q, \partial Q) \otimes_{\mathbb{Z}G} \mathbb{Z} \to H_n(C_*^{\mathrm{sing,inf}}(Q, \partial Q) \otimes_{\mathbb{Z}G} \mathbb{Z})$, where $h \in H_n^{\mathrm{simp,inf}}(K(Q), GK(\partial_1 Q))$ represents $[Q, \partial Q] \in H_n^{\mathrm{sing}}(Q, \partial Q)$. Similarly, (b) means that the element represents the image of $h \otimes 1$ under the canonical homomorphism $H_n^{\mathrm{simp,inf}}(GK(\partial_1 Q)) \otimes_{\mathbb{Z}G} \mathbb{Z} \to H_n(C_*^{\mathrm{simp,inf}}(GK(\partial_1 Q)) \otimes_{\mathbb{Z}G} \mathbb{Z})$, where $h \in H_n^{\mathrm{simp,inf}}(GK(\partial_1 Q))$ represents $[\partial Q] \in H_n^{\mathrm{sing}}(\partial Q)$.

Proof. Since p and r are chain maps, it suffices to check the claim for some chosen representative of $[N, \partial N]$. So let $z \in C_*^{\mathrm{simp,inf}}(\widehat{K}(N), \widehat{K}(\partial N))$ be a representative of $[N, \partial N]$ chosen so that

$$p(z) = z_Q + z_R,$$

where z_Q represents $[Q, \partial Q]$ and z_R represents $[R, \partial R]$, and so that

$$\partial z_Q = w_1 + w_2, \, \partial z_R = -w_2 + w_3$$

with $w_1 \in C_{n-1}^{\mathrm{simp,inf}}(K(\partial_1 Q))$ representing $[\partial_1 Q]$, $w_2 \in C_{n-1}^{\mathrm{simp,inf}}(K(\partial_0 Q))$ representing $[\partial_0 Q]$, and $w_3 \in C_{n-1}^{\mathrm{simp,inf}}(K(\partial_1 R))$ representing $[\partial_1 R]$.

From Lemma 3.5 we have

$$r(p(z) \otimes 1) = z_Q \otimes 1,$$

which implies the first claim, and

$$\partial r(p(z)\otimes 1)=\partial z_Q\otimes 1=w_1\otimes 1+w_2\otimes 1.$$

Since w_1+w_2 represents $[\partial Q]$, this implies the second claim. □

Remark. From the remark after Lemma 3.5 we have $w_2\otimes 1=0$. This implies $\partial r(p(z)\otimes 1)=\partial z_Q\otimes 1=w_1\otimes 1$, that is, $\partial r(p(z)\otimes 1)$ represents at the same time $[\partial Q]\otimes 1$ and $[\partial_1 Q]\otimes 1$.

3H. ***Using amenability.*** The next lemma is well-known in slightly different formulations and we reprove it here only for completeness. (It has of course a relative version as well, but we will not need that for our argument.)

We will apply[2] this lemma in the proof of Theorem 1.1 with $X=\partial Q$, $G=q_*(\Pi(K(\partial_0 Q)))$ and $K=GK^{\mathrm{str}}(\partial_1 Q)$.

Lemma 3.7. *Let X be a closed, orientable manifold and $K\subset S_*(X)$ closed under face maps. Assume that*

- *there is an amenable group G acting on K, such that the action of each $g\in G$ on $|K|$ is homotopic to the identity, and*
- *there is a fundamental cycle $z\in C_*^{\mathrm{simp}}(K)$ such that $z\otimes 1$ is homologous to a cycle $h=\sum_{j=1}^{s} b_j\tau_j\otimes 1\in C_*^{\mathrm{simp}}(K)\otimes_{\mathbb{Z}G}\mathbb{Z}$.*

Then

$$\|X\|\le\sum_{j=1}^{s}|b_j|.$$

Proof. If $\|X\|=0$, there is nothing to prove. Thus we may assume $\|X\|\ne 0$, which implies [Gromov 1982, p. 17] that there is $\beta\in H_b^n(X)$, a bounded cohomology class dual to $[X]\in H_n(X)$, with $\|\beta\|=1/\|X\|$.

Let $p:C_*^{\mathrm{simp}}(K)\to C_*^{\mathrm{simp}}(K)\otimes_{\mathbb{Z}G}\mathbb{Z}$ be the homomorphism defined by $p(\sigma)=\sigma\otimes 1$. Since G is amenable we have, by the proof of [Gromov 1982, Lemma 4b], an "averaging homomorphism" $Av:H_b^*(K)\to H_b^*(C_*(K)\otimes_{\mathbb{Z}G}\mathbb{Z})$ such that Av is left-inverse to p^* and Av is an isometry. Hence

$$\|Av(\beta)\|=\|\beta\|=\frac{1}{\|X\|}.$$

[2]If a group G acts simplicially on a multicomplex M, then $C_*(M)\otimes_{\mathbb{Z}G}\mathbb{Z}$ are abelian groups with well-defined boundary operator $\partial_*\otimes 1$, even though M/G may not be a multicomplex, like for the action of $G=\Pi_X(X)$ on $K(X)$, for a topological space X.

We remark that $C_*(M)\otimes_{\mathbb{Z}G}\mathbb{Z}\simeq C_*(M)\otimes_{\mathbb{R}G}\mathbb{R}$ is just the quotient chain complex for the G-action. In particular, even though $C_*(M)$ is an $\mathbb{R}G$-module, it does not make any difference whether we tensor over $\mathbb{Z}G$ or $\mathbb{R}G$.

Moreover, denoting by $\left[\sum_{j=1}^{s} b_j\tau_j\otimes 1\right]$ the homology class of $\sum_{j=1}^{s} b_j\tau_j\otimes 1$, we have obviously

$$\left|Av(\beta)\left[\sum_{j=1}^{s} b_j\tau_j\otimes 1\right]\right|\leq \|Av(\beta)\|\sum_{j=1}^{s}|b_j|$$

and therefore

$$\|X\|=\frac{1}{\|Av(\beta)\|}\leq\frac{\sum_{j=1}^{s}|b_j|}{\left|Av(\beta)\left[\sum_{j=1}^{s} b_j\tau_j\otimes 1\right]\right|}.$$

It remains to prove that $Av(\beta)\left[\sum_{j=1}^{s} b_j\tau_j\otimes 1\right]=1$. For this we have to look at the definition of Av, which is as follows:

Let $\gamma\in C_b^*(K)$ be a bounded cochain. By amenability there exists a bi-invariant mean $av: B(G)\to\mathbb{R}$ on the bounded functions on G with $\inf_{g\in G}\delta(g)\leq av(\delta)\leq\sup_{g\in G}\delta(g)$ for all $\delta\in B(G)$. Then, given any $p(\sigma)\in C_*(K)\otimes_{\mathbb{Z}G}\mathbb{Z}$, one can fix an identification between G and $G\sigma$, the set of all σ' with $p(\sigma')=p(\sigma)$, and thus consider the restriction of γ to $G\sigma$ as a bounded cochain on G. Define $Av(\gamma)(p(\sigma))$ to be the average av of this bounded cochain on $G\simeq G\sigma$. (This definition is independent of all choices; see [Ivanov 1985].)

Now, if $z=\sum_{j=1}^{s} b_j\tau_j$ is a fundamental cycle, we have $\beta(z)=1$.

If $g\in G$ is arbitrary, then left multiplication with g is a chain map on $C_*^{\text{simp}}(K)$, as well as on $C_*^{\text{sing}}(X)$. Since the action of g on $|K|$ is homotopic to the identity, it induces the identity on the image of $C_*^{\text{simp}}(K)\to C_*^{\text{sing}}(X)$. Thus, for each cycle $z\in C_*^{\text{simp}}(K)$ representing $[X]\in H_*^{\text{sing}}(X)$, the cycle $gz\in C_*^{\text{simp}}(K)$ must also represent $[X]$.

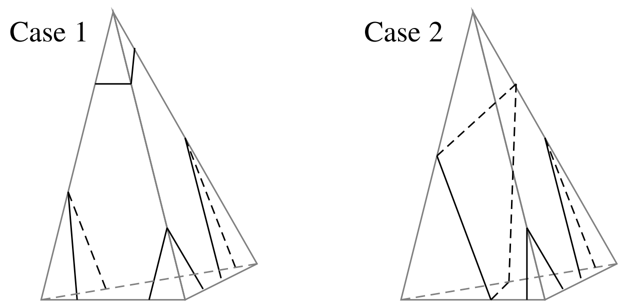
If gz represents $[X]$, then $\beta(gz)=\beta([X])=1$. In conclusion, $\beta(p(z'))=1$ for each z' with $p(z')=p(z)$. By the definition of Av, this implies $Av(\beta)(p(z))=1$ for each fundamental cycle z. In particular, $Av(\beta)\left[\sum_{j=1}^{s} b_j\tau_j\otimes 1\right]=1$, which finishes the proof of the lemma. □

Remark. In the proof of Theorem 1.1, we will work with $C_*^{\text{simp}}(K)\otimes_{\mathbb{Z}G}\mathbb{Z}$ rather than $C_*^{\text{simp}}(K)$. This is analogous to Agol's construction of "crushing the cusps to points" in [Agol 1999]. However $C_*^{\text{simp}}(K(Q))\otimes_{\mathbb{Z}\Pi(\partial_0 Q)}\mathbb{Z}\neq C_*^{\text{simp}}(K(Q/\partial_0 Q))$; thus one cannot simplify our arguments by working directly with $Q/\partial_0 Q$.

4. Disjoint planes in a simplex

In this section, we will discuss the possibilities for how a simplex can be cut by planes without producing parallel arcs in the boundary. (More precisely, we pose the additional condition that the components of the complement can be colored by black and white such that all vertices belong to black components, and we actually want to avoid only parallel arcs in the boundary of white components.)

For example, for the 3-simplex, it will follow that there is essentially only the possibility in Case 1 pictured below; meanwhile, in Case 2, each triangle has a parallel arc with another triangle, regardless how the quadrangle is triangulated.



Let $\Delta^n \subset \mathbb{R}^{n+1}$ be the standard simplex[3] with vertices $v_0, \ldots, v_n$. It is contained in the plane $E = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : x_1 + \cdots + x_{n+1} = 1\}$.

In this section we will be interested in $(n-1)$-dimensional affine planes $P \subset E$ whose intersection with Δ^n either contains no vertex, consists of exactly one vertex, or consists of a face of Δ^n. For such planes we define their type as follows.

Definition 4.1. Let $P \subset E$ be an $(n-1)$-dimensional affine plane such that $P \cap \Delta^n$ contains no vertex, consists of exactly one vertex, or consists of a face of Δ^n.

- If $P \cap \Delta^n = \partial_0 \Delta^n$, we say that P is of type $\{0\}$.
- If $P \cap \Delta^n = \partial_j \Delta^n$ with $j \geq 1$, we say that P is of type $\{01 \ldots \hat{j} \ldots n\}$.
- If $P \cap \{v_0, \ldots, v_n\} = \{v_0\}$, we say that P is of type $\{0\}$.
- If $P \cap \{v_0, \ldots, v_n\} = \varnothing$ or $P \cap \{v_0, \ldots, v_n\} = \{v_j\}$ with $j \geq 1$, we say that P is of type $\{0a_1 \ldots a_k\}$ with $a_1, \ldots, a_k \in \{1, \ldots, n\}$ if the following condition is satisfied: v_i belongs to the same connected component of $\Delta^n - (P \cap \Delta^n)$ as v_0 if and only if $i \in \{a_1, \ldots, a_k\}$.

Observation 4.2. *Let P_1 be a plane of type $\{0a_1 \ldots a_k\}$ and P_2 a plane of type $\{0b_1 \ldots b_l\}$. Assume that $Q_1 := P_1 \cap \Delta^n \neq \varnothing$ and $Q_2 := P_2 \cap \Delta^n \neq \varnothing$. Then $Q_1 \cap Q_2 = \varnothing$ implies that either $\{a_1, \ldots, a_k\} = \{b_1, \ldots, b_l\}$ or exactly one of the following conditions holds*:

- $\{a_1, \ldots, a_k\} \subset \{b_1, \ldots, b_l\}$.
- $\{b_1, \ldots, b_l\} \subset \{a_1, \ldots, a_k\}$.
- $\{a_1, \ldots, a_k\} \cup \{b_1, \ldots, b_l\} = \{1, \ldots, n\}$.

[3]As usual, v_i is the vertex with all coordinates except the i-th equal to zero, and $\partial_i \Delta^n$ denotes the subsimplex spanned by all vertices except v_i. We will occasionally identify singular 1-simplices $\sigma : \Delta^1 \to M$ with paths $e : [0, 1] \to M$ by the rule $e(t) = \sigma(t, 1-t)$. In particular, $e(0) = \sigma(v_0) = \partial_1 \sigma$ and $e(1) = \sigma(v_1) = \partial_0 \sigma$.

Proof. $\Delta^n - Q_1$ consists of two connected components, C_1 and C_2. Similarly, $\Delta^n - Q_2$ consists of two connected components, D_1 and D_2. Choose the numbering so $v_0 \in C_1$ and $v_0 \in C_2$. In particular, $C_1 \cap D_1 \neq \varnothing$.

Since $Q_1 \cap Q_2 = \varnothing$, it follows that Q_2 is contained in one of C_1 or C_2, and Q_1 is contained in one of D_1 or D_2.

If $Q_1 \subset D_1$, either we have $C_1 \subset D_1$, which implies $\{a_1, \ldots, a_k\} \subset \{b_1, \ldots, b_l\}$, or we have $C_2 \subset D_1$, which implies $\{1, \ldots, n\} - \{a_1, \ldots, a_k\} \subset \{b_1, \ldots, b_l\}$, hence $\{a_1, \ldots, a_k\} \cup \{b_1, \ldots, b_l\} = \{1, \ldots, n\}$.

If instead $Q_1 \subset D_2$, we have $Q_2 \subset C_1$. After interchanging Q_1 and Q_2 we are back in the case of the previous paragraph. □

Notational remark. *Arc* will mean the intersection of an $(n-1)$-dimensional affine plane $P \subset E$ (such that $P \cap \Delta^n \neq \varnothing$ either contains no vertex, consists of exactly one vertex or consists of a face) with a 2-dimensional subsimplex $\tau^2 \subset \Delta^n$. If an arc consists of only one vertex, we call it a *degenerate arc*.

Definition 4.3 (parallel arcs). Let $P_1, P_2 \subset E$ be $(n-1)$-dimensional affine planes. Let τ be a 2-dimensional subsimplex of Δ^n with vertices v_r, v_s, v_t. We say that the disjoint arcs e_1, e_2 obtained as intersections of P_1 and P_2, respectively, with τ are parallel arcs if one of the following conditions holds:

- Both are nondegenerate and any two of $\{v_r, v_s, v_t\}$ belong to the same connected component of $\tau - e_1$ if and only if they belong to the same connected component of $\tau - e_2$.
- One, say e_1, is nondegenerate, the other, say with vertices v_s, v_t, is contained in a face, and v_r does not belong to the same connected component of $\tau - e_1$ as either v_s and v_t.
- One, say e_1, is nondegenerate, the other is degenerate, say equal to v_r, and v_s, v_t do not belong to the same connected component of $\tau - e_1$ as v_r.
- Both are degenerate and equal.
- Both are contained in a face and equal.
- One is degenerate, the other is contained in a face.

Lemma 4.4. *Let $\Delta^n \subset \mathbb{R}^{n+1}$ be the standard simplex. Let $P_1, P_2 \subset E$ be $(n-1)$-dimensional affine planes with $Q_i = P_i \cap \Delta^n \neq \varnothing$ for $i = 1, 2$. Let P_1 be of type $\{0a_1 \ldots a_k\}$ with $1 \leq k \leq n-2$ and P_2 of type $\{0b_1 \ldots b_l\}$ with l arbitrary. Then either $Q_1 \cap Q_2 \neq \varnothing$, or Q_1 and Q_2 have a parallel arc.*

Proof. Assume that $Q_1 \cap Q_2 = \varnothing$. By Observation 4.2, there are four possible cases:

- $\{0a_1 \ldots a_k\} = \{0b_1 \ldots b_l\}$. Then we clearly have parallel arcs.

– $\{0a_1 \dots a_k\}$ is a proper subset of $\{0b_1 \dots b_l\}$, i.e., $1 \le k < l \le n-1$ and $a_1 = b_1, \dots, a_k = b_k$. There is at least one index, say i, not contained in $\{0b_1 \dots b_l\}$. The 2-dimensional subsimplex with vertices v_0, v_{a_1}, v_i intersects P_1 and P_2 in parallel arcs, because P_1 and P_2 both separate v_0 and v_{a_k} from v_i.

– $\{0b_1 \dots b_l\}$ is a proper subset of $\{0a_1 \dots a_k\}$, i.e., $0 \le l < k \le n-2$ and $a_1 = b_1, \dots, a_l = b_l$. There are two indices i, j not contained in $\{0a_1 \dots a_k\}$. The 2-dimensional subsimplex with vertices v_0, v_i, v_j intersects P_1 and P_2 in parallel arcs, because P_1 and P_2 both separate v_0 from v_i and v_j.

– $\{a_1, \dots, a_k\} \cup \{b_1, \dots, b_l\} = \{1, \dots, n\}$. Since $k \le n-2$, there are two indices i, j not contained in $\in \{0a_1 \dots a_k\}$. Hence $i, j \in \{b_1, \dots, b_l\}$. There exists an index $h \in \{a_1, \dots, a_k\}$ such that $h \notin \{b_1, \dots, b_l\}$; otherwise, we would have $\{a_1, \dots, a_k\} \subset \{b_1, \dots, b_l\}$, hence $\{1, \dots, n\} = \{a_1, \dots, a_k\} \cup \{b_1, \dots, b_l\} \subset \{b_1, \dots, b_l\}$, contradicting $Q_2 \neq \varnothing$. Now the 2-dimensional subsimplex with vertices v_i, v_j, v_h intersects P_1 and P_2 in parallel arcs, because both P_1 and P_2 separate v_i and v_j from v_h. □

Definition 4.5 (canonical coloring of complementary regions). Let $P_1, P_2, \dots \subset E$ be a (possibly infinite) set of $(n-1)$-dimensional affine planes with $Q_i := P_i \cap \Delta^n$ nonempty and $Q_i \cap Q_j = \varnothing$ for all $i \neq j$. Assume that each Q_i either contains no vertices or consists of exactly one vertex.

A coloring of the connected components of $\Delta^n - \bigcup_i Q_i$ by the colors black and white, and of all the Q_i by black, is called a *canonical coloring* (associated to $P_1, P_2, \dots$) if all the vertices of Δ^n are colored black and each Q_i is incident to at least one white component.

Definition 4.6 (white-parallel arcs). Let $\{P_i : i \in I\}$ be a set of $(n-1)$-dimensional affine planes $P_i \subset E$, with $Q_i := P_i \cap \Delta^n \neq \varnothing$ for $i \in I$. Assume that $Q_i \cap Q_j = \varnothing$ for all $i \neq j \in I$, and that we have a canonical coloring associated to $\{P_i : i \in I\}$. We say that arcs e_i, e_j obtained as intersections of P_i, P_j $(i, j \in I)$ with some 2-dimensional subsimplex of Δ^n are white-parallel arcs if they are parallel arcs and belong to the boundary of the closure of the same white component.

We mention two consequences of Lemma 4.4. They will not be needed for the proof of Lemma 4.13, but they will be necessary for the proof of Theorem 1.1.

Corollary 4.7. *Let $\Delta^n \subset \mathbb{R}^{n+1}$ be the standard simplex. Let $P_1, \dots, P_m \subset E$ be a finite set of $(n-1)$-dimensional affine planes and let $Q_i = P_i \cap \Delta^n$ for $i = 1, \dots, m$. Assume that $Q_i \cap Q_j = \varnothing$ for all $i \neq j$, and that we have an associated canonical coloring such that Q_i and Q_j do not have a white-parallel arc for $i \neq j$.*

Then, unless $m = 0$, we have $m = n+1$ and P_1 is of type $\{0\}$, P_{n+1} is of type $\{0\,1 \dots n-1\}$, and P_i is of type $\{01 \dots \widehat{i-1} \dots n\}$ for $i = 2, \dots, n$.

Proof. If the conclusion were not true, there would exist a plane P_1 of type $\{0a_1 \dots a_k\}$ with $1 \le k \le n-2$. Let W be the white component of the canonical coloring that is incident to P_1. Because, for a canonical coloring, no vertex belongs to a white component, there must be at least one more plane P_2 incident to W. Since $Q_1 \cap Q_2 = \emptyset$, from Lemma 4.4 we get that Q_1 and Q_2 have a parallel arc. Because Q_1 and Q_2 are incident to W, the arc is white-parallel. □

Corollary 4.8. *Let $\Delta^n \subset \mathbb{R}^{n+1}$ be the standard simplex. Let $P_1, P_2, \dots \subset E$ be a (possibly infinite) set of $(n-1)$-dimensional affine planes and let $Q_i = P_i \cap \Delta^n$ for $i = 1, 2, \dots$. Assume that we have an associated canonical coloring.*

Let P_i be of type $\{0a_1^i \dots a_{c(i)}^i\}$, for $i = 1, 2, \dots$. Then either

- *$c(1) \in \{0, n-1\}$, or*
- *whenever, for some $i \in \{2, 3, \dots\}$, P_1 and P_i bound a white component of $\Delta^n - \cup_j Q_j$, then they must have a white-parallel arc.*

Proof. Assume that $c(1) \notin \{0, n-1\}$. The white component W bounded by P_1 is bounded by a finite number of planes; thus we can apply Corollary 4.7, and conclude that P_1 has a white-parallel arc with each other plane adjacent to W. □

Definition 4.9. Let $P \subset E$ be an $(n-1)$-dimensional affine plane and T a triangulation of the polytope $Q := P \cap \Delta^n$. We say that T is *minimal* if all vertices of T are vertices of Q. We say that an edge of some simplex in T is an exterior edge if it is an edge of Q.

Observation 4.10. *Let $P \subset E$ be an $(n-1)$-dimensional affine plane and T a triangulation of the polytope $Q := P \cap \Delta^n$. If T is minimal, each edge of Q is an (exterior) edge of (exactly one) simplex in T.*

Proof. By minimality, the triangulation does not introduce new vertices. Thus every edge of Q is an edge of some simplex. □

Observation 4.11. *Let $P \subset E$ be an $(n-1)$-dimensional affine plane with $Q := P \cap \Delta^n \neq \emptyset$. Assume that P is of type $\{0a_1 \dots a_k\}$. Then either*

(a) *Each vertex of Q arises as the intersection of P with an edge e of Δ^n. The vertices of e are v_i and v_j with $i \in \{0, a_1, \dots, a_k\}$ and $j \notin \{0, a_1, \dots, a_k\}$. (We will denote such a vertex by $(v_i v_j)$.)*

(b) *Two vertices $(v_{i_1} v_{j_1})$ and $(v_{i_2} v_{j_2})$ of Q are connected by an edge of Q (i.e., an exterior edge of any triangulation) if either $i_1 = i_2$ or $j_1 = j_2$.*

Proof. (a) holds because e has to connect vertices in distinct components of $\Delta^n - Q$. Statement (b) holds because the edge of Q has to belong to some 2-dimensional subsimplex of Δ^n, with vertices either $v_{i_1}, v_{j_1}, v_{j_2}$ or $v_{i_1}, v_{i_2}, v_{j_1}$. □

Remark. If, for an affine hyperplane $P \subset E$, $Q = P \cap \Delta^n$ consists of exactly one vertex, then we will consider the minimal triangulation of Q to consist of one (degenerate) $(n-1)$-simplex. This convention helps to avoid needless case distinctions.

Lemma 4.12. *Let $\{P_i \subset E : i \in I\}$ be a set of $(n-1)$-dimensional affine planes and let $Q_i := P_i \cap \Delta^n$ for $i \in I$. Assume that $Q_i \cap Q_j = \varnothing$ for all $i \neq j$ and that we have an associated canonical coloring. Assume that we have fixed, for each $i \in I$, a minimal triangulation $Q_i = \bigcup_a \tau_{ia}$ of Q_i.*

If P_1 is of type $\{0a_1^1 \dots a_{c(1)}^1\}$ with $1 \leq c(1) \leq n-2$, then for each simplex $\tau_{1a} \subset Q_1$ there exists some $j \in I$ and some simplex $\tau_{jb} \subset Q_j$ (of the fixed triangulation of Q_j) such that τ_{ia} and τ_{jb} have a white-parallel arc.

Proof. Let $w_1, \dots, w_n$ be the n vertices of the $(n-1)$-simplex τ_{1k}. By Observation 4.11(a), each w_l arises as intersection of Q_1 with some edge $(v_{r_l} v_{s_l})$ of Δ^n, and the vertices v_{r_l}, v_{s_l} satisfy $r_l \in \{0, a_1^1, \dots, a_{c(1)}^1\}$ and $s_l \notin \{0, a_1^1, \dots, a_{c(1)}^1\}$.

For the canonical coloring, there must be a white component W bounded by P_1. We distinguish the cases whether W and v_0 belong to the same connected component of $\Delta^n - Q_1$ or not.

Case 1: W and v_0 belong to the same connected component of $\Delta^n - Q_1$.

Since $c(1) \leq n-2$, there exist at most $n-1$ possible values for r_l. Hence there exists $l \neq m \in \{1, \dots, n\}$ such that $v_{r_l} = v_{r_m}$.

Let e be the edge of $\tau_{1k} \subset Q_1$ connecting w_l and w_m. By Observation 4.11(b), e is an exterior edge. Consider the 2-dimensional subsimplex $\tau^2 \subset \Delta^n$ with vertices $v_{r_l}, v_{s_l}, v_{s_m}$. We conclude that P_1 intersects τ^2 in e, i.e., in an arc separating v_{r_l} from the other two vertices of τ^2.

Note that $r_l \in \{0, a_1^1, \dots, a_{c(1)}^1\}$; hence v_{r_l} belongs to the same component of $\Delta^n - Q_1$ as v_0. In particular, v_{r_l} belongs to the same component of $\Delta^n - Q_1$ as W. On the other hand, since the coloring is canonical, all vertices are colored black, and v_{r_l} cannot belong to the white component W. Thus there must be some plane P_j such that Q_j bounds W and separates v_{r_l} from Q_1. (The possibility that $P_j \cap \Delta^n = \{v_{r_l}\}$ is allowed.) In particular, some (possibly degenerate) exterior edge f of Q_j separates v_{r_l} from v_{s_l}, v_{s_m}. Thus e and f are white-parallel arcs. By Observation 4.10, f is an edge of some τ_{jl}.

Case 2: W and v_0 don't belong to the same connected component of $\Delta^n - Q_1$.

Since $n - c(1) \leq n-1$, there exist some $l \neq m \in \{1, \dots, n\}$ such that $v_{s_l} = v_{s_m}$.

Let e be the edge of $\tau_{1k} \subset Q_1$ connecting w_l and w_m. e is an exterior edge by Observation 4.11(b). Consider the 2-dimensional subsimplex $\tau^2 \subset \Delta^n$ with vertices $v_{r_l}, v_{r_m}, v_{s_l}$. P_1 intersects τ^2 in e, i.e., in an arc separating v_{s_l} from the other two vertices of τ^2.

We have $s_l \notin \{0, a_1^1, \dots, a_{c(1)}^1\}$; hence v_{s_l} does not belong to the same component of $\Delta^n - Q_1$ as v_0. This implies that v_{s_l} belongs to the same component of $\Delta^n - Q_1$ as W. On the other hand, since the coloring is canonical, v_{s_l} cannot belong to the white component W and there must be some plane P_j such that Q_j bounds W and separates v_{s_l} from Q_1. In particular, some exterior edge f of Q_j separates v_{s_l} from v_{r_l}, v_{r_m}. Thus e and f are white-parallel arcs. By Observation 4.10, f is an edge of some τ_{jl}. □

Lemma 4.13. *Let $\{P_i : i \in I\}$ be a set of $(n-1)$-dimensional affine planes with $Q_i := P_i \cap \Delta^n \neq \varnothing$ for $i \in I$. Let P_i be of type $\{0a_1^{(i)} \dots a_{k_i}^{(i)}\}$ for $i \in I$. Assume that $Q_i \cap Q_j = \varnothing$ for $i \neq j \in I$, and that we have an associated canonical coloring. Assume that for each Q_i one has fixed a minimal triangulation $Q_i = \cup_{k=1}^{t(i)} \tau_{ik}$.*

For each $i \in I$, let

$$D_i = \sharp\{\tau_{ik} \subset Q_i : \text{there is no } \tau_{jl} \subset Q_j \text{ such that } \tau_{ik}, \tau_{jl} \text{ have a white-parallel arc}\}.$$

Then

$$\sum_{i \in I} D_i = 0 \quad \text{or} \quad \sum_{i \in I} D_i = n+1.$$

Proof. First we remark that the number of planes may be infinite, but we may of course remove pairs of planes P_i, P_j whenever they are of the same type and bound the same white component. This removal of P_i, P_j and the common white component does not affect $\sum_{i \in I} D_i$. Since there are only finitely many different types of planes, we may without loss assume that we start with a finite number $P_1, \dots, P_m$ of planes. (It may happen that after this removal no planes and no white components remain. In this case $\sum_{i \in I} D_{i \in I} = 0$.) So we assume now that we have a finite number of planes $P_1, \dots, P_m$, and no two planes of the same type bound a white region.

The first case to consider is that all planes are of type $\{0a_1 \dots a_k\}$ with $k = 0$ or $k = n-1$. Since all vertices are colored black, this means that $m = n+1$ and (upon renumbering) P_1 is of type $\{0\}$, P_{n+1} is of type $\{0\ 1 \dots n-1\}$, and P_i is of type $\{01 \dots \widehat{i-1} \dots n\}$ for $i = 2, \dots, n$. Hence $D_1 = \cdots = D_{n+1} = 1$ and $\sum_{i=1}^{n+1} D_i = n+1$.

Now we assume that there exists P_i, say P_1, of type $\{0a_1^{(1)} \dots a_{k_1}^{(1)}\}$ with $1 \leq c(1) \leq n-2$. Let W be the white component bounded by P_1 and, without loss of generality, let $P_2, \dots, P_l$ be the other planes bounding W. Then Lemma 4.12 says that each simplex in the chosen triangulation of Q_1 has a parallel arc with some simplex in the chosen triangulation of each of $Q_2, \dots, Q_l$. In particular, $D_1 = 0$. For $j \in \{2, \dots, l\}$, if $1 \leq c(j) \leq n-2$, the same argument shows that $D_j = 0$. If $j \in \{2, \dots, l\}$ and $c(j) = 0$ or $c(j) = n-1$, then Q_j consists of only one simplex. By Corollary 4.8, this simplex has a parallel arc with (some exterior edge of) Q_1 and thus (by Observation 4.10) with some simplex of the chosen triangulation of Q_1. This shows that $D_j = 0$ also in this case. Altogether we conclude that $\sum_{j=1}^{l} D_j = 0$ and thus $\sum_{i=1}^{m} D_i = \sum_{i=l+1}^{m} D_i$. Hence we can remove[4] the white component W and its bounding planes $P_1, \ldots, P_l$ to obtain a smaller number of planes and a new canonical coloring without changing $\sum_{i=1}^{m} D_i$. Since we start with finitely many planes, we can repeat this reduction finitely many times and will end up either with an empty set of planes or with a set of planes of type $\{0a_1 \ldots a_k\}$, with $k = 0$ or $k = n-1$. Thus either $\sum_{i=1}^{m} D_i = 0$ or $\sum_{i=1}^{m} D_i = n+1$. □

We have thus proved that, in the presence of a canonical coloring, the number of $(n-1)$-simplices without white-parallel arcs in a minimal triangulation of the Q_i is 0 or $n+1$. We remark that in the proof of Theorem 1.1 we will actually count only those triangles that have neither a white-parallel arc nor a degenerate arc. Thus, in general, we may remain with even less than $n+1$ $(n-1)$-simplices.

5. A straightening procedure

In this section we will always work with the following set of assumptions.

Assumption I. *Q is an aspherical n-dimensional manifold with aspherical boundary ∂Q. We have $(n-1)$-dimensional submanifolds $\partial_0 Q, \partial_1 Q \subset \partial Q$ such that $\partial Q = \partial_0 Q \cup \partial_1 Q$, $\partial\partial_0 Q = \partial\partial_1 Q$, and $\partial_1 Q \neq \varnothing$ is aspherical.*

The example that one should have in mind is a nonpositively curved manifold Q with totally geodesic boundary $\partial_1 Q$ and cusps corresponding to $\partial_0 Q$.

In the case of nonpositively curved manifolds with totally geodesic boundary, there is a well-known straightening procedure that homotopes each relative cycle into a straight relative cycle. It is explained for closed hyperbolic manifolds in [Benedetti and Petronio 1992, Lemma C.4.3].

However, we will need a more subtle straightening procedure, which considers relative cycles with a certain 0-1 labeling of their edges and straightens the 1-labeled edges into certain distinguished 1-simplices. This straightening procedure will be explained in Section 5C. Before that, we explain a construction which will "morally" (although not literally) reduce the proof of Theorem 1.1 to the case that $\partial_0 Q \cap C$ is path-connected, for each path component C of ∂Q.

5A. *Making $\partial_0 Q \cap C$ connected.*

Construction 5.1. Let Assumption I be satisfied. There exists a continuous map of triples $q : (Q, \partial Q, \partial_1 Q) \to (Q, \partial Q, \partial_1 Q)$ that is (as a map of triples) homotopic to the identity and such that, for each path component C of ∂Q, the image $A := q(\partial_0 Q \cap C)$ is path-connected.

[4] To remove a white component means that this component together with the neighboring black components will form *one* new black component.

Moreover, for each path component F of $\partial_1 Q$, the path components of $\partial F \subset \partial_0 Q \cap \partial_1 Q$ can be numbered by $E_0^F, \ldots, E_s^F$ and one can choose points $x_{E_i^F} \in E_i^F$ such that $q(x_{E_i^F}) \equiv x_{E_0^F}$ for $i = 0, \ldots, s$.

Proof. For each path component F of $\partial_1 Q$, number the path components of $\partial F \subset \partial_0 Q \cap \partial_1 Q$ by $E_0^F, \ldots, E_s^F$, where s depends on F. Choose one point $x_E^F \in E$ for each path component $E \subset F$ of $\partial_0 Q \cap \partial_1 Q$. Whenever E_0, E_i is a pair of path components of $\partial_0 Q \cap \partial_1 Q$ adjacent to the same path component F of $\partial_1 Q$, choose a 1-dimensional submanifold $l_{E_0^F E_i^F} \subset \partial_1 Q$ with

$$\partial l_{E_0^F E_i^F} = \{x_{E_0^F}\} \cup \{x_{E_i^F}\}.$$

The $l_{E_0^F E_i^F}$ may be chosen succesively in such a way that they are disjoint from each other (apart from the common vertex $x_{E_0^F}$) and disjoint from $\partial_0 Q$ (apart from the vertices $x_{E_0^F}$ and $x_{E_i^F}$).

For each pair $\{E_0^F, E_i^F\}$ let $h : l_{E_0^F E_i^F} \to \{x_{E_0^F}\}$ be the constant map from $l_{E_0^F E_i^F}$ to $x_{E_0^F}$. For each path component F of $\partial_1 Q$, the union

$$\bigcup_{i=1}^{s} l_{E_0^F E_i^F}$$

is an embedded wedge of arcs in $\partial_1 Q$; hence it is contractible. In particular, h is homotopic to the identity. By the homotopy extension property there exists $g : F \to F$ with

$$g|_{l_{E_0^F E_i^F}} = h \equiv x_{E_0}$$

for all $l_{E_0^F E_i^F}$, and $g \sim \mathrm{id}$ by a homotopy extending the homotopy between h and id.

Thus we defined g on each path component F of $\partial_1 Q$ with $F \cap \partial_0 Q \neq \varnothing$. On path components F of $\partial_1 Q$ with $F \cap \partial_0 Q = \varnothing$ we define $g = \mathrm{id}$. Hence we have defined g on all of $\partial_1 Q$.

On path components C of $\partial_0 Q$ with $C \cap \partial_1 Q = \varnothing$, we define $f = \mathrm{id}$. Again by the homotopy extension property there exists $f : \partial Q \to \partial Q$ with $f|_{\partial_1 Q} = g$, $f|_C = \mathrm{id}$ for path components C of $\partial_0 Q$ with $C \cap \partial_1 Q = \varnothing$, and $f \sim \mathrm{id}$ by a homotopy extending the homotopy of g. (Of course, f does not preserve the path components of $\partial_0 Q$ that intersect $\partial_1 Q$.)

Once again by the homotopy extension property there exists $q : Q \to Q$ with $q \sim \mathrm{id}$ such that q extends f and the homotopy between q and id extends the homotopy between f and id.

Due to the stepwise construction, q is a map of triples, homotopic to the identity by a homotopy of triples. Moreover, $A := q(\partial_0 Q \cap C)$ is path-connected for each component C of ∂Q. Indeed, any two points in $\partial_0 Q \cap C$ can be connected by a sequence of paths which either have image in $\partial_0 Q$ or belong to $\bigcup_{i=1}^{s} l_{E_0^F E_i^F}$ for some path component F of $\partial_1 Q \cap C$. The image of these paths under q, in both cases, is in A. $\square$

Remark. The map q induces a simplicial map $q : K(Q) \to K(Q)$ and a homomorphism $q_* : \Pi(K(\partial_0 Q)) \to \Pi(K(A))$ defined by

$$q_*(\{\gamma_1, \dots, \gamma_n\}) := \{q(\gamma_1), \dots, q(\gamma_n)\},$$

such that $q_*(g)q(\sigma) = q(g\sigma)$ for each $\sigma \in K(Q)$ and $g \in \Pi(K(\partial_0 Q))$.

Proof. Continuous maps $q : Q \to Q$ induce simplicial maps $q : K(Q) \to K(Q)$. (The simplicial map agrees with q on the 0-skeleton, and it maps each 1-simplex $e \in K_1(Q)$ to the unique 1-simplex of $K_1(Q)$ that is in the homotopy class rel $\{0, 1\}$ of $q(e)$.)

Let $e \in K_1(Q)$. By construction, $\{\gamma_1, \dots, \gamma_n\}e = [\alpha * e * \bar{\beta}]$ for some $\alpha, \beta \in \{\gamma_1, \dots, \gamma_n\} \cup \{c_{e(0)}, c_{e(1)}\}$. Thus

$$\{q(\gamma_1), \dots, q(\gamma_n)\}q(e) = [q(\alpha) * q(e) * q(\bar{\beta})] = q(\{\gamma_1, \dots, \gamma_n\}e).$$

This implies the claim for the 1-skeleton, and thus, by the asphericity of $K(Q)$, for all $\sigma \in K(Q)$. $\square$

5B. *Definition of $K^{\mathrm{str}}(Q)$*. Let $Q, \partial Q, \partial_1 Q, \partial_0 Q$ satisfy Assumption I.

Recall that we have defined in Section 3B an aspherical multicomplex $K(Q) \subset S_*(Q)$ with the property that (for aspherical Q) each singular simplex in Q, with boundary in $K(Q)$ and pairwise distinct vertices, is homotopic rel boundary to a unique simplex in $K(Q)$.

The aim of this subsection is to describe a selection procedure yielding a subset $K_*^{\mathrm{str}}(Q) \subset S_*(Q)$. The final purpose of the straightening procedure will be to produce a large number of (weakly) degenerate simplices, in the sense of the following definition.

Definition 5.2. Let Q be an compact manifold with boundary ∂Q. We say that a simplex in $S_*(Q)$ is *degenerate* if one of its edges is a constant loop. We say that it is *weakly degenerate* if it is degenerate or its image is contained in ∂Q.

Notational remark. For subsets $K_*^{\mathrm{str}}(Q) \subset S_*(Q)$ we define

$$K_*^{\mathrm{str}}(\partial_0 Q) := K_*^{\mathrm{str}}(Q) \cap S_*(\partial_0 Q),$$
$$K_*^{\mathrm{str}}(\partial_1 Q) := K_*^{\mathrm{str}}(Q) \cap S_*(\partial_1 Q),$$
$$K_*^{\mathrm{str}}(\partial_0 Q Q) := K_*^{\mathrm{str}}(Q) \cap S_*(\partial_0 Q).$$

Lemma 5.3. *Let $Q, \partial Q, \partial_1 Q, \partial_0 Q$ satisfy Assumption I. Let $K(Q) \subset S_*(Q)$ be as defined in Section 3B. Let $q : Q \to Q$ and $\{x_{E_i^F} \in \partial_0 Q \cap \partial_1 Q : 0 \le i \le s\}$ be given by Construction 5.1.*

Then there exists a subset $K_*^{\mathrm{str}}(Q) \subset S_*(Q)$, closed under face maps, such that:

(i) *If C is a path component of $\partial_0 Q$ with $C \cap \partial_1 Q = \varnothing$, then $K_0^{\mathrm{str}}(Q)$ contains each point in C.*

(ii) *For a path component F of $\partial_1 Q$ with $F \cap \partial_0 Q = \varnothing$, there is exactly one point $x_F \in K_0^{\mathrm{str}}(Q) \cap F$, while*
for a path component F of $\partial_1 Q$ with $F \cap \partial_0 Q \neq \varnothing$, we have $K_0^{\mathrm{str}}(Q) \cap F = \{x_{E_0^F}, \dots, x_{E_s^F}\}$.

(iii) $K_0^{\mathrm{str}}(Q) = K_0^{\mathrm{str}}(\partial Q)$.

(iv) $K_1^{\mathrm{str}}(Q)$ *consists of*
  - *all* 1*-simplices $e \in K(Q)$ with $\partial e \in K_0^{\mathrm{str}}(Q)$,*
  - *exactly one* 1*-simplex for each nontrivial homotopy class* (*rel boundary*) *of loops e with $\partial_0 e = \partial_1 e \in K_0^{\mathrm{str}}(Q)$, and*
  - *the constant loop for the homotopy class of the constant loop at x, if $x \in K_0^{\mathrm{str}}(Q)$.*

(v) *For $n \geq 2$, if $\sigma \in S_n(Q)$ is an n-simplex with $\partial\sigma \in K_{n-1}^{\mathrm{str}}(Q)$, then σ is homotopic rel boundary to a unique $\tau \in K_n^{\mathrm{str}}(Q)$.*

(vi) *If $\sigma \in K_n^{\mathrm{str}}(Q)$ is homotopic rel boundary to some $\tau \in K_n(Q)$, then $\sigma = \tau$.*

(vii) *If $\sigma \in K_n^{\mathrm{str}}(Q)$ is homotopic rel boundary to a simplex $\tau \in S_n(\partial_1 Q)$, then $\sigma \in K_n^{\mathrm{str}}(\partial_1 Q)$; if $\sigma \in K_1^{\mathrm{str}}(Q)$ is homotopic rel boundary to a simplex $\tau \in S_1(\partial_0 Q)$, then $\sigma \in K_1^{\mathrm{str}}(\partial_0 Q)$.*

(viii) *$K_*^{\mathrm{str}}(Q)$ is aspherical*, *i.e.*, *if $\sigma, \tau \in K_*^{\mathrm{str}}(Q)$ have the same* 1*-skeleton*, *then $\sigma = \tau$.*

Proof. $K_*^{\mathrm{str}}(Q)$ is defined by induction on the dimension of simplices as follows.

Definition of $K_0^{\mathrm{str}}(Q)$: Choose $K_0^{\mathrm{str}}(Q)$ such that conditions (i)–(iii) are satisfied. Note that we have chosen a nonempty set of 0-simplices since we are assuming $\partial_1 Q \neq \varnothing$.

Definition of $K_1^{\mathrm{str}}(Q)$: For an ordered pair $(x, y) \in K_0^{\mathrm{str}}(Q) \times K_0^{\mathrm{str}}(Q)$ with $x \neq y$, there exists unique simplex in $K_1(Q)$ in each homotopy class (rel boundary) of arcs e from x to y. Choose these 1-simplices so they belong to $K_1^{\mathrm{str}}(Q)$. (Uniqueness implies that (vi) is true for $n = 1$.) For pairs $(x, x) \in K_0^{\mathrm{str}}(Q) \times K_0^{\mathrm{str}}(Q)$, choose one simplex in each homotopy class (rel boundary) of loops e from x to itself. For the homotopy class of the constant loop, choose the constant loop.

Choose the 1-simplices in $\partial_0 Q$ and/or $\partial_1 Q$ whenever this is possible. (If a 1-simplex is homotopic into both $\partial_0 Q$ and $\partial_1 Q$, then it is necessarily homotopic into $\partial_0 Q \cap \partial_1 Q$. Indeed, a disk realizing a homotopy between 1-simplices in $\partial_0 Q$ and $\partial_1 Q$ can be made transversal to $\partial_0 Q \cap \partial_1 Q$ and then intersects $\partial_0 Q \cap \partial_1 Q$ in an arc or loop.) Hence (vii) is satisfied for $n = 1$.

Definition of $K_n^{\mathrm{str}}(Q)$ for $n \geq 2$, assuming that $K_{n-1}^{\mathrm{str}}(Q)$ is defined: For an $(n+1)$-tuple $\kappa_0, \ldots, \kappa_n$ of $(n-1)$-simplices in $K_{n-1}^{\mathrm{str}}(Q)$, satisfying $\partial_i \kappa_j = \partial_{j-1} \kappa_i$ for all i, j, there are two possibilities:

- If no edge of any κ_i is a loop, then, by the asphericity of Q, there is a unique n-simplex $\sigma \in K_n(Q)$ with $\partial_i \sigma = \kappa_i$ for $i = 0, \ldots, n$. In this case set $\kappa := \sigma$. Uniqueness implies that (vi) is satisfied for n. (By the construction in Section 3B, we have $\kappa \in K_n(\partial_1 Q)$ if κ is homotopic rel boundary into $\partial_1 Q$.)
- Otherwise, choose an n-simplex $\kappa \in S_n(Q)$ with $\partial_i \kappa = \kappa_i$ for $i = 0, \ldots, n$. Since Q is aspherical, κ exists and is unique up to homotopy rel boundary. Choose the simplices in $\partial_1 Q$ whenever this is possible.

By construction, $K_*^{\mathrm{str}}(Q)$ is closed under face maps and satisfies the conditions (i)–(vii). Condition (viii) follows by induction on the dimension of subsimplices of σ and τ from condition (v). $\square$

The simplices in $K_*^{\mathrm{str}}(Q)$ will be called *straight simplices*.

We remark that $K_*^{\mathrm{str}}(Q)$ is not a multicomplex because simplices in $K_*^{\mathrm{str}}(Q)$ need not have pairwise distinct vertices. (Note also that simplices in $K(Q)$ belong to $K^{\mathrm{str}}(Q)$ if and only if all their vertices belong to $K_0^{\mathrm{str}}(Q)$, by construction.)

Observation 5.4. *Let $Q, \partial Q, \partial_1 Q, \partial_0 Q$ satisfy Assumption I. Let $K_*^{\mathrm{str}}(Q) \subset S_*(Q)$ satisfy conditions* (i)–(viii) *from Lemma 5.3. Then $q : Q \to Q$ induces a simplicial map $q : K^{\mathrm{str}}(Q) \to K^{\mathrm{str}}(Q)$, compatible with the simplicial map $q : K(Q) \to K(Q)$ from Section 5A.*

Proof. By construction, q maps $K_0^{\mathrm{str}}(Q)$ to itself. Indeed:

- If C is a path component of $\partial_0 Q$ with $C \cap \partial_1 Q = \varnothing$, then $q(v) = v$ for each $v \in C$.
- If F is a path component F of $\partial_1 Q$ with $F \cap \partial_0 Q = \varnothing$, then $q(v) = v$ for *each* $v \in F$ (in particular for the unique $v \in F \cap K_0^{\mathrm{str}}(Q)$).
- If F is a path component of $\partial_1 Q$ with $F \cap \partial_0 Q \neq \varnothing$, then we have $K_0^{\mathrm{str}}(Q) \cap F = \{x_{E_0^F}, \ldots, x_{E_s^F}\}$, and $q(x_{E_i^F}) = x_{E_0^F}$ for $i = 0, \ldots, s$ by Construction 5.1.

Hence q induces a simplicial map on $K^{\mathrm{str}}(Q)$. (The simplicial map agrees with q on the 0-skeleton, and it maps each 1-simplex $e \in K_1^{\mathrm{str}}(Q)$ to the unique 1-simplex of $K_1^{\mathrm{str}}(Q)$ that is in the homotopy class rel $\{0, 1\}$ of $q(e)$. Since $K^{\mathrm{str}}(Q)$ is aspherical, this determines the simplicial map q uniquely.) $\square$

5C. *Definition of the straightening.*

Definition 5.5. Let $(Q, \partial_1 Q)$ be a pair of topological spaces and let $z = \sum_{i \in I} a_i \tau_i \in C_n^{\mathrm{inf}}(Q)$ be a (possibly infinite) singular chain.

(a) A set of cancellations of z is a symmetric set $\mathscr{C} \subset S_{n-1}(Q) \times S_{n-1}(Q)$ with $(\eta_1, \eta_2) \in \mathscr{C} \Rightarrow \eta_1 = \eta_2$ and $\eta_1 = \partial_k \tau_{i_1}, \eta_2 = \partial_l \tau_{i_2}$ for some $i_1, i_2 \in I$ and $k, l \in \{0, \ldots, n\}$.

(b) Let $z = \sum_{i \in I} a_i \tau_i \in C_n^{\inf}(Q)$. If $\mathscr{C}$ is a set of cancellations for z, the associated simplicial set $\Upsilon_{z,\mathscr{C}}$ is the simplicial set generated[5] by $\{\Delta_i : i \in I\}$, subject to the identifications $\partial_k \Delta_{i_1} = \partial_l \Delta_{i_2}$ if and only if $(\partial_k \tau_{i_1}, \partial_l \tau_{i_2}) \in \mathscr{C}$.

(c) Let $z = \sum_{i \in I} a_i \tau_i \in C_n^{\inf}(Q)$. Choose a minimal presentation for ∂z (meaning that no further cancellation is possible). Define

$$J = J_{\partial z} := \big\{(i, a) \in I \times \{0, \ldots, n\} : \partial_a \tau_i \text{ occurs with a nonzero coefficient in the chosen presentation of } \partial z\big\}.$$

Let $\mathscr{C}$ be a set of cancellations for z. Then the simplicial set $\partial \Upsilon_{z,\mathscr{C}}$ is defined as the set consisting of $|J|$ $(n-1)$-simplices $\Delta_{i,a}$, $(i, a) \in J$, together with all their iterated faces and degenerations, subject to the identifications $\partial_a \partial_{a_1} \tau_{i_1} = \partial_a \partial_{a_2} \tau_{i_2}$ for all $a = 0, \ldots, n-1$, whenever $(\partial_{a_1} \tau_{i_1}, \partial_{a_2} \tau_{i_2}) \in \mathscr{C}$ and $(i_1, a_1) \in J$.

(d) If $z = \sum_{i \in I} a_i \tau_i \in C_n^{\inf}(Q)$ is a relative cycle, then a set of cancellations $\mathscr{C}$ is called sufficient if the formal sum $\sum_{i \in I} \sum_{k=0}^{n} (-1)^k a_i \partial_k \tau_i$ can be reduced to a chain in $C_{n-1}^{\inf}(\partial Q)$ by substracting (possibly infinitely many) multiples of $(\partial_{a_1} \tau_{i_1} - \partial_{a_2} \tau_{i_2})$ with $(\partial_{a_1} \tau_{i_1}, \partial_{a_2} \tau_{i_2}) \in \mathscr{C}$.

Observations 5.6. Let $(Q, \partial_1 Q)$ be a pair of topological spaces.

(a) If $z = \sum_{i \in I} a_i \tau_i \in C_n^{\inf}(Q)$ is a singular chain, $\mathscr{C}$ is a set of cancellations, and $\Upsilon := \Upsilon_{z,\mathscr{C}}$ is the associated simplicial set, the geometric realization $|\Upsilon|$ is obtained from $|I|$ copies of the standard n-simplex Δ_i, $i \in I$, with identifications $\partial_{a_1} \Delta_{i_1} = \partial_{a_2} \Delta_{i_2}$ if and only if $(\partial_{a_1} \tau_{i_1}, \partial_{a_2} \tau_{i_2}) \in \mathscr{C}$. For a minimal presentation of ∂z and $\partial \Upsilon := \partial \Upsilon_{z,\mathscr{C}}$, $|\partial \Upsilon|$ is the subspace of $|\Upsilon|$ containing all simplices $\partial_{a_1} \Delta_{i_1}$ with $(i_1, a_1) \in J$.

(b) There exists an associated continuous map $\tau : |\Upsilon| \to Q$ with $\tau|\Delta_i = \tau_i$ (upon the identification $\Delta_i = \Delta^n$). If z is a relative cycle, i.e., if $\partial z \in C_{n-1}^{\inf}(\partial_1 Q)$, then τ maps $|\partial \Upsilon|$ to $\partial_1 Q$.

(c) Let $z_1 = \sum_{i \in I} a_i \tau_i$, $z_2 = \sum_{i \in I} a_i \sigma_i \in C_n^{\inf}(Q, \partial_1 Q)$ be relative cycles and let $\mathscr{C}_1, \mathscr{C}_2$ be sufficient sets of cancellations of z_1 and z_2, respectively. Assume that $(\partial_{a_1} \tau_{i_1}, \partial_{a_2} \tau_{i_2}) \in \mathscr{C}_1$ if and only if $(\partial_{a_1} \sigma_{i_1}, \partial_{a_2} \sigma_{i_2}) \in \mathscr{C}_2$, and that there exist minimal presentations of $\partial z_1, \partial z_2$ such that $J_{z_1} = J_{z_2}$.

If the associated continuous maps $\tau, \sigma : |\Upsilon| \to Q$ are homotopic, for a homotopy mapping $|\partial \Upsilon|$ to ∂Q, then $\sum_{i \in I} a_i \tau_i$ and $\sum_{i \in I} a_i \sigma_i \in C_*^{\inf}(Q, \partial Q)$ are relatively homologous.

[5] That is, the subset of $S_*^{\text{sing}}(Q)$ containing the $|I|$ n-simplices Δ_i, $i \in I$, together with all simplices obtained by iterated applications of face and degeneracy operators. See [May 1967, Example 1.5].

We emphasize that we do not assume that $\mathscr{C}$ is a complete list of cancellations, and the simplicial map $\tau_* : C_*^{\text{simp}}(\Upsilon) \to C_*^{\text{sing}}(Q)$ need not be injective.

After having set up the necessary notations, we will now define the actual straightening. We first mention that there is of course an analogue of the classical straightening of [Benedetti and Petronio 1992, Lemma C.4.3] in our setting.

Observation 5.7. *Let Q, ∂Q, $\partial_1 Q$, $\partial_0 Q$ satisfy Assumption I. Suppose $K_*^{\text{str}}(Q) \subset S_*(Q)$ satisfy conditions* (i)–(viii) *from Lemma 5.3. Then there exists a "canonical straightening" map*

$$\text{str}_{\text{can}} : C_*^{\text{simp,inf}}(K(Q)) \to C_*^{\text{simp,inf}}(K^{\text{str}}(Q)),$$

mapping $C_*^{\text{simp,inf}}(K(\partial_1 Q))$ to $C_*^{\text{simp,inf}}(K^{\text{str}}(\partial_1 Q))$, with the following properties:

(i) str_{can} *is a chain map.*

(ii) *If $z = \sum_{i \in I} a_i \tau_i \in C_*^{\text{simp,inf}}(K(Q))$ and $\sum_{i\in I} a_i \sigma_i := \sum_{i \in I} a_i \,\text{str}_{\text{can}}(\tau_i)$, then the maps $\tau, \sigma : |\Upsilon| \to Q$* (defined by Observation 5.6(b) after fixing a set of cancellations $\mathscr{C}$ and a minimal presentation of ∂z) *are homotopic.*

Moreover, if $z = \sum_{i \in I} a_i \tau_i$ is a relative cycle with $\partial z \in C_*^{\text{simp,inf}}(K(\partial_1 Q))$, the same is true for $\sum_{i \in I} a_i \sigma_i$, and

$$\tau, \sigma : (|\Upsilon|, |\partial \Upsilon|) \to (Q, \partial_1 Q)$$

are homotopic as maps of pairs.

In particular, $\sum_{i\in I} a_i \,\text{str}_{\text{can}}(\tau_i)$ is relatively homologous to $\sum_{i \in I} a_i \tau_i$.

Proof. We define str_{can}, and the homotopy to the identity, by induction on the dimension of simplices. (During the construction we take care that str_{can} and the homotopy preserve $K(\partial_1 Q)$.)

0-simplices. If C is a path component of $\partial_0 Q$ with $C \cap \partial_1 Q = \varnothing$, we define $\text{str}_{\text{can}}(v) = v$ for each 0-simplex v in C. The homotopy $H(v)$ is for each v given by the constant map.

If C is a path component of $\partial_0 Q$ with $C \cap \partial_1 Q \neq \varnothing$, there is at least one path component F of $\partial_1 Q$ with $C \cap F \neq \varnothing$. By Construction 5.1 and condition (ii) from Lemma 5.3, for each such F, there is a straight 0-simplex $x_{E_i^F} \in C \cap F$. Choose one such straight 0-simplex (among the $x_{E_i^F}$) for each path component C of $\partial_0 Q$, denote it x_C, and for each $v \in C$ we define $\text{str}_{\text{can}}(v) := x_C \in K_0^{\text{str}}(Q) \cap C$ and we choose the homotopy $H(v)$ to belong to C.

If $v \in \partial_1 Q$, then there is (at least) one straight 0-simplex in the same path component F of $\partial_1 Q$, we choose $\text{str}_{\text{can}}(v) \in F \cap K_0^{\text{str}}(Q)$ and there exists $H(v) \in K_1(\partial_1 Q)$ with $\partial H(v) = v - \text{str}_{\text{can}}(v)$.

If $v \notin \partial Q$, then we define $\text{str}_{\text{can}}(v)$ to be some straight 0-simplex in ∂Q and we fix arbitrarily some $H(v) \in K_1(Q)$ with $\partial H(v) = v - \text{str}_{\text{can}}(v)$.

1-simplices. For $e \in K_1(Q)$ we define

$$\mathrm{str}_{\mathrm{can}}(e) := [\overline{H(\partial_1 e)} * e * H(\partial_0 e)],$$

where, as always, $[\,\cdot\,]$ denotes the unique 1-simplex in $K_1^{\mathrm{str}}(Q)$, that is homotopic rel boundary to the path in brackets.

The simplex e is homotopic to $\mathrm{str}_{\mathrm{can}}(e)$ by the canonical homotopy that is inverse to the homotopy moving $\overline{H(\partial_1 e)}$ or $H(\partial_0 e)$ into constant maps. In particular, the restriction of this homotopy to $\partial_1 e$, $\partial_0 e$ gives $H(\partial_1 e)$, $\overline{H(\partial_0 e)}$. Thus, for different edges with common vertices, the homotopies are compatible. We thus have constructed a homotopy for the 1-skeleton Υ_1.

We note that, for $v \in \partial_1 Q$, the homotopy $H(v)$ is either constant or lies in $K_1(\partial_1 Q)$, Thus if $\tau \in K_1(\partial_1 Q)$ then $\mathrm{str}_{\mathrm{can}}(\tau) \in K_1^{\mathrm{str}}(\partial_1 Q)$ and the homotopy between τ and $\mathrm{str}_{\mathrm{can}}(\tau)$ takes place in $\partial_1 Q$.

n-simplices. We assume inductively, that for some $n \geq 1$, we have defined $\mathrm{str}_{\mathrm{can}}$ on $K_{*\leq n}(Q)$, mapping $K_{*\leq n}(\partial_1 Q)$ to $K_{*\leq n}^{\mathrm{str}}(\partial_1 Q)$, and satisfying (i) and (ii).

Let $\tau \in K(Q)$ be an $(n+1)$-simplex. Then we have by (ii) a homotopy between $\partial\tau$ and $\mathrm{str}_{\mathrm{can}}(\partial\tau)$. By Observation 3.1 this homotopy extends to τ. The resulting simplex τ' satisfies $\partial\tau' \in K_n^{\mathrm{str}}(Q)$. Condition (v) from Lemma 5.3 means that τ' is homotopic rel boundary to a unique simplex $\mathrm{str}_{\mathrm{can}}(\tau) \in K_{n+1}^{\mathrm{str}}(Q)$. This proves the inductive step.

If $\tau \in K(\partial_1 Q)$, then we can inductively assume that the homotopy of $\partial\tau$ has image in $\partial_1 Q$. Then condition (vii) from Lemma 5.3 implies $\mathrm{str}_{\mathrm{can}}(\tau) \in K_{n+1}^{\mathrm{str}}(\partial_1 Q)$. Moreover, since $\partial_1 Q$ is aspherical, the homotopy of τ can be chosen to have image in $\partial_1 Q$.

By construction, for any set of cancellations $\mathscr{C}$, the induced maps τ and σ are homotopic. In particular, if we chose a sufficient set of cancellations in the sense of Definition 5.5(d), then Observation 5.6(c) implies that $\sum_{i=1}^r a_i\, \mathrm{str}_{\mathrm{can}}(\tau_i)$ is (relatively) homologous to $\sum_{i=1}^r a_i \tau_i$. □

However, we want to define a more refined straightening, which will be defined only on relative cycles with some kind of additional information.

Before stating the definition of distinguished 1-simplices, we remark that there is a left and right action of the pseudogroup $\Gamma := \Omega(\partial Q)$ (as defined in Section 3D) on $K_1^{\mathrm{str}}(Q)$: if $e \in K_1^{\mathrm{str}}(Q)$, $\gamma_1 \in \pi_1(\partial Q, \partial_1 e)$, $\gamma_2 \in \pi_1(\partial Q, \partial_0 e)$, then let $\gamma_1 e \gamma_2$ be the unique straight 1-simplex homotopic rel $\{0, 1\}$ to $\gamma_1 * e * \gamma_2$. (The left action agrees with the action defined in Section 3D.) The cosets $\Gamma K_1^{\mathrm{str}}(Q)\Gamma$ in Definition 5.8 are with respect to this action.

For $x, y \in K_0^{\mathrm{str}}(Q)$ we will denote $K_{1,xy}^{\mathrm{str}} := \{e \in K_1^{\mathrm{str}}(Q) : \partial_1 e = x,\ \partial_0 e = y\}$.

Definition 5.8. Let Q, ∂Q, $\partial_1 Q$, $\partial_0 Q$ satisfy Assumption I.

Let $q : Q \to Q$ and $\{x_{E_i^F} \in \partial_0 Q \cap \partial_1 Q\}$ be given by Construction 5.1.

Let $K_*^{\mathrm{str}}(Q) \subset S_*(Q)$ satisfy conditions (i)–(viii) from Lemma 5.3.

A set $D \subset K_1^{\mathrm{str}}(Q)$ is called a set of distinguished 1-simplices if it satisfies the following conditions:

(ix) $\partial_0 e, \partial_1 e \in K_0^{\mathrm{str}}(Q)$ for each $e \in D$.

(x) For each $(x, y) \in K_0^{\mathrm{str}}(Q) \times K_0^{\mathrm{str}}(Q)$, the set

$$D_{xy} := \{e \in D : \partial_1 e = x, \partial_0 = y\}$$

contains exactly one element in each double coset $\Gamma f \Gamma \in \Gamma K_{1,xy}^{\mathrm{str}}(Q)\Gamma$, where $\Gamma = \Omega(\partial Q)$.

(xi) For all $x \in K_0^{\mathrm{str}}(Q)$, the constant loop c_x belongs to D.

(xii) If $e \in D$, then $\bar{e} \in D$, where the bar denotes orientation reversal.

(xiii) If F, F' are path components of $\partial_1 Q$ and

$$\{x_{E_i^F} \in \partial_0 Q \cap F\}, \quad \{x_{E_j^{F'}} \in \partial_0 Q \cap F'\}$$

are given by Construction 5.1, then $q(D_{x_{E_i^F} x_{E_j^{F'}}}) = D_{x_{E_0^F} x_{E_0^{F'}}}$ for all $x_{E_i^F}, x_{E_j^{F'}}$.

(xiv) If $x_1, x_2 \in C_1$ and $y_1, y_2 \in C_2$ for some path components C_1, C_2 of ∂Q, then for each $e_1 \in D_{x_1 y_1}$ there exists some $e_2 \in D_{x_2 y_2}$ with $q(e_2) = gq(e_1)$ for some $g \in H := q_*(\Pi(K(\partial_0 Q)))$.

In connection with (xiii) we note that if $F \cap \partial_0 Q = \varnothing$, there is only one straight 0-simplex $x_{E_0^F}$ in F. Similarly, if $F' \cap \partial_0 Q = \varnothing$, there is only one straight 0-simplex $x_{E_0^{F'}}$ in F'. In particular, if $F \cap \partial_0 Q = \varnothing$ and $F' \cap \partial_0 Q = \varnothing$, condition (xiii) is empty.

Observation 5.9. *Let the assumptions of Definition 5.8 be satisfied. Then a set D of distinguished 1-simplices exists.*

Proof. For each path component C of ∂Q we fix some $x_C \in j_0^{\mathrm{str}}(C)$.

For each pair $\{C_1, C_2\}$ of path components, we select for membership in $D_{x_{C_1} x_{C_2}}$ one simplex e with

$$\partial_1 e = x_{C_1}, \quad \partial_0 e = x_{C_2}$$

in each coset of $\Gamma K_{1,x_{C_1} x_{C_2}}^{\mathrm{str}}(Q)\Gamma$. If e is selected for $D_{x_{C_1} x_{C_2}}$, we select $\bar{e}$ for $D_{x_{C_2} x_{C_1}}$. If $C_1 = C_2$, then in particular for the coset of the constant loop we choose the constant loop for $D_{x_{C_1} x_{C_2}}$.

For each path component C of ∂Q and each path component F of $C \cap \partial_1 Q$, we conclude that $q(x_C)$ and $q(x_{E_0^F})$ belong to the path-connected set $q(\partial_0 Q \cap C)$.

Therefore we have a sequence of 1-simplices $\alpha_1, \dots, \alpha_m \in K_1(\partial_0 Q)$ with images in distinct path components of $\partial_0 Q \cap C$, such that

$$\partial_1 q(\alpha_1) = q(x_C), \qquad \partial_0 q(\alpha_1) = \partial_1 q(\alpha_2), \quad \dots,$$
$$\partial_0 q(\alpha_{m-1}) = \partial_1 q(\alpha_m), \quad \partial_0 q(\alpha_m) = q(x_{E_0^F}).$$

To prepare the definition of the $D_{x,y}$, we first describe, for each $x \in C \cap K_0^{\mathrm{str}}(Q)$, a sequence $\{\alpha_1, \dots, \alpha_k\}$ of 1-simplices:

- If $C \cap \partial_1 Q = \varnothing$, then $k = 1$ and for each $x \in C$ we choose arbitrarily a 1-simplex α_1 in C with $\partial_1 \alpha_1 = x_C$, $\partial_0 \alpha_1 = x$.
- If $C \cap \partial_0 Q = \varnothing$, then $C \cap K_0^{\mathrm{str}}(Q) = \{x_C\}$ by condition (ii) of Lemma 5.3, and we let $k = 0$.
- If $C \cap \partial_0 Q \cap \partial_1 Q \neq \varnothing$, again by condition (ii) of Lemma 5.3 we have $x = x_{E_i^F}$ for some path component F of $\partial_1 Q$ and some i; thus we have the sequence $\alpha_1, \dots, \alpha_m$ constructed above with $\partial_1 q(\alpha_1) = q(x_C)$, $\partial_0 q(\alpha_1) = \partial_1 q(\alpha_2)$, …, $\partial_0 q(\alpha_{m-1}) = \partial_1 q(\alpha_m)$, $\partial_0 q(\alpha_m) = q(x_{E_i^F})$, where the last equality holds true because $q(x_{E_i^F}) = x_{E_0^F} = q(x_{E_0^F})$.

Let $x, y \in K_0^{\mathrm{str}}(Q)$. Let C_1, C_2 be the path components of ∂Q with $x \in C_1$ and $y \in C_2$. We have constructed sequences of 1-simplices $\alpha_1, \dots, \alpha_k \in K_1(\partial Q)$ and $\beta_1, \dots, \beta_l \in K_1(\partial Q)$ such that $\partial_1 q(\alpha_1) = q(x_{C_1})$, $\partial_0 q(\alpha_1) = \partial_1 q(\alpha_2)$, …, $\partial_0 q(\alpha_{k-1}) = \partial_1 q(\alpha_k)$, $\partial_0 q(\alpha_k) = q(x)$, and $\partial_1 q(\beta_1) = q(x_{C_2})$, $\partial_0 q(\beta_1) = \partial_1 q(\beta_2)$, …, $\partial_0 q(\beta_{k-1}) = \partial_1 q(\beta_k)$, $\partial_0 q(\beta_k) = q(y)$. Note that all $q(\alpha_i)$ and $q(\beta_i)$ are either constant or contained in $q(K_1(\partial_0 Q))$.

Let $H := q_*(\Pi(K(\partial_0 Q)))$. Define

$$g := \{q(\alpha_1), q(\bar{\alpha}_1)\} \dots \{q(\alpha_k), q(\bar{\alpha}_k)\}\{q(\beta_l), q(\bar{\beta}_l)\} \dots \{q(\beta_1), q(\bar{\beta}_1)\} \in H.$$

(If $k = l = 0$, this just means $g = 1$.)

We have $g = g^{-1}$ and

$$ge \in K^{\mathrm{str}}_{1,q(x)q(y)}(Q) \iff e \in K^{\mathrm{str}}_{1,q(x_{C_1})q(x_{C_2})}(Q).$$

By construction, *the g associated to $x_{E_i^F}, x_{E_j^{F'}}$ agrees with the g associated to $x_{E_0^F}, x_{E_0^{F'}}$.*

We are given $D_{x_{C_1} x_{C_2}}$ and we want to define D_{xy} such that condition (xiii) is satisfied.

First, if $C_1 \cap \partial_1 Q = \varnothing$ or $C_2 \cap \partial_1 Q = \varnothing$, then we can fix an arbitrary choice of $D_{x,y}$ satisfying conditions (x)–(xii). (Condition (xiii) is empty in this case.)

So let us assume $C_1 \cap \partial_1 Q \neq \varnothing$ and $C_2 \cap \partial_1 Q \neq \varnothing$. We note that

$$q : (Q, \partial Q, \partial_1 Q) \to (Q, \partial Q, \partial_1 Q)$$

is homotopic to the identity as a map of triples, by the construction in Section 5A. This implies that cosets of $\Gamma K^{\mathrm{str}}_{1,xy}(Q)\Gamma$ are in one-to-one correspondence (by applying q) with those of $\Gamma K^{\mathrm{str}}_{1,q(x)q(y)}\Gamma$. Thus it suffices to describe $q(D_{xy}) \subset K^{\mathrm{str}}_{1,q(x)q(y)}$.

Let

$$\Gamma f \Gamma \in \Gamma K^{\mathrm{str}}_{1,q(x)q(y)}(Q)\Gamma$$

be a double coset. Then the double coset

$$\Gamma (gf)\Gamma \in \Gamma K^{\mathrm{str}}_{1,q(x_{C_1})q(x_{C_2})}(Q)\Gamma$$

is the image under q of some double coset

$$\Gamma e'\Gamma \in \Gamma K^{\mathrm{str}}_{1,x_{C_1}x_{C_2}}(Q)\Gamma$$

Let e be the unique distinguished simplex in $\Gamma e'\Gamma$. Then we choose $gq(e)$ to be the distinguished simplex in $\Gamma f\Gamma$. This is possible because $gq(e)$ belongs to the double coset $\Gamma f\Gamma$. Indeed,

$$q(e) \in \Gamma(gf)\Gamma$$

means that $q(e) = q_*(\gamma_1)gfq_*(\gamma_2)$ for some loops γ_1 and γ_2 based at x_{C_1} and x_{C_2}, respectively, and this implies $gq(e') = q_*(\gamma_1')fq_*(\gamma_2')$ with

$$\gamma_1' := [\bar{\alpha}_m * \ldots * \bar{\alpha}_1 * \gamma_1 * \alpha_1 * \ldots * \alpha_m],\ \gamma_2' := [\bar{\beta}_n * \ldots * \bar{\beta}_1 * \gamma_2 * \beta_1 * \ldots * \beta_n].$$

This defines D_{xy}. By construction, condition (xiv) is satisfied if $e_1 \in D_{x_{C_1}x_{C_2}}$. In general, if $e_1 \in D_{x_1y_1}$, then we get $e \in D_{x_{C_1}x_{C_2}}$ and $g_1 \in H$ with $q(e_1) = g_1q(e)$ and $e_2 \in D_{x_2y_2}$, $g_2 \in H$ with $q(e_2) = g_2q(e)$; thus $q(e_2) = g_2g_1^{-1}q(e_1)$.

Condition (xiii) is implied because $q(x_{E_i^F}) = x_{E_0^F}$, $q(x_{E_j^{F'}}) = x_{E_0^{F'}}$ and the g associated to $x_{E_i^F}, x_{E_j^{F'}}$ agrees with the g associated to $x_{E_0^F}, x_{E_0^{F'}}$.

One checks easily that (xi) and (xii) are true for D_{xy}, since they are true for $D_{x_{C_1}x_{C_2}}$. $\square$

Definition 5.10. Let $Q, \partial Q, \partial_0 Q, \partial_1 Q$ satisfy Assumption I. Let $z = \sum_{i\in I} a_i\tau_i \in C_n^{\mathrm{inf}}(Q)$ be a singular chain and let Υ be the associated simplicial set (for some set of cancellations $\mathscr{C}$).

We say that a 0-1 labeling of the elements of the 1-skeleton Υ_1 is admissible if $\partial e_1 \cap \partial e_2 = \emptyset$ for all 1-labeled vertices e_1, e_2.

Lemma 5.11. *Let $Q, \partial Q, \partial_1 Q, \partial_0 Q$ satisfy Assumption I. Let $q : Q \to Q$ be given by Construction 5.1.*

Let $K^{\mathrm{str}}_*(Q) \subset S_*(Q)$ satisfy conditions (i)–(viii) *from Lemma 5.3, and let $D \subset K^{\mathrm{str}}_1(Q)$ be a set of distinguished* 1*-simplices.*

Let $z = \sum_{i\in I} a_i\tau_i \in C_*^{\mathrm{simp,inf}}(K(Q))$ be a relative cycle with

$$\partial z \in C_*^{\mathrm{simp,inf}}(K(\partial_1 Q)).$$

Let a set of cancellations $\mathscr{C}$ for z and a minimal presentation of ∂z be given. Let Υ, $\partial\Upsilon$ be the associated simplicial sets, $\tau : (|\Upsilon|, |\partial\Upsilon|) \to (Q, \partial_1 Q)$ the associated continuous mapping.

Assume that we have an admissible 0-1 *labeling of* Υ_1. *Then there exists a relative cycle*

$$z' = \sum_{i\in I} a_i \tau_i' \in C_*^{\mathrm{simp,inf}}(K^{\mathrm{str}}(Q), K^{\mathrm{str}}(\partial_1 Q))$$

satisfying the following conditions:

(i) *The associated continuous mappings*

$$\tau, \tau' : (|\Upsilon|, |\partial\Upsilon|) \to (Q, \partial_1 Q)$$

are homotopic by a homotopy mapping $|\partial\Upsilon|$ *to* ∂Q.

(ii) *If an edge of some* τ_i *is labeled by* 1, *the corresponding edge of* τ_i' *belongs to* D.

Remark. The homotopy in (i) does not necessarily map $|\partial\Upsilon|$ to $\partial_1 Q$, but to ∂Q.

Proof. First we apply the canonical straightening $\mathrm{str}_{\mathrm{can}}$ from Observation 5.7. The resulting chain $\sum_{i\in I} a_i\, \mathrm{str}_{\mathrm{can}}(\tau_i)$ satisfies (i), but not necessarily (ii).

$\sum_{i\in I} a_i\, \mathrm{str}_{\mathrm{can}}(\tau_i)$ inherits the admissible labeling from $\sum_{i\in I} a_i \tau_i$. Thus we can, without loss of generality, restrict ourselves to the case that all τ_i belong to $K^{\mathrm{str}}(Q)$.

Let $e \in K_1^{\mathrm{str}}(Q)$ be a 1-labeled edge, and set $x = \partial_1 e \in K_0^{\mathrm{str}}(Q)$, $y = \partial_0 e \in K_0^{\mathrm{str}}(Q)$. By Definition 5.8, the coset $\Gamma e \Gamma$ contains a unique distinguished 1-simplex $\mathrm{str}(e) \in D_{xy}$. (We use the notation from Definition 5.8; in particular, $\Gamma := \Omega(\partial Q)$.)

That $\mathrm{str}(e) \in \Gamma e \Gamma$ means[6] that there are loops $\gamma_1, \gamma_2 \subset \partial Q$ based at x and y, respectively, such that $\mathrm{str}(e) \sim \gamma_1 * e * \gamma_2$ rel $\{0, 1\}$. There is an obvious homotopy between e and $\gamma_1 * e * \gamma_2$, which moves $\partial_1 e$ along $\overline{\gamma}_1$ and $\partial_0 e$ along γ_2. (Of course, we change the homotopy class relative boundary, so we cannot keep the endpoints fixed during the homotopy.) If e and/or $\partial_0 e$ and/or $\partial_1 e$ have image in $\partial_1 Q$, then their images remain in ∂Q (and end up in $\partial_1 Q$) during the homotopy.

Using Observation 3.1, the homotopy thus constructed between e and $\mathrm{str}(e)$ can be extended to a homotopy from

$$\tau : (|\Upsilon|, |\partial\Upsilon|) \to (Q, \partial_1 Q)$$

to some

$$\hat{\tau} : (|\Upsilon|, |\partial\Upsilon|) \to (Q, \partial_1 Q),$$

such that $\hat{\tau}$ is a simplicial map from Υ to $S_*(Q)$. (If a 0-labeled edge has one or both vertices adjacent to 1-labeled edges, then the 0-labeled edge just follows the

[6] If $\partial_0 e, \partial_1 e \notin \partial_1 Q$, then $\mathrm{str}(e) \in \Gamma e \Gamma$ means, of course, $\mathrm{str}(e) = e$. Similarly, if only one vertex of e belongs to $\partial_1 Q$, then only that vertex is moved during the homotopy.

homotopy of the vertices. Edges labeled with 0 but not adjacent to 1-labeled edges can remain fixed during the homotopy.) The homotopy maps $|\partial\Upsilon|$ to ∂Q.

Next we apply homotopies rel boundary to the (already homotoped images of) all 0-labeled edges $f \in K_1^{\mathrm{str}}(Q)$, to homotope them to edges in $K_1^{\mathrm{str}}(Q)$. If f and/or $\partial_0 f$ and/or $\partial_1 f$ have image in $\partial_1 Q$, then their images remain in ∂Q (and end up in $\partial_1 Q$) during the homotopy.

Now we have a simplicial map $\hat{\tau} : \Upsilon \to S_*(Q)$, such that all 1-simplices are mapped to $K_1^{\mathrm{str}}(Q)$, and such that

$$\hat{\tau}(e) \in D \subset K_1^{\mathrm{str}}(Q)$$

holds for all 1-labeled edges e. Then we can, as in the proof of Observation 5.7, by induction on n, apply homotopies rel boundary to all n-simplices to homotope them into $K_n^{\mathrm{str}}(Q)$. Simplices in $\partial_1 Q$ remain in ∂Q (and end up in $\partial_1 Q$) during the homotopy.

We obtain a homotopy (of pairs), which keeps the 1-skeleton fixed, to a simplicial map

$$\tau' : \Upsilon \to K^{\mathrm{str}}(Q),$$

mapping $\partial\Upsilon$ to $K^{\mathrm{str}}(\partial_1 Q)$ and satisfying conditions (i) and (ii) of Lemma 5.11. $\square$

A somewhat artificial formulation of the conclusion of Lemma 5.11 is that we have constructed a chain map

$$\mathrm{str} : C_*^{\mathrm{simp,inf}}(\Upsilon, \partial\Upsilon) \to C_*^{\mathrm{simp,inf}}(K^{\mathrm{str}}(Q), K^{\mathrm{str}}(\partial_1 Q)).$$

Unfortunately, this somewhat artificial formulation can not be simplified because str depends on the chain $\sum_{i\in I} a_i\tau_i$. That is, we do not get a chain map

$$\mathrm{str} : C_*^{\mathrm{simp,inf}}(K(Q), K(\partial_1 Q)) \to C_*^{\mathrm{simp,inf}}(K^{\mathrm{str}}(Q), K^{\mathrm{str}}(\partial_1 Q)).$$

5D. ***Straightening of crushed cycles.*** Recall from Section 3H that $(\,\cdot\,) \otimes_{\mathbb{Z}G} \mathbb{Z}$ means the tensor product with the trivial $\mathbb{Z}G$-module $\mathbb{Z}$, that is, the quotient under the G-action. We first state obvious generalizations of Observation 5.6 to the case of tensor products with a factor with trivial G-action.

Observation 5.12. Let $(Q, \partial_1 Q)$ be a pair of topological spaces. Let G be a group acting on a pair $(K, \partial K)$ with $K \subset S_*(Q)$ and $\partial K \subset S_*(\partial_1 Q)$ both closed under face maps.

(i) If

$$z = \sum_{i\in I} a_i\tau_i \otimes 1 \in C_*^{\mathrm{simp,inf}}(K, \partial K) \otimes_{\mathbb{Z}G} \mathbb{Z}$$

is a relative cycle, so is

$$\hat{z} = \sum_{i\in I}\sum_{g\in G} a_i(g\tau_i) \in C_*^{\mathrm{simp,inf}}(K, \partial K).$$

If $\mathscr{C}$ is a sufficient set of cancellations for z, there exists a set of cancellations $\widehat{\mathscr{C}}$ for $\hat{z}$ such that $(\eta_1, \eta_2) \in \widehat{\mathscr{C}}$ implies $(\eta_1 \otimes 1, \eta_2 \otimes 1) \in \mathscr{C}$.

If $\partial z = \sum_{a,i} c_{ai} \partial_a \tau_i \otimes 1$ is a minimal presentation for ∂z, then

$$\partial \hat{z} = \sum_{g \in G} \sum_{a,i} c_{ai} \partial_a (g\tau_i)$$

is a minimal presentation for $\hat{z}$.

(ii) Let $\widehat{\Upsilon}$, $\partial\widehat{\Upsilon}$ be the simplicial sets associated to $\hat{z}$, the sufficient set of cancellations $\widehat{\mathscr{C}}$ and the minimal presentation of $\partial\hat{z}$. They come with an obvious G-action. Then we have an associated continuous mapping $\hat{\tau} : (|\widehat{\Upsilon}|, |\partial\Upsilon|) \to (Q, \partial_1 Q)$.

Corollary 5.13. *Let $Q, \partial Q, \partial_1 Q, \partial_0 Q$ satisfy Assumption I. Let $q : Q \to Q$ be given by Construction 5.1. Let $K_*^{\mathrm{str}}(Q) \subset S_*(Q)$ satisfy conditions* (i)–(viii) *from Lemma 5.3, and let $D \subset K_1^{\mathrm{str}}(Q)$ be a set of distinguished* 1*-simplices.*

Let $G := \Pi(K(\partial_0 Q))$ with its action on $K^{\mathrm{str}}(Q)$ defined in Observation 5.4, and let $H := q_*(G)$ as defined in Section 5A. Let

$$\sum_{i \in I} a_i \tau_i \otimes 1 \in C_n^{\mathrm{simp,inf}}(K(Q), GK(\partial_1 Q)) \otimes_{\mathbb{Z}G} \mathbb{Z}$$

be a relative cycle. Fix a sufficient set of cancellations $\mathscr{C}$ and a minimal presentation for ∂z. Let $\widehat{\Upsilon}$, $\partial\widehat{\Upsilon}$ be defined by Observation 5.12. Assume that we have a G-invariant admissible 0-1 *labeling of the edges of $\widehat{\Upsilon}$.*

Then there is a well-defined chain map

$$q \circ \mathrm{str} : C_*^{\mathrm{simp,inf}}(\widehat{\Upsilon}) \otimes_{\mathbb{Z}G} \mathbb{Z} \to C_*^{\mathrm{simp,inf}}(HK^{\mathrm{str}}(Q)) \otimes_{\mathbb{Z}H} \mathbb{Z},$$

mapping $C_*^{\mathrm{simp,inf}}(\partial\widehat{\Upsilon}) \otimes_{\mathbb{Z}G} \mathbb{Z}$ to $C_*^{\mathrm{simp,inf}}(GK^{\mathrm{str}}(\partial_1 Q)) \otimes_{\mathbb{Z}H} \mathbb{Z}$, satisfying the following conditions:

(i) *If $e \in \widehat{\Upsilon}_1$ is a* 1*-labeled edge*, $\mathrm{str}(e \otimes 1) = f \otimes 1$, *then $f \in D$.*

(ii) *If Q is an orientable manifold with boundary ∂Q, and if*

$$\sum_{i \in I} a_i \tau_i \otimes 1 \in C_*^{\mathrm{simp,inf}}(K(Q), GK(\partial_1 Q)) \otimes_{\mathbb{Z}G} \mathbb{Z}$$

represents[7] *the image of* $[Q, \partial Q] \otimes 1$, *then*

$$\sum_{i \in I} a_i q \circ \mathrm{str}(\tau_i \otimes 1) \in C_*^{\mathrm{simp,inf}}(HK^{\mathrm{str}}(Q), HK^{\mathrm{str}}(\partial_1 Q)) \otimes_{\mathbb{Z}H} \mathbb{Z}$$

represents[7] *the image of* $[Q, \partial Q] \otimes 1$ *and*

$$\partial \sum_{i \in I} a_i q \circ \mathrm{str}(\tau_i \otimes 1) \in C_*^{\mathrm{simp,inf}}(HK^{\mathrm{str}}(\partial_1 Q)) \otimes_{\mathbb{Z}H} \mathbb{Z}$$

represents the image of $[\partial Q] \otimes 1$.

[7] See the remark following Lemma 3.6.

Proof. We can apply Lemma 5.11 to the infinite chain $\sum_{i\in I,g\in H} a_i(g\tau_i)$ to obtain a chain map $\mathrm{str}: C_*^{\mathrm{simp,inf}}(\widehat{\Upsilon}) \to C_*^{\mathrm{simp,inf}}(K^{\mathrm{str}}(Q))$, given by

$$\mathrm{str}(g\tau_i) := (g\tau_i)'.$$

The map $q : (K^{\mathrm{str}}(Q), K^{\mathrm{str}}(\partial_1 Q)) \to (K^{\mathrm{str}}(Q), K^{\mathrm{str}}(\partial_1 Q))$ is defined by Observation 5.4. (We actually have $q \circ \mathrm{str}(g\tau_i) \in K^{\mathrm{str}}(Q)$. We need $HK^{\mathrm{str}}(Q)$ in the statement of Corollary 5.13 just to have the tensor product well-defined.)

We are going to define $q \circ \mathrm{str}(\sigma \otimes z) := q(\mathrm{str}(\sigma)) \otimes z$ for each $\sigma \in \widehat{\Upsilon}$ and $z \in \mathbb{Z}$. For this to be well-defined, we need this fact:

Claim. *For each $\sigma \in K$, $g \in G$, there exists $h \in H$ with $q(\mathrm{str}(g\sigma)) = hq(\mathrm{str}(\sigma))$.*

Proof. By condition (viii) from Lemma 5.3 (asphericity of $K^{\mathrm{str}}(Q)$), it suffices to check this for the 1-skeleton.

0-simplices. If $\sigma = v \in S_0(\partial_0 Q)$ then v and gv belong to the same path component C of $\partial_0 Q$, hence $\mathrm{str}(v)$ and $\mathrm{str}(gv)$ belong to the same path component C. Let $\gamma : [0, 1] \to \partial_0 Q$ be a path with $\gamma(0) = \mathrm{str}(v), \gamma(1) = \mathrm{str}(gv)$. Let γ' be the unique 1-simplex in $K(\partial_0 Q)$ which is homotopic rel boundary to γ. Let $g' := \{\gamma', \overline{\gamma'}\} \in G = \Pi(K(\partial_0 Q))$. Then $g' \mathrm{str}(v) = \mathrm{str}(gv)$, which implies $q(\mathrm{str}(gv)) = hq(\mathrm{str}(v))$ with $h = q_*(g') \in H$.

If $\sigma = v \notin \partial_0 Q$, then $gv = v$, hence $q(\mathrm{str}(gv)) = q(\mathrm{str}(v))$.

1-simplices. In a first step we prove that for $e \in K_1(Q)$ and $g \in G$ we have $\mathrm{str}_{\mathrm{can}}(ge) = g' \mathrm{str}_{\mathrm{can}}(e)$ with $g' \in G$. Then we show that, if $e \in K_1^{\mathrm{str}}(Q)$ and $g \in G$, there exists $h \in H$ with $q(\mathrm{str}(ge)) = hq(\mathrm{str}(e))$. Hence altogether we will get $q(\mathrm{str}(ge)) = q(\mathrm{str}(\mathrm{str}_{\mathrm{can}}(ge))) = q(\mathrm{str}(g' \ \mathrm{str}_{\mathrm{can}}(e))) = h\ q(\mathrm{str}(\mathrm{str}_{\mathrm{can}}(e))) = h\ q(\mathrm{str}(e))$.

Step 1. This is basically a case analysis.

First case: If both vertices of e do not belong to $\partial_0 Q$, then also both vertices of $\mathrm{str}_{\mathrm{can}}(e)$ do not belong to $\partial_0 Q$, and we have $ge = e$, $g\,\mathrm{str}_{\mathrm{can}}(e) = \mathrm{str}(e)$, which implies the conclusion.

Second case: If both vertices of e belong to $\partial_0 Q$, then $\mathrm{str}_{\mathrm{can}}(e) \sim \alpha_1 * e * \alpha_2$ and $\mathrm{str}_{\mathrm{can}}(ge) \sim \beta_1 * ge * \beta_2$ for some paths $\alpha_1, \alpha_2, \beta_1, \beta_2$ in $\partial_0 Q$. Moreover, by the definition of the action (Section 3.3) we have $ge \sim \gamma_2 * e * \gamma_1$ for some $\gamma_1, \gamma_2 \in K_1(\partial_0 Q)$. Thus $\mathrm{str}_{\mathrm{can}}(ge) \sim \beta_1 * \gamma_1 * \alpha_1^{-1} * \mathrm{str}_{\mathrm{can}}(e) * \alpha_2^{-1} * \gamma_2 * \beta_2$; in particular $\mathrm{str}_{\mathrm{can}}(ge) = g' \mathrm{str}_{\mathrm{can}}(e)$ for some $g' \in G$.

Third case: Finally we consider the case that one vertex, say $\partial_0 e$, belongs to $\partial_0 Q$, but $\partial_1 e$ does not. Then we are in the situation of the second case with $\gamma_2 = 1$ and $\alpha_2 = \beta_2$, except that α_2 is not contained in $\partial_0 Q$. We get $\mathrm{str}_{\mathrm{can}}(ge) \sim \beta_1 * \gamma_1 * \alpha_1^{-1} * \mathrm{str}_{\mathrm{can}}(e)$. Since $\beta_1 * \gamma_1 * \alpha_1^{-1}$ is contained in $\partial_0 Q$, this implies that $\mathrm{str}_{\mathrm{can}}(ge) = g' \mathrm{str}_{\mathrm{can}}(e)$ for some $g' \in G$.

Step 2. Let $e \in K_1^{\mathrm{str}}(Q)$.

If e is a 1-labeled edge, with $x = \partial_1 e$, $y = \partial_0 e \in K_0^{\mathrm{str}}(Q)$, then we have by condition (xiv) from Definition 5.8 that

$$q(\mathrm{str}(ge)) = hq(e_2)$$

for some $e_2 \in D_{xy}$ and some $h \in H$. But e_2 belongs to the same coset in $\Gamma K_1^{\mathrm{str}}(Q)\Gamma$ as e; thus $e_2 = \mathrm{str}(e)$, which proves the claim for e.

If f is adjacent to one 1-labeled edge e and $q(\mathrm{str}(ge)) = hq(\mathrm{str}(e))$, then $q(\mathrm{str}(gf)) = hq(\mathrm{str}(f))$ because the homotopy of f just followed that of e, and the homotopy of gf just followed that of ge; for example, if $\partial_1 f = \partial_1 e$ and $q(\mathrm{str}(ge)) \sim q_*(\alpha) * q(\mathrm{str}(e)) * q_*(\beta)$ with $\alpha, \beta \in K_1(\partial_0 Q)$, then $q(\mathrm{str}(gf)) \sim q_*(\alpha) * q(\mathrm{str}(f))$. Similarly if f is adjacent to two 1-labeled edges.

Finally, if a 0-labeled straight 1-simplex f is not adjacent to a 1-labeled edge, we have $\mathrm{str}(f) = f$ and $\mathrm{str}(gf) = gf$, which implies that $\mathrm{str}(gf) = g\,\mathrm{str}(f)$ and $q(\mathrm{str}(gf)) = q_*(g)\,\mathrm{str}(f)$.

This concludes the proof of the claim. □

Thus $q \circ \mathrm{str}$ is well-defined; by Lemma 5.11, it satisfies the equation in part (i) of our corollary. To prove part (ii), we first observe that, if $\sum_{i\in I} a_i\tau_i$ represents $[Q, \partial Q]$, then, by Observation 5.6(c) and condition (i) from Lemma 5.11 (together with $q \sim \mathrm{id}$), the element

$$\sum_{i\in I} a_i q \circ \mathrm{str}(\tau_i) = \sum_{i=1}^{r} a_i q(\tau_i')$$

represents $[Q, \partial Q]$ and the claim follows. Thus it suffices to check: if $\sum_{i\in I} a_i\tau_i \otimes 1$ is (relatively) homologous to $\sum_{j\in J} b_j\kappa_j \otimes 1$, then $q \circ \mathrm{str}(\sum_{i\in I} a_i\tau_i \otimes 1)$ is (relatively) homologous to $q \circ \mathrm{str}\left(\sum_{j\in J} b_j\kappa_j \otimes 1\right)$.

So let

$$\sum_{i\in I} a_i\tau_i \otimes 1 - \sum_{j\in J} b_j\kappa_j \otimes 1 = \partial \sum_{k\in K} c_k\eta_k \otimes 1 \bmod C_*^{\mathrm{simp,inf}}(GK(\partial_1 Q)) \otimes_{\mathbb{Z}G} \mathbb{Z}$$

for some chain $\sum_{k\in K} c_k\eta_k \otimes 1 \in C_*^{\mathrm{simp,inf}}(K(Q)) \otimes_{\mathbb{Z}G} \mathbb{Z}$. In complete analogy with Lemma 5.11, we may extend str to the simplicial set built by the $g\eta_k$, their faces and degenerations, and obtain a singular chain $q\bigl(\mathrm{str}\bigl(\sum_{k\in K} c_k\eta_k\bigr)\bigr)$ with boundary

$$\partial q \circ \mathrm{str}\Bigl(\sum_{k\in K} c_k\eta_k\Bigr) = q \circ \mathrm{str}\Bigl(\sum_{i\in I} a_i\tau_i \otimes 1\Bigr) - q \circ \mathrm{str}\Bigl(\sum_{j\in J} b_j\kappa_j \otimes 1\Bigr)$$
$$\bmod C_*^{\mathrm{simp,inf}}(HK^{\mathrm{str}}(\partial_1 Q)) \otimes_{\mathbb{Z}H} \mathbb{Z}.$$

This gives the first claim of (ii). The second claim of (ii) follows because ∂ maps $[Q, \partial Q]$ to $[\partial Q]$. □

5E. *Removal of* 0*-homologous chains.*

Definition 5.14. Let Q be an n-dimensional compact manifold with boundary ∂Q. We define $\mathrm{rmv} : S_*(Q) \to S_*(Q)$ by

$$\mathrm{rmv}(\sigma) = \begin{cases} 0 & \text{if } \sigma \text{ is weakly degenerate (Definition 5.2),} \\ \sigma & \text{else.} \end{cases}$$

Lemma 5.15. *Assume that Q is a n-dimensional compact manifold with boundary ∂Q. Let $K_*^{\mathrm{str}}(Q) \subset S_*(Q)$ satisfy conditions* (i)–(viii) *from Lemma 5.3. Then the map from $C_*^{\mathrm{simp}}(K^{\mathrm{str}}(Q), K^{\mathrm{str}}(\partial_0 Q) \cup K^{\mathrm{str}}(\partial_1 Q))$ to itself defined by*

$$\mathrm{rmv}([\sigma]) := [\mathrm{rmv}(\sigma)]$$

is a well-defined chain map. Moreover, if

$$\sum_{j=1}^{r} a_j \tau_j \in C_*^{\mathrm{simp}}(K^{\mathrm{str}}(Q), K^{\mathrm{str}}(\partial_0 Q) \cup K^{\mathrm{str}}(\partial_1 Q)) \subset C_*^{\mathrm{sing}}(Q, \partial Q)$$

represents $[Q, \partial Q]$, *then* $\sum_{j=1}^{r} a_j \,\mathrm{rmv}(\tau_j)$ *represents* $[Q, \partial Q]$.

Proof. If $\sigma \in K^{\mathrm{str}}(\partial_0 Q) \cup K^{\mathrm{str}}(\partial_1 Q)$, then $\mathrm{rmv}(\sigma) \in K^{\mathrm{str}}(\partial_0 Q) \cup K^{\mathrm{str}}(\partial_1 Q)$; thus rmv is well-defined. We next prove it is a chain map.

Assume that $\mathrm{rmv}(\sigma) = 0$. If σ has image in ∂Q, then $\mathrm{rmv}(\sigma)$ and $\mathrm{rmv}(\partial\sigma)$ both vanish; thus $\partial\, \mathrm{rmv}(\sigma) = \mathrm{rmv}(\partial\sigma)$.

If some edge e of σ, say connecting the i-th and j-th vertices, is a constant loop, then all faces of σ except possibly $\partial_i\sigma$ and $\partial_j\sigma$ have a constant edge. Thus $\mathrm{rmv}(\partial_k\sigma) = 0$ if $k \notin \{i, j\}$. Moreover, since e is constant, corresponding edges of $\partial_i\sigma$ and $\partial_j\sigma$ are homotopic rel boundary and thus agree (possibly up to orientation) by condition (v) from Lemma 5.3. By induction on the dimension of subsimplices we get, again using condition (v) from Lemma 5.3, that $\partial_i\sigma = (-1)^{i-j}\partial_j\sigma$. Altogether we get $\mathrm{rmv}(\partial\sigma) = 0$; thus $\partial\, \mathrm{rmv}(\sigma) = \mathrm{rmv}(\partial\sigma)$.

Assume that $\mathrm{rmv}(\sigma) = \sigma$. Since no edge of σ is a constant loop, of course also no edge of a face $\partial_i\sigma$ is a constant loop. If the image of $\partial_i\sigma$ is not contained in ∂Q, this implies $\mathrm{rmv}(\partial_i\sigma) = \partial_i\sigma = \partial_i\, \mathrm{rmv}(\sigma)$. If $\partial_i\sigma$ has image in ∂Q, then of course $[\partial_i\sigma] = [0] = [\partial_i\, \mathrm{rmv}(\sigma)]$, which implies $\mathrm{rmv}(\partial_i\sigma) = \partial_i\, \mathrm{rmv}(\sigma)$.

Now we prove that rmv sends relative fundamental cycles to relative fundamental cycles. Let $\sum_{j=1}^{r} a_j\tau_j$ be a straight relative cycle representing the relative homology class $[Q, \partial Q]$. We denote by $J_1 \subset \{1, \dots, r\}$ the indices of those τ_j which have a constant edge. The sum $\sum_{j\in J_1} a_j\tau_j$ is a relatively 0-homologous relative cycle. Indeed, each face of $\partial_i\tau_k$ not contained in ∂Q has to cancel against some face of some τ_l, because $\sum_{j=1}^{r} a_j\tau_j$ is a relative cycle. If $\partial_i\tau_k$ is degenerate, then necessarily $l \in J_1$. Moreover, if τ_k is degenerate and $\partial_i\tau_k$ is nondegenerate, it follows from the earlier part of the proof that $\partial_i\tau_k$ cancels against some $\partial_j\tau_k$.

Thus $\sum_{j\in J_1} a_j\tau_j$ represents some relative homology class. The isomorphism $H_n(C_*^{\text{sing}}(Q,\partial Q))\to\mathbb{R}$ is given by pairing with the volume form of an arbitrary Riemannian metric. After smoothing the relative cycle, we can apply Sard's lemma, and conclude that degenerate simplices have volume 0. Thus $\sum_{j\in J_1} a_j\tau_j$ is 0-homologous.

We denote by $J_2\subset\{1,\dots,r\}$ the indices of those τ_j which are contained in ∂Q. For $j\in J_2$ we have $[\tau_j]=[0]\in C_*^{\text{sing}}(Q,\partial Q)$.

Thus $\sum_{j\notin J_1\cup J_2} a_j\tau_j$ is another representative of the homology class $[Q,\partial Q]$. But, by Definition 5.14, it also represents $(\text{rmv})_*([Q,\partial Q])$. $\square$

Consider a subgroup $H\subset\Pi(K(A))$ for some $A\subset\partial Q$. For instance, $A=q(\partial_0 Q)$ in the setting of Construction 5.1, and $H=q_*(\Pi(K(\partial_0 Q)))\subset\Pi(K(A))$.

A 1-simplex e is a constant loop if and only if he is a constant loop for all $h\in H$. This implies that a simplex σ is degenerate if and only if $h\sigma$ is degenerate for all $h\sigma$. Moreover, H maps simplices in ∂Q to simplices in ∂Q. Thus $\text{rmv}(\sigma)=0$ if and only if $\text{rmv}(h\sigma)=0$ for all $h\in H$, that is, rmv is well defined on $C_*^{\text{simp,inf}}(HK^{\text{str}}(Q))\otimes_{\mathbb{Z}H}\mathbb{Z}$ for each subgroup H.

Lemma 5.16. *Assume that Q is a n-dimensional compact manifold with boundary ∂Q. Let the assumptions of Corollary 5.13 be satisfied. Then we can extend* rmv *to a well-defined chain map from $(HK^{\text{str}}(Q),HK^{\text{str}}(\partial_1 Q))\otimes_{\mathbb{Z}H}\mathbb{Z}$ to itself by defining*

$$\text{rmv}(\sigma\otimes z)=\begin{cases}0 & \textit{if } \text{rmv}(\sigma)=0,\\ \sigma\otimes z & \textit{else}.\end{cases}$$

Moreover, if $\sum_{j\in J} a_j\tau_j\otimes 1\in C_*^{\text{simp,inf}}(HK^{\text{str}}(Q),HK^{\text{str}}(\partial_1 Q))\otimes_{\mathbb{Z}H}\mathbb{Z}$ represents the image of $[Q,\partial Q]\otimes 1$, then $\sum_{\in J} a_j\,\text{rmv}(\tau_j\otimes 1)$ represents the image of $[Q,\partial Q]\otimes 1$.

Proof. Well-definedness of rmv follows from the remark before Lemma 5.16. The same proof as for Lemma 5.15 shows that rmv is a chain map.

If $\sum_{j=1}^r a_j\tau_j$ represents $[Q,\partial Q]$, the second claim follows from Lemma 5.15. If $\sum_{j\in J} a_j\tau_j\otimes 1$ is homologous to $\sum_{i=1}^s b_i\kappa_i\otimes 1$ and $\sum_{i=1}^s b_i\kappa_i$ represents $[Q,\partial Q]$, then, because rmv is a chain map, $\text{rmv}\big(\sum_{j\in J} a_j\tau_j\otimes 1\big)$ and $\text{rmv}\big(\sum_{i=1}^s b_i\kappa_i\otimes 1\big)$ are homologous, which implies the second claim. $\square$

The proof of Theorem 1.1 will pursue the idea of straightening a given cycle in such a way that many simplices either become weakly degenerate or will have an edge in $\partial_0 Q$. In the first case, they will disappear after application of rmv. In the second case, they disappear in view of the following observation, which is a variant of an argument used in [Gromov 1982].

Lemma 5.17. (a) *Let Assumption I be satisfied for a manifold Q and consider the action of $G=\Pi(K(\partial_0 Q))$ on $K(Q)$. Let $\sigma\in K(Q)$ be a simplex. If*

$\operatorname{str}(\sigma)$ *has an edge in* $\partial_0 Q$, *then*

$$\operatorname{str}(\sigma \otimes 1) = 0 \in C_*^{\text{simp,inf}}(K(Q)) \otimes_{\mathbb{Z}G} \mathbb{Z}.$$

(b) *If* $q : Q \to Q$ *is given by Construction 5.1,* $H = q_*(G)$, *and* $\sigma \in K(Q)$ *is a simplex such that* $q(\operatorname{str}(\sigma))$ *has an edge in* $q(\partial_0 Q)$, *then*

$$q(\operatorname{str}(\sigma \otimes 1)) = 0 \in C_*^{\text{simp,inf}}(K(Q)) \otimes_{\mathbb{Z}H} \mathbb{Z}.$$

Proof. (a) Let γ be the edge of $\operatorname{str}(\sigma)$ with image in $\partial_0 Q$. Then $g = \{\gamma, \overline{\gamma}\}$ is an element of $G = \Pi(K(\partial_0 Q))$ and $g \operatorname{str}(\sigma) = \overline{\operatorname{str}(\sigma)}$. In the simplicial chain complex $C_*^{\text{simp,inf}}(K(Q))$, one has $\overline{\operatorname{str}(\sigma)} = -\operatorname{str}(\sigma)$. Thus $g \operatorname{str}(\sigma) = -\operatorname{str}(\sigma)$, which implies $\operatorname{str}(\sigma \otimes 1) = \operatorname{str}(\sigma) \otimes 1 = 0$.

(b) Let γ be the edge of $q(\operatorname{str}(\sigma))$ with image in $q(\partial_0 Q)$. Let γ' be the corresponding edge of $\operatorname{str}(\sigma)$. Let $g = \{\gamma', \overline{\gamma}'\} \in G$ and $h = q_*(g) = \{\gamma, \overline{\gamma}\} \in H$. The same argument as in (a) shows $hq(\operatorname{str}(\sigma)) = -q(\operatorname{str}(\sigma))$. □

6. Proof of Main Theorem

As discussed in the introduction, before tackling the proof of Theorem 1.1 in full generality, we prove some particular cases as motivation.

Example 6.1. M is a connected, orientable, hyperbolic n-manifold, F is an orientable, geodesic $(n-1)$-submanifold, and $Q = \overline{M - F}$. For simplicity we assume that M and F are closed; thus Q is a hyperbolic manifold with geodesic boundary $\partial_1 Q \neq \varnothing$, and $\partial_0 Q = \varnothing$.

Outline of proof that $\|M\|_F^{\text{norm}} \geq \|\partial Q\|/(n+1)$. Start with a fundamental cycle $\sum_{i=1}^r a_i \sigma_i$ of M such that $\sigma_1, \dots, \sigma_r$ are normal to F. Since we want to consider laminations without isolated leaves, we replace F by a trivially foliated product neighborhood $\mathcal{F}$. We can assume after a suitable homotopy that each component of $\sigma_i^{-1}(\partial Q)$ either contains no vertex of Δ^n or consists of exactly one vertex, and that each vertex of Δ^n belongs to $\sigma_i^{-1}(\mathcal{F})$, for $i = 1, \dots, r$.

Each $\sigma_i^{-1}(Q)$ consists of polytopes, which can be further triangulated (without introducing new vertices) in a coherent way (i.e., such that boundary cancellations between different σ_i's will remain) into $\tau_{i1}, \dots, \tau_{is(i)}$.

The sum $\sum_{i=1}^r a_i(\tau_{i1} + \dots + \tau_{is(i)})$ is a relative fundamental cycle for Q.

For each σ_i, preimages of the boundary leaves of $\mathcal{F}$ cut Δ^n into regions which we color with black (components of $\sigma_i^{-1}(\mathcal{F})$) and white (components of $\sigma_i^{-1}(Q)$). If $\sigma_i^{-1}(\partial Q)$ contains vertices, these vertices are colored black. This is a canonical coloring (Definition 4.5).

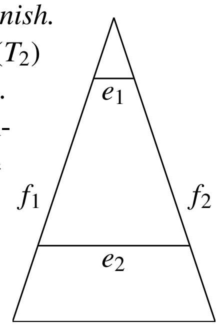
The edges of the simplices $\tau_{i,j}$ fall into two classes: "old edges", i.e., subarcs of edges of σ_i, and "new edges", which are contained in the interior of some subsimplex of σ_i of dimension ≥ 2.

We label the edges of τ_{ij} in such a way that old edges are labeled 1 and new edges are labeled 0. This is an admissible labeling (Definition 5.10). With this labeling, we apply the straightening procedure[8] from Section 5 to get a straight cycle $\sum_{i=1}^{r} a_i(\mathrm{str}(\tau_{i1}) + \cdots + \mathrm{str}(\tau_{is(i)}))$. (Thus old edges are straightened to distinguished 1-simplices.)

After straightening we apply the map rmv from Section 5D to remove all weakly degenerate simplices (simplices contained in ∂Q or having a constant edge). By Lemma 5.15, this does not change the homology class. In particular, the boundary of the relative cycle, $\partial \sum_{i,j} a_i\, \mathrm{rmv}(\mathrm{str}(\tau_{ij}))$ still represents the fundamental class $[\partial Q]$ of ∂Q.

Claim. *For each σ_i, after straightening there remain at most $n+1$ faces of nondegenerate simplices* $\mathrm{str}(\tau_{ij})$ *contributing to* $\partial \sum_{i,j} a_i\, \mathrm{rmv}(\mathrm{str}(\tau_{ij}))$.

Proof of claim. In view of Lemma 4.13, it suffices to show a subclaim: *If, for a fixed i, the faces $T_1 = \partial_{k_1}\tau_{ij_1}$ and $T_2 = \partial_{k_2}\tau_{ij_2}$ of τ_{ij_1} and τ_{ij_2} have a white-parallel arc* (Definition 4.9), *then* $\mathrm{rmv}(\mathrm{str}(\tau_{ij_1}))$ *and* $\mathrm{rmv}(\mathrm{str}(\tau_{ij_2}))$ *vanish.* In particular the corresponding straightened faces[9] $\mathrm{str}(T_1)$, $\mathrm{str}(T_2)$ do not occur (with nonzero coefficient) in $\partial \sum_{i,j} \mathrm{rmv}(\mathrm{str}(\tau_{ij}))$.



To prove the subclaim, let W be the white region of Δ^n containing T_1 and T_2 in its boundary. By the assumption of the subclaim, there is a white square bounded by two arcs $e_1 \subset T_1$, $e_2 \subset T_2$ and two arcs f_1, f_2 which are subarcs of edges of Δ^n. (The square is a formal sum of two triangles, U_1+U_2, which are 2-dimensional faces of some τ_{ij}'s.)

We want to show that all edges of $\mathrm{str}(\tau_{ij_1})$ belong to $S_1^{\mathrm{str}}(\partial Q)$. Note that $T_1, T_2 \subset \partial W$ are mapped to ∂Q. Let $x_1, x_2 \in S_0^{\mathrm{str}}(Q)$ be the unique elements of $S_0^{\mathrm{str}}(Q)$ in the same connected component C_1, C_2 of ∂Q as $\sigma_i(T_1)$ and $\sigma_i(T_2)$, respectively. In particular $\partial_0\, \mathrm{str}(e_1) = x_1 = \partial_1\, \mathrm{str}(e_1)$ and $\partial_0\, \mathrm{str}(e_2) = x_2 = \partial_1\, \mathrm{str}(e_2)$. Thus e_1 and e_2 are straightened to loops $\mathrm{str}(e_1)$ and $\mathrm{str}(e_2)$ based at x_1 and x_2, respectively. The straightenings $\mathrm{str}(f_1)$, $\mathrm{str}(f_2)$ of the other two arcs connect x_1 to x_2, and they are distinguished 1-simplices because they arise as straightenings of old edges. Thus $\mathrm{str}(f_1) = \mathrm{str}(f_2)$, by uniqueness of distinguished 1-simplices in each coset

[8]Under the assumptions of Example 6.1, straight simplices can be chosen to be the totally geodesic simplices with vertices in $S_0^{\mathrm{str}}(Q)$. Distinguished simplices are chosen according to Observation 5.9.

[9]For a subsimplex T of an affine subset $S \subset \Delta^n$ we get a singular simplex $\sigma_i|_T$ by restricting σ_i to T. We denote by $\mathrm{str}(T)$ the straightening of $\sigma_i|_T$.

$\Gamma K_1^{\mathrm{str}}(Q)\Gamma$ of $\Gamma = \Omega(\partial Q)$. This is why we have performed the straightening construction in Section 5 such that there should be only one distinguished 1-simplex, in each coset, for any given pair of connected components.

This means that the square is straightened to a cylinder.

But $(Q, \partial Q)$ is acylindrical; thus either both $\mathrm{str}(e_1)$ and $\mathrm{str}(e_2)$ are constant (in which case $\mathrm{rmv}(\mathrm{str}(\tau_{ij_1})) = \mathrm{rmv}(\mathrm{str}(\tau_{ij_2})) = 0$), or the cylinder must be homotopic into ∂Q. In the latter case, $\mathrm{str}(f_1)$ must be homotopic into, and therefore contained in, ∂Q. In particular, $\partial_0 \mathrm{str}(f_1)$ and $\partial_1 \mathrm{str}(f_1)$ belong to the same component of ∂Q. This implies $\partial_0 \mathrm{str}(f_1) = \partial_1 \mathrm{str}(f_1)$. Since $\mathrm{str}(f_1)$ is a distinguished 1-simplex, this implies that $\mathrm{str}(f_1)$ is constant.

Let P_1, P_2 be the affine planes whose intersections with Δ^n contain T_1 and T_2, respectively. There is an arc f_1 connecting $P_1 \cap \Delta^n$ to $P_2 \cap \Delta^n$ such that $\mathrm{str}(f_1)$ is contained in ∂Q. This implies that for each other arc f connecting $P_1 \cap \Delta^n$ to $P_2 \cap \Delta^n$ its straightening $\mathrm{str}(f)$ must be homotopic into, and therefore contained in, ∂Q.

If P_1 and P_2 are of the same type, then all edges of $\mathrm{str}(\tau_{ij_1})$ connect $P_1 \cap \Delta^n$ to $P_2 \cap \Delta^n$; hence all edges of $\mathrm{str}(\tau_{ij_1})$ belong to $S_1^{\mathrm{str}}(\partial Q)$. If P_1 and P_2 are not of the same type, the existence of a parallel arc implies that at least one of them, say P_1, must be of type $\{0a_1 \ldots a_k\}$ with $k \notin \{0, n-1\}$. Then, if P_3 is any other plane bounding W, it follows from Corollary 4.8 that P_3 has a white-parallel arc with P_1. Repeating the argument in the last paragraph with P_1 and P_3 in place of P_1 and P_2, we conclude that for each arc f connecting $P_1 \cap \Delta^n$ to $P_3 \cap \Delta^n$ its straightening $\mathrm{str}(f)$ must be homotopic into, and therefore contained in, ∂Q. Hence, for each τ_{ij_1} in the chosen triangulation of W, its 1-skeleton is straightened into ∂Q.

Since straight simplices σ (of dimension ≥ 2) with $\partial\sigma$ in the geodesic boundary ∂Q must be in ∂Q, this implies by induction that the k-skeleton of $\mathrm{str}(\tau_{ij_1})$ is in ∂Q for each k. In particular, $\mathrm{str}(\tau_{ij_1}) \in S_n^{\mathrm{str}}(\partial Q)$. Hence $\mathrm{rmv}(\mathrm{str}(\tau_{ij_1})) = 0$. This proves the subclaim.

By Lemma 4.13, the subclaim implies the claim. □

Since $\sum_{i=1}^r a_i \partial \sum_j \mathrm{rmv}(\mathrm{str}(\tau_{ij}))$ represents the fundamental class $[\partial Q]$, we conclude that $\|\partial Q\| \leq (n+1)\sum_{i=1}^r |a_i|$, as desired. □

The simplifications of Example 6.1 in comparison to the general proof below are essentially all due to the fact that $\partial_0 Q = \varnothing$. In the next example, if F is not geodesic, then $Q \neq N$ and thus $\partial_0 Q \neq \varnothing$ (even though ∂M and ∂F are both empty). Thus the generalization to $\partial_0 Q \neq \varnothing$ would be necessary even if one only wanted to consider closed manifolds M and F.

Example 6.2. M is a connected, closed, hyperbolic 3-manifold, $F \subset M$ a closed, incompressible surface, $N = \overline{M - F}$, $Q = \mathrm{Guts}(N)$.

Outline of proof that $\|M\|_F^{\mathrm{norm}} \geq \frac{1}{4}\|\partial Q\|$. Start with a fundamental cycle $\sum_{i=1}^r a_i\sigma_i$ of M, such that $\sigma_1, \dots, \sigma_r$ are normal to F. As in Example 6.1 we get a relative fundamental cycle $\sum_{i=1}^r a_i(\tau_{i1} + \cdots + \tau_{is(i)})$ of N. We cannot apply the argument from Example 6.1 to N because N is not acylindrical. Therefore we would like to work with a relative fundamental cycle for the acylindrical manifold Q.

N is aspherical. Using Lemma 3.2, we can assume that all τ_{ij} belong to $K(N)$. Then we can apply the retraction r from Lemma 3.5. Since r is only defined after tensoring with $\mathbb{Z}$ over $\mathbb{Z}G$, we get $r(\tau_{ij} \otimes 1) = \kappa_{ij} \otimes 1$ with $\kappa_{ij} \in K(Q)$ only determined up to choosing one κ_{ij} in its G-orbit.

Since Q is aspherical, we have $K(Q) = \widehat{K}(Q)$, that is, the κ_{ij} can be considered as simplices in Q and we can apply Lemma 3.6(b) to obtain a fundamental cycle for ∂Q.

The rest of the proof basically boils down to copying the proof of Example 6.1, with τ_{ij} replaced by κ_{ij}; but taking care of the ambiguity in the choice of κ_{ij}. The details can be found in the full-fledged proof of the theorem we're about to give. □

Proof of Theorem 1.1. The theorem is trivially true if $n = 1$. Hence we assume $n \geq 2$.

If $\partial_1 Q$ is empty, the equality $\partial Q = \partial_0 Q$ and the amenability of $\pi_1\partial_0 Q$ would imply $\|\partial Q\| = 0$, and Theorem 1.1 would be trivially true. Hence we assume $\partial_1 Q \neq \varnothing$. In particular, Q satisfies Assumption I from Section 5.

Consider a relative cycle $\sum_{i=1}^r a_i\sigma_i$, representing $[M, \partial M]$, such that $\sigma_1, \dots, \sigma_r$ are normal to $\mathcal{F}$. Our aim is to show that

$$\sum_{i=1}^r |a_i| \geq \frac{1}{n+1}\|\partial Q\|.$$

Define

$$N = \overline{M - \mathcal{F}}.$$

Since each σ_i is normal to $\mathcal{F}$, we have for each $i = 1, \dots, r$ that, after application of a simplicial homeomorphism $h_i : \Delta^n \to \Delta^n$, the image of $\sigma_i^{-1}(N)$ consists of polytopes, which can each be further triangulated in a coherent way (i.e., such that boundary cancellations between different σ_i's will remain) into simplices θ_{ij}, with $j \in \hat{J}_i$. (It is possible that $|\hat{J}_i| = \infty$, because N may be noncompact.) We choose these triangulations of the $\sigma_i^{-1}(N)$ to be minimal (Definition 4.9); that is, we do not introduce new vertices. (Compatible minimal triangulations of the $\sigma_i^{-1}(N)$ do exist: one starts with common minimal triangulations of the common faces and extends them to minimal triangulations of each polytope.)

Because boundary cancellations are preserved, we see that $\sum_{i=1}^r a_i \sum_{j\in\hat{J}_i} \theta_{ij}$ is a countable (possibly infinite) relative cycle representing the fundamental class $[N, \partial N]$ in the sense of Section 3B.

We fix a sufficient set of cancellations $\mathscr{C}^M$ for the relative cycle $\sum_{i=1}^r a_i\sigma_i$, in the sense of Definition 5.5. This induces a sufficient set of cancellations $\mathscr{C}^N$ for the relative cycle $\sum_{i=1}^r \sum_{j\in \hat{J}_i} a_i\theta_{ij}$.

If ∂M is a leaf of $\mathscr{F}$, then all faces of z contributing to ∂z are contained in ∂N. We call these faces exterior faces. We can assume that, for each i,

– each component of $\sigma_i^{-1}(\partial N)$ either contains no vertex of Δ^n, or consists of exactly one vertex, or consists of an exterior face, and

– and each vertex of Δ^n belongs to $\sigma_i^{-1}(\mathscr{F})$.

Indeed, by a small homotopy of the relative fundamental cycle $\sum_{i=1}^r a_i\sigma_i$, preserving normality, we can obtain that no component of $\sigma_i^{-1}(\partial N)$ contains a vertex of Δ^n, except for exterior faces. Afterwards, if some vertices of $\sum_{i=1}^r a_i\sigma_i$ do not belong to $\mathscr{F}$, we may homotope a small neighborhood of the vertex, until the vertex (and no other point of the neighborhood) meets ∂N. This, of course, preserves normality to $\mathscr{F}$.

Since each σ_i is normal to $\mathscr{F}$, in particular each σ_i is normal to the union of boundary leaves

$$\partial_1 N := \overline{\partial N - (\partial M \cap \partial N)}.$$

Thus for each σ_i, after application of a simplicial homeomorphism $h_i : \Delta^n \to \Delta^n$, the image of $\sigma_i^{-1}(\partial_1 N)$ consists of a (possibly infinite) set

$$Q_1, Q_2, \ldots \subset \Delta^n,$$

such that

$$Q_i = P_i \cap \Delta^n$$

for some affine hyperplanes $P_1, P_2, \ldots$. We define a coloring by declaring that (images under h_i of) components of

$$\sigma_i^{-1}(\mathrm{int}(N)) := \sigma_i^{-1}(N - \partial_1 N)$$

are colored white and (images under h_i of) components of $\sigma_i^{-1}(\mathscr{F})$ are colored black. (In particular, all Q_i are colored black.) Since we assume that all vertices of Δ^n belong to $\sigma_i^{-1}(\mathscr{F})$, and since each boundary leaf is adjacent to at least one component of $\sigma_i^{-1}(\mathrm{int}(N))$, this is a canonical coloring (Definition 4.5).

By Lemma 3.2(a), we can homotope the relative cycle $\sum_{i=1}^r \sum_{j\in \hat{J}_i} \theta_{ij}$, which belongs to $C_n^{\mathrm{inf}}(N, \partial N)$, to a relative cycle

$$\sum_{i=1}^r a_i \sum_{j\in \hat{J}_i} \hat{\theta}_{ij}$$

such that each $\hat{\theta}_{ij}$ is a simplex of $\widehat{K}(N)$, as defined in Section 3B, and such that the boundary $\partial \sum_{i=1}^{r} \sum_{j \in \hat{J}_i} \theta_{ij}$ is homotoped into $\widehat{K}(\partial N)$. Then consider

$$\sum_{i=1}^{r} \sum_{j \in \hat{J}_i} a_i \tau_{ij} := \sum_{i=1}^{r} \sum_{j \in \hat{J}_i} a_i \, p(\hat{\theta}_{ij}) \in C_n^{\text{simp,inf}}(K(N)),$$

where $p : \widehat{K}(N) \to K(N)$ is the projection defined at the end of Section 3B, and $\tau_{ij} := p(\hat{\theta}_{ij})$ for all i, j.

Consider $Q \subset N$ as in the assumptions of Theorem 1.1. We define

$$G := \Pi(K(\partial_0 Q)).$$

We have by assumption that $N = Q \cup R$ is an essential decomposition (as defined in the introduction), which means exactly that the assumptions of Lemma 3.5 are satisfied. Thus, according to Lemma 3.5, there exists a retraction

$$r : C_n^{\text{simp,inf}}(K(N)) \otimes_{\mathbb{Z}G} \mathbb{Z} \to C_n^{\text{simp,inf}}(K(Q)) \otimes_{\mathbb{Z}G} \mathbb{Z}$$

for $n \geq 2$, mapping $C_n^{\text{simp,inf}}(GK(\partial N)) \otimes_{\mathbb{Z}G} \mathbb{Z}$ to $C_n^{\text{simp,inf}}(GK(\partial_1 Q)) \otimes_{\mathbb{Z}G} \mathbb{Z}$, such that, for each simplex $\tau_{ij} \in K(N)$, we either have $r(\tau_{ij} \otimes 1) = 0$ or

$$r(\tau_{ij} \otimes 1) = \kappa_{ij} \otimes 1$$

for some simplex $\kappa_{ij} \in K(Q)$. (Recall that we've assumed that $n \geq 2$.) Thus

$$r\Big(\sum_{i=1}^{r} a_i \sum_{j \in \hat{J}_i} \tau_{ij} \otimes 1 \Big) = \sum_{i=1}^{r} a_i \sum_{j \in J_i} \kappa_{ij} \otimes 1,$$

with $J_i \subset \hat{J}_i$ for all i. (It may still be possible that $|J_i| = \infty$.) We remark that κ_{ij} is only determined up to choosing one κ_{ij} in its G-orbit.

Since r is a chain map, we get a sufficient set of cancellations for $\sum_{i=1}^{r} a_i \sum_{j \in J_i} \kappa_{ij} \otimes 1$ by setting

$$\mathscr{C}^Q := \big\{ (\partial_k \kappa_{i_1 j_1} \otimes 1, \partial_l \kappa_{i_2 j_2} \otimes 1) : (\partial_k \tau_{i_1 j_1}, \partial_l \tau_{i_2 j_2}) \in \mathscr{C}^N \big\}.$$

By assumption, Q is aspherical. We can therefore apply Lemma 3.6 and get

$$\partial \Big(\sum_{i=1}^{r} a_i \sum_{j \in J_i} \kappa_{ij} \otimes 1 \Big) \in C_*^{\text{simp,inf}}(GK(\partial_1 Q)) \otimes_{\mathbb{Z}G} \mathbb{Z}$$

represents (the image of) $[\partial Q] \otimes 1$.

Lemma 3.4(a) gives that G is amenable. Together with Lemma 3.7 this implies

$$\|\partial Q\| \leq \sum_{i=1}^{r} |a_i| (n+1) |J_i|.$$

In the remainder of the proof, we will use Lemma 5.16 to improve this inequality, getting rid of the unspecified (possibly infinite) numbers $|J_i|$.

$Q, \partial Q, \partial_0 Q, \partial_1 Q$ satisfy Assumption I (page 135). Thus there exists a simplicial set

$$K_*^{\mathrm{str}}(Q) \subset S_*(Q)$$

satisfying conditions (i)–(viii) from Lemma 5.3, and a set

$$D \subset K_1^{\mathrm{str}}(Q)$$

of distinguished 1-simplices (Definition 5.8).

Recall that, for each i,

$$\sum_{j \in \hat{J}_i} \theta_{i,j}$$

was defined by choosing a triangulation of $\sigma_i^{-1}(N)$. The simplices $\theta_{i,j}$ thus have "old edges", i.e., subarcs of edges of σ_i, and "new edges", whose interior is contained in the interior of some subsimplex of σ_i of dimension ≥ 2.

Associated to $z = \sum_{i=1}^{r} a_i \sum_{j \in \hat{J}_i} \theta_{ij}$ and $\mathscr{C}^N$ (and an arbitrary minimal presentation of ∂z) are, by Definition 5.5, simplicial sets $\Upsilon^N, \partial\Upsilon^N$.

The only possibility that two old edges have a vertex in Υ^N in common is that this vertex is a vertex of σ_i.

So the labeling of edges of $\sum_{i=1}^{r} a_i \sum_{j \in \hat{J}_i} \theta_{ij}$ by labeling old edges not containing a vertex of any σ_i with label 1 and all other edges with label 0 is an admissible labeling (Definition 5.10).

Associated to

$$w = \sum_{i=1}^{r} a_i \sum_{j \in J_i} \kappa_{ij} \otimes 1$$

and $\mathscr{C}^Q$ (and an arbitrary minimal presentation of ∂w) there are simplicial sets $\Upsilon, \partial\Upsilon$. By our definition of $\mathscr{C}^Q$, Υ is isomorphic to a simplicial subset of Υ^N, namely to the subset generated by the set

$$\{\tau \in \Upsilon^N : r(\tau \otimes 1) \neq 0\}$$

together with all iterated faces and degenerations. In particular, the admissible 0-1 labeling of Υ^N induces an admissible 0-1 labeling of Υ.

By Construction 5.1, there is a map of triples $q : (Q, \partial Q, \partial_1 Q) \to (Q, \partial Q, \partial_1 Q)$ which is (as a map of triples) homotopic to the identity, and such that $q(\partial_0 Q \cap C)$ is path-connected for each path component C of ∂Q.

We define

$$A := q(\partial_0 Q), \quad H := q_*(G) = q_*(\Pi(K(\partial_0 Q))) \subset \Pi(K(A)).$$

We observe that H is a quotient of G, hence amenable, even though $\Pi(K(A))$ need not be amenable.

Let $\widehat{\Upsilon}$, $\partial\widehat{\Upsilon}$ be defined by Observation 5.12. By Corollary 5.13, there is a chain map

$$q \circ \mathrm{str} : C_*^{\mathrm{simp,inf}}(\widehat{\Upsilon}) \otimes_{\mathbb{Z}G} \mathbb{Z} \to C_*^{\mathrm{simp,inf}}(HK^{\mathrm{str}}(Q)) \otimes_{\mathbb{Z}H} \mathbb{Z},$$

mapping $C_*^{\mathrm{simp,inf}}(\partial\widehat{\Upsilon}) \otimes_{\mathbb{Z}G} \mathbb{Z}$ to $C_*^{\mathrm{simp,inf}}(HK^{\mathrm{str}}(\partial_1 Q)) \otimes_{\mathbb{Z}H} \mathbb{Z}$ such that

$$\partial \sum_{i=1}^{r} a_i \sum_{j\in J_i} q(\mathrm{str}(\kappa_{ij})) \otimes 1$$

represents (the image of) $[\partial Q] \otimes 1$ and such that 1-labeled edges are mapped to distinguished 1-simplices. (We keep in mind that κ_{ij} is only determined up to G-action; thus $q(\mathrm{str}(\kappa_{ij}))$ is determined only up to choosing one simplex in its H-orbit.)

We then apply Lemma 5.16 to get the cycle

$$\partial \sum_{i=1}^{r} a_i \sum_{j\in J_i} \mathrm{rmv}(q(\mathrm{str}(\kappa_{ij})) \otimes 1) \in C_*^{\mathrm{simp,inf}}(HK^{\mathrm{str}}(\partial_1 Q)) \otimes_{\mathbb{Z}H} \mathbb{Z}$$

representing (the image of) $[\partial Q] \otimes 1$. We want to show that this is actually a finite chain of l^1-norm at most

$$(n+1) \sum_{i=1}^{r} |a_i|.$$

Claim. *For each i,*

$$\partial \sum_{j\in J_i} \mathrm{rmv}(q(\mathrm{str}(\kappa_{ij})) \otimes 1)$$

is the formal sum of at most $n+1$ $(n-1)$-simplices $L \otimes 1$ with coefficient 1.

Proof. This is a consequence of the following subclaim and Lemma 4.13:

Assume that for some fixed $i \in I$, for the chosen triangulation

$$\sigma_i^{-1}(N) = \bigcup_{j\in \hat{J}_i} \theta_{ij}$$

and the associated canonical coloring, there exist $j_1, j_2 \in \hat{J}_i$ and $k_1, k_2 \in \{0, \dots, n\}$ such that the faces

$$T_1 = \partial_{k_1}\theta_{ij_1} \in S_{n-1}(\partial N), \quad T_2 = \partial_{k_2}\theta_{ij_2} \in S_{n-1}(\partial N)$$

have a white-parallel arc (Definition 4.9). *Then*

$$\mathrm{rmv}(q(\mathrm{str}(\kappa_{ij_1})) \otimes 1) = 0, \quad \mathrm{rmv}(q(\mathrm{str}(\kappa_{ij_2})) \otimes 1) = 0.$$

To prove the subclaim, note first that

$$\partial_{k_l}\theta_{ij_l} \in S_{n-1}(\partial N)$$

implies (by Lemma 3.5 and Construction 5.1)

$$\partial_{k_l} q(\mathrm{str}(\kappa_{ij_l})) \in HK_*^{\mathrm{str}}(\partial_1 Q)$$

for $l = 1, 2$. Now assume (for a contradiction) that

$$\mathrm{rmv}(q(\mathrm{str}(\kappa_{ij_1})) \otimes 1) \neq 0.$$

By the subclaim's hypothesis, there are white-parallel arcs e_1, e_2 of T_1 and T_2, respectively. This means that there are arcs e_1, e_2 in a 2-dimensional subsimplex $\tau^2 \subset \Delta^n$ of the standard simplex, and subarcs f_1, f_2 of some edge of τ^2, all satisfying

$$\partial_0 e_1 = \partial_1 f_2, \quad \partial_0 f_2 = \partial_0 e_2, \quad \partial_1 e_2 = \partial_0 f_1, \quad \partial_1 f_1 = \partial_1 e_1$$

and such that e_1, f_2, e_2, f_1 bound a square in the boundary of a white component. (See figure on page 154. We will use the same letter for an affine subset of Δ^n and for the singular simplex obtained by restricting σ_i to this subset.) The square is of the form $U_1 + U_2$, where U_1, U_2 are $(n-2)$-fold iterated faces of some θ_{ij}'s. Hence $\partial U_1 = e_1 + f_2 + \partial_2 U_1$ and $\partial U_2 = -e_2 - f_1 - \partial_2 U_1$, in other words,

$$\partial(U_1 + U_2) = e_1 + f_2 - e_2 - f_1 \quad \text{and} \quad \partial_2 U_1 = -\partial_2 U_2.$$

We emphasize that we assume e_1 and e_2 to be edges of θ_{ij_1} and θ_{ij_2}, respectively, but f_1, f_2 need not be edges of θ_{ij_1} or θ_{ij_2}.

Notational remark. For each iterated face $f = \partial_{k_1} \dots \partial_{k_l} \theta_{ij}$ with $i \in I$, $j \in J_i$, we will denote by f' the $(n-l)$-simplex with

$$f' \otimes 1 = \partial_{k_1} \dots \partial_{k_l} \kappa_{ij} \otimes 1 = r(\partial_{k_1} \dots \partial_{k_l} \tau_{ij} \otimes 1) = r(\partial_{k_1} \dots \partial_{k_l} p(\hat{\theta}_{ij}) \otimes 1).$$

(The last two equations are true because r, p and the homotopy from $\sum_{i,j} a_i \theta_{ij}$ to $\sum_{i,j} \hat{\theta}_{ij}$ are chain maps.) In other words, if f is an iterated face of some τ_{ij}, then f' is, up to the ambiguity by the H-action, the corresponding iterated face of κ_{ij}.

By Lemma 3.5 we have $e_1', e_2' \in GK(\partial_1 Q)$. Thus we can (and will) choose $\kappa_{ij_1}, \kappa_{ij_2}$ in their G-orbits in such a way that $e_1', e_2' \in K(\partial_1 Q)$, which implies that $\mathrm{str}(e_1'), \mathrm{str}(e_2') \in K^{\mathrm{str}}(\partial_1 Q)$.

Since r, p and the homotopy are chain maps, we have

$$\partial_2 U_1' \otimes 1 = -\partial_2 U_2' \otimes 1.$$

That is,

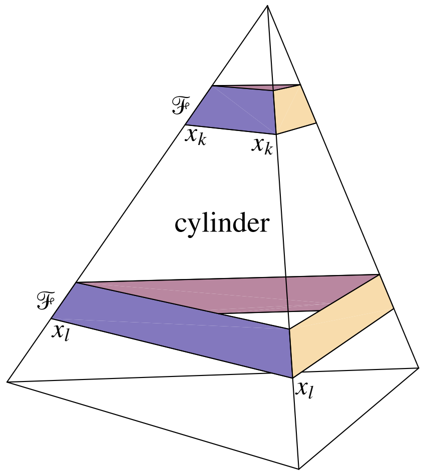
$$\partial_2 U_1' = g \overline{\partial_2 U_2'}$$

for some $g \in G$.

Since U_1' and U_2' belong to different κ_{ij}'s, say κ_{ij_1} and κ_{ij_2}, we can, upon replacing κ_{ij_2} by $g\kappa_{ij_2}$, assume that

$$\partial_2 U_1' = \overline{\partial_2 U_2'},$$

that is, $U_1' + U_2'$ is a square. (Since g maps $\partial e_2'$ to $\partial e_1'$, this second choice of κ_{ij_2} in its G-orbit preserves the condition that $e_2' \in K^{\mathrm{str}}(\partial_1 Q)$.)



Let F and F' be the path components of $\partial_1 Q$ such that $e_1' \subset F$ and $e_2' \subset F'$. Then $\partial_1 \operatorname{str}(f_1'), \partial_0 \operatorname{str}(f_2') \in F$ and $\partial_0 \operatorname{str}(f_1'), \partial_1 \operatorname{str}(f_2') \in F'$.

We note that f_1' and f_2' are edges with label 1. By condition (i) of Corollary 5.13, this implies that $\operatorname{str}(f_1')$ and $\operatorname{str}(f_2')$ are distinguished 1-simplices.

By conditions (ix) and Condition (xiii) of Definition 5.8 we have

$$\partial_1 q(\operatorname{str}(f_1')) = x_{E_0^F} = \partial_0 q(\operatorname{str}(f_2')), \partial_0 q(\operatorname{str}(f_1')) = x_{E_0^{F'}} = \partial_1 q(\operatorname{str}(f_2')).$$

That is, $q(\operatorname{str}(e_1'))$ and $q(\operatorname{str}(e_2'))$ are loops in $\partial_1 Q$, based respectively at $x_{E_0^F}$ and $x_{E_0^{F'}}$.

Since the square $q(\operatorname{str}(U_1' + U_2'))$ realizes a homotopy between $q(\operatorname{str}(f_1'))$ and $q(\operatorname{str}(f_2'))$, we have

$$q(\operatorname{str}(f_1')) = \gamma_1 q(\operatorname{str}(f_2'))\gamma_2$$

with

$$\gamma_1 = q(\operatorname{str}(e_1')), \gamma_2 = q(\operatorname{str}(e_2')) \in \Omega(\partial_1 Q) \subset \Gamma = \Omega(\partial Q).$$

By condition (x) from Definition 5.8 this implies

$$q(\operatorname{str}(f_1')) = q(\operatorname{str}(f_2')).$$

This means that $q(\operatorname{str}(U_1')) + q(\operatorname{str}(U_2'))$ is a cylinder with the boundary circles $q(\operatorname{str}(e_1'))$ and $q(\operatorname{str}(e_2'))$ in $\partial_1 Q$.

(This is why we have performed the straightening construction in Section 5 in such a way that there should be only one distinguished 1-simplex in each coset.)

The assumption $\operatorname{rmv}(q\circ\operatorname{str}(\kappa_{ij_1})\otimes 1)\neq 0$ made at the top of page 161 implies that the loops $q(\operatorname{str}(e_1'))$ and $q(\operatorname{str}(e_2'))$ are not 0-homotopic. Indeed, if one of them is 0-homotopic (and thus constant), so is the other, because they are homotopic through the cylinder. But $q(\operatorname{str}(e_1'))$ and $q(\operatorname{str}(e_2'))$ are edges of $q(\operatorname{str}(\kappa_{ij_1}))$ and $q(\operatorname{str}(\kappa_{ij_2}))$, respectively. In particular, $q(\operatorname{str}(\kappa_{ij_1}))$ and $q(\operatorname{str}(\kappa_{ij_2}))$ then have a constant loop as an edge. By Lemma 5.16 and Definition 5.2, this implies $\operatorname{rmv}(q\circ\operatorname{str}(\kappa_{ij_1})\otimes 1)=0$.

Thus we can assume that $q(\operatorname{str}(e_1'))$ and $q(\operatorname{str}(e_2'))$ are not 0-homotopic, that is, the cylinder

$$q(\operatorname{str}(U_1'))+q(\operatorname{str}(U_2'))$$

is π_1-injective as a map of pairs. Since $(Q, \partial_1 Q)$ is a pared acylindrical manifold, the cylinder must then be homotopic into ∂Q, as a map of pairs

$$(\mathbb{S}^1\times[0,1], \mathbb{S}^1\times\{0,1\})\to(Q,\partial_1 Q).$$

Since $\partial_1 Q$ is acylindrical, the cylinder must then either degenerate (that is, $\mathbb{S}^1\times[0,1]\to\partial Q$ homotopes to a map that factors over the projection $\mathbb{S}^1\times[0,1]\to\mathbb{S}^1$; in particular, $q(\operatorname{str}(e_1'))=q(\operatorname{str}(e_2')))$ or be homotopic into $\partial_0 Q$ (and hence into $q(\partial_0 Q)$, since $q\sim\mathrm{id}$). In the second case the vertices $x_{E_0^F}, x_{E_0^{F'}}$ must belong to $\partial_0 Q$ and we get by condition (vii) from Lemma 5.3 that $q(\operatorname{str}(e_1'))$ and $q(\operatorname{str}(e_2'))$ lie in $K_1^{\mathrm{str}}(\partial_0 Q)$. By Lemma 5.17 this implies that $q(\operatorname{str}(\kappa_{ij_1}))\otimes 1$ and $q(\operatorname{str}(\kappa_{ij_2}))\otimes 1$ vanish.

Thus we can assume that the cylinder degenerates. In particular, $q(\operatorname{str}(f_1'))$ and $q(\operatorname{str}(f_2'))$ lie in $K_1^{\mathrm{str}}(\partial_1 Q)$.

Let P_1, P_2 be the affine planes whose intersections with Δ^n contain T_1 and T_2, respectively. Let W be the white component whose boundary contains the white-parallel arcs of T_1, T_2. We have seen that there are arcs f_1, f_2 connecting $P_1\cap\Delta^n$ to $P_2\cap\Delta^n$ such that

$$q(\operatorname{str}(f_1')),\ q(\operatorname{str}(f_2'))\in K_1^{\mathrm{str}}(\partial_1 Q).$$

This implies that for each other arc f connecting $P_1\cap\Delta^n$ to $P_2\cap\Delta^n$ the straightening $q(\operatorname{str}(f'))$ must be (homotopic into — and therefore, by condition (vii) from Lemma 5.3), contained in — $\partial_1 Q$.

If P_1 and P_2 are of the same type (Definition 4.1), this shows that for all arcs $f\subset W$ we have

$$q(\operatorname{str}(f'))\in K_1^{\mathrm{str}}(\partial_1 Q)$$

If P_1 and P_2 are not of the same type, then the existence of a parallel arc implies that at least one of them, say P_1, must be of type $\{0a_1\dots a_k\}$ with $k\notin\{0,n-1\}$. Then, for each plane $P_3\neq P_1$ with $P_3\cap\Delta^n\subset\partial W$, it follows from Corollary 4.8 that $P_3\cap\Delta^n$ has a white-parallel arc with $P_1\cap\Delta^n$. Thus, repeating the argument with P_1 and P_3 in place of P_1 and P_2, we prove that there are arcs in $\partial_1 Q$ connecting $P_1\cap\Delta^n$

to $P_3 \cap \Delta^n$, and consequently for each arc $f \subset W$ connecting $P_1 \cap \Delta^n$ to $P_3 \cap \Delta^n$, the straightening $\mathrm{str}(f')$ must be homotopic into, and thus contained in, $\partial_1 Q$.

Consequently, also for arcs connecting $P_2 \cap \Delta^n$ to $P_3 \cap \Delta^n$, we conclude that $q(\mathrm{str}(f'))$ must be homotopic into, and therefore contained in, $\partial_1 Q$. This finally shows that the 1-skeleta of $q(\mathrm{str}(\kappa_{ij_1}))$ and $q(\mathrm{str}(\kappa_{ij_2}))$ belong to $K_1^{\mathrm{str}}(\partial_1 Q)$. By the π_1-injectivity of $\partial_1 Q \to Q$, the asphericity of $K(\partial_1 Q)$, and condition (vii) from Lemma 5.3, this implies that the 2-skeleta of $q(\mathrm{str}(\kappa_{ij_1}))$ and $q(\mathrm{str}(\kappa_{ij_2}))$ belong to $K_1^{\mathrm{str}}(\partial_1 Q)$. Inductively, if the k-skeleta of $q(\mathrm{str}(\kappa_{ij_1}))$ and $q(\mathrm{str}(\kappa_{ij_2}))$ belong to $K_k^{\mathrm{str}}(\partial_1 Q)$, then by the asphericity of $K(Q)$ and $K(\partial_1 Q)$ together with condition (vii) from Lemma 5.3 we obtain that the $k+1$-skeleta of $q(\mathrm{str}(\kappa_{ij_1}))$ and $q(\mathrm{str}(\kappa_{ij_2}))$ belong to $K_{k+1}^{\mathrm{str}}(\partial_1 Q)$. This provides the inductive step and thus our inductive proof shows that $q(\mathrm{str}(\kappa_{ij_1}))$ and $q(\mathrm{str}(\kappa_{ij_2}))$ belong to $K^{\mathrm{str}}(\partial_1 Q)$.

By Definitions 5.2 and 5.14 and Lemma 5.16 this implies

$$\mathrm{rmv}(q(\mathrm{str}(\kappa_{ij_1})) \otimes 1) = 0, \quad \mathrm{rmv}(q(\mathrm{str}(\kappa_{ij_2})) \otimes 1) = 0.$$

So we have shown the subclaim: if $T_1 = \partial_{k_1}\theta_{ij_1}$ and $T_2 = \partial_{k_2}\theta_{ij_2}$ have a white-parallel arc, then $\mathrm{rmv}(q(\mathrm{str}(\kappa_{ij_1})) \otimes 1) = 0$ and $\mathrm{rmv}(q(\mathrm{str}(\kappa_{ij_2})) \otimes 1) = 0$. In particular, $q(\mathrm{str}(T_1'))$ and $q(\mathrm{str}(T_2'))$ do not occur (with nonzero coefficient) in

$$\partial \sum_{j \in J_i} \mathrm{rmv}(q(\mathrm{str}(\kappa_{ij})) \otimes 1).$$

By Lemma 4.13, for a canonical coloring associated to a set of affine planes $P_1, P_2, \ldots$, and a fixed triangulation of each $Q_i = P_i \cap \Delta^n$, we have at most $n+1$ $(n-1)$-simplices whose 1-skeleton does not contain a white-parallel arc. This show that the subclaim implies the claim. □

The upshot is that we have presented $[\partial Q] \otimes 1$ as a finite chain of l^1-norm at most $(n+1)\sum_{i=1}^r |a_i|$. By Lemma 3.4(a) we know that $G = \Pi(K(\partial_0 Q))$ is amenable. Hence $H = q_*(G)$ is amenable. Thus Lemma 3.7, applied to $X = \partial Q$ and $K = HK^{\mathrm{str}}(\partial_1 Q)$ with its H-action, implies $\|\partial Q\| \le (n+1)\sum_{i=1}^r |a_i|$. This concludes the proof of Theorem 1.1. □

Theorem 1.1 is not true without assuming the amenability of $\pi_1 \partial_0 Q$. Counterexamples can be found, for example, using [Jungreis 1997] or [Kuessner 2003, Theorem 6.3].

Remark. In [Agol 1999], Theorem 1.1 has been proven for incompressible surfaces in hyperbolic 3-manifolds. We compare the steps of the proof in [Agol 1999] with the arguments in our paper:

Agol's step 1 is the normalization procedure, which we restated in Lemma 2.4.

Step 2 consists in choosing compatible triangulations of the polytopes $\sigma_i^{-1}(N)$.

Step 3 boils down to the statement that, for each component Q_i of Q, there exists a retraction $r : \hat{N} \to p^{-1}(Q_i)$, for the covering $p : \hat{N} \to N$ corresponding to

$\pi_1 Q_i$. Such a statement cannot be correct because it would (together with Agol's step 7) imply $\|N\| \geq \|Q\|$ whenever Q is a π_1-injective submanifold of N. This inequality is true for submanifolds with amenable boundary, but not in general. In fact, one only has the more complicated retraction

$$r : C_*(K(N), K(N')) \otimes_{\mathbb{Z}G} \mathbb{Z} \to C_*(K(Q), K(\partial Q)) \otimes_{\mathbb{Z}G} \mathbb{Z},$$

with $G = \Pi(K(\partial_0 Q))$. This is why much of the latter arguments become notationally awkward, although conceptually not much is changing. Moreover, the action of G is basically the reason why Theorem 1.1 is true only for amenable G.

Basically, the reason why the retraction $r : \hat{N} \to Q$ does not exist, is as follows. Let R_j be the connected components of $\hat{N} - p^{-1}(Q_i)$. Then R_j is homotopy equivalent to each connected component of ∂R_j. If ∂R_j were connected for each j, this homotopy equivalence could be extended to a homotopy equivalence $r : \hat{N} \to p^{-1}(Q_i)$. However, in most cases ∂R_j will be disconnected, and then such an r cannot exist.

We note that also the weaker construction of cutting off simplices does not work. A simplex may intersect Q_i in many components and it is not clear which component to choose.

Step 4 from Agol's proof puts a hyperbolic metric with geodesic boundary on Q.

His step 5 is the straightening procedure, corresponding to Sections 5B–5D in this paper. We remark that the straightening procedure must be slightly more complicated than in [Agol 1999] because it is not possible, as suggested in that same paper, to homotope all edges between boundary components of ∂Q into shortest geodesics. This is the reason why we can only straighten chains with an admissible 0-1 labeling of their edges (and why our straightening homomorphism in Section 5C is only defined on $C_*^{\mathrm{simp}}(|\Upsilon|)$ and not on all of $C_*^{\mathrm{sing}}(Q)$).

Agol's step 6 consists in removing degenerate simplices. This corresponds to Section 5E in this paper.

His step 7 proves that each triangle in $\sigma_i^{-1}(\partial N)$ contributes only once to the constructed fundamental cycle of ∂Q. Since, in our argument, we do not work with the covering $p : \hat{N} \to N$, we have no need for this justification.

His step 8 counts the remaining triangles per simplex (after removing degenerate simplices). It seems to have used the combinatorial arguments which we work out for arbitrary dimensions in Section 4.

We mention that the arguments of Section 4 are the only part of the proof which gets easier if one restricts to 3-manifolds rather than arbitrary dimensions. Moreover, the proof for laminations is the same as for hypersurfaces except for Lemma 2.4. Thus, upon these two points it seems that even in the case of incompressible surfaces in 3-manifolds the proof of Theorem 1.1 cannot be further simplified.

7. Specialization to 3-manifolds

Guts of essential laminations. We start by recalling the guts-terminology. Let M be a compact 3-manifold with (possibly empty) boundary consisting of incompressible tori, and $\mathcal{F}$ an essential lamination transverse or tangential to the boundary. $N = \overline{M - \mathcal{F}}$ is a, possibly noncompact, irreducible 3-manifold with incompressible, aspherical boundary ∂N. We set

$$\partial_0 N = \partial N \cap \partial M, \quad \partial_1 N = \overline{\partial N - \partial_0 N}.$$

(Thus $\partial_1 N$ is the union of boundary leaves of the lamination.) By the proof of [Gabai and Kazez 1998, Lemma 1.3], the noncompact ends of N are essential I-bundles over noncompact subsurfaces of $\partial_1 N$. After cutting off each of these ends along an essential, properly fibered annulus, one obtains a compact 3-manifold to which one can apply the JSJ-decomposition [Jaco and Shalen 1979; Johannson 1979]. Hence we have a decomposition of N into the characteristic submanifold Char(N) (which consists of I-bundles and Seifert fibered solid tori, where the fibrations have to respect boundary patterns as defined in [Johannson 1979, p. 83]) and the guts of N, Guts(N). The I-fibered ends of N will be added to the characteristic submanifold, which thus may become noncompact, while Guts(N) is compact. (We mention that there are different notions of guts in the literature. Our notion is compatible with that of [Agol 1999; Agol et al. 2007], but differs from the definition in [Gabai and Kazez 1998] or [Calegari and Dunfield 2003] by taking the Seifert fibered solid tori into the characteristic submanifold and not into the guts. Thus, solid torus guts in the paper of Calegari–Dunfield is the same as empty guts in our setting.) If $\partial_0 N \cap \partial Q \neq \varnothing$ consists of annuli $A_1, \ldots, A_k$, then, to be consistent with the setting of Theorem 1.1, we add components $A_i \times [0, 1]$ to Char(N) (without changing the homeomorphism type of N), which implies $\partial_0 N \cap \partial Q = \varnothing$.

For $Q = \text{Guts}(N)$ we set

$$\partial_1 Q = \partial_1 N \cap \partial Q = \partial N \cap \partial Q = Q \cap \partial N, \quad \partial_0 Q = \overline{\partial Q - \partial_1 Q}.$$

For $R = \text{Char}(N)$ we set

$$\partial_1 R = \partial N \cap \partial R, \quad \partial_0 R = \overline{\partial R - \partial_1 R}.$$

Then $\partial_0 N \cap \partial Q = \varnothing$ implies $\partial_0 Q = Q \cap R$.

$\partial_0 Q$ consists of essential tori and annuli; in particular $\pi_1 \partial_0 Q$ is amenable. The guts of N has the following properties: the pair $(Q, \partial_1 Q)$ is a pared acylindrical manifold (Definition 4.3), $Q, \partial_1 Q, \partial_1 R$ are aspherical, and the inclusions

$$\partial_0 Q \to Q, \quad \partial_1 Q \to Q, \quad Q \to N, \quad \partial_0 R \to R, \quad \partial_1 R \to R, \quad R \to N$$

are π_1-injective; see [Jaco and Shalen 1979; Johannson 1979]. It follows from Thurston's hyperbolization theorem for Haken manifolds that Q admits a hyperbolic metric with geodesic boundary $\partial_1 Q$ and cusps corresponding to $\partial_0 Q$. (In particular, we have $\chi(\partial Q) \leq 0$; thus ∂Q is aspherical, and $\partial_1 Q$ is a hyperbolic surface, thus acylindrical.)

Theorem 7.1. *Let M be a compact 3-manifold with (possibly empty) boundary consisting of incompressible tori, and let $\mathcal{F}$ be an essential lamination of M. Then*

$$\|M, \partial M\|_{\mathcal{F}}^{\mathrm{norm}} \geq -\chi(\mathrm{Guts}(\mathcal{F})).$$

More generally, if P is a polyhedron with f faces, then

$$\|M, \partial M\|_{\mathcal{F},P}^{\mathrm{norm}} \geq -\frac{2}{f-2}\chi(\mathrm{Guts}(\mathcal{F})).$$

Proof. Let $N = \overline{M - \mathcal{F}}$. Since $\mathcal{F}$ is essential, N is irreducible (hence aspherical, since $\partial N \neq \varnothing$) and has incompressible, aspherical boundary. Let $R = \mathrm{Char}(N)$ be the characteristic submanifold and $Q = \mathrm{Guts}(N)$ be the complement of the characteristic submanifold of N. The discussion before Theorem 7.1 shows that the decomposition $N = Q \cup R$ satisfies the assumptions of Theorem 1.1.

From the computation of the simplicial volume for surfaces [Gromov 1982, section 0.2] and $\chi(Q) = \frac{1}{2}\chi(\partial Q)$ (which is a consequence of Poincaré duality for the closed 3-manifold $Q \cup_{\partial Q} Q$), it follows that

$$-\chi(\mathrm{Guts}(\mathcal{F})) = -\tfrac{1}{2}\chi(\partial\,\mathrm{Guts}(\mathcal{F})) = \tfrac{1}{4}\|\partial\,\mathrm{Guts}(\mathcal{F})\|.$$

Thus, the first claim is obtained as application of Theorem 1.1 to $Q = \mathrm{Guts}(\mathcal{F})$.

The second claim — the generalization to arbitrary polyhedra — is obtained as in [Agol 1999]. Namely, one uses the same straightening as above, and asks again how many nondegenerate 2-simplices may, after straightening, occur in the intersection of ∂Q with some polyhedron P_i. In [Agol 1999, p. 11], it is shown that this number is at most $2f - 4$, where f is the number of faces of P_i. The same argument as above then shows that

$$\sum_{i=1}^{r} |a_i| \geq \frac{1}{2f-4}\|\partial\,\mathrm{Guts}(\mathcal{F})\|,$$

giving the wanted inequality. □

The following corollary applies, for example, to all hyperbolic manifolds obtained by Dehn-filling the complement of the figure-eight knot in $\mathbb{S}^3$. (It was proved in [Hatcher 1992] that each hyperbolic manifold obtained by Dehn-filling the complement of the figure-eight knot in $\mathbb{S}^3$ carries essential laminations.)

Corollary 7.1. *If M is a finite-volume hyperbolic manifold with* $\mathrm{Vol}(M) < 2V_3 = 2.02\ldots$, *then M carries no essential lamination $\mathcal{F}$ with*

$$\|M, \partial M\|^{\mathrm{norm}}_{\mathcal{F},P} = \|M, \partial M\|_P$$

for all polyhedra P, and nonempty guts. In particular, there is no tight essential lamination with nonempty guts.

Proof. The derivation of this corollary from Theorem 1.2 is exactly the same as in [Agol 1999] for the usual (nonlaminated) Gromov norm. Namely, by [Sleator et al. 1988] (or [Agol 1999], end of Section 6) there exists a sequence P_n of straight polyhedra in $\mathbb{H}^3$ with

$$\lim_{n\to\infty} \frac{\mathrm{Vol}(P_n)}{f_n - 2} = V_3,$$

where f_n denotes the number of faces of P_n. Assuming that M carries a lamination $\mathcal{F}$ with $\|M, \partial M\|^{\mathrm{norm}}_{\mathcal{F},P_n} = \|M, \partial M\|_{P_n}$ for all n, one gets

$$-\chi(\mathrm{Guts}(\mathcal{F})) \leq \frac{f_n - 2}{2}\|M, \partial M\|_{\mathcal{F},P_n} = \frac{f_n - 2}{2}\|M, \partial M\|_{P_n} \leq \frac{f_n - 2}{2}\frac{\mathrm{Vol}(M)}{\mathrm{Vol}(P_n)},$$

which tends to

$$\frac{\mathrm{Vol}(M)}{2V_3} < 1.$$

On the other hand, if $\mathrm{Guts}(\mathcal{F})$ is not empty, it is a hyperbolic manifold with nonempty geodesic boundary; hence $\chi(\mathrm{Guts}(\mathcal{F})) \leq -1$, giving a contradiction. □

Definition 7.2. The Weeks manifold is the closed 3-manifold obtained by applying $(-\frac{5}{1}, -\frac{5}{2})$-surgery at the Whitehead link [Rolfsen 1976, p. 68].

It is known that the Weeks manifold is hyperbolic and that its hyperbolic volume is approximately $0.94\ldots$. It is actually the hyperbolic 3-manifold of smallest volume.

Corollary 7.3 [Calegari and Dunfield 2003, Conjecture 9.7]. *The Weeks manifold admits no tight lamination $\mathcal{F}$.*

Proof. According to [Calegari and Dunfield 2003], the Weeks manifold cannot carry a tight lamination with empty guts. Since tight laminations satisfy $\|M\|^{\mathrm{norm}}_{\mathcal{F},P} = \|M\|$ for each polyhedron (see Lemma 2.4), and since the Weeks manifold has volume smaller than $2V_3$, it follows from Corollary 7.1 that it cannot carry a tight lamination with nonempty guts neither. □

The same argument shows that a hyperbolic 3-manifold M of volume less than $2V_3$ and such that there is no injective homomorphism $\pi_1 M \to \mathrm{Homeo}^+(\mathbb{S}^1)$ cannot carry a tight lamination, because it was shown in [Calegari and Dunfield 2003] that the existence of a tight lamination with empty guts implies the existence of an injective homomorphism $\pi_1 M \to \mathrm{Homeo}^+(\mathbb{S}^1)$. Some methods for excluding the existence of injective homomorphisms $\pi_1 M \to \mathrm{Homeo}^+(\mathbb{S}^1)$ have been developed in that same paper (which yielded in particular the nonexistence of such homomorphisms for the Weeks manifold, used in the corollary above), but in general it is still hard to apply this criterion to other hyperbolic 3-manifolds of volume $< 2V_3$.

As indicated in [Calegari 2003], an approach to a generalization of some of the above arguments to essential, nontight laminations, yielding possibly a proof for nonexistence of essential laminations on the Weeks manifold, could consist in trying to define a straightening of cycles (as in the proof of Lemma 2.4) upon possibly changing the essential lamination.

As a consequence of a result of Tao Li, one can at least exclude the existence of transversely orientable essential laminations on the Weeks manifold.

Corollary 7.4. *The Weeks manifold admits no transversely orientable essential lamination $\mathcal{F}$.*

Proof. According to [Li 2006, Theorem 1.1], if a closed, orientable, atoroidal 3-manifold M contains a transversely orientable essential lamination, then it contains a transversely orientable tight essential lamination. Hence Corollary 7.4 is a direct consequence of Corollary 7.3. □

8. Higher dimensions

We want to finish this paper showing that Theorem 1.1 is interesting also in higher dimensions. While in dimension 3 the assumptions of Theorem 1.1 hold for any essential lamination, it is likely that this will not be the case for many laminations in higher dimensions. However, the most straightforward, but already interesting, application of the inequality is Corollary 8.1, which means that, for a given negatively curved manifold M, we can give an explicit bound on the topological complexity of geodesic hypersurfaces. Such a bound seems to be new except, of course, in the 3-dimensional case, where it goes back to [Agol 1999] and (with no explicit constants) to [Hass 1995].

Corollary 8.1. *Let M be a compact Riemannian n-manifold of negative sectional curvature and finite volume. Let $F \subset M$ be a geodesic $(n-1)$-dimensional hypersurface of finite volume. Then*

$$\|F\| \leq \frac{n+1}{2}\|M\|.$$

Proof. Consider $N = \overline{M - F}$. $(N, \partial N)$ is acylindrical. This is well-known and can be seen as follows: assume that N contains an essential cylinder; then the double $DN = N \cup_{\partial_1 N} N$ contains an essential 2-torus. But, since N is a negatively curved manifold with geodesic boundary, we can glue the Riemannian metrics to get a complete negatively curved Riemannian metric on DN. In particular, DN contains no essential 2-torus, giving a contradiction.

Moreover, the geodesic boundary ∂N is π_1-injective and negatively curved, thus aspherical. Therefore we can choose $Q = N$, in which case the other assumptions of Theorem 1.1 are trivially satisfied. From Theorem 1.1 we conclude that

$$\|M\|_F^{\mathrm{norm}} \geq \frac{1}{n+1} \|\partial N\|.$$

The boundary of N consists of two copies of F, hence $\|\partial N\| = 2\|F\|$. The leaf space of $\widetilde{F} \subset \widetilde{M}$ is a Hausdorff tree; thus Lemma 2.4(b) implies $\|M\|_F^{\mathrm{norm}} = \|M\|$. The claim follows. $\square$

This statement should be read as follows: for a given manifold M (with given volume) one has an upper bound on the topological complexity of compact geodesic hypersurfaces.

For hyperbolic manifolds one can use the Chern–Gauss–Bonnet theorem and the proportionality principle to reformulate Corollary 8.1 as follows: If M is a closed hyperbolic n-manifold and F a closed $(n-1)$-dimensional geodesic hypersurface, then $\mathrm{Vol}(M) \geq C_n \chi(F)$ for a constant C_n depending only on n.

Acknowledgements

It is probably obvious that this paper is strongly influenced by Agol's preprint [Agol 1999]. The argument that a generalization of Agol's inequality would imply Corollary 7.3 is from [Calegari and Dunfield 2003].

References

[Agol 1999] I. Agol, "Lower bounds on volumes of hyperbolic Haken 3-manifolds", preprint, University of California, Davis, CA, 1999. arXiv math.GT/9906182

[Agol et al. 2007] I. Agol, P. A. Storm, and W. P. Thurston, "Lower bounds on volumes of hyperbolic Haken 3-manifolds", *J. Amer. Math. Soc.* **20**:4 (2007), 1053–1077. MR 2008i:53086 Zbl 1155.58016

[Benedetti and Petronio 1992] R. Benedetti and C. Petronio, *Lectures on hyperbolic geometry*, Springer, Berlin, 1992. MR 94e:57015 Zbl 0768.51018

[Brittenham 1995] M. Brittenham, "Essential laminations and Haken normal form", *Pac. J. Math.* **168**:2 (1995), 217–234. MR 96c:57028 Zbl 0838.57011

[Calegari 2000] D. Calegari, "The Gromov norm and foliations", *Geom. Funct. Anal.* **10**:6 (2000), 1423–1447. MR 2002c:57026a Zbl 0974.57015

[Calegari 2003] D. Calegari, "Problems in foliations and laminations of 3-manifolds", pp. 297–335 in *Topology and geometry of manifolds* (Athens, GA, 2001), edited by G. Matić and C. McCrory, Proc. Sympos. Pure Math. **71**, AMS, Providence, RI, 2003. MR 2005b:57053 Zbl 1042.57501

[Calegari and Dunfield 2003] D. Calegari and N. M. Dunfield, "Laminations and groups of homeomorphisms of the circle", *Invent. Math.* **152**:1 (2003), 149–204. MR 2005a:57013 Zbl 1025.57018

[Fenley 2007] S. R. Fenley, "Laminar free hyperbolic 3-manifolds", *Comment. Math. Helv.* **82**:2 (2007), 247–321. MR 2008g:57020 Zbl 1136.57015

[Francaviglia 2004] S. Francaviglia, "Hyperbolic volume of representations of fundamental groups of cusped 3-manifolds", *Int. Math. Res. Not.* **2004**:9 (2004), 425–459. MR 2004m:57032 Zbl 1088.57015

[Gabai 1999] D. Gabai, "Essential laminations and Kneser normal form", *J. Differ. Geom.* **53**:3 (1999), 517–574. MR 2001m:57025 Zbl 1028.57012

[Gabai and Kazez 1998] D. Gabai and W. H. Kazez, "Group negative curvature for 3-manifolds with genuine laminations", *Geom. Topol.* **2** (1998), 65–77. MR 99e:57023 Zbl 0905.57011

[Gabai and Oertel 1989] D. Gabai and U. Oertel, "Essential laminations in 3-manifolds", *Ann. Math.* (2) **130**:1 (1989), 41–73. MR 90h:57012 Zbl 0685.57007

[Gromov 1982] M. Gromov, "Volume and bounded cohomology", *Inst. Hautes Études Sci. Publ. Math.* **56** (1982), 5–99. MR 84h:53053 Zbl 0516.53046

[Hass 1995] J. Hass, "Acylindrical surfaces in 3-manifolds", *Mich. Math. J.* **42**:2 (1995), 357–365. MR 96c:57031 Zbl 0862.57011

[Hatcher 1992] A. Hatcher, "Some examples of essential laminations in 3-manifolds", *Ann. Inst. Fourier* (*Grenoble*) **42**:1-2 (1992), 313–325. MR 93e:57026 Zbl 0759.57006

[Ivanov 1985] N. V. Ivanov, "Foundations of the theory of bounded cohomology", *Zap. Nauchn. Semin. Leningr. Otd. Mat. Inst. Steklova* **143** (1985), 69–109. In Russian; translated in *J. Sov. Math.* **37** (1987), 1090–1114. MR 87b:53070 Zbl 0573.55007

[Jaco and Shalen 1979] W. H. Jaco and P. B. Shalen, *Seifert fibered spaces in* 3*-manifolds*, Mem. Amer. Math. Soc. **21**:220, Amer. Math. Soc., Providence, RI, 1979. MR 81c:57010 Zbl 0415.57005

[Johannson 1979] K. Johannson, *Homotopy equivalences of* 3*-manifolds with boundaries*, Lecture Notes in Mathematics **761**, Springer, Berlin, 1979. MR 82c:57005 Zbl 0412.57007

[Jungreis 1997] D. Jungreis, "Chains that realize the Gromov invariant of hyperbolic manifolds", *Ergod. Theory Dyn. Syst.* **17**:3 (1997), 643–648. MR 98c:57013 Zbl 0882.57009

[Kuessner 2003] T. Kuessner, "Efficient fundamental cycles of cusped hyperbolic manifolds", *Pac. J. Math.* **211**:2 (2003), 283–313. MR 2004j:57034 Zbl 1066.57026

[Kuessner 2004] T. Kuessner, "Foliated norms on fundamental group and homology", *Topol. Appl.* **144**:1-3 (2004), 229–254. MR 2005g:57056 Zbl 1054.57032

[Kuessner 2010] T. Kuessner, "Multicomplexes, bounded cohomology and additivity of simplicial volume", preprint, Universität Münster, 2010, available at http://www.math.uni-muenster.de/reine/u/kuessner/preprints/bc.pdf.

[Li 2006] T. Li, "Compression branched surfaces and tight essential laminations", preprint, Oklahoma State University, Stillwater, OK, 2006, available at https://www2.bc.edu/~taoli/tight.pdf.

[May 1967] J. P. May, *Simplicial objects in algebraic topology*, Van Nostrand Math. Studies **11**, Van Nostrand, Princeton, NJ, 1967. Reprinted University of Chicago Press, 1992. MR 36 #5942 Zbl 0165.26004

[Roberts et al. 2003] R. Roberts, J. Shareshian, and M. Stein, "Infinitely many hyperbolic 3-manifolds which contain no Reebless foliation", *J. Amer. Math. Soc.* **16**:3 (2003), 639–679. MR 2004e:57023 Zbl 1012.57022

[Rolfsen 1976] D. Rolfsen, *Knots and links*, Publish or Perish, Berkeley, CA, 1976. MR 58 #24236 Zbl 0339.55004

[Sleator et al. 1988] D. D. Sleator, R. E. Tarjan, and W. P. Thurston, "Rotation distance, triangulations, and hyperbolic geometry", *J. Amer. Math. Soc.* **1**:3 (1988), 647–681. MR 90h:68026 Zbl 0653.51017

[Thurston 1980] W. P. Thurston, “The geometry and topology of 3-manifolds”, lecture notes, Princeton University, 1980, available at http://library.msri.org/books/gt3m.

Received June 15, 2008.

THILO KUESSNER

kuessner@math.uni-muenster.de
MATHEMATISCHES INSTITUT
UNIVERSITÄT MÜNSTER
EINSTEINSTRASSE 62
D-48149 MÜNSTER
GERMANY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 1, 2011

CHERN NUMBERS AND THE INDICES OF SOME ELLIPTIC DIFFERENTIAL OPERATORS

PING LI

Libgober and Wood proved that the Chern number c_1c_{n-1} of a compact complex manifold of dimension n can be determined by its Hirzebruch χ_y-genus. Inspired by the idea of their proof, we show that, for compact, spin, almost-complex manifolds, more Chern numbers can be determined by the indices of some twisted Dirac and signature operators. As a byproduct, we get a divisibility result of certain characteristic number for such manifolds. Using our method, we give a direct proof of the result of Libgober and Wood, which was originally proved by induction.

1. Introduction and main results

Suppose (M, J) is a compact, almost-complex $2n$-manifold with a given almost complex structure J. This J makes the tangent bundle of M into a n-dimensional complex vector bundle T_M. Let $c_i(M, J) \in H^{2i}(M; \mathbb{Z})$ be the i-th Chern class of T_M. Suppose we have a formal factorization of the total Chern class as follows:

$$1 + c_1(M, J) + \cdots + c_n(M, J) = \prod_{i=1}^{n}(1 + x_i),$$

i.e., $x_1, \ldots, x_n$ are formal Chern roots of T_M. The Hirzebruch χ_y-genus of (M, J), $\chi_y(M, J)$, is defined by

$$\chi_y(M, J) = \left(\prod_{i=1}^{n} \frac{x_i(1 + ye^{-x_i})}{1 - e^{-x_i}}\right)[M].$$

Here $[M]$ is the fundamental class of the orientation of M induced by J and y is an indeterminate. If J is specified, we simply denote $\chi_y(M, J)$ by $\chi_y(M)$.

Research supported by Program for Young Excellent Talents in Tongji University.
I would like to express my sincere gratitude to my advisors, Boju Jiang and Haibao Duan, for their consistent guidance and encouragement during the past several years. I also thank the referees for their careful readings of an earlier version of this paper. Their comments have enhanced the quality of the paper.
MSC2000: primary 57R20; secondary 58J20.
Keywords: Chern number, index, Dirac operator, signature operator.

When the almost complex structure J is integrable (equivalently, when M is an n-dimensional compact complex manifold), $\chi_y(M)$ can be obtained by

$$\chi^p(M)=\sum_{q=0}^{n}(-1)^q h^{p,q}(M),\ \ \chi_y(M)=\sum_{p=0}^{n}\chi^p(M)\cdot y^p,$$

where $h^{p,q}(\,\cdot\,)$ is the Hodge number of type (p,q). This is given by the Hirzebruch–Riemann–Roch Theorem, proved in [Hirzebruch 1966] for projective manifolds and in [Atiyah and Singer 1968] in the general case.

The formula

$$\sum_{p=0}^{n}\chi^p(M)\cdot y^p=\Bigl(\prod_{i=1}^{n}\frac{x_i(1+ye^{-x_i})}{1-e^{-x_i}}\Bigr)[M] \tag{1-1}$$

implies that $\chi^p(M)$, the index of the Dolbeault complex, can be expressed as a rational combination of some Chern numbers of M. Conversely, we can address the following question.

Question 1.1. For an n-dimensional compact complex manifold M, given a partition $\lambda=\lambda_1\lambda_2\cdots\lambda_l$ of weight n, can the Chern number $c_{\lambda_1}c_{\lambda_2}\cdots c_{\lambda_l}[M]$ be determined by $\chi^p(M)$, or more generally by the indices of some other elliptic differential operators?

For the simplest case $c_n[M]$, the answer is affirmative and well-known [Hirzebruch 1966, Theorem 15.8.1]:

$$c_n[M]=\chi_y(M)|_{y=-1}=\sum_{p=0}^{n}(-1)^p\chi^p(M).$$

The next-to-simplest case is the Chern number $c_1c_{n-1}[M]$. The answer here is also affirmative, as was proved by Libgober and Wood [1990, pp. 141–143]:

$$\sum_{p=2}^{n}(-1)^p\binom{p}{2}\chi^p(M)=\frac{n(3n-5)}{24}c_n[M]+\frac{1}{12}c_1c_{n-1}[M]. \tag{1-2}$$

The idea of their proof is quite enlightening: expanding both sides of (1-1) at $y=-1$ and comparing the coefficients of the term $(y+1)^2$, one gets (1-2).

Inspired by this idea, in this paper we consider twisted Dirac operators and signature operators on compact, *spin, almost-complex* manifolds and show that the Chern numbers c_n, c_1c_{n-1}, $c_1^2c_{n-2}$ and c_2c_{n-2} of such manifolds can be determined by the indices of these operators.

Remark 1.2. Equation (1-2) was also obtained later by Salamon [1996, p. 144], who applied this result extensively to hyper-Kähler manifolds.

Let M be a compact, *almost-complex* $2n$-manifold. We still use $x_1, \dots, x_n$ to denote the corresponding formal Chern roots of the n-dimensional complex vector bundle T_M. Suppose E is a complex vector bundle over M. Set

$$\hat{A}(M, E) := \left(\mathrm{ch}(E) \cdot \prod_{i=1}^{n} \frac{x_i/2}{\sinh(x_i/2)} \right)[M],$$

$$L(M, E) := \left(\mathrm{ch}(E) \cdot \prod_{i=1}^{n} \frac{x_i(1+e^{-x_i})}{1-e^{-x_i}} \right)[M],$$

where $\mathrm{ch}(E)$ is the Chern character of E. The celebrated Atiyah–Singer index theorem [Hirzebruch et al. 1992, pp. 74–81] states that $L(M, E)$ is the index of the signature operator twisted by E and when M is *spin*, $\hat{A}(M, E)$ is the index of the Dirac operator twisted by E.

Definition 1.3. Set

$$A_y(M) := \sum_{p=0}^{n} \hat{A}(M, \Lambda^p(T_M^*)) \cdot y^p \quad \text{and} \quad L_y(M) := \sum_{p=0}^{n} L(M, \Lambda^p(T_M^*)) \cdot y^p,$$

where $\Lambda^p(T_M^*)$ denotes the p-th exterior power of the dual of T_M.

Our main result is the following:

Theorem 1.4. *Let M be a compact, almost-complex manifold.*

(1) $$\sum_{p=0}^{n} (-1)^p \hat{A}(M, \Lambda^p(T_M^*)) = c_n[M],$$

$$\sum_{p=1}^{n} (-1)^p \cdot p \cdot \hat{A}(M, \Lambda^p(T_M^*)) = \frac{1}{2}\big(nc_n[M] + c_1c_{n-1}[M]\big),$$

$$\sum_{p=2}^{n} (-1)^p \binom{p}{2} \hat{A}(M, \Lambda^p(T_M^*)) = \Big(\frac{n(3n-5)}{24}c_n + \frac{3n-2}{12}c_1c_{n-1} + \frac{1}{8}c_1^2c_{n-2}\Big)[M];$$

(2) $$\sum_{p=0}^{n} (-1)^p L(M, \Lambda^p(T_M^*)) = 2^n c_n[M],$$

$$\sum_{p=1}^{n} (-1)^p \cdot p \cdot L(M, \Lambda^p(T_M^*)) = 2^{n-1}\big(nc_n[M] + c_1c_{n-1}[M]\big),$$

$$\sum_{p=2}^{n} (-1)^p \binom{p}{2} L(M, \Lambda^p(T_M^*))$$
$$= 2^{n-2}\Big(\frac{n(3n-5)}{6}c_n + \frac{3n-2}{3}c_1c_{n-1} + c_1^2c_{n-2} - c_2c_{n-2}\Big)[M].$$

Corollary 1.5. (1) *The Chern numbers* $c_n[M]$, $c_1c_{n-1}[M]$ *and* $c_1^2c_{n-2}[M]$ *can be determined by* $A_y(M)$.

(2) *The characteristic numbers* $c_n[M]$, $c_1c_{n-1}[M]$ *and* $c_1^2c_{n-2}[M] - c_2c_{n-2}[M]$ *can be determined by* $L_y(M)$.

(3) *When* M *is a spin manifold, the Chern numbers* $c_n[M]$, $c_1c_{n-1}[M]$, $c_1^2c_{n-2}[M]$ *and* $c_2c_{n-2}[M]$ *can be expressed by using linear combinations of the indices of some twisted Dirac and signature operators.*

As remarked in [Libgober and Wood 1990, p. 143], it was shown by Milnor [1960] that every complex cobordism class contains a non-singular algebraic variety. Milnor also showed that two almost-complex manifolds are complex cobordant if and only if they have the same Chern numbers. Hence Libgober and Wood's result implies that the characteristic number

$$\frac{n(3n-5)}{24}c_n[N]+\frac{1}{12}c_1c_{n-1}[N]$$

is always an integer for any compact, *almost-complex* $2n$-manifold N.

Note that the right-hand side of the third equality in Theorem 1.4 is

$$\Big(\frac{n(3n-5)}{24}c_n[M]+\frac{1}{12}c_1c_{n-1}[M]\Big)+\frac{1}{8}\Big(2(n-1)c_1c_{n-1}[M]+c_1^2c_{n-2}[M]\Big).$$

Corollary 1.6. *For a compact, spin, almost-complex manifold* M, *the integer*

$$2(n-1)c_1c_{n-1}[M]+c_1^2c_{n-2}[M]$$

is divisible by 8.

Example 1.7. The total Chern class of the complex projective space $\mathbb{C}P^n$ is given by $c(\mathbb{C}P^n) = (1+g)^{n+1}$, where g is the standard generator of $H^2(\mathbb{C}P^n;\mathbb{Z}) \cong \mathbb{Z}$. $\mathbb{C}P^n$ is spin if and only if n is odd, as $c_1(\mathbb{C}P^n) = (n+1)g$. Suppose $n = 2k+1$. Then

$$2(n-1)c_1c_{n-1}[\mathbb{C}P^n]+c_1^2c_{n-2}[\mathbb{C}P^n] = 8(k+1)^2\big(k(2k+1)+\tfrac{1}{3}k(k+1)(2k+1)\big).$$

It is easy to check that $\mathbb{C}P^4$ does not satisfy this divisibility result.

2. Proof of the main result

Lemma 2.1. *Let* M *be a compact, almost-complex manifold. Then*:

$$A_y(M) = \bigg(\prod_{i=1}^{n}\Big(\frac{x_i(1+ye^{-x_i(1+y)})}{1-e^{-x_i(1+y)}}\cdot e^{-x_i(1+y)/2}\Big)\bigg)[M],$$

$$L_y(M) = \bigg(\prod_{i=1}^{n}\Big(\frac{x_i(1+ye^{-x_i(1+y)})}{1-e^{-x_i(1+y)}}\cdot (1+e^{-x_i(1+y)})\Big)\bigg)[M].$$

Proof. From

$$c(T_M) = \prod_{i=1}^{n} (1 + x_i)$$

we have (see, for example, [Hirzebruch et al. 1992, p. 11])

$$c(\Lambda^p(T_M^*)) = \prod_{1 \le i_1 < \cdots < i_p \le n} \big(1 - (x_{i_1} + \cdots + x_{i_p})\big).$$

Hence

$$ch(\Lambda^p(T_M^*))y^p = \sum_{1 \le i_1 < \cdots < i_p \le n} e^{-(x_{i_1} + \cdots + x_{i_p})} y^p = \sum_{1 \le i_1 < \cdots < i_p \le n} \bigg(\prod_{j=1}^{p} y e^{-x_{i_j}} \bigg).$$

Therefore

$$\sum_{p=0}^{n} ch(\Lambda^p(T_M^*))y^p = \sum_{p=0}^{n} \bigg(\sum_{1 \le i_1 < \cdots < i_p \le n} \bigg(\prod_{j=1}^{p} y e^{-x_{i_j}} \bigg) \bigg) = \prod_{i=1}^{n} (1 + y e^{-x_i}).$$

So

$$\begin{aligned} A_y(M) &= \sum_{p=0}^{n} \hat{A}(M, \Lambda^p(T_M^*)) \cdot y^p \\ &= \bigg(\bigg(\sum_{p=0}^{n} ch(\Lambda^p(T_M^*))y^p \bigg) \cdot \prod_{i=1}^{n} \frac{x_i/2}{\sinh(x_i/2)} \bigg)[M] \\ &= \bigg(\prod_{i=1}^{n} \bigg((1 + y e^{-x_i}) \cdot \frac{x_i/2}{\sinh(x_i/2)} \bigg) \bigg)[M] \\ &= \bigg(\prod_{i=1}^{n} \bigg(\frac{x_i(1 + y e^{-x_i})}{1 - e^{-x_i}} \cdot e^{-x_i/2} \bigg) \bigg)[M]. \end{aligned} \tag{2-1}$$

Since for the evaluation only the homogeneous component of degree n in the x_i enters, then we obtain an additional factor $(1 + y)^n$ if we replace x_i by $x_i(1 + y)$ in (2-1). We therefore obtain

$$\begin{aligned} A_y(M) &= \bigg(\frac{1}{(1+y)^n} \prod_{i=1}^{n} \bigg(\frac{x_i(1+y)(1 + y e^{-x_i(1+y)})}{1 - e^{-x_i(1+y)}} \cdot e^{-x_i(1+y)/2} \bigg) \bigg)[M] \\ &= \bigg(\prod_{i=1}^{n} \bigg(\frac{x_i(1 + y e^{-x_i(1+y)})}{1 - e^{-x_i(1+y)}} \cdot e^{-x_i(1+y)/2} \bigg) \bigg)[M]. \end{aligned}$$

Similarly,

$$
\begin{aligned}
L_y(M) &= \Bigg(\prod_{i=1}^{n}\bigg((1+ye^{-x_i})\cdot\frac{x_i(1+e^{-x_i})}{1-e^{-x_i}}\bigg)\Bigg)[M] \\
&= \Bigg(\frac{1}{(1+y)^n}\prod_{i=1}^{n}\bigg(\frac{x_i(1+y)(1+ye^{-x_i(1+y)})}{1-e^{-x_i(1+y)}}\cdot(1+e^{-x_i(1+y)})\bigg)\Bigg)[M] \\
&= \Bigg(\prod_{i=1}^{n}\bigg(\frac{x_i(1+ye^{-x_i(1+y)})}{1-e^{-x_i(1+y)}}\cdot(1+e^{x_i(1+y)})\bigg)\Bigg)[M]. \qquad \square
\end{aligned}
$$

Lemma 2.2. *Set $z=1+y$. We have*

$$
\begin{aligned}
A_y(M) &= \Bigg(\prod_{i=1}^{n}\bigg((1+x_i)-(x_i+\tfrac{1}{2}x_i^2)z+\big(\tfrac{11}{24}x_i^2+\tfrac{1}{8}x_i^3\big)z^2+\cdots\bigg)\Bigg)[M], \\
L_y(M) &= \Bigg(\prod_{i=1}^{n}\bigg(2(1+x_i)-(2x_i+x_i^2)z+\big(\tfrac{7}{6}x_i^2+\tfrac{1}{2}x_i^3\big)z^2+\cdots\bigg)\Bigg)[M].
\end{aligned}
$$

Proof.

$$
\begin{aligned}
\frac{x_i(1+ye^{-x_i(1+y)})}{1-e^{-x_i(1+y)}} &= -x_iy+\frac{x_i(1+y)}{1-e^{-x_i(1+y)}} = -x_i(z-1)+\frac{x_iz}{1-e^{-x_iz}} \\
&= -x_i(z-1)+\big(1+\tfrac{1}{2}x_iz+\tfrac{1}{12}x_i^2z^2+\cdots\big) \\
&= (1+x_i)-\tfrac{1}{2}x_iz+\tfrac{1}{12}x_i^2z^2+\cdots.
\end{aligned}
$$

So we have

$$
\begin{aligned}
A_y(M) &= \Bigg(\prod_{i=1}^{n}\frac{x_i(1+ye^{-x_i(1+y)})}{1-e^{-x_i(1+y)}}\cdot e^{-x_i(1+y)/2}\Bigg)[M] \\
&= \Bigg(\prod_{i=1}^{n}\big((1+x_i)-\tfrac{1}{2}x_iz+\tfrac{1}{12}x_i^2z^2+\cdots\big)\bigg(1-\tfrac{1}{2}x_iz+\tfrac{1}{8}x_i^2z^2+\cdots\bigg)\Bigg)[M] \\
&= \Bigg(\prod_{i=1}^{n}\big((1+x_i)-(x_i+\tfrac{1}{2}x_i^2)z+\big(\tfrac{11}{24}x_i^2+\tfrac{1}{8}x_i^3\big)z^2+\cdots\big)\Bigg)[M].
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
L_y(M) &= \Bigg(\prod_{i=1}^{n}\bigg(\frac{x_i(1+ye^{-x_i(1+y)})}{1-e^{-x_i(1+y)}}\cdot(1+e^{-x_i(1+y)})\bigg)\Bigg)[M] \\
&= \Bigg(\prod_{i=1}^{n}\big((1+x_i)-\tfrac{1}{2}x_iz+\tfrac{1}{12}x_i^2z^2+\cdots\big)\big(2-x_iz+\tfrac{1}{2}x_i^2z^2+\cdots\big)\Bigg)[M] \\
&= \Bigg(\prod_{i=1}^{n}\big(2(1+x_i)-(2x_i+x_i^2)z+\big(\tfrac{7}{6}x_i^2+\tfrac{1}{2}x_i^3\big)z^2+\cdots\big)\Bigg)[M]. \qquad \square
\end{aligned}
$$

Let $f(x_1,\dots,x_n)$ be a symmetric polynomial in $x_1,\dots,x_n$. Then $f(x_1,\dots,x_n)$ can be expressed in terms of $c_1,\dots,c_n$ in a unique way. We use $h(f(x_1,\dots,x_n))$ to denote the homogeneous component of degree n in $f(x_1,\dots,x_n)$. For instance, when $n=3$,

$$\begin{aligned}h(x_1+x_2+x_3+x_1^2x_2+x_1^2x_3+x_2^2x_1+x_2^2x_3+x_3^2x_1+x_3^2x_2)&\\ =x_1^2x_2+x_1^2x_3+x_2^2x_1+x_2^2x_3+x_3^2x_1+x_3^2x_2&\\ =(x_1+x_2+x_3)(x_1x_2+x_1x_3+x_2x_3)-3x_1x_2x_3=c_1c_2-3c_3.&\end{aligned}$$

The next lemma is a crucial technical ingredient in the proof of our main result.

Lemma 2.3.

(1) $h_1:=h\Big(\sum_{i=1}^n\Big(x_i\prod_{j\neq i}(1+x_j)\Big)\Big)=nc_n.$

(2) $h_{11}:=h\Big(\sum_{1\le i<j\le n}\Big(x_ix_j\prod_{k\neq i,j}(1+x_k)\Big)\Big)=\dfrac{n(n-1)}{2}c_n.$

(3) $h_2:=h\Big(\sum_{i=1}^n\Big(x_i^2\prod_{j\neq i}(1+x_j)\Big)\Big)=-nc_n+c_1c_{n-1}.$

(4) $h_{12}:=h\Big(\sum_{1\le i<j\le n}\Big((x_i^2x_j+x_ix_j^2)\prod_{k\neq i,j}(1+x_k)\Big)\Big)=(n-2)(-nc_n+c_1c_{n-1}).$

(5) $h_{22}:=h\Big(\sum_{1\le i<j\le n}\Big(x_i^2x_j^2\prod_{k\neq i,j}(1+x_k)\Big)\Big)=\dfrac{n(n-3)}{2}c_n-(n-2)c_1c_{n-1}+c_2c_{n-2}.$

(6) $h_3:=h\Big(\sum_{i=1}^n\Big(x_i^3\prod_{j\neq i}(1+x_j)\Big)\Big)=nc_n-c_1c_{n-1}+c_1^2c_{n-2}-2c_2c_{n-2}.$

Now we can complete the proof of Theorem 1.4; we postpone the proof of Lemma 2.3 to the end of this section.

Proof. From Lemma 2.2, the constant term in $A_y(M)$ is

$$\Big(\prod_{i=1}^n(1+x_i)\Big)[M]=c_n[M].$$

The coefficient of z is

$$\begin{aligned}\Big(\sum_{i=1}^n\big(-\big(x_i+\tfrac12x_i^2\big)\prod_{j\neq i}(1+x_j)\big)\Big)[M]&=\big(-h_1-\tfrac12h_2\big)[M]\\ &=-\tfrac12\big(nc_n[M]+c_1c_{n-1}[M]\big).\end{aligned}$$

The coefficient of z^2 is

$$\Big(\sum_{i=1}^{n}\Big(\big(\tfrac{11}{24}x_i^2+\tfrac{1}{8}x_i^3\big)\prod_{j\neq i}(1+x_j)\Big)+\sum_{1\leq i<j\leq n}\Big((x_i+\tfrac{1}{2}x_i^2)(x_j+\tfrac{1}{2}x_j^2)\prod_{k\neq i,j}(1+x_k)\Big)\Big)[M]$$
$$=\big(\tfrac{11}{24}h_2+\tfrac{1}{8}h_3+h_{11}+\tfrac{1}{2}h_{12}+\tfrac{1}{4}h_{22}\big)[M]$$
$$=\Big(\frac{n(3n-5)}{24}c_n+\frac{3n-2}{12}c_1c_{n-1}+\frac{1}{8}c_1^2c_{n-2}\Big)[M].$$

Similarly, for $L_y(M)$, the constant term is

$$\Big(2^n\prod_{i=1}^{n}(1+x_i)\Big)[M]=2^nc_n[M].$$

The coefficient of z is

$$\Big(\sum_{i=1}^{n}\Big(-(2x_i+x_i^2)\prod_{j\neq i}2(1+x_j)\Big)\Big)[M]=(-2^nh_1-2^{n-1}h_2)[M]$$
$$=-2^{n-1}(nc_n[M]+c_1c_{n-1}[M]).$$

The coefficient of z^2 is

$$\Big(\sum_{i=1}^{n}\Big(\big(\tfrac{7}{6}x_i^2+\tfrac{1}{2}x_i^3\big)\prod_{j\neq i}2(1+x_j)\Big)+\sum_{1\leq i<j\leq n}\Big((2x_i+x_i^2)(2x_j+x_j^2)\prod_{k\neq i,j}2(1+x_k)\Big)\Big)[M]$$
$$=\Big(\frac{7\cdot 2^{n-2}}{3}h_2+2^{n-2}h_3+2^nh_{11}+2^{n-1}h_{12}+2^{n-2}h_{22}\Big)[M]$$
$$=2^{n-2}\Big(\frac{n(3n-5)}{6}c_n+\frac{3n-2}{3}c_1c_{n-1}+c_1^2c_{n-2}-c_2c_{n-2}\Big)[M].\quad\square$$

Proof of Lemma 2.3. In the following proof, $\hat{x_i}$ means deleting x_i. Parts (1) and (2) are quite obvious. For (3),

$$h_2=\sum_{i=1}^{n}\Big(h\Big(x_i^2\prod_{j\neq i}(1+x_j)\Big)\Big)=\sum_{i=1}^{n}\Big(x_i\sum_{j\neq i}x_1\cdots\hat{x_j}\cdots x_n\Big)=\sum_{i=1}^{n}(x_ic_{n-1}-c_n)$$
$$=-nc_n+c_1c_{n-1}.$$

For (4),

$$h_{12}=\sum_{1\leq i<j\leq n}\Big(h\Big((x_i^2x_j+x_ix_j^2)\prod_{k\neq i,j}(1+x_k)\Big)\Big)$$
$$=\sum_{1\leq i<j\leq n}\Big((x_i+x_j)\sum_{k\neq i,j}x_1\cdots\hat{x_k}\cdots x_n\Big)$$
$$=(n-2)\sum_{i=1}^{n}\Big(x_i\sum_{k\neq i}x_1\cdots\hat{x_k}\cdots x_n\Big)=(n-2)h_2=(n-2)(-nc_n+c_1c_{n-1}).$$

For (5),

$$
\begin{aligned}
c_2c_{n-2} &= \Big(\sum_{1\le i<j\le n} x_ix_j\Big)\Big(\sum_{1\le k<l\le n} x_1\cdots\hat{x}_k\cdots\hat{x}_l\cdots x_n\Big)\\
&= \sum_{1\le i<j\le n}\Big(x_ix_j\sum_{1\le k<l\le n} x_1\cdots\hat{x}_k\cdots\hat{x}_l\cdots x_n\Big)\\
&= \sum_{1\le i<j\le n}\Big(x_1x_2\cdots x_n + (x_i^2x_j + x_ix_j^2)\sum_{k\ne i,j} x_1\cdots\hat{x}_k\cdots\hat{x}_i\cdots\hat{x}_j\cdots x_n\\
&\qquad\qquad + x_i^2x_j^2\sum_{\substack{1\le k<l\le n\\ k\ne i,j\\ l\ne i,j}} x_1\cdots\hat{x}_k\cdots\hat{x}_l\cdots\hat{x}_i\cdots\hat{x}_j\cdots x_n\Big)\\
&= \frac{n(n-1)}{2}c_n + h_{12} + h_{22}.
\end{aligned}
$$

Therefore,

$$
h_{22} = c_2c_{n-2} - \frac{n(n-1)}{2}c_n - h_{12} = \frac{n(n-3)}{2}c_n - (n-2)c_1c_{n-1} + c_2c_{n-2}.
$$

For (6),

$$
\begin{aligned}
&(c_1^2-2c_2)c_{n-2}\\
&= \Big(\sum_{i=1}^n x_i^2\Big)\Big(\sum_{1\le j<k\le n} x_1\cdots\hat{x}_j\cdots\hat{x}_k\cdots x_n\Big) = \sum_{i=1}^n\Big(x_i^2\sum_{1\le j<k\le n} x_1\cdots\hat{x}_j\cdots\hat{x}_k\cdots x_n\Big)\\
&= \sum_{i=1}^n\Big(\Big(x_i^3\sum_{\substack{1\le j<k\le n\\ j\ne i\\ k\ne i}} x_1\cdots\hat{x}_j\cdots\hat{x}_k\cdots\hat{x}_i\cdots x_n) + (x_i^2\sum_{k\ne i} x_1\cdots\hat{x}_k\cdots\hat{x}_i\cdots x_n\Big)\Big)\\
&= h_3 + h_2.
\end{aligned}
$$

Hence $h_3 = (c_1^2-2c_2)c_{n-2} - h_2 = nc_n - c_1c_{n-1} + c_1^2c_{n-2} - 2c_2c_{n-2}$. □

3. Concluding remarks

Libgober and Wood's proof [1990, p. 142, Lemma 2.2] of (1-2) is by induction. Here, using our method, we can give a quite direct proof. We have shown that

$$
\begin{aligned}
\chi_y(M) &= \Big(\prod_{i=1}^n \frac{x_i(1+ye^{-x_i(1+y)})}{1-e^{-x_i(1+y)}}\Big)[M]\\
&= \Big(\prod_{i=1}^n\big((1+x_i) - \tfrac{1}{2}x_iz + \tfrac{1}{12}x_i^2z^2 + \cdots\big)\Big)[M].
\end{aligned}
$$

The coefficient of z^2 is

$$\Big(\sum_{i=1}^{n}\big(\frac{1}{12}x_i^2\prod_{j\neq i}(1+x_j)\big)+\sum_{1\leq i<j\leq n}\Big(\frac{1}{4}x_ix_j\prod_{k\neq i,j}(1+x_k)\Big)\Big)[M]$$

$$=\big(\tfrac{1}{12}h_2+\tfrac{1}{4}h_{11}\big)[M]=\frac{n(3n-5)}{24}c_n[M]+\frac{1}{12}c_1c_{n-1}[M].$$

It is natural to ask what the coefficients are for higher-order terms $(y+1)^p$, for $p\geq 3$. Unfortunately the coefficients become very complicated for such terms. In [Libgober and Wood 1990, pp. 144–145] there is a detailed remark on the coefficients of the higher-order terms of $\chi_y(M)$. Note that the expression of $A_y(M)$ (resp. $L_y(M)$) has an additional factor $e^{-x_i(1+y)/2}$ (resp. $1+e^{x_i(1+y)}$) relative to than that of $\chi_y(M)$. Hence we cannot expect that there are *explicit* expressions of higher-order coefficients similar to Theorem 1.4.

References

[Atiyah and Singer 1968] M. F. Atiyah and I. M. Singer, "The index of elliptic operators, III", *Ann. of Math.* (2) **87** (1968), 546–604. MR 38 #5245 Zbl 0164.24301

[Hirzebruch 1966] F. Hirzebruch, *Topological methods in algebraic geometry*, vol. 131, 3rd ed., Grundlehren der Math. Wiss., Springer, New York, 1966. MR 34 #2573 Zbl 0138.42001

[Hirzebruch et al. 1992] F. Hirzebruch, T. Berger, and R. Jung, *Manifolds and modular forms*, Aspects of Mathematics **E20**, Vieweg, Braunschweig, 1992. MR 94d:57001 Zbl 0767.57014

[Libgober and Wood 1990] A. S. Libgober and J. W. Wood, "Uniqueness of the complex structure on Kähler manifolds of certain homotopy types", *J. Differential Geom.* **32**:1 (1990), 139–154. MR 91g:32039 Zbl 0711.53052

[Milnor 1960] J. Milnor, "On the cobordism ring Ω^* and a complex analogue. I", *Amer. J. Math.* **82** (1960), 505–521. MR 22 #9975 Zbl 0095.16702

[Salamon 1996] S. M. Salamon, "On the cohomology of Kähler and hyper-Kähler manifolds", *Topology* **35**:1 (1996), 137–155. MR 97f:32042 Zbl 0854.58004

Received April 11, 2010. Revised November 4, 2010.

PING LI
DEPARTMENT OF MATHEMATICS
TONGJI UNIVERSITY
SHANGHAI, 200092
CHINA
pingli@tongji.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 1, 2011

BLOCKS OF THE CATEGORY OF CUSPIDAL $\mathfrak{sp}_{2n}$-MODULES

VOLODYMYR MAZORCHUK AND CATHARINA STROPPEL

In this paper we show that every block of the category of cuspidal generalized weight modules with finite dimensional generalized weight spaces over the Lie algebra $\mathfrak{sp}_{2n}(\mathbb{C})$ is equivalent to the category of finite dimensional $\mathbb{C}[[t_1, t_2, \dots, t_n]]$-modules.

1. Introduction and description of the results

Fix the ground field to be the complex numbers. Fix $n \in \{2, 3, \dots\}$ and consider the symplectic Lie algebra $\mathfrak{sp}_{2n} =: \mathfrak{g}$ with a fixed Cartan subalgebra $\mathfrak{h}$ and root space decomposition

$$\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\alpha \in \Delta} \mathfrak{g}_\alpha,$$

where Δ denotes the corresponding root system. For a $\mathfrak{g}$-module V and $\lambda \in \mathfrak{h}^*$ set

$$V_\lambda := \{v \in V : h \cdot v = \lambda(h)v \text{ for any } h \in \mathfrak{h}\},$$
$$V^\lambda := \{v \in V : (h - \lambda(h))^k \cdot v = 0 \text{ for any } h \in \mathfrak{h} \text{ and } k \gg 0\}.$$

A $\mathfrak{g}$-module V is called

- a *weight module* provided that $V = \bigoplus_{\lambda \in \mathfrak{h}^*} V_\lambda$;
- a *generalized weight module* provided that $V = \bigoplus_{\lambda \in \mathfrak{h}^*} V^\lambda$;
- a *cuspidal module* provided that for any $\alpha \in \Delta$ the action of any nonzero element from $\mathfrak{g}_\alpha$ on V is bijective.

If V is a generalized weight module, then the set $\{\lambda \in \mathfrak{h}^* : V_\lambda \neq 0\}$ is called the *support* of V and is denoted by $\mathrm{supp}(V)$.

Denote by $\hat{\mathscr{C}}$ the full subcategory in $\mathfrak{g}$-mod that consists of all cuspidal generalized weight modules with finite dimensional generalized weight spaces, and by $\mathscr{C}$ the full subcategory of $\hat{\mathscr{C}}$ consisting of all weight modules. Understanding the categories $\mathscr{C}$ and $\hat{\mathscr{C}}$ is a classical problem in the representation theory of Lie algebras. The first major step towards the solution of this problem was made in [Mathieu 2000], where all simple objects in $\hat{\mathscr{C}}$ were classified. Britten et al. [2004]

MSC2000: 17B10.
Keywords: cuspidal module, category, block, power series.

showed that the category $\mathscr{C}$ is semisimple, hence completely understood. The aim of the present note is to describe the category $\hat{\mathscr{C}}$.

Apart from $\mathfrak{sp}_{2n}$, cuspidal weight modules with finite dimensional weight spaces exist only for the Lie algebra $\mathfrak{sl}_n$ [Fernando 1990]. In the latter case, simple objects in the corresponding category $\hat{\mathscr{C}}$ are classified in [Mathieu 2000], the category $\mathscr{C}$ is described in [Grantcharov and Serganova 2010] (see also [Mazorchuk and Stroppel 2011]), and the category $\hat{\mathscr{C}}$ is described in [Mazorchuk and Stroppel 2011]. Taking all these results into account, the present paper completes the study of cuspidal generalized weight modules with finite dimensional generalized weight spaces over semisimple finite dimensional Lie algebras.

Let $U(\mathfrak{g})$ be the universal enveloping algebra of $\mathfrak{g}$ and $Z(\mathfrak{g})$ the center of $U(\mathfrak{g})$. The action of $Z(\mathfrak{g})$ on any object from $\hat{\mathscr{C}}$ is locally finite. Using this and the standard support arguments gives the following *block decomposition* of $\hat{\mathscr{C}}$:

$$\hat{\mathscr{C}} \cong \bigoplus_{\substack{\chi:Z(\mathfrak{g})\to\mathbb{C}\\ \xi\in\mathfrak{h}^*/\mathbb{Z}\Delta}} \hat{\mathscr{C}}_{\chi,\xi},$$

where $\hat{\mathscr{C}}_{\chi,\xi}$ consists of all V such that $\mathrm{supp}(V) \subset \xi$ and $(z-\chi(z))^k \cdot v = 0$ for all $v \in V$, $z \in Z(\mathfrak{g})$ and $k \gg 0$. Set

$$\mathscr{C}_{\chi,\xi} := \mathscr{C} \cap \hat{\mathscr{C}}_{\chi,\xi}.$$

From [Mathieu 2000, Section 9] it follows that each nontrivial $\hat{\mathscr{C}}_{\chi,\xi}$ contains a unique (up to isomorphism) simple object. In particular, $\hat{\mathscr{C}}_{\chi,\xi}$ is indecomposable, hence a block. From this and [Britten et al. 2004] we thus get that every nontrivial block $\mathscr{C}_{\chi,\xi}$ is equivalent to the category of finite dimensional $\mathbb{C}$-modules. Our main result is the following:

Theorem 1. *Every nontrivial block $\hat{\mathscr{C}}_{\chi,\xi}$ is equivalent to the category of finite dimensional $\mathbb{C}[[t_1, t_2, \dots, t_n]]$-modules.*

To prove Theorem 1 we use and further develop the technique of extension of the module structure from a Lie subalgebra, originally developed in [Mazorchuk and Stroppel 2011] for the study of categories of singular and nonintegral cuspidal generalized weight $\mathfrak{sl}_n$-modules. The proof of Theorem 1 is given in Section 4. In Section 2 we recall the standard reduction to the case of the so-called simple *completely pointed* modules (that is, simple weight cuspidal modules for which *all* nontrivial weight spaces are one-dimensional) and a realization of such modules using differential operators. In Section 3 we define a functor from the category of finite dimensional $\mathbb{C}[[t_1, t_2, \dots, t_n]]$-modules to any block $\hat{\mathscr{C}}_{\chi,\xi}$ containing a simple completely pointed module. In Section 4 we prove that this functor is an equivalence of categories. In Section 5 we present some consequences of our main result. In particular, we recover the main result of [Britten et al. 2004] stated above.

2. Completely pointed simple cuspidal weight modules

A weight $\mathfrak{g}$-module V is called *pointed* provided that $\dim V_\lambda = 1$ for some $\lambda \in \mathfrak{h}^*$. If V is a pointed simple cuspidal weight $\mathfrak{g}$-module, then, obviously, all nontrivial weight spaces of V are one-dimensional, in which case one says that V is *completely pointed* (see [Britten et al. 2004]). It is enough to consider blocks with completely pointed simple modules because of the following:

Lemma 2. *All nontrivial blocks of $\hat{\mathscr{C}}$ are equivalent.*

Proof. In the case of the category $\mathscr{C}$, this is proved in [Britten et al. 2004, Lemma 2]. The same argument works in the case of the category $\hat{\mathscr{C}}$. □

We recall the explicit realization of completely pointed simple cuspidal modules from [Britten and Lemire 1987]. Denote by W_n the n-th *Weyl algebra*, that is, the algebra of differential operators with polynomial coefficients in variables $x_1, x_2, \dots, x_n$. The algebra W_n is generated by x_i and $\partial/\partial x_i$, $i = 1, \dots, n$, which satisfy the relations $[\partial/\partial x_i, x_j] = \delta_{i,j}$. Let $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ be the vectors of the standard basis in $\mathbb{C}^n$. Identify $\mathbb{C}^n$ with $\mathfrak{h}^*$ such that Δ becomes the following *standard root system* of type C_n:

$$\{\pm(\varepsilon_i \pm \varepsilon_j) : 1 \le i < j \le n\} \cup \{\pm 2\varepsilon_i : 1 \le i \le n\}.$$

Then

$$\boldsymbol{H} = \boldsymbol{H}_n = \{2\varepsilon_1, \varepsilon_2 - \varepsilon_1, \varepsilon_3 - \varepsilon_2, \dots, \varepsilon_n - \varepsilon_{n-1}\}$$

is a basis of Δ. Fix a basis of $\mathfrak{g}$ of the form

$$\boldsymbol{C} := \{X_{\pm\varepsilon_i \pm \varepsilon_j} : 1 \le i < j \le n\} \cup \{X_{\pm 2\varepsilon_i} : i = 1, 2, \dots, n\} \cup \{H_\alpha : \alpha \in \boldsymbol{H}\}$$

such that the following map defines an injective Lie algebra homomorphism from $\mathfrak{g}$ to the Lie algebra associated with W_n:

$$\begin{aligned}
X_{\varepsilon_i - \varepsilon_j} &\mapsto x_i \frac{\partial}{\partial x_j}, && 1 \le i \ne j \le n,\\
X_{\varepsilon_i + \varepsilon_j} &\mapsto x_i x_j, && i, j = 1, 2, \dots, n,\\
X_{-\varepsilon_i - \varepsilon_j} &\mapsto \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j}, && i, j = 1, 2, \dots, n,\\
H_{\varepsilon_{i+1} - \varepsilon_i} &\mapsto x_{i+1} \frac{\partial}{\partial x_{i+1}} - x_i \frac{\partial}{\partial x_i}, && i = 1, 2, \dots, n-1,\\
H_{2\varepsilon_1} &\mapsto \frac{1}{2}\Big(x_1 \frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_1} x_1\Big).
\end{aligned} \tag{1}$$

Set

$$\boldsymbol{B} := \{(b_1, b_2, \dots, b_n) \in \mathbb{Z}^n : b_1 + b_2 + \dots + b_n \in 2\mathbb{Z}\}.$$

For $\boldsymbol{a} = (a_1, a_2, \dots, a_n) \in \mathbb{C}^n$ define $N(\boldsymbol{a})$ to be the linear span of

$$\{\boldsymbol{x}^{\boldsymbol{b}} := x_1^{a_1+b_1} x_2^{a_2+b_2} \cdots x_n^{a_n+b_n} : \boldsymbol{b} \in \boldsymbol{B}\}.$$

First define an action of the elements from $\boldsymbol{C}$ on $N(\boldsymbol{a})$ using the formulae from (1) as follows:

$$
\begin{aligned}
X_{\varepsilon_i-\varepsilon_j}\boldsymbol{x}^{\boldsymbol{b}} &= (a_j+b_j)\boldsymbol{x}^{\boldsymbol{b}+\varepsilon_i-\varepsilon_j} && 1\le i\neq j\le n,\\
X_{\varepsilon_i+\varepsilon_j}\boldsymbol{x}^{\boldsymbol{b}} &= \boldsymbol{x}^{\boldsymbol{b}+\varepsilon_i+\varepsilon_j} && i,j=1,2,\dots,n,\\
X_{-\varepsilon_i-\varepsilon_j}\boldsymbol{x}^{\boldsymbol{b}} &= (a_i+b_i)(a_j+b_j)\boldsymbol{x}^{\boldsymbol{b}-\varepsilon_i-\varepsilon_j} && 1\le i\neq j\le n,\\
X_{-2\varepsilon_i}\boldsymbol{x}^{\boldsymbol{b}} &= (a_i+b_i)(a_i+b_i-1)\boldsymbol{x}^{\boldsymbol{b}-2\varepsilon_i} && i=1,2,\dots,n,\\
H_{\varepsilon_{i+1}-\varepsilon_i}\boldsymbol{x}^{\boldsymbol{b}} &= (a_{i+1}+b_{i+1}-a_i-b_i)\boldsymbol{x}^{\boldsymbol{b}} && i=1,2,\dots,n-1,\\
H_{2\varepsilon_1}\boldsymbol{x}^{\boldsymbol{b}} &= \tfrac{1}{2}(2a_1+2b_1+1)\boldsymbol{x}^{\boldsymbol{b}}.
\end{aligned}
\tag{2}
$$

Theorem 3 [Britten and Lemire 1987]. (i) *For every $\boldsymbol{a}\in\mathbb{C}^n$ the formulae in* (2) *define on $N(\boldsymbol{a})$ the structure of a completely pointed weight $\mathfrak{g}$-module.*

(ii) *If $a_i\notin\mathbb{Z}$ for all $i=1,\dots,n$, then the module $N(\boldsymbol{a})$ is simple and cuspidal.*

(iii) *Every completely pointed simple cuspidal $\mathfrak{g}$-module is isomorphic to $N(\boldsymbol{a})$ for some $\boldsymbol{a}\in\mathbb{C}^n$ such that $a_i\notin\mathbb{Z}$, $i=1,\dots,n$.*

3. The functor F

This section is similar to [Mazorchuk and Stroppel 2011, Section 3.1]. Fix $\boldsymbol{a}\in\mathbb{C}^n$ such that $a_i\notin\mathbb{Z}$, $i=1,\dots,n$. Let $\hat{\mathscr{C}}_{\boldsymbol{a}}$ denote the block of $\hat{\mathscr{C}}$ containing $N(\boldsymbol{a})$. The category $\hat{\mathscr{C}}_{\boldsymbol{a}}$ is closed under extensions. Denote the category of finite dimensional $\mathbb{C}[[t_1,t_2,\dots,t_n]]$-modules by $\mathbb{C}[[t_1,t_2,\dots,t_n]]$-mod. For $V\in\mathbb{C}[[t_1,t_2,\dots,t_n]]$-mod denote by T_i the linear operator describing the action of t_i on V. Set $\mathbf{0}=(0,0,\dots,0)\in\boldsymbol{B}$.

For $\boldsymbol{b}\in\boldsymbol{B}$ consider a copy $V^{\boldsymbol{b}}$ of V. Define

$$\mathrm{F}V := \bigoplus_{\boldsymbol{b}\in\boldsymbol{B}} V^{\boldsymbol{b}}.$$

Define the action of elements from $\boldsymbol{C}$ on the vector space $\mathrm{F}V$ in the following way: for $v\in V^{\boldsymbol{b}}$ set

$$
\begin{cases}
X_{\varepsilon_i-\varepsilon_j}v = (T_j+(a_j+b_j)\,\mathrm{Id}_V)v & \in V^{\boldsymbol{b}+\varepsilon_i-\varepsilon_j},\\
X_{\varepsilon_i+\varepsilon_j}v = v & \in V^{\boldsymbol{b}+\varepsilon_i+\varepsilon_j},\\
X_{-\varepsilon_i-\varepsilon_j}v = (T_i+(a_i+b_i)\,\mathrm{Id}_V)(T_j+(a_j+b_j)\,\mathrm{Id}_V)v & \in V^{\boldsymbol{b}-\varepsilon_i-\varepsilon_j},\\
X_{2\varepsilon_i}v = (T_i+(a_i+b_i)\,\mathrm{Id}_V)(T_i+(a_i+b_i-1)\,\mathrm{Id}_V)v & \in V^{\boldsymbol{b}-2\varepsilon_i},\\
H_{\varepsilon_{i+1}-\varepsilon_i}v = (T_{i+1}-T_i+(a_{i+1}+b_{i+1}-a_i-b_i)\,\mathrm{Id}_V)v & \in V^{\boldsymbol{b}},\\
H_{2\varepsilon_1}v = \tfrac{1}{2}(2T_1+(2a_1+2b_1+1)\,\mathrm{Id}_V)v & \in V^{\boldsymbol{b}},
\end{cases}
\tag{3}
$$

where i and j are as in the respective row of (2). For a homomorphism $f : V \to W$ of $\mathbb{C}[\![t_1, t_2, \dots, t_n]\!]$-modules denote by $\mathrm{F}f$ the diagonally extended linear map from $\mathrm{F}V$ to $\mathrm{F}W$, that is, for every $\boldsymbol{b} \in \boldsymbol{B}$ and $v \in V^{\boldsymbol{b}}$, set

$$\mathrm{F}f(v) = f(v) \in W^{\boldsymbol{b}}. \tag{4}$$

Proposition 4. (i) *The formulae of* (3) *define on* $\mathrm{F}V$ *the structure of a* $\mathfrak{g}$*-module.*

(ii) *Every* $V^{\boldsymbol{b}}$ *is a generalized weight space of* $\mathrm{F}V$. *Moreover, for* $\boldsymbol{b} \neq \boldsymbol{b}'$ *the weights of* $V^{\boldsymbol{b}}$ *and* $V^{\boldsymbol{b}'}$ *are different.*

(iii) *The module* $\mathrm{F}V$ *belongs to* $\hat{\mathscr{C}}_{\boldsymbol{a}}$.

(iv) *Formulae* (3) *and* (4) *turn* F *into a functor*

$$\mathrm{F} : \mathbb{C}[\![t_1, t_2, \dots, t_n]\!]\text{-mod} \to \hat{\mathscr{C}}_{\boldsymbol{a}}.$$

(v) *The functor* F *is exact, faithful and full.*

Proof. Consider the $\mathfrak{g}$-module $N(\boldsymbol{a})$ for $\boldsymbol{a}$ as above. Then, for every $\boldsymbol{b}$, the defining relations of $\mathfrak{g}$ (in terms of elements from $\boldsymbol{C}$) applied to $\boldsymbol{x}^{\boldsymbol{b}}$ can be written as some polynomial equations in the a_i. Since (2) defines a $\mathfrak{g}$-module for any $\boldsymbol{a}$ by Theorem 3(i), these equations hold for any $\boldsymbol{a}$, that is, they are actually formal identities in the a_i. Now write

$$T_j + (a_j + b_j)\,\mathrm{Id}_V = A_j + B_j,$$

a sum of matrices, where $A_j = T_j + a_j\,\mathrm{Id}_V$ and $B_j = b_j\,\mathrm{Id}_V$. All A_i and B_j commute with each other and with all the T_l. For a fixed $\boldsymbol{b}$, the defining relations for $\mathfrak{g}$ on $\mathrm{F}V$ reduce to our formal identities (in the A_i) and hence are satisfied. This proves claim (i). Claim (ii) follows from the last two lines in (3) and the fact that all the T_i are nilpotent (hence zero is the only eigenvalue).

As f commutes with all T_i, the map $\mathrm{F}f$ commutes with the action of all elements from $\boldsymbol{C}$ and hence defines a homomorphism of $\mathfrak{g}$-modules. By construction we also have $\mathrm{F}(f \circ f') = \mathrm{F}f \circ \mathrm{F}f'$, which implies claim (iv).

By construction, F is exact and faithful. It sends the simple one-dimensional $\mathbb{C}[\![t_1, t_2, \dots, t_n]\!]$-module to $N(\boldsymbol{a})$ (as in this case all $T_i = 0$ and hence (3) gives (2)), which is an object of the category $\hat{\mathscr{C}}_{\boldsymbol{a}}$ closed under extensions. Claim (iii) follows.

To complete the proof of claim (v) we are left to show that F is full. Let $\varphi : \mathrm{F}V \to \mathrm{F}W$ be a $\mathfrak{g}$-homomorphism. Then φ commutes with the action of all elements from $\mathfrak{h}$. Using claim (ii), we get that φ induces, by restriction, a linear map $f : V = V^{\mathbf{0}} \to W^{\mathbf{0}} = W$. As φ commutes with all $H_{\varepsilon_{i+1}-\varepsilon_i}$, the map f commutes with all operators $T_{i+1} - T_i$. As φ commutes with $H_{2\varepsilon_1}$, the map f commutes with T_1. It follows that f is a homomorphism of $\mathbb{C}[\![t_1, t_2, \dots, t_n]\!]$-modules. This yields $\varphi = \mathrm{F}f$ and thus the functor F is full. This completes the proof of claim (v) and of the whole proposition. $\square$

4. Proof of Theorem 1

Because of Lemma 2 it is enough to fix one particular block and show there that F is an equivalence. Thus, we may assume that $a_i + a_j \notin \mathbb{Z}$ for all i, j (in particular, $a_i \notin \mathbb{Z}$ for all i). According to Proposition 4, we are only left to show that F is dense (that is, essentially surjective). We establish the density of F by induction on n. We first prove the induction step and then the basis of the induction, which is the case $n = 2$.

Denote by λ the weight of $\boldsymbol{x}^{\boldsymbol{0}} \in N(\boldsymbol{a})$ (see Proposition 4(ii)). Let $M \in \hat{\mathscr{C}}_{\boldsymbol{a}}$. Set $V := M_\lambda$ and denote by M' the $\mathfrak{a}$-module $U(\mathfrak{a})V$.

4.1. *Reduction to the case n = 2.* The main result of this section is the following:

Proposition 5. *If the functor* F *is dense for $n = 2$, then it is dense for any $n \geq 2$.*

Proof. Assume that $n > 2$ and that the functor F is dense in the case of the algebra $\mathfrak{sp}_{2n-2}$. Realize $\mathfrak{sp}_{2n-2}$ as the subalgebra $\mathfrak{a}$ of $\mathfrak{g}$ corresponding to the subset $\boldsymbol{H}_{n-1} \subset \boldsymbol{H}$ of simple roots.

Let $Y_1, Y_2, \ldots, Y_n$ be the linear operators representing the action of the elements $H_{2\varepsilon_1}, H_{\varepsilon_2-\varepsilon_1}, H_{\varepsilon_3-\varepsilon_2}, \ldots, H_{\varepsilon_n-\varepsilon_{n-1}}$ on V, respectively. Set

$$\begin{aligned} T_1 &:= Y_1 - \tfrac{1}{2}(2a_1+1)\,\mathrm{Id}_V, \\ T_2 &:= Y_2 + T_1 - (a_2 - a_1)\,\mathrm{Id}_V, \\ T_3 &:= Y_3 + T_2 - (a_3 - a_2)\,\mathrm{Id}_V, \\ &\ \vdots \\ T_n &:= Y_n + T_{n-1} - (a_n - a_{n-1})\,\mathrm{Id}_V\,. \end{aligned} \tag{5}$$

The T_i are obviously pairwise commuting nilpotent linear operators.

The module M' is a cuspidal generalized weight $\mathfrak{a}$-module with finite dimensional weight spaces. Moreover, as all composition subquotients of M are of the form $N(\boldsymbol{a})$, all composition subquotients of M' are of the form $N(\boldsymbol{a})'$, the latter being a completely pointed simple cuspidal $\mathfrak{a}$-module. By our inductive assumption, the functor F is dense in the case of the algebra $\mathfrak{a}$. Hence $M' \cong N' := \bigoplus_{\boldsymbol{b}} V^{\boldsymbol{b}}$, where $\boldsymbol{b} \in \boldsymbol{B}$ is such that $b_n = 0$, and the action of $\mathfrak{a}$ on N' is given by (3).

Lemma 6. *There is a unique* (*up to isomorphism*) $\mathfrak{g}$-*module* $Q \in \hat{\mathscr{C}}_{\boldsymbol{a}}$ *such that* $Q' = N'$ *and which gives the linear operator* T_n *when computed using* (5).

Proof. The existence statement is clear, so we need only to show uniqueness. Assume that $Q \in \hat{\mathscr{C}}_{\boldsymbol{a}}$ is such that $Q' = N'$ and the formulae in (5) applied to Q produce the linear operator T_n. Since $a_n \notin \mathbb{Z}$, the endomorphism $T_n + (a_n + b_n)\,\mathrm{Id}_V$ is invertible for all $b_n \in \mathbb{Z}$. As the action of $X_{\varepsilon_n-\varepsilon_{n-1}}$ on Q is bijective, we can fix a weight basis in Q such that both the $\mathfrak{a}$-action on $Q' = N'$ and the action of $X_{\varepsilon_n-\varepsilon_{n-1}}$ on the whole Q is given by (3). As $n > 2$, the elements $X_{\pm 2\varepsilon_1}$ commute with $X_{\varepsilon_n-\varepsilon_{n-1}}$ and hence their action extends uniquely to the whole of Q using this commutativity. This holds similarly for all elements $X_{\pm(\varepsilon_i-\varepsilon_{i-1})}$, $i<n-1$, and for the element $X_{\varepsilon_{n-2}-\varepsilon_{n-1}}$. This leaves us with the elements $X_{\varepsilon_{n-1}-\varepsilon_{n-2}}$ and $X_{\varepsilon_{n-1}-\varepsilon_n}$. The simple roots $\varepsilon_{n-1}-\varepsilon_{n-2}$ and $\varepsilon_n-\varepsilon_{n-1}$ corresponding to the elements $X_{\varepsilon_{n-1}-\varepsilon_{n-2}}$ and $X_{\varepsilon_n-\varepsilon_{n-1}}$ generate a root system of type A_2 (this corresponds to the algebra $\mathfrak{sl}_3$). Lemmas 21 and 22 of [Mazorchuk and Stroppel 2011] prove that the actions of $X_{\varepsilon_{n-1}-\varepsilon_{n-2}}$ and $X_{\varepsilon_{n-1}-\varepsilon_n}$ extend uniquely to Q. This completes the proof of Lemma 6. □

The module $\mathrm{F}V$ obviously satisfies $(\mathrm{F}V)'=N'$ and defines the linear operator T_n when computed using (5). Hence Lemma 6 implies $M\cong \mathrm{F}V$. Since $M\in\hat{\mathscr{C}}_{\boldsymbol{a}}$ was arbitrary, the functor F is dense, completing the proof of Proposition 5. □

4.2. *Base of the induction: some $\mathfrak{sl}_2$-theory as preparation.* In this section we will recall (and slightly improve) some classical $\mathfrak{sl}_2$-theory. For details see [Mazorchuk 2010]. Consider the Lie algebra $\mathfrak{sl}_2=\mathfrak{sl}_2(\mathbb{C})$ with standard basis

$$\boldsymbol{e}:=\begin{pmatrix}0&1\\0&0\end{pmatrix},\quad \boldsymbol{f}:=\begin{pmatrix}0&0\\1&0\end{pmatrix},\quad \boldsymbol{h}:=\begin{pmatrix}1&0\\0&-1\end{pmatrix}.$$

Let V be a finite dimensional vector space and A and B be two commuting linear operators on V. For $i\in\mathbb{Z}$ denote by $V^{(i)}$ a copy of V and consider the vector space $\overline{V}:=\bigoplus_{i\in\mathbb{Z}}V^{(i)}$ (a direct sum of copies of V indexed by i). Define the actions of $\boldsymbol{e}$, $\boldsymbol{f}$ and $\boldsymbol{h}$ on $\overline{V}$ as follows: for $v\in V^{(i)}$ set

$$\begin{aligned}\mathbf{v}&:=(P-i\,\mathrm{Id}_V)v &&\in V^{(i+1)},\\ \mathbf{v}&:=(Q+i\,\mathrm{Id}_V)v &&\in V^{(i-1)},\\ \mathbf{v}&:=(Q-P+2i\,\mathrm{Id}_V)v &&\in V^{(i)}.\end{aligned}\tag{6}$$

This can be depicted as follows (here right arrows represent the action of $\boldsymbol{e}$, left arrows represent the action of $\boldsymbol{f}$ and loops represent the action of $\boldsymbol{h}$):

$$\begin{tikzcd}
\cdots \arrow[r, shift left, "P+2\,\mathrm{Id}_V"] & V^{(-1)} \arrow[l, shift left, "Q-\mathrm{Id}_V"] \arrow[r, shift left, "P+\mathrm{Id}_V"] \arrow[loop below, "Q-P-2\,\mathrm{Id}_V"] & V^{(0)} \arrow[l, shift left, "Q"] \arrow[r, shift left, "P"] \arrow[loop below, "Q-P"] & V^{(1)} \arrow[l, shift left, "Q+\mathrm{Id}_V"] \arrow[r, shift left, "P-\mathrm{Id}_V"] \arrow[loop below, "Q-P+2\,\mathrm{Id}_V"] & \cdots \arrow[l, shift left, "Q+2\,\mathrm{Id}_V"]
\end{tikzcd}$$

Proposition 7. (i) *The formulae in* (6) *define on $\overline{V}$ the structure of a generalized weight $\mathfrak{sl}_2$-module with finite dimensional generalized weight spaces.*

(ii) *Every cuspidal generalized weight $\mathfrak{sl}_2$-module with finite dimensional generalized weight spaces is isomorphic to $\overline{V}$ for some V with P and Q as above.*

(iii) *The action of the Casimir element $\boldsymbol{c}:=(\boldsymbol{h}+1)^2+4\boldsymbol{f}\boldsymbol{e}$ on $\overline{V}$ is given by the linear operator $(P+Q+\mathrm{Id}_V)^2$.*

(iv) *Let* $\mathbb{C}^2$ *denote the natural* $\mathfrak{sl}_2$*-module* (*the unique two-dimensional simple* $\mathfrak{sl}_2$*-module*). *Then the linear operator* $(\boldsymbol{c} - (P + Q + 2\,\mathrm{Id}_V)^2)(\boldsymbol{c} - (P + Q)^2)$ *annihilates the* $\mathfrak{sl}_2$*-module* $\mathbb{C}^2 \otimes \overline{V}$.

(v) *Let* $\mathbb{C}^3$ *denote the unique three-dimensional simple* $\mathfrak{sl}_2$*-module. Then the linear operator* $(\boldsymbol{c}-(P+Q+3\,\mathrm{Id}_V)^2)(\boldsymbol{c}-(P+Q+\mathrm{Id}_V)^2)(\boldsymbol{c}-(P+Q-\mathrm{Id}_V)^2)$ *annihilates the* $\mathfrak{sl}_2$*-module* $\mathbb{C}^3 \otimes \overline{V}$.

Proof. The fact that $\overline{V}$ is an $\mathfrak{sl}_2$-module is checked by a direct computation. That $\overline{V}$ is a generalized weight module follows from the fact that the action of $\boldsymbol{h}$ on $\overline{V}$ preserves (by (6)) each V^i and hence is locally finite. Since the category of generalized weight modules is closed under extensions, to prove that $\overline{V}$ has finite dimensional generalized weight spaces it is enough to consider the case when $\boldsymbol{h}$ has a unique eigenvalue on $V^{(0)}$, say λ. However, in this case $\boldsymbol{h}$ has a unique eigenvalue on V^i, namely $\lambda + 2i$, which implies that $\overline{V}^{\lambda} = V$ is finite dimensional. Claim (i) follows. To prove Claim (iii) we observe that the action of $\boldsymbol{c}$ on V^i is given by

$$(Q - P + (2i + 1)\,\mathrm{Id}_V)^2 + 4(Q + (i + 1)\,\mathrm{Id}_V)(P - i\,\mathrm{Id}_V) = (P + Q + \mathrm{Id}_V)^2.$$

Claim (ii) can be found with all details in [Mazorchuk 2010, Chapter 3].

To prove claim (iv) choose a basis $\{v_1, \dots, v_k\}$ in V, which gives rise to a basis $\{v_1^{(i)}, \dots, v_k^{(i)}, i \in \mathbb{Z}\}$ in $\overline{V}$. Choose the standard basis $\{e_1, e_2\}$ in $\mathbb{C}^2$. Since $\boldsymbol{h}e_1 = e_1$, $\boldsymbol{h}e_2 = -e_2$ and $\boldsymbol{h}$ acts by $Q - P + 2i\,\mathrm{Id}_V$ on $V^{(i)}$, we obtain that $\boldsymbol{h}$ acts by $Q - P + (2i + 1)\,\mathrm{Id}_V$ on the vector space $W^{(i)}$ with basis

$$\{e_1 \otimes v_1^{(i)}, \dots, e_1 \otimes v_1^{(i)}, e_2 \otimes v_1^{(i+1)}, \dots, e_2 \otimes v_1^{(i+1)}\}.$$

We have $\mathbb{C}^2 \otimes \overline{V} \cong \bigoplus_{i\in\mathbb{Z}} W^{(i)}$ and one easily computes that in the above basis the actions of $\boldsymbol{e}$ and $\boldsymbol{f}$ on $\mathbb{C}^2 \otimes \overline{V}$ are given by the following picture:

$$\begin{tikzcd}
\cdots \arrow[r, shift left] & W^{(-1)} \arrow[l, shift left] \arrow[r, shift left, "\begin{pmatrix} P+\mathrm{Id} & \mathrm{Id} \\ 0 & P \end{pmatrix}"] & W^{(0)} \arrow[l, shift left, "\begin{pmatrix} Q & 0 \\ \mathrm{Id} & Q+\mathrm{Id} \end{pmatrix}"] \arrow[r, shift left, "\begin{pmatrix} P & \mathrm{Id} \\ 0 & P-\mathrm{Id} \end{pmatrix}"] & W^{(1)} \arrow[l, shift left, "\begin{pmatrix} Q+\mathrm{Id} & 0 \\ \mathrm{Id} & Q+2\,\mathrm{Id} \end{pmatrix}"] \arrow[r, shift left] & \cdots \arrow[l, shift left]
\end{tikzcd}$$

The action of $\boldsymbol{c}$ on $W^{(0)}$ is now easily computed to be given by the linear operator

$$G := \begin{pmatrix} (Q-P+2\,\mathrm{Id})^2 + 4(Q+\mathrm{Id})P & 4(Q+\mathrm{Id}) \\ 4P & (Q-P+2\,\mathrm{Id})^2 + 4(Q+2\,\mathrm{Id})(P-\mathrm{Id}) + 4\,\mathrm{Id} \end{pmatrix}.$$

The characteristic polynomial of G is

$$\chi_G(\lambda) = (\lambda - (P + Q + 2\,\mathrm{Id})^2)(\lambda - (P + Q)^2).$$

Claim (iv) now follows from the Cayley–Hamilton theorem.

We have an isomorphism of $\mathfrak{sl}_2$-modules as follows: $\mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathbb{C}^3 \oplus \mathbb{C}$ (here $\mathbb{C}$ is the trivial module), and hence claim (v) follows applying claim (iv) twice.

Alternatively, one could do a direct calculation, similar to the proof of (iii). The proposition follows. □

The statement of Proposition 7(ii) is a special case of a more general result of Gabriel and Drozd describing blocks of the category of (generalized) weight $\mathfrak{sl}_2$-modules, in particular, simple weight $\mathfrak{sl}_2$-modules (see [Drozd 1983; Dixmier 1996, 7.8.16]). The statements of Proposition 7(iv) and (v) are $\mathfrak{sl}_2$-refinements of a theorem of Kostant [1975, Theorem 5.1] describing possible (generalized) central characters of the tensor product of a finite dimensional module with an infinite dimensional module.

4.3. ***The case $n = 2$.*** Assume now that $n = 2$. We have $a_1, a_2, a_1 + a_2 \notin \mathbb{Z}$. Let $\mathfrak{a}$ denote the Lie subalgebra of $\mathfrak{g}$ generated by $X_{\pm(\varepsilon_2-\varepsilon_1)}$. The algebra $\mathfrak{a}$ is isomorphic to $\mathfrak{sl}_2$.

Let $M \in \hat{\mathscr{C}}_{\boldsymbol{a}}$. Denote by λ the weight of $\boldsymbol{x}^{\mathbf{0}} \in N(\boldsymbol{a})$ and set $V := M_\lambda$. Let Y_1 and Y_2 be the linear operators representing the actions of the elements $H_{\varepsilon_2-\varepsilon_1}$ and $C := (H_{\varepsilon_2-\varepsilon_1} + 1)^2 + 4X_{\varepsilon_1-\varepsilon_2}X_{\varepsilon_2-\varepsilon_1}$ on V. The element C is a Casimir element for $\mathfrak{a}$. In particular, the operators Y_1 and Y_2 commute. Our first observation is the following:

Lemma 8. *The action of C on V is invertible and hence has a square root.*

Proof. From (2) we have that C acts on $\boldsymbol{x}^{\mathbf{0}}$ by

$$(a_2 - a_1 + 1)^2 + 4(a_2 + 1)a_1 = (a_1 + a_2 + 1)^2.$$

Since $a_1 + a_2 \notin \mathbb{Z}$ by our assumptions, $\boldsymbol{x}^{\mathbf{0}}$ is an eigenvector of C with a nonzero eigenvalue. As the module M has a composition series with subquotients isomorphic to $N(\boldsymbol{a})$, the complex number $(a_1 + a_2 + 1)^2 \neq 0$ is the only eigenvalue of C on V. The claim follows. □

Consider the $\mathfrak{a}$-module $M' := U(\mathfrak{a})M_\lambda$. Let Y_2' denote any square root of Y_2, which is a polynomial in Y_2 (it exists by Lemma 8). So Y_2' commutes with Y_1. Set

$$T_1 := \frac{Y_2' - Y_1 - \mathrm{Id}_V}{2} - a_1 \,\mathrm{Id}_V, \quad T_2 := \frac{Y_2' + Y_1 - \mathrm{Id}_V}{2} - a_2 \,\mathrm{Id}_V\,.$$

Then T_1 and T_2 are two commuting nilpotent linear operators (it is easy to check that 0 is the unique eigenvalue for both T_1 and T_2), hence define on V the structure of a $\mathbb{C}[[t_1, t_2]]$-module. The aim of this section is to establish an isomorphism $\mathrm{F}V \cong M$, which would complete the proof of Theorem 1.

Set $R' := U(\mathfrak{a})(\mathrm{F}V)_\lambda$. A direct computation using (3) shows that $H_{\varepsilon_2-\varepsilon_1}$ and C act on $(\mathrm{F}V)_\lambda = V^{\mathbf{0}}$ as the linear operators Y_1 and Y_2, respectively. As any cuspidal generalized weight $\mathfrak{a}$-module is uniquely determined by the actions of $H_{\varepsilon_2-\varepsilon_1}$ and C (see [Drozd 1983; Mazorchuk 2010, 3.7] for full details), it follows that $M' \cong R'$. The isomorphism $\mathrm{F}V \cong M$ now follows from the next proposition:

Proposition 9. *There is at most one* (*up to isomorphism*) $\mathfrak{g}$-*module* $R \in \hat{\mathscr{C}}_{\boldsymbol{a}}$ *such that* $U(\mathfrak{a})R_\lambda = R'$.

Proof. Let $R \in \hat{\mathscr{C}}_{\boldsymbol{a}}$ be such that $U(\mathfrak{a})R_\lambda = R'$. Choose a weight basis in R such that the action of $\mathfrak{a}$ on R' and the action of $X_{2\varepsilon_1}$ on R is given by (3) (in other words these actions coincide with the corresponding actions on $\mathrm{F}V$). Since $X_{\varepsilon_1-\varepsilon_2}$ commutes with $X_{2\varepsilon_1}$, it follows that the action of $X_{\varepsilon_1-\varepsilon_2}$ on R is also given by (3).

It is left to show that the action of $X_{\varepsilon_2-\varepsilon_1}$ extends uniquely from R' to R and then that there is a unique way to define the action of $X_{-2\varepsilon_1}$. This will be done in the Lemmata 10 and 11 below. □

Lemma 10. *There is a unique way to extend the action of* $X_{\varepsilon_2-\varepsilon_1}$ *from* R' *to* R.

Proof. We first show that for every $k \in \{1, 2, \dots\}$, the action of $X_{\varepsilon_2-\varepsilon_1}$ extends uniquely from $X_{2\varepsilon_1}^{k-1}R'$ to $X_{2\varepsilon_1}^{k}R'$ (here $X_{2\varepsilon_1}^{0}R' = R'$).

Consider the following picture:

(7)
$$\begin{tikzcd}
\bullet \arrow[r, dashed, bend left, "X"] & \bullet \arrow[l, bend left, "Q"] & \\
\bullet \arrow[u, "1"] \arrow[r, bend left, "P+1"] & \bullet \arrow[u, "1"'] \arrow[l, bend left, "Q"] \arrow[r, bend left, "P"] & \bullet \arrow[l, bend left, "Q+1"]
\end{tikzcd}$$

Here bullets are weight spaces with some fixed bases. The lower row is a part of $X_{2\varepsilon_1}^{k-1}R'$ where the $\mathfrak{a}$-action is already known by induction. The bases in the weight spaces in the lower row are chosen such that the action of $\mathfrak{a}$ in the lower row is given by (3). The upper row is a part of $X_{2\varepsilon_1}^{k}R'$ where the $\mathfrak{a}$-action is to be determined. Arrows pointing up indicate the action of $X_{2\varepsilon_1}$. The bases of the weight spaces in the upper row are chosen such that the action of $X_{2\varepsilon_1}$ is given by the operator Id_V (as in (3)). Left arrows indicate the action of $X_{\varepsilon_1-\varepsilon_2}$. The latter commutes with the action of $X_{2\varepsilon_1}$ and hence is given by the same linear operator in each column. Right arrows indicate the action of $X_{\varepsilon_2-\varepsilon_1}$ (which is known for $X_{2\varepsilon_1}^{k-1}R'$ and is to be determined for $X_{2\varepsilon_1}^{k}R'$). The part to be determined is given by the dashed arrow. Labels P and Q represent coefficients (which are linear operators on V) appearing in the corresponding parts of formulae (3). Note that P and Q commute. The action of $X_{\varepsilon_2-\varepsilon_1}$ on $X_{2\varepsilon_1}^{k}R'$ which is to be determined is given by some unknown linear operator X.

From $H_{\varepsilon_2-\varepsilon_1} = [X_{\varepsilon_2-\varepsilon_1}, X_{\varepsilon_1-\varepsilon_2}]$ we see that the action of $H_{\varepsilon_2-\varepsilon_1}$ on the middle weight space in the lower row is given by $Q - P$. Using $[H_{\varepsilon_2-\varepsilon_1}, X_{2\varepsilon_1}] = -2X_{2\varepsilon_1}$ we get that $H_{\varepsilon_2-\varepsilon_1}$ acts on the right dot of the upper row via $Q - P - 2$. Using $[H_{\varepsilon_2-\varepsilon_1}, X_{\varepsilon_1-\varepsilon_2}] = -2X_{\varepsilon_1-\varepsilon_2}$ we get that $H_{\varepsilon_2-\varepsilon_1}$ acts on the left dot of the upper row via $Q-P-4$. So the action of C on the upper row is given by $(Q-P-3)^2+4XQ$.

The action of C on the lower row is given by $(Q - P - 1)^2 + 4(P + 1)Q = (Q + P + 1)^2$.

The elements $X_{2\varepsilon_1}$, $X_{2\varepsilon_2}$ and $X_{\varepsilon_1+\varepsilon_1}$ form a weight basis of a simple three-dimensional $\mathfrak{a}$-module $\mathbb{C}^3$ with respect to the adjoint action of $\mathfrak{a}$. Hence the upper row of our picture is a subquotient of the tensor product of the lower row and $\mathbb{C}^3$. Therefore, from Proposition 7(v) we obtain that the linear operator

$$(C - (Q + P - 1)^2)(C - (Q + P + 1)^2)(C - (Q + P + 3)^2)$$

annihilates the upper row. A direct computation using (3) shows that the action of the operators $C - (Q + P - 1)^2$ and $C - (Q + P + 1)^2$ on the part $X^k_{2\varepsilon_1} N(\boldsymbol{a})'$ of the module $N(\boldsymbol{a})$ is invertible. As the $\mathfrak{g}$-module we are working with must have a composition series with subquotients $N(\boldsymbol{a})$, it follows that the action of both $C - (Q + P - 1)^2$ and $C - (Q + P + 1)^2$ on $X^k_{2\varepsilon_1} R'$ is invertible. Hence $C - (Q + P + 3)^2$ annihilates $X^k_{2\varepsilon_1} R'$, which gives us the equation

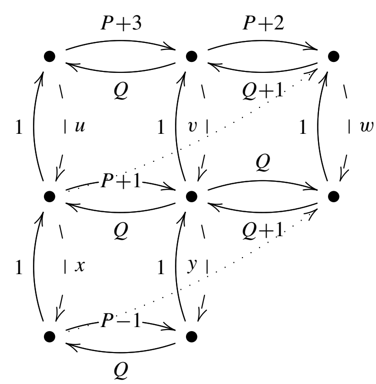
$$(Q - P - 3)^2 + 4XQ = (Q + P + 3)^2.$$

This equation has a unique solution, namely $X = Q + 3$, which gives the required extension.

Similarly one shows that for $k \in \{-1, -2, \dots\}$, the action of $X_{\varepsilon_2-\varepsilon_1}$ extends uniquely from $X^{k+1}_{2\varepsilon_1} R'$ to $X^k_{2\varepsilon_1} R'$ (here again $X^0_{2\varepsilon_1} R' = R'$). $\square$

Lemma 11. *There is a unique way to define the action of $X_{-2\varepsilon_1}$ on N.*

Proof. To determine this action of $X_{-2\varepsilon_1}$ on N we consider the following extension of the picture (7) with the same notation as in the proof of Lemma 10:



Here all right arrows, representing the action of $X_{\varepsilon_2-\varepsilon_1}$, are now determined by Lemma 10 and we have to figure out the down arrows, representing the action of $X_{-2\varepsilon_1}$. The two dotted arrows will be used later on in the proof.

Consider the $\mathfrak{sl}_2$-subalgebra $\mathfrak{c}$ of $\mathfrak{g}$ generated by $e := X_{2\varepsilon_1}$ and $f := X_{-2\varepsilon_1}$. Set $h := [e, f]$. Denote by Z the action of h in the leftmost weight space of the middle row. Then $Z = x - u$. The element h commutes with both h and $H_{\varepsilon_2-\varepsilon_1}$. Therefore, by (3), the operator Z commutes with both T_1 and T_2 and hence with both P and Q.

The algebra $\mathfrak{c}$ has the quadratic Casimir element $C_{\mathfrak{c}}$, whose action on the $\mathfrak{c}$-module given by the leftmost column of our picture is given by $x + f(Z)$, where f is some polynomial of degree two. From (3) it follows that the unique eigenvalue of this action is nonzero, in particular, $x + f(Z)$ is invertible. Let x' be a fixed square root $x + f(Z)$, which is a polynomial in $x + f(Z)$.

The elements $X_{\varepsilon_2-\varepsilon_1}$ and $X_{\varepsilon_2+\varepsilon_1}$ form a basis of a simple two-dimensional $\mathfrak{c}$-module with respect to the adjoint action. Using Proposition 7(iv) and arguments similar to those used in the proof of Lemma 10, we get that $C_{\mathfrak{c}} - (x' + 1)^2$ or $C_{\mathfrak{c}} - (x' - 1)^2$ annihilates the middle column (the sign depends on the original choice of x'). The middle column equals $X_{\varepsilon_2-\varepsilon_1}$ applied to the leftmost column.

Similarly, the elements $X_{\varepsilon_1-\varepsilon_2}$ and $X_{-\varepsilon_2-\varepsilon_1}$ form a basis of a simple two-dimensional $\mathfrak{c}$-module with respect to the adjoint action. Applying the same arguments as in the previous paragraph we get that $C_{\mathfrak{c}} - (x')^2$ annihilates any vector of the form $X_{\varepsilon_1-\varepsilon_2}X_{\varepsilon_2-\varepsilon_1}\mathrm{v}$, where v is from the leftmost column. This implies that the actions of $C_{\mathfrak{c}}$ and $X_{\varepsilon_1-\varepsilon_2}X_{\varepsilon_2-\varepsilon_1}$ and thus the actions of $C_{\mathfrak{c}}$ and C on the leftmost column commute. As the action of H commutes with the action of C, we thus obtain that x commutes with the action of C. This implies that x commutes with $T_1 + T_2$. As it obviously commutes with $T_1 - T_2$, we get that x commutes with both T_1 and T_2 and hence with both P and Q.

Similarly one shows that y, u, v and w commute with both P and Q. From the commutativity of $X_{\varepsilon_2-\varepsilon_1}$ and $X_{-2\varepsilon_1}$ we get the conditions

$$y(P+1) = (P-1)x, \quad V(P+3) = (P+1)u, \quad w(P+2)(P+3) = P(P+1)u.$$

Here everything commutes by the above and $P+1$, $P+2$ and $P+3$ are invertible (as $X_{\varepsilon_2-\varepsilon_1}$ acts bijectively). Therefore

$$y = (P-1)(P+1)^{-1}x, \quad v = (P+1)(P+3)^{-1}u, \quad w = P(P+1)(P+3)^{-1}(P+2)^{-1}u.$$

This implies that y, v and w are uniquely determined by x and u.

Since the actions of both $X_{\varepsilon_2-\varepsilon_1}$ and $X_{2\varepsilon_1}$ are completely determined, we can compute the action of $X_{2\varepsilon_2}$ and see that it is given (similarly to the action of $X_{2\varepsilon_1}$) by Id_V (this is depicted by the dotted arrows in the picture). As $X_{-2\varepsilon_2}$ and $X_{2\varepsilon_2}$ commute, we obtain that $w = x$, that is,

$$x = P(P+1)(P+3)^{-1}(P+2)^{-1}u. \tag{8}$$

Therefore the only parameter left for now is u.

On the one hand, the action of the element h on the middle dot of the second row is given by $y - v = (P-1)(P+1)^{-1}x - (P+1)(P+3)^{-1}u$. On the other hand, from $[h, X_{\varepsilon_2-\varepsilon_1}] = 4X_{\varepsilon_2-\varepsilon_1}$ we have that this action equals $Z+4 = x - u + 4$.

This gives us the equation

$$(P-1)(P+1)^{-1}x-(P+1)(P+3)^{-1}u=x-u+4. \tag{9}$$

Using (9) and (8) we get the equation

$$\frac{P(P-1)}{(P+2)(P+3)}u+\frac{P+1}{P+3}u=\frac{P(P+1)}{(P+2)(P+3)}u-u+4.$$

This is a linear equation with nonzero coefficients and thus it has a unique solution, namely $u=(P+3)(P+2)$. Hence u is uniquely defined. The claim of the lemma follows. □

5. Consequences

Corollary 12. *Let $\boldsymbol{a}\in\mathbb{C}^n$ be such that $a_i\notin\mathbb{Z}$ and $a_i+a_j\notin\mathbb{Z}$ for all i and j. Let $M\in\hat{\mathscr{C}}$ and $\lambda\in\mathrm{supp}(M)$. Denote by U_0 the centralizer of $\mathfrak{h}$ in $U(\mathfrak{g})$. Then for any $A, B\in U_0$ the actions of A and B on M_λ commute.*

Proof. By Proposition 4, we may assume that $M\cong \mathrm{F}V$. For the module $\mathrm{F}V$ the claim follows from the formulae in (3). □

Corollary 13. *For any simple weight cuspidal $\mathfrak{g}$-module L with finite dimensional weight spaces we have* $\dim\mathrm{Ext}^1_{\mathfrak{g}}(L,L)=n$.

Proof. This follows from Theorem 1 and the observation that a similar equality is true for the unique simple $\mathbb{C}[[t_1,t_2,\dots,t_n]]$-module. □

We also recover the main result of [Britten et al. 2004]:

Corollary 14. *The category of all weight cuspidal $\mathfrak{g}$-modules is semisimple.*

Proof. By [Britten et al. 2004, Lemma 2], all blocks of the category of weight cuspidal $\mathfrak{g}$-modules are equivalent. Hence it is enough to prove the claim for the block containing $N(\boldsymbol{a})$ for some $\boldsymbol{a}\in\mathbb{C}^n$ such that $a_i+a_j\notin\mathbb{Z}$ for all i, j. From (3) it follows that the module $\mathrm{F}V$ is weight if and only if all operators T_i are semisimple, hence zero. Therefore from Theorem 1 we get that the block of the category of weight cuspidal modules is equivalent to the category of finite dimensional modules over $\mathbb{C}[[t_1,t_2,\dots,t_n]]/(t_1-0,t_2-0,\dots,t_n-0)\cong\mathbb{C}$. The claim follows. □

Acknowledgments

The first author was partially supported by the Royal Swedish Academy of Sciences and the Swedish Research Council. The paper was written when the second author was visiting the Mathematical Science Research Institute in Berkeley. She deeply acknowledges the excellent working conditions and financial support.

References

[Britten and Lemire 1987] D. Britten and F. W. Lemire, "A classification of simple Lie modules having a 1-dimensional weight space", *Trans. Amer. Math. Soc.* **299**:2 (1987), 683–697. MR 88b:17013 Zbl 0635.17002

[Britten et al. 2004] D. Britten, O. Khomenko, F. Lemire, and V. Mazorchuk, "Complete reducibility of torsion free C_n-modules of finite degree", *J. Algebra* **276**:1 (2004), 129–142. MR 2005b:17008 Zbl 1127.17005

[Dixmier 1996] J. Dixmier, *Enveloping algebras*, Graduate Studies in Math. **11**, Amer. Math. Soc., Providence, RI, 1996. Revised reprint of the 1977 translation. MR 97c:17010 Zbl 0867.17001

[Drozd 1983] Y. A. Drozd, "Representations of Lie algebras $\mathfrak{sl}(2)$", *Vīsnik Kiïv. Unīv. Ser. Mat. Mekh.* **25** (1983), 70–77. In Ukranian. MR 86j:17010 Zbl 0615.17006

[Fernando 1990] S. L. Fernando, "Lie algebra modules with finite-dimensional weight spaces, I", *Trans. Amer. Math. Soc.* **322**:2 (1990), 757–781. MR 91c:17006 Zbl 0712.17005

[Grantcharov and Serganova 2010] D. Grantcharov and V. Serganova, "Cuspidal representations of $\mathfrak{sl}(n+1)$", *Adv. Math.* **224**:4 (2010), 1517–1547. MR 2646303 Zbl 05724629

[Kostant 1975] B. Kostant, "On the tensor product of a finite and an infinite dimensional representation", *J. Functional Analysis* **20**:4 (1975), 257–285. MR 54 #2888 Zbl 0355.17010

[Mathieu 2000] O. Mathieu, "Classification of irreducible weight modules", *Ann. Inst. Fourier (Grenoble)* **50**:2 (2000), 537–592. MR 2001h:17017 Zbl 0962.17002

[Mazorchuk 2010] V. Mazorchuk, *Lectures on $\mathfrak{sl}_2(\mathbb{C})$-modules*, Imperial College Press, London, 2010. MR 2011b:17019 Zbl 05603935

[Mazorchuk and Stroppel 2011] V. Mazorchuk and C. Stroppel, "$\mathfrak{sl}_n$-modules and deformations of certain Brauer tree algebras", preprint, 2011. arXiv 1001.2633

Received April 12, 2010.

VOLODYMYR MAZORCHUK
DEPARTMENT OF MATHEMATICS
UPPSALA UNIVERSITY
BOX 480
751 06 UPPSALA
SWEDEN
mazor@math.uu.se
http://www2.math.uu.se/~mazor/

CATHARINA STROPPEL
MATHEMATIK ZENTRUM
UNIVERSITÄT BONN
ENDENICHER ALLEE 60
D-53115 BONN
GERMANY
stroppel@math.uni-bonn.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 1, 2011

A CONSTANT MEAN CURVATURE ANNULUS TANGENT TO TWO IDENTICAL SPHERES IS DELAUNEY

SUNG-HO PARK

We show that a compact embedded annulus of constant mean curvature in $\mathbb{R}^3$ tangent to two spheres of the same radius along its boundary curves and having nonvanishing Gaussian curvature is part of a Delaunay surface. In particular, if the annulus is minimal, it is part of a catenoid. We also show that a compact embedded annulus of constant mean curvature with negative meeting a sphere tangentially and a plane at a constant contact angle $\geq \pi/2$ (in the case of positive Gaussian curvature) or $\leq \pi/2$ (in the negative case) is part of a Delaunay surface. Thus, if the contact angle is $\geq \pi/2$ and the annulus is minimal, it is part of a catenoid.

Delaunay surfaces are rotational surfaces (surfaces of revolution) of constant mean curvature in $\mathbb{R}^3$. Besides cylinders and spheres, they are divided into unduloids, nodoids, and (allowing the case of zero mean curvature in the definition, for convenience) the catenoid, recognized long ago [Bonnet 1860] as the only nonplanar minimal surface of rotation in $\mathbb{R}^3$.

Thus a Delaunay surface meets every plane perpendicular to the axis of rotation under a constant angle. Conversely, if a compact surface of constant mean curvature meets two parallel planes in constant contact angles, it is part of a Delaunay surface. This can be proved by using Alexandrov's moving plane argument [Alexandrov 1962; Hopf 1989] with planes perpendicular to the parallel planes.

A compact immersed minimal annulus meeting two parallel planes in constant contact angles is also part of a catenoid. This result is not true when the constant mean curvature is nonzero: Wente [1995] constructed examples of immersed constant mean curvature annuli in a slab or in a ball meeting the boundary planes or the boundary sphere perpendicularly. Compared to the above first case, we may ask whether a compact minimal annulus or a compact embedded constant mean curvature annulus meeting two spheres in constant contact angles is part of a catenoid or of a plane. In [Park and Pyo $\geq$ 2011], it is shown that if a compact embedded minimal annulus meets two concentric spheres perpendicularly then the minimal annulus is part of a plane.

Supported by the Hankuk University of Foreign Studies Research Fund of 2011.
MSC2000: 53A10.
Keywords: minimal annulus, contact angle, sphere, parallel surface.

In this paper, we show that a compact embedded constant mean curvature annulus $\mathscr{A}$ in $\mathbb{R}^3$ meeting two spheres S_1 and S_2 of the same radius ρ tangentially and having nonvanishing Gaussian curvature K is part of a Delaunay surface. More precisely, depending on the values of K and the mean curvature H we have three cases: (i) $K < 0$ and $H > -1/\rho$, in which case $\mathscr{A}$ is part of a unduloid if $H < 0$, part of a catenoid if $H = 0$ and part of a nodoid if $H > 0$, (ii) $K > 0$ and $-1/\rho < H < -1/2\rho$, in which case $\mathscr{A}$ is part of a unduloid, and (iii) $K > 0$ and $H < -1/\rho$, in which case $\mathscr{A}$ is part of a nodoid. In the first two cases, $\mathscr{A}$ stays outside of the balls B_1 and B_2 bounded by S_1 and S_2. If (iii) holds, then $\mathscr{A} \subset B_1 \cap B_2$.

We also show that a compact embedded constant mean curvature annulus $\mathscr{B}$ in $\mathbb{R}^3$ with negative (respectively, positive) Gaussian curvature meeting a unit sphere tangentially and a plane in constant contact angle $\geq \pi/2$ (respectively, $\leq \pi/2$) is part of a Delaunay surface. In particular, a compact embedded minimal annulus in $\mathbb{R}^3$ meeting a sphere tangentially and a plane in constant contact angle $\geq \pi/2$ is part of a catenoid.

To prove Theorems 3.1 and 3.2, we use the $-\rho$-parallel surface $\tilde{\mathscr{A}}$ of $\mathscr{A}$ (respectively, $\widetilde{\mathscr{B}}$ of $\mathscr{B}$), that is, the parallel surface of $\mathscr{A}$ (respectively, of $\mathscr{B}$) with distance ρ in the direction to the centers of the spheres. We use Alexandrov's moving plane argument [Alexandrov 1962; Hopf 1989] to prove that $\tilde{\mathscr{A}}$ and $\widetilde{\mathscr{B}}$ are rotational. Since $\tilde{\mathscr{A}}$ and $\widetilde{\mathscr{B}}$ are the parallel surfaces of $\mathscr{A}$ and $\mathscr{B}$ respectively, $\mathscr{A}$ and $\mathscr{B}$ are also rotational and, hence, are part of a Delaunay surface or part of a catenoid.

1. Constant mean curvature annulus meeting spheres tangentially

In the following, we may assume that the spheres have radius 1. Let $\mathscr{A}$ be a compact embedded annulus with constant mean curvature H meeting two unit spheres S_1 and S_2 tangentially along the boundary curves γ_1 and γ_2. We fix the unit normal N of $\mathscr{A}$ in such a way that N points away from the center of S_i along each γ_i. Let $Y : A(1, R) \to \mathbb{R}^3$ be a conformal parametrization of $\mathscr{A}$ from an annulus $A(1, R) = \{(x, y) \in \mathbb{R}^2 : 1 \leq \sqrt{x^2 + y^2} \leq R\}$. We define X by $X = Y \circ \exp$ on the strip $B = \{(u, v) \in \mathbb{R}^2 : 0 \leq u \leq \log R\}$. Then X is periodic with period 2π. Let $z = u + iv$ and $\lambda^2 := |X_u|^2 = |X_v|^2$ with $\lambda > 0$.

Let h_{ij}, $i, j = 1, 2$, be the coefficients of the second fundamental form of X with respect to N. Note that the Hopf differential $\phi(z)\, dz^2 = (h_{11} - h_{22} - 2ih_{12})\, dz^2$ is holomorphic for constant mean curvature surfaces [Hopf 1989]. The theorem of Joachimsthal [do Carmo 1976] says that γ_1 and γ_2 are curvature lines of $\mathscr{A}$. Hence $h_{12} \equiv 0$ on $u = 0$ and $u = \log R$. Since h_{12} is harmonic and periodic, we have $h_{12} \equiv 0$ on B. This implies that z is a conformal curvature coordinate and $h_{11} - h_{22}$ is constant [McCuan 1997]. Let $c = h_{11} - h_{22}$. If $\mathscr{A}$ is minimal, then we have $K < 0$ and $c = 2h_{11} > 0$ by the choice of N. When $H = -1$, $\mathscr{A}$ is part of the unit sphere $S_1 = S_2$ by the boundary comparison principle for the mean curvature operator [Gilbarg and Trudinger 2001]. We assume that $H \neq -1$ in the following. The principal curvatures of $\mathscr{A}$ are

$$\kappa_1 = H + \frac{c}{2\lambda^2} \quad \text{and} \quad \kappa_2 = H - \frac{c}{2\lambda^2}. \tag{1}$$

We use for γ_1 and γ_2 the parametrizations $\gamma_1(v) = X(0, v)$ and $\gamma_2(v) = X(\log R, v)$, for $v \in [0, 2\pi)$. In the following, we assume that $\mathscr{A}$ has nonzero Gaussian curvature.

Lemma 1.1. *Each $\gamma_i(v)$, $i = 1, 2$, has constant speed $\sqrt{c/2(1+H)}$ and κ_2 is -1 on γ_1 and γ_2. As spherical curves, γ_1 and γ_2 are convex. On $\mathscr{A} \setminus \partial\mathscr{A}$, we have $\lambda^2 < c/2(1+H)$ when $K < 0$ and $\lambda^2 > c/2(1+H)$ when $K > 0$.*

Proof. The curvature vector of $\gamma_1(v)$ is

$$\begin{aligned} \vec{\kappa} &= \frac{1}{|X_v|}\frac{d}{dv}\left(\frac{X_v}{|X_v|}\right) = \frac{1}{|X_v|^2}X_{vv} - \frac{X_v}{|X_v|^4}(X_v \cdot X_{vv}) \\ &= \frac{1}{\lambda^2}\left(-\frac{\lambda_u}{\lambda}X_u + h_{22}N\right). \end{aligned} \tag{2}$$

Let the center of S_1 be the origin of $\mathbb{R}^3$. Since $\mathscr{A}$ is tangential to S_1 along γ_1, we have $N(0, v) = X(0, v) = \gamma_1(v)$ on γ_1. Since γ_1 is on the unit sphere S_1, the curvature vector $\vec{\kappa}$ of γ_1 satisfies $(\vec{\kappa} \cdot \gamma_1)(v) = -1$. Hence we have $\kappa_2 = h_{22}/\lambda^2 = -1$ on γ_1. Since $\lambda^2 = |\gamma_{1v}|^2$ on γ_1, we have $|\gamma_{1v}| = \sqrt{c/2(1+H)}$ from (1). By choosing the center of S_2 as the origin of $\mathbb{R}^3$, we get the results for γ_2.

The Gaussian curvature K satisfies

$$\Delta \log \lambda = -K\lambda^2,$$

where $\Delta = \partial^2/\partial u^2 + \partial^2/\partial v^2$. We can rewrite this equation as

$$\lambda \Delta \lambda = |\nabla \lambda|^2 - K\lambda^4. \tag{3}$$

Since $\lambda_v(0, v) = 0$ and $\lambda_v(\log R, v) = 0$ and $K \neq 0$, λ does not have interior maximum when $K < 0$, and does not have interior minimum when $K > 0$. Since $\lambda^2 = c/2(1+H)$ on γ_1 and γ_2, it follows that $\lambda^2 < c/2(1+H)$ on $\mathscr{A} \setminus \partial\mathscr{A}$ when $K < 0$ and $\lambda^2 > c/2(1+H)$ when $K > 0$. Moreover we have $\lambda_u \le 0$ on $u = 0$ and $\lambda_u \ge 0$ on $u = \log R$ when $K < 0$ and $\lambda_u \ge 0$ on $u = 0$ and $\lambda_u \le 0$ on $u = \log R$ when $K > 0$. Since $X_u/|X_u| \in TS_i$ is perpendicular to γ_i, the geodesic curvature of γ_i as a spherical curve is $\vec{\kappa} \cdot (X_u/|X_u|) = -\lambda_u/\lambda^2$. Hence γ_1 and γ_2 are convex as spherical curves. □

Remark 1.2. If $\lambda^2 \equiv c/2(1+H)$ on $\mathscr{A}$, then $K \equiv 0$ and $\mathscr{A}$ is part of a cylinder.

2. The −1-parallel surface

The -1-parallel surface $\tilde{\mathscr{A}}$ of $\mathscr{A}$ is defined by

$$\tilde{X} = X - N.$$

The image of γ_1 (respectively, of γ_2) in $\tilde{\mathscr{A}}$ is a point corresponding to the center of S_1 (respectively, of S_2). We denote the centers of S_1 and S_2 by O and O_2 for simplicity. We fix the unit normal $\tilde{N}$ of $\tilde{\mathscr{A}}$ to be N. Since $z = u + iv$ is a curvature coordinate of X, we have

$$\tilde{X}_u = \left(1 + \frac{h_{11}}{\lambda^2}\right) X_u \quad \text{and} \quad \tilde{X}_v = \left(1 + \frac{h_{22}}{\lambda^2}\right) X_v. \tag{4}$$

Since $\kappa_2 = -1$ on γ_i by Lemma 1.1, $\tilde{X}$ is singular for $u = 0$ and $u = \log R$. By Lemma 1.1, we have $\lambda^2 \neq c/2(1 + H)$ on $\mathscr{A} \setminus \partial\mathscr{A}$, which implies that $1 + \kappa_2 \neq 0$ on $\mathscr{A} \setminus \partial\mathscr{A}$. When $K < 0$, we have $\kappa_1 > 0$ on $\mathscr{A} \setminus \partial\mathscr{A}$. Hence $\tilde{X}$ is regular for $0 < u < \log R$ and we have $H > -1$.

Now suppose that $K > 0$. Since $\kappa_2 = -1$ on γ_i by Lemma 1.1, we have $\kappa_1 < 0$ and $H < -1/2$. We consider two cases separately: $H < -1$ and $-1 < H < -1/2$. If $H < -1$, then $c < 0$ from $\lambda^2 = c/2(1 + H) > 0$ on γ_i. Hence we have $\kappa_1 < -1$, which implies that $\tilde{X}$ is regular for $0 < u < \log R$. If $-1 < H < -1/2$, then we must have $c > 0$. This implies that $1 + \kappa_1 \neq 0$. Otherwise we have $0 < 2\lambda^2(1 + H) = -c$, which contradicts $c > 0$. Hence $\tilde{X}$ is regular for $0 < u < \log R$.

Remark 2.1. When $K < 0$ or $K > 0$ and $-1 < H < -1/2$, $\mathscr{A}$ stays outside of the balls B_1 and B_2 bounded by S_1 and S_2. If $K > 0$ and $H < -1$, then $\mathscr{A} \subset B_1 \cap B_2$.

Lemma 2.2. *The mean curvature $\tilde{H}$ and the Gaussian curvature $\tilde{K}$ of $\tilde{\mathscr{A}}$ satisfies $(1 + H)\tilde{K} = (1 + 2H)\tilde{H} - H$. On $\tilde{\mathscr{A}} \setminus \{O, O_2\}$, we have the following*:

(i) *If $K < 0$ and $H > -1$, then $\tilde{\kappa}_1 > 0$, $\tilde{\kappa}_2 > 1$ and $\tilde{H} > 1$.*

(ii) *If $K > 0$ and $-1 < H < -1/2$, then $0 < c/2\lambda^2(1+H) < \min\{1, -H/(1+H)\}$, $\tilde{\kappa}_1 < 0$, $\tilde{\kappa}_2 < H/(1 + H)$ and $\tilde{H} < H/(1 + H)$.*

(iii) *If $K > 0$ and $H < -1$, then $0 < c/2\lambda^2(1 + H) < 1$, $\tilde{\kappa}_1 > (1 + 2H)/2(1 + H)$, $\tilde{\kappa}_2 > H/(1 + H)$ and $\tilde{H} > H/(1 + H)$.*

Proof. Since

$$\tilde{h}_{12} = N \cdot \tilde{X}_{uv} = \left(1 + \frac{h_{11}}{\lambda^2}\right)(N \cdot X_{uv}) = 0,$$

(u, v) is a curvature coordinate (not conformal) for $\tilde{\mathscr{A}}$ except for O and O_2. We have

$$\tilde{h}_{11} = N \cdot \tilde{X}_{uu} = \left(1 + \frac{h_{11}}{\lambda^2}\right) h_{11}, \quad \tilde{h}_{22} = N \cdot \tilde{X}_{vv} = \left(1 + \frac{h_{22}}{\lambda^2}\right) h_{22}.$$

The principal curvatures of $\tilde{\mathscr{A}}$ are

$$\tilde{\kappa}_1 = \frac{\kappa_1}{1+\kappa_1} = \frac{H/(1+H) + \big(c/2\lambda^2(1+H)\big)}{1+\big(c/2\lambda^2(1+H)\big)},$$
$$\tilde{\kappa}_2 = \frac{\kappa_2}{1+\kappa_2} = \frac{H/(1+H) - \big(c/2\lambda^2(1+H)\big)}{1-\big(c/2\lambda^2(1+H)\big)}.$$

From $\kappa_1 + \kappa_2 = 2H$, we have $H = \dfrac{\tilde{H}-\tilde{K}}{1-2\tilde{H}-\tilde{K}}$ or $(1+H)\tilde{K} = (1+2H)\tilde{H} - H$. It is straightforward to see that

$$\tilde{H} = \frac{H/(1+H) - \big(c/2\lambda^2(1+H)\big)^2}{1-\big(c/2\lambda^2(1+H)\big)^2}.$$

Note that $\kappa_2 < 0$ on $\mathscr{A}$. First suppose that $K < 0$. Then we have $\kappa_1 > 0$, which implies that $\tilde{\kappa}_1 = \kappa_1/(1+\kappa_1) > 0$. Since $c/2\lambda^2(1+H) > 1$ by Lemma 1.1, we have $\tilde{\kappa}_2 > 1$ and $\tilde{H} > 1$.

When $K > 0$, we have $\kappa_1 = H + c/2\lambda^2 < 0$. If $-1 < H < -1/2$, then we have $c > 0$ because $\lambda^2 = c/2(1+H) > 0$ on γ_i. It follows that $c/2\lambda^2(1+H) < -H/(1+H)$. By Lemma 1.1, we also have $c/2\lambda^2(1+H) < 1$. Therefore $0 < c/2\lambda^2(1+H) < \min\{1, -H/(1+H)\}$. It is easy to see that $\tilde{\kappa}_1 < 0$, $\tilde{\kappa}_2 < H/(1+H) < 0$ and $\tilde{H} < H/(1+H) < 0$.

When $K > 0$ and $H < -1$, we have $c < 0$ and $0 < c/2\lambda^2(1+H) < 1$. It is straightforward to see that $\tilde{\kappa}_1 > (1+2H)/(1+H)$, $\tilde{\kappa}_2 > H/(1+H)$ and $\tilde{H} > H/(1+H)$. □

This lemma says that $\tilde{\mathscr{A}}$ is a linear Weingarten surface with two singular points O and O_2 and is positively curved outside O and O_2.

Lemma 2.3. *$\tilde{\mathscr{A}}$ is embedded.*

Proof. Let $\nu(v) = (X_u/|X_u|)(0, v)$. Note that ν is a closed curve in the unit sphere S_1. We claim that ν is *convex as a spherical curve*. Otherwise, there is a great circle η intersecting the image of ν at no less than 3 points $\nu(v_1), \ldots, \nu(v_n)$. (It is possible that ν maps an interval $(v_a, v_b) \subset [0, 2\pi)$ into a single point. We choose the v_i's in such a way that ν maps no two v_i's to the same point.) Each $\nu(v_i)$ determines a great circle $\mathbb{S}^1_{v_i} \subset S_1$ contained in the plane perpendicular to $\nu(v_i)$. At each $\gamma_1(v_i)$, γ_1 is tangent to $\mathbb{S}^1_{v_i}$. Since η and $\mathbb{S}^1_{v_i}$ are perpendicular, γ_1 cannot be convex when $n \geq 3$. Hence ν intersect every geodesic of S_1 at no more than two points. This shows that ν is convex as a spherical curve. Similarly, $(X_u/|X_u|)(\log R, v)$ is also convex as a spherical curve.

Since $\tilde{\mathscr{A}}$ is a parallel surface of $\mathscr{A}$, the tangent cone $\mathrm{Tan}(O, \tilde{\mathscr{A}})$ of $\tilde{\mathscr{A}}$ at O is the cone formed by rays from O through ν. Since ν is a convex spherical curve, $\mathrm{Tan}(O, \tilde{\mathscr{A}})$ is convex. This shows that a small neighborhood of O in $\tilde{\mathscr{A}}$ is embedded and nonnegatively curved as a metric space [Alexandrov 1948]. Similarly, there is a neighborhood of O_2 in $\tilde{\mathcal{A}}$ which is embedded and nonnegatively curved as a metric space.

Hadamard showed that a closed surface S in $\mathbb{R}^3$ with strictly positive Gaussian curvature is the boundary of a convex body [Hopf 1989]. In particular, S is embedded. Alexandrov [1948] generalized Hadamard's theorem to nonnegatively curved metric spaces. Since $\tilde{\mathcal{A}}$ is a nonnegatively curved closed metric space, $\tilde{\mathcal{A}}$ is embedded. □

Remark 2.4. We have $\nu_v = (\lambda_u/\lambda^2)X_v$. At points where $\lambda_u \neq 0$, the curvature vector of ν is

$$\vec{\kappa}_\nu = \frac{1}{\lambda_u}\left(-\frac{\lambda_u}{\lambda}X_u + h_{22}N\right).$$

The geodesic curvature of ν as a spherical curve $\vec{\kappa}_\nu \cdot N = h_{22}/\lambda_u$.

3. Main results

We use Alexandrov's moving plane argument [Alexandrov 1962; Hopf 1989] to prove the theorems.

Theorem 3.1. *A compact embedded constant mean curvature annulus $\mathcal{A}$ with nonvanishing Gaussian curvature meeting two spheres S_1 and S_2 of the same radius tangentially is part of a Delaunay surface. In particular, if $\mathcal{A}$ is minimal, then $\mathcal{A}$ is part of a catenoid.*

Proof. We suppose that the radius of S_1 and S_2 is 1. By Lemma 2.2 and Lemma 2.3, $\tilde{\mathcal{A}}$ is a compact embedded surface with two singular points O and O_2 and satisfying $(1+H)\tilde{K} = (1+2H)\tilde{H} - H$ at regular points. A small neighborhood of a regular point of $\tilde{\mathcal{A}}$ can be represented as the graph of a function $f(x, y)$ satisfying

$$\begin{aligned}(5)\quad & 2(1+H)(f_{xx}f_{yy} - f_{xy}^2) + 2H(1+f_x^2+f_y^2)^2 \\ &= (1+2H)\big((1+f_y^2)f_{xx} - 2f_xf_yf_{xy} + (1+f_x^2)f_{yy}\big)(1+f_x^2+f_y^2)^{1/2}.\end{aligned}$$

This equation can be rewritten as

$$\det\big(2(1+H)D^2f + A(Df)\big) = W^4, \tag{6}$$

where

$$A(Df) = -(1+2H)\begin{pmatrix}(1+f_x^2)W & f_xf_yW \\ f_xf_yW & (1+f_y^2)W\end{pmatrix} \quad \text{and} \quad W = \sqrt{1+f_x^2+f_y^2}.$$

Equation (6) is elliptic with respect to f if $2(1+H)D^2f + A(Df)$ is positive definite. Since $\det\big(2(1+H)D^2f + A(Df)\big) = W^4 > 0$, this happens if

$$(7)\quad \operatorname{Tr}\big(2(1+H)D^2f + A(Df)\big) = 2(1+H)\Delta f - (1+2H)(2+f_x^2+f_y^2)W$$

is strictly positive.

First we consider the case $K < 0$. Since $\tilde{H} > 1$ by Lemma 2.2, we have

$$\Delta f + f_y^2 f_{xx} - 2 f_x f_y f_{xy} + f_x^2 f_{yy} > 2W^{3/2}, \tag{8}$$

for f representing $\tilde{\mathscr{A}}$. We may assume that f is defined on $B(0, \epsilon) \subset T_p\tilde{\mathscr{A}}$ so that $\nabla f(0) = \vec{0}$ and $D^2 f$ is diagonal. For sufficiently small $\epsilon = \epsilon(p)$, (8) implies that (7) is strictly positive. Hence (6) is elliptic with respect to f representing $\tilde{\mathscr{A}}$.

When $-1 < H < -1/2$, (7) is automatically satisfied.

Now we consider the case $K > 0$ and $H < -1$. Since $\tilde{H} > H/(1 + H)$ by Lemma 2.2, we have

$$\Delta f + f_y^2 f_{xx} - 2 f_x f_y f_{xy} + f_x^2 f_{yy} > \frac{2H}{1+H} W^{3/2}. \tag{9}$$

Assuming that f is defined on $B(0, \epsilon) \subset T_p\tilde{\mathscr{A}}$ with $\nabla f(0) = \vec{0}$ and $D^2 f$ is diagonal, (9) implies that

$$\Delta f - \frac{1+2H}{2(1+H)}(2 + f_x^2 + f_y^2) W$$

is strictly positive for sufficiently small ϵ. So $\det\big(-2(1+H)D^2 f - A(Df)\big) = W^4$ is elliptic for f representing $\widetilde{A}$. The ellipticity of (6) for f representing $\widetilde{A}$ enables us to use the maximum principle and the boundary point lemma [Gilbarg and Trudinger 2001].

Since $\tilde{\mathscr{A}}$ is convex and embedded, we can use Alexandrov's moving plane argument [Alexandrov 1962; Hopf 1989] to show that $\tilde{\mathscr{A}}$ is rotational as follows. Let Π_θ be the plane containing the line segment $\overline{OO_2} \subset \mathbb{R}^3$ and making angle θ with a fixed vector $\vec{E}$ which is perpendicular to $\overline{OO_2}$. Fix a positive constant L such that each plane Π_θ^L that is parallel to Π_θ with distance L from Π_θ does not meet $\tilde{\mathscr{A}}$ for all θ. Let Π_θ^l be the plane between Π_θ^L and Π_θ with distance l from Π_θ. When Π_θ^l intersects $\tilde{\mathscr{A}}$, we reflect the Π_θ^L side part of $\tilde{\mathscr{A}}$ about Π_θ^l. Denote this reflected surface by $\tilde{\mathscr{A}}_{l,\theta}^{\text{ref}}$. As we decrease l from L, there might be a first $l_\theta \geq 0$ for which $\tilde{\mathscr{A}}_{l_\theta,\theta}^{\text{ref}}$ is tangent to $\tilde{\mathscr{A}}$ at an interior point or at a boundary point of $\partial\tilde{\mathscr{A}}_{l_\theta,\theta}^{\text{ref}}$. We call this point the *first touch point*. If there is no nonnegative l with the first touch point, we repeat the process for $\Pi_{\theta+\pi}^L$ to find $l_{\theta+\pi}$, which must be positive. At the first touch point, we apply the comparison principles for (5) to see that the part of $\tilde{\mathscr{A}}$ in the Π_θ side and $\tilde{\mathscr{A}}_{l_\theta,\theta}^{\text{ref}}$ are identical and, hence, $l_\theta = 0$. This implies that Π_θ is a symmetry plane for $\tilde{\mathscr{A}}$. Since θ can be chosen arbitrarily, $\tilde{\mathscr{A}}$ should be rotational and, hence, $\mathscr{A}$ is also rotational. Since the Delaunay surfaces and the catenoid are the only nonplanar rotational minimal and constant mean curvature surfaces, $\mathscr{A}$ is part of a Delaunay surface or part of a catenoid. □

We used the embeddedness of $\mathscr{A}$ to prove that $\tilde{\mathscr{A}}$ is embedded. Whether there is a nonembedded minimal or constant mean curvature annulus meeting two unit spheres tangentially is an interesting question. Moreover we raise the following questions.

(1) Is a compact immersed minimal annulus or a compact embedded minimal or constant mean curvature surface meeting a sphere perpendicularly or in constant contact angles part of a catenoid or part of a Delaunay surface? Nitsche showed that an immersed disk type minimal or constant mean curvature surface meeting a sphere in constant contact angle is either a flat disk or a spherical cap [Nitsche 1985].

(2) Is a compact immersed minimal annulus or a compact embedded minimal or constant mean curvature surface meeting two spheres in constant contact angles part of a catenoid or a plane or part of a Delaunay surface?

(3) Is a compact immersed minimal or constant mean curvature annulus or a compact embedded minimal or constant mean curvature surface meeting a sphere and a plane in constant contact angles part of a catenoid or part of a Delaunay surface? We give an affirmative answer to this problem in a special case in the following.

Theorem 3.2. *A compact embedded constant mean curvature annulus $\mathcal{B}$ with negative (respectively, positive) Gaussian curvature meeting a sphere tangentially and a plane in constant contact angle $\geq \pi/2$ (respectively, $\leq \pi/2$) is part of a Delaunay surface. In particular, if $\mathcal{B}$ is minimal and the constant contact angle is $\geq \pi/2$ then $\mathcal{B}$ is part of a catenoid.*

The angle is measured between the outward conormal of $\mathcal{B}$ and the outward conormal of the bounded domain in Π bounded by the boundary curve. Since the proof of this theorem is similar to that of Theorem 3.1, we omit some previously proved details.

Proof. Denote the sphere by S_2 and the plane by Π. We may assume that the radius of S_2 is 1. Let α be the constant contact angle between $\mathcal{B}$ and Π. If $\alpha = \pi/2$, then we can reflect $\mathcal{B}$ about Π to get a constant mean curvature annulus meeting two unit spheres tangentially. Hence $\mathcal{B}$ is part of a catenoid or a Delaunay surface by Theorem 3.1.

In the following, we assume that $\alpha \neq \pi/2$. As in the case for $\mathcal{A}$ in Section 1, there is a conformal parametrization X of $\mathcal{B}$ from a strip $\{(u, v) \in \mathbb{R}^2 : 0 \leq u \leq \log R\}$ for which $z = u + iv$ is a curvature coordinate. We fix the normal N of $\mathcal{B}$ to point away from the center of S_2. Let $c_1(v) = X(0, v)$ be on Π and $c_2(v) = X(\log R, v)$ be on S_2 with $\partial X_3/\partial u > 0$ along c_1. As in Lemma 1.1, c_2 has constant speed $\sqrt{c/2(1+H)}$ and $\kappa_2 = -1$ along c_2. Since $K \neq 0$ on $\mathcal{B}$ and $z = u + iv$ is a curvature coordinate, we have $\kappa_2 < 0$ on c_1. The curvature of c_1 is $|\vec{\kappa}| = -\kappa_2/\sin\alpha > 0$, which shows that c_1 is locally convex. Since c_1 is a Jordan curve, it is convex.

First, we assume that $K < 0$ and $\alpha > \pi/2$. Since $(\vec{\kappa}/|\vec{\kappa}|)\cdot(X_u/X_u|) = \cos\alpha < 0$ on c_1, it follows from (2) that $\lambda_u > 0$ on c_1. Since $\lambda_v(\log R, v) = 0$ (see Lemma 1.1), it follows from (3) that $\lambda_u \geq 0$ on c_2. Otherwise, λ will have an interior maximum, which contradicts (3). Hence we have $\lambda^2 < c/2(1+H)$ on $\mathcal{B} \setminus c_2$. Note that $\kappa_1 > 0$ and $\kappa_2 < 0$ in $\mathcal{B}$. From $\lambda_u \leq 0$ on c_2, we see that c_2 is convex as a spherical curve (see Lemma 1.1). Arguing as in the proof of Lemma 2.3, we see that $(X_u/|X_u|)(\log R, v)$ is also convex as a spherical curve.

When $K > 0$ and $\alpha < \pi/2$, we have $(\vec{\kappa}/|\vec{\kappa}|)\cdot(X_u/|X_u|) = \cos\alpha > 0$ on c_1. Hence $\lambda_u < 0$ on c_1. Since $\lambda_v(\log R, v) = 0$, it follows from (3) that λ does not have interior minimum. Then we have $\lambda_u \leq 0$ on c_2 and $\lambda^2 > c/2(1+H)$ on $\mathcal{B}\setminus c_2$. Note that $\kappa_1 < 0$ and $\kappa_2 < 0$ in $\mathcal{B}$. From $\lambda_u \leq 0$ on c_2, it follows that c_2 is convex as a spherical curve. Moreover $(X_u/|X_u|)(\log R, v)$ is convex as a spherical curve (see Lemma 2.3).

Let $\widetilde{\mathcal{B}}$ be the -1-parallel surface of $\mathcal{B}$. As in Section 2, we can show that $\widetilde{\mathcal{B}}$ is regular except for O_2: the image of c_2, and $H > -1$ when $K < 0$ and $H < -1/2$ when $K > 0$. As in Lemma 2.2, we see that mean curvature $\tilde{H}$ and the Gaussian curvature $\tilde{K}$ of $\widetilde{\mathcal{B}}$ satisfies $(1+H)\tilde{K} = (1+2H)\tilde{H} - H$ and (i) if $K < 0$ and $H > -1$, then $\tilde{\kappa}_1 > 0$, $\tilde{\kappa}_2 > 1$ and $\tilde{H} > 1$, (ii) if $K > 0$ and $-1 < H < -1/2$, then $0 < c/2\lambda^2(1+H) < \min\{1, -H/(1+H)\}$, $\tilde{\kappa}_1 < 0$, $\tilde{\kappa}_2 < H/(1+H)$ and $\tilde{H} < H/(1+H)$, and (iii) if $K > 0$ and $H < -1$, then $0 < c/2\lambda^2(1+H) < 1$, $\tilde{\kappa}_1 > (1+2H)/2(1+H)$, $\tilde{\kappa}_2 > H/(1+H)$ and $\tilde{H} > H/(1+H)$.

The convexity of $(X_u/|X_u|)(\log R, v)$ as a spherical curve implies that there is a neighborhood of O_2 in $\widetilde{\mathcal{B}}$ which is embedded and nonnegatively curved as a metric space. Let $\widetilde{\Pi}$ be the plane parallel to Π and containing $\tilde{c}_1$. The curvature of $\tilde{c}_1$ is $|\tilde{\kappa}_2|/\sin\alpha$, which does not vanish. Hence $\tilde{c}_1$ is locally convex. Using the orthogonal projection onto $\widetilde{\Pi}$, $\tilde{c}_1$ may be considered as a $(\sin\alpha)$-parallel curve of c_1 in $\widetilde{\Pi}$. Hence $\tilde{c}_1$ is also a convex Jordan curve.

Suppose that $K < 0$ and $\alpha > \pi/2$. Since $\kappa_1 > 0$, $\tilde{X}_u$ is a positive multiple of X_u by (4). The positivity of $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$ implies that $\widetilde{\mathcal{B}}$ meets $\widetilde{\Pi}$ in constant angle $\pi - \alpha$. Suppose that $K > 0$ and $\alpha < \pi/2$. If $-1 < H < -1/2$, then we have $c > 0$ and $\kappa_1 > -1$. Hence $\tilde{X}_u$ is a positive multiple of X_u by (4). The negativity of $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$ implies that $\widetilde{\mathcal{B}}$ meets $\widetilde{\Pi}$ in constant angle α. When $K > 0$ and $H < -1$, we have $c < 0$ and $\kappa_1 < -1$. Hence $\tilde{X}_u$ is negative multiple of X_u by (4). In this case, $\widetilde{\mathcal{B}}$ lies below $\widetilde{\Pi}$ and $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$ are both positive. It is straightforward to see that $\widetilde{\mathcal{B}}$ meets $\widetilde{\Pi}$ in constant angle α.

Let $\breve{\mathcal{B}}$ be the singular surface obtained from $\widetilde{\mathcal{B}}$ by attaching the disk in $\widetilde{\Pi}$ bounded by $\tilde{c}_1$ to $\widetilde{\mathcal{B}}$. Since $\widetilde{\mathcal{B}}$ meets $\widetilde{\Pi}$ in acute angle, $\breve{\mathcal{B}}$ is a nonnegatively curved metric space. By Alexandrov's generalization [1948] of Hadamard's theorem, $\breve{\mathcal{B}}$ is the boundary of a convex body. Therefore $\breve{\mathcal{B}}$ is embedded. Note again that $\tilde{H}$, $\tilde{K}$, $\tilde{\kappa_1}$ and $\tilde{\kappa_2}$ satisfy the statements of Lemma 2.2. Hence (5) is elliptic for functions representing $\widetilde{\mathcal{B}}$ locally. We can apply Alexandrov's moving plane argument to $\widetilde{\mathcal{B}}$ using planes perpendicular to $\widetilde{\Pi}$ as in the proof of Theorem 3.1 to see that $\widetilde{\mathcal{B}}$ is rotational. Hence $\mathcal{B}$ is rotational and, as a result, is part of a Delaunay surface or part of a catenoid. $\square$

References

[Alexandrov 1948] A. D. Alexandrov, *Vnutrenniaia geometriia vypuklykh poverkhnostei*, OGIZ, Moscow-Leningrad, 1948. Translated as *Die innere Geometrie der konvexen FlAachen*, Akad. Verl., Berlin, 1955. MR 10,619c

[Alexandrov 1962] A. D. Alexandrov, "Uniqueness theorems for surfaces in the large, V", *Amer. Math. Soc. Transl.* (2) **21** (1962), 412–416. MR 27 #698e Zbl 0119.16603

[Bonnet 1860] O. Bonnet, "Mémoire sur l'emploi d'un nouveau système de variables dans l'étude des surfaces courbes", *J. Math. Pures Appl.* (2) **5** (1860), 153–266.

[do Carmo 1976] M. P. do Carmo, *Differential geometry of curves and surfaces*, Prentice-Hall, Englewood Cliffs, N.J., 1976. MR 52 #15253 Zbl 0326.53001

[Gilbarg and Trudinger 2001] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Classics in Math., Springer, Berlin, 2001. Reprint of the 1998 edition. MR 2001k:35004 Zbl 1042.35002

[Hopf 1989] H. Hopf, *Differential geometry in the large*, 2nd ed., Lecture Notes in Math. **1000**, Springer, Berlin, 1989. Notes taken by P. Lax and J. W. Gray, With a preface by S. S. Chern, With a preface by K. Voss. MR 90f:53001 Zbl 0669.53001

[McCuan 1997] J. McCuan, "Symmetry via spherical reflection and spanning drops in a wedge", *Pacific J. Math.* **180**:2 (1997), 291–323. MR 98m:53013 Zbl 0885.53009

[Nitsche 1985] J. C. C. Nitsche, "Stationary partitioning of convex bodies", *Arch. Rational Mech. Anal.* **89**:1 (1985), 1–19. MR 86j:53013 Zbl 0572.52005

[Park and Pyo $\geq$ 2011] S. Park and J. Pyo, "Embedded minimal surfaces meeting 1 or 2 spheres in constant angle 0 or $\pi/2$", In preparation.

[Wente 1995] H. C. Wente, "Tubular capillary surfaces in a convex body", pp. 288–298 in *Advances in geometric analysis and continuum mechanics* (Stanford, CA, 1993), edited by P. Concus and K. Lancaster, Int. Press, Cambridge, MA, 1995. MR 96j:53009 Zbl 0854.53012

Received February 7, 2010.

SUNG-HO PARK
GRADUATE SCHOOL OF EDUCATION
HANKUK UNIVERSITY OF FOREIGN STUDIES
270 IMUN-DONG
DONGDAEMUN-GU
SEOUL 130-791
SOUTH KOREA

sunghopark@hufs.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 1, 2011

A NOTE ON THE TOPOLOGY OF THE COMPLEMENTS OF FIBER-TYPE LINE ARRANGEMENTS IN $\mathbb{CP}^2$

SHENG-LI TAN, STEPHEN S.-T. YAU AND FEI YE

We prove that $B\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\})$ is a $K(\pi, 1)$ space, where π is the mapping class group of an $(n+1)$-punctured sphere. As a consequence we derive that the center-projecting braid monodromy of a fiber-type projective line arrangement determines the diffeomorphic type of its complement.

1. Introduction

A complex arrangement of hyperplanes $\mathscr{A}$ is a finite collection of $\mathbb{C}$-linear subspaces of dimension $n-1$ in $\mathbb{C}^n$. Denote by $M(\mathscr{A}) = \mathbb{C}^n - \bigcup\{H : H \in \mathscr{A}\}$ the complement of $\mathscr{A}$. The theory of arrangements of hyperplanes is not only closely related to singularity theory, algebraic geometry and hypergeometric function theory, but also has its own interesting questions. For example, one of the central problems is to find the relationship between the topological structure and combinatorial structure of an arrangement. In other words, one wants to understand the topological properties of $M(\mathscr{A})$ and how to classify the arrangements according to their combinatorics. To study such problems, mathematicians have developed many techniques, for example, the lattice-isotopy theorem and braid monodromy method which will be used in this paper. The lattice-isotopy theorem was used in [Jiang and Yau 1994; Wang and Yau 2005; 2007; 2008; Yau and Ye 2009] to derive the structures of so-called nice arrangements and prove that their differential structures are determined by their combinatorics. Braid monodromy method has been widely used to study the topology of complements of plane algebraic curves and line arrangements; see, for example, [Moishezon 1981; Cohen and Suciu 1997; Dung 1999; Kulikov and Taĭkher 2000; Cohen 2001; Artal Bartolo et al. 2003; 2007]. However, there are still many kinds of arrangements for which we are far from understanding the relationship between the topology and combinatorics. This is true even in the case of a fiber-type projective line arrangement, that is, the projectivization of a fiber-type hyperplane arrangement in $\mathbb{C}^3$. Cohen [2001] studied the structure and properties of the fundamental group of the complement of

MSC2000: primary 14N20, 52C35; secondary 57R22, 37E30.
Keywords: fiber-type line arrangement, braid monodromy, differentiable fiber bundle.

a fiber-type arrangement. He showed that the Whitehead group of the fundamental group of the complement of a fiber-type arrangement is trivial, which was conjectured by Aravinda, Farrell and Roushon [2000]. He also proved the conjecture by Xicoténcatl [1997] on the structure of the Lie algebra associated to the lower central series of the fundamental group. Besides that, we still don't know whether the combinatorics of a line arrangement determines the topology of its complement.

It is well known that fiber-type projective line arrangements are the same as supersolvable projective line arrangements (see, for example, [Orlik and Terao 1992]). Moreover, Jiang et al. [2001] studied the geometric characterization of supersolvable line arrangements in $\mathbb{CP}^2$. They showed that any fiber-type line arrangement in $\mathbb{CP}^2$ has a center through which every multiple point of the arrangement has a line in the arrangement passing. The complement of a fiber-type projective line arrangement is a locally trivial fiber bundle with punctured sphere as base and fibers. It is a natural question how to classify the complements of fiber-type line arrangements in $\mathbb{CP}^2$ by center-projecting braid monodromies (see Definition/Construction 4.1). One of the applications of such braid monodromies is that the fundamental group of a fiber-type projective line arrangement is isomorphic to the semidirect product of free groups $\mathbf{F}_m \rtimes_\phi \mathbf{F}_n$, where ϕ is the center-projecting braid monodromy [Cohen 2001]. The purpose of this paper is to use this center-projecting braid monodromy to study the topology of the complement.

It is well known that the braid monodromy determines the homotopy type of the complement of an algebraic curve [Libgober 1986]. In this paper, we prove that for a fiber-type projective line arrangement its center-projecting braid monodromy determines even the diffeomorphic type of its complement, consequently, determines the homotopy type.

Main Theorem. *Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two fiber-type projective line arrangements. If they have the same center-projecting braid monodromies, then their complements are diffeomorphic.*

The key ingredient of the proof is Proposition 3.1. It shows that the classifying space of the structure group of the complement, the orientation-preserving diffeomorphism group $\mathrm{Diff}_+(S^2, \{x_1, x_2, \dots, x_{n+1}\})$ of S^2 fixing the set $\{x_1, x_2, \dots, x_{n+1}\}$, is a $K(\pi, 1)$ space, where π is the mapping class group of a punctured sphere. Morita [1987] explained that $B\mathrm{Diff}_0(\Sigma_g)$, where $\mathrm{Diff}_0(\Sigma_g)$ is the subgroup of diffeomorphisms of a Riemann surface Σ_g which can be deformed to the identity, is contractible for $g \geq 2$, using a result of Earle and Eells [1967]. However, in our case, Earle and Eells' result is not applicable.

2. The complements of fiber-type line arrangements in $\mathbb{CP}^2$

We begin by recalling some definitions which one can find in [Orlik 1992].

Definition 2.1. A hyperplane arrangement $\mathcal{A}$ is called *strictly linear fibered* if, after a suitable linear change of coordinates, the restriction of the projection of $M(\mathcal{A})$ to the first $(n-1)$ coordinates is a fiber bundle projection with base the complement $M(\mathcal{B})$ of an arrangement $\mathcal{B}$ in $\mathbb{C}^{(n-1)}$, and fiber the complement C_* of finitely many points in $\mathbb{C}^3$.

Definition 2.2. A 1-arrangement $\mathcal{A}_1$ of finitely points in $\mathbb{C}$ is fiber-type. An n-arrangement is fiber-type if it is strictly linear fibered over a fiber-type $(n-1)$-arrangement. A fiber-type projective line arrangement $\mathcal{A}^*$ in $\mathbb{CP}^2$ is the projectivization of a fiber-type 3-arrangement $\mathcal{A}_3$ in $\mathbb{C}$.

Definition 2.3. Let $\mathcal{A}^*$ be an arrangement in $\mathbb{CP}^2$ and c be a point in the lattice $L(\mathcal{A}^*)$. The point c is called a center of $\mathcal{A}^*$ if for any multiple point p of $\mathcal{A}^*$ there is a line l in $\mathcal{A}^*$ connecting c and p.

Let $\mathcal{A}^*$ be a fiber-type projective line arrangement with complement $M(\mathcal{A}^*)$. We now recall some geometric characterizations of fiber-type line arrangements.

Theorem 2.4 [Terao 1986]. *An arrangement $\mathcal{A}$ is fiber-type if and only if $L(\mathcal{A})$ is supersolvable.*

Theorem 2.5 [Jiang et al. 2001]. *Let $\mathcal{A}$ be a 3-arrangement. The lattice $L(\mathcal{A})$ is a supersolvable if and only if the projectivization $\mathcal{A}^*$ has a center.*

Using the above two theorems, the structure of the complements of fiber-type projective line arrangements can be characterized as follows.

Remark 2.6 [Jiang et al. 2001]. Let c be the center of $\mathcal{A}^*$. After a suitable linear transformation, we may assume that $c = (0:1:0)$ and that one of the lines passing through c is the line at infinity, $z = 0$. We can view $M(\mathcal{A}^*)$ as a subset of $\mathbb{C}^2$. Assume that the lines passing c are defined by the equations

$$z = 0, \quad x = k_1 z, \quad \dots, \quad x = k_m z,$$

and the rest of the lines in $\mathcal{A}^*$ are

$$y = a_1 x + b_1 z, \quad \dots, \quad y = a_n x + b_n z.$$

Therefore, $M(\mathcal{A}^*)$ is a fiber bundle over base $X = \mathbb{CP}^1 - \{k_1, \dots, k_m, \infty\}$ and with fibers $F_x = \mathbb{CP}^1 - \{a_1 x + b_1, \dots, a_n x + b_n, \infty\}$, $x \in X$, under the first coordinate projection $\mathbb{C}^2 \to \mathbb{C}$. Moreover, this fiber bundle admits a structure group $\mathrm{Diff}_+(S^2, \{x_1, \dots, x_n, x_{n+1}\})$.

Definition 2.7. Let $\mathcal{A}^*$ be a fiber-type projective line arrangement in $\mathbb{CP}^2$. Let $c = (0:1:0)$ be the center of $\mathcal{A}^*$. Denote by $\mathrm{St}(c)$ the set of lines in $\mathcal{A}^*$ passing through c. Define the subarrangement associated to $\mathcal{A}^*$ as $\mathcal{B} = \mathcal{A}^* - \mathrm{St}(c)$.

Note that $\mathcal{B}$ can be viewed as an affine arrangement in $\mathbb{C}^2 = \mathbb{CP}^2 - L_\infty$. We will construct the braid monodromy of $\mathcal{B}$ related to $\mathcal{A}^*$ in Section 4.

3. Classification of the complements of fiber-type line arrangements in $\mathbb{CP}^2$ as fiber bundles

For any differentiable fiber bundle with fiber F, let the group $\mathrm{Diff}_+(F)$ be its structure group, the group generated by all orientation preserving diffeomorphisms of F equipped with topology. It is well-known that $\mathrm{Diff}_+(F)$ is also a manifold. Two differentiable fiber bundles $p_1 : E_1 \to B$ and $p_2 : E_2 \to B$ are isomorphic if there exists an diffeomorphism $h : E_1 \to E_2$ such that the following diagram commutes:

$$\begin{tikzcd} E_1 \arrow[rr, "h"] \arrow[dr, "p_1"'] & & E_2 \arrow[dl, "p_2"] \\ & B & \end{tikzcd}$$

The following natural bijection is a well-known fact:

$$\{\text{isomorphism class of differentiable fiber bundles over } X\} \cong [X, B\mathrm{Diff}_+(F)],$$

where $[X, B\mathrm{Diff}_+(F)]$ is the set of homotopy classes of differentiable maps from X to the classifying space $B\mathrm{Diff}_+(F)$.

Note that the homotopy classes of continuous maps and that of differential maps are canonically the same (see Corollary 3.8.18 in [Conlon 2001]). So the classification of differentiable fiber bundles over X with structure group $\mathrm{Diff}_+(F)$ lies in the set of homotopy classes of continuous maps $X \to B\mathrm{Diff}_+(F)$.

It is well known from obstruction theory (see for example Theorem 11 on page 428 in [Spanier 1981]) that if $B\mathrm{Diff}_+(F)$ is a $K(\pi, 1)$ space, then

$$[X, B\mathrm{Diff}_+(F)] \cong \hom_{\mathrm{conj}}(\pi_1(X), \pi_1(B\mathrm{Diff}_+(F))$$

where $\hom_{\mathrm{conj}}$ means the conjugacy classes of homomorphisms. Two homomorphisms f and g are in the same conjugacy class if and only if there is an inner automorphism a of the target group such that $f = a \circ g \circ a^{-1}$. In the following, we will show that the classifying space of $B\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\})$ is a $K(\pi, 1)$ space and the fundamental group is nothing but the mapping class group of an $(n+1)$-punctured sphere, which is the group $\pi_0(\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\}))$ of path components of $\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\})$; see, for example, Chapter 4 in [Birman 1974].

Proposition 3.1. $B\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\})$ *is a* $K(\pi, 1)$ *space. Moreover*,

$$\pi_1(B\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\})) = \pi_0(\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\}))$$

is the mapping class group of an $(n+1)$*-punctured sphere.*

Proof. Let $\mathrm{Diff}_+(S^2, x_1, \dots, x_{n+1})$ be the subgroup of $\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\})$ consisting of diffeomorphisms fixing the base points x_i, $i = 1, \dots, n+1$. Then $\mathrm{Diff}_+(S^2, x_1, \dots, x_{n+1})$ is a normal subgroup in

$$\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\})$$

with the symmetric group $\mathfrak{S}_{n+1}$ as its quotient. On the classifying space level, it follows the fibration

$$B\mathrm{Diff}_+(S^2, x_1, \dots, x_{n+1}) \to B\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\}) \to B\mathfrak{S}_{n+1};$$

see [Piccinini and Spreafico 1998, Theorem 6.1]. Since $\mathfrak{S}_{n+1}$ is a discrete group, $\pi_i(\mathfrak{S}_{n+1}) = 0$ for $i \geq 1$. Then

$$\pi_i(B\mathfrak{S}_{n+1}) \cong \pi_{i-1}(\mathfrak{S}_{n+1}) = 0$$

for $i \geq 2$, which implies that $B\mathfrak{S}_{n+1}$ is a $K(\mathfrak{S}_{n+1}, 1)$-space. The advantage of working with $\mathrm{Diff}_+(S^2, x_1, x_2, \dots, x_{n+1})$ is that we can take x_{n+1} to be the point at ∞ and identify

$$\mathrm{Diff}_+(S^2, x_1, \dots, x_{n+1}) \cong \mathrm{Diff}_+(S^2 - \{\infty\}, x_1, \dots, x_n)$$

with the group $\mathrm{Diff}_+(\mathbb{R}^2, x_1, \dots, x_n)$ of diffeomorphisms of $\mathbb{R}^2$ that keep the n points $x_1, \dots, x_n$ fixed. The later is a better known group. Following from the well-known criterion for classifying spaces [Steenrod 1999, Theorem 19.4; Cohen 1998, Proposition 2.15], we have another fibration

$$\mathrm{Diff}_+(\mathbb{R}^2)/\,\mathrm{Diff}_+(\mathbb{R}^2, x_1, \dots, x_n) \to B\mathrm{Diff}_+(\mathbb{R}^2, x_1, \dots, x_n) \xrightarrow{f} B\mathrm{Diff}_+(\mathbb{R}^2),$$

where f is defined by forgetting the n points. Consider the configuration space $F_n(\mathbb{R}^2)$ of n points in $\mathbb{R}^2$:

$$F_n(\mathbb{R}^2) = \{\, (x_1, \dots, x_n) \mid x_i \in \mathbb{R}^2 \text{ for } i = 1, 2, \dots, n \text{ and } x_i \neq x_j \text{ if } i \neq j \,\}.$$

It is easy to see that the fiber $\mathrm{Diff}_+(\mathbb{R}^2)/\,\mathrm{Diff}_+(\mathbb{R}^2, x_1, \dots, x_n)$ equals $F_n(\mathbb{R}^2)$, which can be considered as the quotient of the flowing homomorphism

$$\begin{aligned} \mathrm{Diff}_+(\mathbb{R}^2) &\to F_n(\mathbb{R}^2) \\ h &\mapsto (h(x_1), \dots, h(x_n)). \end{aligned}$$

It is well known that the configuration space $F_n(\mathbb{R}^2)$ is a $K(\pi, 1)$-space and its fundamental group is a braid group. On the other hand, by Theorem 1 in [Friberg 1973], $\mathrm{Diff}_+(\mathbb{R}^2)$ has the same homotopy type as $\mathrm{SO}(2)$, which is homeomorphic to the circle S^1. So $\pi_1(\mathrm{Diff}_+(\mathbb{R}^2)) \cong \pi_1(\mathrm{SO}(2)) = \mathbb{Z}$ and $\pi_i(\mathrm{Diff}_+(\mathbb{R}^2)) \cong \pi_i(\mathrm{SO}(2)) = 0$ for $i \geq 2$. Hence to prove that

$$\pi_i(B\mathrm{Diff}_+(\mathbb{R}^2, x_1, \dots, x_n)) = 0$$

for $i \geq 2$, by using the long exact sequence of the fibration

$$\begin{array}{ccccc} \pi_i(F_n(\mathbb{R}^2)) & \longrightarrow & \pi_i(B\mathrm{Diff}_+(\mathbb{R}^2, x_1, \dots, x_n)) & \longrightarrow & \pi_i(B\mathrm{Diff}_+(\mathbb{R}^2)) \\ & & & & \wr\| \\ & & & & \pi_{i-1}(\mathrm{Diff}_+(\mathbb{R}^2)) \\ & & & & \wr\| \\ & & & & \pi_{i-1}(\mathrm{SO}(2)), \end{array}$$

it is enough to prove that the boundary map ∂ in the diagram

$$\begin{tikzcd} \pi_2(B\mathrm{Diff}_+(\mathbb{R}^2)) \arrow[r, "\partial"] \arrow[d, "\cong"'] & \pi_1(F_n(\mathbb{R}^2)) \\ \pi_1(\mathrm{Diff}_+(\mathbb{R}^2)) \arrow[ur, "\varphi"'] & \end{tikzcd}$$

is injective. The map $\pi_1(\mathrm{Diff}_+(\mathbb{R}^2)) \xrightarrow{\varphi} \pi_1(F_n(\mathbb{R}^2))$ can be identified with the induced homomorphism given by

$$\begin{aligned} \mathrm{Diff}_+(\mathbb{R}^2) &\to F_n(\mathbb{R}^2) \\ h &\mapsto (h(x_1), \dots, h(x_n)). \end{aligned}$$

From this interpretation, it is easy to see that a generator of $\pi_1(\mathrm{Diff}_+(\mathbb{R}^2))$ is mapped to a nontrivial element in $\pi_1(F_{n+1}(\mathbb{R}^2))$. Thus ∂ is injective and hence $B\mathrm{Diff}_+(S^2, x_1, \dots, x_{n+1})$ is a $K(\pi, 1)$-space. So $B\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\})$ is also a $K(\pi, 1)$-space and

$$\pi_1(B\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\})) = \pi_0(\mathrm{Diff}_+(S^2, \{x_1, \dots, x_{n+1}\}))$$

is the mapping class group of an $(n+1)$-punctured sphere. □

It follows immediately that:

Theorem 3.2. *Let $B = S^2 \setminus \{k_1, k_2, \dots, k_{m+1}\}$ and $F = S^2 \setminus \{x_1, x_2, \dots, x_{n+1}\}$. The isomorphic classes of differentiable fiber bundles over B with fiber F and structure group $G = \mathrm{Diff}_+(F)$ are in one-to-one correspondence with the conjugacy classes of homomorphisms from $\pi_1(B)$ to $\pi_1(BG) = \mathcal{M}^n$, where $\mathcal{M}^n$ is the mapping class group of an n-punctured sphere.*

4. Application of braid monodromy

Before we prove our Main Theorem, we will give the definition of center-projecting braid monodromy of a fiber-type projective line arrangement and some useful results [Cohen and Suciu 1997; Dung 1999; Artal Bartolo et al. 2003].

Let $f_i(x) = a_i x + b_i$, $1 \le i \le n$, be the linear functions of the lines not passing through the center c of a fiber-type projective line arrangement. Define

$$f : \mathbb{C} \setminus \{k_1, \dots, k_m\} \to F_n(\mathbb{C})$$

to be the map $f(x) = (f_1(x), f_2(x), \dots, f_n(x))$.

Definition/Construction 4.1. Let $\mathcal{A}^*$ be a fiber-type line arrangement in $\mathbb{CP}^2$ with center c and let $\mathcal{B}$ be the subarrangement associated to $\mathcal{A}^*$. Choose the projection from the complement $M(\mathcal{B})$ in $\mathbb{C}^2$ to $\mathbb{C}$ so that it coincides with the projection from $M(\mathcal{A}^*)$ to a $\mathbb{CP}^1$ through the center c. Let $\infty, k_1, k_2, \dots, k_m$ be the points in $\mathbb{CP}^1$ that are the projective images of the lines in $\mathcal{A}^*$ passing through c. The braid monodromy of $\mathcal{B}$ is the homomorphism $\varphi : \pi_1(\mathbb{C} \setminus \{k_1, \dots, k_m\}) \to B_n$ induced by the map f, where B_n is the braid group of n strings (see [Birman 1974]) and n is the number of the lines in $\mathcal{A}^*$ not passing through the center c. Such a braid monodromy is called the *center-projecting braid monodromy* of the fiber-type line arrangement $\mathcal{A}^*$ in $\mathbb{CP}^2$.

One can easily check that the braid monodromy of $\mathcal{B}$ coincides with the monodromy of the fiber bundle $M(\mathcal{A}^*)$.

This fact about the bundle structure of $M(\mathcal{A}^*)$ is a theorem of Cohen [2001]:

Theorem 4.2. *The complement of $\mathcal{A}^*$ with the natural bundle structure is equivalent to the pullback of the bundle of configuration spaces $p_{n+1} : F_{n+1}(\mathbb{C}) \to F_n(\mathbb{C})$ via f.*

The next corollary follows immediately from Theorem 4.2 and Proposition 3.1.

Corollary 4.3. *Let $g : \pi_1(B) \to \mathcal{M}^n$ be a classifying morphism representing the isomorphism class of the bundle $M(\mathcal{A}^*) \to B$ and $q : B_n \to \mathcal{M}^n$ be the classifying morphism representing the isomorphism class the fiber bundle $F_{n+1}(\mathbb{C}) \to F_n(\mathbb{C})$. Then g factors through q via the center-projecting braid monodromy φ.*

Proof. Let $G = \mathrm{Diff}_+(S^2, \{x_1, \dots, x_n, x_{n+1}\})$ be the structure group of the bundle $M(\mathcal{A}^*) \to B$. Let $g' : B \to BG$ be a differentiable map which induces the map g and $q' : F_n(\mathbb{C}) \to BG$ be a differentiable map which induces the map q. Then we have the following bundle isomorphisms: $g'^*EG \cong M(\mathcal{A}^*) \cong f^*(F_{n+1}(\mathbb{C})) \cong f^*(q'^*(EG)) = (q' \circ f)^*(EG)$, where BG is the classifying space of G and EG is the universal fiber bundle over BG. Then $q' \circ f$ and g' are representing the same bundle. Therefore $g = q \circ \varphi$, because the braid monodromy φ is induced by the map f. □

Denote by $\mathbf{F}_m$ the free group generated by m elements.

Definition 4.4. Let $\psi_1, \psi_2 : \pi_1(\mathbf{C} \setminus \{k_1, \dots, k_m\}) = \mathbf{F}_m \to B_n$ be the center-projecting braid monodromies of $\mathcal{A}_1^*$ and $\mathcal{A}_2^*$ respectively. We say that $\mathcal{A}_1^*$ and

$\mathscr{A}_2^*$ have the same braid monodromy if there exists an element $\rho \in B_n$ such that $\psi_2(\alpha) = \rho \cdot \psi_1(\alpha) \cdot \rho^{-1}$ for any $\alpha \in \mathbf{F}_m$.

Main Theorem. *Let $\mathscr{A}_1^*$ and $\mathscr{A}_2^*$ be two fiber-type projective line arrangements. If they have the same center-projecting braid monodromies, then their complements $M(\mathscr{A}_1^*)$ and $M(\mathscr{A}_2^*)$ are diffeomorphic.*

Proof. By Remark 2.6, the complements of the two fiber-type line arrangements are fiber bundles. Since they have the same center-projecting braid monodromy, they have the same base, fiber and structure group. By Theorem 3.2, we know that the isomorphism classes of such fiber bundles over same base with same fiber and structure group are in one-to-one correspondence with the homomorphisms $\pi_1(S^2 \setminus \{x_1, \dots, x_{m+1}\}) \to \mathcal{M}^n$ up to conjugation. By Corollary 4.3, the isomorphism class of the complement of a fiber-type projective line arrangement as a fiber bundle is determined by the braid monodromy. Let the homomorphism $q : B_n \to \mathcal{M}^n$ be a representative of the isomorphism class of the bundle of configurations $F_{n+1}(\mathbb{C}) \to F_n(\mathbb{C})$. If $\psi_1, \psi_2 : \pi_1(\mathbf{C} \setminus \{k_1, \dots, k_m\}) = \mathbf{F}_m \to B_n$ are the same center-projecting braid monodromies associated to $\mathscr{A}_1^*$ and $\mathscr{A}_2^*$ respectively, then there exists a $\rho \in B_n$ such that $\psi_2(\alpha) = \rho \cdot \psi_1(\alpha) \cdot \rho^{-1}$ for any $\alpha \in \mathbf{F}_m$. Thus $q \circ \psi_2(\alpha) = q(\rho) \cdot (q \circ \psi_1(\alpha)) \cdot (q(\rho))^{-1}$ for any $\alpha \in \mathbf{F}_m$. This implies that $q \circ \psi_1$ and $q \circ \psi_2$ determine the same isomorphism class. By the definition of isomorphism of differentiable fiber bundles, any two members in the isomorphism class have diffeomorphic total spaces. This proves the theorem. □

Combined with a theorem of Jiang and Yau [1993], our Main Theorem implies that the center-projecting braid monodromy of a fiber-type projective line arrangement determines its lattice. In fact:

Theorem 4.5 [Cohen and Suciu 1997]. *The braid monodromy of a line arrangement determines its lattice.*

The braid monodromies they considered are generic braid monodromies, that is, projecting from a generic point such that each fiber of the projection contains at most one singularity. However, their method seems also work for nongeneric cases. In fact, when there is more than one singularity in a fiber, the images of the local braid monodromies still record the twists of the braids which reflect the intersecting of lines.

Example 4.6. The complements of any two line arrangements $\mathscr{A}_1^*$ and $\mathscr{A}_2^*$ of six lines with four triple points and three nodes are diffeomorphic. Clearly, any triple point can be viewed as a center for such an arrangement. Assume that the line at infinity passes through the center. After removing the center, the subarrangement in $\mathbb{C}^2$ contains three lines, the three solid lines in Figure 1, and the braid monodromy is uniquely determined. In fact, the center-projecting braid monodromies of $\mathscr{A}_1^*$

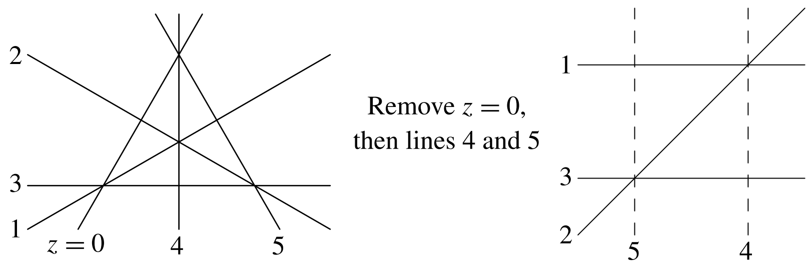


Figure 1. Arrangement of six lines with four triple points and three nodes and its associated subarrangement.

and $\mathscr{A}_2^*$ coincide with the generic braid monodromy of arrangement of 3 lines. Let ξ_1 and ξ_2 be two circles centered at x_1 and x_2, in the base $B = \mathbb{C} \setminus \{x_1, x_2\}$, where x_1 and x_2 are the projections of lines 4 and 5 respectively. Assume that ξ_1 and ξ_2 have a tangent point between x_1 and x_2. Then the fundamental group of the base B is $\pi_1(B) = \langle \xi_1, \xi_2 \rangle$. It is easy to see that the braid monodromy of arrangement of 3 lines as shown in Figure 1 is uniquely determined up to conjugacy by the images of ξ_1 and ξ_2 which are the monodromy generators σ_1^2 (the image of ξ_1) and σ_2^2 (the image of ξ_2), where σ_1 and σ_2 are the two generators of the braid group $\mathbf{B}_3$ on 3 strings as shown in Figure 2 (see, for example, [Cohen and Suciu 1997] on how to calculate braid monodromy generators in general). Hence by our theorem, the complements $M(\mathscr{A}_1^*)$ and $M(\mathscr{A}_2^*)$ are diffeomorphic.

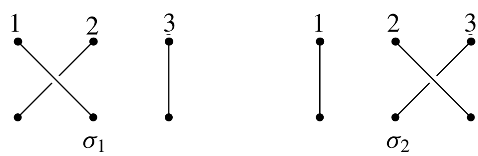


Figure 2. Braid generators of $\mathbf{B}_3$.

Remark 4.7. The arrangement in the example above is well studied in many aspects. For example, it has been shown in a recent paper [Nazir and Yoshinaga 2010] that the moduli space of line arrangements of six lines with four triple points and three nodes is irreducible, so is connected. In fact, it is easy to see that line arrangements of six lines with four triple points and three nodes are of simple C_3 type in the sense of Nazir and Yoshinaga.

Acknowledgement

We gratefully acknowledge professor Ronnie Lee for many useful discussions on Proposition 3.1.

References

[Aravinda et al. 2000] C. S. Aravinda, F. T. Farrell, and S. K. Roushon, "Algebraic K-theory of pure braid groups", *Asian J. Math.* **4**:2 (2000), 337–343. MR 2002a:19002 Zbl 0980.19001

[Artal Bartolo et al. 2003] E. Artal Bartolo, J. Carmona Ruber, and J. I. Cogolludo Agustín, "Braid monodromy and topology of plane curves", *Duke Math. J.* **118** (2003), 261–278. MR 2004k:14015 Zbl 1058.14053

[Artal Bartolo et al. 2007] E. Artal Bartolo, J. Carmona Ruber, and J. I. Cogolludo Agustín, "Effective invariants of braid monodromy", *Trans. Amer. Math. Soc.* **359**:1 (2007), 165–183. MR 2007f:14005 Zbl 1109.14013

[Birman 1974] J. S. Birman, *Braids, links, and mapping class groups*, Annals of Math. Studies **82**, Princeton University Press, Princeton, N.J., 1974. MR 51 #11477 Zbl 0305.57013

[Cohen 1998] R. L. Cohen, "The topology of fiber bundles", Lecture notes, Stanford University, 1998.

[Cohen 2001] D. C. Cohen, "Monodromy of fiber-type arrangements and orbit configuration spaces", *Forum Math.* **13**:4 (2001), 505–530. MR 2002i:52016 Zbl 1018.20032

[Cohen and Suciu 1997] D. C. Cohen and A. I. Suciu, "The braid monodromy of plane algebraic curves and hyperplane arrangements", *Comment. Math. Helv.* **72**:2 (1997), 285–315. MR 98f:52012 Zbl 0959.52018

[Conlon 2001] L. Conlon, *Differentiable manifolds*, 2nd ed., Birkhäuser Advanced Texts: Basler Lehrbücher, Birkhäuser, Boston, 2001. MR 2002b:58001 Zbl 0994.57001

[Dung 1999] N. V. Dung, "Braid monodromy of complex line arrangements", *Kodai Math. J.* **22**:1 (1999), 46–55. MR 2000e:14039 Zbl 0954.14014

[Earle and Eells 1967] C. J. Earle and J. Eells, "The diffeomorphism group of a compact Riemann surface", *Bull. Amer. Math. Soc.* **73** (1967), 557–559. MR 35 #3705 Zbl 0196.09402

[Friberg 1973] B. Friberg, "A topological proof of a theorem of Kneser", *Proc. Amer. Math. Soc.* **39** (1973), 421–426. MR 47 #9657 Zbl 0273.57017

[Jiang and Yau 1993] T. Jiang and S. S.-T. Yau, "Topological invariance of intersection lattices of arrangements in $\mathbf{C}P^2$", *Bull. Amer. Math. Soc.* (*N.S.*) **29**:1 (1993), 88–93. MR 94b:52022 Zbl 0847.52011

[Jiang and Yau 1994] T. Jiang and S. S.-T. Yau, "Diffeomorphic types of the complements of arrangements of hyperplanes", *Compositio Math.* **92**:2 (1994), 133–155. MR 95e:32042 Zbl 0828.57018

[Jiang et al. 2001] T. Jiang, S. S.-T. Yau, and L.-Y. Yeh, "Simple geometric characterization of supersolvable arrangements", *Rocky Mountain J. Math.* **31**:1 (2001), 303–312. MR 2001m:55047 Zbl 1008.32015

[Kulikov and Taĭkher 2000] V. S. Kulikov and M. Taĭkher, "Braid monodromy factorizations and diffeomorphism types", *Izv. Ross. Akad. Nauk Ser. Mat.* **64**:2 (2000), 89–120. MR 2001f:14030 Zbl 1004.14005

[Libgober 1986] A. Libgober, "On the homotopy type of the complement to plane algebraic curves", *J. Reine Angew. Math.* **367** (1986), 103–114. MR 87j:14044 Zbl 0576.14019

[Moishezon 1981] B. G. Moishezon, "Stable branch curves and braid monodromies", pp. 107–192 in *Algebraic geometry* ((Chicago, Ill., 1980)), edited by A. Libgober and P. Wagreich, Lecture Notes in Math. **862**, Springer, Berlin, 1981. MR 83c:14008 Zbl 0476.14005

[Morita 1987] S. Morita, "Characteristic classes of surface bundles", *Invent. Math.* **90**:3 (1987), 551–577. MR 89e:57022 Zbl 0608.57020

[Nazir and Yoshinaga 2010] S. Nazir and M. Yoshinaga, "On the connectivity of the realization spaces of line arrangements", preprint, 2010. arXiv 1009.0202

[Orlik 1992] P. Orlik, "Complements of subspace arrangements", *J. Algebraic Geom.* **1**:1 (1992), 147–156. MR 92h:52014 Zbl 0795.52003

[Orlik and Terao 1992] P. Orlik and H. Terao, *Arrangements of hyperplanes*, Grund. der Math. Wissenschaften **300**, Springer, Berlin, 1992. MR 94e:52014 Zbl 0757.55001

[Piccinini and Spreafico 1998] R. A. Piccinini and M. Spreafico, "The Milgram–Steenrod construction of classifying spaces for topological groups", *Exposition. Math.* **16**:2 (1998), 97–130. MR 99f:55022 Zbl 0941.55016

[Spanier 1981] E. H. Spanier, *Algebraic topology*, Springer, New York, 1981. Corrected reprint. MR 83i:55001 Zbl 0477.55001

[Steenrod 1999] N. Steenrod, *The topology of fibre bundles*, Princeton Landmarks in Math., Princeton University Press, Princeton, NJ, 1999. Reprint of the 1957 edition, Princeton Paperbacks. MR 2000a:55001 Zbl 0942.55002

[Terao 1986] H. Terao, "Modular elements of lattices and topological fibration", *Adv. in Math.* **62**:2 (1986), 135–154. MR 88b:32032 Zbl 0612.05019

[Wang and Yau 2005] S. Wang and S. S.-T. Yau, "Rigidity of differentiable structure for new class of line arrangements", *Comm. Anal. Geom.* **13**:5 (2005), 1057–1075. MR 2007d:32021 Zbl 1115.52010

[Wang and Yau 2007] S. Wang and S. S.-T. Yau, "The diffeomorphic types of the complements of arrangements in $\mathbf{C}P^3$, I, Point arrangements", *J. Math. Soc. Japan* **59**:2 (2007), 423–447. MR 2008i:32041 Zbl 1140.14032

[Wang and Yau 2008] S. Wang and S. S.-T. Yau, "The diffeomorphic types of the complements of arrangements in $\mathbb{CP}^3$, II", *Sci. China Ser. A* **51**:4 (2008), 785–802. MR 2009c:32054 Zbl 1192.14042

[Xicoténcatl 1997] M. Xicoténcatl, *Orbit configuration spaces*, Ph.D. thesis, University of Rochester, 1997.

[Yau and Ye 2009] S. S.-T. Yau and F. Ye, "Diffeomorphic types of complements of nice point arrangements in $\mathbb{CP}^1$", *Sci. China Ser. A* **52**:12 (2009), 2774–2791. MR 2011e:32037 Zbl 1197.14015

Received March 1, 2010.

SHENG-LI TAN
EAST CHINA NORMAL UNIVERSITY
DEPARTMENT OF MATHEMATICS
SHANGHAI 200062
CHINA
sltan@math.ecnu.edu.cn
http://math.ecnu.edu.cn/~sltan/intro.html

STEPHEN S.-T. YAU
UNIVERSITY OF ILLINOIS AT CHICAGO
DEPARTMENT OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE
322 SEO (M/C 249)
851 SOUTH MORGAN STREET
CHICAGO IL 60607
UNITED STATES
yau@uic.edu
http://www.math.uic.edu/~yau/

FEI YE
UNIVERSITY OF ILLINOIS AT CHICAGO
DEPARTMENT OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE
322 SEO (M/C 249)
851 SOUTH MORGAN STREET
CHICAGO 60607
UNITED STATES
fye@math.uic.edu
http://www.math.uic.edu/~fye

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 1, 2011

INEQUALITIES FOR THE NAVIER AND DIRICHLET EIGENVALUES OF ELLIPTIC OPERATORS

QIAOLING WANG AND CHANGYU XIA

This paper studies eigenvalues of elliptic operators on a bounded domain in a Euclidean space. We obtain lower bounds for the eigenvalues of elliptic operators of higher orders with Navier boundary condition. We also prove lower bounds and universal inequalities of Payne–Pólya–Weinberger–Yang type for the eigenvalues of second order elliptic equations in divergence form with Dirichlet boundary condition.

1. Introduction

Let Ω be a bounded domain in an n-dimensional Euclidean space $\mathbb{R}^n$ $(n \geq 2)$ with smooth boundary $\partial\Omega$. Let Δ be the Laplacian of $\mathbb{R}^n$ and consider the fixed membrane or Dirichlet eigenvalue problem

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

Let

$$0 < \lambda_1 < \lambda_2 \leq \cdots \to \infty$$

denote the eigenvalues (repeated with multiplicity) of the problem (1-1). Weyl's asymptotic formula [1912] tells us that

$$\lambda_k \sim C(n)\left(\frac{k}{|\Omega|}\right)^{2/n} \quad \text{as } k \to \infty, \tag{1-2}$$

where $|\Omega|$ is the volume of Ω and $C(n) = (2\pi)^2 \omega_n^{-2/n}$ with ω_n being the volume of the unit ball in $\mathbb{R}^n$. Pólya [1961] showed that for any "plane covering domain" Ω in $\mathbb{R}^2$ (those that tile $\mathbb{R}^2$) this asymptotic relation is a one-sided inequality (his proof

Wang was partially supported by CNPq. Xia was partially supported by CNPq and FAPDF.
MSC2000: primary 35P15, 53C20, 53C42; secondary 35P15, 53C42.
Keywords: eigenvalue, lower bound, universal inequality, elliptic operator, Navier boundary condition.

also works for $\mathbb{R}^n$-covering domains) and conjectured, for any domain $\Omega \subset \mathbb{R}^n$, the inequality

$$\lambda_k \geq C(n)\left(\frac{k}{|\Omega|}\right)^{2/n} \quad \text{for all } k. \tag{1-3}$$

Li and Yau [1983] showed the lower bound

$$\sum_{i=1}^{k} \lambda_i \geq \frac{nkC(n)}{n+2}\left(\frac{k}{|\Omega|}\right)^{2/n}, \tag{1-4}$$

which yields an individual lower bound on λ_k in the form

$$\lambda_k \geq \frac{nC(n)}{n+2}\left(\frac{k}{|\Omega|}\right)^{2/n}. \tag{1-5}$$

Similar bounds for eigenvalues with Neumann boundary condition were proved in [Kröger 1992; 1994; Laptev 1997]. It was pointed out in [Laptev and Weidl 2000] that (1-4) also follows from an earlier result of Berezin [1972] by using the Legendre transformation. Melas [2003] gave an improvement of (1-4):

$$\sum_{i=1}^{k} \lambda_i \geq \frac{nkC(n)}{n+2}\left(\frac{k}{|\Omega|}\right)^{2/n} + d_n k \frac{|\Omega|}{I(\Omega)}, \tag{1-6}$$

where the constant d_n depends only on the dimension and

$$I(\Omega) = \min_{a \in \mathbb{R}^n} \int_\Omega |x-a|^2\, dx$$

is the "moment of inertia" of Ω.

In this paper, we study eigenvalues of elliptic operators of higher orders with Navier boundary condition and of second order elliptic equations in divergence form with Dirichlet boundary condition and prove lower bounds for them which are similar to the inequality (1-6). We will also prove universal inequalities of Yang type for the Dirichlet eigenvalues of second order equations in divergence form. The first two results concern eigenvalues with Navier boundary condition.

Theorem 1.1. *Let Ω be a bounded domain in $\mathbb{R}^n$ and let l be a positive integer. Consider the eigenvalue problem*

$$\begin{cases} (-\Delta)^l u = \lambda u & \text{in } \Omega,\ u \in C^\infty(\Omega), \\ u|_{\partial\Omega} = \Delta u|_{\partial\Omega} = \cdots = \Delta^{l-1} u|_{\partial\Omega} = 0. \end{cases} \tag{1-7}$$

Let

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k \leq \cdots \to \infty$$

be the eigenvalues of (1-7) *and denote by* $\mu_1, \dots, \mu_n$ *the first n nonzero eigenvalues of the Neumann problem*

$$\begin{cases} -\Delta v = \mu v \quad \text{in } \Omega, \\ (\partial v/\partial \nu)|_{\partial\Omega} = 0, \end{cases} \tag{1-8}$$

where ν *is the unit outward normal vector field along* $\partial\Omega$. *Then*

$$\sum_{j=1}^{k} \lambda_j^{1/l} \geq \frac{nkC(n)}{n+2}\left(\frac{k}{|\Omega|}\right)^{2/n} + \frac{d(n)k}{\sum_{i=1}^{n} \mu_i^{-1}}. \tag{1-9}$$

Here $d(n)$ *is a positive constant depending only on* n.

Theorem 1.2. *Let* Ω *be a bounded domain in* $\mathbb{R}^n$ *and let* l *be a fixed positive integer. Let* L *be the elliptic operator given by*

$$Lu = \sum_{m=1}^{l} a_m(-\Delta)^m u, \quad u \in C^\infty(\Omega),$$

where the a_m *are constants with* $a_m \geq 0$, $1 \leq m \leq l$, *and* $a_l = 1$. *Consider the eigenvalue problem*

$$\begin{cases} Lu = \Lambda u \quad \text{in } \Omega, \\ u|_{\partial\Omega} = \Delta u|_{\partial\Omega} = \cdots = \Delta^{l-1} u|_{\partial\Omega} = 0. \end{cases} \tag{1-10}$$

Let

$$0 < \Lambda_1 \leq \Lambda_2 \leq \cdots \leq \Lambda_k \leq \cdots \to \infty$$

be the eigenvalues of (1-10). *Then*

$$\Lambda_k \geq \sum_{m=1}^{l} a_m \left(\frac{nC(n)}{n+2}\left(\frac{k}{|\Omega|}\right)^{2/n} + \frac{d(n)}{\sum_{i=1}^{n} \mu_i^{-1}}\right)^m, \tag{1-11}$$

where $\mu_1, \dots, \mu_n$ *are the first n nonzero Neumann eigenvalues of* Ω *and* $d(n)$ *is a positive constant depending on* n.

Our next results are about second order equations in divergence form with Dirichlet boundary condition. Firstly, we have a Li–Yau type inequality.

Theorem 1.3. *Let* Ω *be a bounded domain in* $\mathbb{R}^n$ *and let* V *be a nonnegative continuous function on* Ω. *Consider the eigenvalue problem*

$$\begin{cases} -\sum_{\alpha,\beta=1}^{n} \frac{\partial}{\partial x_\alpha}\left(a_{\alpha\beta}(x)\frac{\partial u}{\partial x_\beta}\right) + V(x)u = \lambda u \quad \text{in } \Omega, \\ u|_{\partial\Omega} = 0. \end{cases} \tag{1-12}$$

Assume that there is a positive number ξ_0 *such that the symmetric matrix* $[a_{\alpha\beta}]$ *satisfies* $[a_{\alpha\beta}] \geq \xi_0 I$ *in the sense of quadratic forms throughout* Ω, *where* I *is the identity matrix of order* n. *Let*

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k \leq \cdots \to \infty$$

be the eigenvalues of (1-12). *Then*

$$\sum_{j=1}^{k} \lambda_j \geq \xi_0 k \left(\frac{nC(n)}{n+2} \left(\frac{k}{|\Omega|} \right)^{2/n} + \frac{d(n)}{\sum_{i=1}^{n} \mu_i^{-1}} + \frac{V_0}{\xi_0} \right). \tag{1-13}$$

Here $V_0 = \inf_{x\in\Omega} V(x)$, $\mu_1, \ldots, \mu_n$ *and* $d(n)$ *are as in Theorem 1.1.*

We then prove a universal inequality of Payne–Pólya–Weinberger–Yang type [Payne et al. 1956; Yang 1991] for an eigenvalue problem more general than (1-12).

Theorem 1.4. *Let* Ω *be a connected bounded domain in* $\mathbb{R}^n$ *and let* V *be a nonnegative continuous function on* Ω *with* $V_0 = \inf_{x\in\Omega} V(x)$. *Let* ρ *be a continuous function on* Ω *satisfying* $\rho_1 \leq \rho(x) \leq \rho_2$ *for all* $x \in \Omega$, *for some positive constants* ρ_1 *and* ρ_2. *Assume also that there are positive numbers* ξ_1 *and* ξ_2 *such that the symmetric matrix* $[a_{\alpha\beta}]$ *satisfies* $[a_{\alpha\beta}] \geq \xi_1 I$ *in the sense of quadratic forms and* $\sum_{\alpha=1}^{n} a_{\alpha\alpha} \leq n\xi_2$ *throughout* Ω. *Consider the eigenvalue problem*

$$\begin{cases} -\sum_{\alpha,\beta=1}^{n} \frac{\partial}{\partial x_\alpha}\left(a_{\alpha\beta}(x) \frac{\partial u}{\partial x_\beta} \right) + V(x)u = \lambda\rho u & \text{in } \Omega, \\ u|_{\partial\Omega} = 0. \end{cases} \tag{1-14}$$

Let

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k \leq \cdots \to \infty$$

be the eigenvalues of (1-14). *Then*

$$\sum_{i=1}^{k} (\lambda_{k+1} - \lambda_i)^2 \leq \frac{4\xi_2\rho_2^2}{n\xi_1\rho_1^2} \sum_{i=1}^{k} (\lambda_{k+1} - \lambda_i) \left(\lambda_i - \frac{V_0}{\rho_2} \right). \tag{1-15}$$

Remark 1.5. Universal inequalities of Payne–Pólya–Weinberger-Yang type for eigenvalues of elliptic operators on Riemannian manifolds have been studied recently by many mathematicians. One can find various interesting results in this direction, for example, in [Ashbaugh 1999; 2002; Ashbaugh and Benguria 1993a; 1993b; Ashbaugh and Hermi 2004; Cheng and Yang 2005; 2006a; 2006b; 2006c; 2007; El Soufi et al. 2007; Harrell 1993; Harrell and Michel 1994; Harrell and Stubbe 1997; Harrell and Yıldırım Yolcu 2009; Hile and Protter 1980; Hook 1990; Laptev 1997; Levitin and Parnovski 2002; Sun et al. 2008; Wang and Xia 2007a; 2007b; 2008; 2010a; 2010b; 2010c; 2011].

2. An auxiliary result

Before proving 1.1–1.3, we show the following fact.

Theorem 2.1. *Let Ω be a bounded domain in $\mathbb{R}^n$ and let $w_1, \dots, w_k : \Omega \to \mathbb{R}$ be smooth functions satisfying*

$$w_i|_{\partial\Omega} = 0 \quad \text{and} \quad \int_\Omega w_i(x) w_j(x)\, dx = \delta_{ij} \quad \text{for } i, j = 1, \dots, k. \tag{2-1}$$

Then

$$\sum_{j=1}^{k} \int_\Omega |\nabla w_j(x)|^2\, dx \geq \frac{nkC(n)}{n+2}\left(\frac{k}{|\Omega|}\right)^{2/n} + \frac{d(n)k}{\sum_{i=1}^{n} \mu_i^{-1}}, \tag{2-2}$$

where $d(n)$ is a computational positive constant depending only on n and the μ_i are the first n nonzero Neumann eigenvalues of the Laplacian of Ω.

Proof. Let $v_1, \dots, v_n$ be orthonormal eigenfunctions corresponding to $\mu_1, \dots, \mu_n$:

$$-\Delta v_i = \mu_i v_i \quad \text{in } \Omega, \qquad \frac{\partial v_i}{\partial \nu}\Big|_{\partial\Omega} = 0, \qquad \int_\Omega v_i v_j = \delta_{ij} \quad \text{for } i, j = 1, \dots, n.$$

By a translation of the origin and a suitable rotation of axes, we can assume, using [Ashbaugh and Benguria 1993b, p. 563], that

$$\int_\Omega x_i\, dx = 0 \quad \text{for } i = 1, \dots, n,$$
$$\int_\Omega x_j v_i\, dx = 0 \quad \text{for } j = 2, \dots, n,\ i = 1, \dots, j-1.$$

It then follows from inequality (2.8) in the same paper that

$$\sum_{i=1}^{n} \frac{1}{\mu_i} \geq \frac{\int_\Omega |x|^2 dx}{|\Omega|}. \tag{2-3}$$

By a simple rearrangement argument, we have

$$\int_\Omega |x|^2\, dx \geq \frac{n}{n+2}|\Omega|\left(\frac{|\Omega|}{\omega_n}\right)^{2/n}. \tag{2-4}$$

Extend each w_i to $\mathbb{R}^n$ by letting $w_i(x) = 0$ for $x \in \mathbb{R}^n \setminus \Omega$. For a function g on $\mathbb{R}^n$, we will denote by $\mathcal{F}(g)$ the Fourier transformation of g. For any $z \in \mathbb{R}^n$, we have, by definition,

$$\mathcal{F}(w_j)(z) = (2\pi)^{-n/2} \int_\Omega e^{-i\langle x, z\rangle} w_j(x)\, dx. \tag{2-5}$$

Since $\{w_j\}_{j=1}^k$ is an orthonormal set in $L^2(\Omega)$, the Bessel inequality gives

$$\sum_{j=1}^k |\mathcal{F}(w_j)(z)|^2 \le (2\pi)^{-n} \int_\Omega |e^{-i\langle x,z\rangle}|^2\,dx = (2\pi)^{-n}|\Omega|. \tag{2-6}$$

For each $q = 1, \ldots, n$ and $j = 1, \ldots, k$, since w_j vanishes on $\partial\Omega$, one gets from the divergence theorem that

$$\begin{aligned} z_q\mathcal{F}(w_j)(z) &= i(2\pi)^{-n/2} \int_\Omega \frac{\partial e^{-i\langle x,z\rangle}}{\partial x_q} w_j(x)\,dx \\ &= -i(2\pi)^{-n/2} \int_\Omega \frac{\partial w_j(x)}{\partial x_q} e^{-i\langle x,z\rangle}\,dx = -i\mathcal{F}\Big(\frac{\partial w_j}{\partial z_q}\Big)(z). \end{aligned} \tag{2-7}$$

It then follows from the Plancherel formula that

$$\begin{aligned} \int_{\mathbb{R}^n} |z|^2|\mathcal{F}(w_j)(z)|^2\,dz &= \int_{\mathbb{R}^n} \sum_{q=1}^n \Big|\mathcal{F}\Big(\frac{\partial w_j}{\partial z_q}\Big)(z)\Big|^2\,dz \\ &= \int_\Omega \sum_{q=1}^n \Big(\frac{\partial w_j}{\partial x_q}\Big)^2\,dx = \int_\Omega |\nabla w_j(x)|^2\,dx. \end{aligned} \tag{2-8}$$

Since

$$\nabla(\mathcal{F}(w_j))(z) = (2\pi)^{-n/2} \int_\Omega (-ixe^{-i\langle x,z\rangle} w_j(x))\,dx,$$

we have

$$\sum_{j=1}^k |\nabla(\mathcal{F}(w_j))(z)|^2 \le (2\pi)^{-n} \int_\Omega |ixe^{-i\langle x,z\rangle}|^2\,dx = (2\pi)^{-n} \int_\Omega |x|^2\,dx. \tag{2-9}$$

Set

$$G(z) = \sum_{j=1}^k |\mathcal{F}(w_j)(z)|^2.$$

Then $0 \le G(z) \le (2\pi)^{-n}|\Omega|$ and

$$\begin{aligned} |\nabla G(z)| &\le 2\Big(\sum_{j=1}^k |\mathcal{F}(w_j)(z)|^2\Big)^{1/2} \Big(\sum_{j=1}^k |\nabla(\mathcal{F}(w_j))(z)|^2\Big)^{1/2} \\ &\le 2(2\pi)^{-n}\Big(|\Omega| \int_\Omega |x|^2\,dx\Big)^{1/2} \end{aligned} \tag{2-10}$$

for every $z \in \mathbb{R}^n$. We also have

$$\int_{\mathbb{R}^n} G(z)\,dz = \sum_{j=1}^k \int_\Omega w_j(x)^2\,dx = k. \tag{2-11}$$

Let $G^*(z) = g(|z|)$ be the decreasing spherical rearrangement of G. By approximation, we may assume that $g : [0, +\infty) \to [0, (2\pi)^{-n}|\Omega|]$ is absolutely continuous. Setting $\alpha(t) = |\{G^* > t\}| = |\{G > t\}|$, we have

$$\alpha(g(s)) = \omega_n s^n, \tag{2-12}$$

which implies that $n\omega_n s^{n-1} = \alpha'(g(s))g'(s)$ for almost every s. The coarea formula [Chavel 1984] tells us that

$$\alpha'(t) = -\int_{\{G=t\}} \frac{1}{|\nabla G|}\, dA_t. \tag{2-13}$$

Set $\eta = 2(2\pi)^{-n}\big(|\Omega| \int_\Omega |x|^2\, dx\big)^{1/2}$; then one infers from (2-10) and the isoperimetric inequality that

$$-\alpha'(g(s)) \geq \eta^{-1}\mathrm{area}(\{G = g(s)\}) \geq \eta^{-1} n\omega_n s^{n-1}, \tag{2-14}$$

and so

$$-\eta \leq g'(s) \leq 0 \tag{2-15}$$

for almost every s. It follows from (2-11) that

$$k = \int_{\mathbb{R}^n} G(z)\, dz = \int_{\mathbb{R}^n} G^*(z)\, dz = n\omega_n \int_0^\infty s^{n-1} g(s)\, ds \tag{2-16}$$

and, by (2-8),

$$\begin{aligned} \sum_{j=1}^k \int_\Omega |\nabla w_j|^2 = \int_{\mathbb{R}^n} |z|^2 G(z)\, dz &\geq \int_{\mathbb{R}^n} |z|^2 G^*(z)\, dz \\ &= n\omega_n \int_0^\infty s^{n+1} g(s)\, ds, \end{aligned} \tag{2-17}$$

since $z \to |z|^2$ is radial and increasing.

We next apply the following lemma to the function g, with $A = \dfrac{k}{n\omega_n}$ and

$$\eta = 2(2\pi)^{-n}\left(|\Omega| \int_\Omega |x|^2\, dx\right)^{1/2}:$$

Lemma [Melas 2003]. *Let $n \geq 1$ and $\eta,\ A > 0$ and let $h : [0, +\infty) \to [0, +\infty)$ be a decreasing and absolutely continuous function such that*

$$-\eta \leq h'(s) \leq 0 \quad \text{and} \quad \int_0^\infty s^{n-1} h(s)\, ds = A.$$

Then

$$\int_0^\infty s^{n+1} h(s)\, ds \geq \frac{(nA)^{(n+2)/n}}{n+2} h(0)^{-2/n} + \frac{A h(0)^2}{6(n+2)\eta^2}. \tag{2-18}$$

After applying the lemma and using (2-17), we infer that

$$\sum_{j=1}^{k}\int_\Omega |\nabla w_j(x)|^2\,dx \ge \frac{n}{n+2}\,\omega_n^{-2/n}k^{1+2/n}g(0)^{-2/n}+\frac{kg(0)^2}{6(n+2)\eta^2} \ge \frac{n}{n+2}\,\omega_n^{-2/n}k^{1+2/n}g(0)^{-2/n}+\frac{\tau kg(0)^2}{(n+2)\eta^2}, \tag{2-19}$$

where τ is any constant with $0<\tau\le\frac{1}{6}$. From (2-4) we know that

$$\eta \ge 2(2\pi)^{-n}\Big(\frac{n}{n+2}\Big)^{1/2}\omega_n^{-1/n}|\Omega|^{(n+1)/n}. \tag{2-20}$$

Observe that $0<g(0)\le(2\pi)^{-n}|\Omega|$. Let $\tau=\tau(n)$ be the constant given by

$$\tau=\min\Big\{\frac{1}{6},\ \frac{16\pi^2 n}{(n+2)\omega_n^{4/n}}\Big\}.$$

Then one can see by using (2-20) that the function

$$\beta(t)=\frac{n}{n+2}\omega_n^{-2/n}k^{1+2/n}t^{-2/n}+\frac{\tau kt^2}{(n+2)\eta^2}$$

satisfies

$$\beta'((2\pi)^{-n}|\Omega|)\le 0,$$

and so β is decreasing on $(0, (2\pi)^{-n}|\Omega|]$. Hence, choosing $d(n)=\dfrac{\tau}{4(n+2)}$, we have

$$\begin{aligned}\sum_{j=1}^{k}\int_\Omega |\nabla w_j(x)|^2\,dx &\ge \beta(g(0))\ge\beta((2\pi)^{-n}|\Omega|)\\ &=\frac{n}{n+2}\,\omega_n^{-2/n}k^{1+2/n}((2\pi)^{-n}|\Omega|)^{-2/n}+\frac{\tau k((2\pi)^{-n}|\Omega|)^2}{(n+2)\eta^2}\\ &=\frac{n}{n+2}\Big(\frac{2\pi}{\omega_n^{1/n}}\Big)^2k^{1+2/n}|\Omega|^{-2/n}+\frac{d(n)k|\Omega|}{\int_\Omega|x|^2\,dx}.\end{aligned} \tag{2-21}$$

Substituting (2-3) into (2-21), one gets (2-2). completing the proof of Theorem 2.1. □

3. Proof of the main results

Proof of Theorem 1.1. Let $\{u_i\}_{i=1}^k$ be a set of orthonormal eigenfunctions corresponding to $\{\lambda_i\}_{i=1}^k$:

$$(-\Delta)^l u_i = \lambda_i u_i \quad \text{in } \Omega, \qquad u_i|_{\partial\Omega} = \Delta u_i|_{\partial\Omega} = \cdots = \Delta^{l-1} u_i|_{\partial\Omega} = 0,$$
$$\int_\Omega u_i u_j = \delta_{ij} \quad \text{for } i, j = 1, \dots, k.$$

We show that for each $s = 1, \dots, l$ and $i = 1, \dots, k$,

$$0 \le \int_\Omega u_i (-\Delta)^s u_i \le \lambda_i^{s/l}. \tag{3-1}$$

When $l = 1$, (3-1) holds trivially, so assume $l > 1$. when $s \in \{1, \dots, l\}$ is even, we have from the divergence theorem that

$$\int_\Omega u_i (-\Delta)^s u_i = \int_\Omega u_i \Delta^s u_i = \int_\Omega (\Delta^{s/2} u_i)^2 \ge 0.$$

On the other hand, if $s \in \{1, \dots, l\}$ is odd,

$$\begin{aligned}\int_\Omega u_i (-\Delta)^s u_i &= -\int_\Omega u_i \Delta^s u_i \\ &= -\int_\Omega \Delta^{(s-1)/2} u_i \Delta(\Delta^{(s-1)/2} u_i) = \int_\Omega |\nabla(\Delta^{(s-1)/2} u_i)|^2 \ge 0.\end{aligned}$$

Thus the first inequality in (3-1) holds.

We claim now that for any $s = 1, \dots, l-1$,

$$\left(\int_\Omega u_i (-\Delta)^s u_i\right)^{s+1} \le \left(\int_\Omega u_i (-\Delta)^{s+1} u_i\right)^s. \tag{3-2}$$

Since

$$\left(\int_\Omega u_i \Delta u_i\right)^2 \le \int_\Omega u_i^2 \int_\Omega (\Delta u_i)^2 = \int_\Omega u_i \Delta^2 u_i,$$

we know that (3-2) holds when $s = 1$.

Suppose that (3-2) is true for $s - 1$, that is,

$$\left(\int_\Omega u_i (-\Delta)^{s-1} u_i\right)^s \le \left(\int_\Omega u_i (-\Delta)^s u_i\right)^{s-1}. \tag{3-3}$$

When s is even, we have

$$\begin{aligned}\int_\Omega u_i (-\Delta)^s u_i &= \int_\Omega \Delta^{s/2-1} u_i \Delta(\Delta^{s/2} u_i) = -\int_\Omega \nabla(\Delta^{s/2-1} u_i) \nabla(\Delta^{s/2} u_i) \\ &\le \left(\int_\Omega |\nabla(\Delta^{s/2-1} u_i)|^2\right)^{1/2} \left(\int_\Omega |\nabla(\Delta^{s/2} u_i)|^2\right)^{1/2} \\ &= \left(-\int_\Omega \Delta^{s/2-1} u_i \Delta^{s/2} u_i\right)^{1/2} \left(-\int_\Omega \Delta^{s/2} u_i \Delta^{s/2+1} u_i\right)^{1/2} \\ &= \left(\int_\Omega u_i (-\Delta)^{s-1} u_i\right)^{1/2} \left(\int_\Omega u_i (-\Delta)^{s+1} u_i\right)^{1/2}.\end{aligned}$$

On the other hand, when s is odd,

$$\begin{aligned}\int_\Omega u_i(-\Delta)^s u_i &= \int_\Omega (-\Delta)^{(s-1)/2}u_i(-\Delta)^{(s+1)/2}u_i\\ &\le \left(\int_\Omega \left((-\Delta)^{(s-1)/2}u_i\right)^2\right)^{1/2}\left(\int_\Omega \left((-\Delta)^{(s+1)/2}u_i\right)^2\right)^{1/2}\\ &= \left(\int_\Omega u_i(-\Delta)^{s-1}u_i\right)^{1/2}\left(\int_\Omega u_i(-\Delta)^{s+1}u_i\right)^{1/2}.\end{aligned}$$

Thus we always have

$$\int_\Omega u_i(-\Delta)^s u_i \le \left(\int_\Omega u_i(-\Delta)^{s-1}u_i\right)^{1/2}\left(\int_\Omega u_i(-\Delta)^{s+1}u_i\right)^{1/2}. \tag{3-4}$$

Substituting (3-3) into (3-4), we know that (3-2) is true for s. Using (3-2) repeatedly, we get

$$\int_\Omega u_i(-\Delta)^s u_i \le \left(\int_\Omega u_i(-\Delta)^{s+1}u_i\right)^{s/(s+1)} \le \cdots \le \left(\int_\Omega u_i(-\Delta)^{l}u_i\right)^{s/l} = \lambda_i^{s/l}.$$

Thus the second inequality in (3-1) is also true. Consequently,

$$\sum_{j=1}^k \int_\Omega |\nabla u_j|^2 = \sum_{j=1}^k \int_\Omega u_j(-\Delta u_j) \le \sum_{j=1}^k \lambda_j^{1/l} \tag{3-5}$$

which implies (1-9) by applying Theorem 2.1 to the functions $u_1, \dots, u_k$. This completes the proof of Theorem 1.1. □

Proof of Theorem 1.2. Let $\{u_i\}_{i=1}^k$ be a set of orthonormal eigenfunctions of the problem (1-11) corresponding to $\{\lambda_i\}_{i=1}^k$:

$$Lu_i = \Lambda u_i \quad \text{in } \Omega, \qquad u_i|_{\partial\Omega} = \Delta u_i|_{\partial\Omega} = \cdots = \Delta^{l-1}u_i|_{\partial\Omega} = 0,$$
$$\int_\Omega u_i u_j = \delta_{ij} \quad \text{for } i, j = 1, \dots, k.$$

Denote by $\{\lambda_i\}_{i=1}^k$ the first k fixed membrane eigenvalues of the Laplacian of Ω corresponding to the orthonormal eigenfunctions $\{v_i\}_{i=1}^k$:

$$-\Delta v_i = \lambda_i v_i \quad \text{in } \Omega, \qquad v_i|_{\partial\Omega} = 0, \qquad \int_\Omega v_i v_j = \delta_{ij} \quad \text{for } i, j = 1, \dots, k.$$

Let $w = \sum_{j=1}^k \alpha_j u_j \neq 0$ be such that

$$\int_\Omega w v_i = 0 \quad \text{for } i = 1, \dots, k-1. \tag{3-6}$$

Such an element w exists because $\{\alpha_j \mid 1 \le j \le k\}$ is a nontrivial solution of a system of $k-1$ linear equations

$$\sum_{j=1}^{k} \alpha_j \int_\Omega u_j v_i = 0, \quad 1 \le i \le k-1, \tag{3-7}$$

in k unknowns. Also assume without loss of generality that

$$\int_\Omega w^2 = 1. \tag{3-8}$$

Then it follows from the Rayleigh–Ritz inequality that

$$\lambda_k \le \int_\Omega w(-\Delta w). \tag{3-9}$$

By using the arguments similar to those in the proof of (3-2), we have

$$\left(\int_\Omega w(-\Delta)^j w\right)^{j+1} \le \left(\int_\Omega w(-\Delta)^{j+1} w\right)^{j}, \quad j = 1, \dots, l-1. \tag{3-10}$$

Hence

$$\int_\Omega w(-\Delta w) \le \left(\int_\Omega w(-\Delta)^s w\right)^{1/s}, \quad s = 1, \dots, l, \tag{3-11}$$

which, combined with (3-9), gives

$$\lambda_k^s \le \int_M w(-\Delta)^s w, \quad s = 1, 2, \dots, l.$$

Thus we have

$$\begin{aligned} a_1\lambda_k + a_2\lambda_k^2 + \cdots + a_{l-1}\lambda_k^{l-1} + \lambda_k^l &\le \int_\Omega w\big(a_1(-\Delta) + a_2(-\Delta)^2 + \cdots + a_{l-1}(-\Delta)^{l-1} + (-\Delta)^l\big) w \\ &= \int_\Omega wLw = \sum_{i,j=1}^{k} \alpha_i\alpha_j \int_\Omega u_i L u_j = \sum_{i,j=1}^{k} \alpha_i\alpha_j \int_\Omega u_i \Lambda_j u_j \\ &= \sum_{i,j=1}^{k} \alpha_i\alpha_j \Lambda_j \delta_{ij} = \sum_{i=1}^{k} \alpha_i^2 \Lambda_i \le \Lambda_k, \end{aligned} \tag{3-12}$$

where, in the last equality, we have used the fact that

$$\sum_{i=1}^{k} \alpha_i^2 = \int_\Omega w^2 = 1.$$

It is easy to see by taking $l=1$ in (1-9) that

$$\lambda_k \geq \frac{nC(n)}{n+2}\left(\frac{k}{|\Omega|}\right)^{2/n} + \frac{d(n)}{\sum_{i=1}^n \mu_i^{-1}}. \tag{3-13}$$

Substituting (3-13) into (3-12), we get (1-11). Theorem 1.2 follows. □

Proof of Theorem 1.3. Let $\{u_i\}_{i=1}^k$ be a set of orthonormal eigenfunctions of the problem (1-12) corresponding to $\{\lambda_i\}_{i=1}^k$:

$$-\sum_{\alpha,\beta=1}^n \frac{\partial}{\partial x_\alpha}\left(a_{\alpha\beta}(x)\frac{\partial u_i}{\partial x_\beta}\right) + V(x)u_i = \lambda_i u_i \quad \text{in } \Omega, \tag{3-14}$$

$$u_i|_{\partial\Omega} = 0, \quad \int_\Omega u_i u_j = \delta_{ij} \quad \text{for } i, j = 1, \dots, k. \tag{3-15}$$

Multiplying (3-14) by u_i, integrating over Ω, and using the divergence theorem and the inequalities $V \geq V_0$ and $[a_{\alpha\beta}] \geq \xi_0 I$, we obtain

$$\begin{aligned}\lambda_i &= \int_\Omega \left(\sum_{\alpha,\beta=1}^n a_{\alpha\beta}(x)\frac{\partial u_i}{\partial x_\alpha}\frac{\partial u_i}{\partial x_\beta} + V(x)u_i^2\right)\\ &\geq \int_\Omega \xi_0 \sum_{\alpha=1}^n \left(\frac{\partial u_i}{\partial x_\alpha}\right)^2 + V_0\int_\Omega u_i^2 = \xi_0\int_\Omega |\nabla u_i|^2 + V_0,\end{aligned}$$

which gives

$$\sum_{i=1}^k \int_\Omega |\nabla u_i|^2 \leq \frac{1}{\xi_0}\left(\sum_{i=1}^k \lambda_i - kV_0\right). \tag{3-16}$$

Observing (3-15), one gets (1-13) by using Theorem 2.1 applied to $u_1, \dots, u_k$. Theorem 1.3 follows. □

Proof of Theorem 1.4. Let $x_1, \dots, x_n$ be the coordinate functions on $\mathbb{R}^n$. For a function $f : \Omega \to \mathbb{R}$, set $f_{,\alpha} = \partial f/\partial x_\alpha$, $\alpha = 1, \dots, n$. Let $\{u_i\}_{i=1}^k$ be a set of orthonormal eigenfunctions of the problem (1-14) corresponding to $\{\lambda_i\}_{i=1}^k$:

$$-\sum_{\alpha,\beta=1}^n \frac{\partial}{\partial x_\alpha}\left(a_{\alpha\beta}(x)\frac{\partial u_i}{\partial x_\beta}\right) + V(x)u_i = \lambda_i \rho u_i \quad \text{in } \Omega,$$

$$u_i|_{\partial\Omega} = 0, \quad \int_\Omega u_i u_j = \delta_{ij} \quad \text{for } i, j = 1, \dots, k.$$

For each $\alpha = 1, \dots, n$ and $i = 1, \dots, k$, following [Payne et al. 1956], consider the functions $\phi_{\alpha i} : \Omega \to \mathbb{R}$ given by

$$\phi_{\alpha i} = x_\alpha u_i - \sum_{j=1}^k r_{\alpha ij} u_j, \tag{3-17}$$

where

$$r_{\alpha ij} = \int_\Omega \rho x_\alpha u_i u_j. \tag{3-18}$$

Since $\phi_{\alpha i}|_{\partial\Omega} = 0$ and

$$\int_\Omega \rho u_j \phi_{\alpha i} = 0 \quad \text{for } i, j = 1, \dots, k \text{ and } \alpha = 1, \dots, n,$$

it follows from the Rayleigh–Ritz inequality that

$$\begin{aligned} \lambda_{k+1} \int_\Omega \rho \phi_{\alpha i}^2 &\le \int_\Omega \phi_{\alpha i}\Big(-\sum_{\beta,\gamma=1}^n (a_{\beta\gamma}\phi_{\alpha i,\gamma})_{,\beta} + V\phi_{\alpha i}\Big) \\ &= \int_\Omega \phi_{\alpha i}\Big(-\sum_{\beta,\gamma=1}^n (a_{\beta\gamma}(x_\alpha u_i)_{,\gamma})_{,\beta} + V x_\alpha u_i - \sum_{j=1}^k r_{\alpha ij}\lambda_j \rho u_j\Big) \\ &= \int_\Omega \phi_{\alpha i}\Big(-\sum_{\beta,\gamma=1}^n (a_{\beta\gamma}(x_\alpha u_i)_{,\gamma})_{,\beta} + V x_\alpha u_i\Big) \\ &= \int_\Omega \phi_{\alpha i}\Big(\lambda_i \rho x_\alpha u_i - \sum_{\beta=1}^n ((a_{\alpha\beta} u_i)_{,\beta} + a_{\alpha\beta} u_{i,\beta})\Big) \\ &= \lambda_i \int_\Omega \rho\phi_{\alpha i}^2 - \int_\Omega \phi_{\alpha i}\Big(\sum_{\beta=1}^n ((a_{\alpha\beta} u_i)_{,\beta} + a_{\alpha\beta} u_{i,\beta})\Big) \\ &= \lambda_i \int_\Omega \rho\phi_{\alpha i}^2 - \int_\Omega x_\alpha u_i\Big(\sum_{\beta=1}^n ((a_{\alpha\beta} u_i)_{,\beta} + a_{\alpha\beta} u_{i,\beta})\Big) + \sum_{j=1}^k r_{\alpha ij} s_{\alpha ij}, \end{aligned} \tag{3-19}$$

where

$$s_{\alpha ij} = \int_\Omega \Big(\sum_{\beta=1}^n ((a_{\alpha\beta} u_i)_{,\beta} + a_{\alpha\beta} u_{i,\beta})\Big) u_j.$$

Multiplying the equation

$$-\sum_{\beta,\gamma=1}^n (a_{\beta\gamma} u_{j,\beta})_{,\gamma} + V u_j = \lambda_j \rho u_j \tag{3-20}$$

by $x_\alpha u_i$, we have

$$-\sum_{\beta,\gamma=1}^n (a_{\beta\gamma} u_{j,\beta})_{,\gamma} x_\alpha u_i + V x_\alpha u_i u_j = \lambda_j \rho x_\alpha u_i u_j. \tag{3-21}$$

Interchanging the roles of i and j, we get

$$(3\text{-}22)\qquad -\sum_{\beta\gamma=1}^{n}(a_{\beta\gamma}u_{i,\beta})_{,\gamma}x_\alpha u_j+Vx_\alpha u_iu_j=\lambda_i\rho x_\alpha u_iu_j.$$

Subtracting (3-21) from (3-22) and integrating the resulted equation on Ω, we get by using the divergence theorem that

$$\begin{aligned}(3\text{-}23)\qquad (\lambda_i-\lambda_j)r_{\alpha ij}&=\sum_{\beta,\gamma=1}^{n}\int_\Omega\big((a_{\beta\gamma}u_{j,\beta})_{,\gamma}x_\alpha u_i-(a_{\beta\gamma}u_{i,\beta})_{,\gamma}x_\alpha u_j\big)\\&=\sum_{\beta,\gamma=1}^{n}\int_\Omega\big(-a_{\beta\gamma}u_{j,\beta}(x_\alpha u_i)_{,\gamma}+a_{\beta\gamma}u_{i,\beta}(x_\alpha u_j)_{,\gamma}\big)\\&=\sum_{\beta=1}^{n}\int_\Omega\big(-a_{\alpha\beta}u_{j,\beta}u_i+a_{\alpha\beta}u_{i,\beta}u_j\big)\\&=\sum_{\beta=1}^{n}\int_\Omega\big((a_{\alpha\beta}u_i)_{,\beta}+a_{\alpha\beta}u_{i,\beta}\big)u_j=s_{\alpha ij},\end{aligned}$$

which, combined with (3-19), gives

$$\begin{aligned}(3\text{-}24)\qquad (\lambda_{k+1}-\lambda_i)\int_\Omega\rho\phi_{\alpha i}^2&\\&\le-\int_\Omega\phi_{\alpha i}\bigg(\sum_{\beta=1}^{n}((a_{\alpha\beta}u_i)_{,\beta}+a_{\alpha\beta}u_{i,\beta})\bigg)\\&=-\int_\Omega x_\alpha u_i\bigg(\sum_{\beta=1}^{n}((a_{\alpha\beta}u_i)_{,\beta}+a_{\alpha\beta}u_{i,\beta})\bigg)+\sum_{j=1}^{k}(\lambda_i-\lambda_j)r_{\alpha ij}^2.\end{aligned}$$

Set

$$t_{\alpha ij}=\int_\Omega u_ju_{i,\alpha};$$

then $t_{\alpha ij}+t_{\alpha ji}=0$ and

$$(3\text{-}25)\qquad \int_\Omega(-2)\phi_{\alpha i}u_{i,\alpha}=-2\int_\Omega x_\alpha u_iu_{i,\alpha}+2\sum_{j=1}^{k}r_{\alpha ij}t_{\alpha ij}=\|u_i\|^2+2\sum_{j=1}^{k}r_{\alpha ij}t_{\alpha ij}.$$

Multiplying (3-25) by $(\lambda_{k+1}-\lambda_i)^2$ and using the Schwarz inequality and (3-24), we get

$$(\lambda_{k+1}-\lambda_i)^2\bigg(\|u_i\|^2+2\sum_{j=1}^{k}r_{\alpha ij}t_{\alpha ij}\bigg)$$

$$
\begin{aligned}
&= (\lambda_{k+1}-\lambda_i)^2 \int_\Omega (-2)\sqrt{\rho}\phi_{\alpha i}\Big(\frac{1}{\sqrt{\rho}}u_{i,\alpha} - \sum_{j=1}^k t_{\alpha ij}\sqrt{\rho}u_j\Big)\\
&\le \delta(\lambda_{k+1}-\lambda_i)^3\|\sqrt{\rho}\phi_{\alpha i}\|^2 + \frac{(\lambda_{k+1}-\lambda_i)}{\delta}\int_\Omega\Big|\frac{1}{\sqrt{\rho}}u_{i,\alpha} - \sum_{j=1}^k t_{\alpha ij}\sqrt{\rho}u_j\Big|^2\\
&= \delta(\lambda_{k+1}-\lambda_i)^3\|\sqrt{\rho}\phi_{\alpha i}\|^2 + \frac{(\lambda_{k+1}-\lambda_i)}{\delta}\Big(\Big\|\frac{1}{\sqrt{\rho}}u_{i,\alpha}\Big\|^2 - \sum_{j=1}^k t_{\alpha ij}^2\Big)\\
&\le \delta(\lambda_{k+1}-\lambda_i)^2\Big(-\int_\Omega x_\alpha u_i\Big(\sum_{\beta=1}^n((a_{\alpha\beta}u_i)_{,\beta} + a_{\alpha\beta}u_{i,\beta})\Big) + \sum_{j=1}^k(\lambda_i-\lambda_j)r_{\alpha ij}^2\Big)\\
&\qquad\qquad + \frac{(\lambda_{k+1}-\lambda_i)}{\delta}\Big(\Big\|\frac{1}{\sqrt{\rho}}u_{i,\alpha}\Big\|^2 - \sum_{j=1}^k t_{\alpha ij}^2\Big),
\end{aligned}
$$

where δ is any positive constant. Summing over i and noticing that $r_{\alpha ij} = r_{\alpha ji}$ and $t_{\alpha ij} = -t_{\alpha ji}$, we infer that

$$
\begin{aligned}
&\sum_{i=1}^k(\lambda_{k+1}-\lambda_i)^2\|u_i\|^2 - 2\sum_{i,j=1}^k(\lambda_{k+1}-\lambda_i)(\lambda_i-\lambda_j)r_{\alpha ij}t_{\alpha ij}\\
&\le \delta\sum_{i=1}^k(\lambda_{k+1}-\lambda_i)^2\Big(-\int_\Omega x_\alpha u_i\Big(\sum_{\beta=1}^n((a_{\alpha\beta}u_i)_{,\beta} + a_{\alpha\beta}u_{i,\beta})\Big)\Big)\\
&\quad + \sum_{i=1}^k\frac{(\lambda_{k+1}-\lambda_i)}{\delta}\Big\|\frac{1}{\sqrt{\rho}}u_{i,\alpha}\Big\|^2 - \sum_{i,j=1}^k(\lambda_{k+1}-\lambda_i)\delta(\lambda_i-\lambda_j)^2 r_{\alpha ij}^2 - \sum_{i,j=1}^k\frac{(\lambda_{k+1}-\lambda_i)}{\delta}t_{\alpha ij}^2.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\sum_{i=1}^k(\lambda_{k+1}-\lambda_i)^2\|u_i\|^2 &\le \delta\sum_{i=1}^k(\lambda_{k+1}-\lambda_i)^2\Big(-\int_\Omega x_\alpha u_i\Big(\sum_{\beta=1}^n((a_{\alpha\beta}u_i)_{,\beta} + a_{\alpha\beta}u_{i,\beta})\Big)\Big)\\
&\qquad + \sum_{i=1}^k\frac{(\lambda_{k+1}-\lambda_i)}{\delta}\Big\|\frac{1}{\sqrt{\rho}}u_{i,\alpha}\Big\|^2.
\end{aligned}
$$

Summing over α, we infer

$$
\begin{aligned}
\text{(3-26)}\quad & n\sum_{i=1}^k(\lambda_{k+1}-\lambda_i)^2\|u_i\|^2\\
&\le \delta\sum_{i=1}^k(\lambda_{k+1}-\lambda_i)^2\Big(-\sum_{\alpha=1}^n\int_\Omega x_\alpha u_i\Big(\sum_{\beta=1}^n((a_{\alpha\beta}u_i)_{,\beta} + a_{\alpha\beta}u_{i,\beta})\Big)\Big)\\
&\qquad + \sum_{i=1}^k\frac{(\lambda_{k+1}-\lambda_i)}{\delta}\Big\|\frac{1}{\sqrt{\rho}}|\nabla u_i|\Big\|^2.
\end{aligned}
$$

Since $\int_\Omega \rho u_i^2 = 1$ and $\rho_1 \le \rho(x) \le \rho_2$ for $x \in \Omega$, we have

$$\frac{1}{\rho_2} \le \|u_i\|^2 \le \frac{1}{\rho_1}. \tag{3-27}$$

One gets from the divergence theorem that

$$\begin{aligned} -\sum_{\alpha=1}^{n} \int_\Omega x_\alpha u_i \Big(\sum_{\beta=1}^{n} \big((a_{\alpha\beta} u_i)_{,\beta} + a_{\alpha\beta} u_{i,\beta} \big) \Big) \\ = \int_\Omega \Big(\sum_{\alpha,\beta=1}^{n} \big(a_{\alpha\beta} u_i (x_\alpha u_i)_{,\beta} - a_{\alpha\beta} u_{i,\beta} x_\alpha u_i \big) \Big) = \int_\Omega \Big(\sum_{\alpha=1}^{n} a_{\alpha\alpha} \Big) u_i^2 \\ \le n\xi_2 \int_\Omega u_i^2 \le \frac{n\xi_2}{\rho_1}. \end{aligned} \tag{3-28}$$

Multiplying the equation $-\sum_{\alpha,\beta=1}^{n} (a_{\alpha\beta} u_{i,\beta})_{,\alpha} + V(x) u_i = \lambda_i \rho u_i$ by u_i and integrating over Ω, we get

$$\lambda_i = \int_\Omega \Big(\sum_{\alpha,\beta=1}^{n} a_{\alpha\beta}(x) u_{i,\alpha} u_{i,\beta} + V(x) u_i^2 \Big) \ge \int_\Omega \xi_1 |\nabla u_i|^2 + \frac{V_0}{\rho_2}, \tag{3-29}$$

which gives

$$\Big\| \frac{1}{\sqrt{\rho}} |\nabla u_i| \Big\|^2 \le \frac{1}{\rho_1} \int_\Omega |\nabla u_i|^2 \le \frac{1}{\rho_1 \xi_1} \Big(\lambda_i - \frac{V_0}{\rho_2} \Big). \tag{3-30}$$

Substituting (3-27), (3-28) and (3-30) into (3-26), we infer

$$\frac{n}{\rho_2} \sum_{i=1}^{k} (\lambda_{k+1} - \lambda_i)^2 \le \delta \sum_{i=1}^{k} (\lambda_{k+1} - \lambda_i)^2 \cdot \frac{n\xi_2}{\rho_1} + \sum_{i=1}^{k} \frac{(\lambda_{k+1} - \lambda_i)}{\delta} \cdot \frac{1}{\rho_1 \xi_1} \Big(\lambda_i - \frac{V_0}{\rho_2} \Big).$$

Taking

$$\delta = \frac{\Big(\sum_{i=1}^{k} (\lambda_{k+1} - \lambda_i) \Big(\lambda_i - \frac{V_0}{\rho_2} \Big) \Big)^{1/2}}{\Big(\sum_{i=1}^{k} (\lambda_{k+1} - \lambda_i)^2 n \xi_1 \xi_2 \Big)^{1/2}}, \tag{3-31}$$

we get (1-15). This completes the proof of Theorem 1.4. □

References

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

[Ashbaugh 2002] M. S. Ashbaugh, "The universal eigenvalue bounds of Payne–Pólya–Weinberger, Hile–Protter, and H. C. Yang", pp. 3–30 in *Spectral and inverse spectral theory* (Goa, 2000), edited by P. D. Hislop and M. Krishna, Proc. Indian Acad. Sci. Math. Sci. **112**, 2002. MR 2004c:35302 Zbl 1199.35261

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

[Ashbaugh and Benguria 1993b] M. S. Ashbaugh and R. D. Benguria, "Universal bounds for the low eigenvalues of Neumann Laplacians in n dimensions", *SIAM J. Math. Anal.* **24**:3 (1993), 557–570. MR 94b:35191 Zbl 0796.35122

[Ashbaugh and Hermi 2004] M. S. Ashbaugh and L. Hermi, "A unified approach to universal inequalities for eigenvalues of elliptic operators", *Pacific J. Math.* **217**:2 (2004), 201–219. MR 2005k:35305 Zbl 1078.35080

[Berezin 1972] F. A. Berezin, "Covariant and contravariant symbols of operators", *Izv. Akad. Nauk SSSR Ser. Mat.* **36** (1972), 1134–1167. In Russian. MR 50 #2996 Zbl 0259.47004

[Chavel 1984] I. Chavel, *Eigenvalues in Riemannian geometry*, Pure and Applied Math. **115**, Academic Press, Orlando, FL, 1984. MR 86g:58140 Zbl 0551.53001

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

[Cheng and Yang 2006a] Q.-M. Cheng and H. Yang, "Inequalities for eigenvalues of a clamped plate problem", *Trans. Amer. Math. Soc.* **358**:6 (2006), 2625–2635. MR 2006m:35263 Zbl 1096.35095

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

[Cheng and Yang 2006c] Q.-M. Cheng and H. Yang, "Universal bounds for eigenvalues of a buckling problem", *Comm. Math. Phys.* **262**:3 (2006), 663–675. MR 2007f:35056 Zbl 1170.35379

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

[El Soufi et al. 2007] A. El Soufi, E. M. Harrell, and S. Ilias, "Universal inequalities for the eigenvalues of Laplace and Schrödinger operators on submanifolds", preprint, 2007. to appear in *Trans. Amer. Math. Soc.* arXiv 0706.0910

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

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

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

[Harrell and Yıldırım Yolcu 2009] E. M. Harrell, II and S. Yıldırım Yolcu, "Eigenvalue inequalities for Klein–Gordon operators", *J. Funct. Anal.* **256**:12 (2009), 3977–3995. MR 2010e:35198 Zbl 05572175

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

[Hook 1990] S. M. Hook, "Domain-independent upper bounds for eigenvalues of elliptic operators", *Trans. Amer. Math. Soc.* **318**:2 (1990), 615–642. MR 90h:35075 Zbl 0727.35096

[Kröger 1992] P. Kröger, "Upper bounds for the Neumann eigenvalues on a bounded domain in Euclidean space", *J. Funct. Anal.* **106**:2 (1992), 353–357. MR 93d:47091 Zbl 0777.35044

[Kröger 1994] P. Kröger, "Estimates for sums of eigenvalues of the Laplacian", *J. Funct. Anal.* **126**:1 (1994), 217–227. MR 95j:58173 Zbl 0817.35066

[Laptev 1997] A. Laptev, "Dirichlet and Neumann eigenvalue problems on domains in Euclidean spaces", *J. Funct. Anal.* **151**:2 (1997), 531–545. MR 99a:35027 Zbl 0892.35115

[Laptev and Weidl 2000] A. Laptev and T. Weidl, "Recent results on Lieb–Thirring inequalities", pp. XX–1–14 in *Journées "Équations aux Dérivées Partielles"* (La Chapelle sur Erdre, 2000), Univ. Nantes, Nantes, 2000. MR 2001j:81064 Zbl 1135.81337

[Levitin and Parnovski 2002] M. Levitin and L. Parnovski, "Commutators, spectral trace identities, and universal estimates for eigenvalues", *J. Funct. Anal.* **192**:2 (2002), 425–445. MR 2003g:47040 Zbl 1058.47022

[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 84k:58225 Zbl 0554.35029

[Melas 2003] A. D. Melas, "A lower bound for sums of eigenvalues of the Laplacian", *Proc. Amer. Math. Soc.* **131**:2 (2003), 631–636. MR 2003i:35218 Zbl 1015.58011

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

[Pólya 1961] G. Pólya, "On the eigenvalues of vibrating membranes", *Proc. London Math. Soc.* (3) **11** (1961), 419–433. MR 23 #B2256 Zbl 0107.41805

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

[Wang and Xia 2007a] Q. Wang and C. Xia, "Universal bounds for eigenvalues of the biharmonic operator on Riemannian manifolds", *J. Funct. Anal.* **245**:1 (2007), 334–352. MR 2008e:58033 Zbl 1113.58013

[Wang and Xia 2007b] Q. Wang and C. Xia, "Universal inequalities for eigenvalues of the buckling problem on spherical domains", *Comm. Math. Phys.* **270**:3 (2007), 759–775. MR 2007m:35180 Zbl 1112.74017

[Wang and Xia 2008] Q. Wang and C. Xia, "Universal bounds for eigenvalues of Schrödinger operator on Riemannian manifolds", *Ann. Acad. Sci. Fenn. Math.* **33** (2008), 319–336. MR 2009c:35333 Zbl 1171.35091

[Wang and Xia 2010a] Q. Wang and C. Xia, "Inequalities for eigenvalues of the biharmonic operator with weight on Riemannian manifolds", *J. Math. Soc. Japan* **62**:2 (2010), 597–622. MR 2662854 Zbl 1200.53042

[Wang and Xia 2010b] Q. Wang and C. Xia, "Isoperimetric bounds for the first eigenvalue of the Laplacian", *Z. Angew. Math. Phys.* **61**:1 (2010), 171–175. MR 2011e:35246 Zbl 1192.35125

[Wang and Xia 2010c] Q. Wang and C. Xia, "Universal bounds for eigenvalues of the biharmonic operator", *J. Math. Anal. Appl.* **364**:1 (2010), 1–17. MR 2010k:35334 Zbl 1189.35208

[Wang and Xia 2011] Q. Wang and C. Xia, "Inequalities for eigenvalues of a clamped plate problem", *Calc. Var. Partial Differential Equations* **40** (2011), 273–289. MR 2745203 Zbl 1205.35175

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

[Yang 1991] H. C. Yang, "An estimate of the difference between consecutive eigenvalues", preprint IC/91/60, ICTP, Trieste, 1991.

Received April 22, 2009. Revised October 20, 2009.

QIAOLING WANG
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRASÍLIA
ASA NORTE
70910-900 BRASÍLIA
BRAZIL
wang@impa.br

CHANGYU XIA
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRASÍLIA
ASA NORTE
70910-900 BRASÍLIA
BRAZIL
xia@pq.cnpq.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 251, No. 1, 2011

A BEURLING–HÖRMANDER THEOREM ASSOCIATED WITH THE RIEMANN–LIOUVILLE OPERATOR

XUECHENG WANG

We establish an analogue of the Beurling theorem associated with the Riemann–Liouville operator. We also derive some other versions of uncertainty principle theorems associated with this operator.

1. Introduction and the main result

The uncertainty principle, which plays an important role in harmonic analysis, states that a nonzero function and its Fourier transform cannot simultaneously be very small at infinity. This principle has been researched on various aspects and has several versions named after Hardy, Morgan, Cowling and Price, Gelfand, Beurling and others. The Beurling theorem is the most general case since it implies the other uncertainty principles.

The classical Beurling theorem was proved by Hörmander [1991] and generalized to d dimensions by Bonami et al. [2003]. Here we record the general case:

Lemma 1.1. *For $f \in L^2(\mathbb{R}^d)$ and $N \geqq 0$, if*

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|f(x)|\,|\widehat{f}(y)|\,e^{\|x\|\,\|y\|}}{(1+\|x\|+\|y\|)^N}\,\mathrm{d}x\,\mathrm{d}y < \infty,$$

then $f(x) = P(x)\,e^{-a\langle Ax,x\rangle}$, $a > 0$, where A is a real positive definite symmetric matrix and $P(x)$ is a polynomial of degree $< (N-d)/2$. In particular, $f = 0$ when $N \le d$.

In the lemma and the rest of the paper, $\widehat{f}$ is the classic Fourier transform of f in $\mathbb{R}^d$, defined by

$$\widehat{f}(\lambda) = \int_{\mathbb{R}^d} f(x)\,e^{-i\,\lambda\,x}\,\mathrm{d}x, \quad \lambda \in \mathbb{R}^d.$$

The Beurling theorem has been generalized to different settings. L. Bouattour established an analogue in the framework of Chébli–Trimèche hypergroups $(\mathbb{R}_+, *(A))$ (see [Bouattour and Trimèche 2005]). J. Z. Huang and H. P. Liu [2007a;

MSC2000: 33C45, 42B10.
Keywords: uncertainty principle, Riemann–Liouville operator, Beurling–Hörmander theorem, Riemann–Liouville transform.

2007b] gave analogues for the Laguerre hypergroup and the Heisenberg group. R. P. Sarkar and J. Sengupta [2007b] established the analogue of the Beurling theorem on the full group SL(2, $\mathbb{R}$). As for the noncompact semisimple Lie group case, S. Thangavelu [2004] first gave the analogue on rank 1 symmetric spaces with an additional condition like the one required in the Cowling–Price theorem, so he called it the Cowbeurling Theorem; then R. P. Sarkar and J. Sengupta [2007a] removed this additional condition and gave the analogue in rank 1 symmetric spaces; recently, L. Bouattour [2008] generalized this result and gave the analogue for real symmetric spaces of rank d. For more Beurling theorems in different settings, refer to [Kamoun and Trimèche 2005; Parui and Sarkar 2008].

In this paper, for $\alpha \geq 0$ we consider the singular partial differential operators

$$\begin{cases} \Delta_1 = \dfrac{\partial}{\partial x}, \\ \Delta_2 = \dfrac{\partial^2}{\partial r^2} + \dfrac{2\alpha+1}{r}\dfrac{\partial}{\partial r} - \dfrac{\partial^2}{\partial x^2}, \quad (r, x) \in (0, +\infty) \times \mathbb{R},\ \alpha \geqq 0, \end{cases}$$

originally studied in [Baccar et al. 2006; Omri and Rachdi 2008]. The latter authors have proved an uncertainty principle that generalized the Heisenberg–Pauli–Weyl inequality for the classical Fourier transform:

Proposition [Omri and Rachdi 2008]. *For all $f \in L^2(d\upsilon_\alpha)$, we have*

$$\| \, |(r, x)| f \|_{2,\upsilon_\alpha} \, \|(\mu^2 + 2\lambda^2)^{1/2} \mathscr{F}_\alpha(f)\|_{2,\gamma_\alpha} \geqq \frac{2\alpha+3}{2} \|f\|^2_{2,\upsilon_\alpha}$$

with equality if and only if

$$f(r, x) = Ce^{-(r^2+x^2)/2t_0^2} \quad \text{for } (r, x) \in \mathbb{R}_+ \times \mathbb{R},\ t_0 > 0,\ C \in \mathbb{C},$$

where $d\upsilon_\alpha$ is a measure defined on $\mathbb{R}_+ \times \mathbb{R}$ by

$$d\upsilon_\alpha(r, x) = \mathrm{d}c(r) \otimes \mathrm{d}x \quad \text{with } \mathrm{d}c(r) \overset{\text{def}}{=} \frac{r^{2\alpha+1}}{2^\alpha\, \Gamma(\alpha+1)\sqrt{2\pi}} \mathrm{d}r; \tag{1}$$

$dr_\alpha(\mu, \lambda)$ is a measure defined on the set Γ_+

$$\Gamma_+ = \mathbb{R}_+ \times \mathbb{R} \cup \big\{(it, x) : (t, x) \in R_+ \times R,\ t \leqq |x|\big\};$$

$|(r, x)|$ is the Euclidean norm in $\mathbb{R}^2$, that is, $|(r, x)| = (r^2 + x^2)^{1/2}$; and $\mathscr{F}_\alpha(f)$ is the generalized Fourier transform associated with the Riemann–Liouville operator.

Our main result is an analogue of the Beurling–Hörmander theorem for this generalized Fourier transform $\mathscr{F}_\alpha$ associated with the Riemann–Liouville operator:

Theorem 1.2. *Let $K = \mathbb{R}_+ \times \mathbb{R}$, and assume $N \geqq 0$. For $f \in L^2(K, d\nu_\alpha)$, if*

$$\int_{\Gamma_+} \int_K \frac{|f(r,x)|\,|\mathscr{F}_\alpha(f)(\mu,\lambda)|\,e^{|x||\lambda|}}{(1+|x|+|\lambda|)^N}\,\mathrm{d}\nu_\alpha(r,x)\,\mathrm{d}r_\alpha(\mu,\lambda) < \infty,$$

then

$$f(r,x) = e^{-ax^2}\biggl(\sum_{j=0}^{k} \psi_j(r)x^j\biggr),$$

where $a > 0$, $k < \dfrac{N-1}{2}$, and $\psi_j(r) \in L^2\Bigl([0,+\infty), \dfrac{r^{2\alpha+1}}{2^\alpha\Gamma(\alpha+1)}\,\mathrm{d}r\Bigr)$. In particular, when $N \leqq 3$,

$$f(r,x) = e^{-ax^2}\,\psi(r),$$

where $\psi(r) \in L^2([0,+\infty), r^{2\alpha+1}/(2^\alpha\Gamma(\alpha+1))\,\mathrm{d}r)$, and when $N \leqq 1$, we have $f = 0$.

Section 2 contains some preliminary facts about the Riemann–Liouville operator and the generalized Fourier transform. In Section 3, we prove Theorem 1.2. In Section 4, we give some other uncertainty principles. In Section 5, we give a stronger result but at the cost of more strictly constraining the function $f(r,x)$ by utilizing the Riemann–Liouville transform and its dual.

2. Preliminaries

In this section, we set some notation and theorems about the generalized Fourier transform associated with Riemann–Liouville operator. For detailed information, refer to [Baccar et al. 2006; Hamadi and Rachdi 2006; Omri and Rachdi 2008].

From this last reference we know that for all $(\mu,\lambda) \in \mathbb{C}^2$, the system

$$\begin{cases} \Delta_1 u(r,x) = -i\,\lambda\,u(r,x), \\ \Delta_2 u(r,x) = -\mu^2\,u(r,x), \\ u(0,0) = 1,\ \ (\partial u/\partial r)(0,x) = 0, \quad x \in \mathbb{R} \end{cases}$$

admits a unique solution $\varphi_{\mu.\lambda}$, given by

$$\varphi_{\mu,\lambda}(r,x) = j_\alpha(r\sqrt{\mu^2+\lambda^2})\,e^{-i\lambda x} \quad \text{for } (\mu,\lambda) \in \mathbb{R}^2, \tag{2}$$

where

$$j_\alpha(x) = 2^\alpha\,\Gamma(\alpha+1)\frac{J_\alpha(x)}{x^\alpha} = \Gamma(\alpha+1)\sum_0^\infty \frac{(-1)^n}{n!\Gamma(\alpha+n+1)}\Bigl(\frac{x}{2}\Bigr)^{2n}, \tag{3}$$

and $J_\alpha(x)$ is a Bessel function of the first kind of index α. The modified Bessel function j_α has the following integral representation: for all $\mu,\ r \in \mathbb{R}_+$ we have

$$j_\alpha(r\mu) = \begin{cases} \dfrac{2\Gamma(\alpha+1)}{\sqrt{\pi}\,\Gamma(\alpha+1/2)} \displaystyle\int_0^1 (1-t^2)^{\alpha-1/2} \cos(r\mu t)\,\mathrm{d}t & \text{if } \alpha > -1/2, \\ \cos(r\mu) & \text{if } \alpha = -1/2. \end{cases}$$

The Riemann–Liouville integral transform associated with Δ_1, Δ_2 is defined by

$$\mathcal{R}_\alpha(f)(r,x) = \begin{cases} \dfrac{\alpha}{\pi} \displaystyle\int_{-1}^1 \int_{-1}^1 f(rs\sqrt{1-t^2}, x+rt)(1-t^2)^{\alpha-1/2}(1-s^2)^{\alpha-1}\,\mathrm{d}t\,\mathrm{d}s \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \alpha > 0, \\ \dfrac{1}{\pi} \displaystyle\int_{-1}^1 f(r\sqrt{1-t^2}, x+rt)\,\frac{\mathrm{d}t}{\sqrt{1-t^2}} & \text{if } \alpha = 0. \end{cases}$$

Now we give some properties of the eigenfunction $\varphi_{\mu,\lambda}$.

(i) The supremum of $\varphi_{\mu,\lambda}$ satisfies

$$\sup_{(r,x)\in\mathbb{R}^2} |\varphi_{\mu,\lambda}(r,x)| = 1$$

if and only if (μ,λ) belongs to the set

$$\Gamma = \mathbb{R}^2 \cup \{(it,x) : (t,x) \in \mathbb{R}^2,\ |t| \leqq |x|\}.$$

(ii) The eigenfunction $\varphi_{\mu,\lambda}$ has Mehler integral representation

$$\varphi_{\mu,\lambda}(r,x) = \begin{cases} \dfrac{\alpha}{\pi} \displaystyle\int_{-1}^1 \int_{-1}^1 f(rs\sqrt{1-t^2}, x+rt)(1-t^2)^{\alpha-1/2}\,(1-s^2)^{\alpha-1}\,\mathrm{d}t\,\mathrm{d}s \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \alpha > 0, \\ \dfrac{1}{\pi} \displaystyle\int_{-1}^1 f(r\sqrt{1-t^2}, x+rt)\,\frac{\mathrm{d}t}{\sqrt{1-t^2}} & \text{if } \alpha = 0, \end{cases}$$

where f is a continuous function on $\mathbb{R}^2$.

From our definition, we can see that the transform $\mathcal{R}_\alpha$ generalizes the "mean operator" defined by

$$\mathcal{R}_0(f)(r,x) = \frac{1}{\pi} \int_0^{2\pi} f(r\ \sin(\theta), x + r\ \cos(\theta))\,\mathrm{d}\theta.$$

In the remainder of the paper, we use the following notation:

(i) $L^p(d\nu_\alpha)$ denotes the space of measurable functions f on $K = \mathbb{R}_+ \times \mathbb{R}$ such that

$$\|f\|_{p,\nu_\alpha} = \left(\int_0^\infty \int_{\mathbb{R}} |f(r,x)|^p \, d\nu_\alpha(r,x)\right)^{1/p} < \infty \quad \text{if } p \in [1,+\infty),$$
$$\|f\|_{\infty,\nu_\alpha} = \operatorname*{ess\,sup}_{(r,x)\in K} |f(r,x)| < +\infty \quad \text{if } p = +\infty.$$

(ii) $\langle\ ,\ \rangle_{\nu_\alpha}$ is the inner product defined on $L^2(d\nu_\alpha)$ by

$$\langle f, g\rangle_{\nu_\alpha} = \int_0^\infty \int_R f(r,x)\overline{g(r,x)}\, d\nu_\alpha(r,x).$$

(iii) $\Gamma_+ = \mathbb{R}_+ \times \mathbb{R} \cup \big\{(it,x) : (t,x) \in \mathbb{R}_+ \times \mathbb{R},\, t \le |x|\big\}$.

(iv) $\mathscr{B}_{\Gamma_+}$ is a σ-algebra defined on Γ_+ by

$$\mathscr{B}_{\Gamma_+} = \big\{\theta^{-1}(B) : B \in \mathscr{B}(\mathbb{R}_+ \times \mathbb{R})\big\},$$

where θ is the bijective function defined on the set Γ_+ by

$$\theta(\mu,\lambda) = (\sqrt{\mu^2+\lambda^2}, \lambda).$$

(v) Θ is the operator given by $(\Theta \circ f)(\mu,\lambda) = f(\theta(\mu,\lambda))$ for any function f defined on Γ_+.

(vi) $d\gamma_\alpha$ is a measure on $\mathscr{B}_{\Gamma_+}$ given by

$$\gamma_\alpha(A) = \nu_\alpha(\theta(A)) \quad \text{for } A \in \mathscr{B}_{\Gamma_+}.$$

(vii) Let $L^p(d\gamma_\alpha)$ denote the space of measurable functions f on Γ_+ such that

$$\|f\|_{p,\gamma_\alpha} = \left(\iint_{\Gamma_+} |f(\mu,\lambda)|^p d\gamma_\alpha(\mu,\lambda)\right)^{1/p} < \infty \quad \text{if } p \in [1,+\infty),$$
$$\|f\|_{\infty,\gamma_\alpha} = \operatorname*{ess\,sup}_{(\mu,\lambda)\in\Gamma_+} e|f(\mu,\lambda)| < +\infty \quad \text{if } p = +\infty.$$

(viii) $\langle\ ,\ \rangle_{\gamma_\alpha}$ is the inner product defined on $L^2(d\gamma_\alpha)$ by

$$\langle f, g\rangle_{\gamma_\alpha} = \int_{\Gamma_+} f(\mu,\lambda)\overline{g(\mu,\lambda)}\, d\gamma_\alpha(\mu,\lambda).$$

Proposition 2.1. (i) *For all nonnegative measurable functions g on Γ_+, we have*

$$\int_{\Gamma_+} g(\mu,\lambda)\, \mathrm{d}\gamma_\alpha(\mu,\lambda) = \frac{1}{2^\alpha \Gamma(\alpha+1)\sqrt{2\pi}} \bigg(\int_{\mathbb{R}} \int_0^\infty g(\mu,\lambda)(\mu^2+\lambda^2)^\alpha \mu \, \mathrm{d}\mu \, \mathrm{d}\lambda + \int_{\mathbb{R}} \int_0^{|\lambda|} g(i\mu,\lambda)(\lambda^2-\mu^2)^\alpha \mu \, \mathrm{d}\mu \, \mathrm{d}\lambda\bigg).$$

(ii) *For all measurable functions f on K, the function $\Theta \circ f$ is measurable on Γ_+. Furthermore, if f is a nonnegative or integrable function on K with respect to the measure $d\upsilon_\alpha$, then we have*

$$\int_{\Gamma_+} (\Theta \circ f)(\mu, \lambda)\, \mathrm{d}\gamma_\alpha(\mu, \lambda) = \int_0^\infty \int_{\mathbb{R}} f(r, x)\, \mathrm{d}\upsilon_\alpha(r, x). \tag{4}$$

Now we give the definition of the generalized Fourier transform associated with the Riemann–Liouville operator and some relevant properties.

Definition 2.2. For $f \in L^1(d\upsilon_\alpha)$, the Fourier transform $\mathscr{F}_\alpha$ associated with the Riemann–Liouville operator is defined by

$$\mathscr{F}_\alpha(f)(\mu, \lambda) = \int_K f(r, x)\varphi_{\mu,\lambda}(r, x)\, \mathrm{d}\upsilon_\alpha(r, x) \quad \text{for } (\mu, \lambda) \in \Gamma_+.$$

For this generalized Fourier transform, we have an inversion formula and an Plancherel theorem, just as with the classical Fourier transform in Euclidean space.

Theorem 2.3 (inversion formula). *Let $f \in L^1(\mathrm{d}\upsilon_\alpha)$ such that $\mathscr{F}_\alpha(f) \in L^1(\mathrm{d}\gamma_\alpha)$. Then for almost every $(r, x) \in K$, we have*

$$f(r, x) = \int_{\Gamma_+} \mathscr{F}_\alpha(f)(\mu,\ \lambda)\overline{\varphi_{\mu,\lambda}(r, x)}\, \mathrm{d}\gamma_\alpha(\mu, \lambda).$$

Theorem 2.4 (Plancherel). *The Fourier transform $\mathscr{F}_\alpha$ can be extended to an isomorphism from $L^2(\mathrm{d}\upsilon_\alpha)$ onto $L^2(\mathrm{d}\gamma_\alpha)$. In particular, for all $f, g \in L^2(\mathrm{d}\upsilon_\alpha)$, we have a version of Parseval's equality*:

$$\int_{\Gamma_+} \mathscr{F}_\alpha(f)(\mu, \lambda)\, \overline{\mathscr{F}_\alpha(g)(\mu, \lambda)}\, \mathrm{d}\gamma_\alpha(\mu, \lambda) = \int_K f(r, x)\overline{g(r, x)}\, \mathrm{d}\upsilon_\alpha(r, x).$$

The next two important lemmas will be used later in our proof.

Lemma 2.5. *For $m \in \mathbb{N}$, let*

$$\Phi_m(r) = \sqrt{\frac{2^{\alpha+1}\Gamma(\alpha+1)\, m!}{\Gamma(\alpha+m+1)}}\, e^{-r^2/2}\, L_m^\alpha(r^2).$$

The family $\{\Phi_m(r)\}_{m\in N}$ forms an orthonormal basis of the space

$$L^2(\mathbb{R}_+, r^{2\alpha+1}/(2^\alpha\Gamma(\alpha+1))\, \mathrm{d}r)$$

where $L_m^\alpha(x)$ is the Laguerre polynomial of degree m and order α defined by the expansion [Stempak 1988]

$$\sum_{n=0}^\infty t^n L_n^\alpha(x) = \frac{1}{(1-t)^{\alpha+1}} e^{xt/(t-1)}.$$

For the polynomial $L_m^\alpha(x)$, from [Huang and Liu 2007b], we also have the explicit expression for $L_m^\alpha(x)$:

$$L_m^\alpha(x) = \sum_{j=0}^{m} \frac{\Gamma(m+\alpha+1)}{\Gamma(m-j+1)\Gamma(j+\alpha+1)} \frac{(-x)^j}{j!}.$$

From the explicit expression of the Laguerre polynomial of degree m and order α, we know that there exists a function $M : \mathbb{N} \to \mathbb{R}_+$ such that for each $m \in \mathbb{N}$, we have $|\Phi_m(x)| \le M(m)$. The essence of this claim is that the polynomial doesn't grow as rapid as the exponential function when r approaches infinity.

Lemma 2.6 [Omri and Rachdi 2008, page 9]. *For all $m \in N$,*

$$\int_0^\infty e^{-r/2} L_m^\alpha(r) J_\alpha(\sqrt{ry}) r^{\alpha/2}\,\mathrm{d}r = (-1)^m 2\, e^{-y/2} y^{\alpha/2} L_m^\alpha(y).$$

We make the variable replacements $r = a^2$, $y = b^2$, but for simplicity we still use r and y instead of a, b. Then

$$\int_0^\infty e^{-r^2/2} L_m^\alpha(r^2) J_\alpha(ry) r^{\alpha+1}\,\mathrm{d}r = (-1)^m\, e^{-y^2/2} y^\alpha L_m^\alpha(y^2),$$

that is,

$$\int_0^\infty J_\alpha(ry)\, r^{\alpha+1}\, \Phi_m(r)\,\mathrm{d}r = (-1)^m y^\alpha\, \Phi_m(y). \tag{5}$$

3. Proof of the main result

In this section, we will prove Theorem 1.2. From the definition of the generalized Fourier transform, we know that

$$\mathscr{F}_\alpha(f)(\mu, \lambda) = \int_K f(r, x)\varphi_{\mu,\lambda}(r, x)\,\mathrm{d}\nu_\alpha(r, x).$$

Replace $\varphi_{\mu,\lambda}(r, x)$ by the expression in (2) to get

$$\mathscr{F}_\alpha(f)(\mu, \lambda) = \int_0^\infty \int_R f(r, x) j_\alpha(r\sqrt{\mu^2+\lambda^2})\, e^{-i\lambda x}\,\mathrm{d}x\,\mathrm{d}c(r)$$

If we let

$$\widetilde{\mathscr{F}_\alpha(f)}(\mu, \lambda) = \int_0^\infty \int_R f(r, x) j_\alpha(r\mu)\, e^{-i\lambda x}\,\mathrm{d}x\,\mathrm{d}c(r),$$

then $\mathscr{F}_\alpha(f)(\mu, \lambda) = (\Theta \circ \widetilde{\mathscr{F}_\alpha(f)})(\mu, \lambda)$. Thus our condition,

$$\int_K \int_{\Gamma_+} \frac{|f(r, x)|\,|\mathscr{F}_\alpha(f)(\mu, \lambda)|\, e^{|x||\lambda|}}{(1+|x|+|\lambda|)^N}\,\mathrm{d}\nu_\alpha(r, x)\ \mathrm{d}r_\alpha(\mu, \lambda) < \infty,$$

is equivalent to

$$\int_K \int_K \frac{|f(r,x)|\,|\widetilde{\mathscr{F}_\alpha(f)}(\mu,\lambda)|\,e^{|x||\lambda|}}{(1+|x|+|\lambda|)^N}\,\mathrm{d}\upsilon_\alpha(r,x)\,\mathrm{d}\upsilon_\alpha(\mu,\lambda)<\infty$$

by (4) (see Proposition 2.1). Defining

$$f^\lambda(r)=\int_{\mathbb{R}} f(r,x)\,e^{-i\lambda x}\,dx \quad \text{and} \quad f_m(x)=\int_0^\infty f(r,x)\Phi_m(r)\,\mathrm{d}c(r),$$

we obtain

$$\widehat{f_m}(\lambda)=\int_0^\infty f^\lambda(r)\Phi_m(r)\,\mathrm{d}c(r).$$

Before we proceed, we first prove the following useful formula:

$$\left|\int_0^\infty \widetilde{\mathscr{F}_\alpha(f)}(\mu,\lambda)\Phi_m(\mu)\,\mathrm{d}c(\mu)\right|=\frac{1}{\sqrt{2\pi}}\,|\widehat{f_m}(\lambda)|. \tag{6}$$

Indeed,

$$\begin{aligned}(7)\quad &\int_0^\infty \widetilde{\mathscr{F}_\alpha(f)}(\mu,\lambda)\Phi_m(\mu)\,\mathrm{d}c(\mu)\\ &=\int_0^\infty\int_0^\infty\int_{\mathbb{R}} f(r,x)e^{-i\lambda x}\,j_\alpha(r\mu)\Phi_m(\mu)\,\mathrm{d}x\,\mathrm{d}c(r)\,\mathrm{d}c(\mu)\\ &=2^\alpha\,\Gamma(\alpha+1)\int_0^\infty\int_0^\infty f^\lambda(r)\frac{J_\alpha(r\mu)}{(r\mu)^\alpha}\frac{\mu^{2\alpha+1}}{2^\alpha\Gamma(\alpha+1)\sqrt{2\pi}}\Phi_m(\mu)\,\mathrm{d}\mu\,\mathrm{d}c(r).\end{aligned}$$

By (5) (see Lemma 2.6), we know that the right-hand side equals

$$\frac{(-1)^m}{\sqrt{2\pi}}\int_0^\infty f^\lambda(r)\Phi_m(r)\,\mathrm{d}c(r)=\frac{(-1)^m}{\sqrt{2\pi}}\,\widehat{f_m}(\lambda),$$

which proves the claim.

We also need to prove the function $f(r,x)$ is in $L^1(\mathrm{d}\upsilon_\alpha)$. Since

$$\int_{\Gamma_+}\int_K \frac{|f(r,x)|\,|\mathscr{F}_\alpha(f)(\mu,\lambda)|\,e^{|x||\lambda|}}{(1+|x|+|\lambda|)^N}\,\mathrm{d}\upsilon_\alpha(r,x)\,\mathrm{d}r_\alpha(\mu,\lambda)<\infty, \tag{8}$$

there must exist a $\lambda_0\in\mathbb{R}$ such that

$$\int_K \frac{|f(r,x)|e^{|x||\lambda_0|}}{(1+|x|+|\lambda_0|)^N}\,\mathrm{d}\upsilon_\alpha(r,x)<+\infty.$$

Since there exists a constant $C>0$ such that $(1+|x|+|\lambda_0|)^N<Ce^{|x||\lambda_0|}$ for all $x\in\mathbb{R}$, we obtain

$$\int_K |f(r,x)|\,\mathrm{d}\upsilon_\alpha(r,x)<\frac{1}{C}\int_K \frac{|f(r,x)|e^{|x||\lambda_0|}}{(1+|x|+|\lambda_0|)^N}\,\mathrm{d}\upsilon_\alpha(r,x)<+\infty,$$

that is, $f(r,x)\in L^1(\mathrm{d}\upsilon_\alpha(r,x))$.

To proceed, we first prove that for any $m, n \in \mathbb{N}$,

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{|f_m(x)||\widehat{f_n}(\lambda)|e^{|x||\lambda|}}{(1+|x|+|\lambda|)^N}\,\mathrm{d}x\,\mathrm{d}\lambda < +\infty. \tag{9}$$

Since

$$|f_m(x)| = \left|\int_0^\infty f(r,x)\,\Phi_m(r)\,\mathrm{d}c(r)\right| \leqq M(m)\int_0^\infty |f(r,x)|\,\mathrm{d}c(r)$$

and

$$\begin{aligned}|\widehat{f_n}(\lambda)| &= \sqrt{2\pi}\left|\int_0^\infty \widetilde{\mathscr{F}_\alpha(f)}(\mu,\lambda)\Phi_m(\mu)\,\mathrm{d}c(\mu)\right|\\ &\leqq \sqrt{2\pi}\,M(n)\int_0^\infty |\widetilde{\mathscr{F}_\alpha(f)}(\mu,\lambda)|\,\mathrm{d}c(\mu),\end{aligned}$$

we have, for any $m, n \in \mathbb{N}$,

$$\begin{aligned}&\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{|f_m(x)||\widehat{f_n}(\lambda)|e^{|x|\,|\lambda|}}{(1+|x|+|\lambda|)^N}\,\mathrm{d}x\,\mathrm{d}\lambda\\ &\quad\leqq \sqrt{2\pi}M(m)M(n)\int_K\int_K \frac{|f(r,x)||\widetilde{\mathscr{F}_\alpha(f)}(\mu,\lambda)\,e^{|x|\,|\lambda|}}{(1+|x|+|\lambda|)^N}\,\mathrm{d}\upsilon_\alpha(r,x)\,\mathrm{d}\upsilon_\alpha(\mu,\lambda)\\ &\quad= \sqrt{2\pi}M(m)M(n)\int_K\int_{\Gamma_+} \frac{|f(r,x)||\mathscr{F}_\alpha(f)(\mu,\lambda)|\,e^{|x||\lambda|}}{(1+|x|+|\lambda|)^N}\,\mathrm{d}\upsilon_\alpha(r,x)\,\mathrm{d}\gamma_\alpha(\mu,\lambda)\\ &\quad< +\infty.\end{aligned}$$

In particular, setting $m = n$, we get

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{|f_m(x)||\widehat{f_m}(\lambda)|\,e^{|x||\lambda|}}{(1+|x|+|\lambda|)^N}\,\mathrm{d}x\,\mathrm{d}\lambda < +\infty.$$

Then by Lemma 1.1 (in this case $d = 1$), we have

$$f_m(x) = P_m(x)\,e^{-a_m x^2},$$

where a_m is positive and $P_m(x)$ is a polynomial with degree less than $(N-1)/2$. Further we claim that for all $m \in \mathbb{N}$, we have $a_m = a_n = a$. This holds since if there exist $m, n \in \mathbb{N}$ such that $a_m \neq a_n$, then the equation

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{|f_m(x)||\widehat{f_n}(\lambda)|e^{|x|\,|\lambda|}}{(1+|x|+|\lambda|)^N}\,\mathrm{d}x\,\mathrm{d}\lambda < +\infty$$

cannot hold, since it is in contradiction with the same equation derived by exchanging subscripts, which must be equally true. So, by Lemma 2.5,

$$f(r,x) = \sum_{j=0}^{\infty} f_m(x)\,\Phi_m(r) = e^{-ax^2}\left(\sum_{i=0}^{k}\psi_i(r)\,x^i\right),$$

where $k < \dfrac{N-1}{2}$ and

$$\psi_i(r) \in L^2\Big([0, +\infty), \frac{r^{2\alpha+1}}{2^\alpha \Gamma(\alpha+1)}\, \mathrm{d}r\Big).$$

Thus when $N < 3$ we have $f(r, x) = e^{-ax^2} \psi(r)$. In particular, when $N < 1$ we know that $f = 0$, since $f_m(x) = 0$ for each $m \in \mathbb{N}$. This finishes the proof of Theorem 1.2.

4. Some other versions of the uncertainty principle

We now derive other versions of the uncertainty principle as corollaries of our theorem. We start with a Gelfand–Shilov type uncertainty principle, which it is relatively straightforward to prove using Hölder's inequality and reduction to the absurd.

Theorem 4.1 (Gelfand–Shilov type). *Let $N \geqq 0$ and assume $f \in L^2(K, \mathrm{d}\upsilon_\alpha(r, x))$ satisfies*

$$\int_K \frac{|f(r, x)|\, e^{(a^p/p)|x|^p}}{(1+|x|)^N}\, \mathrm{d}\upsilon_\alpha(r, x) < +\infty,$$

$$\int_{\Gamma_+} \frac{|\mathcal{F}_\alpha(f)(\mu, \lambda)|\, e^{(b^q/q)|\lambda|^q}}{(1+|\lambda|)^N}\, \mathrm{d}\gamma_\alpha(\mu, \lambda) < +\infty,$$

where $1 < p, q < \infty$ satisfy $1/p + 1/q = 1$, and a, b are positive numbers such that $ab \geqq 1$. Then $f = 0$ unless $p = q = 2$, $ab = 1$ and $N > 0$, and in this case, we have

$$f(r, x) = e^{-ax^2}\Big(\sum_{j=0}^{m} \varphi_j(r)\, x^j\Big),$$

where $\varphi_j(r) \in L^2(\mathbb{R}_+, \mathrm{d}c(r))$ and $m \leq N - 1$. In particular, when $N \leqq 1$,

$$f(r, x) = e^{-(a^2/2)\, x^2}\, \psi(r),$$

where $\psi(r) \in L^2(\mathbb{R}_+, \mathrm{d}c(r))$, and when $N < 1$, we have $f = 0$.

Proof. Following the same procedure as in the proof of Theorem 1.2, we derive

$$\int_{\mathbb{R}} \frac{|f_m(x)|\, e^{(a^p/p)|x|^p}}{(1 + |x|)^N}\, \mathrm{d}x < \infty, \qquad \int_{\mathbb{R}} \frac{|\widehat{f_m}(\lambda)|\, e^{(b^q/q)|\lambda|^q}}{(1 + |\lambda|)^N}\, \mathrm{d}\lambda < \infty.$$

From Hölder's inequality, we have

$$a\,|x|\, b\,|\lambda| \leq \frac{a^p\, |x|^p}{p} + \frac{b^q\, |\lambda|^q}{q}.$$

Then

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{|f_m(x)|\,|\widehat{f_m}(\lambda)|\,e^{ab\,|x|\,|\lambda|}}{(1+|x|+|\lambda|)^{2N}}\,\mathrm{d}x\,\mathrm{d}\lambda \\ \le \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{|f_m(x)|\,e^{(a^p/p)|x|^p}}{(1+|x|)^N}\,\frac{|\widehat{f_m}(\lambda)|\,e^{(b^q/q)|\lambda|^q}}{(1+|\lambda|)^N}\,\mathrm{d}x\,\mathrm{d}\lambda < \infty.$$

So, when $ab > 1$, we could first derive the exact form of the function $f_m(x)$ from the Beurling theorem. We then know that with this form for $f_m(x)$, the inequality

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{|f_m(x)|\,|\widehat{f_m}(\lambda)|\,e^{ab\,|x|\,|\lambda|}}{(1+|x|+|\lambda|)^{2N}}\,\mathrm{d}x\,\mathrm{d}\lambda < \infty$$

cannot hold if $f_m(x) \neq 0$. When $ab = 1$ and either $p > 2$ or $q > 2$, also from the Beurling theorem, $f_m(x)$ is the product of polynomial and e^{-cx^2}. We deduce that the inequality

$$\int_{\mathbb{R}} \frac{|f_m(x)|\,e^{(a^p/p)|x|^p}}{(1+|x|)^N}\,\mathrm{d}x < \infty$$

cannot hold when $p > 2$ and the inequality

$$\int_{\mathbb{R}} \frac{|\widehat{f_m}(\lambda)|\,e^{(b^q/q)|\lambda|^q}}{(1+|\lambda|)^N}\,\mathrm{d}\lambda < \infty$$

cannot hold when $q > 2$, if $f_m(x) \neq 0$.

The conclusion in the last possible case, when $ab = 1$ and $p = q = 2$, can be derived from the Beurling theorem directly. □

Following the same idea as in Section 3, we can derive a Morgan-type theorem, which also gives a sharp lower bound for the Gelfand–Shilov type uncertainty principle:

Theorem 4.2. *Let $f \in L^2(K, \mathrm{d}\nu_\alpha(r,x))$ and suppose f satisfies*

$$\int_K |f(r,x)|\,e^{a^p|x|^p/p}\,\mathrm{d}\nu_\alpha(r,x) < \infty, \quad \int_{\Gamma_+} |\mathscr{F}_\alpha(f)(\mu,\lambda)|\,e^{b^q|\lambda|^q/q}\,\mathrm{d}\gamma_\alpha(\mu,\lambda) < \infty,$$

where $1 < p < 2$, $1/p + 1/q = 1$, and a, b are positive numbers. Then $f = 0$ if $ab > |\cos(p\pi/2)|^{1/p}$.

Proof. By the same argument as in the proof of our main theorem, we have

$$\int_{\mathbb{R}} |f_m(x)|\,e^{a^p|x|^p/p}\,\mathrm{d}x < \infty \quad \text{and} \quad \int_{\mathbb{R}} |\widehat{f_m}(\lambda)\,e^{b^q|\lambda|^q/q}\,\mathrm{d}\lambda < \infty.$$

Then [Bonami et al. 2003, Theorem 1.4], under the condition $ab > |\cos(p\pi/2)|^{1/p}$, implies that $f_m(x) = 0$ for each m, so we have $f(r,x) = 0$. □

Theorem 4.3 (Hardy type). *Suppose* $f \in L^2(K, \mathrm{d}\upsilon_\alpha(r,x))$ *satisfies*

$$|f(r,x)| \leqq C_1 e^{-a(r^2+x^2)} \quad \textit{and} \quad |\mathcal{F}_\alpha(f)(\mu,\lambda)| \leqq C_2 e^{-b(\mu^2+\lambda^2)},$$

where C_1, C_2 *are positive constants and* a, b *are positive real numbers such that* $ab \geqq \frac{1}{4}$. *If* $ab > \frac{1}{4}$, *then* $f = 0$. *If* $ab = \frac{1}{4}$, *then*

$$f(r,x) = e^{-ax^2}\psi(r),$$

where $\psi(r) \in L^2(\mathbb{R}_+, \mathrm{d}c(r))$.

Proof. To prove this corollary, we recall the well-known classical Hardy's theorem for the classical Fourier transform on $\mathbb{R}$ which says that if

$$|f(x)| \leqq Ce^{-ax^2} \quad \text{and} \quad \widehat{f}(\lambda) \leqq Ce^{-b\lambda^2},$$

where $\widehat{f}$ is the Fourier transform of f, then

(i) $f = 0$ when $ab > \frac{1}{4}$;

(ii) $f(x) = ce^{-ax^2}$ when $ab = \frac{1}{4}$;

(iii) there are infinitely many linearly independent functions satisfying the above conditions when $ab < \frac{1}{4}$.

From the conditions in the corollary and using the same method used in Section 3, we have

$$|f_m(x)| \leqq C\, e^{-ax^2} \quad \text{and} \quad |\widehat{f_m}(\lambda)| \leqq C\, e^{-b\lambda^2}.$$

So from the classical Hardy's theorem, we have $f_m(x) = c_m\, e^{-ax^2}$ if $ab = \frac{1}{4}$ for each $m \in N$. Then

$$f(r,x) = e^{-ax^2}\left(\sum_{m=0}^{\infty} c_m\, \Phi_m(r)\right) = e^{-ax^2}\, \psi(r),$$

where $\psi(r) \in L^2(\mathbb{R}_+, \mathrm{d}\,c(r))$. When $ab > \frac{1}{4}$, each $f_m(x)$ vanishes, so we have $f(r,x) = 0$. □

Theorem 4.4 (Morgan type). *Suppose* $f \in L^2(K, \mathrm{d}\upsilon_\alpha(r,x))$ *satisfies*

$$\int_0^\infty |f(r,x)|\, r^{2\alpha+1}\,\mathrm{d}r \leqq C_1 e^{-a|x|^p}, \quad \int_0^\infty |\widetilde{\mathcal{F}_\alpha(f)}(\mu,\lambda)|\, \mu^{2\alpha+1}\,\mathrm{d}\mu \leqq C_2 e^{-b|\lambda|^q},$$

where C_1, C_2 *are positive constants*, $1 < p < 2$, $1/p + 1/q = 1$, *and* a, b *are positive numbers. Then* $f = 0$ *if* $(a\,p)^{1/p}(b\,q)^{1/q} > |\cos(p\pi/2)|^{1/p}$.

Proof. First let $a = \alpha^p/p$ and $b = \beta^q/q$. Then

$$\alpha\,\beta > |\cos(p\,\pi/2)|^{1/p}.$$

There exists an $\epsilon > 0$, such that $(\alpha - \epsilon)\,(\beta - \epsilon) > |\cos(p\,\pi/2)|^{1/p}$ also holds. Then

$$\int_{\mathbb{R}} |f_m(x)|\, e^{(\alpha-\epsilon)^p |x|^p/p}\,\mathrm{d}x < M(m) \int_{\mathbb{R}} e^{-(\alpha^p-(\alpha-\epsilon)^p)/p\,|x|^p}\,\mathrm{d}x < \infty,$$

$$\int_{\mathbb{R}} |\widehat{f_m}(\lambda)\, e^{(\beta-\epsilon)^q |\lambda|^q/q}\,\mathrm{d}\lambda < M(m) \int_{\mathbb{R}} e^{-(\beta^q-(\beta-\epsilon)^q)/q\,|\lambda|^q}\,\mathrm{d}\lambda < \infty.$$

By [Bonami et al. 2003, Theorem 1.4], we have $f_m(x) = 0$ for each $m \in \mathbb{N}$, so $f = 0$. □

5. More on this topic

We now derive a sharper result than the main theorem, requiring an additional constraint on the function $f(r, x)$.

First we introduce some related notation and propositions about the dual of the Riemann–Liouville operator. For more details, refer to [Baccar et al. 2006]. Let $\mathscr{C}_*(\mathbb{R}^2)$ be the function space of continuous functions on $\mathbb{R}^2$ even with respect to the first variable, and $\mathscr{S}_*(\mathbb{R}^2)$ the space of infinitely differentiable functions on $\mathbb{R}^2$, rapidly decreasing together with all their derivatives even with respect to the first variable. The dual Riemann–Liouville operator (or transform) is defined by

$$\int_0^\infty \int_{\mathbb{R}} \mathscr{R}_\alpha(f)(r,x) g(r,x)\,\mathrm{d}x\, r^{2\alpha+1}\,\mathrm{d}r = \int_0^\infty \int_{\mathbb{R}} f(r,x)\, {}^t\mathscr{R}_\alpha(g)(r,x)\,\mathrm{d}x\; r^{2\alpha+1}\,\mathrm{d}r,$$

where $f \in \mathscr{C}_*(\mathbb{R}^2)$ and $g \in \mathscr{S}_*(\mathbb{R}^2)$. This is also why ${}^t\mathscr{R}_\alpha$ called the "dual". We also have for $f \in \mathscr{S}_*(\mathbb{R}^2)$,

$${}^t\mathscr{R}_\alpha(f)(r,x) = \begin{cases} \dfrac{2\alpha}{\pi} \displaystyle\int_r^\infty \int_{-\sqrt{\mu^2-r^2}}^{\sqrt{\mu^2-r^2}} f(u, x+v)(\mu^2 - v^2 - r^2)^{\alpha-1}\,\mathrm{d}v\,\mu\,\mathrm{d}\mu & \text{if } \alpha > 0, \\ \dfrac{1}{\pi} \displaystyle\int_{\mathbb{R}} f\big(\sqrt{r^2 + (x-y)^2}, y\big)\,\mathrm{d}y & \text{if } \alpha = 0. \end{cases}$$

Some propositions related to the dual Riemann–Liouville transform are needed before going to our main result in this section.

Lemma 5.1 [Baccar et al. 2006, Lemma 3.6, page 9]. *For $f \in \mathscr{S}_*(\mathbb{R}^2)$,*

$$\mathscr{F}_\alpha(f)(\mu.\lambda) = \wedge_\alpha \circ {}^t\mathscr{R}_\alpha(f)\,(\mu,\lambda) \quad \textit{for } (\mu,\lambda) \in \mathbb{R}^2,$$

where $\wedge_\alpha$ is a constant multiple of the classical Fourier transform on $\mathbb{R}^2$ defined by

$$\wedge_\alpha(f)(\mu,\lambda) = \int_0^\infty \int_{\mathbb{R}} f(r,x)\,\cos(r\mu)\exp(-i\lambda\,x)\,\frac{1}{\sqrt{2\pi}\,2^\alpha\Gamma(\alpha+1)}\,\mathrm{d}x\,\mathrm{d}r.$$

Lemma 5.2 [Baccar et al. 2006, Proposition 3.7]. (i) *${}^t\mathscr{R}_\alpha$ is not injective when applied to $\mathscr{S}_*(\mathbb{R}^2)$.*

(ii) ${}^t\mathcal{R}_\alpha(\mathcal{S}_*(R^2)) = \mathcal{S}_*(\mathbb{R}^2)$.

To proceed, we still need to define two special subspaces of $\mathcal{S}_*(\mathbb{R}^2)$. Denote by $\mathcal{S}^0_*(\mathbb{R}^2)$ the subspace of $\mathcal{S}_*(\mathbb{R}^2)$ consisting of functions f such that

$$\operatorname{supp} \widetilde{\mathcal{F}_\alpha(f)} \subset \{(\mu, \lambda) \in \mathbb{R}^2 : |\mu| \geqq |\lambda|\}.$$

Denote by $\mathcal{S}_{*,0}(\mathbb{R}^2)$ the subspace of $\mathcal{S}_*(\mathbb{R}^2)$ consisting of functions f such that

$$\int_0^\infty f(r, x)\, r^{2k}\, \mathrm{d}r = 0 \quad \text{for all } k \in N \text{ and } x \in \mathbb{R}.$$

From Lemma 5.2, we know that ${}^t\mathcal{R}_\alpha$ is not a isomorphism between $\mathcal{S}_*(\mathbb{R}^2)$ and $\mathcal{S}_*(\mathbb{R}^2)$. But things are different on the subspace $\mathcal{S}^0_*(\mathbb{R}^2)$. We have the isomorphism lemma as well as inversion formula for the operator ${}^t\mathcal{R}_\alpha$.

Lemma 5.3. *The dual transform ${}^t\mathcal{R}_\alpha$ is an isomorphism from $\mathcal{S}^0_*(\mathbb{R}^2)$ onto $\mathcal{S}_{*,0}(\mathbb{R}^2)$.*

Lemma 5.4 [Baccar et al. 2006, Theorems 4.5 and 4.6]. *For $g \in \mathcal{S}_{*,0}(\mathbb{R}^2)$ the inversion formula*

$$({}^t\mathcal{R}_\alpha)^{-1}(g) = (K_\alpha^2 \circ \mathcal{R}_\alpha)(g)$$

holds for ${}^t\mathcal{R}_\alpha$, where $\mathcal{R}_\alpha$ is the Riemann–Liouville operator defined in Section 1 and the operator K_α^2 is defined by

$$K_\alpha^2(g)(r, x) = \mathcal{F}_\alpha^{-1}\left(\frac{\pi}{2^{2\alpha+1}\Gamma^2(\alpha+1)}(\mu^2+\lambda^2)^\alpha |\mu| \mathcal{F}_\alpha(g)\right)(r, x).$$

Also K_α^2 is an isomorphism from $\mathcal{S}^0_*(\mathbb{R}^2)$ onto itself.

With the help of these lemmas, we derive our new analogue:

Theorem 5.5. *Suppose $f \in \mathcal{S}^0_*(\mathbb{R}^2)$ satisfies*

$$\int_K \int_{\Gamma_+} \frac{|f(r, x)|\, |\mathcal{F}_\alpha(f)(\mu, \lambda)|\, e^{\|(r,x)\|\|(\mu,\lambda)\|}\ \Xi(\mu, \lambda)}{(1 + \|(r, x)\| + \|(\mu, \lambda)\|)^N}\, \mathrm{d}\gamma_\alpha(\mu, \lambda)\, \mathrm{d}\upsilon_\alpha(r, x) < \infty.$$

Then

$$f(r, x) = ({}^t\mathcal{R}_\alpha)^{-1}\big(P(y)\, e^{-\langle Ay, y\rangle}\big),$$

where $y = (r, x)$, $P(y)$ is a polynomial with degree less than $(N-2)/2$, A is a real positive definite symmetric 2×2 matrix, $\|\cdot\|$ is the usual norm in $\mathbb{C}^n$, and $\Xi(\mu, \lambda)$ is defined by

$$\Xi(\mu, \lambda) = \frac{1}{(\mu^2+\lambda^2)^\alpha |\mu|}.$$

In particular, when $N \leqq 2$, we have $f = 0$.

Proof. We first prove that for all $(\mu, \lambda) \in \mathbb{R}^2$, there exists $C > 0$ such that

$$\int_K \frac{|{}^t\mathcal{R}_\alpha(f)(r,x)|\,|\mathcal{F}_\alpha(f)(\mu,\lambda)|\,e^{\|(r,x)\|\|(\mu,\lambda\|)}}{(1+\|(r,x)\|+\|(\mu,\lambda)\|)^N}\,\mathrm{d}r\,\mathrm{d}x \leqq C\int_K \frac{|f(r,x)||\mathcal{F}_\alpha(f)(\mu,\lambda)\,e^{\|(r,x)\|\,\|(\mu,\lambda)\|}}{(1+\|(r,x)\|+\|(\mu,\lambda)\|)^N}\,\mathrm{d}\upsilon_\alpha(r,x).$$

We first consider the case when $\alpha > 0$; then

$${}^t\mathcal{R}_\alpha(f)(r,x) = \frac{2\alpha}{\pi}\int_r^\infty \int_{-\sqrt{\mu^2-r^2}}^{\sqrt{\mu^2-r^2}} f(\mu, x+v)(\mu^2-v^2-r^2)^{\alpha-1}\,\mathrm{d}v\,\mu\,\mathrm{d}\mu.$$

So we have

$$\begin{aligned}
&\int_K \frac{|{}^t\mathcal{R}_\alpha(f)(r,x)|e^{\|(r,x)\|\|(\mu,\lambda)\|}}{(1+\|(r,x)\|+\|(\mu,\lambda)\|)^N}\mathrm{d}r\,\mathrm{d}x\\
&=\frac{2\alpha}{\pi}\int_0^\infty\int_{\mathbb{R}} \frac{\left|\int_r^\infty\int_{\sqrt{\mu^2-r^2}}^{\sqrt{\mu^2-r^2}} f(\mu,x+v)(\mu^2-v^2-r^2)^{\alpha-1}\mathrm{d}v\,\mu\,\mathrm{d}\mu\right|e^{\|(r,x)\|\|(\mu,\lambda)\|}}{(1+\|(r,x)\|+\|(\mu,\lambda)\|)^N}\mathrm{d}x\,\mathrm{d}r\\
&=\frac{2\alpha}{\pi}\int_0^\infty\int_{\mathbb{R}}\int_r^\infty\int_{\sqrt{\mu^2-r^2}}^{\sqrt{\mu^2-r^2}} \frac{|f(\mu,x+v)|(\mu^2-v^2-r^2)^{\alpha-1}e^{\|(r,x)\|\|(\mu,\lambda)\|}}{(1+\|(r,x)\|+\|(\mu,\lambda)\|)^N}\mathrm{d}v\,\mu\,\mathrm{d}\mu\,\mathrm{d}x\,\mathrm{d}r.
\end{aligned}$$

Changing variables, let $\mu = \mu$, $b = x + v$, $r = r$, $x = x$. For simplicity we will still use v instead of b. Then by a change of variables and integration, we see that the right-hand side above is bounded above by

$$\leqq C_1\int_0^\infty\int_{\mathbb{R}} \frac{|f(r,x)|\,e^{\|(r,x)\|\|(\mu,\lambda)\|}}{(1+\|(r,x)\|+\|(\mu,\lambda)\|)^N}r^{2\alpha+1}\,\mathrm{d}r\,\mathrm{d}x \leqq C_2\,e\int_K \frac{|f(r,x)|\,e^{\|(r,x)\|\|(\mu,\lambda)\|}}{(1+\|(r,x)\|+\|(\mu,\lambda)\|)^N}\,\mathrm{d}\upsilon_\alpha(r,x).$$

For the case $\alpha = 0$, our previous claim also holds by using the same method as in the case $\alpha > 0$, using a different variable replacement by letting $a = \sqrt{r^2 + (x-y)^2}$, $y = y$, and for simplicity still using r instead of a. This proves our claim.

By Proposition 2.1(i), and restricting the integral region Γ_+ to K, we derive the inequality

$$\begin{aligned}
&\int_K\int_K \frac{|{}^t\mathcal{R}_\alpha(f)(r,x)|\,|\mathcal{F}_\alpha(f)(\mu,\lambda)|\,e^{\|(r,x)\|\|(\mu,\lambda\|)}}{(1+\|(r,x)\|+\|(\mu,\lambda)\|)^N}\,\mathrm{d}r\,\mathrm{d}x\,\mathrm{d}\mu\,\mathrm{d}\lambda\\
&\leqq C\times\int_K\int_{\Gamma_+} \frac{|f(r,x)|\,|\mathcal{F}_\alpha(f)(\mu,\lambda)|\,e^{\|(r,x)\|\|(\mu,\lambda\|)}\,\Xi(\mu,\lambda)}{(1+\|(r,x)\|+\|(\mu,\lambda)\|)^N}\,\mathrm{d}\upsilon_\alpha(r,x)\,\mathrm{d}\gamma_\alpha(\mu,\lambda)\\
&< \infty.
\end{aligned}$$

By Lemma 5.1 we know that the above inequality satisfies the conditions of the Beurling theorem (Lemma 1.1) in 2-dimensional Euclidean space. So

$$ {}^t\mathcal{R}_\alpha(f)(r,x) = P(y)\, e^{-\langle Ay,y\rangle}, $$

where $y=(r,x)$, $P(y)$ is a polynomial such that its degree is less than $(N-2)/2$, and A is a positive definite symmetric 2×2 matrix. From $f\in \mathcal{S}^0_*(\mathbb{R}^2)$ and Lemma 5.3 we know that $P(y)\, e^{-\langle Ay,y\rangle}\in \mathcal{S}_{*,0}(\mathbb{R}^2)$ and

$$ f(r,x) = ({}^t\mathcal{R}_\alpha)^{-1}\bigl(P(y)\, e^{-\langle Ay,y\rangle}\bigr). $$

In particular, if $N\leqq 2$, we have

$$ {}^t\mathcal{R}_\alpha(f)(r,x)=0, $$

which implies $f(r,x)=0$ so our proof is finished. □

Remark. In this section, we gave another analogue of the Beurling–Hörmander theorem. When compared with Theorem 1.2, which just gives the precise structure of x but not r since we only know that $\psi_j(r)\in L^2(\mathbb{R}_+, \mathrm{d}c(r))$, the new analogue derived in this section gives the precise structure of both r and x. However, this requires the additional condition that $f\in\mathcal{S}^0_*(\mathbb{R}^2)$ and it's difficult to remove this condition because the dual Riemann–Liouville transform is not injective on the full space $\mathcal{S}_*(\mathbb{R}^2)$. To conquer this difficulty, a different method might be needed.

References

[Baccar et al. 2006] C. Baccar, N. B. Hamadi, and L. T. Rachdi, "Inversion formulas for Riemann–Liouville transform and its dual associated with singular partial differential operators", *Int. J. Math. Math. Sci.* **2006** (2006), Art. ID 86238, 26. MR 2008a:44005 Zbl 1131.44002

[Bonami et al. 2003] A. Bonami, B. Demange, and P. Jaming, "Hermite functions and uncertainty principles for the Fourier and the windowed Fourier transforms", *Rev. Mat. Iberoamericana* **19**:1 (2003), 23–55. MR 2004f:42015 Zbl 1037.42010

[Bouattour 2008] L. Bouattour, "Beurling–Hörmander theorem on noncompact real symmetric spaces", *Commun. Math. Anal.* **4**:1 (2008), 20–34. MR 2009d:43014 Zbl 1172.43003

[Bouattour and Trimèche 2005] L. Bouattour and K. Trimèche, "Beurling–Hörmander's theorem for the Chébli–Trimèche transform", *Glob. J. Pure Appl. Math.* **1**:3 (2005), 342–357. MR 2007h:43002 Zbl 1105.43004

[Hamadi and Rachdi 2006] N. B. Hamadi and L. T. Rachdi, "Weyl transforms associated with the Riemann–Liouville operator", *Int. J. Math. Math. Sci.* **2006** (2006), Art. 94768. MR 2007g:35283 Zbl 1146.35426

[Hörmander 1991] L. Hörmander, "A uniqueness theorem of Beurling for Fourier transform pairs", *Ark. Mat.* **29**:2 (1991), 237–240. MR 93b:42016 Zbl 0755.42009

[Huang and Liu 2007a] J. Huang and H. Liu, "An analogue of Beurling's theorem for the Heisenberg group", *Bull. Austral. Math. Soc.* **76**:3 (2007), 471–478. MR 2009a:43008 Zbl 1154.43001

[Huang and Liu 2007b] J. Huang and H. Liu, "An analogue of Beurling's theorem for the Laguerre hypergroup", *J. Math. Anal. Appl.* **336**:2 (2007), 1406–1413. MR 2009a:43007 Zbl 1132.43003

[Kamoun and Trimèche 2005] L. Kamoun and K. Trimèche, "An analogue of Beurling–Hörmander's theorem associated with partial differential operators", *Mediterr. J. Math.* **2**:3 (2005), 243–258. MR 2006j:43010

[Omri and Rachdi 2008] S. Omri and L. T. Rachdi, "Heisenberg–Pauli–Weyl uncertainty principle for the Riemann–Liouville operator", *JIPAM. J. Inequal. Pure Appl. Math.* **9**:3 (2008), Article 88, 23. MR 2009k:42021 Zbl 1159.42305

[Parui and Sarkar 2008] S. Parui and R. P. Sarkar, "Beurling's theorem and L^p-L^q Morgan's theorem for step two nilpotent Lie groups", *Publ. Res. Inst. Math. Sci.* **44**:4 (2008), 1027–1056. MR 2009j:22014 Zbl 05543221

[Sarkar and Sengupta 2007a] R. P. Sarkar and J. Sengupta, "Beurling's theorem and characterization of heat kernel for Riemannian symmetric spaces of noncompact type", *Canad. Math. Bull.* **50**:2 (2007), 291–312. MR 2008e:43014 Zbl 1134.22007

[Sarkar and Sengupta 2007b] R. P. Sarkar and J. Sengupta, "Beurling's theorem for SL(2, $\mathbb{R}$)", *Manuscripta Math.* **123**:1 (2007), 25–36. MR 2008m:43016 Zbl 1117.43008

[Stempak 1988] K. Stempak, "An algebra associated with the generalized sub-Laplacian", *Studia Math.* **88**:3 (1988), 245–256. MR 89e:47073 Zbl 0672.46025

[Thangavelu 2004] S. Thangavelu, "On theorems of Hardy, Gelfand–Shilov and Beurling for semisimple groups", *Publ. Res. Inst. Math. Sci.* **40** (2004), 311–344. MR 2005e:43014 Zbl 1050.22014

Received February 23, 2010. Revised October 12, 2010.

XUECHENG WANG
CHINA ECONOMICS AND MANAGEMENT ACADEMY
CENTRAL UNIVERSITY OF FINANCE AND ECONOMICS
XUEYUAN SOUTH ROAD #39, HAIDIAN DISTRICT
BEIJING 100081
CHINA

xchwhades@139.com

Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/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 251 No. 1 May 2011

0030-8730(201105)251:1;1-E