Pacific Journal of Mathematics

Volume 259 No. 2 October 2012

PACIFIC JOURNAL OF MATHEMATICS

http://pacificmath.org

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

EDITORS

V. S. Varadarajan (Managing Editor)
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
pacific@math.ucla.edu

Vyjayanthi Chari
Department of Mathematics
University of California
Riverside, CA 92521-0135
chari@math.ucr.edu

Darren Long
Department of Mathematics
University of California
Santa Barbara, CA 93106-3080
long@math.ucsb.edu

Sorin Popa
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
popa@math.ucla.edu

Robert Finn
Department of Mathematics
Stanford University
Stanford, CA 94305-2125
finn@math.stanford.edu

Jiang-Hua Lu
Department of Mathematics
The University of Hong Kong
Pokfulam Rd., Hong Kong
jhlu@maths.hku.hk

Jie Qing
Department of Mathematics
University of California
Santa Cruz, CA 95064
qing@cats.ucsc.edu

Kefeng Liu
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
liu@math.ucla.edu

Alexander Merkurjev
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
merkurev@math.ucla.edu

Jonathan Rogawski
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
jonr@math.ucla.edu

PRODUCTION

pacific@math.berkeley.edu

Silvio Levy, Scientific Editor   Matthew Cargo, Senior Production Editor

SUPPORTING INSTITUTIONS

ACADEMIA SINICA, TAIPEI
CALIFORNIA INST. OF TECHNOLOGY
INST. DE MATEMÁTICA PURA E APLICADA
KEIO UNIVERSITY
MATH. SCIENCES RESEARCH INSTITUTE
NEW MEXICO STATE UNIV.
OREGON STATE UNIV.
STANFORD UNIVERSITY
UNIV. OF BRITISH COLUMBIA
UNIV. OF CALIFORNIA, BERKELEY
UNIV. OF CALIFORNIA, DAVIS
UNIV. OF CALIFORNIA, LOS ANGELES
UNIV. OF CALIFORNIA, RIVERSIDE
UNIV. OF CALIFORNIA, SAN DIEGO
UNIV. OF CALIF., SANTA BARBARA
UNIV. OF CALIF., SANTA CRUZ
UNIV. OF MONTANA
UNIV. OF OREGON
UNIV. OF SOUTHERN CALIFORNIA
UNIV. OF UTAH
UNIV. OF WASHINGTON
WASHINGTON STATE UNIVERSITY

These supporting institutions contribute to the cost of publication of this Journal, but they are not owners or publishers and have no responsibility for its contents or policies.

See inside back cover or pacificmath.org for submission instructions.

The subscription price for 2012 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

Copyright ©2012 by Pacific Journal of Mathematics

PACIFIC JOURNAL OF MATHEMATICS
Vol. 259, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.259.257

FLAG SUBDIVISIONS AND γ-VECTORS

CHRISTOS A. ATHANASIADIS

The γ-vector is an important enumerative invariant of a flag simplicial homology sphere. It has been conjectured by Gal that this vector is nonnegative for every such sphere Δ and by Reiner, Postnikov and Williams that it increases when Δ is replaced by any flag simplicial homology sphere that geometrically subdivides Δ. Using the nonnegativity of the γ-vector in dimension 3, proved by Davis and Okun, as well as Stanley's theory of simplicial subdivisions and local h-vectors, the latter conjecture is confirmed in this paper in dimensions 3 and 4.

1. Introduction

This paper is concerned with the face enumeration of an important class of simplicial complexes, that of flag homology spheres, and their subdivisions. The face vector of a homology sphere (more generally, of an Eulerian simplicial complex) Δ can be conveniently encoded by its γ-vector [Gal 2005], denoted by $\gamma(\Delta)$. Part of our motivation comes from the following two conjectures. (We refer to Section 2 for all relevant definitions.) The first, proposed by Gal [2005, Conjecture 2.1.7], can be thought of as a generalized lower-bound conjecture for flag homology spheres; it strengthens an earlier conjecture by Charney and Davis [1995]. The second, proposed by Postnikov, Reiner and Williams [Postnikov et al. 2008, Conjecture 14.2], is a natural extension of the first.

Conjecture 1.1 [Gal 2005]. *For every flag homology sphere Δ we have $\gamma(\Delta) \geq 0$.*

Conjecture 1.2 [Postnikov et al. 2008]. *For all flag homology spheres Δ and Δ' for which Δ' geometrically subdivides Δ, we have $\gamma(\Delta') \geq \gamma(\Delta)$.*

These statements are trivial for spheres of dimension 2 or less. Conjecture 1.1 was proved for 3-dimensional spheres by Davis and Okun [2001, Theorem 11.2.1] and was deduced from that result for 4-dimensional spheres in [Gal 2005, Corollary 2.2.3]. Conjecture 1.2 can be thought of as a conjectural analogue of the fact [Stanley 1992, Theorem 4.10] that the h-vector (a certain linear transformation of the face

MSC2010: primary 05E45; secondary 05E99.

Keywords: flag complex, homology sphere, simplicial subdivision, flag subdivision, face enumeration, γ-vector.

vector) of a Cohen–Macaulay simplicial complex increases under quasigeometric simplicial subdivision (a class of topological subdivisions that includes all geometric simplicial subdivisions). The main result of this paper proves its validity in three and four dimensions for a new class of simplicial subdivisions, which includes all geometric ones.

Theorem 1.3. *For every flag homology sphere Δ of dimension 3 or 4 and for every flag vertex-induced homology subdivision Δ' of Δ, we have $\gamma(\Delta') \geq \gamma(\Delta)$.*

This result naturally suggests the following stronger version of Conjecture 1.2:

Conjecture 1.4. *For every flag homology sphere Δ and every flag vertex-induced homology subdivision Δ' of Δ, we have $\gamma(\Delta') \geq \gamma(\Delta)$.*

The following structural result on flag homology spheres, which may be of independent interest, will also be proved in Section 4. It implies, for instance, that Conjecture 1.4 is stronger than Conjecture 1.1. Throughout this paper, we will denote by Σ_{d-1} the boundary complex of the d-dimensional cross-polytope (equivalently, the simplicial join of d copies of the 0-dimensional sphere).

Theorem 1.5. *Every flag $(d-1)$-dimensional homology sphere is a vertex-induced (hence quasigeometric and flag) homology subdivision of Σ_{d-1}.*

The proof of Theorem 1.3 relies on the theory of face enumeration for simplicial subdivisions, developed by Stanley [1992]. Given a simplicial complex Δ and a simplicial subdivision Δ' of Δ, the h-vector of Δ' can be expressed in terms of local contributions, one for each face of Δ, and the combinatorics of Δ [Stanley 1992, Theorem 3.2]. The local contributions are expressed in terms of the key concept of a local h-vector, introduced and studied in [Stanley 1992]. When Δ is Eulerian, this formula transforms into one involving γ-vectors (Proposition 5.3) and leads to the concept of a local γ-vector, introduced in Section 5. Using the Davis–Okun theorem [Davis and Okun 2001] mentioned earlier, it is shown that the local γ-vector has nonnegative coefficients for every flag vertex-induced homology subdivision of the 3-dimensional simplex. Theorem 1.3 is deduced from these results in Section 5.

The proof of Theorem 1.5 is motivated by that of [Athanasiadis 2011, Theorem 1.2], stating that the graph of any flag simplicial pseudomanifold of dimension $d-1$ contains a subdivision of the graph of Σ_{d-1}.

We now briefly describe the content and structure of this paper. Sections 2 and 3 provide the necessary background on simplicial complexes, subdivisions and their face enumeration. The notion of a homology subdivision, which is convenient for the results of this paper as well as those of a flag subdivision and vertex-induced (a natural strengthening of quasigeometric) subdivision, are introduced in Section 2C.

Section 3C includes a simple example (see Example 3.4) that shows that there exist quasigeometric subdivisions of the simplex with nonunimodal local h-vector.

Section 4 proves Theorem 1.5 and another structural result on flag subdivisions (Proposition 4.6), stating that every flag vertex-induced homology subdivision of the $(d-1)$-dimensional simplex naturally occurs as a restriction of a flag vertex-induced homology subdivision of Σ_{d-1}. These results are used in Section 5.

Local γ-vectors are introduced in Section 5, where examples and elementary properties are discussed. It is conjectured there that the local γ-vector has nonnegative coordinates for every flag vertex-induced homology subdivision of the simplex (Conjecture 5.4). This statement can be considered as a local analogue of Conjecture 1.1. It is shown to imply both Conjectures 1.1 and 1.4 and to hold in dimension 3. Section 5 concludes with the proof of Theorem 1.3.

Section 6 discusses some special cases of Conjecture 5.4. For instance, the conjecture is shown to hold for iterated edge subdivisions (in the sense of [Charney and Davis 1995, Section 5.3]) of the simplex.

2. Flag complexes, subdivisions and γ-vectors

This section reviews background material on simplicial complexes, in particular on their homological properties and subdivisions. For more information on these topics, the reader is referred to [Stanley 1996]. Throughout this paper, $\Bbbk$ will be a field that we will assume to be fixed. We will denote by $|S|$ the cardinality, and by 2^S the set of all subsets, of a finite set S.

2A. ***Simplicial complexes.*** All simplicial complexes we consider will be abstract and finite. Thus, given a finite set Ω, a *simplicial complex* on the ground set Ω is a collection Δ of subsets of Ω such that $F \subseteq G \in \Delta$ implies $F \in \Delta$. The elements of Δ are called *faces*. The dimension of a face F is defined as one less than the cardinality of F. The dimension of Δ is the maximum dimension of a face and is denoted by $\dim(\Delta)$. Faces of Δ of dimension 0 or 1 are called *vertices* or *edges*, respectively. A *facet* of Δ is a face that is maximal with respect to inclusion. The complex Δ is called *pure* if all its facets have the same dimension. All topological properties of Δ we mention in the sequel will refer to those of the geometric realization $\|\Delta\|$ of Δ [Björner 1995, Section 9], uniquely defined up to homeomorphism. For example, we say that Δ is a simplicial or topological ball or sphere if $\|\Delta\|$ is homeomorphic to a ball or sphere, respectively.

The *open star* of a face $F \in \Delta$, denoted by $\mathrm{st}_\Delta(F)$, is the collection of all faces of Δ that contain F. The *closed star* of $F \in \Delta$, denoted by $\overline{\mathrm{st}}_\Delta(F)$, is the subcomplex of Δ consisting of all subsets of the elements of $\mathrm{st}_\Delta(F)$. The *link* of the face $F \in \Delta$ is the subcomplex of Δ defined as $\mathrm{link}_\Delta(F) = \{G \smallsetminus F : G \in \Delta, F \subseteq G\}$. The *simplicial join* $\Delta_1 * \Delta_2$ of two collections Δ_1 and Δ_2 of subsets of disjoint ground sets is the collection whose elements are the sets of the form $F_1 \cup F_2$, where $F_1 \in \Delta_1$ and $F_2 \in \Delta_2$. If Δ_1 and Δ_2 are simplicial complexes, then so is $\Delta_1 * \Delta_2$. The simplicial join of Δ with the zero-dimensional complex $\{\varnothing, \{v\}\}$ is denoted by $v * \Delta$ and called the *cone over* Δ on the (new) vertex v.

A simplicial complex Δ is called *flag* if every minimal nonface of Δ has two elements. The closed star, the link of any face of a flag complex and the simplicial join of two (or more) flag complexes are also flag complexes. In particular, the simplicial join of d copies of the zero-dimensional complex with two vertices is a flag complex (in fact, a flag triangulation of the $(d-1)$-dimensional sphere), which will be denoted by Σ_{d-1}. Explicitly, Σ_{d-1} can be described as the simplicial complex on the $2d$-element ground set $\Omega_d = \{u_1, u_2, \ldots, u_d\} \cup \{v_1, v_2, \ldots, v_d\}$ whose faces are those subsets of Ω_d that contain at most one element from each of the sets $\{u_i, v_i\}$ for $1 \le i \le d$.

2B. *Homology balls and spheres.* Let Δ be a simplicial complex of dimension $d-1$. We call Δ a *homology sphere* (over $\Bbbk$) if for every $F \in \Delta$ (including $F = \varnothing$) we have

$$\widetilde{H}_i(\mathrm{link}_\Delta(F), \Bbbk) = \begin{cases} \Bbbk & \text{if } i = \dim \mathrm{link}_\Delta(F), \\ 0 & \text{otherwise,} \end{cases}$$

where $\widetilde{H}_*(\Gamma, \Bbbk)$ denotes reduced simplicial homology of Γ with coefficients in $\Bbbk$. We call Δ a *homology ball* (over $\Bbbk$) if there exists a subcomplex $\partial\Delta$ of Δ, called the *boundary* of Δ, so that the following hold:

- $\partial\Delta$ is a $(d-2)$-dimensional homology sphere over $\Bbbk$.
- For every $F \in \Delta$ (including $F = \varnothing$) we have

$$\widetilde{H}_i(\mathrm{link}_\Delta(F), \Bbbk) = \begin{cases} \Bbbk & \text{if } F \notin \partial\Delta \text{ and } i = \dim \mathrm{link}_\Delta(F), \\ 0 & \text{otherwise.} \end{cases}$$

The *interior* of Δ is defined as $\mathrm{int}(\Delta) = \Delta$ if Δ is a homology sphere and as $\mathrm{int}(\Delta) = \Delta \smallsetminus \partial\Delta$ if Δ is a homology ball. For example, the simplicial complex $\{\varnothing, \{v\}\}$ with a unique vertex v is a 0-dimensional homology ball (over any field) with boundary $\{\varnothing\}$ and interior $\{\{v\}\}$. If Δ is a homology ball of dimension $d-1$, then $\partial\Delta$ consists exactly of the faces of those $(d-2)$-dimensional faces of Δ that are contained in a unique facet of Δ.

Remark 2.1. It follows from standard facts [Björner 1995, (9.12)] on the homology of simplicial joins that the simplicial join of a homology sphere and a homology ball or of two homology balls is a homology ball and that the simplicial join of two homology spheres is again a homology sphere. Moreover, in each case the interior of the simplicial join is equal to the simplicial join of the interiors of the two complexes in question.

2C. *Subdivisions.* We will adopt the following notion of homology subdivision of an abstract simplicial complex. This notion generalizes that of topological subdivision of [Stanley 1992, Section 2]. We should point out that the class of homology subdivisions of simplicial complexes is contained in the much broader class of formal subdivisions of Eulerian posets, introduced and studied in [Stanley 1992, Section 7].

Definition 2.2. Let Δ be a simplicial complex. A (finite, simplicial) *homology subdivision* of Δ (over $\Bbbk$) is a simplicial complex Δ' together with a map $\sigma : \Delta' \to \Delta$ such that the following hold for every $F \in \Delta$: (a) the set $\Delta'_F = \sigma^{-1}(2^F)$ is a subcomplex of Δ' that is a homology ball (over $\Bbbk$) of dimension $\dim(F)$, and (b) $\sigma^{-1}(F)$ consists of the interior faces of Δ'_F.

Such a map σ is said to be a *topological* subdivision if the complex Δ'_F is homeomorphic to a ball of dimension $\dim(F)$ for every $F \in \Delta$.

Let $\sigma : \Delta' \to \Delta$ be a homology subdivision of Δ. From the defining properties, it follows that the map σ is surjective and that $\dim(\sigma(E)) \geq \dim(E)$ for every $E \in \Delta'$. Given faces $E \in \Delta'$ and $F \in \Delta$, the face $\sigma(E)$ of Δ is called the *carrier* of E; the subcomplex Δ'_F is called the *restriction* of Δ' to F. The subdivision σ is called *quasigeometric* [Stanley 1992, Definition 4.1(a)] if there do not exist $E \in \Delta'$ and face $F \in \Delta$ of dimension smaller than $\dim(E)$ such that the carrier of every vertex of E is contained in F. Moreover, σ is called *geometric* [Stanley 1992, Definition 4.1(b)] if there exists a geometric realization of Δ' that geometrically subdivides a geometric realization of Δ in the way prescribed by σ.

Clearly, if $\sigma : \Delta' \to \Delta$ is a homology or topological subdivision, then the restriction of σ to Δ'_F is also a homology or topological subdivision of the simplex 2^F for every $F \in \Delta$, respectively. Moreover, if σ is quasigeometric or geometric, respectively, then so are all its restrictions Δ'_F for $F \in \Delta$. As part (c) of the following example shows, the restriction of σ to a face $F \in \Delta$ need not be a flag complex even when Δ' and Δ are flag complexes and σ is quasigeometric.

Example 2.3. Consider a 3-dimensional simplex 2^V with $V = \{a, b, c, d\}$ and set $F = \{b, c, d\}$.

(a) Let Γ be the simplicial complex consisting of the subsets of V and the subsets of $\{b, c, d, e\}$, and let $\sigma : \Gamma \to 2^V$ be the subdivision (considered in part (h) of [Stanley 1992, Example 2.3]) that pushes Γ into the simplex 2^V so that the face F of Γ ends up in the interior of 2^V and e ends up in the interior of 2^F. Formally, for $E \in \Gamma$ we let $\sigma(E) = E$ if $E \in 2^V \smallsetminus \{F\}$, we let $\sigma(E) = V$ if E contains F and otherwise we let $\sigma(E) = F$. Then Γ is a flag complex and the restriction Γ_F of σ is the cone over the boundary of 2^F (with new vertex e), which is not flag. See left half of Figure 1.

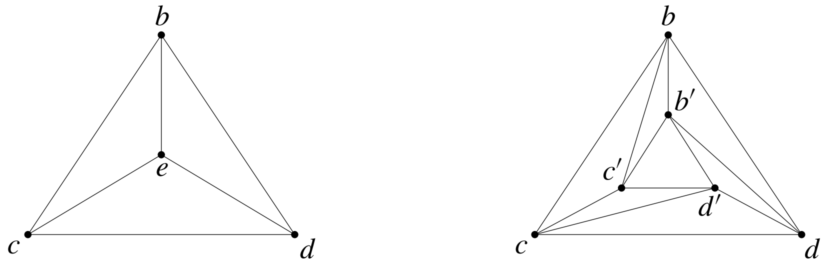


Figure 1. Two nonflag subdivisions of a triangle. Left: see parts (a) and (b) of Example 2.3. Right: see part (c) of Example 2.3.

(b) Let Γ' be the simplicial complex consisting of the faces of the simplex 2^V and those of the cone on a vertex v over the boundary of the simplex with vertex set $\{b, c, d, e\}$. (Note that Γ' is not flag.) Consider the subdivision $\sigma' : \Gamma' \to 2^V$ that satisfies $\sigma'(E) = V$ for every face $E \in \Gamma'$ containing v and otherwise agrees with the subdivision σ of part (a). Then σ' is quasigeometric, and its restriction $\Gamma'_F = \Gamma_F$ is again the nonflag complex shown in the left half of Figure 1.

(c) Let Γ_0 be the simplicial complex on the ground set $F \cup \{b', c', d'\}$ whose faces are F and those of the simplicial subdivision of 2^F, shown in Figure 1, right. Let Γ'' consist of the faces of 2^V and those of the cone over Γ_0 on a new vertex v. We leave to the reader to verify that Γ'' is a flag simplicial complex and that it admits a quasigeometric subdivision $\sigma'' : \Gamma'' \to 2^V$ (satisfying $\sigma''(v) = \sigma''(F) = V$) for which the restriction Γ''_F is the nonflag simplicial complex shown in Figure 1, right.

The previous examples suggest the following definitions:

Definition 2.4. Let Δ' and Δ be simplicial complexes, and let $\sigma : \Delta' \to \Delta$ be a homology subdivision.

(i) We say that σ is *vertex-induced* if for all faces $E \in \Delta'$ and $F \in \Delta$ the following condition holds: if every vertex of E is a vertex of Δ'_F, then $E \in \Delta'_F$.

(ii) We say that σ is a *flag subdivision* if the restriction Δ'_F is a flag complex for every face $F \in \Delta$.

For homology or topological subdivisions, we have the hierarchy of properties: geometric ⇒ vertex-induced ⇒ quasigeometric. The subdivision Γ of Example 2.3 is not quasigeometric while Γ' and Γ'' are quasigeometric but not vertex-induced. (None of the three subdivisions is flag.) Thus, the second implication above is strict. An example discussed in [Chan 1994, p. 468] shows that the first implication is strict as well. We also point out here that if $\sigma : \Delta' \to \Delta$ is a vertex-induced homology subdivision and the simplicial complex Δ' is flag, then σ is a flag subdivision.

Joins and links. The notion of a (vertex-induced or flag) homology subdivision behaves well with respect to simplicial joins and links, as we now explain. Let $\sigma_1 : \Delta'_1 \to \Delta_1$ and $\sigma_2 : \Delta'_2 \to \Delta_2$ be homology subdivisions of two simplicial complexes Δ_1 and Δ_2 on disjoint ground sets. The simplicial join $\Delta'_1 * \Delta'_2$ is naturally a homology subdivision of $\Delta_1 * \Delta_2$ with subdivision map $\sigma : \Delta'_1 * \Delta'_2 \to \Delta_1 * \Delta_2$ defined by $\sigma(E_1 \cup E_2) = \sigma_1(E_1) \cup \sigma_2(E_2)$ for $E_1 \in \Delta'_1$ and $E_2 \in \Delta'_2$. Indeed, given faces $F_1 \in \Delta_1$ and $F_2 \in \Delta_2$, the restriction of $\Delta'_1 * \Delta'_2$ to the face $F = F_1 \cup F_2 \in \Delta_1 * \Delta_2$ is equal to $(\Delta'_1)_{F_1} * (\Delta'_2)_{F_2}$, which, by Remark 2.1, is a homology ball of dimension equal to that of $F_1 \cup F_2$. Moreover, $\sigma^{-1}(F) = \sigma_1^{-1}(F_1) * \sigma_2^{-1}(F_2)$, and hence, $\sigma^{-1}(F)$ is the interior of this ball.

Similarly, let $\sigma : \Delta' \to \Delta$ be a homology subdivision, and let F be a common face of Δ and Δ' (such as a vertex of Δ) that satisfies $\sigma(F) = F$. An easy application of part (ii) of Lemma 4.1 shows that $\mathrm{link}_{\Delta'}(F)$ is a homology subdivision of $\mathrm{link}_\Delta(F)$ with subdivision map $\sigma_F : \mathrm{link}_{\Delta'}(F) \to \mathrm{link}_\Delta(F)$ defined by $\sigma_F(E) = \sigma(E \cup F) \smallsetminus F$. We will refer to this subdivision as the link of σ at F; its restriction to a face $G \in \mathrm{link}_\Delta(F)$ satisfies $(\mathrm{link}_{\Delta'}(F))_G = \mathrm{link}_{\Delta'_{F \cup G}}(F)$.

The following statement is an easy consequence of the relevant definitions; the proof is left to the reader:

Lemma 2.5. *The simplicial join of two vertex-induced or flag homology subdivisions is also vertex-induced or flag*, *respectively. The link of a vertex-induced or flag homology subdivision is also vertex-induced or flag*, *respectively.*

Stellar subdivisions. We recall the following standard way to subdivide a simplicial complex. Given a simplicial complex Δ on the ground set Ω, a face $F \in \Delta$ and an element v not in Ω, the *stellar subdivision* of Δ on F (with new vertex v) is the simplicial complex

$$\Delta' = (\Delta \smallsetminus \mathrm{st}_\Delta(F)) \cup (\{v\} * \partial(2^F) * \mathrm{link}_\Delta(F))$$

on the ground set $\Omega \cup \{v\}$, where $\partial(2^F) = 2^F \smallsetminus \{F\}$. The map $\sigma : \Delta' \to \Delta$, defined by

$$\sigma(E) = \begin{cases} E & \text{if } E \in \Delta, \\ (E \smallsetminus \{v\}) \cup F & \text{otherwise} \end{cases}$$

for $E \in \Delta'$, is a topological (and thus a homology) subdivision of Δ. We leave to the reader to check that if Δ is a flag complex and $F \in \Delta$ is an edge, then the stellar subdivision of Δ on F is again a flag complex.

3. Face enumeration, γ-vectors and local h-vectors

This section reviews the definitions and main properties of the enumerative invariants of simplicial complexes and their subdivisions that will appear in the following sections, namely the h-vector of a simplicial complex, the γ-vector of an Eulerian simplicial complex and the local h-vector of a simplicial subdivision of a simplex. Some new results on local h-vectors are included.

3A. *h-vectors.* A fundamental enumerative invariant of a $(d-1)$-dimensional simplicial complex Δ is the h-polynomial, defined by

$$h(\Delta, x) = \sum_{F \in \Delta} x^{|F|}(1-x)^{d-|F|}.$$

The h-vector of Δ is the sequence $h(\Delta) = (h_0(\Delta), h_1(\Delta), \dots, h_d(\Delta))$, where $h(\Delta, x) = \sum_{i=0}^{d} h_i(\Delta) x^i$. The number

$$(-1)^{d-1} h_d(\Delta) = \sum_{F \in \Delta} (-1)^{|F|-1}$$

is the reduced Euler characteristic of Δ and is denoted by $\widetilde{\chi}(\Delta)$. The polynomial $h(\Delta, x)$ satisfies $h_i(\Delta) = h_{d-i}(\Delta)$ [Stanley 1997, Section 3.14] if Δ is an *Eulerian* complex, meaning that

$$\widetilde{\chi}(\mathrm{link}_\Delta(F)) = (-1)^{\dim \mathrm{link}_\Delta(F)}$$

holds for every $F \in \Delta$. For the simplicial join of two simplicial complexes Δ_1 and Δ_2 we have $h(\Delta_1 * \Delta_2, x) = h(\Delta_1, x)\, h(\Delta_2, x)$. For a homology ball or sphere Δ of dimension $d-1$ we set

$$h(\mathrm{int}(\Delta), x) = \sum_{F \in \mathrm{int}(\Delta)} x^{|F|}(1-x)^{d-|F|}$$

and recall the following well known statement (see, for instance, Theorem 7.1 in [Stanley 1996, Chapter II] and [Athanasiadis 2012, Section 2.1] for additional references).

Lemma 3.1. *Let Δ be a $(d-1)$-dimensional simplicial complex. If Δ is either a homology ball or a homology sphere over $\Bbbk$, then $x^d h(\Delta, 1/x) = h(\mathrm{int}(\Delta), x)$.*

3B. *γ-vectors.* Let $h = (h_0, h_1, \dots, h_d)$ be a vector with real coordinates, and let $h(x) = \sum_{i=0}^{d} h_i x^i$ be the associated real polynomial of degree at most d. We say that $h(x)$ *has symmetric coefficients* and that the vector h is *symmetric* if $h_i = h_{d-i}$ holds for $0 \le i \le d$. It is easy to check [Gal 2005, Proposition 2.1.1] that $h(x)$ has symmetric coefficients if and only if there exists a real polynomial $\gamma(x) = \sum_{i=0}^{\lfloor d/2 \rfloor} \gamma_i x^i$ of degree at most $\lfloor d/2 \rfloor$ satisfying

$$h(x) = (1+x)^d \gamma\left(\frac{x}{(1+x)^2}\right) = \sum_{i=0}^{\lfloor d/2 \rfloor} \gamma_i x^i (1+x)^{d-2i}. \tag{3-1}$$

In that case, $\gamma(x)$ is uniquely determined by $h(x)$ and called the *γ-polynomial* associated with $h(x)$; the sequence $(\gamma_0, \gamma_1, \dots, \gamma_{\lfloor d/2 \rfloor})$ is called the *γ-vector* associated with h. We will refer to the γ-polynomial associated with the h-polynomial of an Eulerian complex Δ as the *γ-polynomial of* Δ and will denote it by $\gamma(\Delta, x)$. Similarly, we will refer to the γ-vector associated with the h-vector of an Eulerian complex Δ as the *γ-vector of* Δ and will denote it by $\gamma(\Delta)$.

3C. *Local h-vectors.* We now recall some of the basics of the theory of face enumeration for subdivisions of simplicial complexes [Stanley 1992; 1996, Section III.10]. The following definition is a restatement of [Stanley 1992, Definition 2.1] for homology (rather than topological) subdivisions of the simplex:

Definition 3.2. Let V be a set with d elements, and let Γ be a homology subdivision of the simplex 2^V. The polynomial $\ell_V(\Gamma, x) = \ell_0 + \ell_1 x + \cdots + \ell_d x^d$ defined by

$$\ell_V(\Gamma, x) = \sum_{F \subseteq V} (-1)^{d-|F|} h(\Gamma_F, x) \tag{3-2}$$

is the *local h-polynomial* of Γ (with respect to V). The *local h-vector* of Γ (with respect to V) is the sequence $\ell_V(\Gamma) = (\ell_0, \ell_1, \dots, \ell_d)$.

The following theorem, stated for homology subdivisions, summarizes some of the main properties of local h-vectors (see Theorems 3.2 and 3.3 and Corollary 4.7 in [Stanley 1992]):

Theorem 3.3. (i) *For every homology subdivision Δ' of a pure simplicial complex Δ we have*

$$h(\Delta', x) = \sum_{F \in \Delta} \ell_F(\Delta'_F, x)\, h(\mathrm{link}_\Delta(F), x). \tag{3-3}$$

(ii) *The local h-vector $\ell_V(\Gamma)$ is symmetric for every homology subdivision Γ of the simplex 2^V.*

(iii) *The local h-vector $\ell_V(\Gamma)$ has nonnegative coordinates for every quasigeometric homology subdivision Γ of the simplex 2^V.*

Proof. Parts (i) and (iii) follow from the proofs of Theorems 3.2 and 4.6, respectively, in [Stanley 1992]. Moreover, Lemma 3.1 implies that every homology subdivision of a simplicial complex is a formal subdivision in the sense of [Stanley 1992, Definition 7.4]. Thus, parts (i) and (ii) are special cases of Corollary 7.7 and Theorem 7.8, respectively, in [Stanley 1992]. □

Example 3.4. The local h-polynomial of the subdivision in part (a) of Example 2.3 was computed in [Stanley 1992] as $\ell_V(\Gamma, x) = -x^2$. This shows that the assumption in Theorem 3.3(iii) that Γ is quasigeometric is essential. For part (b) of Example 2.3 we can easily compute that $\ell_V(\Gamma', x) = x + x^3$. Since Γ' is quasigeometric, this disproves [Stanley 1992, Conjecture 5.4] (see also [Chan 1994, Section 6; Stanley

1996, p. 134]), stating that local h-vectors of quasigeometric subdivisions are unimodal.

The previous example suggests the following question:

Question 3.5. Is the local h-vector $\ell_V(\Gamma)$ unimodal for every vertex-induced homology subdivision Γ of the simplex 2^V?

We now show that local h-vectors also enjoy a locality property. (This will be useful in the proof of Proposition 6.1.)

Proposition 3.6. *Let $\sigma : \Gamma \to 2^V$ be a homology subdivision of the simplex 2^V. For every homology subdivision Γ' of Γ we have*

$$\ell_V(\Gamma', x) = \sum_{E \in \Gamma} \ell_E(\Gamma'_E, x)\, \ell_V(\Gamma, E, x), \tag{3-4}$$

where

$$\ell_V(\Gamma, E, x) = \sum_{\sigma(E) \subseteq F \subseteq V} (-1)^{d-|F|} h(\mathrm{link}_{\Gamma_F}(E), x) \tag{3-5}$$

for $E \in \Gamma$.

Proof. By assumption, Γ'_F is a homology subdivision of Γ_F for every $F \subseteq V$. Thus, using the defining Equation (3-2) for $\ell_V(\Gamma', x)$ and (3-3) to expand $h(\Gamma'_F, x)$ for $F \subseteq V$, we get

$$\begin{aligned} \ell_V(\Gamma', x) &= \sum_{F \subseteq V} (-1)^{d-|F|} h(\Gamma'_F, x) \\ &= \sum_{F \subseteq V} (-1)^{d-|F|} \sum_{E \in \Gamma_F} \ell_E(\Gamma'_E, x)\, h(\mathrm{link}_{\Gamma_F}(E), x) \\ &= \sum_{E \in \Gamma} \ell_E(\Gamma'_E, x) \sum_{F \subseteq V : \sigma(E) \subseteq F} (-1)^{d-|F|} h(\mathrm{link}_{\Gamma_F}(E), x), \end{aligned}$$

and the proof follows. □

Remark 3.7. We call the polynomial $\ell_V(\Gamma, E, x)$ defined by (3-5) the *relative local h-polynomial* of Γ (with respect to V) at E. This polynomial reduces to $\ell_V(\Gamma, x)$ for $E = \varnothing$ and shares many of the important properties of $\ell_V(\Gamma, x)$, established in [Stanley 1992]. For instance, using ideas of [Stanley 1992] and their refinements in [Athanasiadis 2012], one can show that $\ell_V(\Gamma, E, x)$ has symmetric coefficients in the sense that

$$x^{d-|E|} \ell_V(\Gamma, E, 1/x) = \ell_V(\Gamma, E, x)$$

for every homology subdivision Γ of 2^V and $E \in \Gamma$ and that $\ell_V(\Gamma, E, x)$ has nonnegative coefficients for every quasigeometric homology subdivision Γ of 2^V and $E \in \Gamma$. As a consequence of the latter statement and (3-4), we have

$\ell_V(\Gamma', x) \geq \ell_V(\Gamma, x)$ for every quasigeometric homology subdivision Γ of 2^V and every quasigeometric homology subdivision Γ' of Γ. Since these results will not be used in this paper, detailed proofs will appear elsewhere.

4. Flag subdivisions of Σ_{d-1}

This section proves Theorem 1.5 as well as a result on flag subdivisions of the simplex (Proposition 4.6), which will be used in Section 5.

The following lemma gives several technical properties of homology balls and spheres (over the field $\Bbbk$). We will only sketch the proof, which is fairly straightforward and uses standard tools from algebraic topology.

Lemma 4.1. (i) *If Δ is a homology sphere or ball of dimension $d-1$, then $\mathrm{link}_\Delta(F)$ is a homology sphere of dimension $d-|F|-1$ for every $F \in \Delta$ or interior face $F \in \Delta$, respectively.*

(ii) *If Δ is a homology ball of dimension $d-1$ and $F \in \Delta$ is a boundary face, then $\mathrm{link}_\Delta(F)$ is a homology ball of dimension $d-|F|-1$ with interior equal to $\mathrm{link}_\Delta(F) \cap \mathrm{int}(\Delta)$.*

(iii) *If Δ is a homology sphere or ball, then the cone over Δ is a homology ball whose boundary is equal to Δ or the union of Δ with the cone over the boundary of Δ, respectively.*

(iv) *Let Δ_1 and Δ_2 be homology balls of dimension d. If $\Delta_1 \cap \Delta_2$ is a homology ball of dimension $d-1$ that is contained in or equal to the boundary of both Δ_1 and Δ_2, then $\Delta_1 \cup \Delta_2$ is a homology ball or sphere of dimension d, respectively.*

(v) *Let Δ be a homology sphere of dimension $d-1$. If Γ is a subcomplex of Δ that is a homology ball of dimension $d-1$, then the complement of the interior of Γ in Δ is also a homology ball of dimension $d-1$ whose boundary is equal to that of Γ.*

Proof. We first observe that for all faces $F \in \Delta$ and $E \in \mathrm{link}_\Delta(F)$, the link of E in $\mathrm{link}_\Delta(F)$ is equal to $\mathrm{link}_\Delta(E \cup F)$. Moreover, if Δ is a homology ball and F is an interior face, then so is $E \cup F$. Part (i) follows from these facts and the definition of homology balls and spheres. Part (ii) is an easy consequence of part (i) and the relevant definitions. Part (iii) is an easy consequence of the relevant definitions and the fact that cones have vanishing reduced homology. Part (iv) follows by an easy application of the Mayer–Vietoris long exact sequence [Munkres 1984, §25].

For the last part, we let K denote the complement of the interior of Γ in Δ and note that the pairs $(\|\Gamma\|, \|\partial\Gamma\|)$ and $(\|\Delta\|, \|\mathrm{K}\|)$ are compact triangulated relative homology manifolds that are orientable over $\Bbbk$. Applying the Lefschetz duality theorem [Munkres 1984, §70] and the long exact homology sequence [Munkres

1984, §23] to these pairs shows that K has trivial reduced homology over $\Bbbk$. Similar arguments work for the links of faces of K. The details are omitted. □

Remark 4.2. Although not all parts of Lemma 4.1 remain valid if homology balls and spheres are replaced by topological balls and spheres, they do hold for the subclasses of PL balls and PL spheres. (We refer the reader to [Björner et al. 1999, Section 4.7 (d)] for this claim, for a short introduction to PL topology and for additional references.) Thus, the results of this paper remain valid when homology balls and spheres are replaced by PL balls and spheres and the notion of homology subdivision is replaced by its natural PL analogue.

The following lemma will also be essential in the proof of Theorem 1.5. A similar result has appeared in [Barmak 2010, Lemma 3.2].

Lemma 4.3. *Let Δ be a flag $(d-1)$-dimensional homology sphere. For every nonempty face F of Δ, the subcomplex $\bigcup_{v \in F} \overline{\mathrm{st}}_\Delta(v)$ is a homology $(d-1)$-dimensional ball whose interior is equal to $\bigcup_{v \in F} \mathrm{st}_\Delta(v)$.*

Proof. We set $F = \{v_1, v_2, \dots, v_k\}$, $\bigcup_{i=1}^{k} \overline{\mathrm{st}}_\Delta(v_i) = \mathrm{K}$ and $\bigcup_{i=1}^{k} \mathrm{st}_\Delta(v_i) = \mathrm{L}$ and proceed by induction on the cardinality k of F. For $k = 1$, the complex K is the cone over $\mathrm{link}_\Delta(v_1)$ on the vertex v_1. Since Δ is a homology sphere, the result follows from parts (i) and (iii) of Lemma 4.1. Suppose that $k \geq 2$. We will also assume that $d \geq 3$ since the result is trivial otherwise. (We note that the assumption that Δ is flag is essential in the case $d = 2$.) Since by Lemma 4.1(i) links of flag homology spheres are also flag homology spheres, the complex $\Gamma = \mathrm{link}_\Delta(v_k)$ is a flag homology sphere of dimension $d-2$ and $\{v_1, \dots, v_{k-1}\}$ is a nonempty face of Γ. Thus, by the induction hypothesis, the union $\Gamma_1 = \bigcup_{i=1}^{k-1} \overline{\mathrm{st}}_\Gamma(v_i)$ is a homology ball of dimension $d-2$. Let Γ_0 denote the boundary of Γ_1, and let Γ_2 denote the complement of the interior of Γ_1 in Γ. Thus, Γ_0 is a homology sphere of dimension $d-3$, and, by part (v) of Lemma 4.1, Γ_2 is a homology ball of dimension $d-2$ whose boundary is equal to Γ_0.

Consider the union $\mathrm{K}_1 = \bigcup_{i=1}^{k-1} \overline{\mathrm{st}}_\Delta(v_i)$ and the cones $\mathrm{K}_2 = v_k * \Gamma_2$ and $\mathrm{K}_0 = v_k * \Gamma_0$. It is straightforward to verify that $\mathrm{K} = \mathrm{K}_1 \cup \mathrm{K}_2$ and that $\mathrm{K}_1 \cap \mathrm{K}_2 = \mathrm{K}_0$. We note that K_1 is a homology ball of dimension $d-1$ by the induction hypothesis and that K_2 and K_0 are homology balls of dimension $d-1$ and $d-2$, respectively, by part (iii) of Lemma 4.1. By the induction hypothesis, the interior of Γ_1 is equal to $\bigcup_{i=1}^{k-1} \mathrm{st}_\Gamma(v_i)$. Therefore, none of the faces of Γ_0 contains any of $v_1, \dots, v_{k-1}$, and hence, the same holds for K_0. Since by the induction hypothesis the interior of K_1 is equal to $\bigcup_{i=1}^{k-1} \mathrm{st}_\Delta(v_i)$, we conclude that K_0 is contained in the boundary of K_1. Moreover, K_0 is also contained in the boundary of K_2 since Γ_0 is contained in the boundary of Γ_2. It follows from the previous discussion and Lemma 4.1(iv) that K is a homology $(d-1)$-dimensional ball.

We now verify that the interior of K is equal to L. This statement may be derived from the previous inductive argument since the interior of K is equal to the union of the interiors of K_1, K_2 and K_0. We give the following alternative argument: Since K is a homology ball, its boundary consists of all faces of the $(d-2)$-dimensional faces of K that are contained in exactly one facet of K. The validity of the statement for $k = 1$ implies that these $(d-2)$-dimensional faces of K are precisely those that do not contain any of the v_i and that are not contained in more than one of the subcomplexes $\mathrm{link}_\Delta(v_i)$. However, since Δ is $(d-1)$-dimensional and flag, no $(d-2)$-dimensional face of Δ may be contained in more than one of the $\mathrm{link}_\Delta(v_i)$. Thus, the boundary of K consists precisely of its faces that do not contain any of the v_i, and the proof follows. □

Proof of Theorem 1.5. Let Δ be a flag simplicial complex of dimension $d-1$ and Σ_{d-1} be the simplicial join of the zero-dimensional spheres $\{\varnothing, \{u_i\}, \{v_i\}\}$ for $1 \le i \le d$. We fix a facet $\{x_1, x_2, \dots, x_d\}$ of Δ, and for $E \in \Delta$ we define

$$\sigma(E) = \{\, u_i : x_i \in E \,\} \cup \{\, v_i : E \notin \overline{\mathrm{st}}_\Delta(x_i) \,\}. \tag{4-1}$$

Clearly, $\sigma(E)$ cannot contain any of the sets $\{u_i, v_i\}$. Thus, we have $\sigma(E) \in \Sigma_{d-1}$ for every $E \in \Delta$, and hence, we get a map $\sigma : \Delta \to \Sigma_{d-1}$. We will prove that this map is a homology subdivision of Σ_{d-1} if Δ is a homology sphere. Given a face $F \in \Sigma_{d-1}$, we need to show that $\sigma^{-1}(2^F)$ is a subcomplex of Δ of dimension $\dim(F)$ that is a homology ball with interior $\sigma^{-1}(F)$. We denote by S the subset of $\{x_1, x_2, \dots, x_d\}$ consisting of all vertices x_i for which $F \cap \{u_i, v_i\} = \varnothing$ and distinguish two cases:

Case 1: $S = \varnothing$. We may assume that $F = \{u_1, \dots, u_k\} \cup \{v_{k+1}, \dots, v_d\}$ for some $k \le d$. Setting $E_0 = \{x_1, \dots, x_k\}$, the defining Equation (4-1) shows that $\sigma^{-1}(2^F)$ is equal to the intersection of $\bigcap_{i=1}^k \overline{\mathrm{st}}_\Delta(x_i)$ with the complement of $\bigcup_{i=k+1}^d \mathrm{st}_\Delta(x_i)$ in Δ and that $\sigma^{-1}(F)$ consists of those faces of $\sigma^{-1}(2^F)$ that contain E_0 and do not belong to any of the $\mathrm{link}_\Delta(x_i)$ for $k+1 \le i \le d$. Consider the complex $\Gamma = \mathrm{link}_\Delta(E_0)$, and let K denote the complement of the union $\bigcup_{i=k+1}^d \mathrm{st}_\Gamma(x_i)$ in Γ. Since links of homology spheres are also homology spheres (see part (i) of Lemma 4.1), the complex Γ is a homology sphere of dimension $d - |F| - 1$. By Lemma 4.3 and part (v) of Lemma 4.1, K is a homology ball of dimension $d-|F|-1$ whose interior is equal to the set of those faces of K that do not belong to any of the $\mathrm{link}_\Gamma(x_i)$ for $k+1 \le i \le d$. From the above we conclude that $\sigma^{-1}(2^F)$ is equal to the simplicial join of the simplex 2^{E_0} and K and that $\sigma^{-1}(F)$ is equal to the simplicial join of $\{E_0\}$ and the interior of K. The result now follows from part (iii) of Lemma 4.1 and the previous discussion.

Case 2: $S \neq \varnothing$. Then $\sigma^{-1}(2^F)$ is contained in $\mathrm{link}_\Delta(S)$. As a result, replacing Δ by $\mathrm{link}_\Delta(S)$ reduces this case to the previous one.

Finally, we note that (4-1) may be rewritten as $\sigma(E) = \bigcup_{x \in E} f(x)$, where

$$f(x) = \begin{cases} \{u_i\} & \text{if } x = x_i \text{ for some } 1 \le i \le d, \\ \{\, v_i : x \notin \mathrm{link}_\Delta(x_i)\,\} & \text{otherwise} \end{cases}$$

for every vertex x of Δ. This implies that for every $E \in \Delta$, the carrier of E is equal to the union of the carriers of the vertices of E. As a result, σ is vertex-induced and the proof follows. □

Corollary 4.4. *Given any flag homology sphere Δ of dimension $d-1$, there exist simplicial complexes Γ_F, one for each face $F \in \Sigma_{d-1}$, with the following properties*: (a) Γ_F *is a flag vertex-induced homology subdivision of the simplex 2^F for every $F \in \Sigma_{d-1}$, and* (b) *we have*

$$h(\Delta, x) = \sum_{F \in \Sigma_{d-1}} \ell_F(\Gamma_F, x)(1+x)^{d-|F|}. \tag{4-2}$$

Proof. We apply (3-3) to the subdivision of Δ guaranteed by Theorem 1.5 and note that for every $F \in \Sigma_{d-1}$, the restriction Γ_F of this subdivision to F has the required properties and that $h(\mathrm{link}_{\Sigma_{d-1}}(F), x) = (1+x)^{d-|F|}$. □

Remark 4.5. Due to (4-2) and Theorem 3.3(iii), $h(\Delta, x) \ge (1+x)^d$ for every flag $(d-1)$-dimensional homology sphere Δ. This inequality was proved, more generally, for every flag $(d-1)$-dimensional doubly Cohen–Macaulay simplicial complex Δ in [Athanasiadis 2011, Theorem 1.3].

We now fix a d-element set $V = \{v_1, v_2, \dots, v_d\}$ and a homology subdivision Γ of 2^V with subdivision map $\sigma : \Gamma \to 2^V$. We let $U = \{u_1, u_2, \dots, u_d\}$ be a d-element set that is disjoint from V and consider the union Δ of all collections of the form $2^E * \Gamma_G$, where $E = \{\, u_i : i \in I\,\}$ and $G = \{\, v_j : j \in J\,\}$ are subsets of U and V, respectively, and (I, J) ranges over all ordered pairs of disjoint subsets of $\{1, 2, \dots, d\}$. Clearly, Δ is a simplicial complex that contains as a subcomplex Γ (set $I = \varnothing$) and the simplex 2^U (set $J = \varnothing$).

We let Σ_{d-1} be as in the proof of Theorem 1.5 and define the map $\sigma_0 : \Delta \to \Sigma_{d-1}$ by $\sigma_0(E \cup F) = E \cup \sigma(F)$ for all $E \subseteq U$ and $F \in \Gamma$ such that $E \cup F \in \Delta$. The second result of this section is as follows:

Proposition 4.6. *Under the established assumptions and notation, we have*:

(i) *The complex Δ is a $(d-1)$-dimensional homology sphere.*

(ii) *Endowed with the map σ_0, the complex Δ is a homology subdivision of Σ_{d-1}.*

(iii) *If Γ is flag and vertex-induced, then Δ is a flag simplicial complex and a flag, vertex-induced homology subdivision of Σ_{d-1}.*

Proof. We first verify (ii). We consider any face $W \in \Sigma_{d-1}$ so that $W = E \cup G$ for some $E \subseteq U$ and $G \subseteq V$ and recall that Γ_G is a homology ball of dimension $\dim(G)$.

We have $\sigma_0^{-1}(2^W) = 2^E * \Gamma_G$ and $\sigma_0^{-1}(W) = \{E\} * \sigma^{-1}(G) = \{E\} * \mathrm{int}(\Gamma_G)$ by the definition of σ_0. Thus, it follows from part (iii) of Lemma 4.1 that $\sigma_0^{-1}(2^W)$ is a homology ball of dimension $\dim(W)$ and that its interior is equal to $\sigma_0^{-1}(W)$.

Part (i) may be deduced from part (ii) as follows. Let $F_0, F_1, \ldots, F_m$ be a linear ordering of the facets of Σ_{d-1} such that $F_i \cap U \subset F_j \cap U$ implies $i < j$. Thus, we have $m = 2^d$, $F_0 = V$ and $F_m = U$. By assumption, $\Delta_{F_0} = \Gamma_V$ is a $(d-1)$-dimensional homology ball. Moreover, Δ_{F_j} is equal to the simplicial join of a face of 2^U with the restriction of Γ to a face of 2^V for $1 \le j \le m$ and hence a $(d-1)$-dimensional homology ball by part (iii) of Lemma 4.1, and $\Delta_{F_j} \cap \bigcup_{i=0}^{j-1} \Delta_{F_i}$ is equal to the simplicial join of the boundary of this face with the same restriction of Γ. It follows from part (iv) of Lemma 4.1 by induction on j that $\bigcup_{i=0}^{j} \Delta_{F_i}$ is a $(d-1)$-dimensional homology ball for $0 \le j \le m-1$ and a $(d-1)$-dimensional homology sphere for $j = m$. This proves (i) since $\Delta = \bigcup_{i=0}^{m} \Delta_{F_i}$.

To verify (iii), assume that Γ is flag and vertex-induced. It is clear from the definition of σ_0 that the subdivision Δ is also vertex-induced. Since the restriction of Δ to any face of Σ_{d-1} is the join of a simplex with the restriction of Γ to a face of 2^V, the subdivision Δ is flag as well. To verify that Δ is a flag complex, let $E \cup F$ be a set of vertices of Δ that are pairwise joined by edges, where $E = \{ u_i : i \in I \}$ for some $I \subseteq \{1, 2, \ldots, d\}$ and F consists of vertices of Γ. We need to show that $E \cup F \in \Delta$. We set $J = \{1, 2, \ldots, d\} \smallsetminus I$ and $G = \{ v_j : j \in J \}$ and note that the elements of F are vertices of Γ_G by definition of Δ. Since the elements of F are pairwise joined by edges in Γ, our assumptions that Γ is vertex-induced and flag imply that $F \in \Gamma_G$. Therefore, $E \cup F$ belongs to $2^E * \Gamma_G$, which is contained in Δ, and the result follows. □

Remark 4.7. The conclusion in Proposition 4.6 that Δ is a flag complex does not hold under the weaker hypothesis that Γ is quasigeometric rather than vertex-induced. For instance, let Γ be the simplicial complex consisting of the subsets of $V = \{v_1, v_2, v_3\}$ and $\{v_2, v_3, v_4\}$, and let $\sigma : \Gamma \to 2^V$ be the subdivision that pushes Γ into 2^V so that the face $F = \{v_2, v_3\}$ of Γ ends up in the interior of 2^V and v_4 ends up in the interior of 2^F. Then Γ is quasigeometric and flag, but the simplicial complex Δ is not flag since it has $\{u_1, v_2, v_3\}$ as a minimal nonface.

5. Local γ-vectors

This section defines the local γ-vector of a homology subdivision of the simplex, lists examples and elementary properties, discusses its nonnegativity in the special case of flag subdivisions and concludes with the proof of Theorem 1.3. This proof comes as an application of the considerations and results of the present and the previous sections.

Definition 5.1. Let V be a set with d elements, and let Γ be a homology subdivision of the simplex 2^V. The polynomial $\xi_V(\Gamma, x) = \xi_0 + \xi_1 x + \cdots + \xi_{\lfloor d/2 \rfloor} x^{\lfloor d/2 \rfloor}$ defined by

$$\ell_V(\Gamma, x) = (1+x)^d \xi_V\left(\Gamma, \frac{x}{(1+x)^2}\right) = \sum_{i=0}^{\lfloor d/2 \rfloor} \xi_i x^i (1+x)^{d-2i} \tag{5-1}$$

is the *local γ-polynomial* of Γ (with respect to V). The *local γ-vector* of Γ (with respect to V) is the sequence $\xi_V(\Gamma) = (\xi_0, \xi_1, \ldots, \xi_{\lfloor d/2 \rfloor})$.

Thus, $\xi_V(\Gamma, x)$ is the γ-polynomial associated with $\ell_V(\Gamma, x)$, and $\xi_V(\Gamma)$ is the γ-vector associated with $\ell_V(\Gamma)$ in the sense of Section 3B. All formulas in the next example follow from corresponding formulas in [Stanley 1992, Example 2.3] or directly from the relevant definitions.

Example 5.2. (a) For the trivial subdivision $\Gamma = 2^V$ of the $(d-1)$-dimensional simplex 2^V we have

$$\xi_V(\Gamma, x) = \begin{cases} 1 & \text{if } d = 0, \\ 0 & \text{if } d \geq 1. \end{cases} \tag{5-2}$$

(b) Let $\xi_V(\Gamma) = (\xi_0, \xi_1, \ldots, \xi_{\lfloor d/2 \rfloor})$, where Γ and V are as in Definition 5.1. Assuming that $d \geq 1$, we have $\xi_0 = 0$ and $\xi_1 = f_0^\circ$, where f_0° is the number of interior vertices of Γ. Assuming that $d \geq 4$, we also have $\xi_2 = -(2d-3) f_0^\circ + f_1^\circ - \tilde{f}_0$, where f_1° is the number of interior edges of Γ and $\tilde{f}_0$ is the number of vertices of Γ that lie in the relative interior of a $(d-2)$-dimensional face of 2^V.

(c) Suppose that $d \in \{2, 3\}$. As a consequence of (b) we have $\xi_V(\Gamma, x) = tx$ for every homology subdivision Γ of 2^V, where t is the number of interior vertices of Γ.

(d) Let Γ be the cone over the boundary $2^V \smallsetminus \{V\}$ of the simplex 2^V (so Γ is the stellar subdivision of 2^V on the face V). Then $\ell_V(\Gamma, x) = x + x^2 + \cdots + x^{d-1}$, and hence, ξ_2 is negative for $d \geq 4$. For instance, we have $\xi_V(\Gamma, x) = x - x^2$ for $d = 4$.

(e) For the subdivisions of parts (b) and (c) of Example 2.3 we can compute that $\ell_V(\Gamma', x) = \ell_V(\Gamma'', x) = x + x^3$ and hence that $\xi_V(\Gamma', x) = \xi_V(\Gamma'', x) = x - 2x^2$.

The following proposition shows the relevance of local γ-vectors in the study of γ-vectors of subdivisions of Eulerian complexes:

Proposition 5.3. *Let Δ be a pure Eulerian simplicial complex. For every homology subdivision Δ' of Δ we have*

$$\gamma(\Delta', x) = \sum_{F \in \Delta} \xi_F(\Delta'_F, x) \gamma(\mathrm{link}_\Delta(F), x). \tag{5-3}$$

Proof. Since Δ is Eulerian, so is $\mathrm{link}_\Delta(F)$ for every $F \in \Delta$. Thus, applying (3-1) to the h-polynomial of $\mathrm{link}_\Delta(F)$ we get

$$h(\mathrm{link}_\Delta(F), x) = (1+x)^{d-|F|}\gamma\Big(\mathrm{link}_\Delta(F), \frac{x}{(1+x)^2}\Big),$$

where $d-1 = \dim(\Delta)$. Using this and (5-1), Equation (3-3) may be rewritten as

$$h(\Delta', x) = (1+x)^d \sum_{F\in\Delta} \xi_F\Big(\Delta'_F, \frac{x}{(1+x)^2}\Big)\,\gamma\Big(\mathrm{link}_\Delta(F), \frac{x}{(1+x)^2}\Big).$$

The proposed equality now follows from the uniqueness of the γ-polynomial associated with $h(\Delta', x)$. □

The following statement is the main conjecture of this paper:

Conjecture 5.4. *For every flag vertex-induced homology subdivision Γ of the simplex 2^V we have $\xi_V(\Gamma) \ge 0$.*

Parts (d) and (e) of Example 5.2 show that the conclusion of Conjecture 5.4 fails under various weakenings of the hypotheses. We do not know of an example of a flag quasigeometric homology subdivision of the simplex for which the local γ-vector fails to be nonnegative.

We now discuss some consequences of Theorem 1.5 and Proposition 5.3 related to Conjecture 5.4.

Corollary 5.5. *For every flag homology sphere Δ of dimension $d-1$ we have*

$$\gamma(\Delta, x) = \sum_{F\in\Sigma_{d-1}} \xi_F(\Gamma_F, x), \tag{5-4}$$

where Γ_F is as in Corollary 4.4 for each $F \in \Sigma_{d-1}$. In particular, the validity of Conjecture 5.4 for homology subdivisions Γ of dimension at most $d-1$ implies the validity of Conjecture 1.1 for homology spheres Δ of dimension at most $d-1$.

Proof. Setting $\ell_F(\Gamma_F, x) = \sum_i \xi_{F,i} x^i (1+x)^{|F|-2i}$ in (4-2) and changing the order of summation results in (5-4). Alternatively, one can apply (5-3) to the subdivision guaranteed by Theorem 1.5 and note that $\gamma(\mathrm{link}_{\Sigma_{d-1}}(F), x) = 1$ for every $F \in \Sigma_{d-1}$. The last sentence in the statement of the corollary follows from (5-4). □

Corollary 5.6. *The validity of Conjecture 5.4 for homology subdivisions Γ of dimension at most $d-1$ implies the validity of Conjecture 1.4 for homology spheres Δ and subdivisions Δ' of dimension at most $d-1$.*

Proof. We observe that the term corresponding to $F = \varnothing$ in the sum of the right-hand side of (5-3) is equal to $\gamma(\Delta, x)$. Thus, the result follows from (5-3), Corollary 5.5 and the fact that the link of every nonempty face of a flag homology sphere is also a flag homology sphere of smaller dimension. □

Proposition 5.7. *Conjecture 5.4 holds for subdivisions of the* 3*-dimensional simplex.*

Proof. Let Γ be a flag vertex-induced homology subdivision of the $(d-1)$-dimensional simplex 2^V, and let Δ be the homology subdivision of Σ_{d-1} considered in Proposition 4.6. Applying (5-3) to this subdivision and noting that $\gamma(\mathrm{link}_{\Sigma_{d-1}}(F), x) = 1$ for every $F \in \Sigma_{d-1}$, we get

$$\gamma(\Delta, x) = \sum_{F \in \Sigma_{d-1}} \xi_F(\Delta_F, x).$$

By definition of Δ, the restriction Δ_F is a cone over the restriction of Δ to a proper face of F for every $F \in \Sigma_{d-1}$ that is not contained in V. Since every such subdivision has a zero local h-vector [Stanley 1992, p. 821], the previous formula can be rewritten as

$$\gamma(\Delta, x) = \sum_{F \subseteq V} \xi_F(\Gamma_F, x). \tag{5-5}$$

Assume now that $d = 4$ so that $\xi(\Gamma, x) = \xi_0 + \xi_1 x + \xi_2 x^2$ for some integers ξ_0, ξ_1 and ξ_2. Since $\xi_0 = 0$ and $\xi_1 \geq 0$ by part (b) of Example 5.2, it suffices to show that $\xi_2 \geq 0$. For that, we observe that the only contribution to the coefficient of x^2 in the right-hand side of (5-5) comes from the term with $F = V$. As a result, ξ_2 is equal to the coefficient of x^2 in $\gamma(\Delta, x)$. Since Δ is a 3-dimensional flag homology sphere (by Proposition 4.6), this coefficient is nonnegative by the Davis–Okun theorem [Davis and Okun 2001, Theorem 11.2.1], and the result follows. □

Proof of Theorem 1.3. For 3-dimensional spheres the result is due to Proposition 5.7 and Corollary 5.6. Assume now that Δ and Δ' have dimension 4. Then we can write $\gamma(\Delta, x) = 1 + \gamma_1(\Delta)x + \gamma_2(\Delta)x^2$ and $\gamma(\Delta', x) = 1 + \gamma_1(\Delta')x + \gamma_2(\Delta')x^2$. Since $\gamma_1(\Delta) = f_0(\Delta) - 8$ and $\gamma_1(\Delta') = f_0(\Delta') - 8$, where $f_0(\Delta)$ and $f_0(\Delta')$ are the number of vertices of Δ and Δ', respectively, it is clear that $\gamma_1(\Delta') \geq \gamma_1(\Delta)$. As the computation in the proof of [Gal 2005, Corollary 2.2.2] shows, we also have

$$2\gamma_2(\Delta) = \sum_{v \in \mathrm{vert}(\Delta)} \gamma_2(\mathrm{link}_\Delta(v)),$$

where $\mathrm{vert}(\Delta)$ is the set of vertices of Δ. Similarly, we have

$$2\gamma_2(\Delta') = \sum_{v' \in \mathrm{vert}(\Delta')} \gamma_2(\mathrm{link}_{\Delta'}(v')),$$

where we may assume that $\mathrm{vert}(\Delta) \subseteq \mathrm{vert}(\Delta')$. Since $\mathrm{link}_{\Delta'}(v)$ is a flag vertex-induced homology subdivision of $\mathrm{link}_\Delta(v)$ for every $v \in \mathrm{vert}(\Delta)$, by Lemma 2.5, we have $\gamma_2(\mathrm{link}_{\Delta'}(v)) \geq \gamma_2(\mathrm{link}_\Delta(v))$ by the 3-dimensional case, treated earlier, for every such vertex v. Since $\mathrm{link}_{\Delta'}(v')$ is a 3-dimensional flag homology sphere, we also have $\gamma_2(\mathrm{link}_{\Delta'}(v')) \geq 0$ by the Davis–Okun theorem for every $v' \in \mathrm{vert}(\Delta') \smallsetminus \mathrm{vert}(\Delta)$. Hence, $\gamma_2(\Delta') \geq \gamma_2(\Delta)$, and the result follows. □

Question 5.8. Does $\gamma(\Delta') \geq \gamma(\Delta)$ hold for every flag homology sphere Δ and every flag homology subdivision Δ' of Δ?

6. Special cases

This section provides some evidence in favor of the validity of Conjecture 5.4 other than that provided by Proposition 5.7.

Simplicial joins. Let Γ be a homology subdivision of the simplex 2^V and Γ' be a homology subdivision of the simplex $2^{V'}$, where V and V' are disjoint finite sets. Then $\Gamma * \Gamma'$ is a homology subdivision of the simplex $2^V * 2^{V'} = 2^{V \cup V'}$, and given subsets $F \subseteq V$ and $F' \subseteq V'$, the restriction of $\Gamma * \Gamma'$ to the face $F \cup F'$ of this simplex satisfies $(\Gamma * \Gamma')_{F \cup F'} = \Gamma_F * \Gamma'_{F'}$. Since $h(\Gamma_F * \Gamma'_{F'}, x) = h(\Gamma_F, x)\, h(\Gamma'_{F'}, x)$, the defining Equation (3-2) and a straightforward computation show that

$$\ell_{V \cup V'}(\Gamma * \Gamma', x) = \ell_V(\Gamma, x)\, \ell_{V'}(\Gamma', x).$$

This equation and (5-1) imply that

$$\xi_{V \cup V'}(\Gamma * \Gamma', x) = \xi_V(\Gamma, x)\, \xi_{V'}(\Gamma', x). \tag{6-1}$$

From the previous formula and Lemma 2.5 we conclude that if Γ and Γ' satisfy the assumptions and the conclusion of Conjecture 5.4, then so does $\Gamma * \Gamma'$.

Edge subdivisions. Following [Charney and Davis 1995, Section 5.3], we refer to the stellar subdivision on an edge of a simplicial complex Γ as an *edge subdivision*. As mentioned in Section 3B, flagness of a simplicial complex is preserved by edge subdivisions. The following statement describes a class of flag (geometric) subdivisions of the simplex with nonnegative local γ-vectors:

Proposition 6.1. *For every subdivision Γ of the simplex 2^V that can be obtained from the trivial subdivision by successive edge subdivisions, we have $\xi_V(\Gamma) \geq 0$.*

Proof. Let Γ be a subdivision of 2^V and Γ' be the edge subdivision of Γ on $e = \{a, b\} \in \Gamma$. Thus, we have $\Gamma' = (\Gamma \smallsetminus \mathrm{st}_\Gamma(e)) \cup (\{v\} * \partial(e) * \mathrm{link}_\Gamma(e))$, where v is the new vertex added and $\partial(e) = \{\varnothing, \{a\}, \{b\}\}$.

By appealing to (3-4) and noticing that the right-hand side of this formula vanishes except when $E \in \{\varnothing, e\}$ (or by direct computation), we find that

$$\ell_V(\Gamma', x) = \ell_V(\Gamma, x) + x\ell_V(\Gamma, e, x).$$

Thus, it suffices to prove the following claim: if the γ-polynomial corresponding to $\ell_V(\Gamma, E, x)$ has nonnegative coefficients for every face $E \in \Gamma$ of positive dimension (meaning that $\ell_V(\Gamma, E, x)$ can be written as a linear combination of the polynomials $x^i(1+x)^{d-|E|-2i}$ with nonnegative coefficients for every $|E| \geq 2$), then the same holds for Γ'. We consider a face $E \in \Gamma'$ of positive dimension and distinguish the following cases. (We note that E cannot contain e and that if $E \in \Gamma$, then the carrier $\sigma(E) \subseteq V$ of E is the same, whether considered with respect to Γ or Γ'.)

Case 1: $E \in \Gamma \smallsetminus \mathrm{link}_\Gamma(e)$. The links $\mathrm{link}_{\Gamma'_F}(E)$ and $\mathrm{link}_{\Gamma_F}(E)$ are then combinatorially isomorphic for every $F \subseteq V$ that contains the carrier of E (since these two links are equal if $E \cup e \notin \Gamma$) and the defining Equation (3-5) implies that $\ell_V(\Gamma', E, x) = \ell_V(\Gamma, E, x)$.

Case 2: $E \in \mathrm{link}_\Gamma(e)$. For $F \subseteq V$ that contains the carrier of E, the link $\mathrm{link}_{\Gamma'_F}(E)$ is equal to either $\mathrm{link}_{\Gamma_F}(E)$ or to the edge subdivision of $\mathrm{link}_{\Gamma_F}(E)$ on e in case F does not or does contain the carrier of e, respectively. It follows from this and (3-3) that (see also [Gal 2005, Proposition 2.4.3])

$$h(\mathrm{link}_{\Gamma'_F}(E), x) = \begin{cases} h(\mathrm{link}_{\Gamma_F}(E), x) & \text{if } \sigma(e) \not\subseteq F, \\ h(\mathrm{link}_{\Gamma_F}(E), x) + x h(\mathrm{link}_{\Gamma_F}(E \cup e), x) & \text{if } \sigma(e) \subseteq F \end{cases}$$

and then from (3-5) that $\ell_V(\Gamma', E, x) = \ell_V(\Gamma, E, x) + x\ell_V(\Gamma, E \cup e, x)$.

Case 3: $E \notin \Gamma$. Then we must have $E \in \{v\} * \partial(e) * \mathrm{link}_\Gamma(e)$ and, in particular, $v \in E$. We distinguish two subcases:

Suppose first that E intersects e, and set $E' = (E \smallsetminus \{v\}) \cup e$. Then for every $F \subseteq V$ that contains the carrier of E in Γ', $\mathrm{link}_{\Gamma'_F}(E) = \mathrm{link}_{\Gamma_F}(E')$ (and the latter coincides with the carrier of E' in Γ), and hence, $\ell_V(\Gamma', E, x) = \ell_V(\Gamma, E', x)$.

Suppose finally that $E \cap e = \varnothing$, and set $E' = (E \smallsetminus \{v\}) \cup e$. Then for every $F \subseteq V$ that contains the carrier of E in Γ', $\mathrm{link}_{\Gamma'_F}(E) = \mathrm{link}_{\Gamma_F}(E') * \partial(e)$. Therefore, we have $h(\mathrm{link}_{\Gamma'_F}(E), x) = (1+x)\, h(\mathrm{link}_{\Gamma_F}(E'), x)$ for every such F, and hence, $\ell_V(\Gamma', E, x) = (1+x)\, \ell_V(\Gamma, E', x)$.

The expressions obtained for $\ell_V(\Gamma', E, x)$ and our assumption on Γ show that, indeed, the corresponding γ-polynomial has nonnegative coefficients in all cases. □

Barycentric and cluster subdivisions. As a special case of Proposition 6.1, the (first) barycentric subdivision of the simplex 2^V has nonnegative local γ-vector. Several combinatorial interpretations for its entries are given in [Athanasiadis and Savvidou 2011]. Similar results appear there for the simplicial subdivision of a simplex defined by the positive part of the cluster complex associated with a finite root system.

The following special case of Conjecture 5.4 might also be interesting to explore. The notion of a CW-regular subdivision can be defined by replacing the simplicial complex Δ' in the definition of a topological subdivision (Definition 2.2) by a regular CW-complex; see [Stanley 1992, p. 839].

Question 6.2. Does Conjecture 5.4 hold for the barycentric subdivision of any CW-regular subdivision of the simplex?

Acknowledgments

The author wishes to thank an anonymous referee for helpful suggestions and Mike Davis and Volkmar Welker for useful discussions.

References

[Athanasiadis 2011] C. A. Athanasiadis, "Some combinatorial properties of flag simplicial pseudomanifolds and spheres", *Ark. Mat.* **49**:1 (2011), 17–29. MR 2784255 Zbl 1235.52020

[Athanasiadis 2012] C. A. Athanasiadis, "Cubical subdivisions and local h-vectors", *Ann. Comb.* **16**:3 (2012), 421–448. MR 2960014

[Athanasiadis and Savvidou 2011] C. A. Athanasiadis and C. Savvidou, "The local h-vector of the cluster subdivision of a simplex", *Sém. Lothar. Combin.* **66** (2011), Article B66c.

[Barmak 2010] J. A. Barmak, "Star clusters in independence complexes of graphs", preprint, 2010. arXiv 1007.0418

[Björner 1995] A. Björner, "Topological methods", pp. 1819–1872 in *Handbook of combinatorics*, vol. 2, edited by R. L. Graham et al., Elsevier, Amsterdam, 1995. MR 96m:52012 Zbl 0851.52016

[Björner et al. 1999] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. M. Ziegler, *Oriented matroids*, 2nd ed., Encyclopedia of Mathematics and its Applications **46**, Cambridge University Press, Cambridge, 1999. MR 2000j:52016 Zbl 0944.52006

[Chan 1994] C. Chan, "On subdivisions of simplicial complexes: characterizing local h-vectors", *Discrete Comput. Geom.* **11**:4 (1994), 465–476. MR 95b:52015 Zbl 0799.05010

[Charney and Davis 1995] R. Charney and M. Davis, "The Euler characteristic of a nonpositively curved, piecewise Euclidean manifold", *Pacific J. Math.* **171**:1 (1995), 117–137. MR 96k:53066 Zbl 0865.53036

[Davis and Okun 2001] M. W. Davis and B. Okun, "Vanishing theorems and conjectures for the ℓ^2-homology of right-angled Coxeter groups", *Geom. Topol.* **5** (2001), 7–74. MR 2002e:58039 Zbl 1118.58300

[Gal 2005] Ś. R. Gal, "Real root conjecture fails for five- and higher-dimensional spheres", *Discrete Comput. Geom.* **34**:2 (2005), 269–284. MR 2006c:52019 Zbl 1085.52005

[Munkres 1984] J. R. Munkres, *Elements of algebraic topology*, Addison-Wesley, Menlo Park, CA, 1984. MR 85m:55001 Zbl 0673.55001

[Postnikov et al. 2008] A. Postnikov, V. Reiner, and L. Williams, "Faces of generalized permutohedra", *Doc. Math.* **13** (2008), 207–273. MR 2010j:05425 Zbl 1167.05005 arXiv math/0609184

[Stanley 1992] R. P. Stanley, "Subdivisions and local h-vectors", *J. Amer. Math. Soc.* **5**:4 (1992), 805–851. MR 93b:52012 Zbl 0768.05100

[Stanley 1996] R. P. Stanley, *Combinatorics and commutative algebra*, 2nd ed., Progress in Mathematics **41**, Birkhäuser, Boston, 1996. MR 98h:05001 Zbl 0838.13008

[Stanley 1997] R. P. Stanley, *Enumerative combinatorics*, vol. 1, Cambridge Studies in Advanced Mathematics **49**, Cambridge University Press, Cambridge, 1997. MR 98a:05001 Zbl 0889.05001

Received January 11, 2012. Revised May 28, 2012.

Christos A. Athanasiadis
Division of Algebra-Geometry
Department of Mathematics
National and Kapodistrian University of Athens
Panepistimioupolis
15784 Athens
Greece
caath@math.uoa.gr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 259, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.259.279

RAYS AND SOULS IN VON MANGOLDT PLANES

IGOR BELEGRADEK, ERIC CHOI AND NOBUHIRO INNAMI

We study rays in von Mangoldt planes, which has applications to the structure of open complete manifolds with lower radial curvature bounds. We prove that the set of souls of any rotationally symmetric plane of nonnegative curvature is a closed ball, and if the plane is von Mangoldt we compute the radius of the ball. We show that each cone in $\mathbb{R}^3$ can be smoothed to a von Mangoldt plane.

1. Introduction

Let M_m denote $\mathbb{R}^2$ equipped with a smooth, complete, rotationally symmetric Riemannian metric given in polar coordinates as $g_m := dr^2 + m^2(r)\, d\theta^2$; let o denote the origin in $\mathbb{R}^2$. We say that M_m is a *von Mangoldt plane* if its sectional curvature $G_m := -m''/m$ is a nonincreasing function of r.

The Toponogov comparison theorem was extended in [Itokawa et al. 2003] to open complete manifolds with radial sectional curvature bounded below by the curvature of a von Mangoldt plane, leading to various applications in [Shiohama and Tanaka 2002; Kondo and Ohta 2007; Kondo and Tanaka 2011] and generalizations in [Mashiko and Shiohama 2006; Kondo and Tanaka 2010; Machigashira 2010].

A point q in a Riemannian manifold is called a *critical point of infinity* if each unit tangent vector at q makes angle $\leq \pi/2$ with a ray that starts at q. Let $\mathfrak{C}_m$ denote the set of critical points of infinity of M_m; clearly $\mathfrak{C}_m$ is a closed, rotationally symmetric subset that contains every pole of M_m, so that $o \in \mathfrak{C}_m$. One reason for studying $\mathfrak{C}_m$ is the following consequence of the generalized Toponogov theorem of [Itokawa et al. 2003].

Lemma 1.1. *Let $\hat{M}$ be a complete noncompact Riemannian manifold with radial curvature bounded below by the curvature of a von Mangoldt plane M_m, and let $\hat{r}$ and r denote the distance functions to the basepoints $\hat{o}$ and o of $\hat{M}$ and M_m, respectively. If $\hat{q}$ is a critical point of $\hat{r}$, then $\hat{r}(\hat{q})$ is contained in $r(\mathfrak{C}_m)$.*

MSC2010: 53C20, 53C22, 53C45.

Keywords: radial curvature, critical point, von Mangoldt, surface of revolution, ray, nonnegative curvature, soul.

Combined with the critical point theory of distance functions [Grove 1993; Greene 1997, Lemma 3.1; Petersen 2006, §11.1], Lemma 1.1 implies the following.

Corollary 1.2. *In the setting of Lemma 1.1, for any c in $[a, b] \subset r(M_m - \mathfrak{C}_m)$,*

- *the $\hat{r}^{-1}$-preimage of $[a, b]$ is homeomorphic to $\hat{r}^{-1}(a) \times [a, b]$, and the $\hat{r}^{-1}$-preimages of points in $[a, b]$ are all homeomorphic,*
- *the $\hat{r}^{-1}$-preimage of $[0, c]$ is homeomorphic to a compact smooth manifold with boundary, and the homeomorphism maps $\hat{r}^{-1}(c)$ onto the boundary,*
- *if $K \subset \hat{M}$ is a compact smooth submanifold, possibly with boundary, such that $\hat{r}(K) \supset r(\mathfrak{C}_m)$, then $\hat{M}$ is diffeomorphic to the normal bundle of K.*

If M_m is von Mangoldt and $G_m(0) \le 0$, then $G_m \le 0$ everywhere, so every point is a pole, and hence $\mathfrak{C}_m = M_m$ so that Lemma 1.1 yields no information about the critical points of $\hat{r}$. Of course, there are other ways to get this information, as illustrated by classical Gromov's estimate: if M_m is the standard $\mathbb{R}^2$, then the set of critical points of $\hat{r}$ is compact; see, for example, [Greene 1997, p. 109].

The following theorem determines $\mathfrak{C}_m$ when $G_m \ge 0$ everywhere; note that the plane M_m in (i)–(iii) need not be von Mangoldt.

Theorem 1.3. *If $G_m \ge 0$, then*:

(i) *$\mathfrak{C}_m$ is the closed R_m-ball centered at o for some $R_m \in [0, \infty]$.*

(ii) *R_m is positive if and only if $\int_1^\infty m^{-2}$ is finite.*

(iii) *R_m is finite if and only if $m'(\infty) < \frac{1}{2}$.*

(iv) *If M_m is von Mangoldt and R_m is finite, then the equation $m'(r) = \frac{1}{2}$ has a unique solution ρ_m, and the solution satisfies $\rho_m > R_m$ and $G_m(r_m) > 0$.*

(v) *If M_m is von Mangoldt and R_m is finite and positive, then R_m is the unique solution of the integral equation*

$$\int_x^\infty \frac{m(x)\,dr}{m(r)\sqrt{m^2(r) - m^2(x)}} = \pi.$$

Here is a sample application of Theorem 1.3 (iv) and Corollary 1.2:

Corollary 1.4. *Let $\hat{M}$ be a complete noncompact Riemannian manifold with radial curvature from the basepoint $\hat{o}$ bounded below by the curvature of a von Mangoldt plane M_m. If $G_m \ge 0$ and $m'(\infty) < \frac{1}{2}$, then $\hat{M}$ is homeomorphic to the metric ρ_m-ball centered at $\hat{o}$, where ρ_m is the unique solution of $m'(r) = \frac{1}{2}$.*

Theorem 1.3 should be compared with the following results of Tanaka:

- The set of poles in any M_m is a closed metric ball centered at o of some radius R_p in $[0, \infty]$ [Tanaka 1992b, Lemma 1.1].

- $R_p > 0$ if and only if $\int_1^\infty m^{-2}$ is finite and $\liminf_{r\to\infty} m(r) > 0$ [Tanaka 1992a].
- If M_m is von Mangoldt, then R_p is a unique solution of an explicit integral equation [Tanaka 1992a, Theorem 2.1].

It is natural to wonder when the set of poles equals $\mathfrak{C}_m$, and we answer the question when M_m is von Mangoldt.

Theorem 1.5. *If M_m is a von Mangoldt plane, then*:

(a) *If R_p is finite and positive, then the set of poles is a proper subset of the component of $\mathfrak{C}_m$ that contains o.*

(b) $R_p = 0$ *if and only if* $\mathfrak{C}_m = \{o\}$.

Of course $R_p = \infty$ implies $\mathfrak{C}_m = M_m$, but the converse is not true: Theorem 1.11 ensures the existence of a von Mangoldt plane with $m'(\infty) = \frac{1}{2}$ and $G_m \geq 0$, and for this plane $\mathfrak{C}_m = M_m$ by Theorem 1.3, while R_p is finite by Remark 4.7.

We say that a ray γ in M_m *points away from infinity* if γ and the segment $[\gamma(0), o]$ make an angle $< \frac{\pi}{2}$ at $\gamma(0)$. Define $A_m \subset M_m - \{o\}$ as follows: $q \in A_m$ if and only if there is a ray that starts at q and points away from infinity; by symmetry, $A_m \subset \mathfrak{C}_m$.

Theorem 1.6. *If M_m is a von Mangoldt plane, then A_m is open in M_m.*

Any plane M_m with $G_m \geq 0$ has another distinguished subset, namely the set of souls, that is, points produced via the soul construction of Cheeger–Gromoll.

Theorem 1.7. *If $G_m \geq 0$, then $\mathfrak{C}_m$ is equal to the set of souls of M_m.*

Recall that the soul construction takes as input a basepoint in an open complete manifold N of nonnegative sectional curvature and produces a compact totally convex submanifold S without boundary, called a *soul*, such that N is diffeomorphic to the normal bundle to S. Thus if N is contractible, as happens for M_m, then S is a point. The soul construction also gives a continuous family of compact totally convex subsets that starts with S and ends with N, and according to [Mendonça 1997, Proposition 3.7] $q \in N$ is a critical point of infinity if and only if there is a soul construction such that the associated continuous family of totally convex sets drops in dimension at q. In particular, any point of S is a critical point of infinity, which can also be seen directly; see the proof of [Maeda 1974/1975, Lemma 1]. In Theorem 1.7 we prove conversely that every point of $\mathfrak{C}_m$ is a soul; for this M_m need not be von Mangoldt.

In regard to Theorem 1.3 (iii), it is worth mentioning $G_m \geq 0$ implies that m' is nonincreasing, so $m'(\infty)$ exists, and moreover, $m'(\infty) \in [0, 1]$ because $m \geq 0$. As we note in Remark A.5 for any von Mangoldt plane M_m, the limit $m'(\infty)$ exists as a number in $[0, \infty]$. It follows that if $G_m \geq 0$ or if M_m is von Mangoldt, then

M_m admits total curvature, which equals $2\pi(1-m'(\infty))$ and hence takes values in $[-\infty, 2\pi]$; thus $m'(\infty)=\frac{1}{2}$ if and only if M_m has total curvature π. Standard examples of von Mangoldt planes of positive curvature are the one-parametric family of paraboloids, all satisfying $m'(\infty)=0$ [Shiohama et al. 2003, Example 2.1.4], and the one-parametric family of two-sheeted hyperboloids parametrized by $m'(\infty)$, which takes every value in $(0, 1)$ [Shiohama et al. 2003, Example 2.1.4].

A property of von Mangoldt planes, discovered in [Elerath 1980; Tanaka 1992b] and crucial to this paper, is that the cut locus of any $q \in M_m - \{o\}$ is a ray that lies on the meridian opposite q. (If M_m is not von Mangoldt, its cut locus is not fully understood, but it definitely can be disconnected [Tanaka 1992a, p. 266], and known examples of cut loci of compact surfaces of revolution [Gluck and Singer 1979; Sinclair and Tanaka 2006] suggest that it could be complicated.)

As we note in Lemma 3.14, if M_m is a von Mangoldt plane, and if $q \neq o$, then $q \in \mathfrak{C}_m$ if and only if the geodesic tangent to the parallel through q is a ray. Combined with Clairaut's relation this gives the following "choking" obstruction for a point q to belong to $\mathfrak{C}_m$ (see Lemma 3.3):

Proposition 1.8. *If M_m is von Mangoldt and $q \in \mathfrak{C}_m$, then $m'(r_q) > 0$ and $m(r) > m(r_q)$ for $r > r_q$, where r_q is the r-coordinate of q.*

The above proposition is immediate from Lemmas 3.3 and 3.14. We also show in Lemma 3.10 that if M_m is von Mangoldt and $\mathfrak{C}_m \neq o$, then there is ρ such that $m(r)$ is increasing and unbounded on $[\rho, \infty)$.

The following theorem collects most of what we know about $\mathfrak{C}_m$ for a von Mangoldt plane M_m with some negative curvature, where the case $\liminf_{r\to\infty} m(r) = 0$ is excluded because then $\mathfrak{C}_m = \{o\}$ by Proposition 1.8.

Theorem 1.9. *If M_m is a von Mangoldt plane with a point where $G_m < 0$ and such that* $\liminf_{r\to\infty} m(r) > 0$, *then*

(1) *M_m contains a line and has total curvature $-\infty$,*

(2) *if m' has a zero, then neither A_m nor $\mathfrak{C}_m$ is connected,*

(3) *$M_m - A_m$ is a bounded subset of M_m,*

(4) *the ball of poles of M_m has positive radius.*

In Example 6.1 we construct a von Mangoldt plane M_m to which Theorem 1.9 (2) applies. In Example 6.2 we produce a von Mangoldt plane M_m such that neither A_m nor $\mathfrak{C}_m$ is connected while $m' > 0$ everywhere. We do not know whether there is a von Mangoldt plane such that $\mathfrak{C}_m$ has more than two connected components.

Because of Lemma 1.1 and Corollary 1.2, one is interested in subintervals of $(0, \infty)$ that are disjoint from $r(\mathfrak{C}_m)$, as, for example, happens for any interval on which $m' \leq 0$, or for the interval (R_m, ∞) in Theorem 1.3. To this end we prove the following result, which is a consequence of Theorem 6.3.

Theorem 1.10. *Let M_n be a von Mangoldt plane with $G_n \geq 0$, $n(\infty) = \infty$, and such that $n'(x) < \frac{1}{2}$ for some x. Then for any $z > x$ there exists $y > z$ such that if M_m is a von Mangoldt plane with $n = m$ on $[0, y]$, then $r(\mathfrak{C}_m)$ and $[x, z]$ are disjoint.*

In general, if M_m and M_n are von Mangoldt planes with $n = m$ on $[0, y]$, then the sets $\mathfrak{C}_m$ and $\mathfrak{C}_n$ could be quite different. For instance, if M_n is a paraboloid, then $\mathfrak{C}_n = \{o\}$, but by Example 6.2 for any $y > 0$ there is a von Mangoldt M_m with some negative curvature such that $m = n$ on $[0, y]$, and by Theorem 1.9 the set M_m–$\mathfrak{C}_m$ is bounded and $\mathfrak{C}_m$ contains the ball of poles of positive radius.

Basic properties of von Mangoldt planes are described in Appendix A. In particular, in order to construct a von Mangoldt plane with prescribed G_m it suffices to check that 0 is the only zero of the solution of the Jacobi initial value problem (A.7) with $K = G_m$, where G_m is smooth on $[0, \infty)$. Prescribing values of m' is harder. It is straightforward to see that if M_m is a von Mangoldt plane such that m' is constant near infinity, then $G_m \geq 0$ everywhere and $m'(\infty) \in [0, 1]$. We do not know whether there is a von Mangoldt plane with $m' = 0$ near infinity, but all the other values in $(0, 1]$ can be prescribed:

Theorem 1.11. *For every $s \in (0, 1]$ there is $\rho > 0$ and a von Mangoldt plane M_m such that $m' = s$ on $[\rho, \infty)$.*

Thus each cone in $\mathbb{R}^3$ can be smoothed to a von Mangoldt plane, but we do not know how to construct a (smooth) capped cylinder that is von Mangoldt.

Structure of the paper. We collect notations and conventions in Section 2. Properties of von Mangoldt planes are reviewed in Appendix A, while Appendix B contains a calculus lemma relevant to continuity and smoothness of the turn angle. Section 3 contains various results on rays in von Mangoldt planes, including the proofs of Theorem 1.6 and Proposition 1.8. Planes of nonnegative curvature are discussed in Section 4, where Theorems 1.3 and 1.7 are proved. A proof of Theorem 1.11 is in Section 5, and the other results stated in the introduction are proved in Section 6.

2. Notations and conventions

All geodesics are parametrized by arclength. Minimizing geodesics are called *segments*. Let ∂_r and ∂_θ denote the vector fields dual to dr and $d\theta$ on $\mathbb{R}^2$. Given $q \neq o$, denote its polar coordinates by θ_q and r_q. Let γ_q, μ_q, and τ_q denote the geodesics defined on $[0, \infty)$ that start at q in the directions of ∂_θ, ∂_r, and $-\partial_r$, respectively. We refer to $\tau_q|_{(r_q,\infty)}$ as the *meridian opposite* q; note that $\tau_q(r_q) = o$. Also set $\kappa_{\gamma(s)} := \angle(\dot\gamma(s), \partial_r)$.

We write $\dot r$, $\dot\theta$, $\dot\gamma$, and $\dot\kappa$ for the derivatives of $r_{\gamma(s)}$, $\theta_{\gamma(s)}$, $\gamma(s)$, and $\kappa_{\gamma(s)}$ by s, and write m' for dm/dr; similar notations are used for higher derivatives.

Let $\hat{\kappa}(r_q)$ denote the maximum of the angles formed by μ_q and rays emanating from $q \neq o$; let ξ_q denote the ray with $\xi_q(0) = q$ for which the maximum is attained, that is, such that $\kappa_{\xi_q(0)} = \hat{\kappa}(r_q)$.

A geodesic γ in M_m–$\{o\}$ is called *counterclockwise* if $\dot{\theta} > 0$ and *clockwise* if $\dot{\theta} < 0$. A geodesic in M_m is clockwise, counterclockwise, or can be extended to a geodesic through o. If γ is clockwise, then it can be mapped to a counterclockwise geodesic by an isometric involution of M_m.

Convention. Unless stated otherwise, any geodesic in M_m that we consider is either tangent to a meridian or counterclockwise.

Due to this convention the Clairaut constant and the turn angle defined below are nonnegative, which will simplify notations.

3. Turn angle and rays in M_m

This section collects what we know about rays in M_m with emphasis on the cases when $G_m \geq 0$ or $G'_m \leq 0$. Let γ be a geodesic in M_m that does not pass through o, so that γ is a solution of the geodesic equations

$$\ddot{r} = mm'\dot{\theta}^2, \quad \dot{\theta}m^2 = c, \tag{3.1}$$

where c is called the *Clairaut constant of* γ. The equation $\dot{\theta}m^2 = c$ is the so-called *Clairaut's relation*, which, since γ is assumed counterclockwise, can be written as $c = m(r_{\gamma(s)}) \sin \kappa_{\gamma(s)}$. Thus $0 \leq c \leq m(r_\gamma(s))$ where $c = m(r_\gamma(s))$ only at points where γ is tangent to a parallel, and $c = 0$ when γ is tangent to a meridian.

A geodesic is called *escaping* if its image is unbounded; for example, any ray is escaping.

Fact 3.2. (1) A parallel through q is a geodesic in M_m if and only if $m'(r_q) = 0$ [Shiohama et al. 2003, Lemma 7.1.4].

(2) A geodesic γ in M_m is tangent to a parallel at $\gamma(s_0)$ if and only if $\dot{r}_{\gamma(s_0)} = 0$.

(3) If γ is a geodesic in M_m and $\dot{r}_{\gamma(s)}$ vanishes more than once, then γ is invariant under a rotation of M_m about o [Shiohama et al. 2003, Lemma 7.1.6] and hence not escaping.

Lemma 3.3. *If γ_q is escaping, then $m(r) > m(r_q)$ for $r > r_q$, and $m'(r_q) > 0$.*

Proof. Since γ_q is escaping, the image of $s \to r_{\gamma_q}(s)$ contains $[r_q, \infty)$, and q is the only point where γ_q is tangent to a parallel. The Clairaut constant of γ_q is $c = m(r_q)$, hence $m(r) > m(r_q)$ for all $r > r_q$. It follows that $m'(r_q) \geq 0$. Finally, $m'(r_q) \neq 0$ else γ_q would equal the parallel through q. □

Lemma 3.4. *If γ is an escaping geodesic that is tangent to the parallel P_q through q, then $\gamma \setminus \{q\}$ lies in the unbounded component of $M_m \setminus P_q$.*

Proof. By reflectional symmetry and uniqueness of geodesics, γ locally stays on the same side of the parallel P_q through q, that is, γ is the union of γ_q and its image under the reflecting fixing $\mu_q \cup \tau_q$. If γ could cross to the other side of P_q at some point $\gamma(s)$, then $|r_{\gamma(s)} - r_q|$ would attain a maximum between $\gamma(s)$ and q, and at the maximum point γ would be tangent to a parallel. Since γ is escaping, it cannot be tangent to parallels more than once, hence γ stays on the same side of P_q at all times, and since γ is escaping, it stays in the unbounded component of $M_m \setminus P_q$. □

For a geodesic $\gamma : (s_1, s_2) \to M_m$ that does not pass through o, we define the *turn angle* T_γ *of* γ as

$$T_\gamma := \int_\gamma d\theta = \int_{s_1}^{s_2} \dot\theta_{\gamma(s)}\, ds = \theta_{\gamma(s_2)} - \theta_{\gamma(s_1)}.$$

Clairaut's relation reads $\dot\theta = c/m^2 \geq 0$ so the above integral T_γ converges to a number in $[0, \infty]$. Since γ is unit speed, we have $(\dot r)^2 + m^2\dot\theta^2 = 1$. Combining this with $\dot\theta = c/m^2$ gives

$$\dot r = \operatorname{sign}(\dot r)\sqrt{1 - \frac{c^2}{m^2}},$$

which yields a useful formula for the turn angle: if γ is not tangent to a meridian or a parallel on (s_1, s_2), so that $\operatorname{sign}(\dot r_{\gamma(s)})$ is a nonzero constant, then

$$\frac{d\theta}{dr} = \frac{\dot\theta}{\dot r} = \operatorname{sign}(\dot r_{\gamma(s)}) F_c(r) \quad \text{where} \quad F_c := \frac{c}{m\sqrt{m^2 - c^2}}, \tag{3.5}$$

and thus if $r_i := r_{\gamma(s_i)}$, then

$$T_\gamma = \operatorname{sign}(\dot r) \int_{r_1}^{r_2} F_c(r)\, dr. \tag{3.6}$$

Since $c^2 \leq m^2$, this integral is finite except possibly when some r_i is in the set $\{m^{-1}(c), \infty\}$. The integral (3.6) converges at $r_i = m^{-1}(c)$ if and only if $m'(r_i) \neq 0$. Convergence of (3.6) at $r_i = \infty$ implies convergence of $\int_1^\infty m^{-2}\, dr$, and the converse holds under the assumption $\liminf_{r\to\infty} m(r) > c$; this assumption is true when $G_m \geq 0$ or $G'_m \leq 0$, as follows from Lemma 3.10.

Example 3.7. If γ is a ray in M_m that does not pass through o, then $T_\gamma \leq \pi$ else there is s with $|\theta_{\gamma(s)} - \theta_{\gamma(0)}| = \pi$, and by symmetry the points $\gamma(s)$ and $\gamma(0)$ are joined by two segments, so γ would not be a ray.

Example 3.8. If T_{γ_q} is finite, then $m'(r_q) \neq 0$ and m^{-2} is integrable on $[1, \infty)$, as follows immediately from the discussion preceding Example 3.7.

Lemma 3.9. *If $\gamma : [0, \infty) \to M_m$ is a geodesic with finite turn angle, then γ is escaping.*

Proof. Note that γ is tangent to parallels in at most two points, for otherwise γ is invariant under a rotation about o, and hence its turn angle is infinite. Thus after cutting off a portion of γ we may assume it is never tangent to a parallel, so that $r_{\gamma(s)}$ is monotone. By assumption $\theta_{\gamma(s)}$ is bounded and increasing. By Clairaut's relation $m(r_{\gamma(s)})$ is bounded below, so that $m(0) = 0$ implies that $r_{\gamma(s)}$ is bounded below. If γ were not escaping, then $r_{\gamma(s)}$ would also be bounded above, so there would exist a limit of $(r_{\gamma(s)}, \theta_{\gamma(s)})$ and hence the limit of $\gamma(s)$ as $s \to \infty$, contradicting the fact that γ has infinite length. □

Lemma 3.10. *If m^{-2} is integrable on $[1, \infty)$, then*

(1) *the function $(r \log r)^{-\frac{1}{2}} m(r)$ is unbounded,*

(2) *if $G_m \geq 0$, then $m' > 0$ for all r,*

(3) *if M_m is von Mangoldt, then $m' > 0$ for all large r,*

(4) *if either $G_m \geq 0$ or $G'_m \leq 0$, then $m(\infty) = \infty$.*

Proof. Since m^{-2} is integrable, the function $(r \log r)^{-\frac{1}{2}} m(r)$ is unbounded, and in particular, m is unbounded. If $G_m \geq 0$ everywhere, then m' is nonincreasing with $m'(0) = 1$, and the fact that m is unbounded implies that $m' > 0$ for all r. If M_m is von Mangoldt, and $G_m(\rho_0) < 0$, then $G_m < 0$ for $r \geq \rho_0$, that is, m' is nondecreasing on $[\rho_0, \infty)$. Since m is unbounded, there is $\rho > \rho_0$ with $m(\rho) > m(\rho_0)$ so that $\int_{\rho_0}^{\rho} m' = m(\rho) - m(\rho_0) > 0$. Hence m' is positive somewhere on (ρ_0, ρ), and therefore on $[\rho, \infty)$. Finally, since m is an unbounded increasing function for large r, the limit $\lim_{r\to\infty} m(r) = m(\infty)$ exists and equals ∞. □

Lemma 3.11. *If γ_q is escaping, then $\liminf_{r\to\infty} m(r) > m(r_q)$ if and only if there is a neighborhood U of q such that γ_u is escaping for each $u \in U$.*

Proof. First, recall that $m(r) > m(r_q)$ for $r > r_q$ and $m'(r_q) > 0$ by Lemma 3.3. We shall prove the contrapositive: $\liminf_{r\to\infty} m(r) = m(r_q)$ if and only if there is a sequence $u_i \to q$ such that γ_{u_i} is not escaping.

If there is a sequence $z_i \in M_m$ with $r_{z_i} \to \infty$ and $m(r_{z_i}) \to m(r_q)$, then there are points $u_i \to q$ on μ_q with $m(r_{u_i}) = m(r_{z_i})$. If γ_{u_i} is escaping, then it meets the parallel through z_i, so Clairaut's relation implies that γ_{u_i} is tangent to the parallels through u_i and z_i, which cannot happen for an escaping geodesic.

Conversely, suppose there are $u_i \to q$ such that $\gamma_i := \gamma_{u_i}$ is not escaping. Let R_i be the radius of the smallest ball about o that contains γ_i, and let P_i be its boundary parallel. Note that $R_i \to \infty$ as γ_i converges to γ_q on compact sets and γ_q is escaping, and hence $\liminf_{r\to\infty} m(r) = \lim_{r\to\infty} m(R_i)$. For each i there is a sequence $s_{i,j}$ such that the r-coordinates of $\gamma_i(s_{i,j})$ converge to R_i, which implies

$\kappa_{\gamma_i(s_{i,j})} \to \pi/2$ as $j \to \infty$ and i is fixed. (Note that if γ_i is tangent to P_i, then $s_{i,j}$ is independent of j, namely, $\gamma(s_{i,j})$ is the point of tangency.) By Clairaut's relation, $m(R_i) = m(r_{u_i})$, hence $\liminf_{r\to\infty} m(r) = m(r_q)$. □

Lemma 3.12. *If M_m is von Mangoldt, then a geodesic $\gamma : [0, \infty) \to M_m \setminus \{o\}$ is a ray if and only if $T_\gamma \le \pi$.*

Proof. The "only if" direction holds even when M_m is not von Mangoldt by Example 3.7. Conversely, if γ is not a ray, then γ meets the cut locus of q, which by [Tanaka 1992b] is a subset of the opposite meridian $\tau_{\gamma(0)}|_{(r_{\gamma(0)},\infty)}$. Thus $T_\gamma > \pi$. □

Lemma 3.13. *If γ is a ray in a von Mangoldt plane, and if σ is a geodesic with $\sigma(0) = \gamma(0)$ and $\kappa_{\gamma(0)} > \kappa_{\sigma(0)}$, then σ is a ray and $T_\sigma \le T_\gamma$.*

Proof. Set $q = \gamma(0)$. If $\kappa_{\gamma(0)} = \pi$, then $\gamma = \tau_q$, so τ_q is a ray, which in a von Mangoldt plane implies that q is a pole [Shiohama et al. 2003, Lemma 7.3.1], so that σ is also a ray. If $\kappa_{\gamma(0)} < \pi$ and σ is not a ray, then σ is minimizing until it crosses the opposite meridian $\tau_q|_{(r_q,\infty)}$ [Tanaka 1992b]. Near q the geodesic σ lies in the region of M_m bounded by γ and μ_q hence before crossing the opposite meridian σ must intersect γ or μ_q, so they would not be rays. Finally, $T_\sigma \le T_\gamma$ holds as σ lies in the sector between γ and μ_q. □

Lemma 3.14. *If M_m is von Mangoldt and $q \ne o$, then γ_q is a ray if and only if $q \in \mathfrak{C}_m$.*

Proof. If γ_q is a ray, then $q \in \mathfrak{C}_m$ by symmetry. If $q \in \mathfrak{C}_m$, then either q is a pole and there is a ray in any direction, or q is not a pole. In the latter case τ_q is not a ray [Shiohama et al. 2003, Lemma 7.3.1], hence by the definition of $\mathfrak{C}_m$ there is a ray γ with $\kappa_{\gamma(0)} \ge \pi/2$, so γ_q is a ray by Lemma 3.13. □

Recall that $\hat{\kappa}(r_q)$ is the maximum of the angles formed by μ_q and rays emanating from $q \ne o$, and ξ_q is the ray for which the maximum is attained. It is immediate from definitions that $q \in \mathfrak{C}_m$ if and only if $\hat{\kappa}(r_q) \ge \pi/2$. Lemmas 3.15, 3.16, and 3.17 were suggested by the referee.

Lemma 3.15. *$\mathfrak{C}_m \ne \{o\}$ if and only if $\liminf_{r\to\infty} m > 0$ and $\int_1^\infty m^{-2}$ is finite.*

Proof. The "if" direction holds because by the main result of [Tanaka 1992a] the assumptions imply that the ball of poles has a positive radius. Conversely, if $q \in \mathfrak{C}_m - \{o\}$, then ξ_q is a ray different from μ_q. By [Tanaka 1992a, Lemma 1.3 and Proposition 1.7] if either $\liminf_{r\to\infty} m = 0$ or $\int_1^\infty m^{-2} = \infty$, then μ_q is the only ray emanating from q. □

Lemma 3.16. *The limit of the segments $[q, \tau_q(s)]$ as $s \to \infty$ is ξ_q.*

Proof. The segments $[q, \tau_q(s)]$ subconverge to a ray σ that starts at q. Since ξ_q is a ray, it cannot cross the opposite meridian $\tau_q|_{(r_q,\infty)}$. As $[q, \tau_q(s)]$ and ξ_q are minimizing, they only intersect at q, and hence the angle formed by μ_q and $[q, \tau_q(s)]$ is $\geq \hat{\kappa}(r_q)$. It follows that $\kappa_{\sigma(0)} \geq \hat{\kappa}(r_q)$, which must be an equality as $\hat{\kappa}(r_q)$ is a maximum, so $\sigma = \xi_q$. □

Lemma 3.17. *The function $r \to \hat{\kappa}(r)$ is left continuous and upper semicontinuous. In particular, the set $\{q : \hat{\kappa}(r_q) < \alpha\}$ is open for every α.*

Proof. If $\hat{\kappa}$ is not left continuous at r_q, then there exists $\varepsilon > 0$ and a sequence of points q_i on μ_q such that $r_{q_i} \to r_q-$ and either $\hat{\kappa}(r_{q_i}) - \hat{\kappa}(r_q) > \varepsilon$ or $\hat{\kappa}(r_q) - \hat{\kappa}(r_{q_i}) > \varepsilon$. In the former case ξ_{q_i} subconverge to a ray that makes a larger angle with μ_q than ξ_q, contradicting the maximality of $\hat{\kappa}(r_q)$. In the latter case, ξ_{q_i} intersects ξ_q for some i. Therefore, by Lemma 3.16 the segment $[q_i, \tau_q(s)]$ intersects $[q, \tau_q(s)]$ for large enough s at a point $z \neq \tau_q(s)$, so $\tau_q(s)$ is a cut point of z which cannot happen for a segment. This proves that $\hat{\kappa}$ is left continuous. A similar argument shows that

$$\limsup_{r_{q_i} \to r_q^+} \hat{\kappa}(r_{q_i}) \leq \hat{\kappa}(r_q),$$

so that $\hat{\kappa}$ is upper semicontinuous, which implies that $\{q : \hat{\kappa}(r_q) < \alpha\}$ is open for every α. □

Lemmas 3.12 and 3.14 imply that on a von Mangoldt plane $\hat{\kappa}(r_q) \geq \pi/2$ if and only if $T_{\gamma_q} \leq \pi$; the equivalence is sharpened in Theorem 3.24, whose proof occupies the rest of this section.

Lemma 3.18. *If σ is escaping and $0 < \kappa_{\sigma(0)} \leq \pi/2$, then $T_\sigma = \int_{r_q}^{\infty} F_c(r)\, dr$; moreover, if $\kappa_{\sigma(0)} = \pi/2$, then $c = m(r_q)$.*

Proof. This formula for T_σ is immediate from (3.6) once it is shown that $\sigma|_{(0,\infty)}$ is not tangent to a meridian or a parallel. If $\sigma|_{(0,\infty)}$ were tangent to a meridian, $\kappa_{\sigma(0)}$ would be 0 or π, which is not the case. Since σ is escaping, Fact 3.2 implies that σ is tangent to a parallel at most once; that is, $\dot{r}_\sigma$ has at most one zero. If $\kappa_{\sigma(0)} = \pi/2$, then σ is tangent to the parallel through $\sigma(0)$, and so $\sigma|_{(0,\infty)}$ is not tangent to a parallel. Finally, if $\kappa_{\sigma(0)} < \pi/2$, then σ is not tangent to a parallel, else it would be tangent to a parallel through u with $r_u > r_q$, which would imply $r_{\sigma(s)} \leq r_u$ for all s by Lemma 3.4, which cannot happen for an escaping geodesic. □

To better understand the relationship between $\hat{\kappa}(r_q)$ and T_{γ_q}, we study how T_σ depends on σ, or equivalently on $\sigma(0)$ and $\kappa_{\sigma(0)}$, when σ varies in a neighborhood of a ray γ_q.

Lemma 3.19. *If $G_m \geq 0$ or $G'_m \leq 0$, then the function $u \to T_{\gamma_u}$ is continuous at each point u where T_{γ_u} is finite.*

Proof. If T_{γ_u} is finite, then γ_u is escaping by Lemma 3.9, and hence $T_{\gamma_u} = \int_{r_u}^{\infty} F_{m(r_u)}$ by Lemma 3.18. We need to show that this integral depends continuously on r_u.

By Lemmas 3.3 and 3.10 and the discussion preceding Example 3.7, the assumptions on G_m and the finiteness of T_{γ_u} imply that $m(r) > m(r_u)$ for $r > r_u$, m^{-2} is integrable, $m'(r_u) > 0$, and $m(\infty) = \infty$. Hence there exists $\delta > r_u$ with $m'|_{[r_u,\delta]} > 0$, and $m(r) > m(\delta)$ for $r > \delta$; it is clear that small changes in u do not affect δ.

Write $\int_{r_u}^{\infty} F_{m(r_u)} = \int_{r_u}^{\delta} F_{m(r_u)} + \int_{\delta}^{\infty} F_{m(r_u)}$. On $[r_u, \delta]$ we can write $F_{m(r_u)} = h(r, r_u)(r - r_u)^{-1/2}$ for some smooth function h. Since $(r - r_u)^{-1/2}$ is the derivative of $2(r - r_u)^{1/2}$, one can integrate $F_{m(r_u)}$ by parts which easily implies continuous dependence of $\int_{r_u}^{\delta} F_{m(r_u)}$ on r_u.

Continuous dependence of $\int_{\delta}^{\infty} F_{m(r_u)}$ on r_u follows because $F_{m(r_u)}$ is continuous in r_u, and is dominated by Km^{-2} where K is a positive constant independent of small changes of r_u. □

Next we focus on the case when $\sigma(0)$ is fixed, while $\kappa_{\sigma(0)}$ varies near $\pi/2$. To get an explicit formula for T_σ we need the following.

Lemma 3.20. *If M_m is von Mangoldt, and γ_q is a ray, then there is $\varepsilon > 0$ such that every geodesic $\sigma : [0, \infty) \to M_m$ with $\sigma(0) = q$ and $\kappa_{\sigma(0)} \in [\pi/2, \pi/2 + \varepsilon]$ is tangent to a parallel exactly once, and if u is the point where σ is tangent to a parallel, then $m' > 0$ on $[r_u, r_q]$.*

Proof. If $\kappa_{\sigma(0)} = \pi/2$, then $\sigma = \gamma_q$, so it is tangent to a parallel only at q, as rays are escaping. If $\kappa_{\sigma(0)} > \pi/2$, then σ converges to γ_q on compact subsets as $\varepsilon \to 0$, so for a sufficiently small ε the geodesic σ crosses the parallel through q at some point $\sigma(s)$ such that $\kappa_{\sigma(s)} < \pi/2$. Since γ_q is a ray, rotational symmetry and Lemma 3.13 imply that $\sigma|_{[s,\infty)}$ is a ray, so σ is escaping. Thus σ is tangent to a parallel at a point u where $r_{\sigma(s)}$ attains a minimum, and is not tangent to a parallel at any other point by Fact 3.2. Finally, $r_u = \lim_{\varepsilon\to 0} r_q$, and since $m'(r_q) > 0$ by Proposition 1.8, we get $m' > 0$ on $[r_u, r_q]$ for small ε. □

Under the assumptions of Lemma 3.20 the Clairaut constant c of σ equals $m(r_u) = m(r_q) \sin \kappa_{\sigma(0)}$, and the turn angle of σ is given by

$$(3.21) \qquad T_\sigma = \int_{r_q}^{\infty} F_{m(r_q)}(r)\, dr \quad \text{if} \quad \kappa_{\sigma(0)} = \frac{\pi}{2} \quad \text{and}$$

$$(3.22) \qquad T_\sigma = \int_{r_u}^{\infty} F_c(r)\, dr - \int_{r_q}^{r_u} F_c(r)\, dr = \int_{r_q}^{\infty} F_c(r)\, dr + 2\int_{r_u}^{r_q} F_c(r)\, dr$$

if $\pi/2 < \kappa_{\sigma(0)} < \pi/2 + \varepsilon$. These integrals converge, that is, T_σ is finite, as follows from Example 3.8 and Lemmas 3.10 and 3.20.

Since any geodesic σ with $\sigma(0) = q$ and $\kappa_{\sigma(0)} \in [0, \pi/2 + \varepsilon]$ has finite turn angle, one can think of T_σ as a function of $\kappa_{\sigma(0)}$ where σ varies over geodesics with $\sigma(0) = q$ and $\kappa_{\sigma(0)} \in [0, \pi/2 + \varepsilon]$.

Lemma 3.23. *If M_m is von Mangoldt, and γ_q is a ray, then there is $\delta > \pi/2$ such that the function $\kappa_{\sigma(0)} \to T_\sigma$ is continuous and strictly increasing on $[\pi/2, \delta]$, and continuously differentiable on $(\pi/2, \delta]$; moreover, the derivative of T_σ is infinite at $\pi/2$.*

Proof. The Clairaut constant c of σ equals $m(r_u) = m(r_q)\sin\kappa_{\sigma(0)}$, so the assertion is immediate from (elementary but nontrivial) Lemma B.2 about continuity and differentiability of the integrals (3.21) and (3.22). □

Theorem 3.24. *If M_m is von Mangoldt and $q \neq o$, then*

(1) $\hat{\kappa}(r_q) > \pi/2$ *if and only if* $T_{\gamma_q} < \pi$,

(2) $\hat{\kappa}(r_q) = \pi/2$ *if and only if* $T_{\gamma_q} = \pi$.

Proof. (1) If $\hat{\kappa}(r_q) > \pi/2$, then any geodesic σ with $\sigma(0) = q$ and $\kappa_{\sigma(0)} \in [\pi/2, \hat{\kappa}(r_q)]$ is a ray, and so has turn angle $\le \pi$. By Lemma 3.23 the turn angle is increasing at $\pi/2$, so $T_{\gamma_q} < \pi$. Conversely, if $T_{\gamma_q} < \pi$, then by Lemma 3.23 the turn angle is continuous at $\pi/2$, so any geodesic σ with $\sigma(0) = q$ and $\kappa_{\sigma(0)}$ near $\pi/2$ has turn angle $< \pi$, and is therefore a ray, so $\hat{\kappa}(r_q) > \pi/2$.

(2) This follows from (1) and the fact that $\hat{\kappa}(r_q) \ge \pi/2$ if and only if $T_{\gamma_q} \le \pi$. □

Proof of Theorem 1.6. By Theorem 3.24 we know that $q \in A_m$ if and only if $T_{\gamma_q} < \pi$, and by Lemma 3.19 the map $u \to T_{\gamma_u}$ is continuous at q, so the set $\{u \in M_m \mid T_{\gamma_u} < \pi\}$ is open, and hence so is A_m. □

Another proof of Theorem 1.6. Fix $q \in A_m$ so that $T_{\gamma_q} < \pi$ by Theorem 3.24. Fix $\varepsilon > 0$ such that $\varepsilon + T_{\gamma_q} < \pi$. Let P_q be the parallel through q. Then there is a ray γ with $\gamma(0) = q$ and $\kappa_{\gamma(0)} > \pi/2$ such that γ intersects P_q at points q and $\gamma(t)$, and the turn angle of $\gamma|_{(0,t)}$ is $< \varepsilon$.

For an arbitrary sequence $q_i \to q$ we need to show that $q_i \in A_m$ for all large i. Let $\gamma_i : [0, \infty) \to M_m$ be the geodesic with $\gamma_i(0) = q_i$ and $\kappa_{\gamma_i(0)} = \kappa_{\gamma(0)}$. Since γ_i converge to γ on compact sets, for large i there are $t_i > 0$ such that $\gamma_i(t_i) \in P_q$ and $t_i \to t$. The angle formed by γ and $\mu_{\gamma(t)}$ is $< \pi/2$. Rotational symmetry and Lemma 3.13 imply that if i is large, then $\gamma_i|_{[t_i,\infty)}$ is a ray whose turn angle is $\le T_{\gamma_q}$. The turn angles of $\gamma_i|_{(0,t_i)}$ converge to the turn angle of $\gamma|_{(0,t)}$, which is $< \varepsilon$. Thus $T_{\gamma_i} < T_{\gamma_q} + \varepsilon < \pi$ for large i, so that γ_i is a ray by Lemma 3.12, and hence $q_i \in A_m$. □

4. Planes of nonnegative curvature

A key consequence of $G_m \ge 0$ is monotonicity of the turn angle and of $\hat{\kappa}$.

Proposition 4.1. *Suppose that M_m has $G_m \ge 0$. If $0 < r_u < r_v$ and γ_u has finite turn angle, then $T_{\gamma_u} \le T_{\gamma_v}$ with equality if and only if G_m vanishes on $[r_u, \infty]$.*

Proof. The result is trivial when G is everywhere zero. Since γ_u has finite turn angle, m^{-2} is integrable, and hence m is a concave function with $m' > 0$ and $m(\infty) = \infty$ by Lemma 3.10.

Set $x := r_q$, so that the turn angle of γ_q is $\int_x^\infty F_{m(x)}$. As $m' > 0$, we can change variables by $t := m(r)/m(x)$ or $r = m^{-1}(tm(x))$ so that

$$\int_x^\infty F_{m(x)}(r)\,dr = \int_1^{m(\infty)/m(x)} \frac{dt}{l(t,x)t\sqrt{t^2-1}} = \int_1^\infty \frac{dt}{l(t,x)t\sqrt{t^2-1}},$$

where $l(t,x) := m'(r)$. Computing

$$\frac{\partial l(t,x)}{\partial x} = m''(r)\frac{\partial r}{\partial x} = \frac{m''(r)tm'(x)}{m'(r)} = -G(r)\frac{tm'(x)}{m'(r)} \le 0$$

we see that $l(t,x)$ is nonincreasing in x. Thus if $r_u < r_v$, then $l(t,r_u) \ge l(t,r_v)$ for all t implying $T_{\gamma_u} \le T_{\gamma_v}$. The equality occurs precisely when $l(t,x)$ is constant on $[1,\infty)\times[r_u,r_v]$, or equivalently, when $G(m^{-1}(tm(x)))$ vanishes on $[1,\infty)\times[r_u,r_v]$, which in turn is equivalent to $G = 0$ on $[r_u,\infty)$, because $tm(x)$ takes all values in $(m(r_u),\infty)$ so $m^{-1}(tm(x))$ takes all values in (r_u,∞). □

Lemma 4.2. *If $G_m \ge 0$, then $\hat{\kappa}$ is nonincreasing in r.*

Proof. Let u_1, u_2, and v be points on μ_v with $0 < r_{u_1} < r_{u_2} < r_v$. By Lemma 3.16 the ray ξ_{u_i} is the limit of geodesic segments that join u_i with points $\tau_v(s)$ as $s \to \infty$. The segments $[u_1, \tau_v(s)]$ and $[u_2, \tau_v(s)]$ only intersect at the endpoint $\tau_v(s)$ for if they intersect at a point z, then z is a cut point for $\tau_v(s)$, so $[\tau_v(s), u_i]$ cannot be minimizing. Hence the geodesic triangle with vertices u_1, v, and $\tau_v(s)$ contains the geodesic triangle with vertices u_2, v, and $\tau_v(s)$. Since $G_m \ge 0$, the former triangle has larger total curvature, which is finite as M_m has finite total curvature. As m only vanishes at 0, concavity of m implies that m is nondecreasing.

If m is unbounded, Clairaut's relation implies that the angles at $\tau_v(s)$ tend to zero as $s \to \infty$. By the Gauss–Bonnet theorem $\kappa_{\xi_1(0)} - \kappa_{\xi_2(0)}$ equals the total curvature of the "ideal" triangle with sides ξ_1, ξ_2, and $[u_1, u_2]$. Thus $\hat{\kappa}(r_{u_1}) \ge \hat{\kappa}(r_{u_2})$ with equality if and only if G_m vanishes on $[r_{u_1}, \infty)$.

If m is bounded, then $\int_1^\infty m^{-2} = \infty$, so by [Tanaka 1992a, Proposition 1.7] the only ray emanating from q is μ_q so that $\hat{\kappa} = 0$ on $M_m \setminus \{o\}$. For future use note that in this case the angle formed by $\mu_q = \xi_q$ and $[q, \tau_q(s)]$ tends to zero as $s \to \infty$, so Clairaut's relation together with the boundedness of m imply that the angle at $\tau_q(s)$ in the bigon with sides $[q, \tau_q(s)]$ and τ_q also tends to zero as $s \to \infty$. □

Remark 4.3. By the above proof if $G_m \ge 0$ and m^{-2} is integrable on $[1,\infty)$, then $\hat{\kappa}(r_1) = \hat{\kappa}(r_2)$ for some $r_2 > r_1$ if and only if G_m vanishes on $[r_1, \infty)$.

Proof of Theorem 1.3. (i) Since rays converge to rays, $\mathfrak{C}_m$ is closed. As $o \in \mathfrak{C}_m$, rotational symmetry and Lemma 4.2 imply that $\mathfrak{C}_m$ is a closed ball.

(ii) Since m is concave and positive, it is nondecreasing, so $\liminf_{r\to\infty} m > 0$, and the claim follows from Lemma 3.15.

(iii) We prove the contrapositive that $M_m = \mathfrak{C}_m$ if and only if $m'(\infty) \geq \frac{1}{2}$. Note that the latter is equivalent to $c(M_m) \leq \pi$, where $c(Z)$ denotes the total curvature of a subset $Z \subseteq M_m$ which varies in $[0, 2\pi]$.

Suppose $c(M_m) \leq \pi$. Fix $q \neq o$, and consider the segments $[q, \tau_q(s)]$ that by Lemma 3.16 converge to ξ_q as $s \to \infty$. Consider the bigon bounded by $[q, \tau_q(s)]$ and its symmetric image under the reflection that fixes $\tau_q \cup \mu_q$. As in the proof of Lemma 4.2 we see that the angle at $\tau_q(s)$ goes to zero as $s \to \infty$, so the sum of angles in the bigon tends to $2(\pi - \hat{\kappa}(r_q))$, which by the Gauss–Bonnet theorem cannot exceed $c(M_m) \leq \pi$. We conclude that $\hat{\kappa}(r_q) \geq \pi/2$, so $q \in \mathfrak{C}_m$.

Conversely, suppose that $\mathfrak{C}_m = M_m$. Given $\varepsilon > 0$ find a compact rotationally symmetric subset $K \subset M_m$ with $c(K) > c(M_m) - \varepsilon$. Fix $q \neq o$ and consider the rays $\xi_{\mu_q(s)}$ as $s \to \infty$. If all these rays intersect K, then they subconverge to a line [Shiohama et al. 2003, Lemma 6.1.1], so by the splitting theorem M_m is the standard $\mathbb{R}^2$, and $c(M_m) = 0 < \pi$. Thus we can assume that there is v on the ray μ_q such that ξ_v is disjoint from K. Therefore, if s is large enough, then K lies inside the bigon bounded by $[v, \tau_v(s)]$ and its symmetric image under the reflection that fixes $\tau_q \cup \mu_q$. The sum of angles in the bigon tends to $2(\pi - \hat{\kappa}(r_v))$, and by the Gauss–Bonnet theorem it is bounded below by $c(K)$. Since $v \in \mathfrak{C}_m$, we have $\hat{\kappa}(r_v) \geq \pi/2$, and hence $c(K) \leq \pi$. Thus $c(M_m) < \pi + \varepsilon$, and since ε is arbitrary, we get $c(M_m) \leq \pi$, which completes the proof of (iii).

(iv) Since R_m is finite, $m'(\infty) < \frac{1}{2}$ by (iii). As $m'(0) = 1$, the equation $m'(x) = \frac{1}{2}$ has a solution ρ_m. As $G_m \geq 0$, the function m' is nonincreasing, so uniqueness of the solution is equivalent to positivity of $G_m(\rho_m)$. Since M_m is von Mangoldt, $G_m(\rho_m) > 0$ for otherwise G_m would have to vanish for $r \geq \rho_m$, implying $m'(\infty) = m'(\rho_m) = \frac{1}{2}$, so R_m would be infinite.

Now we show that $\rho_m > R_m$. This is clear if $R_m = 0$ because $\rho_m \geq 0$ and $m'(0) = 1 \neq \frac{1}{2} = m'(\rho_m)$. Suppose $R_m > 0$. Then m^{-2} is integrable by Lemma 3.15, so $m' > 0$ everywhere by the proof of Lemma 3.10. Hence for any $r_v \geq \rho_m$ we have $m(r_v) \geq m(\rho_m)$, which implies $tm(r_v) > m(\rho_m)$ for all $t > 1$. Thus $m^{-1}(tm(r_v)) > m^{-1}(m(\rho_m)) = \rho_m$. Applying m' to the inequality, we get in notations of Proposition 4.1 that $l(t, r_v) < m'(\rho_m) = \frac{1}{2}$, where the inequality is strict because $G_m(r_m) > 0$ by (iv). Now (4.5) below implies

$$T_{\gamma_v} = \int_1^\infty \frac{dt}{l(t, r_v)\, t\sqrt{t^2 - 1}} > \int_1^\infty \frac{2\, dt}{t\sqrt{t^2 - 1}} = \pi.$$

Since M_m is von Mangoldt, $v \notin \mathfrak{C}_m$ by Lemma 3.14. In summary, if $r_v \geq \rho_m$, then $v \notin \mathfrak{C}_m$, so $\rho_m > R_m$.

(v) Since R_m is positive and finite, and M_m is von Mangoldt, there are geodesics tangent to parallels whose turn angles are $\le \pi$ and $> \pi$. By Proposition 4.1, the turn angle is monotone with respect to r, so let r_q be the (finite) supremum of all x such that $\int_x^\infty F_{m(x)} < \pi$. Since $\mathfrak{C}_m$ is closed, $q \in \mathfrak{C}_m$ so that $T_{\gamma_q} \le \pi$. In fact, $T_{\gamma_q} = \pi$ for if $T_{\gamma_q} < \pi$, then r_q is not maximal because by Theorems 1.6 and 3.24 the set of points q with $T_{\gamma_q} < \pi$ is open in M_m. If $G_m(r_q) > 0$, then by monotonicity r_q is a unique solution of $T_{\gamma_q} = \pi$. If $G_m(r_q) = 0$, then $G_m|_{[r_q,\infty)} = 0$ as M_m is von Mangoldt, so (4.5) implies that the turn angle of each γ_v with $r_v \ge r_q$ equals $\pi/(2m'(r_q))$. So $m'(r_q) = \frac{1}{2}$ but this case cannot happen as R_m is infinite by (iii). □

In preparation for a proof of Theorem 1.7 we recall that the Cheeger–Gromoll soul construction with basepoint q, described, for example, in [Sakai 1996, Theorem V.3.4], starts by deleting the horoballs associated with all rays emanating from q, which results in a compact totally convex subset. The next step is to consider the points of this subset which are at maximal distance from its boundary, and these points in turn form a compact totally convex subset, and after finitely many iterations the process terminates in a subset with empty boundary, called a soul. As we shall see below, if $G_m \ge 0$, then the soul construction with basepoint $q \in \mathfrak{C}_m \setminus \{o\}$ takes no more than two steps; more precisely, deleting the horoballs for rays emanating from q results either in $\{q\}$ or in a segment with q as an endpoint. In the latter case the soul is the midpoint of the segment.

In what follows we let B_σ denote the (open) horoball for a ray σ with $\sigma(0) = q$, that is, the union over $t \in [0, \infty)$ of the metric balls of radius t centered at $\sigma(t)$. Let H_σ denote the complement of B_σ in the ambient complete Riemannian manifold.

Lemma 4.4. *Let σ be a ray in a complete Riemannian manifold M, and let $q = \sigma(0)$. Then for any nonzero $v \in T_q M$ that makes an acute angle with σ, the point $\exp_q(tv)$ lies in the horoball B_σ for all small $t > 0$.*

Proof. This follows from the definition of a horoball for if Υ denotes the image of $t \to \exp_q(tv)$, then

$$\lim_{s\to+0} \frac{d(\sigma(s), \Upsilon)}{d(\sigma(s), q)} = \sin\angle(v'(0), \sigma'(0)) < 1,$$

so B_σ contains a subsegment of $\Upsilon - \{q\}$ that approaches q. □

Proof of Theorem 1.7. For $q \in \mathfrak{C}_m$, let C_q denote the complement in M_m of the union of the horoballs for rays that start at q; note that C_q is compact and totally convex. If C_q equals $\{q\}$, then q is a soul. Otherwise, C_q has positive dimension and $q \in \partial C_q$. Set $\gamma := \xi_q$; thus γ is a ray.

Case 1. Suppose $\pi/2 < \hat{\kappa}(r_q) < \pi$. Let $\bar{\gamma}$ be the clockwise ray that is mapped to γ by the isometry fixing the meridian through q. We next show that q is the intersection of the complements of the horoballs for rays μ_q, γ, and $\bar\gamma$, implying that q is a soul for the soul construction that starts at q. As $\kappa_{\gamma(0)} > \pi/2$, any nonzero $v \in T_q M_m$ forms angle $< \pi/2$ with one of $\mu'(0)$, $\gamma'(0)$, or $\bar\gamma'(0)$, so $\exp_q(tv)$ cannot lie in the intersection of H_{μ_q}, H_γ, and $H_{\bar\gamma}$ for small t, and since the intersection is totally convex, it is $\{q\}$.

Case 2. Suppose $\hat\kappa(r_q) = \pi/2$, so that $\gamma = \gamma_q$, and suppose that G_m does not vanish along γ. By symmetry and Lemma 4.4, it suffices to show that every point of the segment $[o, q)$ near q lies in B_γ. Let α be the ray from o passing through q. The geodesic γ is orthogonal to α, and it suffices to show that there is a focal point w of α along γ (for this would imply that there is a family of geodesics of the same length that minimize the distance from w to α, and since the geodesics cannot minimize beyond the focal point, all points near q on α, except q, are in B_γ [Sakai 1996, Lemma III.2.11]).

Any α-Jacobi field along γ is of the form jn where n is a parallel nonzero normal vector field along γ and j solves $j''(t) + G_m(r_{\gamma(t)})j(t) = 0$, $j(0) = 1$, $j'(0) = 0$. Since $G_m \geq 0$, the function j is concave, so due to its initial values, j must vanish unless it is constant. The point where j vanishes is focal. If j is constant, then $G_m = 0$ along γ, which is ruled out by assumption.

Case 3. Suppose $\hat\kappa(r_q) = \pi$, that is, $\gamma = \tau_q$. For any vector $v \in T_q M_m$ pointing inside C_q, for small t the point $\exp_q(tv)$ is not in the horoballs for μ_q and τ_q, and hence v is tangent to a parallel, that is, C_q is a subsegment of the geodesic α tangent to the parallel through q. As C_q lies outside the horoballs for μ_q and τ_q, these rays there cannot contain focal points of α, implying that G_m vanishes along μ_q and τ_q, and hence everywhere, by rotational symmetry, so that M_m is the standard $\mathbb{R}^2$, and q is a soul.

Case 4. Suppose $\hat\kappa(r_q) = \pi/2$, so that $\gamma = \gamma_q$, and suppose that G_m vanishes along γ. By rotational symmetry $G_m(r) = 0$ for $r \geq r_q$, so $m(r) = ar + m(0)$ for $r \geq r_q$ where $a > 0$, as m only vanishes at 0. The turn angle of γ can be computed explicitly as

$$(4.5)\quad \int_x^\infty \frac{dr}{m(r)\sqrt{\dfrac{m(r)^2}{m(x)^2} - 1}} = \int_1^\infty \frac{dt}{at\sqrt{t^2-1}} = -\frac{1}{a}\operatorname{arccot}\big(\sqrt{t^2-1}\big)\Big|_1^\infty = \frac{\pi}{2a},$$

where $x := r_q$. Since γ is a ray, we deduce that $a \geq \frac{1}{2}$.

Let $z \leq x$ be the smallest number such that $m'|_{[z,\infty)} = a$; thus there is no neighborhood of z in $(0, \infty)$ on which G_m is identically zero.

Note that $m(r) = a(r - z) + m(z)$ for $r \geq z$, so the surface $M_m – B(o, z)$ is isometric to $C – B(\bar o, m(r_q)/a)$ where C is the cone with apex $\bar o$ such that cutting C

along the meridian from $\bar{o}$ gives a sector in $\mathbb{R}^2$ of angle $2\pi a$ with the portion inside the radius $m(r_q)/a$ removed.

Since γ_q is a ray, Lemma 4.4 implies the existence of a neighborhood U_q of q such that each point in U_p–$[o, q]$ lies in a horoball for a ray from q.

We now check that o lies in the horoball of γ_q. Concavity of m implies that the graph of m lies below its tangent line at z, so evaluating the tangent line at $r = 0$ and using $m(0) = 0$ gives $m(z)/a > z$. The Pythagorean theorem in the sector in $\mathbb{R}^2$ of angle $2\pi a$ implies that

$$d_{M_m}(\gamma_q(s), o) = \sqrt{s^2 + \left(x - z + \frac{m(z)}{a}\right)^2} + z - \frac{m(z)}{a},$$

which is $< s$ for large s, implying that o is in the horoball of γ_q.

To realize q as a soul, we need to look at the soul construction with arbitrary basepoint v, which starts by considering the complement in M_m of the union of the horoballs for all rays from v, which by the above is either v or a segment $[u, v]$ contained in $(o, v]$, where u is uniquely determined by v. It will be convenient to allow for degenerate segments for which $u = v$; with this convention the soul is the midpoint of $[u, v]$. Since z is the smallest such that $G_m|_{[z,\infty)} = 0$, the focal point argument of Case 2 shows that $u = v$ when $0 < r_v < z$. Set $y := r_v$, and let $e(y) := r_u$; note that $0 < e(y) \leq y$, and the midpoint of $[u, v]$ has r-coordinate $h(y) := (y + e(y))/2$.

To realize each point of M_m as a soul, it suffices to show that each positive number is in the image of h. Since h approaches zero as $y \to 0$ and approaches infinity as $y \to \infty$, it is enough to show that h is continuous and then apply the intermediate value theorem.

Since $e(y) = y$ when $0 < y < z$, we only need to verify continuity of e when $y \geq z$. Let v_i be an arbitrary sequence of points on α converging to v, where as before α is the ray from o passing through q. Set $v_i := r_{v_i}$. Arguing by contradiction suppose that $e(y_i)$ does not converge to $e(y)$. Since $0 < e(y_i) \leq y_i$ and $y_i \to y$, we may pass to a subsequence such that $e(y_i) \to e_\infty \in [0, y]$. Pick any w such that r_w lies between e_∞ and $e(y)$. Thus there is i_0 such that either $e(y_i) < r_w < e(y)$ for all $i > i_0$, or $e(y) < r_w < e(y_i)$ for all $i > i_0$. As $y \geq z$, we know that G_m vanishes along γ_v, so every α-Jacobi field along γ_v is constant. Therefore, the rays γ_{v_i} converge uniformly (!) to γ_v, as $v_i \to v$, and hence their Busemann functions b_i and b converge pointwise. Thus $b_i(w) \to b(w)$, but we have chosen w so that $b(w)$ and $b_i(w)$ are all nonzero, and $\operatorname{sign}(b(w)) = -\operatorname{sign}(b_i(w))$, which gives a contradiction proving the theorem. □

Remark 4.6. In Cases 1, 2, and 3 the soul construction terminates in one step, namely, if $q \in \mathfrak{C}_m$, then $\{q\}$ is the result of removing the horoballs for all rays that start at q. We do not know whether the same is true in Case 4 because the basepoint v needed to produce the soul q is found implicitly, via the intermediate value theorem, and it is unclear how v depends on q, and whether $v = q$.

Remark 4.7. Let M_m be as in Case 4 with $m'|_{[z,\infty)} = \frac{1}{2}$. If M_m is von Mangoldt, then no point q with $r_q \geq z$ is a pole because by (4.5) the turn angle of γ_q is π, which by Theorem 3.24 cannot happen for a pole.

5. Smoothed cones made von Mangoldt

Proof of Theorem 1.11. It is of course easy to find a von Mangoldt plane g_{m_x} that has zero curvature near infinity, but prescribing the slope of m' there takes more effort. We exclude the trivial case $x = 1$ in which $m(r) = r$ works.

For $u \in [0, \frac{1}{4}]$ set $K_u(r) = 1/(4(r+1)^2) - u$, and let m_u be the unique solution of (A.7) with $K = K_u$. Then g_{m_u} is von Mangoldt. For $u > 0$ let $z_u \in [0, \infty)$ be the unique zero of K_u; note that z_u is the global minimum of m'_u, and $z_u \to \infty$ as $u \to 0$.

Lemma 5.1. *The function $u \to m'_u(z_u)$ takes every value in $(0, 1)$ as u varies in $(0, \frac{1}{4})$.*

Proof. One verifies that $m_0(r) = \ln(r+1)\sqrt{r+1}$, that is, the right hand side solves (A.7) with $K = K_0$. Then $m'_0 = (2 + \ln(r+1))/(2\sqrt{r+1})$ is a positive function converging to zero as $r \to \infty$. By Sturm comparison $m_u \geq m_0 > 0$ and $m'_u \geq m'_0 > 0$.

We now show that $m'_u(z_u) \to 0$ as $u \to +0$. To this end fix an arbitrary $\varepsilon > 0$. Fix t_ε such that $m'_0(t_\varepsilon) < \varepsilon$. By continuous dependence on parameters (m_u, m'_u) converges to (m_0, m'_0) uniformly on compact sets as $u \to 0$. So for all small u we have $m'_u(t_\varepsilon) < \varepsilon$ and also $t_\varepsilon < z_u$. Since m'_u decreases on $(0, z_u)$, we conclude that $0 < m'_u(z_u) < m'_u(t_\varepsilon) < \varepsilon$, proving that $m'_u(z_u) \to 0$ as $u \to +0$.

On the other hand, $m'_{1/4}(z_{1/4}) = 1$ because $z_{1/4} = 0$ and by the initial condition $m'_{1/4}(0) = 1$. Finally, the assertion of the lemma follows from continuity of the map $u \to m'_u(z_u)$, because then it takes every value within $(0, 1)$ as u varies in $(0, \frac{1}{4})$. (To check continuity of the map fix u_*, take an arbitrary $u \to u_*$ and note that $z_u \to z_{u_*}$, so since m'_u converges to m'_{u_*} on compact subsets, it does so on a neighborhood of z_{u_*}, so $m'_u(z_u)$ converges to $m'_{u_*}(z_{u_*})$.) □

Continuing the proof of the theorem, fix an arbitrary $u > 0$. The continuous function $\max(K_u, 0)$ is decreasing and smooth on $[0, z_u]$ and equal to zero on $[z_u, \infty)$. So there is a family of nonincreasing smooth functions $G_{u,\varepsilon}$ depending on the small parameter ε such that $G_{u,\varepsilon} = \max(K_u, 0)$ outside the ε-neighborhood of z_u. Let $m_{u,\varepsilon}$ be the unique solution of (A.7) with $K = G_{u,\varepsilon}$; thus $m'_{u,\varepsilon}(r) = m'_{u,\varepsilon}(z_u + \varepsilon)$ for all $r \geq z_u + \varepsilon$. If ε is small enough, then $G_{u,\varepsilon} \leq K_0$, so $m_{u,\varepsilon} \geq m_0 > 0$ and

$m'_{u,\varepsilon} \geq m'_0 > 0$. By continuous dependence on parameters, the function $(u, \varepsilon) \to m'_{u,\varepsilon}$ is continuous, and moreover $m'_{u,\varepsilon}(z_u + \varepsilon) \to m'_u(z_u)$ as $\varepsilon \to 0$, and u is fixed.

Fix $x \in (0, 1)$. By Lemma 5.1 there are positive v_1 and v_2 such that $m'_{v_1}(z_{v_1}) < x < m'_{v_2}(z_{v_2})$. Letting u of the previous paragraph to be v_1, v_2, we find ε such that $m'_{v_1,\varepsilon}(z_{v_1} + \varepsilon) < x < m'_{v_2,\varepsilon}(z_{v_2} + \varepsilon)$, so by the intermediate value theorem there is u with $m'_{u,\varepsilon}(z_u + \varepsilon) = x$. Then the metric $g_{m_{u,\varepsilon}}$ has the asserted properties for $\rho = z_u + \varepsilon$. □

6. Other applications

Proof of Lemma 1.1. Assuming $\hat{r}(\hat{q}) \notin r(\mathfrak{C}_m)$ we will show that $\hat{q}$ is not a critical point of $\hat{r}$. Since $\hat{M}$ is complete and noncompact, there is a ray $\hat{\gamma}$ emanating from $\hat{q}$. Consider the comparison triangle $\Delta(o, q, q_i)$ in M_m for any geodesic triangle with vertices $\hat{o}$, $\hat{q}$, and $\hat{\gamma}(i)$. Passing to a subsequence, arrange so that the segments $[q, q_i]$ subconverge to a ray, which we denote by γ. Since $q \notin \mathfrak{C}_m$, the angle formed by γ and $[q, o]$ is $> \pi/2$, and hence for large i the same is true for the angles formed by $[q, q_i]$ and $[q, o]$. By comparison, $\hat{\gamma}$ forms angle $> \pi/2$ with any segment joining $\hat{q}$ to $\hat{o}$, that is, $\hat{q}$ is not a critical point of $\hat{r}$. □

Proof of Theorem 1.5. (a) Let P_m denote the set of poles; it is a closed metric ball [Tanaka 1992b, Lemma 1.1]. Moreover, P_m clearly lies in the connected component A^o_m of $A_m \cup \{o\}$ that contains o, and hence in the component of $\mathfrak{C}_m$ that contains o. By Theorem 1.6 A_m is open in M_m, so $A_m \cup \{o\}$ is locally path-connected, and hence A^o_m is open in M_m. If P_m were equal to A^o_m, the latter would be closed, implying $A^o_m = M_m$, which is impossible as the ball has finite radius.

(b) The "if" direction is trivial as $P_m \subset \mathfrak{C}_m$. Conversely, if $\mathfrak{C}_m \neq \{o\}$, then by Lemma 3.15 m^{-2} is integrable and $\liminf_{r\to\infty} m(r) > 0$, so $R_p > 0$ [Tanaka 1992a]. □

Proof of Theorem 1.9. By assumption there is a point of negative curvature, and since the curvature is nonincreasing, outside a compact subset the curvature is bounded above by a negative constant. As $\liminf_{r\to\infty} m(r) > 0$, m is bounded below by a positive constant outside any neighborhood of 0, so $\int_0^\infty m = \infty$. Hence the total curvature $2\pi \int_0^\infty G_m(r) m(r)\, dr$ is $-\infty$.

Hence there is a metric ball B of finite positive radius centered at o such that the total curvature of B is negative, and such that no point of $G_m \geq 0$ lies outside B. By [Shiohama et al. 2003, Theorem 6.1.1, p. 190], for any $q \in M_m$ the total curvature of the set obtained from M_m by removing all rays that start at q is in $[0, 2\pi]$. So for any q there is a ray that starts at q and intersects B.

If q is not in B, then the ray points away from infinity, so $q \in A_m$ and any point on this ray is in $\mathfrak{C}_m$. Thus $M_m - A_m$ lies in B. Since $\mathfrak{C}_m \neq \{o\}$, Theorem 1.5 implies that $R_p > 0$. Letting q run to infinity the rays subconverge to a line that intersects B; see, for example, [Shiohama et al. 2003, Lemma 6.1.1, p. 187].

If $m'(r_p) = 0$, the parallel through p is a geodesic but not a ray, so Lemma 3.14 implies that no point on the parallel through p is in $\mathfrak{C}_m$. Since $\mathfrak{C}_m$ contains o and all points outside a compact set, $\mathfrak{C}_m$ is not connected; the same argument proves that A_m is not connected. □

Example 6.1. Here we modify [Tanaka 1992b, Example 4] to construct a von Mangoldt plane M_m such that m' has a zero, and neither A_m nor $\mathfrak{C}_m$ is connected. Given $a \in (\pi/2, \pi)$ let $m_0(r) = \sin r$ for $r \in [0, a]$, and define m_0 for $r \geq a$ so that m_0 is smooth, positive, and $\liminf_{r\to\infty} m_0 > 0$. Thus $K_0 := -m_0''/m_0$ equals 1 on $[0, a]$. Let K be any smooth nonincreasing function with $K \leq K_0$ and $K|_{[0,a]} = 1$. Let m be the solution of (A.7); note that $m(r) = \sin(r)$ for $r \in [0, a]$ so that m' vanishes at $\pi/2$. By Sturm comparison $m \geq m_0 > 0$, and hence M_m is a von Mangoldt plane. Since $m'(a) < 0$ and $m > 0$ for all $r > 0$, the function m cannot be concave, so $K = G_m$ eventually becomes negative, and Theorem 1.9 implies that A_m and $\mathfrak{C}_m$ are not connected.

Example 6.2. Here we construct a von Mangoldt plane such that $m' > 0$ everywhere but A_m and $\mathfrak{C}_m$ are not connected. Let M_n be a von Mangoldt plane such that $G_n \geq 0$ and $n' > 0$ everywhere, and R_n is finite (where R_n is the radius of the ball $\mathfrak{C}_n$). This happens, for example, for any paraboloid, any two-sheeted hyperboloid with $n'(\infty) < \frac{1}{2}$, or any plane constructed in Theorem 1.11 with $n'(\infty) < \frac{1}{2}$. Fix $q \notin \mathfrak{C}_n$. Then γ_q has turn angle $> \pi$, so there is $R > r_q$ such that $\int_{r_q}^{R} F_{n(r_q)} > \pi$. Let G be any smooth nonincreasing function such that $G = G_n$ on $[0, R]$ and $G(z) < 0$ for some $z > R$. Let m be the solution of (A.7) with $K = G$. By Sturm comparison $m \geq n > 0$ and $m' \geq n' > 0$ everywhere; see Remark A.10. Since $m = n$ on $[0, R]$, on this interval we have $F_{m(r_q)} = F_{n(r_q)}$, so in the von Mangoldt plane M_m the geodesic γ_q has turn angle $> \pi$, which implies that no point on the parallel through q is in $\mathfrak{C}_m$. Now Theorem 1.9 (3) and (4) imply that A_m and $\mathfrak{C}_m$ are not connected.

Theorem 6.3. *Let M_m be a von Mangoldt plane such that $m'|_{[0,y]} > 0$ and $m'|_{[x,y]} < \frac{1}{2}$. Set $f_{m,x}(y) := m^{-1}(\cos(\pi b) m(y))$, where b is the maximum of m' on $[x, y]$. If $x \leq f_{m,x}(y)$, then $r(\mathfrak{C}_m)$ and $[x, f_{m,x}(y)]$ are disjoint.*

Proof. Set $f := f_{m,x}$. Arguing by contradiction assume there is $q \in \mathfrak{C}_m$ with $r_q \in [x, f(y)]$. Then γ_q has turn angle $\leq \pi$, so if $c := m(r_q)$, then

$$\pi \geq \int_{r_q}^{\infty} \frac{c\,dr}{m\sqrt{m^2 - c^2}} > \int_{r_q}^{y} \frac{c\,dr}{m\sqrt{m^2 - c^2}}$$
$$= \int_{c}^{m(y)} \frac{c\,dm}{m'(r) m\sqrt{m^2 - c^2}} \geq \int_{c}^{m(y)} \frac{c\,dm}{bm\sqrt{m^2 - c^2}} = \frac{1}{b} \arccos\left(\frac{c}{m(y)}\right),$$

so that $\pi b > \arccos(c/(m(y)))$, which is equivalent to $\cos(\pi b)m(y) < m(r_q)$.

On the other hand, $m(f(y))$ is in the interval $[0, m(y)]$ on which m^{-1} is increasing, so $f(y) < y$, and therefore m is increasing on $[x, f(y)]$. Hence $r_q < f(y)$ implies $m(r_q) < m(f(y)) = \cos(\pi b)m(y)$, which is a contradiction. □

Proof of Theorem 1.10. We use the notation of Theorem 6.3. The assumptions on n imply $n' > 0$, $n'|_{[x,\infty)} < \frac{1}{2}$, and $b = n'(x)$. Hence $f_{n,x}$ is an increasing smooth function of y with $f_{n,x}(\infty) = \infty$. In particular, if y is large enough, then $f_{n,x}(y) > z > x$; fix y that satisfies the inequality. Now if M_m is any von Mangoldt plane with $m = n$ on $[0, y]$, then $f_{m,x}(y) = f_{n,x}(y)$, so M_m satisfies the assumptions of Theorem 6.3, so $[x, z]$ and $r(\mathfrak{C}_m)$ are disjoint. □

Appendix A: Von Mangoldt planes

The purpose of this appendix is to discuss what makes von Mangoldt planes special among arbitrary rotationally symmetric planes.

For a smooth function $m : [0, \infty) \to [0, \infty)$ whose only zero is 0, let g_m denote the rotationally symmetric inner product on the tangent bundle to $\mathbb{R}^2$ that equals the standard Euclidean inner product at the origin and elsewhere is given in polar coordinates by $dr^2 + m(r)^2\,d\theta^2$. It is well known (see, for example, [Shiohama et al. 2003, §7.1]) that:

- Any rotationally symmetric complete smooth Riemannian metric on $\mathbb{R}^2$ is isometric to some g_m. (As before, M_m denotes $(\mathbb{R}^2, g_m)$.)
- If $\bar{m} : \mathbb{R} \to \mathbb{R}$ denotes the unique odd function such that $\bar{m}|_{[0,\infty)} = m$, then g_m is a smooth Riemannian metric on $\mathbb{R}^2$ if and only if $m'(0) = 1$ and $\bar{m}$ is smooth.
- If g_m is a smooth metric on $\mathbb{R}^2$, then g_m is complete, and the sectional curvature of g_m is a smooth function on $[0, \infty)$ that equals $-m''/m$.

It is easier to visualize M_m as a surface of revolution in $\mathbb{R}^3$, so we recall:

Lemma A.1. (1) *M_m is isometric to a surface of revolution in $\mathbb{R}^3$ if and only if $|m'| \le 1$.*

(2) *M_m is isometric to a surface of revolution $(r\cos\phi, r\sin\phi, g(r))$ in $\mathbb{R}^3$ if and only if $0 < m' \le 1$.*

Proof. (1) Consider a unit speed curve $s \to (x(s), 0, z(s))$ in $\mathbb{R}^3$ where $x(s) \ge 0$ and $s \ge 0$. Rotating the curve about the z-axis gives the surface of revolution

$$(x(s)\cos\phi, x(s)\sin\phi, z(s))$$

with metric $ds^2 + x(s)^2\,d\phi^2$. The meridians starting at the origin are rays, so for this metric to be equal to $ds^2 + m(s)^2\,d\phi^2$ we must have $m(s) = x(s)$. Since the curve has unit speed, $|x'(s)| \le 1$, so a necessary condition for writing the metric as a surface of revolution is $|m'(s)| \le 1$. It is also sufficient for if $|m'(s)| \le 1$, then we could let $z(s) := \int_0^s \sqrt{1-(m'(s))^2}\,ds$, so that now $(m(s), z(s))$ has unit speed.

(2) If, furthermore, $m' > 0$ for all s, then the inverse function of $m(s)$ makes sense, and we can write the surface of revolution $(m(s)\cos\phi, m(s)\sin\phi, z(s))$ as $(x\cos\phi, x\sin\phi, g(x))$ where $x := m(s)$ and $g(x) := z(m^{-1}(x))$. Conversely, given the surface $(x\cos\phi, x\sin\phi, g(x))$, the orientation-preserving arclength parametrization $x = x(s)$ of the curve $(x, 0, g(x))$ satisfies $x' > 0$. □

Example A.2. The standard $\mathbb{R}^2$ is the only von Mangoldt plane with $G_m \le 0$ that can be embedded into $\mathbb{R}^3$ as a surface of revolution because $m'(0) = 1$ and m' is nondecreasing afterwards.

Example A.3. If $G_m \ge 0$, then $m' \in [0, 1]$ because $m > 0$, m' is nonincreasing, and $m'(0) = 1$, so that M_m is isometric to a surface of revolution in $\mathbb{R}^3$. In fact, if $m'(s_0) = 0$, then $m|_{[s_0,\infty)} = m(s_0)$, that is, outside the s_0-ball about the origin M_m is a cylinder. Thus except for such surfaces M_m can be written as

$$(x\cos\phi, x\sin\phi, g(x)) \quad \text{for } g(x) = \int_0^{m^{-1}(x)} \sqrt{1-(m'(s))^2}\,ds.$$

Paraboloids and two-sheeted hyperboloids are von Mangoldt planes of positive curvature [Shiohama et al. 2003, p. 234–235] and are of the form $(x\cos\phi, x\sin\phi, g(x))$.

The defining property $G'_m \le 0$ of von Mangoldt planes clearly restricts the behavior of m'. Let $Z(G_m)$ denote the set where G_m vanishes; as M_m is von Mangoldt, $Z(G_m)$ is closed and connected, and hence it could be equal to the empty set, a point, or an interval, while m' behaves as follows.

(i) If $G_m > 0$, then m' is decreasing and takes values in $(0, 1]$.

(ii) If $G_m \le 0$, then m' is nondecreasing and takes values in $[1, \infty)$.

(iii) If $Z(G_m)$ is a positive number z, then m' decreases on $[0, z)$ and increases on (z, ∞), and m' may have two, one, or no zeros.

(iv) If $Z(G_m) = [a, b] \subset (0, \infty]$, then m' decreases on $[0, a)$, is constant on $[a, b]$, and increases on (b, ∞) if $b < \infty$. Also either $m'|_{[a,b]} = 0$ or else m' has two, or no zeros.

Remark A.4. All the above possibilities occur with one possible exception: in Cases (iii) and (iv) we are not aware of examples where m' vanishes on $Z(G_m)$.

Remark A.5. Thus if M_m is von Mangoldt, then m' is monotone near infinity, so $m'(\infty)$ exists; moreover, $m'(\infty) \in [0, \infty]$, for otherwise m would vanish on $(0, \infty)$. It follows that M_m admits total curvature, which equals

$$\int_0^{2\pi}\int_0^{\infty} G_m m\,dr\,d\theta = -2\pi\int_0^{\infty} m'' = 2\pi(1-m'(\infty)) \in [-\infty, 2\pi].$$

Here the *total curvature of a subset* $A \subset M_m$ is the integral of G_m over A with respect to the Riemannian area form $m\, dr d\theta$, provided the integral converges to a number in $[-\infty, \infty]$, in which case we say that A *admits total curvature*.

Remark A.6. The zeros of m' correspond to parallels that are geodesics and are of interest. In contrast with restrictions on the zero set of m' for von Mangoldt planes, if M_m is not necessarily von Mangoldt, then any closed subset of $[0, \infty)$ that does not contain 0 can be realized as the set of zeros of m'. (Indeed, for any closed subset of a manifold there is a smooth nonnegative function that vanishes precisely on the subset [Bröcker and Jänich 1982, Whitney's Theorem 14.1]. It follows that if C is a closed subset of $[0, \infty)$ that does not contain 0, then there is a smooth function $g : [0, \infty) \to [0, \infty)$ that is even at 0, satisfies $g(0) = 1$, and is such that $g(s) = 0$ if and only if $s \in C$. If m is the solution of $m' = g$ and $m(0) = 0$, then M_m has the promised property.)

A common way of constructing von Mangoldt planes involves the Jacobi initial value problem

$$m'' + Km = 0, \quad m(0) = 0, \quad m'(0) = 1, \tag{A.7}$$

where K is smooth on $[0, \infty)$. It follows from the proof of [Kazdan and Warner 1974, Lemma 4.4] that g_m is a complete smooth Riemannian metric on $\mathbb{R}^2$ if and only if the following condition holds:

($\star$) the (unique) solution m of (A.7) is positive on $(0, \infty)$.

Remark A.8. A basic tool that produces solutions of (A.7) satisfying condition ($\star$) is the Sturm comparison theorem that implies that if m_1 is a positive function that solves (A.7) with $K = K_1$, and if K_2 is any nonincreasing smooth function with $K_2 \le K_1$, then the solution m_2 of (A.7) with $K = K_2$ satisfies $m_2 \ge m_1$, so that g_{m_2} is a von Mangoldt plane.

Example A.9. If K is a smooth function on $[0, \infty)$ such that $\max(K, 0)$ has compact support, then a positive multiple of K can be realized as the curvature G_m of some M_m; of course, if K is nonincreasing, then M_m is von Mangoldt. (Indeed, in [Kazdan and Warner 1974, Lemma 4.3] Sturm comparison was used to show that if $\int_t^\infty \max(K, 0) \le 1/(4t + 4)$ for all $t \ge 0$, then K satisfies ($\star$), and in particular, if $\max(K, 0)$ has compact support, then there is a constant $\varepsilon > 0$ such that the above inequality holds for εK.)

Remark A.10. A useful addendum to Remark A.8 is that the additional assumption $m_1' \ge 0$ implies $m_2' \ge m_1' > 0$. (Indeed, the function $m_1' m_2 - m_1 m_2'$ vanishes at 0 and has nonpositive derivative $(-K_1 + K_2) m_1 m_2$, so $m_1' m_2 \le m_1 m_2'$. As m_1, m_2, and m_1' are nonnegative, so is m_2'. Hence, $m_1 m_2' \le m_2 m_2'$, which gives $m_1' m_2 \le m_2 m_2'$, and the claim follows by canceling m_2.)

Question A.11. *Let* $m_0 : [r_0, \infty) \to (0, \infty)$ *be a smooth function such that* $r_0 > 0$ *and* $-m_0''/m_0$ *is nonincreasing. What are sufficient conditions for (or obstructions to) extending* m_0 *to a function* m *on* $[0, \infty)$ *such that* g_m *is a von Mangoldt plane?*

Appendix B: A calculus lemma

This appendix contains an elementary lemma on continuity and differentiability of the turn angle, which is needed for Theorem 3.24.

Given numbers $r_q > r_0 > 0$, let m be a smooth self-map of $(0, \infty)$ such that

- $m' > 0$ on $[r_0, r_q]$,
- $m(r) > m(r_q)$ for $r > r_q$,
- m^{-2} is integrable on $(1, \infty)$,
- $\liminf_{r\to\infty} m(r) > m(r_q)$.

Example B.1. Suppose $G_m \geq 0$ or $G'_m \leq 0$. If γ_q is a ray on M_m, and r_0 is sufficiently close to r_q, then m satisfies the above properties by Lemmas 3.3, 3.8, and 3.10.

Set $c_0 := m(r_0)$ and $c_q := m(r_q)$. Let $T = T(c)$ be the function given by the integral (3.21) for $c = c_q$, and by the sum of integrals (3.22) for $c_0 \leq c \leq c_q$, where F_c is given by (3.5) and $r_u := m^{-1}(c)$, where m^{-1} is the inverse of $m|_{[r_0, r_q]}$.

Lemma B.2. *Under the assumptions of the previous paragraph,* T *is continuous on* $(c_0, c_q]$*, continuously differentiable on* (c_0, c_q)*, and* $T'(c)\sqrt{c_q^2 - c^2}$ *converges to* $-1/(m'(r_q)) < 0$ *as* $c \to c_q-$.

Proof. By definition T equals $\int_{r_q}^{\infty} F_c + \int_{r_u}^{r_q} F_c$ if $c \in [c_0, c_q)$ and $T = \int_{r_q}^{\infty} F_c$ if $c = c_q$. Step 1 shows that $\int_{r_q}^{\infty} F_c$ depends continuously on $c \in [c_0, c_q]$, while Step 2 establishes continuity of T at c_q. In Steps 3 and 4 we prove continuous differentiability and compute the derivatives of integrals $\int_{r_q}^{\infty} F_c$ and $\int_{r_u}^{r_q} F_c$ with respect to $c \in (c_0, c_q)$. Step 5 investigates the behavior of $T'(c)$ as $c \to c_q$.

Recall that the integral $\int_a^b H_c(r)\,dr$ depends continuously on c if for each $r \in (a, b)$ the map $c \to H_c(r)$ is continuous, and every c has a neighborhood U_0 in which $|H_c| \leq h_0$ for some integrable function h_0. If in addition each map $c \to H_c(r)$ is C^1, and every c has a neighborhood U_1 where $|\partial H_c/\partial c| \leq h_1$ for an integrable function h_1, then $\int_a^b H_c(r)\,dr$ is C^1 and differentiation under the integral sign is valid; the same conclusion holds when H_c and $\partial H_c/\partial c$ are continuous in the closure of $U_1 \times (a, b)$.

Step 1. The integrand F_c is smooth over (r_u, ∞), because the assumptions on m imply that $m(r) > c$ for $r > r_u$.

Since $0 < c \leq c_q$ we have $F_c \leq F_{c_q} = c_q/(m\sqrt{m^2 - c_q^2})$ which is integrable on (r_q, ∞). Indeed, fix $\delta > r_q$ and note that since m^{-2} is integrable on (δ, ∞), so is

F_{c_q}. To prove integrability of F_{c_q} on (r_q, δ), note that

$$h(r) := \frac{m(r) - m(r_q)}{r - r_q}$$

is positive on $[r_q, \infty)$, as $h(r_q) = m'(r_q) > 0$ and $m(r) > m(r_q)$ for $r > r_q$. Then F_{c_q} is the product of $(r - r_q)^{-1/2}$ and a function that is smooth on $[r_q, \delta]$, and hence F_{c_q} is integrable on (r_q, δ).

Thus the integrals $\int_{r_q}^{\delta} F_c(r)\,dr$ and $\int_{\delta}^{\infty} F_c(r)\,dr$ depend continuously on $c \in (0, c_q]$, and hence so does their sum $\int_{r_q}^{\infty} F_c(r)\,dr$.

Step 2. As $c \to c_q$, the integral $\int_{r_u}^{r_q} F_c$ converges to zero, for if K is the maximum of $(mm'\sqrt{m+c})^{-1}$ over the points with $r \in [r_0, r_q]$ and $c \in [c_0, c_q]$, then

$$\int_{r_u}^{r_q} F_c \le K \int_{r_u}^{r_q} \frac{m'\,dr}{\sqrt{m-c}} = K \int_0^{c_q - c} \frac{dt}{\sqrt{t}},$$

which goes to zero as $c \to c_q$. Thus T is continuous at $c = c_q$.

Step 3. To find an integrable function dominating $\partial F_c/\partial c$ on (r_q, ∞) locally in c, note that every $c \in (c_0, c_q)$ has a neighborhood of the form $(c_0, c_q - \delta)$ with $\delta > 0$, and over this neighborhood

$$\frac{\partial F_c}{\partial c} = \frac{m}{(m^2 - c^2)^{3/2}} \le \frac{m}{(m^2 - (c_q - \delta)^2)^{3/2}},$$

where the right hand side is integrable over $[r_q, \infty)$, as m^{-2} is integrable at ∞; thus

$$\frac{d}{dc} \int_{r_q}^{\infty} F_c = \int_{r_q}^{\infty} \frac{m}{(m^2 - c^2)^{3/2}}\,dr$$

is continuous with respect to $c \in (c_0, c_q)$. This integral diverges if $c = m(r_q)$.

Step 4. To check continuity of $\int_{r_u}^{r_q} F_c$ change variables via $t := m/c$ so that $r = m^{-1}(tc)$. Thus $dt = m'(r)\,dr/c = n(tc)\,dr/c$ where $n(r) := m'(m^{-1}(r))$, and

$$\int_{r_u}^{r_q} F_c(r)\,dr = \int_1^{c_q/c} \bar{F}_c(t)\,dt \quad \text{where} \quad \bar{F}_c(t) = \frac{1}{n(tc)\,t\sqrt{t^2 - 1}}.$$

Since $m' > 0$ on $[r_0, r_q]$ and $n(tc) = m'(r)$, the function $\bar{F}_c$ is smooth over $(1, c_q/c)$. To prove the continuity of $\int_1^{c_q/c} \bar{F}_c$, fix an arbitrary $(u, v) \subset (c_0, c_q)$. If $c \in (u, v)$ and $t \in (1, c_q/c)$, then $m^{-1}(tc)$ lies in the m^{-1}-image of $(u, (v/u)c_q)$, which by taking the interval (u, v) sufficiently small can be made to lie in an arbitrarily small neighborhood of $[r_0, r_q]$, so we may assume that $m' > 0$ on that neighborhood. It follows that the maximum K of $1/(n(tc))$ over $c \in [u, v]$ and $t \in [1, c_q/c]$ is finite, and $|\bar{F}_c| \le K/(t\sqrt{t^2 - 1})$ for $c \in (u, v)$, that is, $|F_c|$ is locally dominated by an integrable function that is independent of c; for the same reason the conclusion also holds for

$$\frac{\partial \bar{F}_c}{\partial c} = -\frac{n'(tc)}{n(tc)^2\sqrt{t^2-1}}.$$

Finally, given $c_* \in (c_0, c_q)$, fix $\delta \in (1, c_q/c_*)$ and write $\int_1^{c_q/c} \bar{F}_c = \int_1^{\delta} \bar{F}_c + \int_{\delta}^{c_q/c} \bar{F}_c$ for c varying near c_*. The first summand is C^1 at c_*, as the integrand and its derivative are dominated by the integrable function near c_*. The second summand is also C^1 at c_* as the integrand is C^1 on a neighborhood of $\{c_*\} \times [\delta, c_q/c]$. By the integral Leibniz rule

$$\frac{d}{dc}\int_1^{c_q/c} \bar{F}_c = -\frac{c_q}{c^2}\bar{F}_c\left(\frac{c_q}{c}\right) - \int_1^{c_q/c} \frac{n'(tc)\,dt}{n(tc)^2\sqrt{t^2-1}}.$$

The first summand equals $-\left(m'(r_q)\sqrt{c_q^2-c^2}\,\right)^{-1}$, and the second summand is bounded.

Step 5. Let us investigate the behavior of $\int_{r_q}^{\infty}(m/(m^2-c^2)^{3/2})\,dr$ from Step 3 as $c \to c_q-$. Fix $\delta > r_q$ such that $m' > 0$ on $[r_0, \delta]$ and write the above integral as the sum of the integrals over (r_q, δ) and (δ, ∞). The latter one is bounded. Integrate the former integral by parts as

$$\int_{r_q}^{\delta} \frac{mm'}{m'(m^2-c^2)^{3/2}}\,dr = -\int_{r_q}^{\delta}\frac{1}{m'}\,d\left(\frac{1}{\sqrt{m^2-c^2}}\right)$$
$$= \frac{1}{m'(r_q)\sqrt{c_q^2-c^2}} - \frac{1}{m'(\delta)\sqrt{\delta^2-c^2}} - \int_{r_q}^{\delta}\frac{m''\,dr}{(m')^2\sqrt{m^2-c^2}}.$$

Only the first summand is unbounded as $c \to c_q-$. The terms from Steps 4 and 5 enter into T' with coefficients 2 and 1, respectively, so as $c \to c_q-$

$$T'(c)\sqrt{c_q^2-c^2} \to -\frac{1}{m'(r_q)} < 0$$

as the bounded terms multiplied by $\sqrt{c_q^2-c^2}$ disappear in the limit. □

Acknowledgments

Belegradek is grateful to NSF for support (DMS-0804038). This paper will be a part of Choi's Ph.D. thesis at Emory University.

References

[Bröcker and Jänich 1982] T. Bröcker and K. Jänich, *Introduction to differential topology*, Cambridge University Press, 1982. MR 83i:58001 Zbl 0486.57001

[Elerath 1980] D. Elerath, "An improved Toponogov comparison theorem for nonnegatively curved manifolds", *J. Differential Geom.* **15**:2 (1980), 187–216. MR 83b:53039 Zbl 0526.53043

[Gluck and Singer 1979] H. Gluck and D. A. Singer, "Scattering of geodesic fields, II", *Ann. of Math.* (2) **110**:2 (1979), 205–225. MR 80k:53073 Zbl 0436.53047

[Greene 1997] R. E. Greene, "A genealogy of noncompact manifolds of nonnegative curvature: history and logic", pp. 99–134 in *Comparison geometry* (Berkeley, CA, 1993–94), edited by K. Grove and P. Petersen, Math. Sci. Res. Inst. Publ. **30**, Cambridge Univ. Press, 1997. MR 98g:53069 Zbl 0884.53029

[Grove 1993] K. Grove, "Critical point theory for distance functions", pp. 357–385 in *Differential geometry, III: Riemannian geometry* (Los Angeles, CA, 1990), edited by R. Greene, Proc. Sympos. Pure Math. **54**, American Mathematical Society, Providence, RI, 1993. MR 94f:53065 Zbl 0806.53043

[Itokawa et al. 2003] Y. Itokawa, Y. Machigashira, and K. Shiohama, "Generalized Toponogov's theorem for manifolds with radial curvature bounded below", pp. 121–130 in *Explorations in complex and Riemannian geometry*, edited by J. Bland et al., Contemp. Math. **332**, Amer. Math. Soc., Providence, RI, 2003. MR 2004j:53051 Zbl 1046.53017

[Kazdan and Warner 1974] J. L. Kazdan and F. W. Warner, "Curvature functions for open 2-manifolds", *Ann. of Math.* (2) **99** (1974), 203–219. MR 49 #7950 Zbl 0278.53031

[Kondo and Ohta 2007] K. Kondo and S.-i. Ohta, "Topology of complete manifolds with radial curvature bounded from below", *Geom. Funct. Anal.* **17**:4 (2007), 1237–1247. MR 2008k:53069 Zbl 1144.53050

[Kondo and Tanaka 2010] K. Kondo and M. Tanaka, "Total curvatures of model surfaces control topology of complete open manifolds with radial curvature bounded below, II", *Trans. Amer. Math. Soc.* **362**:12 (2010), 6293–6324. MR 2011f:53061 Zbl 1225.53034

[Kondo and Tanaka 2011] K. Kondo and M. Tanaka, "Total curvatures of model surfaces control topology of complete open manifolds with radial curvature bounded below, I", *Math. Ann.* **351**:2 (2011), 251–266. MR 2836657 Zbl 05963443

[Machigashira 2010] Y. Machigashira, "Generalized Toponogov comparison theorem for manifolds of roughly non-negative radial curvature", *Information* **13**:3B (2010), 835–841. MR 2011h:53034

[Maeda 1974/1975] M. Maeda, "On the total curvature of noncompact Riemannian manifolds", *Kōdai Math. Sem. Rep.* **26** (1974/1975), 95–99. MR 52 #6611 Zbl 0302.53021

[Mashiko and Shiohama 2006] Y. Mashiko and K. Shiohama, "Comparison geometry referred to warped product models", *Tohoku Math. J.* (2) **58**:4 (2006), 461–473. MR 2007k:53034 Zbl 1135.53023

[Mendonça 1997] S. J. Mendonça, "The asymptotic behavior of the set of rays", *Comment. Math. Helv.* **72**:3 (1997), 331–348. MR 98g:53060 Zbl 0910.53026

[Petersen 2006] P. Petersen, *Riemannian geometry*, 2nd ed., Graduate Texts in Mathematics **171**, Springer, New York, 2006. MR 2007a:53001 Zbl 1220.53002

[Sakai 1996] T. Sakai, *Riemannian geometry*, Translations of Mathematical Monographs **149**, American Mathematical Society, Providence, RI, 1996. MR 97f:53001 Zbl 0886.53002

[Shiohama and Tanaka 2002] K. Shiohama and M. Tanaka, "Compactification and maximal diameter theorem for noncompact manifolds with radial curvature bounded below", *Math. Z.* **241**:2 (2002), 341–351. MR 2003i:53048 Zbl 1020.53019

[Shiohama et al. 2003] K. Shiohama, T. Shioya, and M. Tanaka, *The geometry of total curvature on complete open surfaces*, Cambridge Tracts in Mathematics **159**, Cambridge University Press, 2003. MR 2005c:53037 Zbl 1086.53056

[Sinclair and Tanaka 2006] R. Sinclair and M. Tanaka, "A bound on the number of endpoints of the cut locus", *LMS J. Comput. Math.* **9** (2006), 21–39. MR 2006i:53039 Zbl 1111.53004

[Tanaka 1992a] M. Tanaka, "On a characterization of a surface of revolution with many poles", *Mem. Fac. Sci. Kyushu Univ. Ser. A* **46**:2 (1992), 251–268. MR 93m:53033 Zbl 0781.53034

[Tanaka 1992b] M. Tanaka, "On the cut loci of a von Mangoldt's surface of revolution", *J. Math. Soc. Japan* **44**:4 (1992), 631–641. MR 93h:53034 Zbl 0789.53023

Received November 14, 2011. Revised June 5, 2012.

IGOR BELEGRADEK
SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
686 CHERRY STREET
ATLANTA, GA 30332-0160
UNITED STATES
ib@math.gatech.edu

ERIC CHOI
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
EMORY UNIVERSITY
400 DOWMAN DR., W401
ATLANTA, GA 30322
UNITED STATES
ericchoi314@gmail.com

NOBUHIRO INNAMI
DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
NIIGATA UNIVERSITY
NIIGATA 950-2181
JAPAN
innami@math.sc.niigata-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 259, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.259.307

ISOPERIMETRIC SURFACES WITH BOUNDARY, II

ABRAHAM FRANDSEN, DONALD SAMPSON AND NEIL STEINBURG

Following our previous work with Dorff and Lawlor, we extend results for the so-called equitent problem of fixed boundary and fixed volume. We define sufficient conditions, which in $\mathbb{R}^2$ and $\mathbb{R}^3$ are also necessary, for local minima to be piecewise spherical, and we show that these are area-minimizing in their homotopy class. We also give new examples of these surfaces in $\mathbb{R}^2$ and $\mathbb{R}^3$.

1. Introduction

Equitent problems, first introduced in the paper "Isoperimetric surfaces with boundary" [Dorff et al. 2011], ask what is the area-minimizing surface enclosing a given volume and spanning a given boundary. In this way, equitent problems represent a combination of isoperimetry and boundary conditions, such as in Steiner problems and minimal surfaces. Our previous approach, which we extend here, uses the technique of metacalibration. *Metacalibration* is a version of the calibration methods popularized by Harvey and Lawson [1982], adapted to use on isoperimetric problems. In particular, we use a combination of the mapping of [Gromov 1986], after [Knothe 1957], and the paired calibrations of [Lawlor and Morgan 1994].

In our original results, we construct various classes of surfaces bounded by the dual figures of uniform polytopes and enclosing a prescribed volume and prove that these surfaces are minimizing in their homotopy class. The results, however, turn out to be limited in scope, as shown in [Ross et al. 2011]. Consequently, there remains much ground to be covered.

In this paper, we will extend previous results by considering equitent systems generated by polytopes whose edges are all of a given length. This results in a much wider range of equitent surfaces than those bounded by uniform polytopes. We construct the conjectured minimizing surface using a refinement of previous methods and prove that this surface is indeed area-minimizing in its homotopy class.

Further, we show that any homotopically area-minimizing equitent surface with piecewise spherical faces and simplex vertex figures is equivalent to one generated

MSC2010: primary 53C38; secondary 49Q10.
Keywords: metacalibration, calibration, isoperimetric, equitent.

by our construction. We conclude with a discussion of new equitent surfaces and a survey of open problems.

2. The surfaces

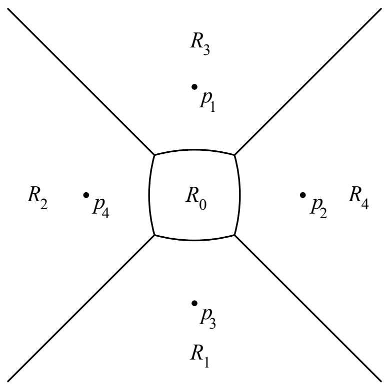
Let Γ be a convex polytope of dimension $m \leq n$ with equal edge lengths, r, embedded in $\mathbb{R}^n$. Let $p_1, \ldots, p_k$ be the vertices of Γ. For each p_i let R_i be the region farthest from p_i:

$$R_i = \{x \in \mathbb{R}^n : r < \|x - p_i\| \text{ and } \|x - p_j\| < \|x - p_i\| \text{ for all } j \neq i\}.$$

Note that if p_i and p_j share an edge in Γ, $\partial R_i \cap \partial R_j$ is a subset of the perpendicular bisecting hyperplane of that edge. Now, define

$$R_0 = \{x \in \mathbb{R}^n : \|x - p_i\| < r \text{ for } 1 \leq i \leq k\}.$$

See figure.



This region represents the enclosed volume. We suppose that $R_0 \neq \varnothing$ and $\mathcal{H}^{n-1}(\partial R_0 \cap \partial R_i) \neq 0$ for all $i > 0$. Let $V_0 = \mathcal{H}^n(R_0)$. Then let

$$M = \bigcup_{i=0}^{k} \partial R_i.$$

Notice that ∂R_0 is the portion of the surface that encloses the volume R_0. In order to have a nontrivial result, we require $R_0 \neq \varnothing$. The condition

$$\mathcal{H}^{n-1}(\partial R_0 \cap \partial R_i) \neq 0$$

for all $i > 0$ ensures that all smooth subsurfaces of M meet at 120 degree angles. For $m = 2$, the only viable generating figures, Γ, are equilateral triangles, rhombi with interior angles strictly greater than 60°, and small perturbations of regular pentagons. For $m = 3$, the valid generating figures include all but two of the eight convex deltahedra (polyhedra where all faces are equilateral triangles), as well as other polytopes with faces of higher degree. It is worth noting, however, that those generating figures, Γ, whose faces are not equilateral triangles produce surfaces which are locally minimal within their homotopy class, but not globally minimal. As will be seen in the proof, this construction gives sufficient conditions for minimizing surfaces to be piecewise spherical. Furthermore, due to the regularity properties of soap films proved by Taylor and Almgren [Taylor 1976], these are also necessary conditions in $\mathbb{R}^2$ and $\mathbb{R}^3$. In higher dimensions nonsimplicial vertex figures may be minimizing, but are not considered in this paper. See for example [Brakke 1991].

3. The minimization theorem

In this section we prove that the surfaces constructed are homotopically minimizing in the following sense: Let U be a bounded open set that contains $\overline{R_0}$, and let $M_0 = M \cap U$.

Theorem 1. *The surface M_0 is area-minimizing among all compact surfaces* (*rectifiable sets*) *in U with boundary $\partial U \cap M$ that enclose the fixed volume $\mathcal{H}^n(R_0)$ and are homotopically equivalent to M_0. This also holds with the weaker assumption that competitor surfaces are not necessarily homotopic to M_0 but separate space into the same regions as M_0 and these regions share boundary nontrivially* (*on a set of positive $\mathcal{H}^{n-1}$ measure*) *only if the corresponding regions do in M_0.*

Our proof uses a metacalibration argument that compares figures according to their flux on specially crafted vector fields. In particular, we use a paired-calibration approach with one vector field defined for each separated region.

Let N be any competitor surface and let S_i be the separated regions that correspond to each R_i respectively. (Then $\mathcal{H}^n(S_0) = \mathcal{H}^n(R_0)$ by the volume condition.) Define $v_i : S_i \to \mathbb{R}^n$ for $1 \le i \le m$ to be the constant vector field $-p_i/r$. Let $\phi : S_0 \to R_0$ be the Knothe–Rosenblatt rearrangement and let $v_0 : S_0 \to \mathbb{R}^n$ be given by $v_0 = \phi/r$.

At this point a few simple results would be useful:

Proposition 2. *If S_i and S_j share boundary nontrivially, then $v_i - v_j$ is a unit vector. If $N = M_0$ then $v_i - v_j$ is the unit normal to $\partial R_i \cap \partial R_j$.*

Proof. Note that S_i and S_j share boundary nontrivially if and only if p_i and p_j are adjacent in Γ. Thus $\|v_i - v_j\| = (1/r)\|p_i - p_j\| = 1$. Also if $N = M_0$, $v_i - v_j$ is the unit normal to $\partial R_i \cap \partial R_j$ since $\partial R_i \cap \partial R_j$ lies on the hyperplane equidistant to p_i and p_j. □

Proposition 3. *The matrix Dv_0 is triangular. If $N = M$ then v_0 is the identity scaled by* $1/r$.

Proof. Follows from the definition of v_0. See [Dorff et al. 2011] for details. □

Proposition 4. *For* $i \neq 0$, $\int_{N\cap\partial S_i} v_i \cdot n \, d\mathcal{H}^{n-1} = \int_{M\cap\partial R_i} v_i \cdot n \, d\mathcal{H}^{n-1}$, *where* n *is the unit normal to the surface of integration, outward pointing with respect to* S_i *or* R_i.

Proof. Follows from the divergence theorem since v_i is divergence free and $\partial(M \cap \partial R_i) = \partial(N \cap \partial S_i)$. □

Proof of Theorem 1. For any competitor surface N, let $G(N) = \sum_i \int_{N\cap\partial S_i} v_i \cdot n \, d\mathcal{H}^{n-1}$. Letting $P(N) = \sum_i \int_{N\cap\partial S_i} d\mathcal{H}^{n-1}$ be our objective function, we find that

$$\begin{aligned} G(N) &= \sum_i \int_{N\cap\partial S_i} v_i \cdot n \, d\mathcal{H}^{n-1} \\ &= \sum_{i\neq j} \int_{N\cap(\partial S_i \cap \partial S_j)} (v_i - v_j) \cdot n \, d\mathcal{H}^{n-1} \\ &\leq \sum_{i\neq j} \int_{N\cap(\partial S_i \cap \partial S_j)} \|v_i - v_j\| \, \|n\| \, d\mathcal{H}^{n-1} \\ &\leq \sum_{i\neq j} \int_{N\cap(\partial S_i \cap \partial S_j)} d\mathcal{H}^{n-1} \\ &= \sum_i \int_{N\cap\partial S_i} d\mathcal{H}^{n-1} = P(N), \end{aligned}$$

with equality if $N = M_0$.

Now also note that

$$\begin{aligned} \int_{N\cap\partial S_0} v_0 \cdot n \, d\mathcal{H}^{n-1} &= \int_{S_0} \operatorname{div} v_0 \, d\mathcal{H}^{n-1} \\ &= \int_{S_0} \frac{1}{r} \left(\frac{\partial \phi_1}{\partial x_1} + \frac{\partial \phi_2}{\partial x_2} + \cdots + \frac{\partial \phi_n}{\partial x_n} \right) d\mathcal{H}^{n-1} \\ &\geq \int_{S_0} \frac{n}{r} \sqrt[n]{\frac{\partial \phi_1}{\partial x_1} \frac{\partial \phi_2}{\partial x_2} \cdots \frac{\partial \phi_n}{\partial x_n}} \, d\mathcal{H}^{n-1} \\ &= \int_{S_0} \frac{n}{r} \sqrt[n]{1} \, d\mathcal{H}^{n-1} \\ &= \frac{n}{r} \mathcal{H}^{n-1}(S_0) = \frac{n}{r} \mathcal{H}^{n-1}(R_0), \end{aligned}$$

with equality if $N = M_0$. This follows from the AM-GM inequality and the equality

$$\frac{\partial \phi_1}{\partial x_1} \frac{\partial \phi_2}{\partial x_2} \cdots \frac{\partial \phi_n}{\partial x_n} = \det(D\phi) = 1,$$

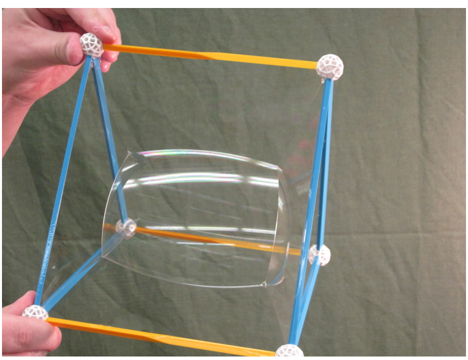
which is valid since ϕ is volume-preserving.

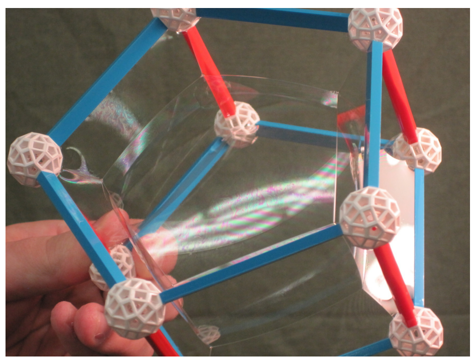
Combining these results we find

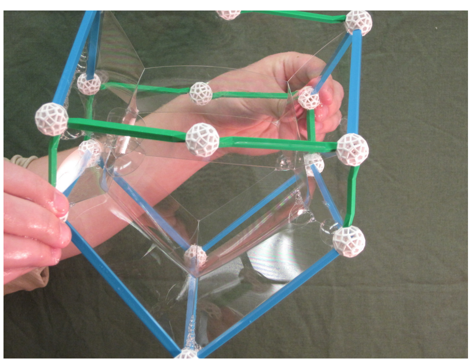
$$\begin{aligned}P(M_0) = G(M_0) &= \sum_{i\neq 0} \int_{M_0 \cap \partial R_i} v_i \cdot n \, d\mathcal{H}^{n-1} + \int_{M_0 \cap \partial R_0} v_0 \cdot n \, d\mathcal{H}^{n-1} \\ &= \sum_{i\neq 0} \int_{N \cap \partial S_i} v_i \cdot n \, d\mathcal{H}^{n-1} + \frac{n}{r}\mathcal{H}^{n-1}(R_0) \\ &\leq \sum_{i\neq 0} \int_{N \cap \partial S_i} v_i \cdot n \, d\mathcal{H}^{n-1} + \int_{N \cap \partial S_0} v_0 \cdot n \, d\mathcal{H}^{n-1} \\ &= G(N) \leq P(N). \qquad \square\end{aligned}$$

4. Soap films in $\mathbb{R}^3$

In [Dorff et al. 2011] and [Ross et al. 2011] we identified the regular tetrahedron, the regular octahedron, and the regular icosahedron as polytopes that generate realizable soap films. The only other three-dimensional polytopes, Γ, that generate surfaces realizable as soap films are the triangular dipyramid, the pentagonal dipyramid, and the snub disphenoid. The generated soap films are shown below.







This is due to the conditions proven by Jean Taylor [1976], namely that each face not intersecting with the boundary must meet with exactly two other faces in 120-degree angles. Thus, any generating figure, Γ, with nontriangular faces will not yield a surface realizable as a soap film. The remaining two deltahedra fail to meet the conditions of our construction because of their large circumradius.

Every surface generated by our construction will have piecewise spherical faces and simplicial vertex figures. The converse is also true. Given any area-minimizing equitent surface with piecewise spherical faces and simplicial vertex figures, we can recover the generating polytope, Γ, as the set of centers of each spherical face. In higher dimensions there may exist area-minimizing equitent surfaces with nonsimplicial vertex figures.

5. Conclusion

We have characterized all piecewise spherical equitent surfaces in two and three dimensions, and proven them to be area minimizing. Several interesting and open problems arise. An especially intriguing question deals with equitent surfaces that have negatively curved bubbles, meaning each face of the bubble region bends inward. Such a surface can be created using soap films, but our methods are not yet able to address this case. Similarly, equitent surfaces with nonspherical faces fall outside the scope of our approach. Finally, our construction generates surfaces whose vertices are cones over simplices. In spaces of dimension greater than three, however, minimal surfaces need not have simplicial vertex figures, and we may yet find interesting new equitent surfaces. Extensions of the metacalibration methods outlined in this paper show great promise in solving these open problems.

References

[Brakke 1991] K. A. Brakke, "Minimal cones on hypercubes", *J. Geom. Anal.* **1**:4 (1991), 329–338. MR 92k:49082 Zbl 0738.49022

[Dorff et al. 2011] R. Dorff, D. Johnson, G. R. Lawlor, and D. Sampson, "Isoperimetric surfaces with boundary", *Proc. Amer. Math. Soc.* **139**:12 (2011), 4467–4473. MR 2012f:53101 Zbl 05988131

[Gromov 1986] M. Gromov, "Appendix I: Isoperimetric inequalities in Riemannian manifolds", pp. 114–129 in *Asymptotic theory of finite-dimensional normed spaces,* by V. D. Milman and G. Schechtman, Lecture Notes in Mathematics **1200**, Springer, Berlin, 1986.

[Harvey and Lawson 1982] R. Harvey and H. B. Lawson, Jr., "Calibrated geometries", *Acta Math.* **148** (1982), 47–157. MR 666108 (85i:53058) Zbl 0584.53021

[Knothe 1957] H. Knothe, "Contributions to the theory of convex bodies", *Michigan Math. J.* **4** (1957), 39–52. MR 18,757b Zbl 0077.35803

[Lawlor and Morgan 1994] G. Lawlor and F. Morgan, "Paired calibrations applied to soap films, immiscible fluids, and surfaces or networks minimizing other norms", *Pacific J. Math.* **166**:1 (1994), 55–83. MR 95i:58051 Zbl 0830.49028

[Ross et al. 2011] J. Ross, D. Sampson, and N. Steinburg, "Soap film realization of isoperimetric surfaces with boundary", *Involve* **4**:3 (2011), 271–276.

[Taylor 1976] J. E. Taylor, "The structure of singularities in soap-bubble-like and soap-film-like minimal surfaces", *Ann. of Math.* (2) **103**:3 (1976), 489–539. MR 55 #1208a Zbl 0335.49032

Received January 26, 2012.

ABRAHAM FRANDSEN
BRIGHAM YOUNG UNIVERSITY
PROVO, UT 84604
UNITED STATES
abefrandsen@gmail.com

DONALD SAMPSON
BRIGHAM YOUNG UNIVERSITY
PROVO, UT 84604
UNITED STATES
dsampson@byu.net

NEIL STEINBURG
BRIGHAM YOUNG UNIVERSITY
PROVO, UT 84604
UNITED STATES
neil.steinburg@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 259, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.259.315

CYCLIC BRANCHED COVERINGS OF KNOTS AND QUANDLE HOMOLOGY

YUICHI KABAYA

We give a construction of quandle cocycles from group cocycles, especially, for any integer $p \geq 3$, quandle cocycles of the dihedral quandle R_p from group cocycles of the cyclic group $\mathbb{Z}/p$. We show that a group 3-cocycle of $\mathbb{Z}/p$ gives rise to a nontrivial quandle 3-cocycle of R_p. When p is an odd prime, since $\dim_{\mathbb{F}_p} H^3_Q(R_p; \mathbb{F}_p) = 1$, our 3-cocycle is a constant multiple of the Mochizuki 3-cocycle up to coboundary. Dually, we construct a group cycle represented by a cyclic branched covering branched along a knot K from the quandle cycle associated with a colored diagram of K.

1. Introduction

A quandle, which was introduced by Joyce [1982], is an algebraic object whose axioms are motivated by knot theory and conjugation in a group. Carter, Jelsovsky, Kamada, Langford and Saito [Carter et al. 2003] introduced a quandle homology theory, and they defined the quandle cocycle invariants for classical knots and surface knots. The quandle homology is defined as the homology of a chain complex generated by cubes whose edges are labeled by elements of a quandle. On the other hand, the group homology is defined as the homology of a chain complex generated by tetrahedra whose edges are labeled by elements of a group. So it is natural to ask for a relation between quandle homology and group homology. This question also arises from the fact that the quandle cocycle invariants were defined as an analogue of the Dijkgraaf–Witten invariants, which are defined using group cocycles.

In [Inoue and Kabaya 2010], we defined a simplicial version of quandle homology and constructed a homomorphism from the usual quandle homology to the simplicial quandle homology. The important point is that this homomorphism gives a triangulation of a knot complement in algebraic fashion. This enables us to relate the quandle homology to the topology of knot complements.

The author is supported by JSPS Research Fellowships for Young Scientists.
MSC2010: 57M05, 57M10, 57M12, 57M25, 57M27.
Keywords: quandle homology, group homology, cyclic branched covering.

In this paper, we apply the results of [Inoue and Kabaya 2010] to construct quandle cocycles from group cocycles. First, we demonstrate how to give a quandle cocycle of the dihedral quandle R_p from a group cocycle of the cyclic group $\mathbb{Z}/p$ for any integer $p \geq 3$ in Section 7. We show that a generator of $H^3(\mathbb{Z}/p; \mathbb{Z}/p)$ gives rise to a nontrivial quandle 3-cocycle of $H^3_Q(R_p; \mathbb{Z}/p)$. When p is an odd prime, since $\dim_{\mathbb{F}_p} H^3_Q(R_p; \mathbb{F}_p) = 1$, our quandle 3-cocycle is equal to a constant multiple of the Mochizuki 3-cocycle [2003] up to coboundary.

Then we generalize the construction to wider classes of quandles. Let G be a group and let h be an element of G; then the set $\mathrm{Conj}(h) = \{g^{-1}hg \mid g \in G\}$ forms a quandle by conjugation. (It is known from [Joyce 1982] that any faithful homogeneous quandle has such a presentation.) When some obstruction in second cohomology vanishes, we construct a quandle cocycle of $\mathrm{Conj}(h)$ from a group cocycle of G.

Dually, we relate the quandle cycle associated with an arc and region coloring (shadow coloring) of a knot K to a group cycle represented by a cyclic branched covering branched along K. Let D be a diagram of K. We can define the notion of arc and region colorings of D by a quandle $\mathrm{Conj}(h)$. A pair of an arc and a region coloring is called a *shadow coloring*. We can associate a cycle of a quandle homology group to a shadow coloring of D. Using the homomorphism constructed in [Inoue and Kabaya 2010], we construct a group cycle of G represented by a cyclic branched covering branched along K. This reveals a close relationship between the shadow cocycle invariant of a knot and the Dijkgraaf–Witten invariant of the cyclic branched cover.

Hatakenaka and Nosaka [2012] defined an invariant of 3-manifolds called the *4-fold symmetric quandle homotopy invariant*, based on the fact that any 3-manifold can be represented as a 4-fold simple branched covering of S^3 along a link. As an application, they showed that the shadow cocycle invariant of a link for the Mochizuki 3-cocycle is equal to a scalar multiple of the Dijkgraaf–Witten invariant of the double branched cover along the link.

This paper is organized as follows. In Section 2, we recall the definition of group homology and show how to represent a group cycle by a triangulation with a labeling of its 1-simplices. We give a presentation of the fundamental group of a cyclic branched covering branched along a knot in Section 3, which is independent from the other sections. In Section 4, we construct a group cycle represented by a cyclic branched cover. We recall the definition of quandles and their homology theory in Section 5. We review some results from [Inoue and Kabaya 2010] in Section 6 and apply them to construct quandle cocycles of the dihedral quandle R_p in Section 7. The reader interested in the form of the 3-cocycle should consult (7-5) (and (7-4)). We generalize the construction to wider classes of quandles in Section 8. In Section 9, we construct a group cycle represented by a cyclic branched covering from the quandle cycle associated with a shadow coloring. The reader who is only interested in the construction of quandle cocycles from group cocycles may skip Sections 2B 2C 3, and 4,

2. Group homology

In this section, we collect basic facts on group homology, starting with a review of definitions; see [Brown 1982] for details. The material discussed in Sections 2B and 2C was developed in [Neumann 2004].

2A. ***Group homology.*** Let G be a group. Let $C_n(G)$ be the free $\mathbb{Z}[G]$-module generated by n-tuples $[g_1|\dots|g_n]$ of elements of G. Define the boundary map $\partial : C_n(G) \to C_{n-1}(G)$ by

$$\partial([g_1|\dots|g_n]) \\ = g_1[g_2|\dots|g_n] + \sum_{i=1}^{n-1}(-1)^i\big[g_1|\dots|g_i g_{i+1}|\dots|g_n\big] + (-1)^n[g_1|\dots|g_{n-1}].$$

We remark that the chain complex $\{\cdots \to C_1(G) \to C_0(G) \to \mathbb{Z} \to 0\}$ is acyclic, where $C_0(G) \cong \mathbb{Z}[G] \to \mathbb{Z}$ is the augmentation map. So the chain complex $C_*(G)$ gives a free resolution of $\mathbb{Z}$. Let M be a right $\mathbb{Z}[G]$-module. The homology of $C_n(G; M) = M \otimes_{\mathbb{Z}[G]} C_n(G)$ is called the *group homology* of M and denoted by $H_n(G; M)$. In other words, $H_n(G; M) = \mathrm{Tor}_n^{\mathbb{Z}[G]}(M, \mathbb{Z})$.

Let $C'_n(G)$ be the free $\mathbb{Z}$-module generated by $(g_0, \dots, g_n) \in G^{n+1}$. Then $C'_n(G)$ is a left $\mathbb{Z}[G]$-module by the action $g(g_0, \dots, g_n) = (gg_0, \dots, gg_n)$. Define the boundary operator of $C'_n(G)$ by

$$\partial(g_0, \dots, g_n) = \sum_{i=0}^{n}(-1)^i(g_0, \dots, \widehat{g_i}, \dots, g_n).$$

$C_*(G)$ and $C'_*(G)$ are isomorphic as chain complexes. In fact, the following correspondence gives an isomorphism:

$$[g_1|g_2|\dots|g_n] \leftrightarrow (1, g_1, g_1g_2, \dots, g_1\cdots g_n),$$

or, equivalently,

$$g_0\big[g_0^{-1}g_1|g_1^{-1}g_2|\dots|g_{n-1}^{-1}g_n\big] \leftrightarrow (g_0, \dots, g_n).$$

The notation that uses $(g_0, \dots, g_n)$ is called *homogeneous*, and the one that uses $[g_1|\dots|g_n]$ is called *inhomogeneous*.

Factoring out $C_n(G)$ by the degenerate subcomplex generated by $[g_1|\dots|g_n]$ such that $g_i = 1$ for some i, we obtain the *normalized* chain complex and its homology group. It is known that the group homology using the normalized chain complex coincides with the homology using the unnormalized one. In the homogeneous notation, we factor out $C'_n(G)$ by the subcomplex generated by $(g_0, \dots, g_n)$ such that $g_i = g_{i+1}$ for some i to define the normalized chain complex.

For a left $\mathbb{Z}[G]$-module N, the group cohomology $H^n(G; N)$ is defined as the cohomology of the cochain complex $C^n(G; N) = \mathrm{Hom}_{\mathbb{Z}[G]}(C_n(G), N)$. Let A be an abelian group. A cocycle of $C^n(G; A)$ in the homogeneous notation is a function $f : G^{n+1} \to A$ satisfying the following conditions:

(1) $\displaystyle\sum_{i=0}^{n+1}(-1)^i f(x_0, \dots, \widehat{x_i}, \dots, x_{n+1}) = 0,$

(2) $f(gx_0, \dots, gx_n) = f(x_0, \dots, x_n)$ for any $g \in G$ (left invariance).

If f also satisfies

(3) $f(x_0, \dots, x_n) = 0$ if $x_i = x_{i+1}$ for some i,

then f is a normalized n-cocycle. We can show that any n-cocycle is cohomologous to a normalized n-cocycle.

2B. ***Cycles represented by triangulations.*** Let Δ be an n-dimensional simplex. We label the vertices of Δ by $0, 1, \dots, n$. A face of Δ is presented by a subset of $\{0, 1, \dots, n\}$. Let $\langle i_0, \dots, i_k\rangle$ be the face spanned by $i_0, \dots, i_k \in \{0, \dots, n\}$, with a vertex ordering given by $i_0, \dots, i_k$. Any face inherits a vertex ordering from the vertex ordering of Δ, that is, $\langle i_0, \dots i_k\rangle$ with $i_0 < i_1 < \cdots < i_k$.

Definition 2.1. Let T be a CW-complex obtained by gluing a finite number of n-dimensional simplices along their $(n-1)$-dimensional faces in pairs by simplicial homeomorphisms. We denote the k-skeleton of T by $T^{(k)}$. We assume that the gluing maps preserve the vertex orderings of the faces. Then $T - T^{(n-3)}$ is homeomorphic to a topological n-manifold (not orientable in general). When $T - T^{(n-3)}$ is oriented, we call T an *ordered n-cycle*.

Consider an n-cycle σ of $C_n(G; \mathbb{Z})$. Then σ is represented by a sum

$$\sum_j \epsilon_j [g_{j1}| \dots |g_{jn}],$$

where $\epsilon_j = \pm 1$ and $g_{jk} \in G$. For each $[g_{j1}| \dots |g_{jn}]$, take an n-simplex Δ_j. Then label the edge $\langle i_1 i_2\rangle$ of Δ_j by $g_{ji_1} g_{j(i_1+1)} \dots g_{ji_2}$ for $i_1 < i_2$. In particular, we have

$$\langle 0, 1\rangle \leftrightarrow g_{j1}, \quad \langle 1, 2\rangle \leftrightarrow g_{j2}, \quad \dots, \quad \langle n-1, n\rangle \leftrightarrow g_{jn}.$$

We denote the label of $\langle i_1, i_2\rangle$ by $\lambda\langle i_1, i_2\rangle$. For $i_1 > i_2$, label the oriented edge $\langle i_1, i_2\rangle$ by $\lambda\langle i_2, i_1\rangle^{-1}$. For any 2-dimensional face $\langle i_0, i_1, i_2\rangle$, we have

$$\lambda\langle i_0, i_1\rangle \lambda\langle i_1, i_2\rangle = \lambda\langle i_0, i_2\rangle.$$

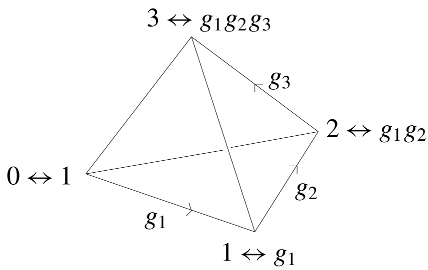


Figure 1. A labeling of a simplex.

Rewriting them in the homogeneous notation, we assign labels to the vertices of Δ_j as

$$0 \leftrightarrow 1, \quad 1 \leftrightarrow g_{j1}, \quad 2 \leftrightarrow g_{j1}g_{j2}, \quad \dots, \quad n \leftrightarrow g_{j1}\dots g_{jn},$$

up to the left action of G (Figure 1).

Since $\partial\sigma = 0$, $(n-1)$-dimensional faces cancel in pairs. Gluing the Δ_j's along their faces according to such pairings, we obtain an n-cycle T. At any $(n-1)$-simplex of T, there exist exactly two adjacent n-simplices. The labelings of the $(n-1)$-simplex derived from these two n-simplices coincide. Thus we have a well-defined labeling of 1-simplices λ : {oriented 1-simplices of T} $\to G$ satisfying:

(1) $\lambda\langle i_0, i_1\rangle\lambda\langle i_1, i_2\rangle = \lambda\langle i_0, i_2\rangle$ for any 2-dimensional face $\langle i_0, i_1, i_2\rangle$,

(2) $\lambda\langle i_1, i_0\rangle = \lambda\langle i_0, i_1\rangle^{-1}$.

We call a labeling of 1-simplices satisfying the conditions (1) and (2) a *G-valued 1-cocycle*. Orient Δ_j positively if $\epsilon_j = 1$ and negatively if $\epsilon_j = -1$. Since these orientations agree on face pairings, we thus have an orientation on T. Therefore T is an ordered n-cycle with a G-valued 1-cocycle λ. In general, T may not be connected, but we assume that T is connected because we can treat each connected component separately in our arguments. Conversely, any ordered n-cycle T with a G-valued 1-cocycle λ represents an n-cycle of $C_n(G; \mathbb{Z})$.

2C. *Group cycles and representations.* Suppose a cycle $\sigma \in C_n(G; \mathbb{Z})$ is represented by an ordered n-cycle T with a G-valued 1-cocycle λ. Then σ induces a homomorphism from $\pi_1(T)$ to G as follows. Let $\widetilde{T}$ be the universal covering of T and let $p : \widetilde{T} \to T$ be the covering map. Then the simplices of T lift to simplices of $\widetilde{T}$, and each lift has an induced vertex ordering compatible with adjacent n-simplices. The G-valued 1-cocycle λ of T induces a G-valued 1-cocycle of $\widetilde{T}$. Consider a *fundamental domain* of T, that is, a contractible subcomplex D of $\widetilde{T}$ such that

- $\widetilde{T} = \displaystyle\bigcup_{\gamma\in\pi_1(M)} \gamma D$,
- $D \cap \gamma D =$ (lower-dimensional simplices), for any $\gamma \neq 1$,

where we regard $\pi_1(T)$ as the group of deck transformations. By definition, the number of n-simplices in D coincides with the number of n-simplices in T; in particular, both are finite. We fix a base point $\widetilde{*}$ in the interior of D. Each $(n-1)$-simplex on ∂D is glued to another $(n-1)$-simplex on ∂D. We denote such a pair of faces by $F_i^{\pm}$. Let x_i be a path in T that starts at $* = p(\widetilde{*})$, traverses $p(F_i)$ in the direction from F_i^+ to F_i^-, and ends at $*$. These paths form a system of generators of the fundamental group $\pi_1(T, *)$. The relations are given at any $(n-2)$-simplices, around which there are a finite number of $p(F_i)$'s.

Fix an n-simplex Δ in D and a labeling of vertices of Δ derived from λ. Then n-simplices adjacent to Δ inherit labelings of vertices from λ. In this way, all vertices of D are labeled by elements of G. Now consider the labeling of vertices of F_i^+ and F_i^-. Since these reduce to the same labeling of edges, they coincide up to left multiplication. Therefore there exists an element of G that sends the labeling of vertices of F_i^- to the one of F_i^+. Denote the element by $\rho(x_i)$. This ρ induces a homomorphism $\rho : \pi_1(T, *) \to G$.

Conversely, if we have an ordered n-cycle T and a homomorphism

$$\rho : \pi_1(T, *) \to G,$$

we can construct a G-valued 1-cocycle λ and then a cycle of $C_n(G; \mathbb{Z})$ up to boundary as follows. Since ρ induces a map

$$T \to K(\pi_1(T, *), 1) \to BG,$$

we obtain a labeling of 1-simplices λ of T. This gives rise to a G-valued 1-cocycle λ and a homology class in $H_n(G; \mathbb{Z})$. The G-valued 1-cocycle λ is well-defined up to the coboundary action. A map $\mu : \{0\text{-simplices of } T\} \to G$ acts on a G-valued 1-cocycle λ as a coboundary action by

$$\langle i_1, i_2 \rangle \mapsto \mu(i_1)^{-1} \lambda \langle i_1, i_2 \rangle \mu(i_2).$$

We can show that the homology class obtained from λ does not change under the coboundary action. For any $g \in G$, the cycle corresponding to the representation $g^{-1}\rho g$ is obtained from λ by the coboundary action by $\mu \equiv g$. As a result, the homology class obtained from ρ depends only on the conjugacy class of ρ.

For a closed oriented n-manifold M and a representation $\rho : \pi_1(M) \to G$, we have a homology class defined by the image of the fundamental class $[M]$ under the map $H_n(M) \to H_n\big(K(\pi_1(M), 1)\big) \to H_n(G; \mathbb{Z})$. When M is homeomorphic to an ordered n-cycle T, the homology class is represented by a G-valued 1-cocycle λ of T associated with ρ. In this situation, we say that the homology class defined by M and ρ is *represented* by T and λ.

Figure 2. The relation is given by $x_k = x_j^{-1} x_i x_j$.

3. Cyclic branched covering

In this section, we give a presentation of the fundamental group of a cyclic branched covering from the Wirtinger presentation.

3A. ***Presentation of the fundamental group of the branched cover.*** Let K be a knot in S^3 and let D be a diagram of K. Then $\pi_1(S^3 - K)$ is presented by generators and relations called the Wirtinger presentation. Let $x_1, \dots, x_n$ be the generators of the Wirtinger presentation that correspond to the arcs of D. Each crossing (Figure 2) gives rise to a relation $x_k = x_j^{-1} x_i x_j$.

For any integer $l > 1$, let C_l be the l-fold cyclic covering of K, that is, the manifold corresponding to the kernel of

$$\pi_1(S^3 - K) \to H_1(S^3 - K) \cong \mathbb{Z} \to \mathbb{Z}/l.$$

Putting back the knot K to C_l, we obtain the l-fold cyclic branched covering $\widehat{C}_l$ of K.

Proposition 3.1. *$\pi_1(C_l)$ has the following presentation*:

Generators: $x_{i,s}$ (*for* $i = 1, 2, \dots, n$ *and* $s = 0, 1, \dots, l-1$),

Relations: $x_{k,s} = x_{j,s-1}^{-1} x_{i,s-1} x_{j,s}$ (*for each crossing and* $s = 0, 1, \dots, l-1$),

$$x_{1,0} = x_{1,1} = \cdots = x_{1,l-2} = 1.$$

The inclusion map $\pi_1(C_l) \to \pi_1(S^3 - K)$ is given by

$$x_{i,s} \mapsto x_1^{s-1} x_i x_1^{-s},$$

if we take appropriate base points. By adding a relation $x_{1,l-1} = 1$, we obtain a presentation of $\pi_1(\widehat{C}_l)$.

A method for obtaining a presentation of the fundamental group of a branched covering is given in [Rolfsen 1976]. But we give a proof here because some techniques are used to construct a group cycle represented by $\widehat{C}_l$ later.

Proof. First, we construct a handle decomposition of the knot complement associated with the Wirtinger presentation (Figure 3). Then we lift the handle decomposition

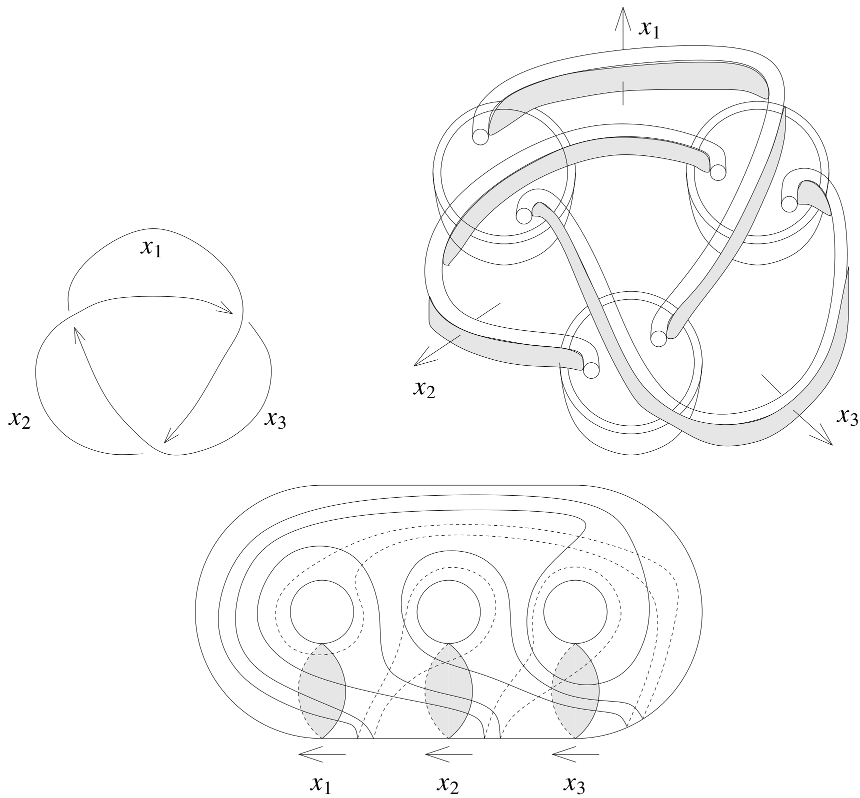


Figure 3. A handle decomposition of a knot complement.

to a handle decomposition of C_l (Figure 4). After that, we read the relations given by attaching 2-handles.

Let $N(K)$ be a regular neighborhood of K. We give a handle decomposition of $S^3 - N(K)$. We represent S^3 by the one-point compactification of

$$\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}.$$

Let $B_+ = \{z \geq 0\} \cup \{\infty\}$ and $B_- = \{z \leq 0\} \cup \{\infty\}$. We denote the equatorial sphere of S^3 by $S_0 = \{z = 0\} \cup \{\infty\}$. Put K in a position such that the projection to S_0 has only double points. Let n be the number of crossings of this projection. We deform K in the z-direction so that K intersects S_0 at $2n$ points and each of $B_\pm \cap K$ consists of n arcs. We call $B_+ \cap K$ over-crossing arcs and $B_- \cap K$ under-crossing arcs. Index the arcs of $B_+ \cap K$ by x_i $(i = 1, 2, \ldots, n)$.

Now $B_+ - N(K)$ is homeomorphic to a handlebody of genus n (Figure 3). Projecting the over-crossing arcs to S_0, we obtain a meridian disk system of the handlebody. We denote the meridian disk corresponding to x_i by D_i. For each under-crossing arc, attach a 2-handle $D^2 \times D^1$ along $\partial D^2 \times D^1$ to $B_+ - N(K)$

(Figure 3). Then the resulting manifold is homeomorphic to $(S^3 - N(K)) - B^3$, where B^3 is a 3-ball. Attaching B^3 along the boundary to $\partial((S^3 - N(K)) - B^3)$, the resulting manifold is homeomorphic to $S^3 - N(K)$. So we have a handle decomposition of $S^3 - N(K)$ into one 0-handle, n 1-handles, n 2-handles and one 3-handle. In this handle decomposition, 1-handles correspond to the Wirtinger generators and 2-handles to the Wirtinger relations. We denote the set of i-handles by h^i and $X^{(i)} = h^0 \cup \cdots \cup h^i$.

Next we consider the preimage of $X^{(1)}$ in C_l. Cut $X^{(1)}$ along the meridian disks D_i, and denote the resulting manifold by B, which is homeomorphic to a 3-ball. Let $*$ be a point in $B \subset S^3 - N(K)$. We take a loop in $X^{(1)}$ that starts at $*$, intersects the meridian disk D_i, and ends at $*$. Orient the loop so that it corresponds to the generator of $H_1(S^3 - K)$. By abuse of notation, we also denote this loop by x_i. Take a lift $B_0 \subset C_l$ of B and denote the preimage of $* \in B$ in B_0 by $\tilde{*}$. Then there exists a unique lift $\tilde{x}_i$ of the loop x_i starting at $\tilde{*}$. Since x_i corresponds to the generator of $H_1(S^3 - K)$, $\tilde{x}_i$ ends at another lift of B, which we denote by B_1. Similarly, the lift of x_i starting at B_1 ends at another lift, which we denote by B_2. Continuing this, we see that $B_l = B_0$ and all lifts of B will appear. Therefore the preimage of $X^{(1)}$ is decomposed into l 3-balls $B_0 \cup B_1 \cup \cdots \cup B_{l-1}$. The intersection of B_s and B_{s+1} consists of n disks, each of which is a lift of a meridian disk D_i. We denote this lifted disk by $D_{i,s}$ (Figure 4). It is easy to check that

$$\{D_{1,l-1}\} \cup \{D_{i,s} : i = 2, \ldots, n;\ s = 0, \ldots, l-1\}$$

forms a meridian disk system of the preimage of $X^{(1)}$. Denote by $\tilde{x}_{i,s}$ the lift of x_i starting at B_s and ending at B_{s+1}.

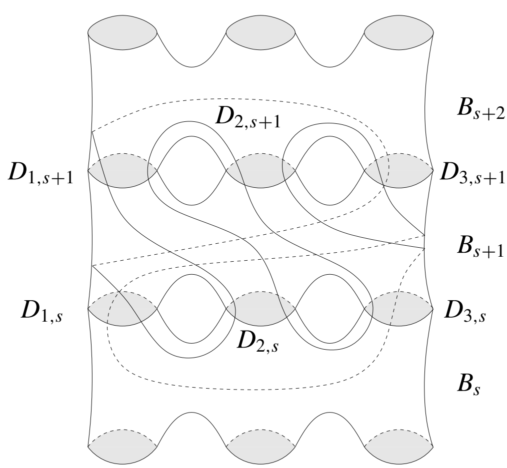


Figure 4. A handle decomposition of C_l.

The generator of $\pi_1(C_l, \widetilde{*})$ corresponding to the meridian disk $D_{i,s}$ is given by

$$\tilde{x}_{1,0}\tilde{x}_{1,1}\ldots\tilde{x}_{1,s-1}\tilde{x}_{i,s}\tilde{x}_{1,s}^{-1}\tilde{x}_{1,s-1}^{-1}\ldots\tilde{x}_{1,0}^{-1}. \tag{3-1}$$

We denote this element by $x_{i,s}$. To give a simple presentation of $\pi_1(C_l, \widetilde{*})$, we add extra generators $x_{1,0}, x_{1,1}, \ldots, x_{1,l-2}$ corresponding to $\tilde{x}_{1,0}, \tilde{x}_{1,1}, \ldots \tilde{x}_{1,l-2}$, respectively, and relations $x_{1,0} = x_{1,1} = \cdots = x_{1,l-2} = 1$.

Finally, we consider the relations given by the lifts of 2-handles. We see that the relation $x_k = x_j^{-1}x_ix_j$ lifts to

$$x_{k,s} = x_{j,s-1}^{-1}x_{i,s-1}x_{j,s} \quad (s = 0, 1, \ldots, l-1);$$

see Figure 4.

Since the generator $x_{i,s}$ is represented by (3-1), the inclusion map $\pi_1(C_l, \tilde{*}) \to \pi_1(S^3 - K, *)$ is given by $x_{i,s} \mapsto x_1^{s-1}x_sx_1^{-s}$. This proves the second statement.

By adding a 2-handle to C_l along $\tilde{x}_{1,0}\tilde{x}_{1,1}\ldots\tilde{x}_{1,l-1}$ and capping off the resulting sphere, we obtain a manifold homeomorphic to the cyclic branched covering $\widehat{C}_l$. Therefore a presentation of the cyclic branched covering is obtained by adding a relation $x_{1,l-1} = 1$. □

4. Cycle represented by cyclic branched covering

For a representation $\rho : \pi_1(S^3 - K) \to G$, we have the restriction map $\rho|_{\pi_1(C_l)} : \pi_1(C_l) \to G$ given by

$$\rho|_{\pi_1(C_l)}(x_{i,s}) = \rho(x_1)^{s-1}\rho(x_i)\rho(x_1)^{-s}. \tag{4-1}$$

If $\rho(x_1)^l = 1$, it reduces to a representation $\widehat{\rho} : \pi_1(\widehat{C}_l) \to G$ and there is a group cycle given by $\widehat{C}_l$ and $\widehat{\rho}$. In this section, we construct an explicit ordered 3-cycle and its G-valued 1-cocycle representing the homology class given by $\widehat{C}_l$ and $\widehat{\rho}$. First we give a triangulation of $S^3 - N(K)$ associated with the Wirtinger presentation.

Let K be a knot and fix a diagram of K. As in the proof of Proposition 3.1, we define $B_\pm$ and give a Heegaard splitting of $S^3 - N(K)$ with the meridian disks D_i. Cutting the handlebody $B_+ - N(K)$ along the meridian disks D_i, the result is a ball with $2n$ 2-cells on the boundary. We denote the resulting 3-ball by B and the pair of 2-cells corresponding to D_i by F_i^+ and F_i^- so that the Wirtinger generator corresponding to x_i runs from F_i^+ to F_i^- (Figure 5(A)). Now consider the attaching regions of the two handles corresponding to under-crossing arcs, which consist of annuli $S^1 \times D^1 (\subset D^2 \times D^1)$ on $\partial(B_+ - N(K))$. Each annulus is divided by the D_i's into four rectangles on ∂B. We define a graph on ∂B with vertices consisting of $F_i^\pm$ and edges consisting of these rectangles (Figure 5(B)). Each vertex of this graph has valency at least four. We can make all vertices of the graph into trivalent vertices (Figure 5(C)) by adding extra 1-handles and 2-handles (stabilizations of

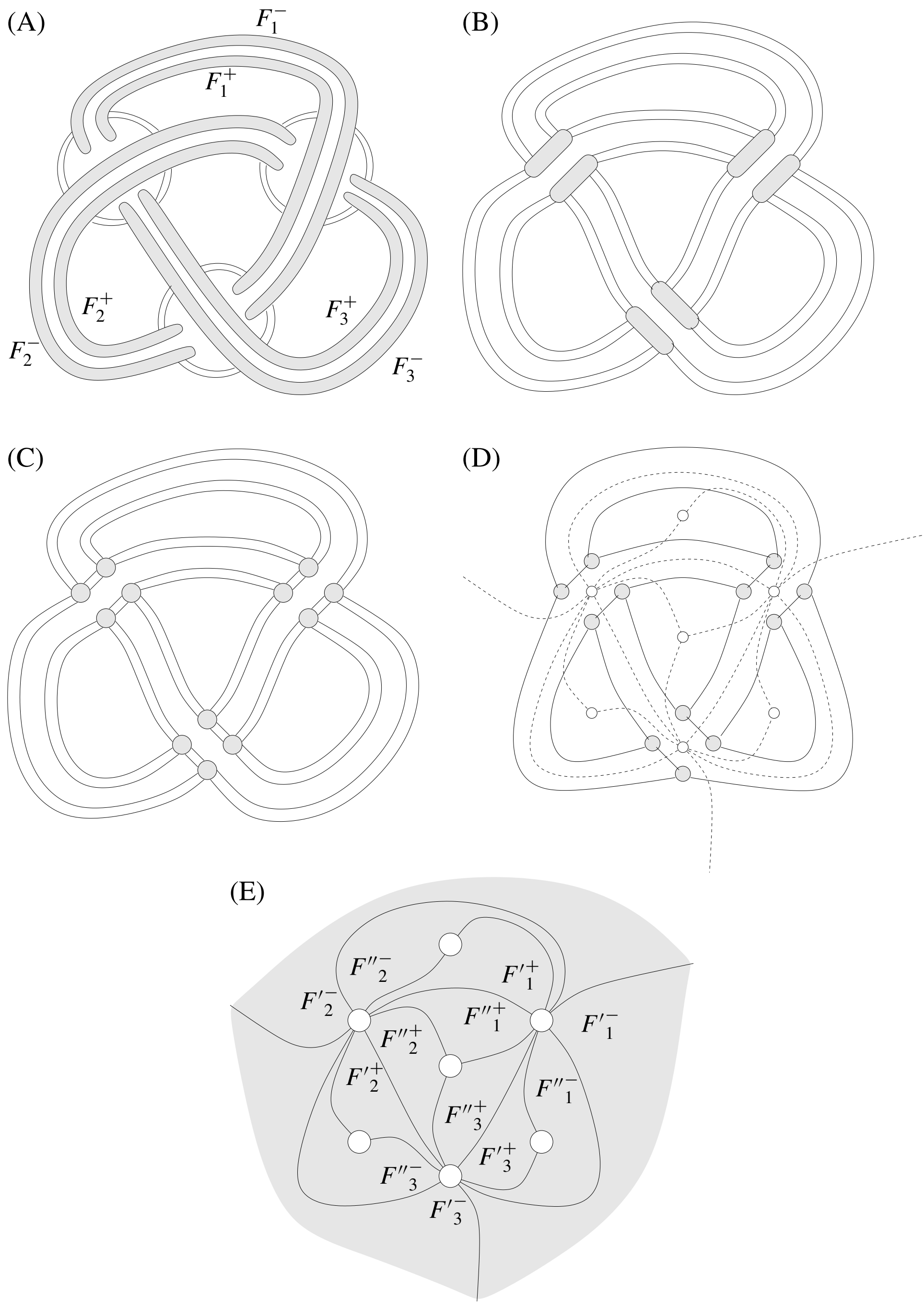


Figure 5. A polyhedral decomposition of $S^3 - N(K)$. We are looking from inside of B_+. The white rectangles in (A) and (B) correspond to the attaching regions of the 2-handles.

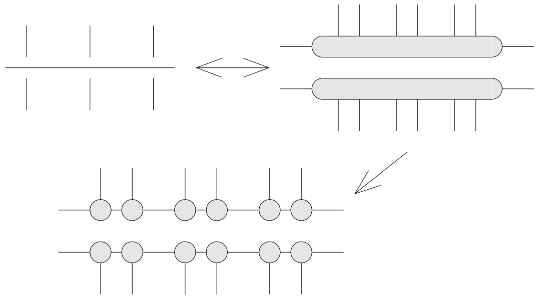


Figure 6. Stabilizations of the handle decomposition.

the Heegaard splitting; see Figure 6). We denote the new vertices, which originally belonged to $F_i^{\pm}$, by ${F'}_i^{\pm}$, ${F''}_i^{\pm}$, ... The dual of the graph gives a triangulation of ∂B (Figures 5(D) and 5(E)). By abuse of notation, we denote the triangles dual to the vertices ${F'}_i^{\pm}$, ${F''}_i^{\pm}$, ... by the same symbols. Taking a cone from an interior point of B, we obtain a triangulation T of B into $4n$ tetrahedra. Regluing the triangles ${F'}_i^{+}$, ${F''}_i^{+}$, ... to ${F'}_i^{-}$, ${F''}_i^{-}$, ..., we obtain a triangulation of $S^3 - N(K)$ into $4n$ tetrahedra, which was explained in [Weeks 2005]. We remark that this is not a triangulation in the usual sense: it is not a simplicial complex, and moreover, the link of some 0-simplex is not homeomorphic to the 2-sphere. Actually there exist only three 0-simplices; one is the cone point in B (the north pole in [Weeks 2005]), the second is the 0-simplex corresponding to the complementary regions of the diagram (the south pole), and the last one is a 0-simplex whose small neighborhood is homeomorphic to the cone over the torus $\partial N(K)$.

Using this triangulation, we construct a triangulation of the cyclic branched covering $\widehat{C}_l$. Let $B_0, B_1, \ldots, B_{l-1}$ be l copies of B and let T_s be the triangulation of B_s we have constructed. We denote the triangles ${F'}_i^{\pm}$, ${F''}_i^{\pm}$, ... on ∂B_s by ${F'}_{i,s}^{\pm}$, ${F''}_{i,s}^{\pm}$, ... respectively. By abuse of notation, we regard ${F'}_{i,s}^{\pm}$, ${F''}_{i,s}^{\pm}$, ... simply as $F_{i,s}^{\pm}$. Glue the T_s's along their boundary triangles by the identification maps

$$F_{i,s}^{-} \to F_{i,s-1}^{+} \quad (i = 1, 2, \ldots, n, \quad s = 0, 1, \ldots, l-1).$$

Denote this triangulation by $\widehat{T}$. We define an ordering of each tetrahedron by assigning 0 to the interior vertex of B_s, 1 to the vertex corresponding to the complementary region of the diagram, 2 to the vertex corresponding to the under-crossing arc, and 3 to the vertex corresponding to the over-crossing arc, respectively (Figure 7). The orderings are compatible under the gluing maps. So $\widehat{T}$ is an ordered 3-cycle. Here $\widehat{T}$ is a triangulation of $\widehat{C}_l$ except near the 0-simplex corresponding to K, whose small neighborhood is homeomorphic to a cone over a torus. We can resolve this singularity by inserting a suspension of a $2l$-gon around K (Figure 8). As a

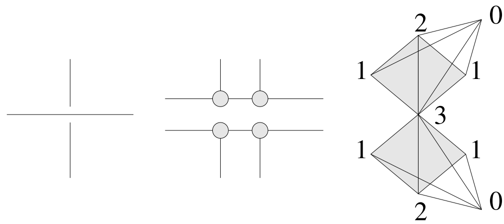


Figure 7. The orderings of the tetrahedra of B.

result, we obtain an ordered 3-cycle homeomorphic to $\widehat{C}_l$. (This procedure is called "blowing up" at a 0-simplex in [Neumann 2004].)

We construct a group cycle given by $\widehat{C}_l$ and a representation $\widehat{\rho} : \pi_1(\widehat{C}_l) \to G$ using the triangulation $\widehat{T}$. Here $\widehat{\rho}$ is given by the set $\{\widehat{\rho}(x_{i,s})\} \subset G$ satisfying

$$\widehat{\rho}(x_{k,s}) = \widehat{\rho}(x_{j,s-1})^{-1}\widehat{\rho}(x_{i,s-1})\widehat{\rho}(x_{j,s}),$$
$$\widehat{\rho}(x_{1,0}) = \widehat{\rho}(x_{1,1}) = \widehat{\rho}(x_{1,2}) = \cdots = \widehat{\rho}(x_{1,l-1}) = 1,$$

with i, j, k as in Figure 2 and $s = 0, 1, \ldots, l-1$. Give a labeling of vertices on T_s for each s. Let (g_1, g_2, g_3) be the labeling of vertices on $F^-_{i,s+1}$ and let (g'_1, g'_2, g'_3) be the labeling of vertices on $F^+_{i,s}$. If these are related by

$$\widehat{\rho}(x_{i,s})(g_1, g_2, g_3) = (g'_1, g'_2, g'_3),$$

then we obtain a G-valued 1-cocycle λ on $\widehat{T}$ by gluing T_s using $\widehat{\rho}$. To obtain a group cycle given by $\widehat{C}_l$ and $\widehat{\rho}$, we insert a suspension over a $2l$-gon to $\widehat{T}$ at the 0-simplex corresponding to K. We give an ordering on the vertices of each tetrahedron of the suspension compatible with the ordering on the $2l$-gon; for example, order the central 0-simplex maximal. These tetrahedra inherit a vertex labeling by G on the boundary faces of the suspension. We assign any labeling at the central 0-simplex of the suspension. Then upper and lower tetrahedra have the same labelings with different orientations. So these cancel out in pairs and give no contribution to the group cycle represented by $\widehat{C}_l$ and $\widehat{\rho}$. Therefore the homology class given by $\widehat{T}$ and λ represents the homology class given by $\widehat{C}_l$ and $\widehat{\rho}$.

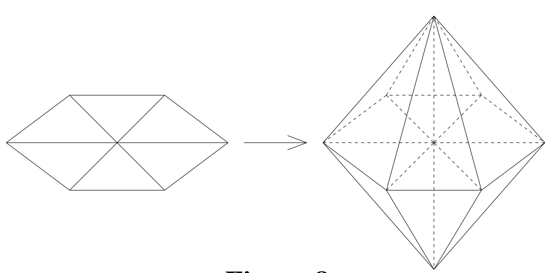


Figure 8

5. Quandle homology

In this section, we review the definitions of quandles, rack (co)homology and quandle (co)homology. Our treatment of quandle (or rack) homology follows that of [Etingof and Graña 2003]. In Sections 5B and 5C, we recall the notions of colorings and quandle cocycle invariants defined in [Carter et al. 2003].

5A. ***Quandle and quandle homology.*** A *quandle* X is a set with a binary operation $*$ satisfying the following axioms:

(Q1) $x * x = x$ for any $x \in X$,

(Q2) the map $*y : X \to X$ defined by $x \mapsto x * y$ is a bijection for any $y \in X$, and

(Q3) $(x * y) * z = (x * z) * (y * z)$ for any $x, y, z \in X$.

We denote the inverse of $*y$ by $*^{-1}y$. For a quandle X, we define the *associated group* G_X by

$$G_X = \langle x \in X \mid y^{-1}xy = x * y \ (x, y \in X) \rangle.$$

A quandle X has a right G_X-action in the following way. Let $g = x_1^{\epsilon_1} x_2^{\epsilon_2} \cdots x_n^{\epsilon_n}$ be an element of G_X where $x_i \in X$ and $\epsilon_i = \pm 1$. Define

$$x * g = \big(\dots ((x *^{\epsilon_1} x_1) *^{\epsilon_2} x_2) \dots \big) *^{\epsilon_n} x_n.$$

One can easily check that this is a right action of G_X on X. So the free abelian group $\mathbb{Z}[X]$ generated by X is a right $\mathbb{Z}[G_X]$-module.

Let $C_n^R(X)$ be the free left $\mathbb{Z}[G_X]$-module generated by X^n. We define the boundary map $C_n^R(X) \to C_{n-1}^R(X)$ by

$$\begin{aligned} &\partial(x_1, x_2, \dots, x_n) \\ &\quad = \sum_{i=1}^{n} (-1)^i \big((x_1, \dots, \widehat{x_i}, \dots, x_n) - x_i (x_1 * x_i, \dots, x_{i-1} * x_i, x_{i+1}, \dots, x_n) \big). \end{aligned}$$

Figure 9 shows a graphical picture of the boundary map. Let $C_n^D(X)$ be the $\mathbb{Z}[G_X]$-submodule of $C_n^R(X)$ generated by $(x_1, \dots, x_n)$ with $x_i = x_{i+1}$ for some i. Now $C_n^D(X)$ is a subcomplex of $C_n^R(X)$. Let $C_n^Q(X) = C_n^R(X)/C_n^D(X)$. For a right $\mathbb{Z}[G_X]$-module M, we define the *rack homology* of M by the homology of $C_n^R(X; M) = M \otimes_{\mathbb{Z}[G_X]} C_n^R(X)$ and denote it by $H_n^R(X; M)$. We also define the *quandle homology* of M by the homology of $C_n^Q(X; M) = M \otimes_{\mathbb{Z}[G_X]} C_n^Q(X)$ and denote it by $H_n^Q(X; M)$. The homology $H_n^Q(X; \mathbb{Z})$, where $\mathbb{Z}$ is the trivial $\mathbb{Z}[G_X]$-module, is equal to the usual quandle homology $H_n^Q(X)$. Let Y be a set with a right G_X-action. For any abelian group A, the abelian group $A[Y]$ freely generated by Y over A is a right $\mathbb{Z}[G_X]$-module. The homology group $H_n^Q(X; A[Y])$ is usually denoted by $H_n^Q(X; A)_Y$ [Kamada 2007].

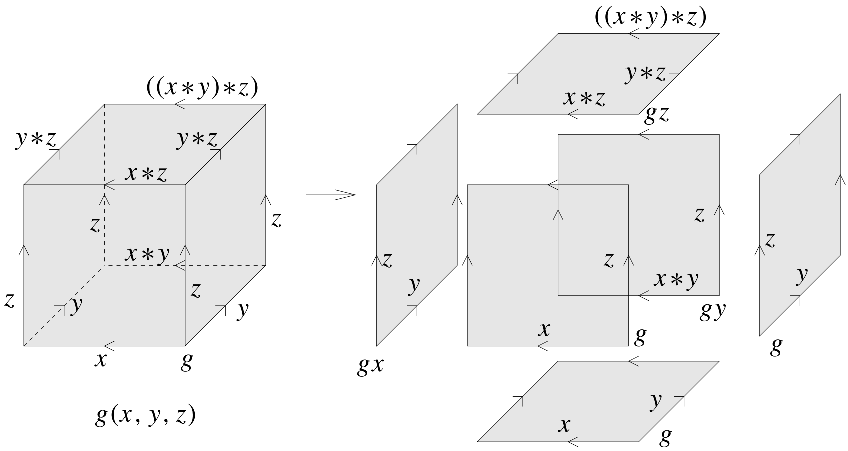
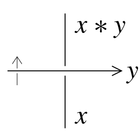


Figure 9. The boundary map $\partial(g(x,y,z)) = -(g(y,z) - gx(y,z)) + (g(x,z) - gy(x*y,z)) - (g(x,y) - gz(x*z, y*z))$. Here $x, y, z \in X$ and $g \in G_X$. Edges are labeled by elements of X and vertices are labeled by elements of G_X.

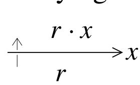
Let N be a left $\mathbb{Z}[G_X]$-module. We define the *rack cohomology* $H^n_R(X; N)$ by the cohomology of $C^n_R(X; N) = \mathrm{Hom}_{\mathbb{Z}[G_X]}(C^R_n(X), N)$. The *quandle cohomology* $H^n_Q(X; N)$ is defined in a similar way. For a set Y with a right G_X-action and an abelian group A, we let $\mathrm{Func}(Y, A)$ be the left $\mathbb{Z}[G_X]$-module generated by functions $\phi : Y \to A$, where the action is defined by $(g\phi)(y) = \phi(yg)$ for $y \in Y$ and $g \in G_X$. The cohomology group $H^n_Q(X; \mathrm{Func}(Y, A))$ is usually denoted by $H^n_Q(X; A)_Y$ [Kamada 2007].

5B. ***Shadow coloring and associated quandle cycle.*** Let X be a quandle. Let L be an oriented link in S^3 and let D be a diagram of L. An *arc coloring* of D is an assignment of elements of X to arcs of D satisfying the following relation at each crossing:



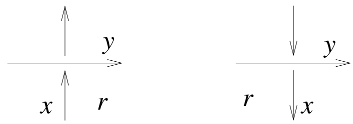
where $x, y \in X$. By the Wirtinger presentation of the knot complement, an arc coloring determines a representation of $\pi_1(S^3 - L)$ into the associated group G_X. This is obtained by sending each meridian to its color.

Let Y be a set with a right G_X action. A *region coloring* of D is an assignment of elements of Y to regions of D satisfying the relation



for any pair of adjacent regions, where $r \in Y$ and $x \in X$. A pair $\mathcal{S} = (\mathcal{A}, \mathcal{R})$ is called a *shadow coloring*. If we fix a color of a region of D, then colors of other regions are uniquely determined. Therefore there always exists a region coloring compatible with a given arc coloring.

Define a cycle $C(\mathcal{S})$ of $C_2^Q(X; \mathbb{Z}[Y])$ for a shadow coloring $\mathcal{S}$. Put $+r \otimes (x, y)$ for a positive crossing and $-r \otimes (x, y)$ for a negative crossing colored by



respectively. Then define

$$C(\mathcal{S}) = \sum_{c:\,\text{crossing}} \epsilon_c r_c \otimes (x_c, y_c) \in C_2^Q(X; \mathbb{Z}[Y]),$$

where $\epsilon_c = \pm 1$. We can easily check the following:

Proposition 5.1 [Inoue and Kabaya 2010]. *$C(\mathcal{S})$ is a cycle and the homology class $[C(\mathcal{S})]$ is invariant under Reidemeister moves. Moreover it does not depend on the choice of the region coloring if the action of G_X on Y is transitive.*

So the homology class $[C(\mathcal{S})]$ is an invariant of the arc coloring $\mathcal{A}$ in many cases. There are two important sets with right G_X-action: one is when Y consists of one point $\{*\}$ and the other is when $Y = X$. Eisermann [2003; 2007] showed that the cycle $[C(\mathcal{S})]$ for $Y = \{*\}$ is essentially described by the monodromy of some representation of the knot group along the longitude. So we concentrate on the invariant $[C(\mathcal{S})]$ in the case of $Y = X$ from now on.

5C. ***Quandle cocycle invariant.*** Let X be a quandle with $|X| < \infty$. Let A be an abelian group and f be a cocycle of $H_Q^2(X; \mathrm{Func}(X, A))$. We define the *(shadow) quandle cocycle invariant* by

$$\frac{1}{|X|} \sum_{\mathcal{S}:\,\text{shadow colorings}} \langle f, C(\mathcal{S})\rangle \in \mathbb{Z}[A].$$

Here the sum is finite because there are only a finite number of shadow colorings of D. This is an invariant of oriented knots by Proposition 5.1. If G_X acts on X transitively (that is, X is connected), $\langle f, C(\mathcal{S})\rangle$ does not depend on the choice of a region coloring $\mathcal{R}$ by Proposition 5.1, and thus

$$\sum_{\substack{\mathcal{S}=(\mathcal{A},\mathcal{R}),\\ \mathcal{A}:\,\text{arc coloring}}} \langle f, C(\mathcal{S})\rangle$$

coincides with the quandle cocycle invariant, where $\mathcal{R}$ is a region coloring compatible with $\mathcal{A}$.

We can regard $f \in C_Q^n(X; A)$ as an element of $C_Q^{n-1}(X; \mathrm{Func}(X, A))$ by

$$f(x_1, x_2, \ldots, x_{n-1})(r) = f(r, x_1, x_2, \ldots, x_{n-1}).$$

This gives a homomorphism $H_Q^n(X; A) \to H_Q^{n-1}(X; \mathrm{Func}(X, A))$. Therefore a quandle 3-cocycle $f \in H_Q^3(X; A)$ gives rise to a quandle cocycle invariant. Explicitly, the cocycle invariant has the form

$$\sum_{\substack{\mathcal{S}=(\mathcal{A},\mathcal{R}),\\ \mathcal{A}\text{: arc coloring}}} \sum_{c\text{: crossing}} \epsilon_c f(r_c, x_c, y_c) \in \mathbb{Z}[A],$$

where $x_c, y_c \in X$ are given by $\mathcal{A}$ and $r_c \in X$ are given by $\mathcal{R}$.

6. $H_n^\Delta(X; \mathbb{Z})$ and the map $H_n^R(X; \mathbb{Z}[X]) \to H_{n+1}^\Delta(X; \mathbb{Z})$

Let X be a quandle. Let $C_n^\Delta(X) = \mathrm{span}_\mathbb{Z}\{(x_0, \ldots, x_n) \mid x_i \in X\}$. We define the boundary operator of $C_n^\Delta(X)$ by

$$\partial(x_0, \ldots, x_n) = \sum_{i=0}^{n} (-1)^i (x_0, \ldots, \widehat{x_i}, \ldots, x_n).$$

Since X has a right action of G_X, the chain complex $C_n^\Delta(X)$ has a right action of G_X by $(x_0, \ldots, x_n) * g = (x_0 * g, \ldots, x_n * g)$. Let M be a left $\mathbb{Z}[G_X]$-module. We denote the homology of $C_n^\Delta(X) \otimes_{\mathbb{Z}[G_X]} M$ by $H_n^\Delta(X; M)$. For any abelian group A, we can also define the cohomology group $H_\Delta^n(X; A)$ in a similar way.

Let I_n be the set consisting of maps $\iota : \{1, 2, \ldots, n\} \to \{0, 1\}$. We let $|\iota|$ denote the cardinality of the set $\{i \mid \iota(i) = 1,\ 1 \le i \le n\}$. For each generator $r \otimes (x_1, x_2, \ldots, x_n)$ of $C_n^R(X; \mathbb{Z}[X])$, where $r, x_1, \ldots, x_n \in X$, we define

$$r(\iota) = r * \left(x_1^{\iota(1)} x_2^{\iota(2)} \cdots x_n^{\iota(n)}\right) \in X,$$
$$x(\iota, i) = x_i * \left(x_{i+1}^{\iota(i+1)} x_{i+2}^{\iota(i+2)} \cdots x_n^{\iota(n)}\right) \in X,$$

for any $\iota \in I_n$. Fix an element $q \in X$. For each $n \ge 1$, we define a homomorphism

$$\varphi : C_n^R(X; \mathbb{Z}[X]) \to C_{n+1}^\Delta(X) \otimes_{\mathbb{Z}[G_X]} \mathbb{Z}$$

by

$$\varphi\big(r \otimes (x_1, x_2, \ldots, x_n)\big) = \sum_{\iota \in I_n} (-1)^{|\iota|} \big(q, r(\iota), x(\iota, 1), x(\iota, 2), \ldots, x(\iota, n)\big). \tag{6-1}$$

For example, in the case $n = 2$ (Figure 10),

$$\begin{aligned}\varphi(r \otimes (x, y)) \\ &= (q, r, x, y) - (q, r * x, x, y) - (q, r * y, x * y, y) + (q, (r * x) * y, x * y, y),\end{aligned}$$

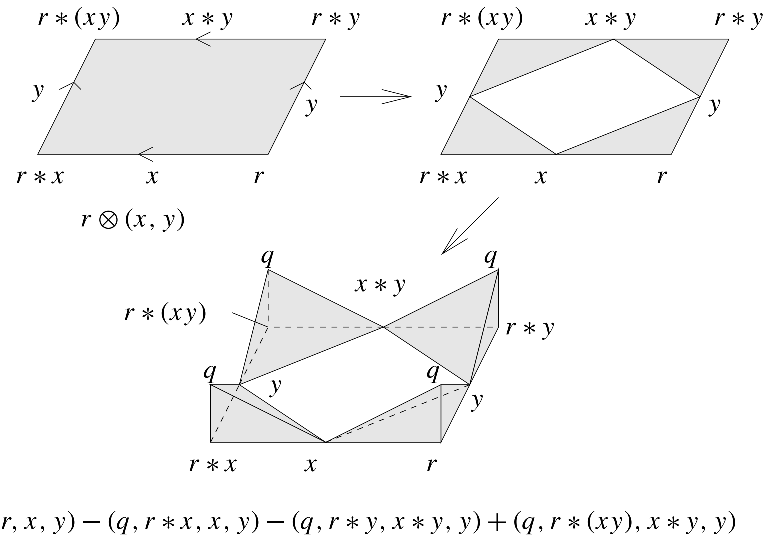


Figure 10

and in the case $n = 3$,

$$\begin{aligned}\varphi(r \otimes (x, y, z)) = (q, r, x, y, z) &- (q, r * x, x, y, z)\\ &- (q, r * y, x * y, y, z) - (q, r * z, x * z, y * z, z)\\ &+ (q, (r * x) * y, x * y, y, z) + (q, (r * x) * z, x * z, y * z, z)\\ &+ (q, (r * y) * z, (x * y) * z, y * z, z)\\ &- \big(q, ((r * x) * y) * z, (x * y) * z, y * z, z\big).\end{aligned}$$

Theorem 6.1 [Inoue and Kabaya 2010]. *The map*

$$\varphi : C_n^R(X; \mathbb{Z}[X]) \to C_{n+1}^{\Delta}(X) \otimes_{\mathbb{Z}[G_X]} \mathbb{Z}$$

is a chain map.

Therefore φ induces a homomorphism $\varphi_* : H_n^R(X; \mathbb{Z}[X]) \to H_{n+1}^{\Delta}(X; \mathbb{Z})$. We remark that the induced map $\varphi_* : H_n^R(X; \mathbb{Z}[X]) \to H_{n+1}^{\Delta}(X; \mathbb{Z})$ does not depend on the choice of $q \in X$ [Inoue and Kabaya 2010].

In general, it is easier to construct cocycles of $H_{n+1}^{\Delta}(X)$ from group cocycles of some group related to X than those of $H_n^R(X; \mathbb{Z}[X])$. If we have a function f from X^{k+1} to some abelian group A satisfying

(1) $\sum_{i=0}^{k+1} (-1)^i f(x_0, \ldots, \widehat{x_i}, \ldots, x_{k+1}) = 0$,

(2) $f(x_0 * y, \ldots, x_k * y) = f(x_0, \ldots, x_k)$ for any $y \in X$, and

(3) $f(x_0, \ldots, x_k) = 0$ if $x_i = x_{i+1}$ for some i,

then f is a cocycle of $H^k_\Delta(X; A)$ and $\varphi^* f$ is a cocycle of $H^{k-1}_Q(X; \mathrm{Func}(X, A))$. Moreover, $\varphi^* f$ can be regarded as a cocycle in $H^k_Q(X; A)$ by

$$(\varphi^* f)(r, x_1, \dots, x_{k-1}) = (\varphi^* f)(x_1, \dots, x_{k-1})(r).$$

We will construct functions satisfying these three conditions from group cocycles.

7. Cocycles of dihedral quandles

For any integer $p > 2$, let R_p denote the cyclic group $\mathbb{Z}/p$ with quandle operation defined by $x * y = 2y - x \mod p$. Actually this operation satisfies the quandle axioms. The quandle R_p is called the *dihedral quandle*. In this section, we construct quandle cocycles of R_p from group cocycles of $G = \mathbb{Z}/p$. In the next section, we will propose a general construction of quandle cocycles from group cocycles.

7A. ***Group cohomology of cyclic groups.*** Let G be the cyclic group $\mathbb{Z}/p$ (where p is an integer greater than 2). The first cohomology $H^1(G; \mathbb{Z}/p) = \mathrm{Hom}(\mathbb{Z}/p, \mathbb{Z}/p)$ is generated by the 1-cocycle b_1 defined by

$$b_1(x) = x.$$

The connecting homomorphism $\delta : H^1(G, \mathbb{Z}/p) \to H^2(G; \mathbb{Z})$ of the long exact sequence corresponding to $0 \to \mathbb{Z} \to \mathbb{Z} \to \mathbb{Z}/p \to 0$ maps b_1 to a generator of $H^2(G; \mathbb{Z})$, and the reduction $H^2(G; \mathbb{Z}) \to H^2(G; \mathbb{Z}/p)$ maps it to a generator b_2 of $H^2(G; \mathbb{Z}/p)$. Explicitly, we have

$$b_2(x, y) = \frac{1}{p}(\overline{y} - \overline{x+y} + \overline{x}) = \begin{cases} 1 & \text{if } \overline{x} + \overline{y} \geq p, \\ 0 & \text{otherwise,} \end{cases} \tag{7-1}$$

where $\overline{x}$ is an integer $0 \leq \overline{x} < p$ such that $\overline{x} \equiv x \mod p$. Cup products of b_1's and b_2's are also cocycles. Moreover, when p is an odd prime, it is known that any element of $H^*(G; \mathbb{Z}/p)$ can be presented by a cup product of b_1's and b_2's; see, for example, [Benson 1991, Proposition 3.5.5]. We remark that b_1 and b_2 and their products are normalized cocycles.

7B. ***Cocycle of R_p.*** For an integer $p > 2$, let f be a normalized k-cocycle of $H^k(G, \mathbb{Z}/p)$. Regarding R_p as $G = \mathbb{Z}/p$, we obtain a map $f : (R_p)^{k+1} \to \mathbb{Z}/p$ satisfying

(1) $\sum_{i=0}^{k+1} (-1)^i f(x_0, \dots, \widehat{x_i}, \dots, x_{k+1}) = 0$,

(3) $f(x_0, \dots, x_k) = 0$ if $x_i = x_{i+1}$ for some i,

by using homogeneous notation (Section 2A). If f also satisfies the condition

(2) $f(x_0 * y, \dots, x_k * y) = f(x_0, \dots, x_k)$ for any $y \in R_p$,

then f gives rise to a quandle k-cocycle of $H^k_Q(R_p; \mathbb{Z}/p)$ by the construction of Section 6. Define $\tilde{f} : (R_p)^{k+1} \to \mathbb{Z}/p$ by

$$\tilde{f}(x_0, \ldots, x_k) = f(x_0, \ldots, x_k) + f(-x_0, \ldots, -x_k). \tag{7-2}$$

Then $\tilde{f}$ satisfies condition (2) by the left invariance of the homogeneous group cocycle. It is easy to check that $\tilde{f}$ also satisfies conditions (1) and (3). So we obtain a quandle k-cocycle.

We give an explicit presentation of the 3-cocycle arising from $b_1 b_2 \in H^3(G; \mathbb{Z}/p)$. Let

$$d(x, y) = b_2(x, y) - b_2(-x, -y);$$

then d is a 2-cocycle. (We can check that d is cohomologous to $2b_2$.) Then by the defining equation (7-2), $\widetilde{b_1 b_2}$ is given by

$$[x|y|z] \mapsto x \cdot d(y, z).$$

By definition we have

$$d(-x, -y) = -d(x, y) \tag{7-3}$$

and

$$d(x, y) = \begin{cases} 1 & \text{if } \overline{x} + \overline{y} > p, \\ -1 & \text{if } \overline{x} + \overline{y} < p, \, x \neq 0 \text{ and } y \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{7-4}$$

We remark that the cocycle d can be easily understood geometrically. Identify $i \in \mathbb{Z}/p$ with the complex number ζ^i, where $\zeta = \exp(2\pi\sqrt{-1}/p)$. Then $d(x, y) = -1$ if $(0, x, x+y)$ is counterclockwise, $d(x, y) = +1$ if $(0, x, x+y)$ is clockwise, and $d(x, y) = 0$ if $(0, x, x+y)$ is degenerate (Figure 11). This interpretation and Equation (7-3) make various calculations easy.

Proposition 7.1. *The quandle* 3*-cocycle arising from* $b_1 b_2 \in H^3(G; \mathbb{Z}/p)$ *has the following presentation*:

$$(x, y, z) \mapsto 2z\big(d(y-x, z-y) + d(y-x, y-z)\big), \quad (x, y, z \in R_p). \tag{7-5}$$

This is a nontrivial quandle 3*-cocycle of* R_p *with* $\mathbb{Z}/p$ *coefficients.*

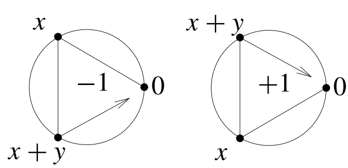


Figure 11. The value of $d(x, y)$.

Proof. In (6-1), since the map φ_* does not depend on the choice of $q \in R_p$, we let $q = 0$. Then we have

$$\begin{aligned}\varphi(x, y, z) &= (0, x, y, z) - (0, x * y, y, z) \\ &\qquad - (0, x * z, y * z, z) + (0, (x * y) * z, y * z, z) \\ &= (0, x, y, z) - (0, 2y - x, y, z) \\ &\qquad - (0, 2z - x, 2z - y, z) + (0, 2z - 2y + x, 2z - y, z),\end{aligned}$$

for $x, y, z \in R_p$. Rewriting in inhomogeneous notation, this is equal to

$$\begin{aligned}&[x \mid y - x \mid z - y] - [2y - x \mid x - y \mid z - y] \\ &\qquad - [2z - x \mid x - y \mid y - z] + [2z - 2y + x \mid y - x \mid y - z].\end{aligned}$$

The evaluation of $\widetilde{b_1 b_2}$ on this chain is

$$\begin{aligned}&x \cdot d(y - x, z - y) - (2y - x)d(x - y, z - y) \\ &\qquad - (2z - x)d(x - y, y - z) + (2z - 2y + x)d(y - x, y - z) \\ &= x \cdot d(y - x, z - y) + (2y - x)d(y - x, y - z) \\ &\qquad + (2z - x)d(y - x, z - y) + (2z - 2y + x)d(y - x, y - z) \\ &= 2z \cdot d(y - x, z - y) + 2z \cdot d(y - x, y - z).\end{aligned}$$

We will see that this cocycle is nontrivial because the evaluation on the cycle given by a shadow coloring $\mathcal{S}$ of the $(2, p)$-torus link (Figure 12) is nonzero. Color two arcs by $x, y \in R_p$ as in Figure 12. Then other arcs must be colored by $(i+1)y - ix$ by the relations at the crossings. Let r be the color of the central region. Then we have

$$C(\mathcal{S}) = \sum_{i=0}^{p-1} r \otimes \bigl(iy - (i-1)x, (i+1)y - ix\bigr).$$

We assume that $r = 0$ and $x = 0$. Evaluation of the cocycle on $C(\mathcal{S})$ yields

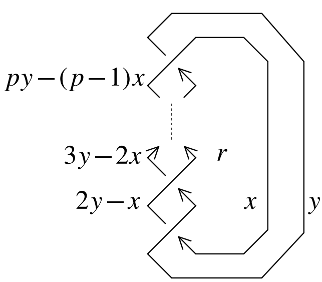


Figure 12. A shadow coloring of the $(2, p)$-torus knot by R_p (for any $x, y, r \in R_p$).

$$\begin{aligned}\sum_{i=0}^{p-1} 2(i+1)y\big(d(iy,y)+d(iy,-y)\big)\\ &= 2y\sum_{i=0}^{p-1}(i+1)\big(d(iy,y)-d(-iy,y)\big) \quad \text{(by (7-3))}\\ &= 2y\sum_{i=0}^{p-1}\big((i+1)d(iy,y)+(i-1)d(iy,y)\big) = 4y\sum_{i=0}^{p-1} i\cdot d(iy,y).\end{aligned}$$

By the next lemma, this is equal to $-4y^2 \mod p$. □

Lemma 7.2. *Let $p > 2$ be an integer. For $0 < y < p$, we have*

$$\sum_{i=0}^{p-1} i\cdot d(iy,y) = \begin{cases} -y & p\text{: odd},\\ \frac{p}{2}-y & p\text{: even}.\end{cases}$$

When $y = 0$, the left-hand side is 0.

Proof. When $y = 0$, this is straightforward since $d(x,0) = 0$ for any $0 \le x < p$. When y is a unit in $\mathbb{Z}/p$, by (7-4) we have

$$\begin{aligned}\sum_{i=0}^{p-1} i\cdot d(iy,y) &= \sum_{i=0}^{p-1}\frac{i}{y}\cdot d(i,y)\\ &= \frac{1}{y}\big(-1-2-\cdots-(p-y-1)+(p-y+1)+(p-y+2)+\cdots+(p-1)\big)\\ &= \frac{1}{y}\Big(\frac{(p-y-1)(y-p)}{2}+\frac{(y-1)(2p-y)}{2}\Big)\\ &= \frac{1}{y}\Big(-y^2-\frac{p}{2}+2py-\frac{p^2}{2}\Big) \equiv -y-\frac{p}{2}\cdot\frac{1}{y}.\end{aligned}$$

This is equal to $-y$ when p is odd. When p is even, $-p/2\cdot 1/y \equiv p/2 \mod p$ since $1/y$ is a unit in $\mathbb{Z}/p$.

When y is not a unit in $\mathbb{Z}/p$, let c be the greatest common divisor of y and p. Then

$$\sum_{i=0}^{p-1} i\cdot d(iy,y) = c\sum_{j=0}^{p/c-1}\frac{j}{y/c}\cdot d(jc,y) = \frac{c}{y/c}\sum_{j=0}^{p/c-1} j\cdot d(j,y/c).$$

Since y/c is a unit in $\mathbb{Z}/(p/c)$, this reduces to the previous case. □

Since 2 is divisible in $\mathbb{Z}/p$ when p is odd, we have:

Corollary 7.3. *When $p > 2$ is an odd number,*

$$(x,y,z) \mapsto z\big(d(y-x,z-y)+d(y-x,y-z)\big) \tag{7-6}$$

is a nontrivial quandle 3-cocycle of R_p with $\mathbb{Z}/p$ coefficients.

When p is prime, it is known that $\dim_{\mathbb{F}_p} H^3_Q(R_p; \mathbb{F}_p) = 1$. Therefore our cocycle is a constant multiple of the Mochizuki 3-cocycle [2003]. We remark that when p is prime, $\dim_{\mathbb{F}_p} H^n_Q(R_p; \mathbb{F}_p)$ was calculated for any n by Nosaka [2009], who gave a system of generators of $H^n_Q(R_p; \mathbb{F}_p)$.

When p is an odd integer, Nosaka [2010] showed that $H^Q_3(R_p; \mathbb{Z}) \cong \mathbb{Z}/p$. Since $H^Q_2(R_p; \mathbb{Z})$ is zero, we have $H^3_Q(R_p; \mathbb{Z}/p) \cong \mathbb{Z}/p$. This means that there exists a nontrivial quandle 3-cocycle of R_p with $\mathbb{Z}/p$ coefficients.

8. General construction

In this section, we generalize the construction in the previous section to wider classes of quandles. We construct a quandle cocycle of a faithful homogeneous quandle X from a group cocycle of $\mathrm{Aut}(X)$ when an obstruction living in the second cohomology of $\mathrm{Aut}(X)$ vanishes.

8A. Let G be a group. Fix an element $h \in G$. Let $\mathrm{Conj}(h) = \{g^{-1}hg \mid g \in G\}$. Now $\mathrm{Conj}(h)$ has a quandle operation by $x * y = y^{-1}xy$. In this section, we construct a quandle cocycle of $\mathrm{Conj}(h)$ from a group cocycle of G. First we shall show that this class of quandles is not so special.

Let X be a quandle. We denote the group of the quandle automorphisms of X by $\mathrm{Aut}(X)$. We consider an automorphism that acts on X from the right. For $x \in X$, let $S(x)$ be the map that sends y to $y * x$. By the axioms (Q2) and (Q3), $S(x)$ is a quandle automorphism. X is called *faithful* if $S : X \to \mathrm{Aut}(X)$ is injective. A quandle X is *homogeneous* if $\mathrm{Aut}(X)$ acts on X transitively. The following lemma was essentially shown in Theorem 7.1 of [Joyce 1982], but we include a proof for completeness.

Lemma 8.1. *Every faithful homogeneous quandle X is represented by* $\mathrm{Conj}(h)$ *with some group G and $h \in G$.*

Proof. For $x \in X$ and $g \in \mathrm{Aut}(X)$, we have $S(xg) = g^{-1}S(x)g$. In fact, $(y)S(xg) = y * (xg) = (yg^{-1} * x)g = (y)g^{-1}S(x)g$ for any $y \in X$.

Let $G = \mathrm{Aut}(X)$ and fix an element $x_0 \in X$. Put $h = S(x_0)$. Because X is homogeneous, for any $x \in X$, there exists $g \in G$ such that $x = x_0 g$. So we have $S(x) = S(x_0 g) = g^{-1}hg$, that is, $S(x) \in \mathrm{Conj}(h)$. Therefore we obtain a homomorphism $S : X \to \mathrm{Conj}(h)$. This is surjective since $g^{-1}S(x_0)g = S(x_0 g)$, and injective since X is faithful. □

Let $Z(h) = \{g \in G \mid gh = hg\}$ be the centralizer of h in G.

Lemma 8.2. *A map*

$$\begin{array}{ccc} \mathrm{Conj}(h) & \to & Z(h)\backslash G \\ \cup\!\!\!\!\!\!\;\shortmid & & \cup\!\!\!\!\!\!\;\shortmid \\ g^{-1}hg & \mapsto & Z(h)g \end{array}$$

is well-defined and bijective.

Proof. Let $g_1^{-1}hg_1 = g_2^{-1}hg_2$. Then $(g_1g_2^{-1})^{-1}h(g_1g_2^{-1}) = h$, so $g_1g_2^{-1} \in Z(h)$ and $g_1 \in Z(h)g_2$. This means that g_1 and g_2 belong to the same right coset. Therefore the map is well-defined. By a similar argument, we can show the injectivity. Surjectivity is trivial by definition. □

Now we study the quandle structure on $Z(h)\backslash G$ and construct a section of the projection $\pi : G \to Z(h)\backslash G$. The quandle operation on $\mathrm{Conj}(h)$ induces a quandle operation on $Z(h)\backslash G$:

$$\begin{aligned} Z(h)g_1 * Z(h)g_2 \leftrightarrow (g_1^{-1}hg_1) * (g_2^{-1}hg_2) &= (g_2^{-1}hg_2)^{-1}(g_1^{-1}hg_1)(g_2^{-1}hg_2) \\ &= (g_1g_2^{-1}hg_2)^{-1}h(g_1g_2^{-1}hg_2) \leftrightarrow Z(h)g_1(g_2^{-1}hg_2). \end{aligned}$$

This quandle operation on $Z(h)\backslash G$ lifts to a quandle operation $\tilde{*}$ on G by

$$g_1 \tilde{*} g_2 = h^{-1}g_1(g_2^{-1}hg_2), \quad (g_1, g_2 \in G). \tag{8-1}$$

We can easily check that $\tilde{*}$ satisfies the quandle axioms (the inverse operation is given by $g_1 \tilde{*}^{-1} g_2 = hg_1g_2^{-1}h^{-1}g_2$) and that the projection map $\pi : G \to Z(h)\backslash G$ is a quandle homomorphism. We remark that the quandle operation given by (8-1) has been already studied by Joyce [1982] and Eisermann [2003].

Let $s : Z(h)\backslash G \to G$ be a section of π, that is, a map (not a homomorphism) satisfying $\pi \circ s = \mathrm{id}$. Since $s(x * y)$ and $s(x) \tilde{*} s(y)$ are in the same coset in $Z(h)\backslash G$, there exists an element $c(x, y) \in Z(h)$ satisfying

$$s(x) \tilde{*} s(y) = c(x, y)s(x * y).$$

Lemma 8.3. *If $Z(h)$ is an abelian group, $c : Z(h)\backslash G \times Z(h)\backslash G \to Z(h)$ is a quandle 2-cocycle. If the cocycle c is cohomologous to zero, we can change the section s to satisfy $s(x * y) = s(x) \tilde{*} s(y)$.*

Proof. For $c_1, c_2 \in Z(h)$ and $g_1, g_2 \in G$, we have

$$(c_1g_1) \tilde{*} (c_2g_2) = h^{-1}c_1g_1g_2^{-1}c_2^{-1}hc_2g_2 = c_1h^{-1}g_1g_2^{-1}hg_2 = c_1(g_1 \tilde{*} g_2).$$

Therefore,

$$\begin{aligned} \big(s(x) \tilde{*} s(y)\big) \tilde{*} s(z) &= \big(c(x, y)s(x * y)\big) \tilde{*} s(z) \\ &= c(x, y)\big(s(x * y) \tilde{*} s(z)\big) \\ &= c(x, y)c(x * y, z)s\big((x * y) * z\big) \end{aligned} \tag{8-2}$$

and

$$\begin{aligned} \big(s(x) \tilde{*} s(z)\big) \tilde{*} \big(s(y) \tilde{*} s(z)\big) &= \big(c(x, z)s(x * z)\big) \tilde{*} \big(c(y, z)s(y * z)\big) \\ &= c(x, z)c(x * z, y * z)s\big((x * z) * (y * z)\big) \end{aligned} \tag{8-3}$$

for any $x, y, z \in Z(h)\backslash G$. Comparing (8-2) and (8-3), we have

$$c(x,z)c(x*y,z)^{-1}c(x,y)^{-1}c(x*z,y*z)=1.$$

By $s(x)\,\widetilde{*}\,s(x)=s(x)$, we also have $c(x,x)=1$.

If c is cohomologous to zero; then there exists a map $b : Z(h)\backslash G \to Z(h)$ satisfying $c(x,y) = b(x)b(x*y)^{-1}$. Put $s'(x) = b(x)^{-1}s(x)$, then s' satisfies $s'(x)\,\widetilde{*}\,s'(y) = s'(x*y)$. □

Remark 8.4. The 2-cocycle c has already appeared in [Eisermann 2003] in a similar context.

Example 8.5. Let G be the dihedral group

$$D_{2p} = \langle h, x \mid h^2 = x^p = hxhx = 1\rangle,$$

where p is an odd number greater than 2. Then we have $Z(h) = \{1, h\}$ and $\mathrm{Conj}(h) = \{x^{-i}hx^i \mid i = 0, 1, \ldots, p-1\} = \{hx^{2i} \mid i = 0, \ldots, p-1\}$. We can identify $x^{-i}hx^i \in \mathrm{Conj}(h)$ with $i \in R_p = \{0, 1, 2, \ldots, p-1\}$. Define a section $s : Z(h)\backslash G \to G$ by

$$\begin{array}{ccccc} \mathrm{Conj}(h) & \cong & Z(h)\backslash G & \xrightarrow{s} & G \\ \Cup & & \Cup & & \Cup \\ x^{-i}hx^i & \leftrightarrow & Z(h)x^i & \mapsto & hx^i \end{array}$$

Then we have

$$\begin{aligned} s\big(Z(h)x^i * Z(h)x^j\big) &= s(Z(h)x^{2j-i}) = hx^{2j-i} \\ &= h^{-1}(hx^i)(x^{-j}hx^j) = s(Z(h)x^i)\,\widetilde{*}\,s(Z(h)x^j). \end{aligned}$$

Therefore $c(x,y) = 0$ for any $x, y \in R_p$.

8B. Let G be a group. Fix $h \in G$ with $h^l = 1$ $(l > 1)$. In the following we assume:

Assumption 8.6. $Z(h)$ is abelian and the 2-cocycle corresponding to $G \to Z(h)\backslash G$ is cohomologous to zero.

Under this assumption, we can take a section $s : Z(h)\backslash G \to G$ satisfying

$$s(x*y) = s(x)\,\widetilde{*}\,s(y),$$

by Lemma 8.3. Let $f : G^{k+1} \to A$ be a normalized group k-cocycle of G in the homogeneous notation, where A is an abelian group. Then f satisfies

(1) $\displaystyle\sum_{i=0}^{k+1}(-1)^i f(x_0, \ldots, \widehat{x_i}, \ldots, x_{k+1}) = 0,$

(2) $f(gx_0, \ldots, gx_k) = f(x_0, \ldots, x_k)$ for any $g \in G$ (left invariance), and

(3) $f(x_0, \ldots, x_k) = 0$ if $x_i = x_{i+1}$ for some i.

Since it is convenient to use a right-invariant function in the following construction, we modify condition (2) by replacing $f(x_0, \dots, x_k)$ with $f(x_0^{-1}, \dots, x_k^{-1})$:

(2′) $f(x_0 g, \dots, x_k g) = f(x_0, \dots, x_k)$ for any $g \in G$ (right invariance).

Define $\tilde{f} : \mathrm{Conj}(h)^{k+1} \to A$ by

$$\tilde{f}(x_0, \dots, x_k) = \sum_{i=0}^{l-1} f\bigl(h^i s(x_0), \dots, h^i s(x_k)\bigr),$$

for $x_0, \dots, x_k \in \mathrm{Conj}(h)$.

Proposition 8.7. *The function $\tilde{f}$ satisfies conditions (1), (2) and (3) of Section 6. Therefore $\tilde{f}$ gives rise to a k-cocycle of $H^k_\Delta(\mathrm{Conj}(h); A)$.*

Proof. It is clear that (1) and (3) are satisfied from the conditions on a normalized group cocycle in homogeneous notation. We only have to check the second property.

$$\begin{aligned}
&\tilde{f}(x_0 * y, \dots, x_k * y) \\
&\quad = \sum_{i=0}^{l-1} f\bigl(h^i s(x_0 * y), \dots, h^i s(x_k * y)\bigr) \\
&\quad = \sum_{i=0}^{l-1} f\bigl(h^i s(x_0) \,\tilde{*}\, s(y), \dots, h^i s(x_k) \,\tilde{*}\, s(y)\bigr) \\
&\quad = \sum_{i=0}^{l-1} f\bigl(h^{i-1} s(x_0)(s(y)^{-1} h s(y)), \dots, h^{i-1} s(x_k)(s(y)^{-1} h s(y))\bigr) \\
&\quad = \sum_{i=0}^{l-1} f\bigl(h^{i-1} s(x_0), \dots, h^{i-1} s(x_k)\bigr) \quad \text{(right invariance)} \\
&\quad = \tilde{f}(x_0, \dots, x_k).
\end{aligned}$$

□

Combining with the arguments of Section 6, we have:

Corollary 8.8. *If $Z(h)$ is abelian and the 2-cocycle corresponding to $G \to Z(h)\backslash G$ is cohomologous to zero, then there is a homomorphism*

$$H^n(G; A) \to H^n_Q(\mathrm{Conj}(h); A)$$

for any abelian group A.

8C. We return to the case of R_p discussed in the previous section. We assume that p is an odd integer greater than 2. Let G be the dihedral group

$$D_{2p} = \langle h, x \mid h^2 = x^p = hxhx = 1 \rangle.$$

Consider the short exact sequence

(8-4) $$0 \to \mathbb{Z}/p \to D_{2p} \to \mathbb{Z}/2 \to 0.$$

We regard $\mathbb{Z}/2$ as $\{1, h\}$ by taking coset representatives in D_{2p}. Then $\mathbb{Z}/2$ acts on $\mathbb{Z}/p$ by $h(x^i) = x^{-i}$. This induces the restriction map

$$H^*(D_{2p}; \mathbb{Z}/p) \to H^*(\mathbb{Z}/p; \mathbb{Z}/p)^{\mathbb{Z}/2}.$$

We can show that this homomorphism is an isomorphism [Brown 1982, Proposition III.10.4]. To obtain a group cocycle of D_{2p} from a group cocycle of $\mathbb{Z}/p$, we need the inverse map, which is called the *transfer*. The transfer map is described as follows (see also [Brown 1982]). Let r be the map $D_{2p} \to \mathbb{Z}/p$ defined by $r(x^i) = x^i$ and $r(hx^i) = x^{-i}$. For a cocycle f of $H^n(\mathbb{Z}/p; \mathbb{Z}/p)^{\mathbb{Z}/2}$, the image f' of the transfer is given by

$$f'(x_0, \dots, x_n) = f(rx_0, \dots, rx_n) + f(hrx_0, \dots, hrx_n)$$

in homogeneous notation. When restricted to the image of $s : Z(h)\backslash G \to G$, this is equal to the map defined in (7-2). Applying our construction for this group cocycle, we obtain a quandle 3-cocycle of R_p, which is twice the cocycle constructed in the previous section.

8D. We end this section by giving another homomorphism from $H^n(G; A)$ to $H^n_Q(\mathrm{Conj}(h); A)$ arising from a more general context.

Let X be a quandle and let M be a right $\mathbb{Z}[G_X]$-module. We can construct a map from the rack homology $H^R_n(X; M)$ to the group homology $H_n(G_X; M)$. The following lemma is well-known; for example, see Lemma 7.4 of Chapter I of [Brown 1982].

Lemma 8.9. *Let $\cdots \to P_1 \to P_0 \to M \to 0$ be a chain complex where the P_i are projective (for example, free). Let $\cdots \to C_1 \to C_0 \to N \to 0$ be an acyclic complex. Any homomorphism $M \to N$ can be extended to a chain map from $\{P_*\}$ to $\{C_*\}$. Moreover, such a chain map is unique up to chain homotopy.*

So there exists a unique chain map from $C^R_*(X)$ to $C_*(G_X)$ up to homotopy. This map induces $M \otimes_{\mathbb{Z}[G_X]} C^R_*(X) \to M \otimes_{\mathbb{Z}[G_X]} C_*(G_X)$ and then $H^R_n(X; M) \to H_n(G_X; M)$. Using normalized chains in group homology, we can also construct a map $H^Q_n(X; M) \to H_n(G_X; M)$. We give an explicit chain map. Let $(x_1, \dots, x_n)$ be a generator of $C^R_n(X)$. We define a map ψ by

$$\psi((x_1, \dots, x_n)) = \sum_{\sigma \in \mathfrak{S}_n} \mathrm{sgn}(\sigma)\big[y_{\sigma,1}|\cdots|y_{\sigma,i}|\cdots|y_{\sigma,n}\big],$$

where $y_{\sigma,i} \in X$ is defined for a permutation σ and $i \in \{1, \dots, n\}$ as follows. Let

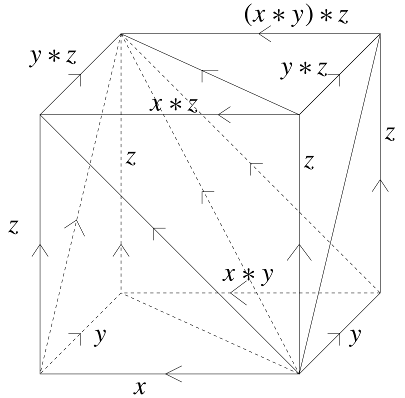


Figure 13

$j_1, \dots, j_i < i$ be the maximal set of numbers satisfying

$$\sigma(i) < \sigma(j_1) < \sigma(j_2) < \cdots < \sigma(j_i).$$

Then define

$$y_{\sigma,i} = x_{\sigma(i)} * (x_{\sigma(j_1)} x_{\sigma(j_2)} \cdots x_{\sigma(j_i)}).$$

A graphical picture of this map is given in Figure 13. For example, when $n = 3$,

$$\psi((x, y, z)) = [x|y|z] - [x|z|y*z] \\ + [y|z|(x*y)*z] - [y|x*y|z] + [z|x*z|y*z] - [z|y*z|(x*y)*z],$$

for $(x, y, z) \in C_3^R(X)$. Dually, we also have a map $H^n(G_X; M) \to H_Q^n(X; M)$.

We apply this map for $\mathrm{Conj}(h)$. Since there exists a natural homomorphism from the associated group $G_{\mathrm{Conj}(h)}$ to G, we have a homomorphism

$$H^n(G; A) \to H^n(G_{\mathrm{Conj}(h)}; A) \to H_Q^n(\mathrm{Conj}(h); A).$$

Fenn, Rourke and Sanderson [Fenn et al. 1995] defined the *rack space* BX. Since $\pi_1(BX)$ is isomorphic to G_X, there exists a unique map, up to homotopy, from BX to the Eilenberg–MacLane space $K(G_X, 1)$ that induces the isomorphism between their fundamental groups. This map induces a homomorphism $H^n(G_X; M) \to H_Q^n(X; M)$, which is equal to the map we have constructed in this subsection. Clauwens [2011, Proposition 25] showed that this map vanishes under some conditions on X and M. In particular, when p is odd prime,

$$H^n(D_{2p}; \mathbb{Z}/p) \to H_Q^n(R_p; \mathbb{Z}/p)$$

vanishes for $n > 0$.

9. Quandle cycle and branched cover

In this section, we study the dual of the previous construction. We show that the cycle $C(\mathcal{S})$ associated with a shadow coloring $\mathcal{S} = (\mathcal{A}, \mathcal{R})$ of a knot K gives rise to a group cycle represented by a cyclic branched covering along K and the representation induced from the arc coloring $\mathcal{A}$.

9A. Let X be a quandle. Let D be a diagram of a knot K. For a shadow coloring $\mathcal{S} = (\mathcal{A}, \mathcal{R})$ of D whose arcs and regions are colored by X, define $\mathcal{A} * a$ and $\mathcal{R} * a$ for $a \in X$ by

$$(\mathcal{A} * a)(x) = \mathcal{A}(x) * a, \quad (\mathcal{R} * a)(r) = \mathcal{R}(r) * a \quad \text{(for any arc } x \text{ and region } r).$$

By the axiom (Q3), $\mathcal{S} * a = (\mathcal{A} * a, \mathcal{R} * a)$ is also a shadow coloring.

In the following, we assume that $X = \mathrm{Conj}(h)$ for some group G and that h is in G and satisfies Assumption 8.6. As in the previous section, let $s : \mathrm{Conj}(h) \cong Z(h)\backslash G \to G$ be a section satisfying $s(a * b) = s(a) \mathbin{\widetilde{*}} s(b)$ for $a, b \in X$. Let

$$\iota : C_n^{\Delta}(X) \to C_n(G; \mathbb{Z}) : (a_0, \ldots, a_n) \mapsto \big(s(a_0), \ldots, s(a_n)\big).$$

Composing with $\varphi : C_n^{Q}(X; \mathbb{Z}[X]) \to C_{n+1}^{\Delta}(X)$, we have a map

$$C_2^{Q}(X; \mathbb{Z}[X]) \xrightarrow{\varphi} C_3^{\Delta}(X) \xrightarrow{\iota} C_3(G; \mathbb{Z}).$$

For a shadow coloring $\mathcal{S}$, $\iota\varphi(C(\mathcal{S}))$ is not a cycle of $C_3(G; \mathbb{Z})$ in general. But we can show:

Theorem 9.1. *Let $\mathcal{S} = (\mathcal{A}, \mathcal{R})$ be a shadow coloring of a diagram D of a knot K by* $\mathrm{Conj}(h)$. *Let $a \in \mathrm{Conj}(h)$ be the color of an arc of D. If $h^l = 1$, then*

$$\iota\varphi(C(\mathcal{S})) + \iota\varphi(C(\mathcal{S} * a)) + \iota\varphi(C(\mathcal{S} * a^2)) + \cdots + \iota\varphi(C(\mathcal{S} * a^{l-1})) \in C_3(G; \mathbb{Z}) \tag{9-1}$$

is a group cycle represented by the l-fold cyclic branched covering $\widehat{C}_l$ along the knot K and the representation $\pi_1(\widehat{C}_l) \to G$ induced from the arc coloring $\mathcal{A}$.

Proof. Let $x_1, \ldots, x_n$ be the arcs of the diagram D such that $a = \mathcal{A}(x_1)$. We denote the color $\mathcal{A}(x_i)$ by a_i ($a_1 = a$). The arc coloring $\mathcal{A}$ induces a representation $\rho : \pi_1(S^3 - K) \to G$. Using $s : Z(h)\backslash G \to G$, ρ is given by

$$\rho(x_i) = s(a_i)^{-1} h s(a_i) \in \mathrm{Conj}(h) \subset G.$$

We have

$$s(b * a_i) = h^{-1} s(b) s(a_i)^{-1} h s(a_i) = h^{-1} s(b) \rho(x_i) \in G,$$
$$s(b *^{-1} a_i) = h s(b) s(a_i)^{-1} h^{-1} s(a_i) = h s(b) \rho(x_i)^{-1} \in G,$$

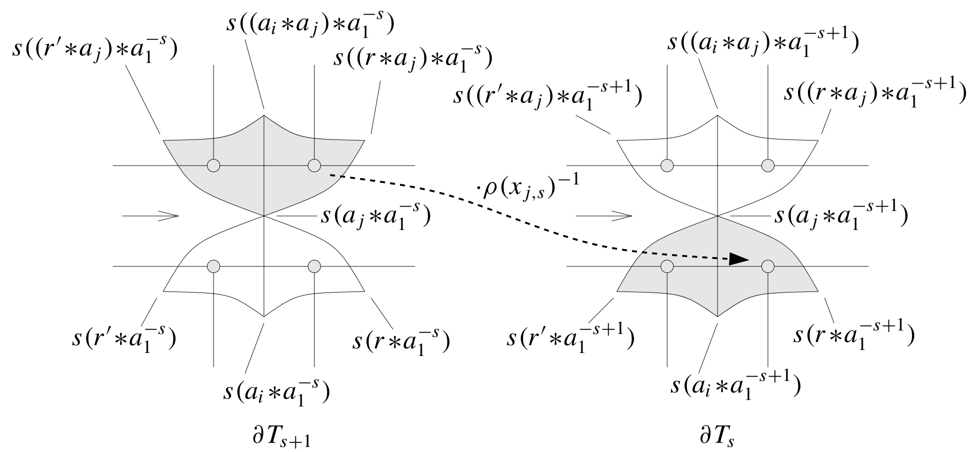


Figure 14. A labeling at a crossing. ($r' = r*^{\pm 1} a_i$.)

for any $b \in \mathrm{Conj}(h)$. So we have

$$s((b*a_i)*a_1^{-s}) = h^{s-1}s(b)\rho(x_i)\rho(x_1)^{-s},$$
$$s(b*a_1^{-s+1}) = h^{s-1}s(b)\rho(x_1)^{-s+1}.$$

Since

$$\rho(x_{i,s}) = \rho(x_1^{s-1}x_i x_1^{-s}) = \rho(x_1)^{s-1}\rho(x_i)\rho(x_1)^{-s}$$

by (4-1), we have

(9-2) $$s(b*a_1^{-s+1}) = s((b*a_i)*a_1^{-s})\rho(x_{i,s})^{-1}.$$

Let $T_s\,(s=0,1,\dots l-1)$ be copies of the triangulation of a 3-ball with a vertex ordering on each tetrahedron constructed in Section 4 (Figures 5 and 7). Then we define a labeling of vertices of T_s by $\iota\varphi(C(\mathcal{S}*a_1^{-s+1}))$ (Figure 14). The vertices of the face $F'^{-}_{j,s+1} \subset T_{s+1}$ are labeled by

$$\big(s((r*a_j)*a_1^{-s}),\quad s((a_i*a_j)*a_1^{-s}),\quad s(a_j*a_1^{-s})\big),$$

and the vertices of the face $F'^{+}_{j,s} \subset T_s$ are labeled by

$$\big(s(r*a_1^{-s+1}),\quad s(a_i*a_1^{-s+1}),\quad s(a_j*a_1^{-s+1})\big),$$

where r is the color of the region as indicated in Figure 14. By (9-2), the labelings of the face $F'^{-}_{j,s+1}$ and $F'^{+}_{j,s}$ are related by

$$\begin{aligned}\big(s(r*a_1^{-s+1}),\quad s(a_i*a_1^{-s+1}),\quad s(a_j*a_1^{-s+1})\big)\\ = \big(s((r*a_j)*a_1^{-s}),\quad s((a_i*a_j)*a_1^{-s}),\quad s(a_j*a_1^{-s})\big)\rho(x_{j,s})^{-1}.\end{aligned}$$

Therefore this gives a G-valued 1-cocycle on $\widehat{T}$ as constructed in Section 4. Thus the chain given by (9-1) is a cycle represented by the cyclic branched cover $\widehat{C}_l$ and the representation induced from the arc coloring $\mathscr{A}$. □

9B. We end this section by comparing the shadow cocycle invariant of the $(2, p)$-torus knot (p: odd) for the 3-cocycle obtained in Section 7 and the Dijkgraaf–Witten invariant of the lens space $L(p, 1)$.

Let G be a finite group and α be a cocycle of $H^3(G; A)$, where A is an abelian group. Dijkgraaf and Witten defined an invariant of closed oriented 3-manifolds for each α. For an oriented closed manifold M, it is defined by

$$\sum_{\rho:\pi_1(M)\to G} \langle \rho^*\alpha, [M]\rangle \in \mathbb{Z}[A],$$

where $\rho^*\alpha$ is the pull-back of α by the classifying map $M \to BG$ corresponding to ρ. Since $L(p, 1)$ is a double branched cover along the $(2, p)$-torus knot, the cycle obtained from a shadow coloring by R_p gives rise to a group 3-cycle of D_{2p} represented by $L(p, 1)$ by Theorem 9.1. Since every representation of $\pi_1(L(p, 1)) \cong \mathbb{Z}/p$ into D_{2p} reduces to a representation into $\mathbb{Z}/p$, it is natural to ask whether the shadow cocycle invariant for our quandle 3-cocycle coincides with the Dijkgraaf–Witten invariant for $b_1b_2 \in H^3(\mathbb{Z}/p; \mathbb{Z}/p)$. As we remarked in the introduction, these invariants coincide up to some constant by the result of Hatakenaka and Nosaka [2012] when p is prime.

Let $\mathscr{S}$ be the shadow coloring indicated in Figure 12. Since the homology class $[C(\mathscr{S})]$ does not change under the action of R_p on the shadow coloring $\mathscr{S}$ (see Lemma 4.5 of [Inoue and Kabaya 2010]), we assume that $x = 0$. As computed in Section 7B, the evaluation of the $C(\mathscr{S})$ at the 3-cocycle derived from b_1b_2 is equal to $-4y^2 \mod p$. Therefore the shadow cocycle invariant is

$$p\sum_{y=0}^{p-1} t^{-4y^2} = p\sum_{y=0}^{p-1} t^{-y^2} \in \mathbb{Z}[t]/(t^p - 1) \cong \mathbb{Z}[\mathbb{Z}/p].$$

We compute the Dijkgraaf–Witten invariant of the lens space $L(p, q)$ for the group 3-cocycle $b_1b_2 \in H^3(\mathbb{Z}/p; \mathbb{Z}/p)$. Although this was computed in [Murakami et al. 1992], we give a proof based on a triangulation. We represent $L(p, q)$ by an ordered 3-cycle and give a labeling of 1-simplices as indicated in Figure 15. Here the triple a, b and c must satisfy $a^p = 1$ and $b = a^q bc$. Using multiplicative notation, the evaluation of b_1b_2 on this cycle is

$$\sum_{i=0}^{p-1} b_1b_2\big([a|ia+b|-qa]\big) = \sum_{i=0}^{p-1} a\cdot b_2(ia+b, -qa) = -qa^2.$$

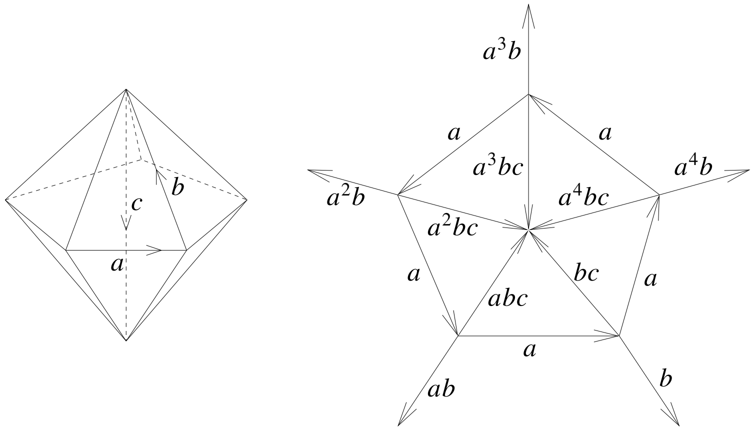


Figure 15. A triangulation of $L(5, q)$ and a labeling where $a^5 = 1$ and $b = a^q bc$.

The second equality follows from the fact that $ia + b$ runs over $\mathbb{Z}/p$ and from (7-1) when a is a unit, and also when a is not a unit by a similar argument. Therefore the Dijkgraaf–Witten invariant of $L(p, q)$ is equal to

$$\sum_{a=0}^{p-1} t^{-qa^2} \in \mathbb{Z}[t]/(t^p - 1) \cong \mathbb{Z}[\mathbb{Z}/p].$$

Therefore this coincides with the shadow cocycle invariant up to a constant.

Acknowledgements

The author thanks Takefumi Nosaka for useful discussions. He also thanks the referees for helpful comments for improving the exposition.

References

[Benson 1991] D. J. Benson, *Representations and cohomology, I: Basic representation theory of finite groups and associative algebras*, Cambridge Studies in Advanced Mathematics **30**, Cambridge University Press, 1991. MR 92m:20005 Zbl 0718.20001

[Brown 1982] K. S. Brown, *Cohomology of groups*, Graduate Texts in Mathematics **87**, Springer, New York, 1982. MR 83k:20002 Zbl 0584.20036

[Carter et al. 2003] J. S. Carter, D. Jelsovsky, S. Kamada, L. Langford, and M. Saito, "Quandle cohomology and state-sum invariants of knotted curves and surfaces", *Trans. Amer. Math. Soc.* **355**:10 (2003), 3947–3989. MR 2005b:57048 Zbl 1028.57003

[Clauwens 2011] F. Clauwens, "The algebra of rack and quandle cohomology", *J. Knot Theory Ramifications* **20**:11 (2011), 1487–1535. MR 2854230 Zbl 05998125

[Eisermann 2003] M. Eisermann, "Homological characterization of the unknot", *J. Pure Appl. Algebra* **177**:2 (2003), 131–157. MR 2003j:57009 Zbl 1013.57002

[Eisermann 2007] M. Eisermann, "Knot colouring polynomials", *Pacific J. Math.* **231**:2 (2007), 305–336. MR 2008j:57014 Zbl 1152.57010

[Etingof and Graña 2003] P. Etingof and M. Graña, "On rack cohomology", *J. Pure Appl. Algebra* **177**:1 (2003), 49–59. MR 2004e:55006 Zbl 1054.16028

[Fenn et al. 1995] R. Fenn, C. Rourke, and B. Sanderson, "Trunks and classifying spaces", *Appl. Categ. Structures* **3**:4 (1995), 321–356. MR 96i:57023 Zbl 0853.55021

[Hatakenaka and Nosaka 2012] E. Hatakenaka and T. Nosaka, "Some topological aspects of 4-fold symmetric quandle invariants of 3-manifolds", *Int. J. Math.* **23**:7 (2012), Art. ID 1250064.

[Inoue and Kabaya 2010] A. Inoue and Y. Kabaya, "Quandle homology and complex volume", preprint, 2010. arXiv 1012.2923v1

[Joyce 1982] D. Joyce, "A classifying invariant of knots, the knot quandle", *J. Pure Appl. Algebra* **23**:1 (1982), 37–65. MR 83m:57007 Zbl 0474.57003

[Kamada 2007] S. Kamada, "Quandles with good involutions, their homologies and knot invariants", pp. 101–108 in *Intelligence of low dimensional topology 2006* (Hiroshima, 2006), edited by J. S. Carter et al., Ser. Knots Everything **40**, World Sci., Hackensack, NJ, 2007. MR 2009a:57042 Zbl 1145.57008

[Mochizuki 2003] T. Mochizuki, "Some calculations of cohomology groups of finite Alexander quandles", *J. Pure Appl. Algebra* **179**:3 (2003), 287–330. MR 2004b:55013 Zbl 1036.18009

[Murakami et al. 1992] H. Murakami, T. Ohtsuki, and M. Okada, "Invariants of three-manifolds derived from linking matrices of framed links", *Osaka J. Math.* **29**:3 (1992), 545–572. MR 93h:57013 Zbl 0776.57009

[Neumann 2004] W. D. Neumann, "Extended Bloch group and the Cheeger–Chern–Simons class", *Geom. Topol.* **8** (2004), 413–474. MR 2005e:57042 Zbl 1053.57010

[Nosaka 2009] T. Nosaka, "On quandle homology groups of Alexander quandles of prime order", preprint, 2009, available at http://www.kurims.kyoto-u.ac.jp/preprint/file/RIMS1680.pdf. To appear in *Trans. Amer. Math. Soc.*

[Nosaka 2010] T. Nosaka, "Quandle homotopy invariants of knotted surfaces", preprint, 2010. arXiv 1011.6035

[Rolfsen 1976] D. Rolfsen, *Knots and links*, Mathematics Lecture Series **7**, Publish or Perish, Berkeley, CA, 1976. MR 58 #24236 Zbl 0339.55004

[Weeks 2005] J. Weeks, "Computation of hyperbolic structures in knot theory", pp. 461–480 in *Handbook of knot theory*, edited by W. Menasco and M. Thistlethwaite, Elsevier, Amsterdam, 2005. MR 2006k:57027 Zbl 1096.57015

Received November 9, 2011. Revised April 13, 2012.

YUICHI KABAYA
DEPARTMENT OF MATHEMATICS
OSAKA UNIVERSITY
TOYONAKA, OSAKA 560-0043
JAPAN

y-kabaya@cr.math.sci.osaka-u.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 259, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.259.349

ON A CLASS OF SEMIHEREDITARY CROSSED-PRODUCT ORDERS

JOHN S. KAUTA

Let F be a field, let V be a valuation ring of F of arbitrary Krull dimension (rank), let K be a finite Galois extension of F with group G, and let S be the integral closure of V in K. Let $f : G \times G \mapsto K \setminus \{0\}$ be a normalized two-cocycle such that $f(G \times G) \subseteq S \setminus \{0\}$, but we do not require that f should take values in the group of multiplicative units of S. One can construct a crossed-product V-algebra $A_f = \sum_{\sigma \in G} S x_\sigma$ in a natural way, which is a V-order in the crossed-product F-algebra $(K/F, G, f)$. If V is unramified and defectless in K, we show that A_f is semihereditary if and only if for all $\sigma, \tau \in G$ and every maximal ideal M of S, $f(\sigma, \tau) \notin M^2$. If in addition $J(V)$ is not a principal ideal of V, then A_f is semihereditary if and only if it is an Azumaya algebra over V.

1. Introduction

In this paper we study certain orders over valuation rings in central simple algebras. If R is a ring, then $J(R)$ will denote its Jacobson radical, $U(R)$ its group of multiplicative units, and $R^{\#}$ the subset of all the nonzero elements. The residue ring $R/J(R)$ will be denoted by $\bar{R}$. Given the ring R, it is called *primary* if $J(R)$ is a maximal ideal of R. It is called *hereditary* if one-sided ideals are projective R-modules. It is called *semihereditary* (respectively *Bézout*) if finitely generated one-sided ideals are projective R-modules (respectively are principal). Let V be a valuation ring of a field F. If Q is a finite-dimensional central simple F-algebra, then a subring R of Q is called an order in Q if $RF = Q$. If in addition $V \subseteq R$ and R is integral over V, then R is called a V-order. If a V-order R is maximal among the V-orders of Q with respect to inclusion, then R is called a maximal V-order (or just a maximal order if the context is clear). A V-order R of Q is called an *extremal* V-order (or simply *extremal* when the context is clear) if for every V-order B in Q with $B \supseteq R$ and $J(B) \supseteq J(R)$, we have $B = R$. If R is an order in Q, then it is

MSC2010: 13F30, 16H05, 16E60, 16S35, 16H10.
Keywords: crossed-product orders, semihereditary orders, hereditary orders, Azumaya algebras, Dubrovin valuation rings.

called a *Dubrovin valuation ring* of Q (or a *valuation ring* of Q in short) if it is semihereditary and primary (see [Dubrovin 1982; 1984]).

In this paper, V will denote a commutative valuation ring of *arbitrary* Krull dimension (rank). Let F be its field of quotients, let K/F be a finite Galois extension with group G, and let S be the integral closure of V in K. If $f \in Z^2(G, U(K))$ is a normalized two-cocycle such that $f(G \times G) \subseteq S^{\#}$, then one can construct a "crossed-product" V-algebra

$$A_f = \sum_{\sigma \in G} S x_\sigma,$$

with the usual rules of multiplication ($x_\sigma s = \sigma(s) x_\sigma$ for all $s \in S$, $\sigma \in G$ and $x_\sigma x_\tau = f(\sigma, \tau) x_{\sigma\tau}$). Then A_f is associative, with identity $1 = x_1$, and center $V = V x_1$. Further, A_f is a V-order in the crossed-product F-algebra $\Sigma_f = \sum_{\sigma \in G} K x_\sigma = (K/F, G, f)$. Following [Haile 1987], we let $H = \{\sigma \in G \mid f(\sigma, \sigma^{-1}) \in U(S)\}$. Then H is a subgroup of G.

In this paper, we will *always* assume that V is unramified and defectless in K (for the definitions of these terms, see [Endler 1972]). By [Endler 1972, Theorem 18.6], S is a finitely generated V-module, hence A_f is always finitely generated over V. If V_1 is a valuation ring of K lying over V then $\{\sigma \in G \mid \sigma(x) - x \in J(V_1)\ \forall x \in V_1\}$ is called the *inertial group* of V_1 over F. By [Kauta 2001, Lemma 1], the condition that V is unramified and defectless in K is equivalent to saying that the inertial group of V_1 over F is trivial, since K/F is a finite Galois extension.

These orders were first studied in [Haile 1987], and later in [Haile and Morandi 1993; Kauta 2012]. In [Haile 1987; Kauta 2012], only the case when V is a discrete valuation ring (DVR) was considered. In [Kauta 2012], hereditary properties of crossed-product orders were examined. In [Haile 1987; Haile and Morandi 1993], valuation ring properties of the crossed-product orders were explored, and the latter considered the cases when either V has arbitrary Krull dimension but is indecomposed in K, or V is a discrete finite-rank valuation ring, that is, its value group is $\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}$. When V is a DVR, then any V-order in Σ_f containing S is a crossed-product order of the form A_g for some two-cocycle $g : G \times G \mapsto S^{\#}$, with g cohomologous to f over K, by [Haile 1987, Proposition 1.3], but this need not be the case in general. While [Haile and Morandi 1993] considered any V-order in Σ_f containing S, some of which were not of the type described above and so in that sense its scope was wider than ours, in this paper we shall only be concerned with crossed-product orders A_g where g is either f (almost always), or is cohomologous to f over K, that is, if there are elements $\{c_\sigma \mid \sigma \in G\} \subseteq K^{\#}$ such that $g(\sigma, \tau) = c_\sigma \sigma(c_\tau) c_{\sigma\tau}^{-1} f(\sigma, \tau)$ for all $\sigma, \tau \in G$, a fact denoted by $g \sim_K f$.

The purpose of this paper is to generalize the results of [Kauta 2012] to the case when V is not necessarily a DVR. The main results of this paper are as follows:

A_f is semihereditary if and only if for all $\sigma, \tau \in G$ and every maximal ideal M of S, $f(\sigma, \tau) \notin M^2$; if $J(V)$ is not a principal ideal of V, then A_f is semihereditary if and only if it is an Azumaya algebra over V. As in [Kauta 2012], the utility of these criteria lie in their simplicity.

Although in our case the valuation ring V need not be a DVR, some of the steps in the proofs in [Haile 1987; Kauta 2012] remain valid, mutatis mutandis, owing to the theory developed in [Kauta 1997a; 1997b]. We shall take full advantage of this whenever the opportunity arises. Aside from the difficulties inherent when dealing with V-orders that are not necessarily noetherian, the hurdles encountered in this theory arise mainly due to the fact that the two-cocycle f is not assumed to take on values in $U(S)$.

2. Preliminaries

In this section, we gather together various results that will help us prove the main results of this paper, which are in the next section. For the convenience of the reader, we have included complete proofs whenever it warrants, although the arguments are sometimes routine.

The following lemma is essentially embedded in the proof of [Kauta 1997a, Proposition 1.8], and the remark that follows it.

Lemma 2.1. *Let A be a finitely generated extremal V-order in a finite-dimensional central simple F-algebra Q.*

(1) *If B is a V-order of Q containing A, then B is also a finitely generated extremal order. If in addition B is a maximal V-order, then it is a valuation ring of Q.*

(2) *If W is an overring of V in F with $V \subsetneqq W$, then WA is a valuation ring of Q with center W.*

Proof. Let B be a V-order containing A. By [Kauta 1997a, Proposition 1.8], A is semihereditary, hence B is semihereditary by [Morandi 1992, Lemma 4.10], and therefore B is extremal by [Kauta 1997a, Theorem 1.5]. Since $[B/J(B) : V/J(V)] \leq [\Sigma_f : F] < \infty$, there exists $a_1, a_2, \ldots, a_m \in B$ such that $B = a_1V + a_2V + \cdots + a_mV + J(B)$. But by [Kauta 1997a, Proposition 1.4], $J(B) \subseteq J(A)$, since A is extremal. Therefore $B = a_1V + a_2V + \cdots + a_mV + A$, a finitely generated V-order. If, in addition, B is a maximal V-order, then by the remark after [Kauta 1997a, Proposition 1.8], B is a valuation ring of Q.

Now let W be a proper overring of V in F. Let C be a maximal V-order containing A. Then C is a valuation ring of Q, as seen above, hence WC is a valuation ring of Q with center W. Since A is an extremal V-order, we have $J(C) \subseteq J(A)$, thus $WC = WJ(V)C \subseteq WJ(C) \subseteq WA \subseteq WC$, so that $WA = WC$. Thus WA is always a valuation ring of Q. □

Since A_f is finitely generated over V, we immediately have the following lemma, because of [Kauta 1997a, Proposition 1.8], the remark that follows it, and the fact that Bézout V-orders are maximal orders by [Morandi 1992, Theorem 3.4].

Lemma 2.2. *Given the crossed-product order A_f,*

(1) *it is an extremal order if and only if it is semihereditary and*

(2) *it is a maximal order if and only if it is a valuation ring, if and only if it is Bézout.*

Lemma 2.3. *Let W be a valuation ring of F such that $V \subsetneq W$, and let $R = WS$.*

(1) *Then R is the integral closure of W in K, and W is also unramified and defectless in K.*

(2) *Let $t \in S$ satisfy $t \notin M^2$ for every maximal ideal M of S. Then $t \in U(R)$. If in addition $J(V)$ is a nonprincipal ideal of V, then $t \in U(S)$.*

Proof. The ring R is obviously integral over W. Since it contains S, it is also integrally closed in K, hence it is the integral closure of W in K.

Now let $V_1 \subseteq W_1$ be valuation rings of K lying over V and W respectively. Then $J(W_1) \subseteq J(V_1)$; hence the inertial group of W_1 over F, namely

$$\{\sigma \in G \mid \sigma(x) - x \in J(W_1)\ \forall x \in W_1\},$$

is contained in the inertial group of V_1 over F, $\{\sigma \in G \mid \sigma(x) - x \in J(V_1)\ \forall x \in V_1\}$. Since V is unramified and defectless in K, the latter group is trivial, forcing W to be unramified and defectless in K.

Let W_1 be a valuation ring of K lying over W, and let V_1 be a valuation ring of K lying over V such that $V_1 \subseteq W_1$, as in the preceding paragraph. Let $M = J(V_1) \cap S$, a generic maximal ideal of S. We claim that $M^2 = J(V_1)^2 \cap S$. To see this, note that

$$M^2 = (J(V_1) \cap S)(J(V_1) \cap S) \subseteq J(V_1)^2 \cap S$$

and

$$M^2 V_1 = (J(V_1) \cap S)(J(V_1) \cap S)V_1 = J(V_1)^2 = (J(V_1)^2 \cap S)V_1.$$

If V' is an extension of V to K different from V_1, then $M^2 V' = V' = (J(V_1)^2 \cap S)V'$. Thus $M^2 = J(V_1)^2 \cap S$ as desired. If $t \in S$ satisfies $t \notin M^2$, then $t \notin J(V_1)^2$. Since $J(W_1) \subsetneq J(V_1)^2$, we have $t \in U(W_1)$. Since W_1 was an arbitrary extension of W in K, we conclude that $t \in U(R)$. If $J(V)$ is a nonprincipal ideal of V, then $J(V_1)^2 = J(V_1)$, hence $t \in U(V_1)$ for every such extension V_1 of V to K, and we conclude that $t \in U(S)$. □

Part (4) of the following lemma was originally proved in [Haile 1987] when V is a DVR. The same arguments work when V is an arbitrary valuation ring.

Lemma 2.4. *Given a $\sigma \in G$, let $I_\sigma = \bigcap M$, where the intersection is taken over those maximal ideals M of S for which $f(\sigma, \sigma^{-1}) \notin M$. Then*:

(1) $I_\sigma = \{x \in S \mid xf(\sigma, \sigma^{-1}) \in J(V)S\}$.

(2) $I_\sigma^{\sigma^{-1}} = I_{\sigma^{-1}}$.

(3) *If $f(\sigma, \sigma^{-1}) \notin M^2$ for every maximal ideal M of S, then $I_\sigma f(\sigma, \sigma^{-1}) = J(V)S$.*

(4) $J(A_f) = \sum_{\sigma \in G} I_\sigma x_\sigma$.

Proof. Let $x \in S$. Clearly, if $x \in I_\sigma$ then $xf(\sigma, \sigma^{-1}) \in J(V)S$. On the other hand, if $x \notin I_\sigma$ then there exists a maximal ideal M of S such that $x, f(\sigma, \sigma^{-1}) \notin M$, hence $xf(\sigma, \sigma^{-1}) \notin M$, and thus $xf(\sigma, \sigma^{-1}) \notin J(V)S$.

The second statement is proved in the same manner as [Kauta 2012, Sublemma]. To see that the third statement holds, we note that $I_\sigma f(\sigma, \sigma^{-1}) \subseteq J(V)S$. We claim that $I_\sigma f(\sigma, \sigma^{-1}) = J(V)S$. To see this, let M be a maximal ideal of S. If $f(\sigma, \sigma^{-1}) \notin M$, then $(I_\sigma f(\sigma, \sigma^{-1}))S_M = J(S_M) = (J(V)S)S_M$. On the other hand, if $f(\sigma, \sigma^{-1}) \in M$ then, since $f(\sigma, \sigma^{-1}) \notin M^2$, we have $J(S_M)^2 \subsetneq I_\sigma f(\sigma, \sigma^{-1})S_M \subseteq J(S_M)$, hence $I_\sigma f(\sigma, \sigma^{-1})S_M = J(S_M) = (J(V)S)S_M$, and thus $I_\sigma f(\sigma, \sigma^{-1}) = J(V)S$. By [Haile and Morandi 1993, Lemma 1.3], $J(A_f) = \sum_{\sigma \in G}(J(A_f) \cap Sx_\sigma)$. Therefore the fourth statement can be verified in exactly the same manner as [Haile 1987, Proposition 3.1(b)], because of the observations made above. □

The following lemma is a generalization of [Haile 1987, Proposition 1.3].

Lemma 2.5. *Let $B \subseteq \Sigma_f$ be a V-order. There is a normalized cocycle $g : G \times G \mapsto S^\#$, $g \sim_K f$, such that $B = A_g$ (viewed as a subalgebra of Σ_f in a natural way) if and only if $B \supseteq S$ and B is finitely generated over V. When this occurs, $B = \sum_{\sigma \in G} Sk_\sigma x_\sigma$ for some $k_\sigma \in K^\#$.*

Proof. Suppose $B \supseteq S$. By [Haile and Morandi 1993, Lemma 1.3], $B = \sum_{\sigma \in G} B_\sigma x_\sigma$, where each B_σ is a nonzero S-submodule of K. If in addition B is finitely generated over V, then each B_σ is finitely generated over V: if $B = \sum_{i=1}^n Vy_i$ then, if we write $y_i = \sum_{\tau \in G} k_\tau^{(i)} x_\tau$ with $k_\tau^{(i)} \in K$, we see that B_σ is generated by $\{k_\sigma^{(i)}\}_{i=1}^n$ over V. Since S is a commutative Bézout domain with K as its field of quotients, $B_\sigma = Sk_\sigma$ for some $k_\sigma \in K^\#$. Thus we get $B = \sum_{\sigma \in G} Sk_\sigma x_\sigma$. Since B is integral over V, $B_1 = S$ and so we can choose $k_1 = 1$. Define $g : G \times G \mapsto S^\#$ by $g(\sigma, \tau)k_{\sigma\tau}x_{\sigma\tau} = (k_\sigma x_\sigma)(k_\tau x_\tau)$, as in [Haile 1987, Proposition 1.3]. Since $k_1 = 1$, g is also a normalized two-cocycle. The converse is obvious. □

Lemma 2.6. *Suppose S is a valuation ring of K. Then*

$$J(V)A_f \text{ is a maximal ideal of } A_f \iff H = G \iff A_f \text{ is Azumaya over } V.$$

Proof. Suppose $J(V)A_f$ is a maximal ideal of A_f. Note that $A_f/J(V)A_f = \sum_{\sigma\in G}\bar{S}\tilde{x}_\sigma$. By [Haile et al. 1983, Theorem 10.1(c)], $J=\sum_{\sigma\notin H}\bar{S}\tilde{x}_\sigma$ is an ideal of $A_f/J(V)A_f$. Since $A_f/J(V)A_f$ is simple, $J=0$, hence $H=G$. □

We set up additional notation, following [Haile 1987; Kauta 2012]. Let L be an intermediate field of F and K, let G_L be the Galois group of K over L, let U be a valuation ring of L lying over V, and let T be the integral closure of U in K. Then one can obtain a two-cocycle $f_{L,U}: G_L\times G_L \mapsto T^{\#}$ from f by restricting f to $G_L\times G_L$, and embedding $S^{\#}$ in $T^{\#}$. As before, $A_{f_{L,U}}=\sum_{\sigma\in G_L}Tx_\sigma$ is a U-order in $\Sigma_{f_{L,U}}=\sum_{\sigma\in G_L}Kx_\sigma=(K/L, G_L, f_{L,U})$, and U is unramified and defectless in K. If M is a maximal ideal of S, and L is the decomposition field of M and $U=L\cap S_M$, then we will denote $f_{L,U}$ by f_M, $A_{f_{L,U}}$ by A_{f_M}, $\Sigma_{f_{L,U}}$ by Σ_{f_M}, L by K_M, and the decomposition group G_L by D_M, as in [Haile 1987]. Further, we let $H_M=\{\sigma\in D_M \mid f_M(\sigma,\sigma^{-1})\in U(S_M)\}$, a subgroup of D_M.

Given a maximal ideal M of S, let $M=M_1, M_2, \dots, M_r$ be the complete list of maximal ideals of S, let $U_i=S_{M_i}\cap K_{M_i}$ with $U=U_1$, and let (K_i, S_i) be a Henselization of (K, S_{M_i}). Let (F_h, V_h) be the unique Henselization of (F, V) contained in (K_1, S_1). We note that (F_h, V_h) is also a Henselization of (K_M, U). By [Haile et al. 1995, Proposition 11], we have $S\otimes_V V_h\cong S_1\oplus S_2\oplus\cdots\oplus S_r$.

Part (1) of the following lemma was originally proved in [Haile 1987] in the case when V is a DVR. Virtually the same proof holds in the general case. Part (2c) is a generalization of [Haile 1987, Corollary 3.11].

Lemma 2.7. *Let the notation be as above.*

(1) *The crossed-product order A_f is primary if and only if for every maximal ideal M of S there is a set of right coset representatives $g_1, g_2, \dots, g_r$ of D_M in G (that is, G is the disjoint union $\bigcup_i D_M g_i$) such that for all i, $f(g_i, g_i^{-1})\notin M$.*

(2) *Assume the crossed-product order A_f is primary. Then*:
   (a) $A_f\otimes_V V_h\cong M_r(A_{f_M}\otimes_U V_h)$.
   (b) *As a result of* (a), $A_f/J(A_f)\cong M_r(A_{f_M}/J(A_{f_M}))$.
   (c) *Also as a result of* (a), *A_f is a valuation ring of Σ_f if and only if A_{f_M} is a valuation ring of Σ_{f_M} for some maximal ideal M of S. When this occurs, A_{f_M} is a valuation ring of Σ_{f_M} for every maximal ideal M of S.*
   (d) *A_f is Azumaya over V if and only if $H_M=D_M$ for some maximal ideal M of S. When this occurs, $H_M=D_M$ for every maximal ideal M of S.*

Proof. The proof of [Haile 1987, Theorem 3.2], appropriately adapted, works here as well to establish part (1). We outline the argument, for the convenience of the reader: For a $\sigma\in G$, let I_σ be as in Lemma 2.4, and, for a maximal ideal M of S, set $\hat{M}:=\bigcap_{N \text{ max}, N\neq M}N$. If I is an ideal of A_f then, by [Haile and Morandi 1993, Lemma 1.3], $I=\sum_{\sigma\in G}(I\cap Sx_\sigma)$, so A_f is primary if and only if the following condition holds: if $\sigma \in G$ and T is an ideal of S such that $T \not\subseteq I_\sigma$, then $A_f T x_\sigma A_f = A_f$.

If A_f is primary and M is a maximal ideal of S, then $A_f = A_f \hat{M} x_1 A_f$. Therefore if $G = \bigcup_{j=1}^{r} h_j D_M$ is a left coset decomposition, then

$$S = \sum_j \hat{M}^{h_j} \Big(\sum_{d \in D_M} f(h_j d, d^{-1} h_j^{-1}) \Big),$$

as in the proof of [Haile 1987, Theorem 3.2], so that, if we fix i, $1 \le i \le r$, and localize at M^{h_i}, we get

$$S_{M^{h_i}} = \sum_{j \neq i} J(S_{M^{h_i}}) \Big(\sum_{d \in D_M} f(h_j d, d^{-1} h_j^{-1}) \Big) + S_{M^{h_i}} \Big(\sum_{d \in D_M} f(h_i d, d^{-1} h_i^{-1}) \Big),$$

and hence $\sum_{d \in D_M} f(h_i d, d^{-1} h_i^{-1}) \notin M^{h_i}$. So there is an element $d_i \in D_M$ such that $f(h_i d_i, d_i^{-1} h_i^{-1}) \notin M^{h_i}$. Let $g_i = d_i^{-1} h_i^{-1}$. Then $g_1, g_2, \ldots, g_r$ have the desired properties.

For the converse, suppose $\sigma \in G$ and T is an ideal of S such that $T \not\subseteq I_\sigma$. We need to show that $A_f T x_\sigma A_f = A_f$. Since $T \not\subseteq I_\sigma$, there is a maximal ideal M of S such that $f(\sigma, \sigma^{-1}) \notin M$ and $T \not\subseteq M$. The argument in [Haile 1987, Theorem 3.2] shows that $A_f T x_\sigma A_f \supseteq \sum_{i=1}^{r} T_i$, where $T_i = T^{g_i^{-1}} f^{g_i^{-1}}(\sigma, \sigma^{-1} g_i) f(g_i^{-1}, g_i)$ are ideals of S satisfying the condition $T_i \not\subseteq M^{g_i^{-1}}$. Inasmuch as $g_1^{-1}, g_2^{-1}, \ldots, g_r^{-1}$ form a complete set of *left* coset representatives of D_M in G, the ideal $\sum_{i=1}^{r} T_i$ is not contained in any maximal ideal of S. Therefore $\sum_{i=1}^{r} T_i = S$, and so $A_f T x_\sigma A_f = A_f$.

Using part (1) and the fact that $S \otimes_V V_h \cong S_1 \oplus S_2 \oplus \cdots \oplus S_r$, we can construct a full set of matrix units in $A_f \otimes_V V_h$ and hence verify part (2a), as in the proof of [Haile 1987, Theorem 3.12] (see also the remark following that theorem). Part (2b) follows from (2a) and [Kauta 1997a, Lemma 3.1]; part (2c) follows from (2a); and (2d) follows from (2a) and Lemma 2.6. □

3. The main results

We now give the main results of this paper. There are essentially two parallel theories: one takes effect when $J(V)$ is a principal ideal of V, and the other when it is not. In the former case, the order A_f displays characteristics akin to the situation when V is a DVR. Our theory, however, yields surprising results in the latter case. It turns out in this case that the property that A_f is Azumaya over V is equivalent to a much weaker property: that it is an extremal V-order in Σ_f.

Proposition 3.1. *The order A_f is Azumaya over V if and only if $H = G$.*

Proof. Suppose A_f is Azumaya over V. Let M be a maximal ideal of S. By Lemma 2.7(1), there is a set of right coset representatives $g_1, g_2, \ldots, g_r$ of D_M

in G such that $f(g_i, g_i^{-1}) \notin M$. If $\sigma \in G$, then $\sigma = hg_i$ for some $h \in D_M$ and some i. Since A_f is Azumaya, $H_M = D_M$ by Lemma 2.7(2d), hence we have $f(h^{-1}, h) \notin M$. Because

$$f^{h^{-1}}(hg_i, g_i^{-1}h^{-1}) f^{h^{-1}}(h, g_i) f^{g_i}(g_i^{-1}, h^{-1}) = f(h^{-1}, h) f(g_i, g_i^{-1}),$$

we conclude that $f(\sigma, \sigma^{-1}) \notin M$. Since M is arbitrary, $f(\sigma, \sigma^{-1}) \in U(S)$ for every $\sigma \in G$, so that $H = G$.

The converse is well-known and straightforward to demonstrate. □

It is perhaps instructive to compare the above proposition to [Kauta 2001, Theorem 3].

Recall that $J(V)$ is a nonprincipal ideal of V if and only if $J(V)^2 = J(V)$.

Proposition 3.2. *Suppose $J(V)$ is a nonprincipal ideal of V. Then the following statements about the crossed-product order A_f are equivalent*:

(1) *A_f is an extremal V-order in Σ_f.*

(2) *A_f is a semihereditary V-order.*

(3) *A_f is a maximal V-order in Σ_f.*

(4) *A_f is a Bézout V-order.*

(5) *A_f is a valuation ring of Σ_f.*

(6) *A_f is Azumaya over V.*

Proof. By Lemma 2.2, it suffices to demonstrate that (1)⇒(5)⇒(6). So suppose A_f is an extremal V-order. Let B be a maximal V-order containing A_f. By Lemma 2.1, B is a valuation ring finitely generated over V. By Lemma 2.5, we get that $B = \sum_{\sigma \in G} Sk_\sigma x_\sigma$ for some $k_\sigma \in K^\#$. Since A_f is extremal, we have $J(B) \subseteq J(A_f)$ by [Kauta 1997a, Proposition 1.4], so $J(V)B \subseteq A_f$. Therefore $\sum_{\sigma \in G} J(S)k_\sigma x_\sigma = J(V)B = J(V)^2 B \subseteq J(V)A_f = \sum_{\sigma \in G} J(S)x_\sigma$, so that $J(S)k_\sigma \subseteq J(S)$. Hence for each maximal ideal M of S, we have $S_M J(S)k_\sigma \subseteq S_M J(S)$, that is, $J(S_M)k_\sigma \subseteq J(S_M)$. This shows that $k_\sigma \in S_M$ for all M and so $k_\sigma \in S$ for every $\sigma \in G$, and thus $A_f = B$, a valuation ring.

Now suppose A_f is a valuation ring of Σ_f. By Lemma 2.7(2), to show that A_f is Azumaya over V, we may as well assume S is a valuation ring of K. By [Dubrovin 1984, §2, Theorem 1], $J(A_f) = J(V)A_f$, and so A_f is Azumaya over V by Lemma 2.6. □

Remark. It follows from Lemma 2.3(2) and Proposition 3.1 that, if $J(V)$ is a nonprincipal ideal of V, then the crossed-product order A_f is extremal if and only if for all $\tau, \gamma \in G$ and every maximal ideal M of S, $f(\tau, \gamma) \notin M^2$.

If W is a valuation ring of F such that $V \subsetneq W$, then we will denote by B_f the W-order $WA_f = \sum_{\sigma \in G} Rx_\sigma$, where $R = WS$ is the integral closure of W in K by Lemma 2.3. Recall that W is also unramified and defectless in K.

Proposition 3.3. *Suppose $J(V)$ is a principal ideal of V. Then A_f is semihereditary if and only if for all $\tau, \gamma \in G$ and every maximal ideal M of S, $f(\tau, \gamma) \notin M^2$.*

Proof. The result holds when the Krull dimension of V is one, by [Kauta 2012, Corollary], since V is a DVR in this case. So let us assume from now on that the Krull dimension of V is greater than one.

Let $p = \bigcap_{n \geq 1} J(V)^n$. Then p is a prime ideal of V, $W = V_p$ is a minimal overring of V in F, and $\tilde{V} = V/J(W)$ is a DVR of $\overline{W}$. Set $B_f = WA_f$, as above.

Suppose A_f is semihereditary. We will show that for each $\tau \in G$ and each maximal ideal M of S, $f(\tau, \tau^{-1}) \notin M^2$.

First, assume that V is indecomposed in K. By [Haile and Morandi 1993, Proposition 2.6], A_f is primary, hence it is a valuation ring of Σ_f. Therefore B_f is Azumaya over W, by [Haile and Morandi 1993, Proposition 2.10], and $f(G \times G) \subseteq U(R)$, by Proposition 3.1. Observe that R is a valuation ring of K lying over W and $\bar{R}$ is Galois over $\overline{W}$, with group G, and $B_f/J(B_f) = \sum_{\sigma \in G} \bar{R}\tilde{x}_\sigma$ is a crossed-product $\overline{W}$-algebra. Further, $A_f/J(B_f)$ has center $\tilde{V}$, a DVR of $\overline{W}$, and is a crossed-product $\tilde{V}$-order in $B_f/J(B_f)$ of the type under consideration in this paper, since $\tilde{V}$ is unramified in $\bar{R}$ and $f(G \times G) \subseteq S \cap U(R)$. Since the crossed-product $\tilde{V}$-order $A_f/J(B_f)$ is a valuation ring of $B_f/J(B_f)$ and hence hereditary, it follows from [Kauta 2012, Theorem] that for each $\tau \in G$, $f(\tau, \tau^{-1}) \notin J(S)^2$.

Suppose V is not necessarily indecomposed in K, but assume A_f is a valuation ring. Fix a maximal ideal M of S. By Lemma 2.7(1), there is a set of right coset representatives $g_1, g_2, \ldots, g_r$ of D_M in G such that $f(g_i, g_i^{-1}) \notin M$. If $\tau \in G$, then $\tau = hg_i$ for some $h \in D_M$ and some i. By Lemma 2.7(2), A_{f_M} is a valuation ring of Σ_{f_M}. Hence, by the preceding paragraph, $f_M(h^{-1}, h) \notin M^2$, and thus $f(h^{-1}, h) \notin M^2$. But the following holds:

$$f^{h^{-1}}(hg_i, g_i^{-1}h^{-1}) f^{h^{-1}}(h, g_i) f^{g_i}(g_i^{-1}, h^{-1}) = f(h^{-1}, h) f(g_i, g_i^{-1}).$$

Therefore we must have $f(\tau, \tau^{-1}) \notin M^2$.

Now suppose that A_f is not necessarily a valuation ring. To show that for each $\tau \in G$ and each maximal ideal M of S we have $f(\tau, \tau^{-1}) \notin M^2$, one only needs to emulate the corresponding steps in the proof of [Kauta 2012, Theorem], equipped with the following four observations:

1. Any maximal V-order containing A_f is a valuation ring, by Lemma 2.1, hence A_f is the intersection of finitely many valuation rings all with center V, since $J(V)$ is a principal ideal of V, by [Kauta 1997b, Theorem 2.5].

2. If B is one such valuation ring containing A_f, then $B = A_g = \sum_{\tau \in G} Sk_\tau x_\tau$ for some $k_\tau \in K^\#$, where $g : G \times G \mapsto S^\#$ is some normalized two-cocycle, by Lemma 2.1(1) and Lemma 2.5. Fix $\sigma \in G$ and a maximal ideal N of S. We may choose B such that $k_\sigma \in U(S_N)$, as in the proof of [Kauta 2012, Theorem].

3. Both $J(A_f)$ and $J(A_g)$ are as in Lemma 2.4, that is, $J(A_f) = \sum_{\sigma \in G} I_\sigma x_\sigma$ (respectively $J(B_f) = \sum_{\sigma \in G} J_\sigma k_\sigma x_\sigma$) where $I_\sigma = \bigcap M$ (respectively $J_\sigma = \bigcap M$), as M runs through all maximal ideals of S for which $f(\sigma, \sigma^{-1}) \notin M$ (respectively $g(\sigma, \sigma^{-1}) \notin M$). We have $J(A_g) \subseteq J(A_f)$ by [Kauta 1997a, Theorem 1.5].

4. By Lemma 2.4, $I_\sigma^{\sigma^{-1}} = I_{\sigma^{-1}}$, $J_\sigma^{\sigma^{-1}} = J_{\sigma^{-1}}$, and $J_{\sigma^{-1}} g(\sigma^{-1}, \sigma) = J(V)S$.

We conclude, as in the proof of [Kauta 2012, Theorem], that

$$J(V)S \subseteq k_\sigma I_\sigma f(\sigma, \sigma^{-1}). \tag{1}$$

Since $k_\sigma \in U(S_N)$, if $f(\sigma, \sigma^{-1}) \in N^2$ then, localizing both sides of (1) at N we get $J(S_N) \subseteq J(S_N)^2$, a contradiction, since $J(V)$ is a principal ideal of V. Therefore for each $\tau \in G$ and each maximal ideal M of S, $f(\tau, \tau^{-1}) \notin M^2$. Since the cocycle identity $f^\tau(\tau^{-1}, \tau\gamma) f(\tau, \gamma) = f(\tau, \tau^{-1})$ holds, we conclude that for all $\tau, \gamma \in G$ and every maximal ideal M of S, $f(\tau, \gamma) \notin M^2$.

Conversely, suppose $f(\tau, \gamma) \notin M^2$ for all $\tau, \gamma \in G$, and every maximal ideal M of S. Let $O_l(J(A_f)) = \{x \in \Sigma_f \mid xJ(A_f) \subseteq J(A_f)\}$. We will first establish that $O_l(J(A_f)) = A_f$, again emulating the relevant steps in the proof of [Kauta 2012, Theorem]. To achieve this, it suffices to show that $O_l(J(A_f)) = \sum_{\tau \in G} Sk_\tau x_\tau$ for some $k_\tau \in K^\#$, and that $I_\tau f(\tau, \tau^{-1}) = J(V)S$ for each $\tau \in G$, where I_τ is as in Lemma 2.4. The second assertion follows from Lemma 2.4(3). As for the first one, we first note that $O_l(J(A_f))$ is a V-order in Σ_f, by [Kauta 1997a, Corollary 1.3]. By Lemma 2.5, $O_l(J(A_f)) = \sum_{\tau \in G} Sk_\tau x_\tau$ for some $k_\tau \in K^\#$ if and only if it is finitely generated over V.

Since for all $\tau, \gamma \in G$ and every maximal ideal M of S we have $f(\tau, \gamma) \notin M^2$, we conclude from Lemma 2.3 that $f(G \times G) \subseteq U(R)$, hence B_f is Azumaya over W. Therefore $J(B_f) = J(W)B_f = J(W)(WA_f) = J(W)A_f \subseteq J(A_f)$, and $A_f/J(B_f)$ is a $\tilde{V}$-order in $B_f/J(B_f)$. Since $O_l(J(A_f))$ is a V-order containing A_f, $O_l(J(A_f))W$ is a W-order containing B_f, so $O_l(J(A_f))W = B_f$, since B_f is a maximal W-order in Σ_f, and hence $O_l(J(A_f)) \subseteq B_f$. Therefore $O_l(J(A_f))/J(B_f)$ is a $\tilde{V}$-order in $B_f/J(B_f)$, a central simple $\overline{W}$-algebra. Since $\tilde{V}$ is a DVR of $\overline{W}$, $O_l(J(A_f))/J(B_f)$ must be finitely generated over $\tilde{V}$, by [Reiner 2003, Theorem 10.3], hence there exists $a_1, a_2, \ldots, a_n \in O_l(J(A_f))$ such that $O_l(J(A_f)) = a_1V + a_2V + \cdots + a_nV + J(B_f) = a_1V + a_2V + \cdots + a_nV + A_f$, a finitely generated V-module. Thus $O_l(J(A_f)) = A_f$.

As in the proof of [Morandi 1992, Lemma 4.11], we have $O_l(J(A_f/J(B_f))) = O_l(J(A_f)/J(B_f)) = O_l(J(A_f))/J(B_f) = A_f/J(B_f)$, where $O_l(J(A_f/J(B_f)))$

and $O_l(J(A_f)/J(B_f))$ are defined accordingly. Since $\tilde{V}$ is a DVR of $\overline{W}$, $A_f/J(B_f)$ is a hereditary $\tilde{V}$-order in the central simple $\overline{W}$-algebra $B_f/J(B_f)$, hence A_f is semihereditary by [Morandi 1992, Lemma 4.11]. □

We summarize these results as follows.

Theorem 3.4. *Given a crossed-product order A_f:*

(1) *It is semihereditary if and only if for all $\tau, \gamma \in G$ and every maximal ideal M of S, $f(\tau, \gamma) \notin M^2$; if and only if for each $\gamma \in G$ and each maximal ideal M of S, $f(\tau, \tau^{-1}) \notin M^2$.*

(2) *If $J(V)$ is a nonprincipal ideal of V, then A_f is semihereditary if and only if it is Azumaya over V, if and only if $H = G$.*

We now lump together several corollaries of the theorem above, generalizing results in [Kauta 2012].

Corollary 3.5. (1) *Given a crossed-product order A_f:*

(a) *It is a valuation ring if and only if given any maximal ideal M of S, $f(\tau, \tau^{-1}) \notin M^2$ for each $\tau \in G$, and there exists a set of right coset representatives $g_1, g_2, \ldots, g_r$ of D_M in G (that is, G is the disjoint union $\bigcup_i D_M g_i$) such that for all i, $f(g_i, g_i^{-1}) \notin M$.*

(b) *If V is indecomposed in K, then it is a valuation ring if and only if for each $\tau \in G$, $f(\tau, \tau^{-1}) \notin J(S)^2$.*

(2) *Suppose the crossed-product order A_f is primary. Then it is a valuation ring if and only if there exists a maximal ideal M of S such that for each $\tau \in D_M$, $f(\tau, \tau^{-1}) \notin M^2$.*

(3) *Suppose the crossed-product order A_f is semihereditary. Then $A_{f_{L,U}}$ is a semihereditary order in $\Sigma_{f_{L,U}}$ for each intermediate field L of F and K, and every valuation ring U of L lying over V.*

(4) *Suppose the crossed-product order A_f is semihereditary. Then A_{f_M} is a valuation ring of Σ_{f_M} for each maximal ideal M of S.*

We end by observing yet another peculiarity of these crossed-product orders. The proposition below not only strengthens Lemma 2.1(2) when the V-order A is taken to be the crossed-product order A_f, but also generalizes [Haile and Morandi 1993, Proposition 2.10] to the case where V is not necessarily indecomposed in K.

Proposition 3.6. *Suppose the crossed-product order A_f is extremal and W is a valuation ring of F with $V \subsetneq W$. Then WA_f is Azumaya over W.*

Proof. This follows from Lemma 2.2(1), Theorem 3.4(1), Lemma 2.3, and Proposition 3.1. □

References

[Dubrovin 1982] N. I. Dubrovin, "Noncommutative valuation rings", *Trudy Moskov. Mat. Obshch.* **45** (1982), 265–280. In Russian; translated in *Trans. Moskow Math. Soc.* **45** (1984), 273–287. MR 85d:16002 Zbl 0543.16010

[Dubrovin 1984] N. I. Dubrovin, "Noncommutative valuation rings in simple finite-dimensional algebras over a field", *Mat. Sb.* (*N.S.*) **123**(165):4 (1984), 496–509. In Russian; translated in *Math. USSR-Sb.* **51**:2 (1985), 493–505. MR 85j:16020 Zbl 0573.16012

[Endler 1972] O. Endler, *Valuation theory*, Springer, New York, 1972. MR 50 #9847 Zbl 0257.12111

[Haile 1987] D. E. Haile, "Crossed-products orders over discrete valuation rings", *J. Algebra* **105**:1 (1987), 116–148. MR 88b:16013 Zbl 0609.16004

[Haile and Morandi 1993] D. E. Haile and P. J. Morandi, "On Dubrovin valuation rings in crossed product algebras", *Trans. Amer. Math. Soc.* **338**:2 (1993), 723–751. MR 93j:16010 Zbl 0798.16031

[Haile et al. 1983] D. E. Haile, R. G. Larson, and M. E. Sweedler, "A new invariant for $\mathbb{C}$ over $\mathbb{R}$: almost invertible cohomology theory and the classification of idempotent cohomology classes and algebras by partially ordered sets with a Galois group action", *Amer. J. Math.* **105**:3 (1983), 689–814. MR 84k:12014 Zbl 0523.13005

[Haile et al. 1995] D. E. Haile, P. J. Morandi, and A. R. Wadsworth, "Bézout orders and Henselization", *J. Algebra* **173**:2 (1995), 394–423. MR 96i:16025 Zbl 0826.16010

[Kauta 1997a] J. S. Kauta, "Integral semihereditary orders, extremality, and Henselization", *J. Algebra* **189**:2 (1997), 226–252. MR 98d:16030 Zbl 0871.16009

[Kauta 1997b] J. S. Kauta, "Integral semihereditary orders inside a Bézout maximal order", *J. Algebra* **189**:2 (1997), 253–272. MR 98d:16031 Zbl 0871.16010

[Kauta 2001] J. S. Kauta, "Crossed product orders over valuation rings", *Bull. London Math. Soc.* **33**:5 (2001), 520–526. MR 2002d:16023 Zbl 1029.16012

[Kauta 2012] J. S. Kauta, "On a class of hereditary crossed-product orders", 2012. To appear in *Proc. Amer. Math. Soc.* arXiv 1210.1065

[Morandi 1992] P. J. Morandi, "Maximal orders over valuation rings", *J. Algebra* **152**:2 (1992), 313–341. MR 93k:16028 Zbl 0781.16007

[Reiner 2003] I. Reiner, *Maximal orders*, London Mathematical Society Monographs. New Series **28**, Clarendon, Oxford, 2003. MR 2004c:16026 Zbl 1024.16008

Received October 5, 2011. Revised March 27, 2012.

JOHN S. KAUTA
DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
UNIVERSITI BRUNEI DARUSSALAM
BANDAR SERI BEGAWAN BE1410
BRUNEI DARUSSALAM

john.kauta@ubd.edu.bn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 259, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.259.361

AN EXPLICIT FORMULA FOR SPHERICAL CURVES WITH CONSTANT TORSION

DEMETRE KAZARAS AND IVAN STERLING

We give an explicit formula for all curves of constant torsion in the unit two-sphere. Our approach uses hypergeometric functions to solve relevant ordinary differential equations.

1. Introduction

The purpose of this article is to give an explicit formula for all curves of constant torsion τ in the unit two-sphere $S^2(1)$. These curves and their basic properties have been known since the 1890s, and some of these properties are discussed in the Appendix. Some example curves, computed with a standard ODE package, with $\tau = 0.1, 0.5, 1, 2$ are shown in Figure 1. Though their existence and some

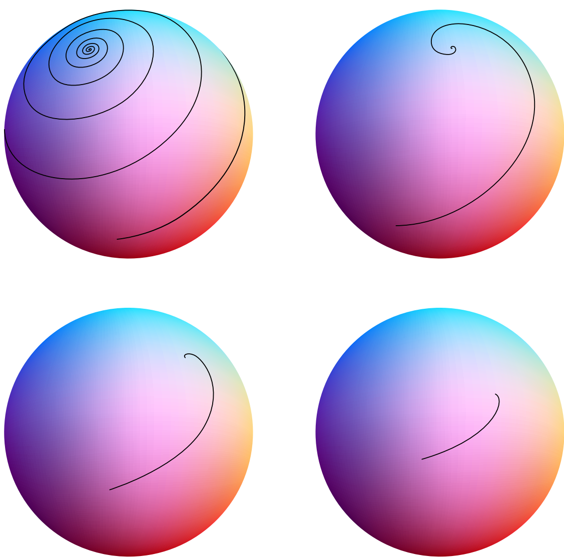


Figure 1. The curves of torsion $\tau = 0.1, 0.5, 1, 2$ on the unit sphere.

MSC2010: 53A04.
Keywords: constant torsion curves, spherical curves.

of their general properties were known, our explicit formulas for them, in terms of hypergeometric functions, are new.

Curves of constant torsion are also of interest because all asymptotic curves on pseudospherical surfaces (that is, surfaces in $\mathbb{R}^3$ with constant negative Gauss curvature) are of constant torsion. Furthermore, any pair of curves with constant torsion $\pm\tau$, intersecting at one point, define an essentially unique pseudospherical surface. A complete classification of curves of constant torsion in $\mathbb{R}^3$, in the context of integrable geometry, is a work in progress and is related to the corresponding unfinished classification of pseudospherical surfaces.

The authors would like to thank the referees and the editor for their suggestions to improve the original version of this paper.

2. General setting

2.1. ***General curves in*** $\mathbb{R}^3$**.** Let

$$\gamma : (a, b) \longrightarrow \mathbb{R}^3$$

be a regular (that is, nonzero speed) C^∞ curve in $\mathbb{R}^3$ with nonzero curvature. The speed v, curvature κ and torsion τ of γ are given by

$$v = \|\gamma'\|, \qquad \kappa = \frac{\|\gamma' \times \gamma''\|}{\|\gamma'\|^3}, \qquad \tau = \frac{[\gamma'\gamma''\gamma''']}{\|\gamma' \times \gamma''\|^2}.$$

The unit tangent T is given by

$$T = \frac{\gamma'}{v}. \tag{1}$$

The unit normal and unit binormal are given by

$$N = \frac{T'}{v\kappa} \quad \text{and} \quad B = T \times N.$$

These are related by the Frenet formulas

$$\begin{aligned} T' &= && v\kappa N && \\ N' &= -v\kappa T && && +v\tau B \\ B' &= && -v\tau N. && \end{aligned} \tag{2}$$

Curves γ with prescribed differentiable curvature $\kappa > 0$ and torsion τ can be found by integrating (2) and (1). Up to reparametrization (see below) a curve in $\mathbb{R}^3$ is determined, up to a rigid motion of $\mathbb{R}^3$, by its curvature κ and torsion τ.

2.2. ***Changing parametrizations.*** Given $\gamma(t)$, the arc length function $s(t)$ of $\gamma(t)$ is given by

$$s(t) = \int_a^t \|\gamma'(u)\|\, du.$$

Note that since $s(t)$ is increasing, it has an inverse $t(s)$. To obtain a unit-speed reparametrization γ^{unit} of γ we let

$$\gamma^{\text{unit}}(s) := \gamma(t(s)).$$

We denote the parameter of a unit-speed curve by the letter s.

On the other hand, if we are given a unit-speed curve $\gamma^{\text{unit}}(s)$, we may wish to find a reparametrization γ of γ^{unit} by letting $t(s)$ be some special monotone function. In that case we have

$$\gamma(t) := \gamma^{\text{unit}}(s(t)),$$

where $s(t)$ is the inverse of $t(s)$.

2.3. ***Spherical curves.*** We call γ a *spherical curve* if $\gamma(t) \in S^2(r)$ for all t (for fixed $r > 0$). One can show (see [Gray et al. 2006], for example) that the speed, curvature and torsion of a spherical curve satisfy

$$\kappa^2\tau^2(\kappa^2 r^2 - 1) = \kappa'^2 v^2. \tag{3}$$

2.4. ***Effect of homothety on curvature and torsion.*** If two curves γ, $\tilde{\gamma}$ are related by $\tilde{\gamma}(t) := \lambda\gamma(t)$, then $\tilde{\kappa}(t) = \kappa(t)/\lambda$ and $\tilde{\tau}(t) = \tau(t)/\lambda$. Thus, a curve of constant torsion τ_1 on a sphere of radius r_1 corresponds by homothety to a curve of constant torsion $\tau_2 = (r_1/r_2)\,\tau_1$ on a sphere of radius r_2.

In other words, any spherical curve of constant positive torsion corresponds to precisely one spherical curve with $\tau = 1$ as well as to precisely one curve of constant positive torsion on the unit sphere. Without loss of generality, we consider only spherical curves of constant positive torsion on the unit sphere.

2.5. ***Constant torsion unit-speed curves on the unit sphere.*** Let $r = 1$. If τ is a positive constant and $\gamma : (a, b) \to S^2(1)$ is of unit speed, then (3) is an ordinary differential equation in κ:

$$\kappa'^2 = \kappa^2\tau^2(\kappa^2 - 1). \tag{4}$$

(Notice that $\kappa \geq 1$ holds for any curve on the unit sphere.) The general solution to (4) is given by

$$\kappa = \csc(\tau s + C), \qquad \frac{-C}{\tau} < s < \frac{-C+\pi}{\tau}. \tag{5}$$

Notice we use the parameter s instead of t since γ is a unit-speed curve. Furthermore, $\kappa(s)$ is decreasing on

$$\left(\frac{-C}{\tau}, \frac{-C+\pi/2}{\tau}\right).$$

2.6. ***Our goal.*** As mentioned, a unit-speed curve γ is determined up to rigid motion by its curvature and torsion. However, in general it is not possible to explicitly solve for γ given $\kappa > 0$ and $\tau > 0$. Spherical curves of constant torsion provide an interesting and natural example to study. They were considered by classical geometers and the formula (5) was known. Even though the formula for κ is so simple, no explicit solutions for γ were found. This is most likely because the integration methods that we found necessary were not developed until decades later. By choosing a special reparametrization and using functions defined in the 1940s, we were successful in obtaining an explicit formula for γ involving hypergeometric functions.

3. Explicit formulas

3.1. ***The radius of curvature parametrization.*** In curve theory parametrization by the curvature is called the "natural parametrization". In our case, when the natural parametrization is used, the domain of definition lies outside the radius of convergence of the resulting hypergeometric solutions. To avoid having to deal with the problem of analytically continuing hypergeometric functions beyond their radii of convergence we instead parametrize by the reciprocal of the curvature, which is called the radius of curvature.

We seek unit-speed curves $\gamma^{\text{unit}} : (-C/\tau, (-C+\frac{\pi}{2})/\tau) \to S^2(1)$ of constant torsion $\tau > 0$ on the unit sphere. In order to simplify the Frenet equations, we reparametrize γ^{unit} by $t(s) = 1/\kappa(s) = \sin(\tau s + C)$. Since $1/\kappa(s)$ is increasing on its domain, the inverse $s(t)$, $s : (0, 1) \longrightarrow (-C/\tau, (-C+\frac{\pi}{2})/\tau)$, exists and we have

$$\gamma(t) = \gamma^{\text{unit}}(s(t)) = \gamma^{\text{unit}}\Big(\frac{\sin^{-1}(t) - C}{\tau}\Big), \quad 0 < t < 1.$$

One can recover γ^{unit} from γ by reversing the process. Note that

$$v = \|\gamma'\| = \|\gamma^{\text{unit}\prime}\| |s'(t)| = |s'(t)| = \frac{1}{\tau\sqrt{1-t^2}}.$$

With $\kappa = 1/t$ the Frenet equations (2) become

$$T' = vN/t \tag{6a}$$

$$N' = -vT/t + v\tau B \tag{6b}$$

$$B' = -v\tau N. \tag{6c}$$

Recall $\gamma' = vT$. Thus as a preliminary step we will compute T. Namely, we want to solve (6) for T.

From (6a) and (6b) we have

$$N = t\sqrt{1-t^2}\,\tau T', \tag{7}$$

$$B = \sqrt{1-t^2}\,N' + \frac{1}{t\tau}T. \tag{8}$$

Equations (7) and (6c) yield

$$B' = -\tau t T'. \tag{9}$$

On the other hand differentiating (7) yields

$$N' = \frac{\tau}{\sqrt{1-t^2}}\big(t(1-t^2)T'' + (1-2t^2)T'\big). \tag{10}$$

Plugging this into (8) yields

$$B = \frac{-1}{t\tau}\big(t^2(t^2-1)\tau^2 T'' + t(2t^2-1)\tau^2 T' - T\big). \tag{11}$$

Hence

$$B' = \frac{-1}{t^2\tau}\big(t^3(t^2-1)\tau^2 T''' + t^2(5t^2-2)\tau^2 T'' + t(4t^2\tau^2-1)T' + T\big). \tag{12}$$

Equating (9) and (12) and simplifying we arrive at

$$t^3(t^2-1)\tau^2 T''' + t^2(5t^2-2)\tau^2 T'' + t(3t^2\tau^2-1)T' + T = 0. \tag{13}$$

This is a third-order linear homogeneous differential equation with nonconstant coefficients. In general it is not possible to find a closed form solution for such an equation. However, this is one of the special cases where one can find hypergeometric type solutions. These methods were developed in the 1940s, and hence were not available to the classical (1890s) geometers.

3.2. ***Initial conditions.*** To arrive at initial conditions for our ODE, we find initial conditions for T, N, and B and use the Frenet equations (6) to arrive at initial conditions for T, T', and T''. We let $T = (T_1, T_2, T_3)$, $N = (N_1, N_2, N_3)$, and $B = (B_1, B_2, B_3)$. For $V = T, N, B, T'$, or T'' we use the notation $V_{i_0} := V_i(t_0)$.

T and N are unit vectors ($\|T\| = 1$ and $\|N\| = 1$) so we have

$$|T_{1_0}| \le 1, \quad |T_{2_0}| \le \sqrt{1-T_{1_0}}, \quad T_{3_0} = \sqrt{1-T_{1_0}^2 - T_{2_0}^2},$$
$$|N_{1_0}| \le 1, \quad |N_{2_0}| \le \sqrt{1-N_{1_0}}, \quad N_{3_0} = \sqrt{1-N_{1_0}^2 - N_{2_0}^2}.$$

Also T is orthogonal to N ($T \cdot N = 0$),

$$N_{1_0}T_{1_0} + N_{2_0}T_{2_0} + N_{3_0}T_{3_0} = 0.$$

Since $B = T \times N$, we have

$$B_{1_0} = T_{2_0} N_{3_0} - T_{3_0} N_{2_0}, \quad B_{2_0} = T_{3_0} N_{1_0} - T_{1_0} N_{3_0}, \quad B_{3_0} = T_{1_0} N_{2_0} - T_{2_0} N_{1_0}.$$

We will, without loss of generality and up to rigid motion, choose $N_0 = (0, 1, 0)$, $T_0 = (T_{1_0}, T_{2_0}, T_{3_0}) = (1, 0, 0)$, and $t_0 = \frac{1}{2}$. Now that we have initial conditions for T, N, and B, we will use the Frenet equations to express T_0' and T_0'' in terms of T_0, N_0, and B_0,

$$T_0' = \frac{v_0 N_0}{t_0}, \quad T_0'' = (v_0' t_0 + v_0) N_0 - v_0^2 t_0^2 T_0 + v_0^2 \tau t_0 B_0.$$

The set of initial conditions $t_0 = \frac{1}{2}$, $T_0 = (T_{1_0}, T_{2_0}, T_{3_0}) = (1, 0, 0)$, and $N_0 = (0, 1, 0)$ yields

$$T_0' = \left(0, \frac{4}{\sqrt{3}\tau}, 0\right) \quad \text{and} \quad T_0'' = \left(\frac{-16}{3\tau^2}, \frac{-16}{3\sqrt{3}\tau}, \frac{8}{3\tau}\right).$$

3.3. ***Solving for T via hypergeometric functions.*** The Barnes generalized hypergeometric function [Itō 1987] is defined by

$${}_pF_q(a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; t^a) := \sum_{n=1}^{\infty} \frac{(a_1)_n \ldots (a_p)_n}{(b_1)_n \ldots (b_q)_n} \frac{t^{an}}{n!}.$$

Note the use of the Pochhammer symbols $(x)_n := \Gamma(x+n)/\Gamma(x)$. We will also use

$${}_2F_1^{\text{reg}}(a, b, c, t^a) := \frac{{}_2F_1(a, b; c; t^a)}{\Gamma[c]}.$$

By direct substitution (see for example Section 46 of [Rainville 1971]) it is straightforward to check that the following is a solution to (13):

$$T = (T_1, T_2, T_3), \quad T_j = \sum_{l=1}^{3} c_{jl} S_l,$$

where

$$S_1 = it\, {}_3F_2\Big(\frac{1}{2}, \frac{1}{2}, \frac{3}{2}; \frac{3}{2} - \frac{i}{2\tau}, \frac{3}{2} + \frac{i}{2\tau}; t^2\Big),$$
$$S_2 = (-1)^{-i/(2\tau)} t^{-i/(2\tau)}\, {}_3F_2\Big(1 - \frac{i}{2\tau}, -\frac{i}{2\tau}, -\frac{i}{2\tau}; \frac{1}{2} - \frac{i}{2\tau}, 1 - \frac{i}{\tau}; t^2\Big),$$
$$S_3 = (-1)^{i/(2\tau)} t^{i/(2\tau)}\, {}_3F_2\Big(1 + \frac{i}{2\tau}, \frac{i}{2\tau}, \frac{i}{2\tau}; \frac{1}{2} + \frac{i}{2\tau}, 1 + \frac{i}{\tau}; t^2\Big),$$

the c_{jl} are constants, and $i = \sqrt{-1}$. Note that S_1 is pure imaginary and that S_3 is the complex conjugate of S_2. For proper complex constants c_{jl}, T is a real valued vector function. By plugging in the initial conditions of the last section we can solve for the c_{jl}.

3.4. ***Solving for γ.*** Recall that $\gamma(t) = \int vT\,dt$. Since we have found T in terms of hypergeometric functions, we must compute the following type of integrals:

$$\int \frac{h(t)\ {}_pF_q(a_1, a_2, \dots, a_p; b_1, b_2, \dots, b_q; , t^2)}{\tau\sqrt{1-t^2}}\,dt$$

$$:= \int \frac{h(t) \sum_{n=1}^{\infty} \frac{(a_1)_n \dots (a_p)_n}{(b_1)_n \dots (b_q)_n} \frac{t^{2n}}{n!}}{\tau\sqrt{1-t^2}}\,dt = \frac{\alpha}{\tau} \sum_{n=1}^{\infty} \frac{(a_1)_n \dots (a_p)_n}{(b_1)_n \dots (b_q)_n n!} \int \frac{t^{\beta n}}{\sqrt{1-t^2}}\,dt.$$

Where α, β are constants. We repeat this process for each S_l, using the notation $\gamma = (U_1, U_2, U_3)$. For S_1,

$$U_1 := \int \frac{S_1}{\tau\sqrt{1-t^2}}\,dt = \int \frac{it\ {}_3F_2\big(\frac{1}{2}, \frac{1}{2}, \frac{3}{2}; \frac{3}{2} - \frac{i}{2\tau}, \frac{3}{2} + \frac{i}{2\tau}; t^2\big)}{\tau\sqrt{1-t^2}}dt$$
$$= \sum_{n=0}^{\infty} d_{1n} \int \frac{t^{2n+1}}{\sqrt{1-t^2}}\,dt,$$

where

$$d_{1n} = \frac{i\big(1+\tau^2\big)\Gamma\big(\frac{1}{2}+n\big)^2\Gamma\big(\frac{3}{2}+n\big)\operatorname{sech}\big(\frac{\pi}{2\tau}\big)}{2\sqrt{\pi}\tau^3 n!\Gamma\big(\frac{3}{2}+n-\frac{i}{2\tau}\big)\Gamma\big(\frac{3}{2}+n+\frac{i}{2\tau}\big)}.$$

For S_2,

$$U_2 := \int \frac{S_2}{\tau\sqrt{1-t^2}}\,dt$$
$$= \int \frac{(-1)^{-i/(2\tau)} t^{-i/(2\tau)}\ {}_3F_2\big(1-\frac{i}{2\tau}, -\frac{i}{2\tau}, -\frac{i}{2\tau}; \frac{1}{2}-\frac{i}{2\tau}, 1-\frac{i}{\tau}; t^2\big)}{\tau\sqrt{1-t^2}}\,dt$$
$$= \sum_{n=0}^{\infty} d_{2n} \int \frac{t^{2n-i/\tau}}{\sqrt{1-t^2}}dt,$$

where

$$d_{2n} = \frac{e^{\frac{\pi}{2\tau}} 2^{-\frac{i}{\tau}} \Gamma\big(n-\frac{i}{2\tau}\big)^2 \Gamma\big(1+n-\frac{i}{2\tau}\big)\Gamma\big(\frac{-i+\tau}{2\tau}\big)^2}{\sqrt{\pi}\tau\Gamma\big(1+n\big)\Gamma\big(1+n-\frac{i}{\tau}\big)\Gamma\big(-\frac{i}{2\tau}\big)^2\Gamma\big(n+\frac{-i+\tau}{2\tau}\big)}.$$

For S_3,

$$U_3 := \int \frac{S_3}{\tau\sqrt{1-t^2}}\,dt$$
$$= \int \frac{(-1)^{i/(2\tau)} t^{i/(2\tau)}\ {}_3F_2\big(1+\frac{i}{2\tau}, \frac{i}{2\tau}, \frac{i}{2\tau}; \frac{1}{2}+\frac{i}{2\tau}, 1+\frac{i}{\tau}; t^2\big)}{\tau\sqrt{1-t^2}}\,dt$$
$$= \sum_{n=0}^{\infty} d_{3n} \int \frac{t^{2n+\frac{i}{\tau}}}{\sqrt{1-t^2}}\,dt,$$

where

$$d_{3n} = \frac{e^{-\frac{\pi}{2\tau}} 2^{\frac{i}{\tau}} \Gamma\left(n+\frac{i}{2\tau}\right)^2 \Gamma\left(1+n+\frac{i}{2\tau}\right) \Gamma\left(\frac{i+\tau}{2\tau}\right)^2}{\sqrt{\pi}\,\tau\,\Gamma\left(1+n\right)\Gamma\left(1+n+\frac{i}{\tau}\right)\Gamma\left(\frac{i}{2\tau}\right)^2\Gamma\left(n+\frac{i+\tau}{2\tau}\right)}.$$

Once again we are lucky and for each U_l we can evaluate the integrals. In each case they are hypergeometric,

$$U_1 = \sum_{n=0}^{\infty} \frac{1}{2} n!\, d_{1n}\, {}_2F_1^{\text{reg}}\left(\frac{1}{2}, 1+n, 2+n, t^2\right),$$

$$U_2 = \sum_{n=0}^{\infty} \frac{1}{2} \Gamma\left(n+\frac{-i+\tau}{2\tau}\right) d_{2n}\, {}_2F_1^{\text{reg}}\left(\frac{1}{2}, n+\frac{-i+\tau}{2\tau}, \frac{3}{2}+n-\frac{i}{2\tau}, t^2\right),$$

$$U_3 = \sum_{n=0}^{\infty} \frac{1}{2} \Gamma\left(n+\frac{i+\tau}{2\tau}\right) d_{3n}\, {}_2F_1^{\text{reg}}\left(\frac{1}{2}, n+\frac{i+\tau}{2\tau}, \frac{3}{2}+n+\frac{i}{2\tau}, t^2\right).$$

Each ${}_2F_1^{\text{reg}}$ also has a power series,

$${}_2F_1^{\text{reg}}\left(\frac{1}{2}, 1+n, 2+n, t^2\right) = \sum_{m=0}^{\infty} e_{1m} t^{2m},$$

$${}_2F_1^{\text{reg}}\left(\frac{1}{2}, n+\frac{-i+\tau}{2\tau}, \frac{3}{2}+n-\frac{i}{2\tau}, t^2\right) = \sum_{m=0}^{\infty} e_{2m} t^{2m},$$

$${}_2F_1^{\text{reg}}\left(\frac{1}{2}, n+\frac{i+\tau}{2\tau}, \frac{3}{2}+n+\frac{i}{2\tau}, t^2\right) = \sum_{m=0}^{\infty} e_{3m} t^{2m},$$

where

$$e_{1m} = \frac{\Gamma\left(\frac{1}{2}+m\right)}{(n+m+1)\sqrt{\pi}\,\Gamma(1+n)\Gamma(1+m)},$$

$$e_{2m} = \frac{2\tau\,\Gamma\left(\frac{1}{2}+m\right)}{\sqrt{\pi}\,\Gamma(2n\tau+2m\tau+\tau-i)\Gamma(1+m)\Gamma(n+(-i+\tau)/2\tau)},$$

$$e_{3m} = \frac{2\tau\,\Gamma\left(\frac{1}{2}+m\right)}{\sqrt{\pi}\,\Gamma(2n\tau+2m\tau+\tau+i)\Gamma(1+m)\Gamma(n+(i+\tau)/2\tau)}.$$

Thus

$$U_1 = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{1}{2} n!\, d_{1n}\, e_{1m}\, t^{2m+2n+2},$$

$$U_2 = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{1}{2} \Gamma\left(n+\frac{-i+\tau}{2\tau}\right) d_{2n}\, e_{2m}\, t^{2m+2n+2},$$

$$U_3 = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{1}{2} \Gamma\left(n+\frac{i+\tau}{2\tau}\right) d_{3n}\, e_{3m}\, t^{2m+2n+2}.$$

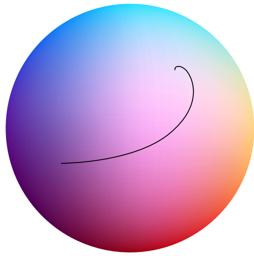


Figure 2. The curve of torsion $\tau = 1$ on the unit sphere.

These complicated double sums combine nicely and simplify as

$$U_1 = \frac{i}{2\sqrt{\pi}\tau} \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}+k\right)}{\Gamma(2+k)} \times {}_4F_3\left(\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, -k; \frac{1}{2}-k, \frac{3}{2}-\frac{i}{2\tau}, \frac{3}{2}+\frac{i}{2\tau}; 1\right) t^{2+2k},$$

$$U_2 = \frac{e^{\pi/(2\tau)}}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}+k\right)}{(-i+(1+2k)\tau)\Gamma(1+k)} \times {}_4F_3\left(-k, 1-\frac{i}{2\tau}, -\frac{i}{2\tau}, -\frac{i}{2\tau}; \frac{1}{2}-k, \frac{1}{2}-\frac{i}{2\tau}, 1-\frac{i}{\tau}; 1\right) t^{1-i/\tau+2k},$$

$$U_3 = \frac{e^{-\pi/(2\tau)}}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}+k\right)}{(i+(1+2k)\tau)\Gamma(1+k)} \times {}_4F_3\left(-k, 1+\frac{i}{2\tau}, \frac{i}{2\tau}, \frac{i}{2\tau}; \frac{1}{2}-k, \frac{1}{2}+\frac{i}{2\tau}, 1+\frac{i}{\tau}; 1\right) t^{1+i/\tau+2k}.$$

Thus we can write $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ as a power series in t, where

$$\gamma_j = \sum_{l=1}^{3} c_{jl} U_l.$$

The curve with $\tau = 1$ is given in Figure 2, this time using the explicit formula.

Appendix

The purpose of this appendix is to address questions and issues about the curves raised by the referees and the editor.

For $\tau = 0$, the curves are also planar and are precisely the set of circles lying on the sphere. If we consider curves corresponding to solutions (5) with $C = 0$, and $k(\frac{\pi}{2}/\tau) = \csc((\frac{\pi}{2}/\tau)\tau) = 1$, then as τ varies from 0 to ∞ the curves numerically appear to vary (in a nonuniform way) from an infinitely covered great circle, through a family of spiral "clothoid" like curves.

Figure 3. Curves of constant torsion approaching a circle.

The editor pointed out that if we consider those solutions to (5) with $C = 0$, and $k(s_0) = \csc(s_0\tau) > 1$, then the corresponding curves approach a “small circle” on $S^2(1)$ of constant curvature $k(s_0)$. This is an interesting example of nonuniform convergence. The curves as a whole converge pointwise to an infinitely covered great circle, while it is still possible to find sequences of “tails” that converge to infinitely covered small circles. This phenomenon is indicated in Figure 3, where one sees a sequence of curves converging to a small circle. More details of this simple yet interesting behavior will be written up elsewhere.

We will mention a few of the qualitative properties of these curves. Let us consider the case of curves in $S^2(1)$ with a fixed initial point and varying τ. All curves of constant torsion differ from one of these by a rigid motion. In [Cesàro 1926, page 185], it is shown that the curves are embedded, spiral infinitely often about a limiting endpoint, and are reflectionally symmetric through the initial point. (In Figure 1 we show only the upper half of the curves.) Cesàro [1926, page 185] also shows that as τ varies from 0 to ∞, the length varies from ∞ to 0; see also Equation (5).

One referee asked if it would be possible to foliate $S^2(1)$ with curves of constant torsion (other than by the just using circles). It may be possible to foliate $S^2(1)$, in some convoluted way, by packing $S^2(1)$ with pieces of curves of constant torsion; however our conjecture would be that it is not possible to foliate it in any “reasonable” way.

The reasoning is as follows. It seems to be a difficult problem to find an explicit formula for the upper endpoint in terms of the τ and the initial point. Nevertheless, numerically as τ varies from 0 to ∞ the upper endpoint steadily moves downward from the north pole to the initial point. In particular this would imply that the curves corresponding to an infinitesimal change in τ would (repeatedly) intersect. It would follow that any foliation of the $S^2(1)$ by curves of constant torsion would have to include curves with common endpoints that differ by a rigid motion; a rotation about the upper endpoint. This type of foliation could only work in some radius about the upper endpoint, because the effect of a rotation on the opposite lower endpoint would result in (repeated) intersections. In summary, the numerics strongly indicate that there is no foliation (singular or not) of $S^2(1)$ by curves of constant torsion.

Weiner [1977] proved that there exist arbitrarily short closed constant torsion curves in $\mathbb{R}^3$. More recently, Musso [2001] studied those curves of constant torsion in $\mathbb{R}^3$ whose normal vectors sweep out elastic curves in $S^2(1)$. Ivey [2000] generalizes Musso's results and gives examples of closed constant torsion curves of various knot types. The examples in the current paper complement these known examples.

References

[Cesàro 1926] E. Cesàro, *Vorlesungen über natülicher Geometrie*, 2nd ed., Teubner, Leipzig, 1926. JFM 52.0686.03

[Gray et al. 2006] A. Gray, E. Abbena, and S. Salamon, *Modern differential geometry of curves and surfaces with Mathematica*, 3rd ed., Chapman & Hall/CRC, Boca Raton, 2006. MR 2007d:53001 Zbl 0212.31802

[Itō 1987] K. Itō (editor), *Encyclopedic dictionary of mathematics*, 2nd ed., MIT Press, Cambridge, MA, 1987. MR 89b:00033 Zbl 0674.00025

[Ivey 2000] T. A. Ivey, "Minimal curves of constant torsion", *Proc. Amer. Math. Soc.* **128**:7 (2000), 2095–2103. MR 2001j:58021 Zbl 0954.53004

[Musso 2001] E. Musso, "Elastic curves and the Delaunay problem for curves with constant torsion", *Rend. Circ. Mat. Palermo* (2) **50**:2 (2001), 285–298. MR 2002k:58028 Zbl 1072.37043

[Rainville 1971] E. D. Rainville, *Special functions*, Chelsea Publishing, New York, 1971. MR 52 #14399 Zbl 0231.33001

[Weiner 1977] J. L. Weiner, "Closed curves of constant torsion, II", *Proc. Amer. Math. Soc.* **67**:2 (1977), 306–308. MR 57 #1370 Zbl 0382.53001

Editor's note

I was pleased to receive this interesting and unusual new work submitted to the *Pacific Journal of Mathematics*. It has also found referee approval for the main substance of the new material it offers. I am however uneasy with some assertions in the Appendix; to the best I can interpret them, I find myself in some disagreement. The authors have indicated that they prefer not to alter the statements, and have correctly pointed out that the impact of the appendix material on the main thrust of the paper is peripheral. In the view that differences over ancillary matters should not impede publication of new discoveries, I have recommended publication coupled with this subordinate comment.

Robert Finn

Received October 22, 2011. Revised July 3, 2012.

DEMETRE KAZARAS
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF OREGON
FENTON HALL
EUGENE, OR 97403-1222
UNITED STATES

dkazaras@gmail.com

IVAN STERLING
ST. MARY'S COLLEGE OF MARYLAND
18952 E. FISHER ROAD
ST. MARY'S CITY, MD 20686-3001
UNITED STATES

isterling@smcm.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 259, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.259.373

COMPARING SEMINORMS ON HOMOLOGY

JEAN-FRANÇOIS LAFONT AND CHRISTOPHE PITTET

We compare the l^1-seminorm $\|\cdot\|_1$ and the manifold seminorm $\|\cdot\|_{\mathrm{man}}$ on n-dimensional integral homology classes. Crowley and Löh showed that for any topological space X and any $\alpha \in H_n(X;\mathbb{Z})$, with $n \neq 3$, the equality $\|\alpha\|_{\mathrm{man}} = \|\alpha\|_1$ holds. We compute the simplicial volume of the 3-dimensional Tomei manifold and apply Gaifullin's desingularization to establish the existence of a constant $\delta_3 \approx 0.0115416$, with the property that for any X and any $\alpha \in H_3(X;\mathbb{Z})$, one has the inequality

$$\delta_3 \|\alpha\|_{\mathrm{man}} \le \|\alpha\|_1 \le \|\alpha\|_{\mathrm{man}}.$$

1. Introduction

Let X be a topological space and let K be either the field of rational numbers or the field of real numbers. Let $\alpha \in H_n(X, K)$ be a class in the n-dimensional singular homology of X with coefficients in K. By definition there is a finite linear combination of continuous maps $\sigma_i : \Delta \to X$ defined on the standard n-dimensional simplex, with coefficients a_i in K, which represents α. The l^1-*(semi)norm* on singular homology is defined as

$$\|\alpha\|_1 = \inf\Big\{\sum |a_i| : \Big[\sum a_i\sigma_i\Big] = \alpha\Big\};$$

see [Gromov 1982, 0.2].

If $\alpha \in H_n(X, \mathbb{Z})$ is an *integral* class, we may apply to it the natural change-of-coefficients morphism

$$H_*(X, \mathbb{Z}) \to H_*(X, \mathbb{R})$$

and view it as a *real* class (which may vanish) and consider its l^1-norm, also denoted $\|\alpha\|_1$. This measures the optimal "size" (in the l^1-norm) of a real representative

Lafont was supported by the NSF, under grants DMS-0906483, DMS-1207782, and by an Alfred P. Sloan Research Fellowship. Pittet was supported by a Research Membership in Quantitative Geometry offered by the Mathematical Sciences Research Institute. The authors would also like to thank the MRI for providing financial support for a collaborative visit by Pittet to the Ohio State University.
MSC2010: primary 53C23; secondary 57M50.
Keywords: l^1-norm, simplicial volume, singular homology, manifold norm, Steenrod's realization problem, Thurston norm, Tomei manifold.

for the integral class. When M is a closed oriented manifold, the l^1-norm of its fundamental class $[M] \in H_n(M;\mathbb{Z})$ is called the *simplicial volume* of M, and will be denoted by $\|M\|$.

Rather than looking at *all* chains representing the class α, one could instead restrict oneself to chains which satisfy some additional geometric constraint. To this end, let us consider the set of all closed smooth oriented manifolds and continuous maps $(M, f : M \to X)$ such that f sends the fundamental class of M to α. Recall [Thom 1954, Théorème III.9] that if $n \geq 7$, this set may be empty, even if X is a finite polyhedron. On integral homology, we consider the subadditive function

$$\mu(\alpha) = \inf\{\|M\| : f_*[M] = \alpha\},$$

(with the usual convention that the infimum of the empty set is $+\infty$) and the corresponding *manifold* (*semi*)*norm*

$$\|\alpha\|_{\text{man}} = \inf_{m\in\mathbb{N}} \left\{ \frac{\mu(m \cdot \alpha)}{m} \right\}.$$

Thom [1954, Théorème III.4] has shown that the manifold norm is finite when X is a finite polyhedron. Since any homology class can be represented as the image of a finite polyhedron, it follows from Thom's result that the manifold norm is finite for any topological space.

It is immediate from the definitions that $\|\cdot\|_1 \leq \|\cdot\|_{\text{man}}$ holds on $H_n(X,\mathbb{Z})$, for any n, and any topological space X.

Theorem 1.1. *For each degree n, there exists a constant $\delta_n > 0$, such that for any topological space X and any class $\alpha \in H_n(X,\mathbb{Z})$, we have*

$$\delta_n \|\alpha\|_{man} \leq \|\alpha\|_1 \leq \|\alpha\|_{man}.$$

One can take $\delta_n = 1$ if $n \neq 3$, and $\delta_3 \approx 0.0115416$.

After some preliminary material in Sections 2 and 3, we provide a proof of Theorem 1.1 in Sections 4 and 5. Section 4 shows the existence of the δ_n, whereas Section 5 is devoted to identifying the optimal values of the δ_n. It is straightforward to show that the norms are equal if $n \leq 2$ (that is, one can take $\delta_2 = 1$). Crowley and Löh [2012, Proposition 4.3] showed that for degree $n \geq 4$, one can take $\delta_n = 1$ (see Proposition 5.1 below). So in all cases except possibly in degree $= 3$, one actually has the equality $\|\alpha\|_1 = \|\alpha\|_{\text{man}}$. We do not know if the optimal value of δ_3 is 1.

Shortly after this paper was written, Gaifullin posted a preprint [2012a] containing some closely related results. In fact, our Theorem 1.1 can be deduced from the results in [Gaifullin 2012a, Section 6], though without an explicit estimate for δ_3.

2. Gluing simplices along their faces

Our first goal is to realize an integral class β as the image of a Δ-complex [Hatcher 2002, Section 2.1] which is a disjoint union of n-dimensional pseudomanifolds [Spanier 1981, Chapter 3, Example C] whose number of n-simplices is controlled in terms of β. The precise statement we need is the following.

Proposition 2.1. *Let X be a topological space and $\beta \in H_n(X, \mathbb{Z})$ an integral class on X of degree n represented by a singular cycle $\sum_i m_i \sigma_i$, $m_i \in \mathbb{Z}$. Then there is a Δ-complex Q and a continuous map $g : Q \to X$ with the following properties.*

(1) *The number of n-dimensional simplices of Q is $\sum_i |m_i|$.*

(2) *The Δ-complex Q is topologically a finite disjoint union of oriented n-dimensional pseudomanifolds without boundary.*

(3) *$g_*[Q] = \beta$, that is, with appropriate orientations on each pseudomanifold, g sends the sum of the fundamental classes of the pseudomanifolds forming Q to the class β.*

Remark 2.2. If $n \leq 2$, we can choose Q so that the pseudomanifolds are manifolds.

All this is well-known and can be deduced from [Hatcher 2002, Chapter 2]. We sketch the proof for the convenience of the reader.

Proof. The statement is trivial if $n = 0$, hence we assume $n \geq 1$. In the cycle $\sum_i m_i \sigma_i$, we consider each singular n-simplex σ_i whose coefficient m_i is negative. We precompose σ_i with an affine automorphism of the standard n-simplex that reverses the orientation and changes the sign of m_i. This leads to a representative of the same class β with positive coefficients $m_i \in \mathbb{N}$. Let us define

$$T = \sum_i m_i,$$

and let U be the disjoint union of T standard n-simplices. Repeating m_i times each singular simplex σ_i, we write our cycle

$$\sum_{i=1}^{T} \sigma_i$$

and we obtain a continuous map

$$\sigma : U \to X$$

whose restriction to the i-th copy of the standard n-simplex is σ_i. Each term of the boundary

$$\partial\Big(\sum_{i=1}^{T} \sigma_i\Big)$$

is the restriction of some σ_i to an $(n-1)$-face of the i-th n-simplex of U (times a coefficient which is either 1 or -1 because we repeat the terms). If two such singular $(n-1)$-simplices are equal (as maps defined on the standard $(n-1)$-simplex) and if their coefficients are opposite, they form what we call a canceling pair. We choose a maximal collection of canceling pairs, and for each pair we identify the two $(n-1)$-faces of U on which the two terms of the pair coincide. The topological space defined as the quotient of U with respect to the equivalence relation defined by these identifications has a Δ-complex structure Q with T n-simplices. It has no boundary because we chose a maximal family of canceling pairs and because $\sum_{i=1}^{T} \sigma_i$ is a cycle. This also implies that each connected component of Q is an n-dimensional oriented pseudomanifold. The map $\sigma : U \to X$ factors through Q. The quotient map $g : Q \to X$ is continuous and $g_*[Q] = \beta$. This proves the proposition.

If $n \leq 2$, one checks that each link of each vertex of Q is a sphere. This proves the remark. □

3. Gaifullin's desingularization

We need a result of Gaifullin, which provides a *constructive* desingularization of an oriented pseudomanifold (see [[2008]; 2012b] for a more detailed explanation). Let us briefly describe this result. Gaifullin establishes the existence, in each dimension n, of a closed oriented n-manifold M having the following universal property. Given any oriented n-dimensional pseudomanifold P with K top-dimensional simplices, and with a regular coloring of the vertex set by $(n+1)$ colors (that is, any adjacent vertices are of different colors), there exists

- a finite cover $\pi : \widehat{M} \to M$, of degree $\frac{1}{2} K \, \Pi_\omega \, |P_\omega|$,
- a map $f : \widehat{M} \to P$ with the property that

$$f_*[\widehat{M}] = 2^{n-1} \, \Pi_\omega \, |P_\omega| \cdot [P] \in H_n(P; \mathbb{Z}).$$

The degrees of the maps involve the integer $\Pi_\omega |P_\omega|$ (which is the product of the cardinalities of the finite sets P_ω), whose precise definition [Gaifullin 2008, page 563] we will not need. We merely point out that the term $\Pi_\omega |P_\omega|$ depends *solely* on the combinatorics of P, and appears in the expressions for *both* the degree of the covering map π, *and* of the "desingularization" map f.

The universal manifolds M are explicitly described, and are the *Tomei manifolds*. For the convenience of the reader, we provide some discussion of the Tomei manifolds in the Appendix, which also establishes some specific properties of the 3-dimensional Tomei manifold which are used in the proof of Proposition 5.2.

Finally, we make a brief comment concerning simplicial complexes versus Δ-complexes. The difference between these two classes is that, for Δ-complexes, one does not restrict the gluing of simplices to be along a single face of distinct simplices. While Gaifullin's result is stated in the setting where P is a simplicial complex, the constraint on the gluings of simplices is not used in his proofs. As such, his desingularization process works equally well when applied to Δ-complexes (assuming of course that there exists a regular vertex $(n+1)$-coloring). We thank the anonymous referee for pointing this out to us.

4. Existence of the δ_n

In this section, we show that there exist constants δ_n satisfying the conclusion of Theorem 1.1.

Let $\alpha \in H_n(X, \mathbb{Z})$ and let $\epsilon > 0$. The change-of-coefficients morphism

$$H_n(X, \mathbb{Z}) \to H_n(X, \mathbb{R})$$

factors through $H_n(X, \mathbb{Q})$, and the map

$$H_n(X, \mathbb{Q}) \to H_n(X, \mathbb{R})$$

is an isometric injection. Hence we can find a representative

$$\sum_i r_i \sigma_i$$

of α with $r_i \in \mathbb{Q}$ such that

$$\sum_i |r_i| \leq \|\alpha\|_1 + \epsilon. \tag{1}$$

Let m be the least common multiple of all the denominators of the reduced fractions of the r_i. The chain

$$\sum_i m r_i \sigma_i$$

is an integral chain representing the class

$$\beta = m\alpha \in H_n(X, \mathbb{Z}).$$

Now we apply Proposition 2.1 to the integral class β. This gives us a Δ-complex Q and a continuous map $g : Q \to X$ with the following properties:

(i) The number of n-dimensional simplices of Q is

$$m \sum_i |r_i| \leq m(\|\alpha\|_1 + \epsilon).$$

(ii) Q consists of a finite disjoint union of oriented n-dimensional pseudomanifolds without boundary.

(iii) g maps the sum of the fundamental classes of the pseudomanifolds in Q to the class β, that is, $g_*[Q] = \beta$.

Notice that in the case where Q is a manifold (that is automatic if $n = 2$, as explained at the end of the proof of Proposition 2.1), the inequality

$$\|\alpha\|_{\text{man}} \leq \|\alpha\|_1$$

follows, since for any $\epsilon > 0$ we have

$$\|Q\|/m \leq \|\alpha\|_1 + \epsilon.$$

If Q is not a manifold — that is, if at least one of the connected components of Q is not a manifold but only a pseudomanifold — a desingularization process is needed to produce a manifold. We first consider the case when Q is connected. Let P denote the first barycentric subdivision of the Δ-complex Q. The number of n-dimensional simplices of the barycentric division of the standard n-simplex is $(n+1)!$, so the number K of top-dimensional simplices in P is

$$K = (n+1)! m \sum_i |r_i|.$$

Moreover, the vertex set of P clearly has a regular coloring by $(n+1)$ colors: each vertex v lies in the interior of a unique cell σ_v from the original Δ-complex Q, and we can color the vertex v with the color $1 + \dim(\sigma_v) \in \{1, \ldots, n+1\}$. So we can now apply Gaifullin's desingularization process to the pseudomanifold P, obtaining the following diagram of spaces and maps:

$$M \xleftarrow{\pi} \widehat{M} \xrightarrow{f} P \xrightarrow{g} X .$$

We also know that

(a) $g_*[P] = \beta = m \cdot \alpha \in H_n(X; \mathbb{Z})$,

(b) $f_*[\widehat{M}] = 2^{n-1} \prod_\omega |P_\omega| \cdot [P] \in H_n(P; \mathbb{Z})$.

The map π is a covering map of degree $\frac{1}{2} K \prod_\omega |P_\omega|$, so we can also compute the simplicial volume of $\widehat{M}$:

$$\|\widehat{M}\| = \tfrac{1}{2} K \prod_\omega |P_\omega| \, \|M\|.$$

Combining (a) and (b), we see that the composite map $g \circ f : \widehat{M} \to X$ allows us to represent the homology class $[m \cdot 2^{n-1} \prod_\omega |P_\omega|] \cdot \alpha \in H_n(X; \mathbb{Z})$ as the image of the fundamental class of the oriented manifold $\widehat{M}$. From the definition of the manifold seminorm, we obtain

$$\|\alpha\|_{\mathrm{man}} \leq \frac{1}{m \cdot 2^{n-1}\,\Pi_\omega |P_\omega|} \|\widehat{M}\| = \frac{\frac{1}{2}\,K\,\Pi_\omega\,|P_\omega|}{m \cdot 2^{n-1}\,\Pi_\omega\,|P_\omega|} \|M\|$$
$$= \frac{(n+1)!\,m \sum_i |r_i|}{m \cdot 2^n} \|M\| \leq \|M\| \, \frac{(n+1)!}{2^n} \, (\|\alpha\| + \epsilon).$$

Letting ϵ go to zero completes the proof, with the explicit value

$$\delta_n = \frac{2^n}{(n+1)!\,\|M\|}$$

where M is the n-dimensional Tomei manifold appearing in Gaifullin's desingularization procedure. In the case where $P = \bigsqcup_i P_i$ has several connected components P_i, let d be the least common multiple of the $\Pi_\omega |(P_i)_\omega|$, and for each i, let $m_i = d/\Pi_\omega |(P_i)_\omega|$. Exactly the same proof applies with $\widehat{M} = \bigsqcup_i \bigsqcup_{m_i} \widehat{M}_i$, $f = \bigsqcup_i \bigsqcup_{m_i} f_i$, and $\pi = \bigsqcup_i \bigsqcup_{m_i} \pi_i$.

5. Estimating the δ_n

In this section, we complete the proof of Theorem 1.1 by estimating the δ_n. As explained in the previous section, one can take $\delta_2 = 1$. Crowley and Löh [2012] have shown that for $n \geq 4$, one can take $\delta_n = 1$. Their result is stated in the a priori more restrictive setting of finite CW-complexes, but it is straightforward to deduce the general case from that special case. For completeness, we include a proof of this result.

Proposition 5.1. *In degrees $n \geq 4$, we can take $\delta_n = 1$, that is, for any topological space X and any class $\alpha \in H_n(X, \mathbb{Z})$ of degree $n \geq 4$, one has the equality*

$$\|\alpha\|_1 = \|\alpha\|_{man}.$$

Proof. The inequality $\|\alpha\|_1 \leq \|\alpha\|_{\mathrm{man}}$ is immediate from the definitions, so let us focus on the converse. Proceeding as in the proof of Theorem 1.1, given any $\epsilon > 0$, we can find a corresponding *integral* chain

$$\sum_i m r_i \sigma_i$$

representing a class

$$\beta = m\alpha \in H_n(X, \mathbb{Z})$$

and where the rational numbers r_i satisfy

$$\sum_i |r_i| \leq \|\alpha\|_1 + \epsilon/2 \tag{2}$$

Now apply Proposition 2.1 to the integral class β, obtaining a Δ-complex Q and a continuous map $g : Q \to X$ such that $g_*[Q] = \beta$. As Q itself is a finite CW-complex of dimension $n \geq 4$, [Crowley and Löh 2012, Prop. 4.3] implies that $\|[Q]\|_1 = \|[Q]\|_{\text{man}}$. Since we have a realization of Q as a Δ-complex with exactly $m \sum_i |r_i|$ top-dimensional simplices, we obtain

$$\|[Q]\|_{\text{man}} = \|[Q]\|_1 \leq m \sum_i |r_i|.$$

Consider the positive real number $m\epsilon/2 > 0$. From the definition of the manifold norm, we can find a closed oriented manifold N, and a continuous map $h : N \to Q$ of degree d, with the property that $h_*[N] = d \cdot [Q]$, and satisfying

$$\frac{\|N\|}{d} \leq \|Q\|_{\text{man}} + m\epsilon/2 \leq m \sum_i |r_i| + m\epsilon/2. \tag{3}$$

The composite map $g \circ h : N \to X$ sends the fundamental class $[N]$ to $d \cdot \beta = d \cdot m\alpha$. Using this map to estimate the manifold norm of α, we obtain

$$\begin{aligned} \|\alpha\|_{\text{man}} &\leq \frac{\|N\|}{d\,m} \\ &\leq \frac{1}{m}\Big(m \sum_i |r_i| + m\epsilon/2\Big) \\ &\leq \sum_i |r_i| + \epsilon/2 \\ &\leq \|\alpha\|_1 + \epsilon, \end{aligned}$$

where the second inequality was deduced from (3), and the last inequality from (2). Finally, letting $\epsilon > 0$ go to zero, we obtain $\|\alpha\|_{\text{man}} \leq \|\alpha\|_1$, completing the proof. □

It is tempting to guess that the optimal value of δ_3 is also 1. Our method of proof gives a substantially lower value of δ_3, which is explicitly given by the following.

Proposition 5.2. *The optimal value of* δ_3 *is* $\geq V_3/(24V_8) \approx 0.0115416$, *where* V_3 *and* V_8 *are the volumes of the* 3*-dimensional regular ideal hyperbolic tetrahedron and octahedron, respectively.*

Proof. The proof of Theorem 1.1 yields the general value

$$\delta_n = \frac{2^n}{(n+1)!\,\|M\|}$$

where M is the n-dimensional Tomei manifold. Specializing to dimension $n = 3$, and using the fact that $\|M^3\| = 8V_8/V_3$ (see Lemma A.2 below), we obtain the claim. □

Appendix: Tomei manifolds

The universal manifolds M used in Gaifullin's desingularization are the *Tomei manifolds*. For the convenience of the reader, we provide a brief description of these manifolds. We also establish some results concerning the 3-dimensional Tomei manifold that are used in estimating the constant δ_3 arising in our proof of Theorem 1.1 (see Proposition 5.2).

A matrix $A = [a_{ij}]$ is *tridiagonal* if $a_{ij} = 0$ for all indices satisfying $|i - j| > 1$. The n-dimensional Tomei manifold consists of all $(n+1) \times (n+1)$ real symmetric tridiagonal matrices, with fixed simple spectrum $\lambda_0 < \lambda_1 < \cdots < \lambda_n$ (the manifold is independent of the choice of simple spectrum). These manifolds were introduced by Tomei [1984] and further studied by Davis [1987]. An important result of Tomei is that these manifolds support a very natural cellular decomposition, which we now describe.

First, recall the definition of the n-dimensional permutahedron Π^n. The permutahedron is an n-dimensional, simple, convex polytope, obtained as the convex hull of a specific configuration of points in $\mathbb{R}^{n+1}$. If the symmetric group S_{n+1} acts on $\mathbb{R}^{n+1}$ by permuting the coordinates, the permutahedron Π^n is defined to be the convex hull of the S_{n+1}-orbit of the point $(1, 2, \ldots, n+1) \in \mathbb{R}^{n+1}$. Denote by $\mathcal{S}$ this specific S_{n+1}-orbit, so that $\Pi^n = \mathrm{Conv}(\mathcal{S})$ (see Figure 1 for an illustration of Π^3).

The facets (codimension one faces) of the permutahedron Π^n are indexed by the $2^{n+1} - 2$ nonempty proper subsets $\omega \subsetneq \{1, \ldots, n+1\}$, as follows. Given a subset ω, define the subset $\mathcal{S}_\omega \subset \mathcal{S}$ by

$$\mathcal{S}_\omega := \{\vec{x} \in \mathcal{S} \mid \forall i \in \omega, \forall j \notin \omega,\ x_i < x_j\}.$$

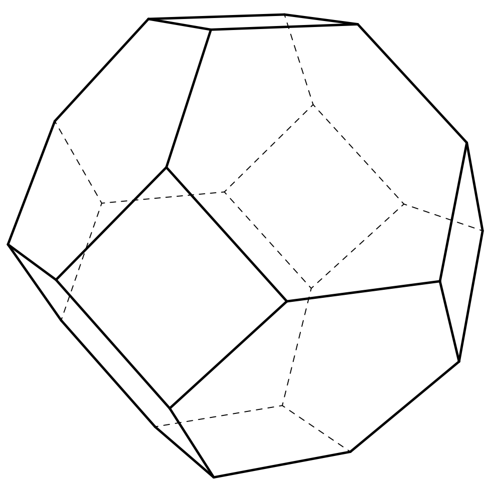


Figure 1. The 3-dimensional permutahedron Π^3.

In other words, a vertex $\vec{x} \in \mathcal{S}$ lies in $\mathcal{S}_\omega$ if the integers $\{1, \dots, |\omega|\}$ occur precisely in the coordinates whose index lies in ω. The facet F_ω is then defined to be the convex hull $\mathrm{Conv}(\mathcal{S}_\omega)$. From this, it easily follows that two distinct facets $F_{\omega_1}, F_{\omega_2}$ intersect if and only if $\omega_1 \subsetneq \omega_2$ or $\omega_2 \subsetneq \omega_1$. One also has that any codimension k face of Π^n, being of the form $F_{\omega_1} \cap \dots \cap F_{\omega_k}$ for some choice of distinct facets, corresponds (after possibly reindexing) to a unique length k chain $\omega_1 \subsetneq \omega_2 \subsetneq \dots \subsetneq \omega_k$ of nonempty proper subsets of $\{1, \dots, n+1\}$.

Tomei [1984] showed that the n-dimensional Tomei manifold M has a particularly simple tiling by 2^n copies of the n-dimensional permutahedron Π^n. Let $e_1, \dots, e_n$ be the standard generators for $\mathbb{Z}_2^n$. Then the n-dimensional Tomei manifold can be identified with $(\mathbb{Z}_2^n \times \Pi^n)/\sim$, where the equivalence relation is given by $(g, x) \sim (e_{|\omega|} g, x)$ whenever $x \in F_\omega$.

Example. For a concrete example, when $n = 3$, the permutahedron Π^3 is the truncated octahedron (see Figure 1). It has 6 square facets (parametrized by subsets $\omega \subsetneq \{1, 2, 3, 4\}$ with $|\omega| = 2$) and 8 hexagonal facets (parametrized by the ω with $|\omega| = 1, 3$). Figure 2 includes some vertex coordinates and labels some of the facets with the corresponding subset of $\{1, 2, 3, 4\}$.

In the corresponding Tomei manifold M^3, tessellated by eight copies of Π^3, one can easily see that each edge of the tessellation lies on exactly four copies of Π^3. Now consider the 24 squares appearing in the tessellation of M. The union of all these squares forms a collection of six tori embedded in M, each tessellated by four squares. Note that, from the definition of the gluings, each square bounds two copies of Π^3, whose indices in $\mathbb{Z}^3$ differ in the middle coordinate (corresponding to the generator e_2). This implies that the collection of six tori separate M^3 into two copies of a manifold N. Each of the two copies of N is tessellated by four copies of Π^3, and there is a $\mathbb{Z}_2$-involution on M^3 which fixes the collection of tori and interchanges the two copies of N. The involution can be easily described in terms of the description $M = (\mathbb{Z}_2^3 \times \Pi^3)/\sim$: it sends each element (g, x) to $(e_2 \cdot g, x)$.

A nice consequence of Gaifullin's work is the following elementary result.

Lemma A.1. *If M is a Tomei manifold*, $\|M\| > 0$.

Proof. Let N be a closed hyperbolic manifold of the same dimension as M. It follows from work of Gromov and Thurston that $\|N\| > 0$ (see [Thurston 1980, Chapter 6]). Take an arbitrary triangulation of N, pass to the barycentric subdivision, and apply Gaifullin's desingularization. This gives us a finite cover $\widehat{M} \to M$ with a map $f : \widehat{M} \to N$, of degree $d \neq 0$. Since $\|N\| > 0$, the obvious inequality $\|\widehat{M}\|/d \geq \|N\|$ immediately forces $\|\widehat{M}\| > 0$. But the simplicial volume scales under covering maps, so we conclude that $\|M\| > 0$, as desired. □

In general, the computation of the exact value of the simplicial volume is an extremely difficult problem. For the 3-dimensional Tomei manifold, we can, however, give an exact computation. Let V_8 denote the volume of a regular ideal hyperbolic octahedron and V_3 the volume of a regular ideal hyperbolic tetrahedron. These volumes can be expressed in terms of the Lobachevsky function

$$\Lambda(\theta) := -\int_0^\theta \log|2\sin t|\,dt$$

and are exactly equal to $V_8 = 8\Lambda(\pi/4)$ and $V_3 = 2\Lambda(\pi/6)$ (see [Thurston 1980, Section 7.2]). Up to five decimal places, $V_8 \approx 3.66386$ and $V_3 \approx 1.01494$.

Lemma A.2. *The* 3*-dimensional Tomei manifold* M^3 *has simplicial volume* $\|M\| = 8V_8/V_3$ (*which is* ≈ 28.8794).

Proof. Closed 3-manifolds are one of the few classes of manifolds for which the simplicial volume is known. Recall that for hyperbolic 3-manifolds, the simplicial volume is proportional to the hyperbolic volume, with constant of proportionality $1/V_3$. For Seifert fibered 3-manifolds, the existence of an S^1-action immediately implies that the simplicial volume is zero. For a general closed, orientable 3-manifold, the validity of Thurston's geometrization conjecture (recently established

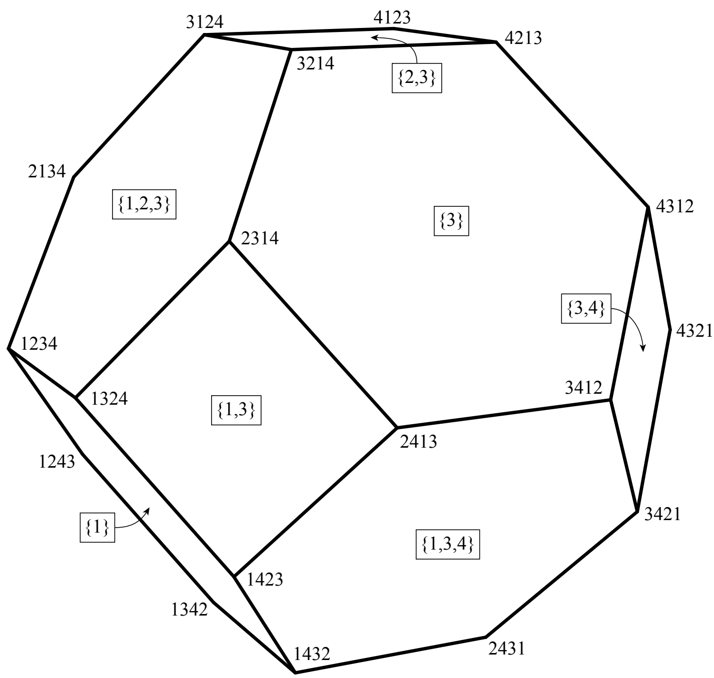


Figure 2. A portion of Π^3. Vertices are labeled by their coordinates in $\mathbb{R}^4$ (parentheses and commas omitted to avoid cluttering the picture). Facets are labeled with the corresponding subset $\omega \subset \{1, 2, 3, 4\}$.

by Perelman) implies that there is a decomposition into geometric pieces. Since simplicial volume is additive under connected sums (in dimensions ≥ 3) and under gluings along tori (see [Gromov 1982, Section 3.5]), this implies that the simplicial volume of any closed, orientable 3-manifold is proportional (with constant $1/V_3$) to the sum of the (hyperbolic) volumes of the hyperbolic pieces in its geometric decomposition.

Let us apply this procedure to the Tomei manifold M. Recall that M is the double of a 3-manifold N with ∂N consisting of four tori. From the gluing formula we deduce that $\|M\| = 2\|N\|$. To compute $\|N\|$, recall that N is tessellated by four copies of the 3-dimensional permutahedron Π^3, with the collection of square faces of all the Π^3 forming the boundary tori of N. This implies that the interior of N is tessellated by copies of Π^3 *with the square boundary faces removed*. Next we claim that $\mathrm{Int}(N)$ supports a finite volume hyperbolic metric.

Under this tessellation, each interior edge of N lies on exactly *four* of the Π^3. Let $\mathbb{O} \subset \mathbb{H}^3$ denote the regular ideal hyperbolic octahedron. This octahedron has all six vertices on the boundary at infinity of $\mathbb{H}^3$, and has all incident pairs of faces forming angles of $\pi/2$. A copy of the permutahedron Π^3 can be obtained by removing small horoball neighborhoods of each of the ideal vertices. Each hexagonal face of Π^3 corresponds to a triangular face of $\mathbb{O}$. So one can form a manifold N^0 by gluing together four copies of $\mathbb{O}$, using the same gluing pattern as in the formation of N. Using isometries to glue together the sides of $\mathbb{O}$, one obtains a metric on N^0 which is hyperbolic, except possibly along the 1-skeleton of N^0. To check whether or not one has a singularity along the edges of N^0, one just needs to calculate the total angle transverse to the edge. But recall that along each edge in N^0, one has four copies of $\mathbb{O}$ coming together. Since each edge in $\mathbb{O}$ has an internal angle of $\pi/2$, the total angle transverse to each edge of N^0 is equal to 2π. We conclude that N^0 supports a complete hyperbolic metric, with hyperbolic volume $= 4V_8$.

N is obtained from N^0 by removing a neighborhood of the ideal vertices in each $\mathbb{O}$ in the tessellation of N^0. This means that N is obtained from the noncompact, finite volume, hyperbolic manifold N^0 by truncating the cusps. It follows that $\mathrm{Int}(N)$ is diffeomorphic to N^0. Since cutting M open along the collection of tori results in two copies of $\mathrm{Int}(N) = N^0$, a manifold supporting a hyperbolic metric, we have that this is exactly the geometric decomposition of M predicted by Thurston's geometrization conjecture (cf. [Davis 1987, page 105, footnote 2]). Our discussion above implies that $\|M\| = 2\,\mathrm{Vol}(N^0)/V_3 = 8V_8/V_3$, completing the proof. $\square$

Acknowledgments

The authors thank Mike Davis for helpful discussions on the topology of Tomei manifolds, Allen Hatcher for explanations about Δ-complexes, Clara Löh for useful comments on terminology, and Guido Mislin for his interest in this work. Finally, the authors are indebted to the anonymous referee for several valuable suggestions.

References

[Crowley and Löh 2012] D. Crowley and C. Löh, "Functorial semi-norms on singular homology and (in)flexible manifolds", preprint, 2012. arXiv 1103.4139

[Davis 1987] M. W. Davis, "Some aspherical manifolds", *Duke Math. J.* **55**:1 (1987), 105–139. MR 88j:57044 Zbl 0631.57019

[Gaifullin 2008] A. A. Gaifullin, "Realisation of cycles by aspherical manifolds", *Uspekhi Mat. Nauk* **63**:3 (2008), 157–158. In Russian; translated in *Russian Math. Surveys* **63**:3 (2008), 562–564. MR 2009k:57051 Zbl 1175.55002

[Gaifullin 2012a] A. A. Gaifullin, "Combinatorial realisation of cycles and small covers", preprint, 2012. arXiv 1204.0208

[Gaifullin 2012b] A. A. Gaifullin, "Universal realisators for homology classes", preprint, 2012. arXiv 1201.4823

[Gromov 1982] M. Gromov, "Volume and bounded cohomology", *Inst. Hautes Études Sci. Publ. Math.* **56** (1982), 5–99. MR 84h:53053 Zbl 0516.53046

[Hatcher 2002] A. Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, 2002. MR 2002k:55001 Zbl 1044.55001

[Spanier 1981] E. H. Spanier, *Algebraic topology*, Springer, New York, 1981. MR 83i:55001 Zbl 0477.55001

[Thom 1954] R. Thom, "Quelques propriétés globales des variétés différentiables", *Comment. Math. Helv.* **28** (1954), 17–86. MR 15,890a Zbl 0057.15502

[Thurston 1980] W. P. Thurston, "The geometry and topology of three-manifolds", preprint, Princeton Univesity, 1980, Available at http://library.msri.org/books/gt3m.

[Tomei 1984] C. Tomei, "The topology of isospectral manifolds of tridiagonal matrices", *Duke Math. J.* **51**:4 (1984), 981–996. MR 86d:58091 Zbl 0558.57006

Received March 28, 2012. Revised August 2, 2012.

JEAN-FRANÇOIS LAFONT
DEPARTMENT OF MATHEMATICS
THE OHIO STATE UNIVERSITY
100 MATH TOWER
231 WEST 18TH AVENUE
COLUMBUS, OHIO 43210-1174
UNITED STATES
jlafont@math.ohio-state.edu

CHRISTOPHE PITTET
CENTRE DE MATHÉMATIQUES ET INFORMATIQUE
AIX-MARSEILLE UNIVERSITÉ
39 RUE FRÉDÉRIC JOLIOT-CURIE
13453 MARSEILLE CEDEX 13
FRANCE
pittet@cmi.univ-mrs.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 259, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.259.387

RELATIVELY MAXIMUM VOLUME RIGIDITY IN ALEXANDROV GEOMETRY

NAN LI AND XIAOCHUN RONG

Given compact metric spaces X and Z with Hausdorff dimension n, if there is a distance-nonincreasing onto map $f : Z \to X$, then the Hausdorff n-volumes satisfy $\mathrm{vol}(X) \le \mathrm{vol}(Z)$. The *relatively maximum volume conjecture* says that if X and Z are both Alexandrov spaces and $\mathrm{vol}(X) = \mathrm{vol}(Z)$, X is isometric to a gluing space produced from Z along its boundary ∂Z and f is length-preserving. We partially verify this conjecture and give a further classification for compact Alexandrov n-spaces with relatively maximum volume in terms of a fixed radius and space of directions. We also give an elementary proof for a pointed version of the Bishop–Gromov relative volume comparison with rigidity in Alexandrov geometry.

Introduction

Let Z be a compact metric space with Hausdorff dimension α. Consider all compact metric spaces X with Hausdorff dimension α such that there is a distance-nonincreasing onto map $f : Z \to X$. We let "vol" denote the Hausdorff measure (or volume) in the top dimension. Then $\mathrm{vol}\, X \le \mathrm{vol}\, Z$. A natural question is to determine X (in terms of Z) when $\mathrm{vol}\, X = \mathrm{vol}\, Z$. We refer to this as a *relatively maximum volume rigidity problem*.

A possible answer to the relatively maximum volume rigidity problem is closely related to the regularity of underlying geometric and topological structures. For instance, if Z and X are closed Riemannian n-manifolds, f is an isometry (see Corollary 0.2). On the other hand, taking any measure-zero subset S in Z (a Riemannian manifold) and identifying S with a point $p \in S$, the projection map, $Z \to X = Z/(S \sim p)$, is a distance-nonincreasing onto map, and it is hopeless to have some rigidity on Y in terms of X.

In this paper, we will study the relatively maximum volume rigidity problem in Alexandrov geometry, partly because an Alexandrov space X has a "right" geometric structure for this problem (see Conjecture 0.1 below). For instance, for

Both authors are partially supported by NSF grant DMS 0805928 and by a research fund from Capital Normal University.
MSC2010: 53C21, 53C23, 53C24.
Keywords: volume, radius, Alexandrov space, rigidity, stability .

$p \in X$, the gradient-exponential map, $g\exp_p : T_pX \to X$, becomes a distance-nonincreasing map, when T_pX is equipped with the κ-cone metric via the cosine law on the space form S^2_κ; see [Burago et al. 1992]. When taking Z to be a closed r-ball at the vertex (for $\kappa > 0$, $r \le \pi/(2\sqrt{\kappa})$ or $r = \pi/\sqrt{\kappa}$), the relatively volume rigidity problem (see Theorem B) indeed extends the (absolutely) maximum radius-volume rigidity theorem proved in [Grove and Petersen 1992]; see Theorem 0.3.

The recent study of Alexandrov spaces was initiated by Burago, Gromov, and Perelman [Burago et al. 1992] and has gotten a lot of attention lately. An Alexandrov space with curvature curv $\ge \kappa$ is a length metric space such that each point has a neighborhood in which the Toponogov triangle comparison holds with respect to the space form of constant curvature κ. In the rest of the paper, we will freely use basic notions on an Alexandrov space from [Burago et al. 1992] and [Petrunin 2007]; for example, the space of directions, the gradient-exponential maps, and (n, δ)-strained points, among others. Let $\mathrm{Alex}^n(\kappa)$ denote the collection of compact Alexandrov n-spaces with curv $\ge \kappa$.

Note that the boundary gluing will automatically yield a distance-nonincreasing onto (projection) map, which also preserves the volume (see Examples 2.14 and 2.15). We propose the following relatively maximum volume rigidity conjecture for Alexandrov spaces.

Conjecture 0.1. *Consider $Z, X \in \mathrm{Alex}^n(\kappa)$, and let $f : Z \to X$ be a distance-nonincreasing onto map. If $\mathrm{vol}\, Z = \mathrm{vol}\, X$, X is isometric to a gluing space produced from Z along its boundary ∂Z and f is length-preserving. In particular, Z is isometric to X if $\partial Z = \varnothing$ or if f is injective.*

Our goal in this paper is to partially verify Conjecture 0.1 and give a classification for the boundary gluing maps in a special case (see Theorem A, Corollary 0.2, and Theorem B).

We now begin to state the main results. Throughout this paper, $\tau(\delta)$ denotes a function in δ such that $\tau(\delta) \to 0$ as $\delta \to 0$. Our first result verifies Conjecture 0.1 for the case where f preserves non-(n, δ)-strained points up to an error $\tau(\delta)$. For $X \in \mathrm{Alex}^n(\kappa)$ and $\delta > 0$, let $X^\delta \subseteq X$ denote the set of all (n, δ)-strained points. Then a small ball centered at an (n, δ)-strained point is almost isometric to an open subset in $\mathbb{R}^n$ [Burago et al. 1992].

Theorem A. *Let Z, X be Alexandrov n-spaces (not necessarily complete) with curvature* curv $\ge \kappa$ *and* $\mathrm{vol}\, Z = \mathrm{vol}\, X$. *Suppose that $f : Z \to X$ is a distance-nonincreasing onto map such that for any $\delta > 0$, $f^{-1}(X^\delta) \subseteq Z^{\tau(\delta)}$. Then f is an isometry.*

A point z in Z is called *regular* if the space of directions Σ_x is isometric to a unit sphere. Clearly, the space Z with all points regular is a topological manifold, but

Z may not be isometric to any Riemannian manifold (for example, the doubling of two flat disks). Theorem A includes the following case:

Corollary 0.2. *Let* $Z, X \in \mathrm{Alex}^n(\kappa)$ *with* $\mathrm{vol}\, Z = \mathrm{vol}\, X$ *and all points in* Z *regular* (*for example*, Z *is a Riemannian manifold*). *If* $f : Z \to X$ *is a distance-nonincreasing onto map*, f *is an isometry.*

In Alexandrov geometry, perhaps the most natural distance-nonincreasing onto map is the gradient-exponential map $g\exp_p : C_\kappa(\Sigma_p) \to X$, $p \in X \in \mathrm{Alex}^n(\kappa)$, where $C_\kappa(\Sigma_p)$ denotes the tangent cone T_pX equipped with a κ-cone metric via the cosine law in S^2_κ [Burago et al. 1992]. Since $g\exp_p$ is distance-nonincreasing and preserves any r-ball, we immediately get the pointed version of the Bishop type volume comparison:

$$\mathrm{vol}\, B_R(p) \le \mathrm{vol}\, C^R_\kappa(\Sigma_p),$$

where $C^R_\kappa(\Sigma_p)$ denotes the *open* R-ball in $C_\kappa(\Sigma_p)$ at the vertex $\tilde{o}$. We show that when the equality holds, $g\exp_p$ will satisfy the conditions in Theorem A (Lemmas 2.4 and 2.5) and thus open ball $C^R_\kappa(\Sigma_p)$ is isometric to $B_R(p)$ with respect to intrinsic metrics (see Theorem 2.1).

We prove an important case of Conjecture 0.1, which gives a classification of Alexandrov spaces with relatively maximum volume: given any κ, $R > 0$ and $\Sigma \in \mathrm{Alex}^{n-1}(1)$, let $\mathcal{A}^R_\kappa(\Sigma)$ be the collection of Alexandrov n-spaces $X \ni p$ satisfying

$$\mathrm{curv} \ge \kappa, \quad X = \bar{B}_R(p), \quad \Sigma_p = \Sigma.$$

Then $\mathrm{vol}\, X \le \mathrm{vol}\, C^R_\kappa(\Sigma) = v(\Sigma, \kappa, R)$. When $\mathrm{vol}\, X = v(\Sigma, \kappa, R)$, we say that X has the relatively maximum volume.

Theorem B (relatively maximum volume rigidity). *Let* $X \in \mathcal{A}^R_\kappa(\Sigma)$ *such that* $\mathrm{vol}\, X = v(\Sigma, \kappa, R)$. *Then* X *is isometric to* $\overline{C}^R_\kappa(\Sigma)/x \sim \phi(x)$ *and* $R \le \pi/(2\sqrt{\kappa})$ *or* $R = \pi/\sqrt{\kappa}$ *for* $\kappa > 0$, *where* $\phi : \Sigma \times \{R\} \to \Sigma \times \{R\}$ *is an isometric involution* (*which can be trivial*). *Conversely*, *given any isometric involution* ϕ *on* Σ, $\overline{C}^R_\kappa(\Sigma)/x \sim \phi(x) \in \mathcal{A}^R_\kappa(\Sigma)$ *and has the relatively maximum volume.*

Theorem B verifies Conjecture 0.1 for the case $f = g\exp_p : Z = \overline{C}^R_\kappa(\Sigma_p) \to X$, together with a further classification for the boundary identification. Note that Theorem B implies that if $k > 0$ and $\pi/(2\sqrt{\kappa}) < R < \pi/\sqrt{\kappa}$,

$$\max\{\mathrm{vol}\, X,\ X \in \mathcal{A}^R_\kappa(\Sigma)\} < v(\Sigma, \kappa, R).$$

For the case where X is a limit of Riemannian manifolds, a classification was given in [Grove and Petersen 1992]. A general classification is more complicated, and we wish to discuss it elsewhere.

As mentioned earlier, Theorem B extends this radius-volume rigidity theorem:

Theorem 0.3 [Grove and Petersen 1992]. *Let $M_i \xrightarrow{d_{GH}} X$ be a Gromov–Hausdorff convergent sequence of Riemannian n-manifolds such that*

$$\sec_{M_i} \geq \kappa, \quad \operatorname{rad}(M_i) = R, \quad \operatorname{vol} M_i \to \operatorname{vol} C_\kappa^R(S_1^{n-1}),$$

where $\operatorname{rad}(M_i) = \min\{r,\ \bar{B}_r(p) = M_i,\ p \in M_i\}$. *Then $R \leq \pi/(2\sqrt{\kappa})$ or $R = \pi/\sqrt{\kappa}$ for* $\kappa > 0$, *and X is isometric to the quotient of $\bar{C}_\kappa^R(S_1^{n-1})$ by the equivalence relation $x \sim \phi(x)$, where $\phi : \partial \bar{C}_\kappa^R(S_1^{n-1}) \to \partial \bar{C}_\kappa^R(S_1^{n-1})$ is either the antipodal map or a reflection in a totally geodesic hypersurface. Moreover, each M_i is homeomorphic to an n-sphere or a real projective n-space.*

Note that $\operatorname{vol} X = \operatorname{vol} C_\kappa^R(S_1^{n-1})$. Choosing $p_i \in M_i$ such that $M_i = \bar{B}_R(p_i)$, $p_i \to p \in X$ and $\Sigma_p = S_1^{n-1}$. By now Theorem B implies the rigidity part of Theorem 0.3 (a generalization of the homeomorphic rigidity in Theorem 0.3 will be given in Theorem C). Theorem B also implies the following extension of Theorem 0.3.

Theorem 0.4 [Shteingold 1994]. *Let $X \in \mathcal{A}_\kappa^r(S_1^{n-1})$ with* $\operatorname{vol} X = v(S_1^{n-1}, \kappa, r)$. *Then $X = \bar{C}_\kappa^r(S_1^{n-1})/x \sim \phi(x)$, $x \in S_1^{n-1} \times \{r\}$, where ϕ is the reflection on an ℓ-dimensional totally geodesic subsphere,* $1 \leq \ell \leq n$ *(ϕ is trivial for $\ell = n$.)*

A further problem concerning Theorem B is to determine the homeomorphic type of X. We have solved this problem for X being a topological manifold (see Theorem 0.3).

Theorem C. *Given* $\Sigma \in \operatorname{Alex}^{n-1}(1)$, *$\kappa$ and* $R > 0$, *there exists a constant $\epsilon = \epsilon(\Sigma, \kappa, R) > 0$ such that if $X \in \mathcal{A}_\kappa^R(\Sigma)$ with* $\operatorname{vol} X > v(\Sigma, \kappa, R) - \epsilon$ *and X is a closed topological manifold, X is homeomorphic to S_1^n or a real projective space $\mathbb{R}P^n$.*

Note that Σ in Theorem C is not necessarily a topological manifold; for instance, $X = C_1(C_1(N))$, the twice spherical suspensions over a Poincaré sphere N, satisfies Theorem C, but $\Sigma = C_1(N)$ is not a topological manifold. However, X is homeomorphic to a 5-sphere, by [Kapovitch 2002].

In the proof of Theorem B, we establish a pointed version of the Bishop volume comparison with rigidity (Theorem 2.1). In general, we will prove the following pointed version of the Bishop–Gromov relative volume comparison with rigidity.

For $p \in X \in \operatorname{Alex}^n(\kappa)$, let $A_R^r(p)$ denote the annulus $\{x \in X : r < |px| < R\}$, $0 \leq r < R$, and let $A_R^r(\Sigma_p)$ denote the corresponding annulus in $C_\kappa(\Sigma_p)$.

Theorem D (pointed Bishop–Gromov relative volume comparison). *Let*

$$X \in \operatorname{Alex}^n(\kappa).$$

Then, for any $p \in X$ and $R_3 > R_2 > R_1 \geq 0$,

$$\frac{\operatorname{vol} A_{R_3}^{R_1}(p)}{\operatorname{vol} A_{R_3}^{R_2}(p)} \geq \frac{\operatorname{vol} A_{R_3}^{R_1}(\Sigma_p)}{\operatorname{vol} A_{R_3}^{R_2}(\Sigma_p)}, \quad \textit{or equivalently}, \quad \frac{\operatorname{vol} A_{R_2}^{R_1}(p)}{\operatorname{vol} A_{R_3}^{R_2}(p)} \geq \frac{\operatorname{vol} A_{R_2}^{R_1}(\Sigma_p)}{\operatorname{vol} A_{R_3}^{R_2}(\Sigma_p)}.$$

In particular,

$$\frac{\operatorname{vol} B_{R_1}(p)}{\operatorname{vol} B_{R_3}(p)} \geq \frac{\operatorname{vol} C_\kappa^{R_1}(\Sigma_p)}{\operatorname{vol} C_\kappa^{R_3}(\Sigma_p)}.$$

If any of these inequalities becomes an equality, the open ball $B_{R_3}(p)$ is isometric to $C_\kappa^{R_3}(\Sigma_p)$ with respect to the intrinsic metrics.

Remark 0.5. The Riemannian version of the Bishop–Gromov relative comparison for Alexandrov spaces (that is, the model space is S_κ^n) was stated in [Burago et al. 1992]; compare [Burago et al. 2001]. A notable difference between Theorem D and the Riemannian version is in the rigidity part: the latter is the *absolute maximum* volume rigidity and its model space is *unique*, while the former may be viewed as the *relatively maximum* volume rigidity (relatively to Σ_p), whose model spaces are of *infinitely many* possibilities. Moreover, the proof of Theorem D is considerably difficult; for instance, a dimension-inductive argument (which works in the Riemannian version) does not work.

Remark 0.6. By Lemma 2.1 in [Li 2010], we see that

$$\frac{\operatorname{vol} C_\kappa^R(\Sigma_p)}{\operatorname{vol} C_\kappa^r(\Sigma_p)} = \frac{\operatorname{vol} B_R(S_\kappa^n)}{\operatorname{vol} B_r(S_\kappa^n)},$$

and thus the monotonicity part of Theorem D coincides with that in the Riemannian version. We point out that our proof of the volume ratio monotonicity in Theorem D is different from one suggested in [Burago et al. 1992]; we take an elementary (calculus) approach via finding an (unconventional) partition suitable for triangle comparison arguments, while a proof in [Burago et al. 2001] relies on a coarea formula for Alexandrov spaces. We point out that in the case where $\kappa \leq 0$, a weak form of the above monotonicity was previously obtained in [Liu and Shen 1994, Proposition 1].

We now give some indication on our approach to Theorem A and Theorem B. In the proof of Theorem A, we show that f is a homeomorphism and f preserves the length of curves. Based on basic properties of an Alexandrov space (not necessarily complete), any curve c in X can be approximated by piecewise geodesics c_i in X^{δ_i} ($\delta_i \to 0$) such that lengths $L(c_i) \to L(c)$. Thus, it suffices to show that when restricting to $f^{-1}(X^\delta)$ and X^δ, respectively, f is injective and f^{-1} preserves the length of any geodesic up to an error $\tau(\delta) \to 0$ as $\delta \to 0$, respectively. We derive this with a volume formula for tube-like ϵ-balls in X^δ, which can be treated as a replacement of the volume formula of a thin tube around a curve. The proof of the volume formula is based on the fact that a small ball at an (n, δ)-strained point can be almost isometrically embedded into $\mathbb{R}^n$; see [Burago et al. 1992].

Our approach to Theorem B consists of two steps: first, establishing the open ball rigidity: the gradient-exponential map $g\exp_p : C_\kappa^R(\Sigma_p) \to B_R(p) \subset X$ is an isometry with respect to the intrinsic distance. We achieve this by showing that $g\exp_p$ satisfies the condition in Theorem A; see Lemmas 2.4 and 2.5. Consequently, $X = \bar{C}_\kappa^R(\Sigma_p)/\sim$, where $\sim$ is a relation on $\Sigma_p \times \{R\}$: $\tilde{x} \sim \tilde{y}$ if and only if $g\exp_p(\tilde{x}) = g\exp_p(\tilde{y})$. Observe that if $\tilde{x} \neq \tilde{y} \in \Sigma_p \times \{R\}$ with $\tilde{x} \sim \tilde{y}$, then the $g\exp_p$-images of the two geodesics $[\tilde{o}\tilde{x}]$ and $[\tilde{o}\tilde{y}]$ together form a local geodesic at $g\exp_p \tilde{x} = g\exp_p \tilde{y}$. Because a geodesic does not bifurcate, any equivalent class contains at most two points and thus we obtain an involution $\phi : \Sigma_p \times \{R\} \to \Sigma_p \times \{R\}$ such that $X = \bar{C}_\kappa^R(\Sigma)/\tilde{x} \sim \phi(\tilde{x})$, $\tilde{x} \in \Sigma_p \times \{R\}$. The main difficulty is to show that ϕ is an isometry. Our main technical lemma says that ϕ is almost 1-bi-Lipschitz up to a uniform error:

$$\left| \frac{|\phi(\tilde{x})\phi(\tilde{y})|}{|\tilde{x}\tilde{y}|} - 1 \right| \leq 20\, \tilde{x}\tilde{y}|$$

for $|\tilde{x}\tilde{y}|$ small (see Lemma 2.12). This implies that ϕ is continuous and preserves the length of a path, and thus ϕ is distance-nonincreasing. Consequently, ϕ is an isometry since ϕ is an involution. Note that without the curvature lower bound, this does not, in general, imply that the metric on $X = \bar{C}_\kappa^R(\Sigma)/\tilde{x} \sim \phi(\tilde{x})$ coincides with the induced metric. For example, $X = \bar{C}_0^1(\mathbb{S}_1^1)/(\tilde{x} \sim \tilde{x}) = \bar{B}_1(\mathbb{R}^2)$ is equipped with the length metric and coincides with the Euclidean metric when restricted to the interior, and $L(\gamma)$ is half of the Euclidean arc length for any $\gamma \subset \partial X$. Our proof relies on the curvature lower bound as well as the cone metric.

Let $L_p(X) = g\exp_p(\Sigma \times \{R\})$, which locally divides a tubular neighborhood of $L_p(X)$ into two components U_1, U_2. The main difficulty in proving the above inequality is that a geodesic in X connecting two points $a, b \in L_p(X)$ may intersect with $L_p(X)$ at many points other than a, b (called *crossing points*). We show that if a geodesic is not contained in $L_p(X)$, the crossing points are discrete (Corollary 2.9). Thus we can reduce the proof to the case where $c_1 = [ab] \subset U_1$ has no crossing point. It's sufficient to construct a noncrossing piecewise intrinsic geodesic $c_2 \subset U_2$ connecting a, b, and show that $\mathrm{length}(c_2)$ is close to $\mathrm{length}(c_1) = |ab|$ up to a second order error (Lemma 2.12).

We remark that the present proof, in an essential way, relies on the κ-cone metric structure; and we believe that establishing a similar inequality in general will be the main obstacle in Conjecture 0.1.

Theorems A, B, C and D are proved in Sections 1, 2, 3 and 4, respectively.

1. Proof of Theorem A: (n, δ)-strained isometries

Let $f : Z \to X$ be as in Theorem A. We will establish that f is an isometry through the following properties:

(i) If a distance-nonincreasing onto map f preserves the volume of the total spaces, then f and f^{-1} preserve volumes of any subsets (see Lemma 1.1).

(ii) Based on a local bi-Lipschitz embedding property (see Lemma 1.2), we show that for δ suitably small, f is injective on $f^{-1}(X^\delta) \subseteq Z^{\tau(\delta)}$. In particular, for any curve $c \subset X^\delta$, $f^{-1}(c) \subseteq Z^{\tau(\delta)}$ is a curve (see Lemma 1.3).

(iii) Our main technical lemma is a volume formula for a tube of ϵ-balls (which can be treated as a replacement for an ϵ-tube around a curve, see Lemma 1.4). Together with (i) and (ii), this formula implies that f^{-1} preserves the length of any geodesic in X^δ up to an error $\tau(\delta)$. Because for any small $\delta < 1/(8n)$, the set X^δ is dense in X (see Lemma 1.6), we are able to show that f is also distance nondecreasing and thus f is an isometry.

Lemma 1.1. *Let $f : Z \to X$ be a distance-nonincreasing onto map of two metric spaces of equal Hausdorff dimension. If* $\operatorname{vol} X = \operatorname{vol} Z$, *then, for any subset $A \subseteq Z$ and $B \subseteq X$,*

$$\operatorname{vol} A = \operatorname{vol} f(A), \quad \operatorname{vol} B = \operatorname{vol} f^{-1}(B).$$

Proof. We argue by contradiction. If $\operatorname{vol} A > \operatorname{vol} f(A)$, then

$$\operatorname{vol} Z = \operatorname{vol} A + \operatorname{vol}(Z - A) > \operatorname{vol} f(A) + \operatorname{vol} f(Z - A) \geq \operatorname{vol} f(Z) = \operatorname{vol} X,$$

a contradiction. Similarly, one can check that $\operatorname{vol} f^{-1}(B) = \operatorname{vol} B$. □

Let $X^\delta(\rho)$ denote the union of points with an (n, δ)-strainer $\{(a_i, b_i)\}$ of radius $\rho > 0$, where $\rho = \min_{1 \leq i \leq n} \{|pa_i|, |pb_i|\} > 0$.

Lemma 1.2 [Burago et al. 1992, Theorem 9.4]. *Let $X \in \mathrm{Alex}^n(\kappa)$. If $p \in X^\delta(\rho)$, the map $\psi : X \to \mathbb{R}^n$ defined by $\psi(x) = (|a_1 x|, \dots, |a_n x|)$ maps a small neighborhood U of p $\tau(\delta, \delta_1)$-almost isometrically onto a domain in $\mathbb{R}^n$, that is,*

$$\big| |\psi(x)\psi(y)| - |xy| \big| < \tau(\delta, \delta_1)|xy|$$

for any $x, y \in U$, where $\delta_1 = \rho^{-1} \operatorname{diam}(U)$. In particular, ψ is a $\tau(\delta)$-almost isometric embedding when restricting to $B_{\delta\rho}(p)$.

A consequence of Lemma 1.2 is that

$$1 - \tau(\delta) \leq \frac{\operatorname{vol} B_\epsilon(p)}{\operatorname{vol} B_\epsilon(\mathbb{R}^n)} \leq 1 + \tau(\delta)$$

for any $p \in X^\delta(\rho)$ and $\epsilon \leq \delta\rho$.

Lemma 1.3. *Let the assumptions be as in Theorem A. Then* $f : f^{-1}(X^\delta) \to X^\delta$ *is injective. Consequently, if* $\gamma \subset X^\delta$ *is a continuous curve,* $f^{-1}(\gamma)$ *is also a continuous curve.*

Proof. We argue by contradiction, assuming $z_1 \neq z_2 \in f^{-1}(X^\delta)$ such that $f(z_1) = f(z_2) = x$. We may assume that z_1 and z_2 have $\tau(\delta)$-strainers of radius $\rho > 0$. Choose $4\epsilon < |z_1 z_2|$ and $\epsilon < \delta\rho$. By Lemma 1.1 and the above consequence of Lemma 1.2, we get

$$1 = \frac{\operatorname{vol} f^{-1}(B_\epsilon(x))}{\operatorname{vol} B_\epsilon(x)} \geq \frac{\operatorname{vol} B_\epsilon(z_1) + \operatorname{vol} B_\epsilon(z_2)}{\operatorname{vol} B_\epsilon(x)} \geq 2(1-\tau(\delta)),$$

a contradiction. □

We now develop a formula which estimates the volume of an ϵ-ball tube with a higher order error. Let $x_1, x_2, \dots, x_{N+1}$ be $N+1$ points in $X^\delta(\rho)$. We first give an estimate of the volume of the ϵ-ball tube $\bigcup_{i=1}^{N+1} B_\epsilon(x_i)$ in terms of $\sum_{i=1}^{N} |x_i x_{i+1}|$ and ϵ, δ with errors.

Lemma 1.4 (volume of an ϵ-ball tube). *Let* $X \in \mathrm{Alex}^n(\kappa)$ *and* $x_i \in X^\delta(\rho)$, $i = 1, 2, \dots, N+1$ *satisfy that* $0 < |x_i x_{i+1}| < 2\epsilon \ll \delta\rho$ *and* $B_\epsilon(x_i) \cap B_\epsilon(x_j) \cap B_\epsilon(x_k) = \varnothing$ *for* $i \neq j \neq k$. *Then the volume of the* ϵ*-ball tube* $\bigcup_{i=1}^{N+1} B_\epsilon(x_i)$ (see Figure 1) *satisfies*

$$(1\text{-}1) \quad (1+\tau(\delta)) \operatorname{vol} \bigcup_{i=1}^{N+1} B_\epsilon(x_i) = \operatorname{vol} B_\epsilon(\mathbb{R}^n) + 2\epsilon \operatorname{vol} B_\epsilon(\mathbb{R}^{n-1}) \sum_{i=1}^{N} \int_{\theta_i}^{\pi/2} \sin^n(t)\, dt,$$

where $\theta_i \in [0, \pi/2]$ *such that* $\cos\theta_i = |x_i x_{i+1}|/(2\epsilon)$. *If, in addition,* $|x_i x_{i+1}| \leq \epsilon^2$ *for all* $1 \leq i \leq N$,

$$(1\text{-}2) \quad (1+\tau(\delta)) \operatorname{vol} \bigcup_{i=1}^{N+1} B_\epsilon(x_i) = \operatorname{vol} B_\epsilon(\mathbb{R}^n) + \operatorname{vol} B_\epsilon(\mathbb{R}^{n-1}) \sum_{i=1}^{N} |x_i x_{i+1}| + O(\epsilon^{n+1}) \sum_{i=1}^{N} |x_i x_{i+1}|.$$

Because $B_\epsilon(x_{i-1}) \cup B_\epsilon(x_i) \cup B_\epsilon(x_{i+1}) \subset B_{\delta\rho}(x_i)$, which is $\tau(\delta)$-almost isometrically embedded into $\mathbb{R}^n$, one can divide $\bigcup_{i=1}^{N+1} B_\epsilon(x_i)$ into small pieces $\Gamma^\pm(x_i)$, whose volumes are $(1+\tau(\delta))$-proportional to the volumes of the following "trapezoidal balls"

$$\Gamma_\epsilon^{h_i^\pm}(\mathbb{R}^n)$$

in $\mathbb{R}^n$. This allows us to reduce the calculation to Euclidean space.

We define the trapezoidal ball $\Gamma_r^h(\mathbb{R}^n)$ in $\mathbb{R}^n_+ = \{(x_1, x_2, \dots, x_n) : x_n \geq 0\}$ in the following way. Let $u \in \mathbb{R}^n_+$ be a point with $|ou| = h \leq r$. Then the hyper plane H

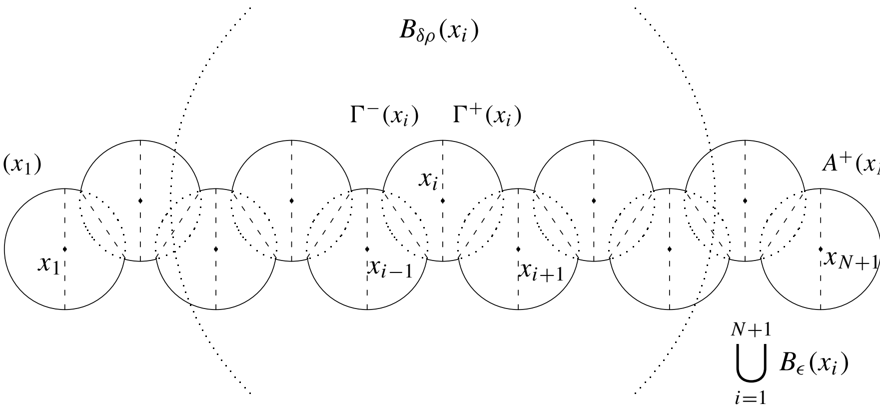


Figure 1

passing through u and perpendicular to $\overrightarrow{ou}$ divides the half ball $B_r(\mathbb{R}^n) \cap \mathbb{R}^n_+$ into two subsets. Let $\Gamma^h_r(\mathbb{R}^n)$ be the subset which contains the origin (see Figure 3). It's easy to see that vol $\Gamma^h_r(\mathbb{R}^n)$ depends only on h and r, and not on the direction $\overrightarrow{ou}$, as long as $H \cap B_r(\mathbb{R}^n) \subset \mathbb{R}^n_+$.

Lemma 1.5. *Let $\Gamma^h_r(\mathbb{R}^n)$ be a trapezoidal ball defined as above. Then*

$$\text{vol}\, \Gamma^h_r(\mathbb{R}^n) = r \,\text{vol}\, B_r(\mathbb{R}^{n-1}) \int_\theta^{\pi/2} \sin^n(t)\, dt,$$

where $\theta \in [0, \pi/2]$ such that $r \cos\theta = h$.

Proof. Let $s = r \cos t \in [0, h]$ be the parameter for the height with the corresponding angle $t \in [\theta, \pi/2]$. Then

$$\begin{aligned}\text{vol}\, \Gamma^h_r(\mathbb{R}^n) &= \int_0^h \text{vol}\, B_{r \sin t}(\mathbb{R}^{n-1})\, ds = \int_\theta^{\pi/2} \text{vol}\, B_{r\sin t}(\mathbb{R}^{n-1}) r \sin(t)\, dt \\ &= r \,\text{vol}\, B_r(\mathbb{R}^{n-1}) \int_\theta^{\pi/2} \sin^n(t)\, dt. \qquad \square\end{aligned}$$

Proof of the volume formula, Lemma 1.4. Because $B_\epsilon(x_i) \cap B_\epsilon(x_{i+1}) \neq \varnothing$ and $B_\epsilon(x_i) \cap B_\epsilon(x_j) \cap B_\epsilon(x_k) = \varnothing$ for any $i \neq j \neq k$, we can decompose $\bigcup_{i=1}^{N+1} B_\epsilon(x_i)$ as the following (see Figure 2): let

$$A^+(x_i) = \{q \in B_\epsilon(x_i) : |qx_i| \le |qx_{i+1}|\}, \quad A^-(x_i) = \{q \in B_\epsilon(x_i) : |qx_i| \le |qx_{i-1}|\}.$$

For $i = 2, 3, \ldots, N$, let

$$\begin{aligned} H^+(x_i) &= A^+(x_i) \cap A^-(x_{i+1}) = \{q \in B_\epsilon(x_i) \cap B_\epsilon(x_{i+1}) : |qx_i| = |qx_{i+1}|\}, \\ H^-(x_i) &= A^-(x_i) \cap A^+(x_{i-1}) = \{q \in B_\epsilon(x_i) \cap B_\epsilon(x_{i-1}) : |qx_i| = |qx_{i-1}|\}, \end{aligned}$$

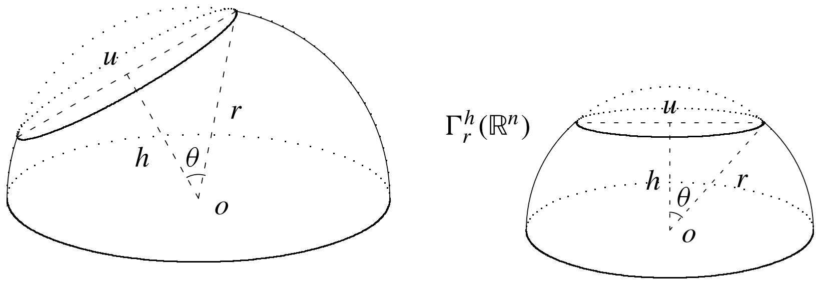


Figure 2

and

$$\Gamma^+(x_i) = \big\{q \in A^+(x_i) \cap A^-(x_i) : d(q, H^+(x_i)) \le d(q, H^-(x_i))\big\},$$
$$\Gamma^-(x_i) = \big\{q \in A^+(x_i) \cap A^-(x_i) : d(q, H^+(x_i)) \ge d(q, H^-(x_i))\big\}.$$

By the construction,

$$\bigcup_{i=1}^{N+1} B_\epsilon(x_i) = A^-(x_1) \cup \left(\bigcup_{i=2}^{N} \Gamma^\pm(x_i)\right) \cup A^+(x_{N+1}).$$

Note that $H^\pm(x_i)$, $i = 2, \dots, N$ consist of all the possible intersections of any two of $A^-(x_1)$, $\Gamma^\pm(x_i)$, $i = 2, \dots, N$, and $A^+(x_{N+1})$ and $\operatorname{vol} H^\pm(x_i) = 0$. We have

$$\text{(1-3)} \quad \operatorname{vol} \bigcup_{i=1}^{N+1} B_\epsilon(x_i) = \operatorname{vol} A^-(x_1) + \operatorname{vol} A^+(x_{N+1}) + \sum_{i=2}^{N} \operatorname{vol} \Gamma^+(x_i) + \sum_{i=2}^{N} \operatorname{vol} \Gamma^-(x_i).$$

Because $B_\epsilon(x_{i-1}) \cup B_\epsilon(x_i) \cup B_\epsilon(x_{i+1}) \subset B_{\delta\rho}(x_i)$, which is homeomorphically and $\tau(\delta)$-almost isometrically embedded into $\mathbb{R}^n$, we have that

$$(1+\tau(\delta)) \operatorname{vol} \Gamma^\pm(x_i) = \operatorname{vol} \Gamma_\epsilon^{h_i^\pm}(\mathbb{R}^n),$$
$$(1+\tau(\delta)) \operatorname{vol} A^+(x_1) = \tfrac{1}{2} \operatorname{vol} B_\epsilon(\mathbb{R}^n) + \operatorname{vol} \Gamma_\epsilon^{h_1^+}(\mathbb{R}^n),$$
$$(1+\tau(\delta)) \operatorname{vol} A^-(x_{N+1}) = \tfrac{1}{2} \operatorname{vol} B_\epsilon(\mathbb{R}^n) + \operatorname{vol} \Gamma_\epsilon^{h_{N+1}^-}(\mathbb{R}^n),$$

where $h_i^+ = \frac{1}{2}|x_i x_{i+1}|$, $h_i^- = \frac{1}{2}|x_i x_{i-1}|$. Note that it's our convention that the same symbol $\tau(\delta)$ may represent different functions of δ, as long as $\tau(\delta) \to 0$ as $\delta \to 0$. Together with (1-3) and the fact that $h_i^+ = h_{i+1}^-$, we get

$$\text{(1-4)} \qquad (1+\tau(\delta)) \operatorname{vol} \bigcup_{i=1}^{N+1} B_\epsilon(x_i) = \operatorname{vol} B_\epsilon(\mathbb{R}^n) + 2\sum_{i=1}^{N} \operatorname{vol} \Gamma_\epsilon^{h_i^+}(\mathbb{R}^n).$$

Let $\theta_i \in [0, \pi/2]$ such that $\cos\theta_i = h_i^+/\epsilon = |x_i x_{i+1}|/(2\epsilon)$. By Lemma 1.5, we have

$$\operatorname{vol}\Gamma_\epsilon^{h_i^+}(\mathbb{R}^n) = \epsilon \operatorname{vol} B_\epsilon(\mathbb{R}^{n-1}) \int_{\theta_i}^{\pi/2} \sin^n(t)\,dt.$$

Plugging this into (1-4), we get (1-1).

To get (1-2), we need to write $\int_{\theta_i}^{\pi/2} \sin^n(t)\,dt$ in terms of $|x_i x_{i+1}|$. Let

$$g(s) = \int_{\theta}^{\pi/2} \sin^n(t)\,dt,$$

where $\theta \in [0, \pi/2]$ with $\cos\theta = s/(2\epsilon)$. Noting that $\theta = \pi/2$ if and only if $s = 0$, we have $g(0) = 0$. Furthermore,

$$\begin{aligned} g'(s) &= -\sin^n\theta \cdot \frac{d\theta}{ds} = -\sin^n\theta \cdot \frac{1}{-2\epsilon\sin\theta} = \frac{\sin^{n-1}\theta}{2\epsilon}; \\ g''(s) &= \frac{1}{2\epsilon}(n-1)\sin^{n-2}\theta\cos\theta \cdot \frac{1}{-2\epsilon\sin\theta} = \frac{n-1}{-4\epsilon^2}\sin^{n-3}\theta\cos\theta; \end{aligned}$$

and thus $g'(0) = 1/(2\epsilon)$, $g''(0) = 0$, and $g'''(0) = c_n/\epsilon^3$. The Taylor expansion of g at $s = 0$ is

$$g(s) = \int_{\theta}^{\pi/2} \sin^n(t)\,dt = 0 + \frac{s}{2\epsilon} + \frac{1}{\epsilon^3} \cdot O(s^3).$$

Letting $s = |x_i x_{i+1}| \le \epsilon^2$, we get

$$\int_{\theta_i}^{\pi/2} \sin^n(t)\,dt = \frac{1}{2\epsilon}|x_i x_{i+1}| + O(\epsilon)|x_i x_{i+1}|.$$

Plugging this into (1-1), we get (1-2). □

In the rest of this section we assume that $f : Z \to X$ is a distance-nonincreasing onto map such that $f^{-1}(X^\delta) \subset Z^{\tau(\delta)}$. By Lemma 1.3, f is homeomorphic on $f^{-1}(X^\delta)$.

Lemma 1.6. *Let the assumptions be as in Theorem A. Let $x, y \in X^\delta$. For $\delta > 0$ sufficiently small, there exists a small constant $c = c(\rho, \delta) > 0$ such that if $|xy| \le c$, $|f^{-1}(x) f^{-1}(y)| \le 2|xy|$.*

Proof. Assume that $|xy| = \epsilon \ll \delta\rho$ and $|f^{-1}(x) f^{-1}(y)| > 2\epsilon$. Consider the metric balls $B_\epsilon(x)$ and $B_\epsilon(y)$. By Lemma 1.4,

$$\begin{aligned} &(1+\tau(\delta))\, vol(B_\epsilon(x) \cup B_\epsilon(y)) \\ &\qquad = \operatorname{vol} B_\epsilon(\mathbb{R}^n) + 2\epsilon \operatorname{vol} B_\epsilon(\mathbb{R}^{n-1}) \int_{\pi/3}^{\pi/2} \sin^n(t)\,dt + O(\epsilon^{n+1}). \end{aligned}$$

Since $B_\epsilon(f^{-1}(x)) \cap B_\epsilon(f^{-1}(y)) = \varnothing$, we have

$$(1+\tau(\delta))\, vol(B_\epsilon(f^{-1}(x)) \cup B_\epsilon(f^{-1}(y))) = 2 \operatorname{vol} B_\epsilon(\mathbb{R}^n).$$

Because f is distance-nonincreasing,

$$B_\epsilon(f^{-1}(x)) \cup B_\epsilon(f^{-1}(y)) \subset f^{-1}(B_\epsilon(x) \cup B_\epsilon(y)).$$

Together with the fact that f^{-1} is volume-preserving, we get

$$\begin{aligned} 1 &= \frac{\operatorname{vol} f^{-1}(B_\epsilon(x) \cup B_\epsilon(y))}{vol(B_\epsilon(x) \cup B_\epsilon(y))} \\ &\geq \frac{(1-\tau(\delta)) \cdot 2 \operatorname{vol} B_\epsilon(\mathbb{R}^n)}{\operatorname{vol} B_\epsilon(\mathbb{R}^n) + 2\epsilon \operatorname{vol} B_\epsilon(\mathbb{R}^{n-1}) \int_{\pi/3}^{\pi/2} \sin^n(t)\,dt + O(\epsilon^{n+1})} \\ &= \frac{(1-\tau(\delta)) \cdot 2 \int_0^{\pi/2} \sin^n(t)\,dt}{\int_0^{\pi/2} \sin^n(t)\,dt + \int_{\pi/3}^{\pi/2} \sin^n(t)\,dt + O(\epsilon)}. \end{aligned}$$

(See Lemma 1.5, $\theta = 0$.) This leads to a contradiction for sufficiently small ϵ and δ. □

In the proof of Theorem A, we will need the following result.

Lemma 1.7 [Burago et al. 1992, 10.6.1]. *Let $X \in \operatorname{Alex}^n(\kappa)$. For a fixed sufficiently small $\delta > 0$, the union of interior points which do not admit any (n, δ)-strainer has Hausdorff dimension $\leq n-2$. In particular, X^δ is dense.*

Proof of Theorem A. Since f is distance-nonincreasing, it suffices to show that f is distance nondecreasing, that is, for any $\tilde{a}, \tilde{b} \in Z$, $|ab| \geq |\tilde{a}\tilde{b}|$, where $a = f(\tilde{a})$ and $b = f(\tilde{b})$.

For any small ϵ_1, by Lemma 1.7, there are $\tilde{a}_{\epsilon_1}, \tilde{b}_{\epsilon_1} \in Z^{\tau(\delta)}$, $a_{\epsilon_1} = f(\tilde{a}_{\epsilon_1})$, $b_{\epsilon_1} = f(\tilde{b}_{\epsilon_1}) \in X^\delta$, such that $|aa_{\epsilon_1}| \leq |\tilde{a}\tilde{a}_{\epsilon_1}| < \epsilon_1$, $|bb_{\epsilon_1}| \leq |\tilde{b}\tilde{b}_{\epsilon_1}| < \epsilon_1$.

Case 1. Assume that there exists a minimal geodesic $[a_{\epsilon_1} b_{\epsilon_1}] \subset X$. Then, because the spaces of directions are isometric along the interior of a geodesic, $[a_{\epsilon_1} b_{\epsilon_1}] \subset X^{2\delta}$ [Petrunin 1998]. By Lemma 1.3 (which will be frequently used without mention), $f^{-1}([a_{\epsilon_1} b_{\epsilon_1}])$ is also a continuous curve. Because $[a_{\epsilon_1} b_{\epsilon_1}]$ is compact, we may let $\rho > 0$ such that $[a_{\epsilon_1} b_{\epsilon_1}] \subset X^{2\delta}(\rho)$ and $f^{-1}([a_{\epsilon_1} b_{\epsilon_1}]) \subset Z^{\tau(\delta)}(\rho)$. Let $\{x_i\}_{i=1}^{N+1}$ be an ϵ-partition of $[a_{\epsilon_1} b_{\epsilon_1}]$, where $x_1 = a_{\epsilon_1}$, $x_{N+1} = b_{\epsilon_1}$ for $\epsilon \ll \delta\rho$. Because $[a_{\epsilon_1} b_{\epsilon_1}]$ is a geodesic, Lemma 1.4 can be applied on the partition $\{x_i\}_{i=1}^{N+1}$. Thus we get

$$\begin{aligned} (1+\tau(\delta)) \operatorname{vol} \bigcup_{i=1}^{N+1} B_\epsilon(x_i) & \\ &= \operatorname{vol} B_\epsilon(\mathbb{R}^n) + \operatorname{vol} B_\epsilon(\mathbb{R}^{n-1}) \sum_{i=1}^{N} |x_i x_{i+1}| + O(\epsilon^{n+1}) \sum_{i=1}^{N} |x_i x_{i+1}| \\ &= \operatorname{vol} B_\epsilon(\mathbb{R}^{n-1})\, |a_{\epsilon_1} b_{\epsilon_1}| + O(\epsilon^n). \end{aligned}$$

Let $z_i = f^{-1}(x_i)$. By Lemma 1.6, $|z_i z_{i+1}| \leq 2|x_i x_{i+1}| = 2\epsilon$. Together with the fact that f is distance-nonincreasing, one can easily check that $\bigcup_{i=1}^{N+1} B_\epsilon(z_i)$ satisfies the condition of Lemma 1.4. Then we have

$$(1+\tau(\delta))\ \mathrm{vol} \bigcup_{i=1}^{N+1} B_\epsilon(z_i) = \mathrm{vol}\, B_\epsilon(\mathbb{R}^{n-1}) \sum_{i=1}^{N} |z_i z_{i+1}| + O(\epsilon^n).$$

Because f is distance-nonincreasing and volume-preserving,

$$\begin{aligned} 1 &= \frac{\mathrm{vol}\, f^{-1}\big(\bigcup_{i=1}^{N+1} B_\epsilon(x_i)\big)}{\mathrm{vol} \bigcup_{i=1}^{N+1} B_\epsilon(x_i)} \geq \frac{\mathrm{vol} \bigcup_{i=1}^{N+1} B_\epsilon(z_i)}{\mathrm{vol} \bigcup_{i=1}^{N+1} B_\epsilon(x_i)} \\ &= (1-\tau(\delta)) \frac{\mathrm{vol}\, B_\epsilon(\mathbb{R}^{n-1}) \sum_{i=1}^{N} |z_i z_{i+1}| + O(\epsilon^n)}{\mathrm{vol}\, B_\epsilon(\mathbb{R}^{n-1})\, |a_{\epsilon_1} b_{\epsilon_1}| + O(\epsilon^n)}, \\ &= (1-\tau(\delta)) \frac{\sum_{i=1}^{N} |z_i z_{i+1}| + O(\epsilon)}{|a_{\epsilon_1} b_{\epsilon_1}| + O(\epsilon)} \\ &\geq (1-\tau(\delta)) \frac{|\tilde{a}_{\epsilon_1} \tilde{b}_{\epsilon_1}| + O(\epsilon)}{|a_{\epsilon_1} b_{\epsilon_1}| + O(\epsilon)}. \end{aligned}$$

Letting $\epsilon \to 0$, we get

$$|a_{\epsilon_1} b_{\epsilon_1}| \geq (1-\tau(\delta)) |\tilde{a}_{\epsilon_1} \tilde{b}_{\epsilon_1}|.$$

Case 2. Assume that there is no minimal geodesic in X^δ from a_{ϵ_1} to b_{ϵ_1} (since X may not be complete). Because spaces of directions along the interior of a geodesic are isometric to each other [Petrunin 1997], we may assume a curve c_1 in X^δ from a_{ϵ_1} to b_{ϵ_1} such that $L(c_1) < |a_{\epsilon_1} b_{\epsilon_1}| + \epsilon_1$. Since $c_1(t)$ is a compact subset in the open set X^δ, we may assume $\eta > 0$ such that an η-tube of c_1 is also contained in X^δ. Consequently, we may assume a piecewise geodesic c in X^δ such that $L(c) \leq L(c_1) \leq |a_{\epsilon_1} b_{\epsilon_1}| + \epsilon_1$. Applying Case 1 to each geodesic segment of c, we conclude that

$$|a_{\epsilon_1} b_{\epsilon_1}| \geq L(c) - \epsilon_1 \geq (1-\tau(\delta)) |\tilde{a}_{\epsilon_1} \tilde{b}_{\epsilon_1}| - \epsilon_1.$$

In either Case 1 or Case 2, we have

$$\begin{aligned} |ab| &\geq |a_{\epsilon_1} b_{\epsilon_1}| - 2\epsilon_1 \geq (1-\tau(\delta)) |\tilde{a}_{\epsilon_1} \tilde{b}_{\epsilon_1}| - 3\epsilon_1 \\ &\geq (1-\tau(\delta))\,(|\tilde{a}\tilde{b}| - 2\epsilon_1) - 3\epsilon_1. \end{aligned}$$

Letting $\delta \to 0$, $\epsilon_1 \to 0$, we get $|ab| \geq |\tilde{a}\tilde{b}|$. □

2. Proof of Theorem B: Relatively maximum volume

Our proof of the classification part in Theorem B is divided into the following two theorems: open ball rigidity (Theorem 2.1) and isometric involution (Theorem 2.2). Recall that $\tilde{o}$ denotes the vertex of the cone $\bar{C}_\kappa^R(\Sigma_p)$ and thus $g\exp_p(\tilde{o}) = p$.

Theorem 2.1. *Under the assumptions of Theorem B,*

$$g\exp_p : C^R_\kappa(\Sigma) \to B_R(p)$$

is an isometry with respect to the intrinsic metrics. In particular, $g\exp_p = \exp_p$.

By Theorem 2.1, $X = \bar{C}^R_\kappa(\Sigma_p)/x \sim x'$, where the equivalent relation $x \sim x'$ if and only if $\exp_p x = \exp_p x'$ and $x, x' \in \Sigma_p \times \{R\}$.

Theorem 2.2. *Let* $X = \bar{C}^R_\kappa(\Sigma_p)/x \sim x' \in \mathrm{Alex}^n(\kappa)$ *be defined as above. Then each equivalent class contains at most two points. Moreover, the induced involution* $\phi : \Sigma_p \times \{R\} \to \Sigma_p \times \{R\}$, $\phi(x) = x'$ (*where* $x \sim x'$) *is an isometry.*

Recall that the induced gradient-exponential map $g\exp_p : \bar{C}^R_\kappa(\Sigma) \to \bar{B}_R(p) = X$ is distance-nonincreasing and onto. Indeed, the open ball rigidity is essentially a consequence of Theorem A and the general property that $\exp_p^{-1} : X \to T_pX : \exp_p^{-1}$ preserves (n,δ)-strained points up to a constant depending on δ (see Lemma 2.4). In the proof, let's recall the following property from [Burago et al. 1992]:

Lemma 2.3 [Burago et al. 1992, Lemmas 7.5 and 11.2]. *Let* $p \in X \in \mathrm{Alex}^n(\kappa)$. *Then, for any* $\delta > 0$, *there is a small neighborhood* U_p *of* p *such that, for any triangle* $\triangle pab$ *with* $a, b \in U_p$, *each angle of* $\triangle pab \subset X$ *differs from the comparison angle of* $\tilde{\triangle} pab \subset \mathbb{S}^2_\kappa$ *by less than* δ.

Lemma 2.4. *Let* $q \in X^\delta$. *Then for any* $p \in X$, $\uparrow_p^q \in \Sigma_p^{\tau(\delta)}$. *Consequently,*

$$\exp_p^{-1}(q) \in \bar{C}^R_\kappa(\Sigma_p)^{\tau(\delta)}.$$

Proof. Since $q \in X^\delta$, by Lemma 1.2, we may assume an $(n, 2\delta)$-strainer $\{(a_i, b_i)\}$ for $q_1 \in [pq]$ and near q, such that $b_n = q$, $a_n \in [pq_1]$. Because the spaces of directions are isometric along the interior of a geodesic [Petrunin 1998], there is $q' \in [pq] \cap U_p$ which has an $(n, \tau(\delta))$-strainer $\{(a_i', b_i')\}$. By the same reason as above, we can assume that $a_n' \in [pq']$ and $b_n' \in [q'q]$.

In addition, we can assume that $|q'a_i'|$, $|q'b_i'|$ are short so that $a_i', b_i' \in U_p$ and $\measuredangle a_i'pq', \measuredangle b_i'pq' < 5\delta$. We claim that

$$\left\{\left(\uparrow_p^{a_i'}, \uparrow_p^{b_i'}\right)\right\}_{i=1}^{n-1}$$

forms an $(n-1, \tau(\delta))$-strainer at $\uparrow_p^q \in \Sigma_p$. It's easy to see that

$$\measuredangle a_i'pq' = \tilde{\measuredangle} a_i'pq' + \tau(\delta) = \frac{|a_i'q'|}{|pq'|} + \tau(\delta).$$

Thus

$$\cos\tilde{\measuredangle} \uparrow_p^{a_i'} \uparrow_p^{q'} \uparrow_p^{x_j} = \frac{|a_i'q'|^2 + |x_jq'|^2 - |a_i'x_j|}{2|a_i'q'|\,|x_jq|} + \tau(\delta) = \cos\tilde{\measuredangle} a_i'q'x_j + \tau(\delta),$$

where $i, j = 1, 2, \ldots, n-1$, $x_j = a_j'$ or b_j'. □

To conclude the open ball rigidity by applying Theorem A, we need to check that $g\exp_p^{-1}(X^\delta) \subseteq \bar{C}_\kappa^R(\Sigma_p)^{\tau(\delta)}$. We do this by showing that $g\exp_p = \exp_p$ when $\operatorname{vol} X = v(\Sigma_p, \kappa, R)$.

Lemma 2.5. *If* $\operatorname{vol} B_R(p) = \operatorname{vol} C_\kappa^R(\Sigma_p)$, *the gradient exponential map is actually an exponential map* $\exp_p : \bar{C}_\kappa^R(\Sigma_p) \to \bar{B}_R(p)$ *which preserves the distance along the radial direction.*

Proof. Clearly, the map $\exp_p^{-1} : \bar{B}_R(p) \to \bar{C}_\kappa^R(\Sigma_p)$ (If there is more than one image, we will pick one) is distance nondecreasing. Because

$$\operatorname{vol} C_\kappa^R(\Sigma_p) = \operatorname{vol} X \le \operatorname{vol}\exp_p^{-1}(X) \le \operatorname{vol} C_\kappa^R(\Sigma_p),$$

$\exp_p^{-1}(X)$ is dense in $C_\kappa^R(\Sigma_p)$. For any $z \in C_\kappa^R(\Sigma_p)$, there is a sequence $x_i \in X$, such that $\exp_p^{-1}(x_i) = z_i \to z$. Let $\exp_p : C_\kappa^R(\Sigma_p) \to X$; $\exp_p(z) = \lim_{i\to\infty} x_i$. Such an $\exp_p$ is well defined, since if there is another sequence $\exp_p^{-1}(x_i') = z_i' \to z$,

$$d(\lim_{i\to\infty} x_i, \lim_{i\to\infty} x_i') = \lim_{i\to\infty} d(x_i, x_i') \le \lim_{i\to\infty} d(z_i, z_i') = 0.$$

It's clear that $\exp_p$, defined as an extension of $\exp_p^{-1}$, is distance-nonincreasing. Moreover, it preserves the distance along the radial direction.

We now show that any geodesic from $p = \exp_p(\tilde{o})$ to $q = \exp_p(\tilde{q}) \in B_R(p)$ can be extended. Therefore $\exp_p$ is a bijection, since geodesics do not bifurcate. Let $[\tilde{o}\tilde{q}]$ be the geodesic in $C_\kappa^R(\Sigma_p)$ such that $\exp_p([\tilde{o}\tilde{q}]) = [pq]$, and $\tilde{q}' \in C_\kappa^R(\Sigma_p)$ the extended point of $[\tilde{o}\tilde{q}]$. Then

$$|pq| + |qq'| \le |\tilde{o}\tilde{q}| + |\tilde{q}\tilde{q}'| = |\tilde{o}\tilde{q}'| = |pq'|,$$

which forces $[pq] \cup [qq']$ to be a geodesic. □

Proof of Theorem 2.1. For $X \in \mathscr{A}_\kappa^R(\Sigma)$ with $\operatorname{vol} X = v(\Sigma, \kappa, R)$, by Lemmas 2.4 and 2.5, we see that $\exp_p : C_\kappa^R(\Sigma) \to B_R(p)$ is a distance-nonincreasing onto map that satisfies the assumptions in Theorem A (note that $\exp_p : \bar{C}_\kappa^R(\Sigma_p) \to \bar{B}_R(p) = X$ may not satisfy the assumptions of Theorem A). □

In the proof of Theorem 2.2, our main technical lemma is Lemma 2.12. Let $\phi : \Sigma \times \{R\} \to \Sigma \times \{R\}$ be defined as in Theorem 2.2. We first observe that ϕ is an involution. Let $L_p(X) = \exp_p(\Sigma \times \{R\}) = \{x \in X : |px| = R\}$.

Lemma 2.6. *Let* $X = \bar{C}_\kappa^R(\Sigma)/x \sim x' \in \operatorname{Alex}^n(\kappa)$ *be defined as in Theorem 2.2. For any* $q \in L_p(X)$, *if* $\tilde{q}_1 \ne \tilde{q}_2$ *with* $\exp_p(\tilde{q}_1) = \exp_p(\tilde{q}_2) = q$, *then the loop* $\exp_p([\tilde{o}\tilde{q}_1]) \cup \exp_p([\tilde{o}\tilde{q}_2])$ *forms a local geodesic at* q. *Consequently*, $\exp_p^{-1}(q)$ *contains at most two points.*

Proof. It's clear that $\exp_p([\tilde{o}\tilde{q}_i])$ are minimal geodesics, $i = 1, 2$. Let $x_i \in X$ be a point on $\exp_p([\tilde{o}\tilde{q}_i])$ and $\tilde{x}_i = \exp_p^{-1}(x_i)$, $i = 1, 2$. We claim that if x_1, x_2

are both close enough to q, the geodesic $[x_1x_2]$ intersects with $L_p(X)$. If not, $[x_1x_2] \subset B_R(p)$. By the assumption, $|x_1x_2|_X = |\tilde{x}_1\tilde{x}_2|_{\bar{C}^R_\kappa(\Sigma)}$. Let $x_1, x_2 \to q$. We get that $|x_1x_2|_X \to 0$ and $|\tilde{x}_1\tilde{x}_2|_{\bar{C}^R_\kappa(\Sigma)} \to |\tilde{q}_1\tilde{q}_2|_{\bar{C}^R_\kappa(\Sigma)} > 0$, a contradiction.

Let $a \in [x_1x_2] \cap L_p(X)$. It remains to show that $a = q$. For $i = 1, 2$,

$$|x_ia| \geq |pa| - |px_i| = |pq| - |px_i| = |x_iq|.$$

Thus

$$|x_1q| + |x_2q| \leq |x_1a| + |x_2a| = |x_1x_2|,$$

which forces both of the above inequalities to be equalities, and thus $a = q$. □

As a corollary of Lemma 2.6, we conclude that for $X \in \mathcal{A}^R_\kappa(\Sigma)$, $\kappa > 0$, and $\pi/(2\sqrt{\kappa}) < R < \pi/\sqrt{\kappa}$, $\operatorname{vol} C^R_\kappa(\Sigma)$ is not the optimal upper bound for $\operatorname{vol} X$; see [Grove and Petersen 1992]. Equivalently, we have:

Corollary 2.7. *Assume $X \in \mathcal{A}^R_\kappa(\Sigma)$ with $\operatorname{vol}(X) = \operatorname{vol} \bar{C}^R_\kappa(\Sigma)$ and $\kappa > 0$. Then $R \leq \pi/(2\sqrt{\kappa})$ or $R = \pi/\sqrt{\kappa}$. In the second case, $X = C_\kappa(\Sigma)$ which is the k-suspension of Σ.*

Proof. Assume $\pi/(2\sqrt{\kappa}) < R < \pi/\sqrt{\kappa}$. Let $p \in X$ such that $\Sigma_p = \Sigma$. It's clear that $\operatorname{rad}_p(X) = R$. We claim that $L_p(X) = \{q\}$ has only one point. Then by Lemma 2.6, $\Sigma_p \times \{R\} = \exp_p^{-1}(q)$ contains at most two points, a contradiction. Let $a \neq b \in L_p(X)$. Consider the triangle $\triangle pab$ and the compared triangle $\widetilde{\triangle} pab \in S^2_\kappa$. Take $c \in [ab]$ and the corresponding $\tilde{c} \in [\tilde{a}\tilde{b}]$ with $|ac| = |\tilde{a}\tilde{c}|$. By the triangle comparison, $|pc| \geq |\tilde{p}\tilde{c}| > R$, a contradiction. Note that the case where $R = \pi/\sqrt{\kappa}$ follows from Theorem 2.1. □

It remains to show that ϕ is an isometry. The following lemma plays an important role in the study of the angles in the gluing space X.

Lemma 2.8. *Let $a, b \in C_\kappa(\Sigma)$. Then $\measuredangle apb = \widetilde{\measuredangle} apb$ and $\measuredangle pab = \widetilde{\measuredangle} pab$.*

Proof. The proofs are essentially the same for different κ. For simplicity, we only give a proof for $\kappa = 0$. Note that $\measuredangle apb = \widetilde{\measuredangle} apb$ by the definition of $C_\kappa(\Sigma)$.

To see $\measuredangle pab = \widetilde{\measuredangle} pab$, shortly extend the geodesic $[pa]$ to a' and apply the cosine law to the triangles $\triangle aa'b$, $\triangle pa'b$, and $\triangle pab$. We get

$$|a'b|^2 = |aa'|^2 + |ab|^2 - 2|aa'|\,|ab| \cos \widetilde{\measuredangle} a'ab, \tag{2-1}$$

$$\begin{aligned} |a'b|^2 &= |pa'|^2 + |pb|^2 - 2|pa'|\,|pb| \cos \measuredangle apb \\ &= (|pa| + |aa'|)^2 + |pb|^2 - 2(|pa| + |aa'|)|pb| \cos \measuredangle apb, \end{aligned} \tag{2-2}$$

$$|ab|^2 = |pa|^2 + |pb|^2 - 2|pa|\,|pb| \cos \measuredangle apb. \tag{2-3}$$

Calculating (2-1) + (2-3) − (2-2), we get

$$\begin{aligned}0 &= |ab|\cos\tilde{\measuredangle}a'ab + |pa| - |pb|\cos\measuredangle apb\\ &\geq |ab|\cos\measuredangle a'ab + |pa| - |pb|\cos\measuredangle apb\\ &= -|ab|\cos\measuredangle pab + |pa| - |pb|\cos\measuredangle apb.\end{aligned}$$

Since $\measuredangle pab \geq \tilde{\measuredangle}pab$ and $\measuredangle apb = \tilde{\measuredangle}apb$, the above inequality implies

$$\begin{aligned}|pa| &\leq |ab|\cos\measuredangle pab + |pb|\cos\measuredangle apb\\ &\leq |ab|\cos\tilde{\measuredangle}pab + |pb|\cos\tilde{\measuredangle}apb = |pa|,\end{aligned}$$

which forces $\measuredangle pab = \tilde{\measuredangle}pab$. □

Corollary 2.9. *Let $x, y \in X$ be two points. If $[xy] \cap L_p(X) \neq \varnothing$, then either $[xy] \subset L_p(X)$ or $[xy] \cap L_p(X)$ is finite.*

Proof. Let $x \notin L_p(X)$. We show that $[xy] \cap L_p(X)$ is finite. Let $a \in [xy] \cap L_p(X)$ be the accumulation point which is closest to x. Clearly $a \neq x$ since $x \notin L_p(X)$. Thus there is a geodesic segment $[ba]$ of $[xy]$ with $[ba] - \{a\} \subset B_R(p)$. Since $|pb| < |pa| = R$, by Lemma 2.8,

$$\measuredangle pab = \tilde{\measuredangle}pab < \frac{\pi}{2}.$$

On the other hand, because there are $a_i \in [xy] \cap L_p(X)$ with $a_i \to a$ as $i \to \infty$ and $|pa| = |pa_i| = R$, by the first variation formula, we get

$$\measuredangle pay = \frac{\pi}{2}.$$

Therefore $\pi = \measuredangle pab + \measuredangle pay < \pi$, a contradiction. □

As another corollary, we prove Theorem 2.2 for the special case $\kappa > 0$ and $R = \pi/(2\sqrt{\kappa})$.

Corollary 2.10. *Theorem 2.2 holds for the case $\kappa > 0$ and $R = \pi/(2\sqrt{\kappa})$.*

Proof. Let $x, y \in L_p(X)$, $\tilde{x}_1, \tilde{x}_2, \tilde{y}_1, \tilde{y}_2 \in \Sigma \times \{R\}$ with $\exp_p(\tilde{x}_1) = \exp_p(\tilde{x}_2) = x$, $\exp_p(\tilde{y}_1) = \exp_p(\tilde{y}_2) = y$. We will show that $|\tilde{x}_1\tilde{y}_1|_{\bar{C}^R_\kappa(\Sigma)} = |\tilde{x}_2\tilde{y}_2|_{\bar{C}^R_\kappa(\Sigma)}$. Assume $|\tilde{x}_1\tilde{y}_1|_{\bar{C}^R_\kappa(\Sigma)} > |\tilde{x}_2\tilde{y}_2|_{\bar{C}^R_\kappa(\Sigma)}$. Then there is a point $a \notin L_p(X)$ (take $\exp_p^{-1}(a)$ close to x_1) such that $[ay] \cap L_p(X)$ contains a point $b \neq y$. Because $\exp_p$ is distance-nonincreasing and $\Sigma \times \{\pi/(2\sqrt{\kappa})\}$ is totally geodesic, $[by] \subset L_p(X)$, which contradicts Corollary 2.9. □

Let $\mathrm{Fix}(\phi) = \{\tilde{x} \in \Sigma \times \{R\} : \phi(\tilde{x}) = \tilde{x}\}$ be the fixed points set. Let $L^1_p(X) = \exp_p(\mathrm{Fix}(\phi))$ denote the image. Due to Lemma 2.6, let $L^2_p(X) = L_p(X) - L^1_p(X)$ denote the points that are identified from exactly two points, that is, for any

$$x \in L^2_p(X),$$

$\exp^{-1}(x) = \{\tilde{x}^+, \tilde{x}^-\}$ contains exactly two points.

In the rest of the proof of Theorem 2.2, by Corollaries 2.9 and 2.10 and their proofs, we can always assume $R < \pi/(2\sqrt{\kappa})$ for $\kappa > 0$ and that for any $x, y \in X$, $[xy] \cap L_p(X)$ is finite if it is not empty. Moreover, the following corollary shows that $]xy[\cap L_p(X) \subset L_p^2(X)$, where $]xy[$ denotes the geodesic connecting x, y without the end points.

Corollary 2.11. *Let the assumption be as in Theorem 2.2. Assume $R < \pi/(2\sqrt{\kappa})$ when $\kappa > 0$. For any $x, y \in X$, if $q \in]xy[\cap L_p(X)$, $q \in L_p^2(X)$.*

Proof. Without losing generality, we assume $x, y \notin L_p(X)$ and $]xy[\cap L_p(X) = \{q\}$. If $q \in L_p^1(X)$, by Lemma 2.8, $\measuredangle xqp = \widetilde{\measuredangle} xqp < \pi/2$ and $\measuredangle yqp = \widetilde{\measuredangle} yqp < \pi/2$. Thus $\measuredangle xqp + \measuredangle yqp < \pi$, which contradicts the fact that $[xy]$ is a geodesic. □

Now we are ready to prove our main technical lemma. Let $x \in L_p^2(X)$ and $\{\tilde{x}^+, \tilde{x}^-\} = \exp_p^{-1}(x)$ denote the preimage. Then there are exactly two geodesics $\exp_p([\tilde{o}\tilde{x}^+])$, $\exp_p([\tilde{o}\tilde{x}^-])$ connecting x to p. To distinguish geodesics and angles, we use the following notation.

- Let $[px^+]$ and $[px^-]$ denote $\exp_p([\tilde{o}\tilde{x}^+])$ and $\exp_p([\tilde{o}\tilde{x}^-])$ respectively.

In addition, for $y \in L_p^2(X)$ and $\exp_p^{-1}(y) = \{\tilde{y}^+, \tilde{y}^-\}$:

- let $[x^\pm y^\pm]$ denote $\exp_p([\tilde{x}^\pm \tilde{y}^\pm])$;
- let $|x^\pm y^\pm|$ denote the length of the geodesics $[x^\pm y^\pm]$;
- let $\measuredangle x^\pm p y^\pm$ denote the angle between $[px^\pm]$ and $[py^\pm]$ at p;
- let $\measuredangle p x^\pm y^\pm$ denote the angle between $[px^\pm]$ and $[x^\pm y^\pm]$ at x.

Lemma 2.12. *Let the assumptions be as in Theorem 2.2. Assume $R < \pi/(2\sqrt{\kappa})$ when $\kappa > 0$. Then, for any $\tilde{x} \neq \tilde{y} \in \Sigma \times \{R\}$ with $|\tilde{x}\tilde{y}|$ sufficiently small,*

$$\left| \frac{|\phi(\tilde{x})\phi(\tilde{y})|}{|\tilde{x}\tilde{y}|} - 1 \right| \leq 20\,|\tilde{x}\tilde{y}|.$$

Proof. For simplicity, we give a proof for the case $\kappa = 0$. The other cases can be carried out similarly. Throughout the proof, we will frequently use lemmas 2.6 and 2.8 and Corollary 2.11 without mentioning it. We will also assume that for any $a, b \in X$, $[ab] \cap L_p(X)$ is finite if it is not empty.

Clearly, ϕ preserves the distance when x and y are both in $L_p^1(X)$. Let

$$x \in L_p^2(X), \quad y \in L_p(X)$$

(if $y \in L_p^1(X)$, $\tilde{y}^+ = \tilde{y}^-$ will denote the same point and the argument will still go through). Because $[xy] \cap L_p(X)$ is finite, not losing generality, assume $[xy] =$

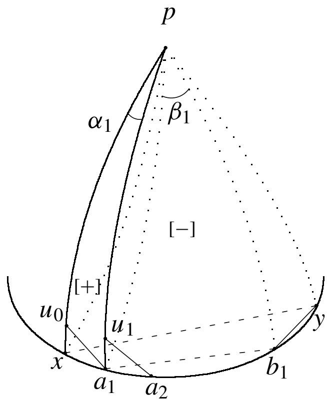


Figure 3

$[x^- y^-]$. Thus $\measuredangle x^- p y^- \leq \measuredangle x^+ p y^+$. Let $\beta_0 = \measuredangle x^- p y^-$. Since $|x^- y^-| = 2R \sin \dfrac{\beta_0}{2}$ and $|x^+ y^+| = 2R \sin(\measuredangle x^+ p y^+/2)$, it's sufficient to show that

$$10\beta_0^2 + \beta_0 \geq \measuredangle x^+ p y^+. \tag{2-4}$$

Take $u_0 \in [p x^+]$ with $|u_0 x^+| = \epsilon$. Let $[u_0 y]$ be a geodesic. If

$$[(u_0 y)] \cap L_p(X) \neq \varnothing,$$

let $a_1 (\neq y)$ and b_1 (b_1 can be y) be the first and second intersection points in $[u_0 y] \cap L_p(X)$ along the direction $\uparrow_{u_0}^{y}$ (see Figure 3). Assign $\pm$ to $\exp_p^{-1}(a_1)$, $\exp_p^{-1}(b_1)$ such that $\measuredangle p a_1^+ u_0 < \pi/2$. Let $\alpha_1 = \measuredangle x^+ p a_1^+$ and $\beta_1 = \measuredangle a_1^- p b_1^-$. In the case of $[(u_0 y)] \cap L_p(X) = \varnothing$, we take $b_1 = a_1 = y$ and $\beta_1 = 0$.

Because $[u_0 a_1^+] * [a_1^- b_1^-] * [b_1^+ y]$ is a minimal geodesic, by triangle inequality,

$$|u_0 x| + |xy| \geq |u_0 a_1^+| + |a_1^- b_1^-| + |b_1 y|.$$

This implies

$$\epsilon + 2R \sin \frac{\beta_0}{2} \geq |u_0 a_1^+| + 2R \sin \frac{\beta_1}{2}. \tag{2-5}$$

Applying the cosine law (the form in Lemma 4.7(5)) in $\triangle p u_0 a_1$ with the angle $\measuredangle u_0 p a_1^+ = \alpha_1$, we get that

$$|u_0 a_1^+| = \sqrt{\epsilon^2 + 4R(R-\epsilon)\sin^2 \frac{\alpha_1}{2}} \geq 2(R-\epsilon)\sin \frac{\alpha_1}{2}.$$

Thus

$$\epsilon + 2R \sin \frac{\beta_0}{2} \geq 2(R-\epsilon)\sin \frac{\alpha_1}{2} + 2R \sin \frac{\beta_1}{2}. \tag{2-6}$$

If $[(u_0y)]\cap L_p(X)=\varnothing$, we stop here. If $[(u_0y)]\cap L_p(X)\neq\varnothing$, we proceed with $u_1\in[pa_1^+]$ and $|u_1a_1|=\epsilon$. Let $[u_1b_1]$ be a geodesic. Again, if $[(u_1b_1)]\cap L_p(X)\neq\varnothing$, let $a_2(\neq y)$ and b_2 (can be b_1) be the first and second intersection points in $[u_1b_1]\cap L_p(X)$ along the direction $\uparrow_{u_1}^{b_1}$. Assign $\pm$ to $\exp_p^{-1}(a_2)$, $\exp_p^{-1}(b_2)$ such that $\measuredangle pa_2^+u_1<\pi/2$. Let $\alpha_2=\measuredangle a_1^+pa_2^+$ and $\beta_2=\measuredangle a_2^-pb_2^-$. If $[(u_1b_1)]\cap L_p(X)=\varnothing$, $a_2=b_2=b_1$, $\beta_2=0$, and we stop the process. Proceed inductively until $[(u_Nb_N)]\cap L_p(X)=\varnothing$, which yields that $a_{N+1}=b_{N+1}=b_N$ and $\beta_{N+1}=0$. We claim that N is finite, and, moreover,

$$(N+1)\epsilon<5R\,\beta_0^2. \tag{2-7}$$

For each $0\le i\le N$, we have

$$\epsilon+2R\sin\frac{\beta_i}{2}\ge|u_ia_{i+1}^+|+2R\sin\frac{\beta_{i+1}}{2}, \tag{2-8}$$

$$\epsilon+2R\sin\frac{\beta_i}{2}\ge 2(R-\epsilon)\sin\frac{\alpha_{i+1}}{2}+2R\sin\frac{\beta_{i+1}}{2}, \tag{2-9}$$

where $\alpha_i=\measuredangle a_i^+pa_{i+1}^+$, $\beta_i=\measuredangle a_i^-pb_i^-$. Summing up (2-9) for $i=0,1,\dots,N$ and applying (2-7), we get

$$\begin{aligned}5R\,\beta_0^2+2R\sin\frac{\beta_0}{2}&\ge(N+1)\epsilon+2R\sin\frac{\beta_0}{2}\\&\ge 2(R-\epsilon)\sum_{i=1}^{N}\sin\frac{\alpha_i}{2}\ge 2(R-\epsilon)\sin\frac{\sum_{i=1}^{N}\alpha_i}{2}\\&\ge 2(R-\epsilon)\sin\frac{\measuredangle x^+pb_N}{2}.\end{aligned}$$

Since $b_N\to b_1\to y^+$ when taking $\epsilon\to 0$, (2-4) follows.

It remains to show (2-7). A sum of (2-8) for $i=0,1,\dots,N$ indicates that the upper bound of N relies on an estimate of $|u_ia_{i+1}^+|$ in terms of ϵ and β_{i+1}. Noting that $a_{i+1}=[u_ib_{i+1}]\cap([pa_{i+1}^+]*[pa_{i+1}^-])$ and $[pa_{i+1}^+]*[pa_{i+1}^-]$ is a local geodesic at a_{i+1}, we have $\measuredangle pa_{i+1}^+u_i=\measuredangle pa_{i+1}^-b_{i+1}=\pi/2-\beta_{i+1}/2$. Applying the cosine law in triangle $\triangle pu_ia_{i+1}^+$, we get

$$(R-\epsilon)^2=R^2+|u_ia_{i+1}^+|^2-2R|u_ia_{i+1}^+|\sin\frac{\beta_{i+1}}{2},$$

that is,

$$|u_ia_{i+1}^+|^2-2R\sin\frac{\beta_i}{2}\,|u_ia_{i+1}^+|+R\epsilon-\epsilon^2=0.$$

Solving for $|u_ia_{i+1}^+|$ and taking into account that $\epsilon>0$ is small, we have

$$|u_ia_{i+1}^+|\ge R\sin\frac{\beta_{i+1}}{2}-\sqrt{\Big(R\sin\frac{\beta_{i+1}}{2}\Big)^2-(R\epsilon-\epsilon^2)}>\frac{\epsilon}{4\sin(\beta_{i+1}/2)}.$$

Note that β_i is decreasing, which is implied by (2-8) and $|u_i a_{i+1}^+| > |u_i a_i^+| = \epsilon$. We get

$$|u_i a_{i+1}^+| > \frac{\epsilon}{4\sin(\beta_0/2)}. \tag{2-10}$$

Plugging (2-10) into (2-8), we get

$$\epsilon + 2R\sin\frac{\beta_i}{2} > \frac{\epsilon}{4\sin(\beta_0/2)} + 2R\sin\frac{\beta_{i+1}}{2}. \tag{2-11}$$

Summing up (2-11) for $i = 0, 1, \ldots, N$, we get

$$(N+1)\epsilon + 2R\sin\frac{\beta_0}{2} > (N+1)\frac{\epsilon}{4\sin(\beta_0/2)}.$$

Therefore

$$(N+1)\epsilon < \frac{8R\sin^2(\beta_0/2)}{1-4\sin(\beta_0/2)} < 5R\,\beta_0^2. \qquad \square$$

Proof of Theorem 2.2, assuming $R < \pi/(2\sqrt{\kappa})$ when $\kappa > 0$. By Lemma 2.12, ϕ is a continuous involution and thus a homeomorphism. It reduces to show that $\phi : \Sigma \times \{R\} \to \Sigma \times \{R\}$ preserves the length of any curve $c : [0, 1] \to \Sigma \times \{R\}$. Given $\delta, \epsilon > 0$, we may assume a partition $P : 0 = t_0 < t_1 < \cdots < t_N = 1$ with $|c(t_i)c(t_{i+1})| \le \delta$ such that the length of the curve satisfies

$$L(c) < \sum_{i=0}^{N-1} |c(t_i)c(t_{i+1})| + \frac{\epsilon}{2}, \quad L(\phi(c)) < \sum_{i=0}^{N-1} |\phi(c(t_i))\phi(c(t_{i+1}))| + \frac{\epsilon}{2}.$$

Then

$$\begin{aligned} |L(c) - L(\phi(c))| &\le \sum_{i=0}^{N-1} \big||c(t_i)c(t_{i+1})| - |\phi(c(t_i))\phi(c(t_{i+1}))|\big| + \epsilon \\ &\le \sum_{i=0}^{N-1} 20\,|c(t_i)c(t_{i+1})|^2 + \epsilon \\ &\le 20\,\delta \sum_{i=0}^{N-1} |c(t_i)c(t_{i+1})| + \epsilon \\ &\le 20\,\delta\, L(c) + \epsilon. \end{aligned}$$

Since $\epsilon > 0$ and $\delta > 0$ can be chosen arbitrarily small, we get the desired result. $\square$

Completion of Proof of Theorem B. By Theorems 2.1 and 2.2, we identify X with $\bar{C}_\kappa^R(\Sigma_p)/x \sim \phi(x)$. We show that the metric on X coincides with the metric induced from the identification $x \sim \phi(x)$. It's equivalent to show that

$$\exp_p : \bar{C}_\kappa^R(\Sigma_p) \to X$$

preserves lengths of geodesics. Let $\gamma \subset \overline{C}_\kappa^R(\Sigma_p)$ be a geodesic and $\sigma = f(\gamma)$. Since $L(\gamma) \geq L(\sigma)$, it remains to show that $L(\sigma) \geq L(\gamma)$. Because either $\gamma \subset \Sigma \times \{R\}$ or $\gamma \cap (\Sigma \times \{R\})$ has at most two points, we need only check for the case $\gamma \subset \Sigma \times \{R\}$, that is, $\sigma \subset L_p(X)$. For any $\epsilon > 0$, let $\{x_i\}_{i=0}^{2N+1} \subset \sigma$ be an ϵ-partition and

$$L(\sigma) = \lim_{\epsilon \to 0} \sum_{i=0}^{2N} |x_i x_{i+1}|.$$

Let $a_i \in \gamma$ so that $\exp_p(a_i) = x_i$. Choose $b_{2k} \in C_\kappa^R(\Sigma)$, $k = 0, 1, \ldots, N$, with $|a_{2k} - b_{2k}| < \epsilon^4$. Let $b_{2k+1} = a_{2k+1}$ for $k = 0, 1, \ldots, N$ and $y_i = \exp_p(b_i)$ for $i = 0, 1, \ldots 2N+1$. Then $|y_i - x_i| \leq |b_i - a_i| < \epsilon^4$ and thus

$$L(\sigma) = \lim_{\epsilon \to 0} \sum_{i=0}^{2N} |y_i y_{i+1}|.$$

We claim that $[y_i y_{i+1}] \cap L_p(X)$ is either y_i or y_{i+1}. By Corollary 2.9, let

$$u, v \in [y_i y_{i+1}] \cap L_p(X)$$

and there is no crossing point in between. Without losing generality, we assume $y_i \notin L_p(X)$ and $|y_i u| < |y_i v|$. Let $[u^- v^-] \subset [y_i y_{i+1}]$. Because the involution ϕ is an isometry (Theorem 2.2), $L([u^+ v^+]) = L([u^- v^-])$. Thus $[y_i u] \cup [u^+ v^+] \neq [y_i u] \cup [u^- v^-]$ is also a geodesic, which yields a bifurcation of geodesics.

By the claimed property, we have that $|y_i y_{i+1}| = |b_i b_{i+1}|$. Since $\sum_{i=0}^{2N} |b_i b_{i+1}| \geq L(\gamma)$, we have

$$L(\sigma) = \lim_{\epsilon \to 0} \sum_{i=0}^{2N} |y_i y_{i+1}| = \lim_{\epsilon \to 0} \sum_{i=0}^{2N} |b_i b_{i+1}| \geq L(\gamma).$$

It remains to show that for $\Sigma \in \mathrm{Alex}^{n-1}(1)$, if $\phi : \Sigma \times \{R\} \to \Sigma \times \{R\}$ is an isometric involution, $X = \overline{C}_\kappa^R(\Sigma)/(x \sim \phi(x)) \in \mathrm{Alex}^n(\kappa)$.

Case 1. Assume $\partial \Sigma = \varnothing$. Take two copies of $\overline{C}_\kappa^R(\Sigma)$, marked as $\overline{C}_\kappa^R(\Sigma)_1$ and $\overline{C}_\kappa^R(\Sigma)_2$, whose vertices are p_1 and p_2, respectively. Gluing along their boundaries by ϕ, we obtain a double space $\widehat{X} = \overline{C}_\kappa^R(\Sigma)_1 \cup_\phi \overline{C}_\kappa^R(\Sigma)_2$. By the gluing theorem [Petrunin 1997], $\widehat{X} \in \mathrm{Alex}^n(\kappa)$.

Now we extend the isometric $\mathbb{Z}_2$-action by ϕ on Σ to an isometric $\hat{\mathbb{Z}}_2$-action on $\widehat{X}$ such that $X = \widehat{X}/\hat{\mathbb{Z}}_2$, and thus $X \in \mathrm{Alex}^n(\kappa)$. For any $u \in \overline{C}_\kappa^R(\Sigma)_1$, extend the geodesic $[p_1 u]_{\overline{C}_\kappa^R(\Sigma)_1}$ to $u_1 \in (\Sigma \times \{R\})_1$. Let $\hat{\phi}(u)$ be the point on the geodesic $[p_2 \phi(u_1)]_{\overline{C}_\kappa^R(\Sigma)_2}$ such that $|p_2 \hat{\phi}(u)| = |p_1 u|$ (so $\hat{\phi} : \overline{C}_\kappa^R(\Sigma)_1 \to \overline{C}_\kappa^R(\Sigma)_2$). Switching the roles of $\overline{C}_\kappa^R(\Sigma)_1$ and $\overline{C}_\kappa^R(\Sigma)_2$, we extend ϕ to an isometric involution $\hat{\phi} : \overline{C}_\kappa^R(\Sigma)_2 \to \overline{C}_\kappa^R(\Sigma)_1$. Clearly, $\hat{\phi} : \hat{X} \to \widehat{X}$ is an isometric involution such that $X = \widehat{X}/\hat{\phi}$.

Case 2. Assume $\partial\Sigma \neq \varnothing$. Let $\hat{\Sigma} = \Sigma^+ \cup \Sigma^-$ denote the double of Σ. We first extend the isometric involution ϕ on Σ to $\hat{\phi} : \hat{\Sigma} \to \hat{\Sigma}$ by $\hat{\phi}(x_\pm) = \phi(x)_\mp$, where $x_+ = x_- \in \Sigma$. We then define another isometric involution $\psi : \hat{\Sigma} \to \hat{\Sigma}$ by the reflection on $\partial\Sigma$, $\psi(x_\pm) = x_\mp$. Then $\hat{\psi}(\hat{\phi}(x_\pm)) = \hat{\psi}(\phi(x)_\pm) = \phi(x)_\mp = \hat{\phi}(x_\mp) = \hat{\phi}(\hat{\psi}(x_\pm))$. This implies that $\hat{\Sigma}$ admits a $\mathbb{Z}_2 \oplus \mathbb{Z}_2$-action. Clearly, the $\mathbb{Z}_2 \oplus \mathbb{Z}_2$-action extends uniquely to an isometric $\mathbb{Z}_2 \oplus \mathbb{Z}_2$-action on $\bar{C}^r_\kappa(\hat{\Sigma})$. By Case 1, we extend only the $\hat{\phi}$-action to $\widehat{X}$ such that $\bar{C}^r_\kappa(\hat{\Sigma})/x \sim \hat{\phi}(x) \in \mathrm{Alex}^n(\kappa)$. Then $X = [\bar{C}^r_\kappa(\hat{\Sigma})/x \sim \hat{\phi}(x)]/\hat{\psi} \in \mathrm{Alex}^n(\kappa)$. □

By Theorem B, the isometric classification of $X \in \mathscr{A}^r_\kappa(\Sigma)$ with relatively maximum volume reduces to the isometric classification of all $(n-1)$-dimensional Alexandrov spaces Σ with curv ≥ 1 and the equivariant isometric $\mathbb{Z}_2$-actions on Σ. For $n = 2$, one easily gets a complete list:

Corollary 2.13. *Any* 2*-dimensional compact Alexandrov space with* curv $\geq \kappa$ *and relatively maximum volume is isometric to one of the following*:

$$\bar{C}^r_\kappa(S^1_\theta)/\phi_i \quad (i = 1, 2, 3) \quad \text{or} \quad \bar{C}^r_\kappa([0,\theta])/\psi_i \quad (i = 1, 2),$$

where S^1_θ denotes a circle of length 2θ with $0 < \theta \leq \pi$ and $\phi_i : S^1_\theta \to S^1_\theta$ (respectively $\psi_i : [0,\theta] \to [0,\theta]$) is trivial, a reflection or the antipodal map respectively for $i = 1, 2$ and 3 (respectively $i = 1$ and 2).

Example 2.14 (cf. [Grove and Petersen 1992]). Let $Z = \mathbb{D}^2$ be a 2-dimensional flat unit disk. Then $\partial Z = \mathbb{S}^1(1)$ is a unit circle. Let $\phi : \partial Z \to \partial Z$ be a one-to-one map and $X = \mathbb{D}^2/x \sim \phi(x)$ the glued space via identification $z \sim \phi(z)$. By Theorem B, X is an Alexandrov space if and only if ϕ is an isometric involution, that is, ϕ is a reflection, antipodal map, or identity, where X is homeomorphic to $\mathbb{S}^2$, $\mathbb{R}P^2$, and $\mathbb{D}^2$, respectively.

Example 2.15. Consider a 2-dimensional simplex. We identify points on each side via a reflection about the mid point. Then we get a tetrahedron, in which one vertex is glued from the three vertices of the simplex.

3. Proof of Theorem C: Relatively almost maximum volume

In the proof of Theorem C, we need the following result.

Theorem 3.1 [Bredon 1972, Theorem 5.5]. *Let M be a G-manifold. G is a finite group. Assume that, for a given prime p and all p-subgroups, $P \subseteq G$ satisfies*

$$H_i(M^P; \mathbb{Z}_p) = 0, \quad i \leq q \ (\textit{including } P = \{e\}).$$

Then $H_i(M/G; \mathbb{Z}_p) = 0$ for all $i \leq q$. Moreover, if this holds for all prime p and $H_i(M; \mathbb{Z}) = 0$ for $i \leq q$, then $H_i(M/G; \mathbb{Z}) = 0$ for $i \leq q$.

Proof of Theorem C. We first show that if $X \in \mathcal{A}^r_\kappa(\Sigma)$ with $\operatorname{vol} X = v(\Sigma, \kappa, r)$, X is homeomorphic to S^n or $\mathbb{C}P^n$.

By Theorem B, X is isometric to $\bar{C}^R_\kappa(\Sigma))/x \sim \phi(x)$ and $\phi : \Sigma \to \Sigma$ is an isometric involution. To determine the homeomorphism type of X, we consider the double space $\widehat{X} = \bar{C}^R_\kappa(\Sigma))^+ \cup_\phi \bar{C}^R_\kappa(\Sigma))^-$. As seen in the proof of Theorem B, $\hat{X} \in \operatorname{Alex}^n(\kappa)$ and ϕ extends an isometric $\mathbb{Z}_2$-action on $\hat{X}$ such that $\hat{X}/\mathbb{Z}_2$.

We claim that $\hat{X}$ is a homeomorphism sphere. First, $\hat{X}$ is a topological manifold if every point $\hat{q} \in \partial\bar{C}^R_\kappa(\Sigma)) \hookrightarrow \hat{X}$ is a manifold point. According to [Wu 1997], a point x in an Alexandrov space is a manifold point if and only if Σ_x is simply connected. Because $\Sigma_{\hat{q}}$ is a suspension of $\Sigma_{\hat{q}}(\Sigma)$, $\hat{q}$ is a manifold point. By the Poincaré conjecture (in all dimensions), our claim reduces to the following: $\hat{X}$ is an integral homotopy sphere. Because $\hat{X}$ is a suspension, $\hat{X}$ is simply connected, and thus it suffices to show that $\hat{X}$ is a homology sphere. Because $\bar{C}^R_\kappa(\Sigma)$ is contractible, from the Mayer–Vietoris exact sequence of $\left(\bar{C}^R_\kappa(\Sigma)^+, \bar{C}^R_\kappa(\Sigma)^-\right)$, it is easy to see that $\hat{X}$ is an integral homology sphere.

If the $\mathbb{Z}_2$-action is free, $X = \hat{X}/\mathbb{Z}_2$ is homeomorphic to $\mathbb{R}P^n$. Otherwise, X is a simply connected topological manifold (the induced map, $\pi_1(\hat{X}) \to \pi_1(X)$ is an onto map). Again, it suffices to show that X is an integral homology sphere. By the Smith theorem, the $\mathbb{Z}_2$-fixed point set $\hat{X}^{\mathbb{Z}_2}$ is a $\mathbb{Z}_2$-homology sphere. By now we can apply Theorem 3.1 to conclude the claim.

We prove Theorem C by contradiction; assume a sequence $X_i \in \mathcal{A}^r_\kappa(\Sigma)$ such that $\operatorname{vol} X_i > \operatorname{vol} C^R_\kappa(\Sigma) - \epsilon_i$ ($\epsilon_i = i^{-1}$) and none of X_i is homeomorphic to S^n or $\mathbb{R}P^n$. Without loss of generality, we may assume that

$$(X_i, p_i) \xrightarrow{d_{GH}} (X, p) \in \operatorname{Alex}^n(\kappa),$$

where $X_i = \bar{B}_r(p_i)$. By Perelman's stability theorem [Kapovitch 2007; Perelman 1991], X_i is homeomorphic to X for i large. In particular, X is a topological manifold. We claim that $X \in \mathcal{A}^r_\kappa(\Sigma_p)$ satisfies $\operatorname{vol} X = v(\Sigma_p, \kappa, r)$. By the above, we then conclude that X is homeomorphic to S^n or $\mathbb{R}P^n$, and thus X_i is homeomorphic to X for i large, a contradiction.

To see the claim,

$$\operatorname{vol} X = \lim_{i\to\infty} \operatorname{vol} X_i = \lim_{i\to\infty} (\operatorname{vol} C^R_\kappa(\Sigma) - \epsilon_i) = \operatorname{vol} C^R_\kappa(\Sigma).$$

On the other hand, we shall construct a distance-nonincreasing map, $\phi : \Sigma \to \Sigma_p$. Consequently, $\operatorname{vol} \Sigma_p \le \operatorname{vol} \Sigma$ and thus

$$\operatorname{vol} X \le \operatorname{vol} C^R_\kappa(\Sigma_p) \le \operatorname{vol} C^R_\kappa(\Sigma) \le \operatorname{vol} X.$$

Let $A = \{v_i\} \subset \Sigma$ be a countable dense subset and $f_i : (X_i, p_i) \to (X, p)$ a sequence of ϵ_i-Gromov–Hausdorff approximations, $\epsilon_i \to 0$. For v_1, the sequence

$\{f_i(g \exp_{p_i} v)\} \subset X$ contains a converging subsequence

$$f_{i_1}(g \exp_{p_{i_1}} q(v)) \to x_1 \in X.$$

Then $[px_1] = w_1 \in \Sigma_p$ (which may not be unique). We define $\phi(v_1) = w_1$. For v_2 and $\{f_{i_1}\}$, repeating the above, we obtain $w_2 \in \Sigma_p$ and define $\phi(v_2) = w_2$. Iterating this process, we define a map $\phi : A \to \Sigma_p$, $\phi(v_i) = w_i$. It is easy to check that ϕ is distance-nonincreasing and thus ϕ extends uniquely to a distance-nonincreasing map from Σ to Σ_p. □

4. Proof of Theorem D: Pointed Bishop–Gromov relative volume comparison

Assuming the monotonicity in Theorem D, the rigidity part follows by Lemma 4.3 and Theorem 2.1. For $p \in X \in \mathrm{Alex}^n(\kappa)$, let $A^r_R(p)$ (or briefly A^r_R) denote the annulus $\{x \in X : r < |px| \le R\}$, $0 \le r < R$, and let $A^r_R(\Sigma_p)$ (or briefly $\widetilde{A}^r_R$) denote the corresponding annulus in $C_\kappa(\Sigma_p)$. Let B_r denote A^0_r and let $\widetilde{B}_r$ denote $\widetilde{A}^0_r$. Let's recall the following two lemmas.

Lemma 4.1 [Li 2010, Lemma 2.1]. *Let $\Sigma \in \mathrm{Alex}^{n-1}(1)$ and $0 < r \le \pi/\sqrt{\kappa}$. Then*

$$\mathrm{vol}\, C^r_\kappa(\Sigma) = \mathrm{vol}\,\Sigma \int_0^r \mathrm{sn}^{n-1}_\kappa(t)\, dt.$$

Lemma 4.2 [Li 2010, Theorem B]. *Let U be an open subset in $X \in \mathrm{Alex}^n(\kappa)$. Then there is a constant $c(n)$ depending only on n such that*

$$V_{r_n}(\overline{U}) = V_{r_n}(U) = c(n)\,\mathrm{Haus}_n(U) = c(n)\,\mathrm{Haus}_n(\overline{U}),$$

where V_{r_n} and Haus_n represent the n-dimensional rough volume and Hausdorff measure, respectively.

Lemma 4.3. *If the monotonicity in Theorem B holds,*

$$\frac{\mathrm{vol}\, B_r}{\mathrm{vol}\, \widetilde{B}_r} = \frac{\mathrm{vol}\, B_R}{\mathrm{vol}\, \widetilde{B}_R}$$

for some $0 < r < R$ ($R \le \pi/\sqrt{\kappa}$ for $\kappa > 0$) if and only if $\mathrm{vol}\, B_R = \mathrm{vol}\, \widetilde{B}_R$.

Proof. Assume $\mathrm{vol}\, B_R = \mathrm{vol}\, \widetilde{B}_R$. The desired equation follows by the monotonicity:

$$1 = \frac{\mathrm{vol}\, B_R}{\mathrm{vol}\, \widetilde{B}_R} \le \frac{\mathrm{vol}\, B_r}{\mathrm{vol}\, \widetilde{B}_r} \le \lim_{r \ge t \to 0} \frac{\mathrm{vol}\, B_t}{\mathrm{vol}\, \widetilde{B}_t} = 1.$$

Assume

$$\frac{\mathrm{vol}\, B_r}{\mathrm{vol}\, \widetilde{B}_r} = \frac{\mathrm{vol}\, B_R}{\mathrm{vol}\, \widetilde{B}_R}$$

for some $0 < r < R$. Then for any $t < r$,

$$\frac{\mathrm{vol}\, B_t}{\mathrm{vol}\, A_R^r} + \frac{\mathrm{vol}\, A_R^t}{\mathrm{vol}\, A_R^r} = \frac{\mathrm{vol}\, B_R}{\mathrm{vol}\, A_R^r} = \frac{\mathrm{vol}\, \widetilde{B}_R}{\mathrm{vol}\, \widetilde{A}_R^r} = \frac{\mathrm{vol}\, \widetilde{B}_t}{\mathrm{vol}\, \widetilde{A}_R^r} + \frac{\mathrm{vol}\, \widetilde{A}_R^t}{\mathrm{vol}\, \widetilde{A}_R^r}.$$

By the monotonicity, we have

$$\frac{\mathrm{vol}\, A_R^t}{\mathrm{vol}\, A_R^r} \geq \frac{\mathrm{vol}\, \widetilde{A}_R^t}{\mathrm{vol}\, \widetilde{A}_R^r}.$$

Also,

$$\frac{\mathrm{vol}\, B_t}{\mathrm{vol}\, A_R^r} = \frac{\mathrm{vol}\, B_t}{\mathrm{vol}\, A_r^t} \cdot \frac{\mathrm{vol}\, A_r^t}{\mathrm{vol}\, A_R^r} \geq \frac{\mathrm{vol}\, \widetilde{B}_t}{\mathrm{vol}\, \widetilde{A}_r^t} \cdot \frac{\mathrm{vol}\, \widetilde{A}_r^t}{\mathrm{vol}\, \widetilde{A}_R^r} = \frac{\mathrm{vol}\, \widetilde{B}_t}{\mathrm{vol}\, \widetilde{A}_R^r}.$$

Consequently

$$\frac{\mathrm{vol}\, B_t}{\mathrm{vol}\, A_R^r} = \frac{\mathrm{vol}\, \widetilde{B}_t}{\mathrm{vol}\, \widetilde{A}_R^r}, \quad \text{orequivalently,} \quad \frac{\mathrm{vol}\, B_t}{\mathrm{vol}\, \widetilde{B}_t} = \frac{\mathrm{vol}\, A_R^r}{\mathrm{vol}\, \widetilde{A}_R^r}.$$

Letting $t \to 0$, we get $\mathrm{vol}\, A_R^r = \mathrm{vol}\, \widetilde{A}_R^r$. Thus

$$1 \geq \frac{\mathrm{vol}\, B_R}{\mathrm{vol}\, \widetilde{B}_R} \geq \frac{\mathrm{vol}\, A_R^r}{\mathrm{vol}\, \widetilde{A}_R^r} = 1. \qquad \square$$

Now it remains to show the monotonicity in Theorem D. We take an elementary approach by expressing the monotonicity as a form of "Riemann sum" (see (4-5)) and using the Toponogov triangle comparison to bound each term in terms of the desired form (see Corollary 4.6). To achieve this goal, we choose a special infinite partition (see (4-5) and (4-6)).

We start the proof of Theorem D by deriving an equivalent form of the monotonicity. For $0 \leq R_1 < R_2 < R_3$ ($< \pi/\sqrt{\kappa}$ when $\kappa > 0$), and $p \in X$, by Lemma 4.1, the monotonicity has the following integral form:

$$\frac{\mathrm{vol}\, A_{R_3}^{R_1}}{\mathrm{vol}\, A_{R_2}^{R_1}} \leq \frac{\int_{R_1}^{R_3} \mathrm{sn}_\kappa^{n-1}(t)\, dt}{\int_{R_1}^{R_2} \mathrm{sn}_\kappa^{n-1}(t)\, dt},$$

which is equivalent to

$$I_1 = \log\left[\frac{\mathrm{vol}\, A_{R_3}^{R_1}}{\mathrm{vol}\, A_{R_2}^{R_1}}\right] \leq \log\left[\frac{\int_{R_1}^{R_3} \mathrm{sn}_\kappa^{n-1}(t)\, dt}{\int_{R_1}^{R_2} \mathrm{sn}_\kappa^{n-1}(t)\, dt}\right] = I_2. \tag{4-1}$$

Fixing a small $\delta > 0$, let $m = [(R_3 - R_2)/\delta] + 1$, $\Delta = (R_3 - R_2)/m \approx \delta$, and $r_j = R_2 + j\Delta$, $0 \leq j \leq m$. Then

$$A_{R_2}^{R_1} = A_{r_0}^{R_1} \subset A_{r_1}^{R_1} \subset \cdots \subset A_{r_m}^{R_1} = A_{R_3}^{R_1}.$$

Using the Taylor expansion $\log(1/x) = 1 - x + O((1-x)^2)$, we may rewrite the left hand side of (4-1) as

$$I_1 = \sum_{j=1}^{m} \log \frac{\operatorname{vol} A_{R_1}^{r_j}}{\operatorname{vol} A_{R_1}^{r_{j-1}}} = \sum_{j=1}^{m} \left[\left(1 - \frac{\operatorname{vol} A_{R_1}^{r_{j-1}}}{\operatorname{vol} A_{R_1}^{r_j}} \right) + O(\delta^2) \right] \tag{4-2}$$
$$= \sum_{j=1}^{m} \frac{\operatorname{vol} A_{r_{j-1}}^{r_j}}{\operatorname{vol} A_{R_1}^{r_j}} + O(\delta).$$

Let $\phi(r) = \int_{R_1}^{r} \operatorname{sn}_\kappa^{n-1}(t)\,dt$. Then the right hand side of (4-1) can be written as

$$I_2 = \log \frac{\phi(R_3)}{\phi(R_2)} = \int_{R_2}^{R_3} \frac{\phi'(t)}{\phi(t)}\,dt \tag{4-3}$$
$$= \sum_{j=1}^{m} \frac{\phi'(r_j)}{\phi(r_j)} \delta + \tau(\delta)$$
$$= \sum_{j=1}^{m} \frac{\delta \operatorname{sn}_\kappa^{n-1}(r_j)}{\int_{R_1}^{r_j} \operatorname{sn}_\kappa^{n-1}(t)\,dt} + \tau(\delta).$$

Comparing (4-1) to (4-2) and (4-3), it's sufficient to show

$$\frac{\operatorname{vol} A_{r_{j-1}}^{r_j}}{\operatorname{vol} A_{R_1}^{r_j}} \le \frac{\delta \operatorname{sn}_\kappa^{n-1}(r_j)}{\int_{R_1}^{r_j} \operatorname{sn}_\kappa^{n-1}(t)\,dt}. \tag{4-4}$$

We further divide $A_{R_1}^{r_j}$ into thinner annuli: given a monotonic sequence

$$\{a_i\}_{i=1}^{\infty} \subset [0, 1]$$

such that $a_j \to 0$, $\{a_i r_j\}_{i=1}^{\infty}$ is an infinite partition for $[0, r_j]$, and (4-4) is equivalent to

$$\frac{\operatorname{vol} A_{R_1}^{r_j}}{\operatorname{vol} A_{r_{j-1}}^{r_j}} = \sum_{i=1}^{\infty} \frac{\operatorname{vol} A_{a_i r_j}^{a_{i+1} r_j}}{\operatorname{vol} A_{r_{j-1}}^{r_j}} \ge \frac{\int_{R_1}^{r_j} \operatorname{sn}_\kappa^{n-1}(t)\,dt}{\delta \operatorname{sn}_\kappa^{n-1}(r_j)}. \tag{4-5}$$

To show (4-5), we need an estimate for $\operatorname{vol} A_{a_i r_j}^{a_{i+1} r_j} / \operatorname{vol} A_{r_{j-1}}^{r_j}$ from below (see Corollary 4.6). Assume δ is so small that $R - \delta > 0$ and $r - \lambda\delta > 0$. Let $x \in A_{R-\delta}^{R}$. We define a map, $\phi : A_{R-\delta}^{R} \to A_{r-\lambda\delta}^{r}$, where $f(x)$ is the point on a minimal geodesic $[px]$ (if not unique, we pick one of them) such that

$$|pf(x)| = r - \lambda(R - |px|).$$

Because a geodesic in X does not branch, ϕ is well-defined and is injective.

In the proof of Theorem D, the following is a main technical lemma, which asserts that ϕ behaves like a bi-Lipschitz function.

Lemma 4.4. *Let $\delta > 0$ be sufficiently small, $\lambda = (\mathrm{sn}_\kappa\, r/(\mathrm{sn}_\kappa\, R))$, and*

$$\phi : \widetilde{A}^R_{R-\delta} \to \widetilde{A}^r_{r-\lambda\delta}$$

be defined as above. Then

$$c(\kappa,\delta)\,\lambda \le \frac{\mathrm{sn}_\kappa(|\phi(x)\phi(y)|/2)}{\mathrm{sn}_\kappa(|xy|/2)} \le c(\kappa,\delta)^{-1}\lambda,$$

where

$$c(\kappa,\delta) = \begin{cases} 1 & \kappa = 0, \\ 1 - \dfrac{2\delta}{\mathrm{sn}_\kappa\, R + \delta} & \kappa > 0, \\ 1 - \delta\,\dfrac{\cosh_\kappa\, R}{R} & \kappa < 0 \end{cases}.$$

Because the proof of Lemma 4.4 is technical and somewhat tedious, we will delay it to the end of this section.

Lemma 4.5. *Let U and V be two open subsets of $X \in \mathrm{Alex}^n(\kappa)$, and let $\phi : V \to U$ be an injection. If ϕ satisfies $\mathrm{sn}_\kappa(|\phi(x)\phi(y)|/2) \ge c\,\mathrm{sn}_\kappa(|xy|/2)$ for any $x, y \in V$, $\mathrm{vol}\, U \ge c^n\, \mathrm{vol}\, V$, where c is a constant.*

Proof. By Lemma 4.2, it suffices to prove for rough volume. Recall that the n-dimensional rough volume of a subset V is

$$V_{r_n}(V) = \lim_{\epsilon\to 0} \epsilon^n \beta_V(\epsilon),$$

where $\beta_V(\epsilon)$ denotes the number of points in an ϵ-net $\{x_i\}$ on V.

By the assumption, $\{\phi(x_i)\}$ is a $2\,\mathrm{sn}_\kappa^{-1}(c\,\mathrm{sn}_\kappa(\epsilon/2))$-net in U. We get

$$\beta_U\Big(2\,\mathrm{sn}_\kappa^{-1}\Big(c\,\mathrm{sn}_\kappa\,\frac{\epsilon}{2}\Big)\Big) \ge \beta_V(\epsilon),$$

or in another form,

$$\frac{\epsilon^n}{\big(2\,\mathrm{sn}_\kappa^{-1}(c\,\mathrm{sn}_\kappa(\epsilon/2))\big)^n}\Big(2\,\mathrm{sn}_\kappa^{-1}\Big(c\,\mathrm{sn}_\kappa\,\frac{\epsilon}{2}\Big)\Big)^n \beta_U\Big(2\,\mathrm{sn}_\kappa^{-1}\Big(\mathrm{sn}_\kappa\,\frac{\epsilon}{2}\Big)\Big) \ge \epsilon^n\beta_V(\epsilon).$$

Letting $\epsilon \to 0$, we get $(1/c^n)V_{r_n}(U) \ge V_{r_n}(V)$. □

Corollary 4.6. *Let $p \in X \in \mathrm{Alex}^n(\kappa)$, $\delta > 0$ small. Then*

$$\frac{\mathrm{vol}\, A^r_{r-\lambda\delta}}{\mathrm{vol}\, A^R_{R-\delta}} \ge (1 - \tau(\delta))\Big(\frac{\mathrm{sn}_\kappa\, r}{\mathrm{sn}_\kappa\, R}\Big)^n.$$

Proof. Consider the map $\phi : A^R_{R-\delta} \to A^r_{r-\lambda\delta}$ and $\tilde\phi : \widetilde{A}^R_{R-\delta} \to \widetilde{A}^r_{r-\lambda\delta}$ defined as above. For any $x, y \in A^R_{R-\delta}$, take two points $\tilde x, \tilde y \in C_\kappa(\Sigma_p)$ such that $|\tilde o\tilde x| = |px|$,

$|\tilde{o}\tilde{y}| = |py|$, and $|\tilde{x}\tilde{y}| = |xy|$. By condition B (see [Burago et al. 1992]), it's easy to see that $|f(x)f(y)| \geq |\tilde{\phi}(\tilde{x})\tilde{\phi}(\tilde{y})|$. Thus, by Lemma 4.4, we have

$$\mathrm{sn}_\kappa \frac{|f(x)f(y)|}{2} \geq \mathrm{sn}_\kappa \frac{|\tilde{\phi}(\tilde{x})\tilde{\phi}(\tilde{y})|}{2} \geq (1-\tau(\delta))\,\mathrm{sn}_\kappa \frac{|\tilde{x}\tilde{y}|}{2} = (1-\tau(\delta))\,\mathrm{sn}_\kappa \frac{|xy|}{2}.$$

Then we get the desired estimate by Lemma 4.5. □

Proof of the monotonicity in Theorem D. Continuing from the earlier discussion, the proof reduces to verifying (4-5). We now take $\delta > 0$ sufficiently small, and choose the sequence $\{a_i\}_{i=0}^{\infty}$ as

$$a_0 = 1, \quad a_{i+1} = a_i - \frac{\mathrm{sn}_\kappa(a_i r_j)}{r_j\,\mathrm{sn}_\kappa r_j}\,\delta, \quad i = 0, 1, \ldots \tag{4-6}$$

Then

$$0 < a_{i+1} \leq \begin{cases} \Big(1 - \dfrac{\delta}{r_j}\Big) a_i, & \text{if } \kappa \geq 0, \\ \Big(1 - \dfrac{\delta}{\mathrm{sn}_\kappa r_j}\Big) a_i, & \text{if } \kappa < 0, \end{cases}$$

and thus $a_i \to 0$ and is monotonically decreasing. For each $0 \leq i < \infty$ and $0 \leq j \leq m$, consider the map, $\phi : A_{r_j-\delta}^{r_j} \to A_{a_i r_j - \lambda_i \delta}^{a_i r_j} = A_{a_{i+1} r_j}^{a_i r_j}$, with $\lambda_i = \mathrm{sn}_\kappa(a_i r_j)/\mathrm{sn}_\kappa(r_j)$. By Corollary 4.6, we obtain that

$$\frac{\mathrm{vol}\, A_{a_{i+1}r_j}^{a_i r_j})}{\mathrm{vol}\, A_{r_j-\delta}^{r_j}} \geq (1-\tau(\delta)) \left(\frac{\mathrm{sn}_\kappa(a_i r_j)}{\mathrm{sn}_\kappa r_j}\right)^n.$$

Observe that for $\delta \to 0$, $\{a_i\}$ will become more dense, and thus we can take $N_\delta > 0$ such that $a_{N_\delta} r_j \geq R_1$ and $a_{N_\delta} r_j \to R_1$ as $\delta \to 0$. Summing up for $i = 0, 1, \ldots, N_\delta$, we get

$$\begin{aligned} \frac{\mathrm{vol}\, A_{r_j}^{R_1}}{\mathrm{vol}\, A_{r_j-\delta}^{r_j}} &\geq \frac{\sum_{i=0}^{N_\delta} \mathrm{vol}\, A_{a_{i+1}r_j}^{a_i r_j})}{\mathrm{vol}\, A_{r_j-\delta}^{r_j}} \\ &\geq \sum_{i=0}^{N_\delta} (1-\tau(\delta)) \left(\frac{\mathrm{sn}_\kappa(a_i r_j)}{\mathrm{sn}_\kappa r_j}\right)^n \\ &\geq (1-\tau(\delta)) \frac{1}{\delta\,\mathrm{sn}_\kappa^{n-1}(r_j)} \sum_{i=0}^{N_\delta} \mathrm{sn}_\kappa^{n-1}(a_i r_j) \frac{\delta\,\mathrm{sn}_\kappa(a_i r_j)}{\mathrm{sn}_\kappa r_j} \\ &= (1-\tau(\delta)) \frac{1}{\delta\,\mathrm{sn}_\kappa^{n-1}(r_j)} \left(\int_{R_1}^{r_j} \mathrm{sn}_\kappa^{n-1}(t)\,dt + \tau(\delta)\right) \\ &= (1-\tau(\delta)) \frac{\int_{R_1}^{r_j} \mathrm{sn}_\kappa^{n-1}(t)\,dt}{\delta\,\mathrm{sn}_\kappa^{n-1}(r_j)}, \end{aligned}$$

or the following equivalent form:

$$\frac{\operatorname{vol} A_{r_j-\delta}^{r_j}}{\operatorname{vol} A_{r_j}^{R_1}} \le (1+\tau(\delta))\,\frac{\delta\,\operatorname{sn}_\kappa^{n-1}(r_j)}{\int_{R_1}^{r_j}\operatorname{sn}_\kappa^{n-1}(t)\,dt}.$$

Summing up for all j and together with (4-2) and (4-3), we get

$$I_1 + O(\delta) \le (1+\tau(\delta))I_2 + \tau(\delta).$$

Letting $\delta \to 0$, we get the desired inequality. □

The rest of this section is devoted to a proof of Lemma 4.4. The following are some properties used in the proof.

Lemma 4.7. (1) *For* $\lambda \in [0, 1]$ *and* $x \in [0, \pi]$, $\sin\lambda x \ge \lambda \sin x$.

(2) *For* $\lambda \in [0, 1]$ *and* $x \ge 0$, $\sinh\lambda x \le \lambda\sinh x$.

(3) *For* $\lambda \ge 0$ *and* $x \ge 0$, $\sin\lambda x/(\lambda\sin x) \ge 1-(\lambda x)^2/6$.

(4) *For* $\lambda \ge 0$ *and* $x \ge 0$, $\sinh\lambda x/(\lambda\sinh x) \ge x/\sinh x \ge 1-x$.

(5) *Let* $\triangle pab$ *be a triangle in* S_κ^2. *The cosine law can be written as*

$$sn_\kappa^2\frac{|ab|}{2} = \operatorname{sn}_\kappa^2\frac{|pa|-|pb|}{2} + \sin^2\frac{\measuredangle apb}{2}\operatorname{sn}_\kappa|pa|\operatorname{sn}_\kappa|pb|.$$

Proof.

(1) Let $h(x) = \sin\lambda x - \lambda\sin x$. Then

$$h'(x) = \lambda\cos\lambda x - \lambda\cos x = \lambda(\cos\lambda x - \cos x) \ge 0,$$

since $0 \le \lambda x \le x \le \pi$.

(2) Let $h(x) = \sinh\lambda x - \lambda\sinh x$. Then

$$h'(x) = \lambda\cosh\lambda x - \lambda\cosh x = \lambda(\cosh\lambda x - \cosh x) \le 0,$$

since $0 \le \lambda x \le x$.

(3) For $x > 0$, one can show that $x \ge \sin x \ge x - x^3/6$. Then

$$\frac{\sin\lambda x}{\lambda\sin x} \ge \frac{\lambda x-(\lambda x)^3/6}{\lambda x} = 1-(\lambda x)^2/6.$$

(4) The first equality is easy to see through $\sinh\lambda x \ge \lambda x$. Obviously, the second equality is true for $x \ge 1$. For $0 < x < 1$,

$$\sinh x = x + \frac{x^3}{6} + \cdots \le x(1+x+x^2+\cdots) = \frac{x}{1-x}.$$

(5) Follows by trigonometric metric identities. □

Proof of Lemma 4.4. By scaling, we only need to check for $\kappa = 1, -1$ and $\kappa = 0$.

Case 1 ($\kappa = 1$). Noting that

$$\frac{|px'|-|py'|}{|px|-|py|} = \frac{\lambda(|px|-|py|)}{|px|-|py|} = \lambda,$$

by Lemma 4.7(3) and $0 \le \big||px|-|py|\big| \le \delta < \frac{1}{2}\sin R$, we have

$$\begin{aligned}\sin\frac{\big||px'|-|py'|\big|}{2} &= \sin\Big(\lambda\,\frac{\big||px|-|py|\big|}{2}\Big)\\ &\ge \Big(1-\frac{(\lambda\delta)^2}{6}\Big)\lambda\sin\frac{\big||px|-|py|\big|}{2}\\ &\ge \Big(1-\frac{\delta^2}{6\sin^2 R}\Big)\lambda\sin\frac{\big||px|-|py|\big|}{2}\\ &\ge \Big(1-\frac{2\delta}{\sin R+\delta}\Big)\lambda\sin\frac{\big||px|-|py|\big|}{2}\\ &= \tau_1\lambda\sin\frac{\big||px|-|py|\big|}{2}.\end{aligned}$$

Thus

$$\tau_1\lambda \le \frac{\sin(\big||px'|-|py'|\big|/2)}{\sin(\big||px|-|py|\big|/2)} \le \frac{\lambda\big||px|-|py|\big|/2}{\sin(\big||px|-|py|\big|/2)} \le \lambda\,\frac{\delta}{\sin\delta} \le \tau_1^{-1}\lambda. \tag{4-7}$$

For any $x \in \widetilde{A}^R_{R-\delta}$, by Lemma 4.7(1), we have

$$\sin|px'| \ge \frac{|px'|}{r}\sin r \ge \frac{r-\lambda\delta}{r}\sin r = \frac{r-(\sin r/\sin R)\,\delta}{r}\sin r \ge \Big(1-\frac{\delta}{\sin R}\Big)\sin r,$$

which, together with

$$\sin|px'|-\sin r = 2\sin\frac{|px'|-r}{2}\cos\frac{|px'|+r}{2} \le r-|px'| \le \lambda\delta,$$

gives us

$$\Big(1-\frac{\delta}{\sin R}\Big)\sin r \le \sin|px'| \le \sin r+\lambda\delta = \Big(1+\frac{\delta}{\sin R}\Big)\sin r.$$

Similarly,

$$\sin|px| \ge \frac{|px|}{R}\sin R \ge \frac{R-\delta}{R}\sin R \ge \Big(1-\frac{\delta}{\sin R}\Big)\sin R$$

and

$$\sin|px|-\sin R = 2\sin\frac{|px|-R}{2}\cos\frac{|px|+R}{2} \le R-|px| \le \delta,$$

hence

$$\Big(1-\frac{\delta}{\sin R}\Big)\sin R \le \sin|px| \le \sin R+\delta = \Big(1+\frac{\delta}{\sin R}\Big)\sin R.$$

So

$$c_1 \frac{\sin r}{\sin R} \le \frac{\sin|px'|}{\sin|px|} \le c_1^{-1} \frac{\sin r}{\sin R}. \tag{4-8}$$

Let $\theta = \measuredangle xpy$. Since $|xy|/2 \le \pi/2$, by the cosine law and inequalities (4-7), (4-8),

$$\begin{aligned} c_1^2\lambda^2 &\le \frac{\sin^2(|x'y'|/2)}{\sin^2(|xy|/2)} \\ &= \frac{\sin^2\big((|px'|-|py'|)/2\big)+\sin^2(\theta/2)\sin|px'|\sin|py'|}{\sin^2\big((|px|-|py|)/2\big)+\sin^2(\theta/2)\sin|px|\sin|py|} \le c_1^{-2}\lambda^2. \end{aligned}$$

Case 2 ($\kappa = -1$). By Lemma 4.7(2),

$$\lambda\delta = \frac{\sinh r}{\sinh R}\cdot\frac{R}{\cosh R} < \frac{r}{R}\,R = r,$$

which, together with Lemma 4.7(4), gives

$$\lambda \ge \frac{\sinh\big(\big||px'|-|py'|\big|/2\big)}{\sinh\big(\big||px|-|py|\big|/2\big)} = \frac{\sinh\big(\lambda\,\big||px|-|py|\big|/2\big)}{\sinh\big(\big||px|-|py|\big|/2\big)} \ge (1-\delta)\lambda \ge c_{-1}\lambda, \tag{4-9}$$

since

$$\frac{\cosh R}{R} \ge \frac{1+R^2/2}{R} > 1.$$

If $\delta < R/\cosh R < R$, we have

$$\frac{\lambda\delta}{2r} < \frac{r}{R}\cdot\frac{\delta}{2r} = \frac{\delta}{2R} < 1.$$

Hence we can apply Lemma 4.7(2) with $\lambda = (\sinh r)/\sinh R \le r/R$, to get

$$\frac{\sinh r - \sinh(r-\lambda\delta)}{\sinh r} \le \frac{2\sinh(\lambda\delta/2)\cosh r}{\sinh r} \le \frac{\lambda\delta}{r}\cosh r \le \frac{\delta\cosh R}{R}.$$

Thus

$$\sinh(r-\lambda\delta) \ge \Big(1-\delta\frac{\cosh R}{R}\Big)\sinh r.$$

For $x' \in \widetilde{A}^r_{r-\lambda\delta}$,

$$\Big(1-\delta\frac{\cosh R}{R}\Big)\sinh r \le \sinh(r-\lambda\delta) \le \sinh|px'| \le \sinh r.$$

For $x \in \widetilde{A}^R_{R-\lambda\delta}$,

$$\frac{\sinh R - \sinh(R-\delta)}{\sinh R} \le \frac{2\sinh(\delta/2)\cosh R}{\sinh R} \le \frac{\delta\cosh R}{R},$$

and

$$\Big(1-\delta\frac{\cosh R}{R}\Big)\sinh R \le \sinh(R-\lambda\delta) \le \sinh|px| \le \sinh R.$$

Then

$$c_{-1}\frac{\sinh r}{\sinh R} \le \frac{\sinh|px'|}{\sinh|px|} \le c_{-1}^{-1}\frac{\sinh r}{\sinh R}. \tag{4-10}$$

By inequalities (4-9), (4-10), and the cosine law, we get

$$\begin{aligned} c_{-1}^2\lambda^2 &\le \frac{\sinh^2(|x'y'|/2)}{\sinh^2(|xy|/2)} \\ &= \frac{\sinh^2((|px'|-|py'|)/2)+\sin^2(\theta/2)\sinh|px'|\sinh|py'|}{\sinh^2((|px|-|py|)/2)+\sin^2(\theta/2)\sinh|px|\sinh|py|} \le c_{-1}^{-2}\lambda^2. \end{aligned}$$

Case 3 ($\kappa = 0$). This is straightforward. □

References

[Bredon 1972] G. E. Bredon, *Introduction to compact transformation groups*, Pure and Applied Mathematics **46**, Academic, New York, 1972. MR 54 #1265 Zbl 0246.57017

[Burago et al. 1992] Y. Burago, M. Gromov, and G. Perelman, "A. D. Aleksandrov spaces with curvatures bounded below", *Uspekhi Mat. Nauk* **47**:2(284) (1992), 3–51. In Russian; translated in *Russian Math. Surveys* **47**:2 (1992), 1-58. MR 93m:53035 Zbl 0802.53018

[Burago et al. 2001] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics **33**, American Mathematical Society, Providence, RI, 2001. MR 1835418 (2002e:53053) Zbl 0981.51016

[Grove and Petersen 1992] K. Grove and P. Petersen, V, "Volume comparison à la Aleksandrov", *Acta Math.* **169**:1-2 (1992), 131–151. MR 93j:53057

[Kapovitch 2002] V. Kapovitch, "Regularity of limits of noncollapsing sequences of manifolds", *Geom. Funct. Anal.* **12**:1 (2002), 121–137. MR 2003b:53043 Zbl 1013.53046

[Kapovitch 2007] V. Kapovitch, "Perelman's stability theorem", pp. 103–136 in *Metric and comparison geometry*, edited by J. Cheeger and K. Grove, Surv. Differ. Geom. **11**, International Press, Somerville, MA, 2007. MR 2009g:53057 Zbl 1151.53038

[Li 2010] N. Li, *Controlled geometry via volumes on Alexandrov spaces*, thesis, Rutgers University, 2010, Available at http://wwwlib.umi.com/dissertations/fullcit/3434721. MR 2827253

[Liu and Shen 1994] Z.-d. Liu and Z. M. Shen, "On the Betti numbers of Alexandrov spaces", *Ann. Global Anal. Geom.* **12**:2 (1994), 123–133. MR 95b:53055 Zbl 0824.53041

[Perelman 1991] G. Perelman, "Alexandrov spaces with curvatures bounded from below, II", preprint, 1991.

[Petrunin 1997] A. Petrunin, "Applications of quasigeodesics and gradient curves", pp. 203–219 in *Comparison geometry* (Berkeley, CA, 1993–94), edited by K. Grove and P. Petersen, Math. Sci. Res. Inst. Publ. **30**, Cambridge University Press, 1997. MR 98m:53061 Zbl 0892.53026

[Petrunin 1998] A. Petrunin, "Parallel transportation for Alexandrov space with curvature bounded below", *Geom. Funct. Anal.* **8**:1 (1998), 123–148. MR 98j:53048 Zbl 0903.53045

[Petrunin 2007] A. Petrunin, "Semiconcave functions in Alexandrov's geometry", pp. 137–201 in *Metric and comparison geometry*, edited by J. Cheeger and K. Grove, Surv. Differ. Geom. **11**, International Press, Somerville, MA, 2007. MR 2010a:53052 Zbl 1166.53001

[Shteingold 1994] S. Shteingold, "Volume comparison for Alexandrov spaces", *Indiana Univ. Math. J.* **43**:4 (1994), 1349–1357. MR 96c:53065 Zbl 0833.53039

[Wu 1997] J.-Y. Wu, "Topological regularity theorems for Alexandrov spaces", *J. Math. Soc. Japan* **49**:4 (1997), 741–757. MR 1466363 (98h:53071) Zbl 0915.53021

Received December 10, 2011.

NAN LI
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NOTRE DAME
255 HURLEY HALL
NOTRE DAME, IN 46556
UNITED STATES
lilinanan@gmail.com

XIAOCHUN RONG
DEPARTMENT OF MATHEMATICS
CAPITAL NORMAL UNIVERSITY
105 XI SAN HUAN BEI RD.
HAIDIAN DISTRICT
BEIJING 100048
CHINA
and
DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
HILL CENTER FOR THE MATHEMATICAL SCIENCES
110 FRELINGHUYSEN RD
PISCATAWAY, NEW JERSEY 08854
UNITED STATES
rong@math.rutgers.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 259, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.259.421

PROPERNESS, CAUCHY INDIVISIBILITY AND THE WEIL COMPLETION OF A GROUP OF ISOMETRIES

ANTONIOS MANOUSSOS AND POLYCHRONIS STRANTZALOS

Investigating the impact of local compactness and connectedness in the theory of proper actions on locally compact and connected spaces, we introduce a new class of isometric actions on separable metric spaces called *Cauchy-indivisible actions*. The new class coincides with that of proper actions on locally compact metric spaces, without assuming connectivity, and, as examples show, may be different in general. In order to provide some basic theory for this new class of actions, we embed a Cauchy-indivisible action in a proper action of a semigroup in the completion of the underlying space. We show that, if this semigroup is a group, there are remarkable connections between Cauchy indivisibility and properness, while the original group has a Weil completion and vice versa. Further connections in this direction establish a relation between *Borel sections* for Cauchy-indivisible actions and *fundamental sets* for proper actions. Some open questions are added.

1. Introduction

In the paper at hand, having in mind the fruitful theory of proper transformation groups on locally compact and connected spaces, we propose an analogous class of actions, not necessarily proper, without assuming local compactness and connectedness of the underlying spaces. So, we introduce a new, rather natural, class of metric actions on separable (not necessarily connected) metric spaces called *Cauchy-indivisible*. Note that isometric actions constitute nowadays an important part of the theory of proper actions and that the group of isometries of a locally compact and connected metric space acts properly on it.

As the following definition shows, Cauchy-indivisible actions are characterized by an *isotropic* behavior of divergent nets of the acting group with respect to the basic metric notion of a Cauchy sequence. Recall that $z_i \to \infty$ in Z means that the net $\{z_i\}$ does not have any convergent subnet in the space Z.

During this research the first author was fully supported by SFB 701 "Spektrale Strukturen und Topologische Methoden in der Mathematik" at the University of Bielefeld, Germany.
MSC2010: primary 37B05, 54H20; secondary 54H15.
Keywords: proper action, Weil completion, Cauchy indivisibility, Borel section, fundamental set.

Definition 1.1. Let (G, X) be a continuous action of a topological group G on a metric space X. The action is said to be *Cauchy-indivisible* if the following holds: If $\{g_i\}$ is a net in G such that $g_i \to \infty$ in G and $\{g_i x\}$ is a Cauchy net in X for some $x \in X$ then $\{g_i x\}$ is a Cauchy net for every $x \in X$.

It turns out that a Cauchy-indivisible action on a locally compact or complete metric space is proper and vice versa (see Section 3), and that in general the two notions may differ (see Section 4). In both cases the underlying space is not assumed to be connected. The omission of this assumption in the locally compact case, as well the omission of local compactness in the main part of the paper at hand is an advantage coming from the fact that Cauchy indivisibility essentially reflects the *global* character of self-maps of X compared with the *local* properties or the *connectedness* of the underlying space. So we can generalize the framework of proper actions and go beyond, provided that this new framework leads

(a) to interesting results in the non locally compact case, and

(b) enables a better understanding of proper actions on locally compact spaces.

Concerning requirement (b) we note that in Theorem 3.3 we give an answer to the open question of characterizing proper actions on nonconnected locally compact metric spaces and in Theorem 7.4 we establish an interconnection between Borel sections (which occur in Cauchy-indivisible actions on separable spaces, see Proposition 7.1) and fundamental sets that characterize proper isometric actions. Recall that a *section* of an action (G, X) is a subset of X which contains only one point from each orbit. A *Borel section* is a section that is a Borel subset of X (useful, for example, in measure theory).

Theorem 7.4. *Let G be a group which acts properly on a locally compact space X, and suppose that the orbit space $G\backslash X$ is paracompact. Let S be a section for the action (G, X).*

(i) *For every open neighborhood U of S we can construct a closed fundamental set F_c and an open fundamental set F_o such that $F_c \subset F_o \subset U$.*

(ii) *If, in addition, (X, d) is a separable metric space, in which case the action (G, X) is Cauchy-indivisible, then there exists a Borel section S_B, which is also a fundamental set, such that $S_B \subset F_c \subset F_o \subset U$.*

Note that S_B in (ii) of the above theorem is a "minimal" fundamental set, because of its construction, and as such may lead to applications.

The new notion of "like properness" seems to be suitable for structure theorems, as our first results indicate. Concerning requirement (a) above, in Section 5, which is the main part of the paper at hand, we consider a separable metric space (X, d) such that the natural evaluation action of the group of isometries $\mathrm{Iso}(X)$ on X is Cauchy-indivisible. Let $\widehat{X}$ denote the completion of (X, d) and let E be the *Ellis* *semigroup* of the lifted group $\widehat{\mathrm{Iso}(X)}$ in $C(\widehat{X}, \widehat{X})$, that is, the pointwise closure of $\widehat{\mathrm{Iso}(X)}$ in $C(\widehat{X}, \widehat{X})$. Let

$$
\begin{aligned}
H &= \{h \in C(\widehat{X}, \widehat{X}) \mid \text{there exists a sequence } \{g_n\} \subset \mathrm{Iso}(X) \\
&\qquad\qquad \text{with } g_n \to \infty \text{ in } \mathrm{Iso}(X) \text{ and } \hat{g}_n \to h \text{ in } C(\widehat{X}, \widehat{X})\},\\
X_l &= \{hx \mid h \in H, x \in X\},\\
X_p &= \{hx \mid h \in H \cap \mathrm{Iso}(\widehat{X}), x \in X\}.
\end{aligned}
$$

With this notation and the previously mentioned assumptions, among other results we show the following.

Theorem 5.13. *The set $X \cup X_p$ is the maximal subset of $X \cup X_l$ that contains X such that the map*

$$\omega : E \times (X \cup X_p) \to (X \cup X_p) \times \widehat{X},$$

with $\omega(f, y) = (y, fy)$, $f \in E$, and where $y \in X \cup X_p$ is proper.

The interest in this theorem lies in the fact that an action (G, X) is proper if the map $G \times X \to X \times X$ defined by $(g, x) \mapsto (x, gx)$ is proper, see [Bourbaki 1966a, Definition 1, p. 250].

We recall that a topological group has a *Weil completion* with respect to the uniformity of pointwise convergence if it can be embedded densely in a complete group with respect to its left uniform structure.

Proposition 5.18. *The following are equivalent*:

(i) *The map $\omega : E \times (X \cup X_l) \to (X \cup X_l) \times \widehat{X}$ is proper.*

(ii) *E is a group* (*precisely a closed subgroup of* $\mathrm{Iso}(\widehat{X})$).

(iii) $\mathrm{Iso}(X)$ *has a Weil completion.*

Corollary 5.20. *If E is a group the action* $(\mathrm{Iso}(X), X)$ *is embedded densely in the proper action* $(E, X \cup X_l)$ *such that the following equivariant diagram commutes*:

$$
\begin{array}{ccc}
(\mathrm{Iso}(X), X) & \longrightarrow & X \\
\downarrow & & \downarrow \\
(E, X \cup X_l) & \longrightarrow & \widehat{X},
\end{array}
$$

where $X \to X \cup X_l$ is the inclusion map and the map $\mathrm{Iso}(X) \to E$ *is defined by $g \mapsto \hat{g}$ for every $g \in \mathrm{Iso}(X)$. By "densely" we mean that X is dense in $X \cup X_l$ and $\widehat{\mathrm{Iso}(X)}$ is dense in E.*

The above result may lead to further structure theorems, see Question 5.21.

Proposition 7.1. *If the Ellis semigroup E is a group then the action $(E, X \cup X_l)$ has a Borel section.*

As Theorem 7.4(ii) mentioned above indicates, the notion of a Borel section, which according to the above result is a feature of the Cauchy-indivisible actions on separable metric spaces, is remarkably related to that of a fundamental set in the locally compact case and may be, similarly, used for structural theorems. So, it is interesting to ask whether the existing Borel section for the action $(E, X \cup X_l)$ can be reduced to a Borel section for the initial action $(\mathrm{Iso}(X), X)$, see Question 7.7.

In order to indicate or to exclude possible directions for further investigation concerning Cauchy-indivisible actions, we study various examples, see Examples 5.14, 5.17, 5.23, and 7.3. Among them, an example that may be of independent interest is the following (see Section 6): Consider the action $(\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z})), \mathrm{Iso}(\mathbb{Z}))$, where $\mathbb{Z}$ is the discrete space of the integers, with suitable metrics on the acting group $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ and the underlying space $\mathrm{Iso}(\mathbb{Z})$. We show that this action is proper and Cauchy-indivisible while the Ellis semigroup is not a group and $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ has no Weil completion.

2. Basic notions and notation

For what follows, in addition to the notation established in the introduction, (X, d) will denote a metric space with metric d and $\mathrm{Iso}(X)$ will denote its group of (surjective) isometries of X endowed with the topology of pointwise convergence. With this topology $\mathrm{Iso}(X)$ is a topological group [Bourbaki 1966b, Chapter X, §3.5, Corollary]. Let $(\widehat{X}, \hat{d})$ stand for the completion of (X, d). For a Cauchy sequence $\{x_n\}$ in X let $[x_n] \in \widehat{X}$ denote the limit point of $\{x_n\}$ in $\widehat{X}$. We denote by $\hat{g}$ and $\widehat{\mathrm{Iso}(X)}$ the lift of $g \in \mathrm{Iso}(X)$ and the lift of the group $\mathrm{Iso}(X)$, respectively, in $C(\widehat{X}, \widehat{X})$, the space of the continuous self-maps of $\widehat{X}$ (which is considered with the topology of pointwise convergence).

A continuous action of a topological group G on a topological space X is a continuous map $G \times X \to X$ with $(g, x) \mapsto gx$, $g \in G$, $x \in X$ such that $(e, g) \mapsto x$, for every $x \in X$ where e denotes the unit element of G, and $(h, (g, x)) \mapsto (hg)x$ for every $h, g \in G$, and $x \in X$. When the action map is known we will denote the action simply by (G, X). Let $U \subset X$, then $GU := \{gx \mid g \in G, x \in U\}$. Especially, if $U = \{x\}$ then the set $Gx := G\{x\}$ is called the orbit of $x \in X$ under G. The subgroup $G_x := \{g \in G \mid gx = x\}$ of G is called the isotropy group of $x \in X$. The natural evaluation action of $\mathrm{Iso}(X)$ on X (denoted by $(\mathrm{Iso}(X), X)$) is the map $\mathrm{Iso}(X) \times X \to X$ with $(g, x) \mapsto g(x)$, $g \in \mathrm{Iso}(X)$, and $x \in X$. If we endow $\mathrm{Iso}(X)$ with the topology of pointwise convergence this action is always continuous. As usual, $S(x, \varepsilon)$ will denote the open ball centered at x with radius $\varepsilon > 0$.

Definition 2.1. A continuous action (G, X) is (equivalently to the Bourbaki definition) proper if $J(x) = \varnothing$, for every $x \in X$, where

$$J(x) = \{ y \in X \mid \text{there exist nets } \{x_i\} \text{ in } X, \text{ and } \{g_i\} \text{ in } G \text{ with } g_i \to \infty, \lim x_i = x \text{ and } \lim g_i x_i = y \}$$

denotes the extended (prolongational) limit set of $x \in X$.

It is easily seen that in the special case of actions by isometries $J(x) = L(x)$ holds for every $x \in X$, where

$$L(x) = \{ y \in X \mid \text{there exists a net } \{g_i\} \text{ in } G \text{ with } g_i \to \infty \text{ and } \lim g_i x = y \}$$

denotes the limit set of $x \in X$ under the action of G on X. Hence an action by isometries (G, X) is proper if and only if $L(x) = \varnothing$ for every $x \in X$.

3. Cauchy indivisibility and proper actions on locally compact metric spaces

In this section we show that for group actions on locally compact metric spaces the notions of properness and Cauchy indivisibility coincide. We start with the following easily proved observation.

Lemma 3.1. *Let (X, d) be a locally compact metric space and $\{g_i\} \subset \mathrm{Iso}(X)$ be a net such that $\{g_i x\}$ is a Cauchy net for some $x \in X$. Then there exists a point $y \in X$ such that $g_i x \to y$.*

Proposition 3.2. *Let (X, d) be a locally compact metric space. The action* $(\mathrm{Iso}(X), X)$ *is proper if and only if it is Cauchy-indivisible.*

Proof. Assume that $(\mathrm{Iso}(X), X)$ is Cauchy-indivisible. We will show that the limit sets $L(x)$ are empty for every $x \in X$. Assume the contrary, that is, there exist a net $\{g_i\}$ in $\mathrm{Iso}(X)$ and $x, y \in X$ such that $g_i \to \infty$ and $g_i x \to y$. We will show that $g_i \to h$ for some $h \in \mathrm{Iso}(X)$, which is a contradiction of the assumption $g_i \to \infty$. Since $(\mathrm{Iso}(X), X)$ is Cauchy-indivisible then $\{g_i x\}$ is a Cauchy net, for every $x \in X$. Therefore, by the previous lemma, there is a map $h : X \to X$ defined by $h(x) := y$ such that $g_i \to h$ pointwise on X and h preserves the metric d. Observe that $g_i^{-1} y \to x$ since $d(g_i^{-1} y, x) = d(y, g_i x)$. Applying Cauchy indivisibility for the action $(\mathrm{Iso}(X), X)$ and the previous lemma again, we conclude that there exists a map $f : X \to X$ such that $g_i^{-1} \to f$ pointwise on X and f preserves the metric d. Obviously f is the inverse map of h, hence $h \in \mathrm{Iso}(X)$.

The converse implication follows easily in a similar way. □

If X is locally compact and G acts properly on X (hence G is a locally compact group), it is well known, see, for example, [Koszul 1965], that there exists a G-invariant compatible metric on X. Compatible means that this metric induces the topology of X. Hence, the previous proposition states the following result that characterizes the properness of actions on locally compact metric spaces independently of the connectedness of the underlying space.

Theorem 3.3. *Let (X, d) be a locally compact metric space. An action (G, X) is proper if and only if it is Cauchy-indivisible.*

Remark 3.4. The previous theorem also holds, and can be similarly proved, if we replace the full group of isometries of X by a closed subgroup of it or if we replace the local compactness of X by completeness.

4. Cauchy indivisibility vs. properness

In this section we provide examples showing that Cauchy indivisibility and properness are distinct notions for isometric actions on separable and non locally compact metric spaces. We also provide some criteria for the coexistence of Cauchy-indivisible and proper actions on the basis of the dynamical behavior of the lifting of the action $(\mathrm{Iso}(X), X)$ in the completion of the underlying space.

Remark 4.1. The example in Section 6 shows that the two notions may coexist also in the case when X is neither locally compact nor complete.

The following example shows that the action $(\mathrm{Iso}(X), X)$ can be proper and not Cauchy-indivisible.

Example 4.2. Let X be the set $\mathbb{Q}$ of the rational numbers endowed with the Euclidean metric. It is easy to see that the action $(\mathrm{Iso}(X), X)$ is proper. Take a sequence of rational numbers $\{q_n\}$ such that $q_n \to a$, where a is an irrational. Let $\{g_n\} \subset \mathrm{Iso}(X)$ with $g_n x := (-1)^n x + q_n$ for every $x \in X$, then $g_n \to \infty$ in $\mathrm{Iso}(X)$. Since $g_n 0 = q_n$ for every $n \in \mathbb{N}$, the sequence $\{g_n 0\}$ is Cauchy. But for $x \neq 0$ the sequence $\{g_n x\}$ has two limit points in $\mathbb{R}$ and hence cannot be a Cauchy sequence.

The next example shows that the action $(\mathrm{Iso}(X), X)$ can be Cauchy-indivisible and not proper.

Example 4.3. Let X be the set $\mathbb{Q} + \sqrt{2}\mathbb{N}$ endowed with the Euclidean metric. Its group of isometries is $\mathbb{Q}$ acting by translations (reflections are excluded because of the addend $\sqrt{2}\mathbb{N}$). Therefore, $(\mathrm{Iso}(X), X)$ is Cauchy-indivisible. However, the action $(\mathrm{Iso}(X), X)$ is not proper. To see that take a sequence of rational numbers $\{q_n\}$ such that $q_n \to \sqrt{2}$. Let $\{g_n\} \subset \mathrm{Iso}(X)$ with $g_n x := x + q_n$. Observe that $g_n^{-1}\sqrt{2} \to 0 \notin X$. Therefore $g_n \to \infty$ in $\mathrm{Iso}(X)$. Since $g_n\sqrt{2} \to 2\sqrt{2} \in X$ the limit set $L(\sqrt{2})$ is not empty, so the action $(\mathrm{Iso}(X), X)$ is not proper.

Motivated by these examples we give necessary and sufficient conditions for a Cauchy-indivisible action $(\mathrm{Iso}(X), X)$ to be proper and vice versa:

Proposition 4.4. *Let $\mathrm{Iso}(X)$ be Cauchy-indivisible. The following are equivalent*:

(i) *The action* $(\mathrm{Iso}(X), X)$ *is proper.*

(ii) *If h is in the pointwise closure of* $\widehat{\mathrm{Iso}(X)}$ *in* $C(\widehat{X}, \widehat{X})$ *then either* $h(X) \subset X$ *or* $h(X) \subset \widehat{X} \setminus X$.

Proof. Assume that the action $(\mathrm{Iso}(X), X)$ is proper and h is in the pointwise closure of $\widehat{\mathrm{Iso}(X)}$ in $C(\widehat{X}, \widehat{X})$. Then there is a net $\{\hat{g}_i\}$ in $\widehat{\mathrm{Iso}(X)}$ such that $\hat{g}_i \to h$ pointwise in $\widehat{X}$. If $h(X) \cap X \neq \varnothing$ then there is some $x \in X$ such that $\hat{g}_i x \to hx \in X$. Since the action $(\mathrm{Iso}(X), X)$ is proper the net $\{g_i\}$ has a convergent subnet in $\mathrm{Iso}(X)$. Then it is easy to see that $h \in \widehat{\mathrm{Iso}(X)}$, hence $h(X) \subset X$.

Assume now that condition (ii) holds. We will show that the limit sets $L(x)$ are empty for every $x \in X$, hence the action $(\mathrm{Iso}(X), X)$ is proper. We will proceed by contradiction. Assume that there exist $x, y \in X$ and a net $\{g_i\}$ in $\mathrm{Iso}(X)$ with $g_i x \to y$ and $g_i \to \infty$ in $\mathrm{Iso}(X)$. Since $\{g_i x\}$ is a Cauchy net in X and $\mathrm{Iso}(X)$ is Cauchy-indivisible then $\{g_i x\}$ is a Cauchy net for every $x \in X$, hence $\{g_i x\}$ converges in $\widehat{X}$ for every $x \in X$. So, we can define a map $h : X \to \widehat{X}$ by letting $hx := \lim \hat{g}_i x$. It is easy to see that h preserves the metric $\hat{d}$ on X. Thus, if $w \in \widehat{X}$ and $\{x_n\} \subset X$ is a sequence in X such that $x_n \to w$ in $\widehat{X}$ then $\{hx_n\}$ is a Cauchy sequence in X, hence it converges to a point in $\widehat{X}$ which is independent of the choice of the sequence $\{x_n\}$. Then, by [Bourbaki 1966a, Chapter I, §8.5, Theorem 1], the map $h : X \to \widehat{X}$ has a unique continuous extension on $\widehat{X}$. It is easy to see that $\hat{g}_i \to h$ pointwise on $\widehat{X}$, thus h is in the pointwise closure of $\widehat{\mathrm{Iso}(X)}$ in $C(\widehat{X}, \widehat{X})$. Since $g_i x \to y$ then $hx = y$ where $x, y \in X$. So using our hypothesis $h(X) \subset X$. Since g_i preserves the metric d then $g_i^{-1} y \to x$. Using the same arguments as before we have that $h \in \widehat{\mathrm{Iso}(X)}$ hence the net $\{g_i\}$ converges in $\mathrm{Iso}(X)$, a contradiction of the assumption $g_i \to \infty$ in $\mathrm{Iso}(X)$. □

Proposition 4.5. *Assume that* $(\mathrm{Iso}(X), X)$ *is a proper action. The following are equivalent*:

(i) $\mathrm{Iso}(X)$ *is Cauchy-indivisible.*

(ii) *Let* $\{g_i\} \subset \mathrm{Iso}(X)$ *a net with* $g_i \to \infty$ *and* $\{g_i x\}$ *be a Cauchy net for some* $x \in X$. *If* $y \in X$ *then the net* $\{g_i y\}$ *cannot have more than one limit point in the completion* $\widehat{X}$ *of* X.

Proof. The direction from (i) to (ii) is trivial. If the converse implication does not hold, then there is a Cauchy net $\{g_i x\}$ such that there is $y \in X$, an $\varepsilon > 0$, and subnets $\{g_{i_k} y\}$ and $\{g_{i_l} y\}$ of $\{g_i y\}$ such that $d(g_{i_k} y, g_{i_l} y) \geq \varepsilon$ for every index k, l. Since $\{g_i x\}$ is a Cauchy net in X then we may assume that $d(g_{i_k} x, g_{i_l} x) \to 0$. Hence, $d(g_{i_k}^{-1} g_{i_l} x, x) \to 0$. We can define a new net $\{h_{i,j}\} \subset \mathrm{Iso}(X)$ by letting $h_{i,j} := g_j^{-1} g_i$ for every pair of indices (i, j), with direction defined by $(i_1, j_1) \leq (i_2, j_2)$ if and only if $i_1 \leq i_2$ and $j_1 \leq j_2$. Therefore, $h_{i_k, i_l} x \to x$. Since $(\mathrm{Iso}(X), X)$ is proper there is a subnet $\{h_{i_{k_m}, i_{l_m}}\}$ and some $g \in \mathrm{Iso}(X)$ such that $h_{i_{k_m}, i_{l_m}} \to g$. Hence

$\{h_{i_{k_m}, i_{l_m}} y\}$ is a Cauchy net in X, therefore for every $\varepsilon' > 0$ there exists an index m_0 such that

$$d(g_{i_{k_m}}^{-1} g_{i_{l_m}} y, g_{i_{k_n}}^{-1} g_{i_{l_n}} y) < \varepsilon' \quad \text{for every } m, n \geq m_0.$$

By taking $m = n \geq m_0$ it is easy to see that $\{g_{i_{l_m}} y\}$ is a Cauchy net and if we follow the same procedure we can also show that $\{g_{i_{k_m}} y\}$ is a Cauchy net. Since $d(g_{i_{k_m}} y, g_{i_{l_m}} y) \geq \varepsilon$ for every index m the net $\{g_i y\}$ has two limit points in the completion $\widehat{X}$ of X, a contradiction of our hypothesis. □

5. Cauchy-indivisible isometric actions on separable metric spaces

In this section (X, d) *will denote a separable metric space such that the action* $(\mathrm{Iso}(X), X)$ *is Cauchy-indivisible.*

We show the adequacy of sequences in the definition of Cauchy indivisibility:

Proposition 5.1. *In the definition of Cauchy indivisibility for isometric actions nets can be replaced by sequences.*

Proof. Assume that if $\{g_n\}$ is a sequence in $\mathrm{Iso}(X)$ such that $g_n \to \infty$ and $\{g_n x\}$ is a Cauchy sequence in X for some $x \in X$ then $\{g_n x\}$ is a Cauchy sequence for every $x \in X$. Let $\{f_i\}$ be a net in $\mathrm{Iso}(X)$ such that $f_i \to \infty$ and $\{f_i x\}$ is a Cauchy net in X for some $x \in X$. We will show that $\{f_i x\}$ is a Cauchy net in X for every $x \in X$. We argue by contradiction. Suppose that there exists $y \in X$ such that $\{f_i y\}$ is not a Cauchy net. Hence, there is an $\varepsilon > 0$ and subnets $\{f_{i_k}\}$ and $\{f_{i_l}\}$ such that $d(f_{i_k} y, f_{i_l} y) \geq \varepsilon$ for every k, l. Since $\{f_i x\}$ is a Cauchy net in X there is a point $z \in \widehat{X}$ such that $\hat{f}_i x \to z$. Hence, the subnets $\{\hat{f}_{i_k} x\}$ and $\{\hat{f}_{i_l} x\}$ also converge to z. So we may find sequences $\{\hat{f}_{i_{k_n}} x\}$ and $\{\hat{f}_{i_{l_n}} x\}$ such that $\hat{f}_{i_{k_n}} x \to z$ and $\hat{f}_{i_{l_n}} x \to z$. Therefore, $\{f_{i_{k_n}} x\}$ and $\{f_{i_{l_n}} x\}$ are Cauchy sequences in X and $d(f_{i_{k_n}} y, f_{i_{l_n}} y) \geq \varepsilon$ for every $n \in \mathbb{N}$. Let $\{h_n\} \subset \mathrm{Iso}(X)$ with

$$h_{4n-3} = f_{i_{k_{2n-1}}}, \quad h_{4n-2} = f_{i_{l_{2n-1}}}, \quad h_{4n-1} = f_{i_{l_{2n}}}, \quad \text{and} \quad h_{4n} = f_{i_{k_{2n}}},$$

$n = 1, 2, \ldots$. It is easy to see that $\hat{h}_n x \to z$, hence $\{h_n x\}$ is a Cauchy sequence in X. Moreover, $\{h_n y\}$ is not a Cauchy sequence in X since $d(f_{i_{k_n}} y, f_{i_{l_n}} y) \geq \varepsilon$ for every $n \in \mathbb{N}$ and for the same reason $h_n \to \infty$ in $\mathrm{Iso}(X)$, which is a contradiction of our hypothesis. □

Definition 5.2. Fix a dense sequence $D = \{x_i\} \subset X$ in $\widehat{X}$. Since the metric $\hat{d}/(1+\hat{d})$ is an equivalent metric to $\hat{d}$ on X (and also gives the same groups of isometries on X and $\widehat{X}$ and the same Cauchy sequences) we may assume that $\hat{d}$ is bounded by 1.

We define $\delta : \mathrm{Iso}(\widehat{X}) \times \mathrm{Iso}(\widehat{X}) \to \mathbb{R}^+$ by

$$\delta(f, g) = \sum_{i=1}^{\infty} \frac{1}{2^i} \hat{d}(f x_i, g x_i)$$

for every $f, g \in \mathrm{Iso}(\widehat{X})$. It is easy to see that δ is a left-invariant metric on $\mathrm{Iso}(\widehat{X})$.

Proposition 5.3. *The uniformity of pointwise convergence, the left uniformity, and the uniformity induced by δ on $\mathrm{Iso}(\widehat{X})$ and $\mathrm{Iso}(X)$ coincide, independently of Cauchy indivisibility.*

Proof. The proof is similar to the proof of [Hjorth 2008, Lemma 2.11]. □

Proposition 5.4. *The pointwise closures of $\mathrm{Iso}(X)$ in $C(X, X)$ and of $\mathrm{Iso}(\widehat{X})$ in $C(\widehat{X}, \widehat{X})$ endowed with the metric δ are separable metric spaces.*

Proof. It follows easily using the same arguments as in the proof of [Hjorth 2008, Lemma 2.11] and [Bourbaki 1966b, Chapter X, §3, Exercise 6(b), p. 327]. □

The following lemma will be used often in the sequel.

Lemma 5.5. *Let $\{g_n\}$ be a sequence in $\mathrm{Iso}(X)$ such that $\{g_n x\}$ is a Cauchy sequence in X for some $x \in X$ and $g_n \to \infty$. Then*

(i) *$\{g_n x_n\}$ is a Cauchy sequence for every Cauchy sequence $\{x_n\}$ in X and*

(ii) *if $\{x_k\}$ is Cauchy sequence in X then $\hat{g}_n[x_k] \to [g_k x_k]$ in $\widehat{X}$.*

Proof. (i) The proof follows by the triangle inequality and the fact that $\{g_n x_{n_0}\}$ is a Cauchy sequence, for suitable $n_0 \in \mathbb{N}$.

(ii) By (i), $\{g_k x_k\}$ is a Cauchy sequence in X, hence $[g_k x_k] \in \widehat{X}$. The rest of the proof is similar to that of (i). □

Corollary 5.6. *If $\{g_n\}$ is a sequence in $\mathrm{Iso}(X)$ such that $g_n \to \infty$ andi $\{g_n x\}$ is a Cauchy sequence in X for some $x \in X$, then $\{\hat{g}_n\}$ converges pointwise on $\widehat{X}$ to some $h \in C(\widehat{X}, \widehat{X})$ which preserves the metric $\hat{d}$. In addition, if $\{g_n^{-1} y\}$ is a Cauchy sequence for some $y \in X$, then $\{\hat{g}_n\}$ converges pointwise on $\widehat{X}$ to some $h \in \mathrm{Iso}(\widehat{X})$.*

Proof. The proof is an immediate consequence of Lemma 5.5(ii) if we set $h : \widehat{X} \to \widehat{X}$ with $h[x_k] := [g_k x_k]$ for every $[x_k] \in \widehat{X}$. □

Corollary 5.6 enables the following equivalent expressions of the corresponding sets defined in the introduction:

Notation 5.7. We have

$$H = \big\{ h \in C(\widehat{X}, \widehat{X}) \,\big|\, \text{there exists a sequence } \{g_n\} \subset \mathrm{Iso}(X) \text{ with } g_n \to \infty \text{ in } \mathrm{Iso}(X),\ \{g_n x\} \text{ is a Cauchy sequence for some } x \in X \text{ and } \hat{g}_n \to h \text{ in } C(\widehat{X}, \widehat{X}) \big\}.$$

X_l denotes the set of the limit points of the action $(\mathrm{Iso}(X), X)$ in $\widehat{X}$; specifically,

$$X_l = \big\{ y \in \widehat{X} \;\big|\; \text{there exists a sequence } \{g_n\} \subset \mathrm{Iso}(X) \text{ with } g_n \to \infty \text{ in } \mathrm{Iso}(X), \\ \text{such that } \{g_n x\} \text{ is a Cauchy sequence for some } x \in X \text{ and } y = [g_k x] \big\}.$$

X_p denotes the set of the special limit points of $(\mathrm{Iso}(X), X)$ in $\widehat{X}$; specifically,

$$X_p = \big\{ y \in \widehat{X} \;\big|\; \text{there exists a sequence } \{g_n\} \subset \mathrm{Iso}(X) \text{ with } g_n \to \infty \text{ in } \mathrm{Iso}(X), \\ \text{such that } \{g_n x\} \text{ and } \{g_n^{-1} x\} \text{ are Cauchy sequences for some } x \in X \text{ and } y = [g_k x] \big\}.$$

Proposition 5.8. *If $\{g_n\}$ is a sequence in $\mathrm{Iso}(X)$ such that $g_n \to f$ on X for some f in $C(X, X)$, then $\hat{g}_n \to \hat{f}$ on $\widehat{X}$ and $\hat{f} \in \mathrm{Iso}(\widehat{X})$.*

Proof. If $\{g_n\}$ has a convergent subsequence $\{g_{n_k}\}$ to some point $g \in \mathrm{Iso}(X)$ then $f = g$ on X and it is easily seen that $\hat{g}_n \to \hat{g}$ pointwise on $\widehat{X}$. □

With the notation established in the introduction, we have

Proposition 5.9. *The set E is*

(i) *the union $\widehat{\mathrm{Iso}(X)} \cup H$,*

(ii) *complete with respect to the uniformity of pointwise convergence on $\widehat{X}$, and*

(iii) *a semigroup, the Ellis semigroup of $\widehat{\mathrm{Iso}(X)}$ in $C(\widehat{X}, \widehat{X})$, that is, the pointwise closure of $\widehat{\mathrm{Iso}(X)}$ in $C(\widehat{X}, \widehat{X})$.*

Proof. For part (i), take a sequence $\{\hat{g}_n\}$ in $\widehat{\mathrm{Iso}(X)}$ such that $\hat{g}_n \to h$ for some $h \in C(\widehat{X}, \widehat{X})$. If $\{g_n\}$ has a convergent subsequence to some $g \in \mathrm{Iso}(X)$ then, by Proposition 5.8, $h = \hat{g} \in \widehat{\mathrm{Iso}(X)}$. Let $g_n \to \infty$ in $\mathrm{Iso}(X)$ and take some $x \in X$. Since $\hat{g}_n x \to hx$, then $\{g_n x\}$ is a Cauchy sequence in X and therefore $h \in H$. Parts (ii) and (iii) follow from [Hjorth 2008, Lemmata 2.10 and 2.11] by noticing that a sequence $\{g_n\}$ in $\mathrm{Iso}(X)$ is Cauchy with respect to the left uniformity of $\mathrm{Iso}(X)$ if and only if $\{g_n x\}$ is Cauchy in X for every $x \in X$. □

Remark 5.10. As the example in Section 6 shows, the Ellis semigroup E is not, in general, a group. However:

Proposition 5.11. *The Ellis semigroup E is a group if and only if $X_l = X_p$.*

Proof. Assume that E is a group and let $y \in X_l$. Hence, there is a sequence $\{g_n\} \subset \mathrm{Iso}(X)$ with $g_n \to \infty$ in $\mathrm{Iso}(X)$ and a map $h \in C(\widehat{X}, \widehat{X})$ such that $\hat{g}_n \to h$ pointwise on $\widehat{X}$ and $y = hx$ for some $x \in X$. Since E is a group then h has an inverse h^{-1}. Thus $\hat{g}_n^{-1} \to h^{-1}$. The last implies that $\{g_n^{-1} x\}$ is a Cauchy sequence in X, therefore $y \in X_p$.

To show the converse implication, assume that $X_l = X_p$ and take some $h \in E$. By Proposition 5.9(i),

$$h \in \widehat{\mathrm{Iso}(X)} \cup H.$$

So, if $h \in \widehat{\mathrm{Iso}(X)}$ obviously it has an inverse. Assume that $h \in H$. Hence, there is a sequence $\{g_n\} \subset \mathrm{Iso}(X)$ with $g_n \to \infty$ in $\mathrm{Iso}(X)$ such that $\hat{g}_n \to h$ pointwise on $\widehat{X}$. So $[g_n x] \in X_l$ for every $x \in X$. But $X_l = X_p$, hence $\{g_n^{-1} x\}$ is a Cauchy sequence. Applying Corollary 5.6, $h \in \mathrm{Iso}(\widehat{X})$ so it has an inverse in E. □

Lemma 5.12. *The set $X \cup X_l$ is E-invariant.*

Proof. It is easy to verify that X and X_l are $\widehat{\mathrm{Iso}(X)}$-invariant. We will show that they are also H-invariant. Let $h \in H$ and $x \in X$. By the definition of H there is some sequence $\{g_n\}$ in $\mathrm{Iso}(X)$ such that $g_n \to \infty$ in $\mathrm{Iso}(X)$ and $\hat{g}_n \to h$ pointwise on $\widehat{X}$. If $[f_n x] \in X_l$, for some sequence $\{f_n\} \subset \mathrm{Iso}(X)$ and $x \in X$ then, by Corollary 5.6, $h[f_n x] = [g_n f_n x]$. If the sequence $\{g_n f_n\}$ has a convergent subsequence in $\mathrm{Iso}(X)$ then the Cauchy sequence $\{g_n f_n x\}$ has a convergent subsequence in X, so it converges in X. So $h[f_n x] = [g_n f_n x] \in X$. Otherwise $g_n f_n \to \infty$ and $h[f_n x] = [g_n f_n x] \in X_l$. □

Theorem 5.13. *The set $X \cup X_p$ is the maximal subset of $X \cup X_l$ that contains X such that the map*

$$\omega : E \times (X \cup X_p) \to (X \cup X_p) \times \widehat{X},$$

with $\omega(f, y) = (y, fy)$, $f \in E$, and $y \in X \cup X_p$, is proper.

Proof. We first show that the map $\omega : E \times (X \cup X_p) \to (X \cup X_p) \times \widehat{X}$ is proper. Since the evaluation map $E \times (X \cup X_p) \to \widehat{X}$ is isometric and action-like, according to Section 2, it suffices to show that the limit sets $L(x)$ are empty for every $x \in X \cup X_p$. Let $\{f_n\}$ be a sequence in E such that $f_n y \to z$ for some $y \in X \cup X_p$ and $z := [z_k] \in \widehat{X}$.

Case I. Assume that $y \in X$. If $\{f_n\}$ has a subsequence $\{f_{n_k}\}$ in $\widehat{\mathrm{Iso}(X)}$ then either the restriction of $\{f_{n_k}\}$ on X has a convergent subsequence in $\mathrm{Iso}(X)$ hence, by Proposition 5.8, the sequence $\{f_{n_k}\}$ converges pointwise to some point of $\widehat{\mathrm{Iso}(X)} \subset E$, or $f_n \to \infty$ in $\mathrm{Iso}(X)$. In this case, since $\{f_n y\}$ is a Cauchy sequence in X, the sequence $\{f_n\}$ converges pointwise to some point of $H \subset E$ by Corollary 5.6.

Assume, now, that $\{f_n\}$ is in H and consider the dense sequence $D = \{x_i\}$ in X which we used to define the metric δ; see Definition 5.2. So, there is a sequence $\{x_{i_n}\}$ in D such that $x_{i_n} \to y$. By the definition of H and Proposition 5.3, there is a sequence $\{g_n\}$ in $\mathrm{Iso}(X)$ such that

$$\delta(\hat{g}_n, f_n) < \frac{1}{i_n 2^{i_n}}. \tag{5-1}$$

Hence, using the form of the metric δ, we conclude that

$$\hat{d}(\hat{g}_n x_{i_n}, f_n x_{i_n}) < \frac{1}{i_n}.$$

Moreover,

$$\begin{aligned}\hat{d}(\hat{g}_n y, z) &\leq \hat{d}(\hat{g}_n x_{i_n}, f_n x_{i_n}) + \hat{d}(f_n x_{i_n}, f_n y) + \hat{d}(f_n y, z)\\ &= \hat{d}(\hat{g}_n x_{i_n}, f_n x_{i_n}) + \hat{d}(x_{i_n}, y) + \hat{d}(f_n y, z).\end{aligned}$$

Therefore, $g_n y \to z$. Arguing as in the beginning of the proof, $\{g_n\}$ has a convergent subsequence to a point of E, hence by (5-1), the same holds for the sequence $\{f_n\}$.

Case II. Assume that $y \in X_p$. Hence, there exists a sequence $\{p_k\} \subset \mathrm{Iso}(X)$ with $p_k \to \infty$ in $\mathrm{Iso}(X)$, an isometry $h_1 \in \mathrm{Iso}(\widehat{X})$ such that $\hat{p}_k \to h_1$ pointwise on $\widehat{X}$, and $h_1 x := [p_k x] = y$ for some $x \in X$. If $\{f_n\}$ has a subsequence $\{f_{n_k}\}$ in $\widehat{\mathrm{Iso}(X)}$ then either the restriction of $\{f_{n_k}\}$ on X has a convergent subsequence in $\mathrm{Iso}(X)$ hence, by Proposition 5.8, the sequence $\{f_{n_k}\}$ converges pointwise to some point of $\widehat{\mathrm{Iso}(X)} \subset E$, or $f_n \to \infty$ in $\mathrm{Iso}(X)$. If the latter holds, then we will show that there is a Cauchy sequence of the form $\{f_{n_i} p_{k_i} x\}$ in X for some subsequences $\{f_{n_i}\}$ and $\{p_{k_i}\}$ of $\{f_n\}$ and $\{p_k\}$, respectively (the problem is that we do not know if $\{f_n x\}$ or $\{f_n p_n x\}$ is a Cauchy sequence in X for some $x \in X$).

Let i be a positive integer. Since $f_n[p_k x] \to z$ and $z := [z_k] \in \widehat{X}$, there is a positive integer n_0 that depends only on i such that

$$\hat{d}(f_n[p_k x], [z_k]) < \frac{1}{i}$$

for every $n \geq n_0(i)$. Therefore

$$\lim_k d(f_n p_k x, z_k) := \hat{d}(f_n[p_k x], [z_k]) < \frac{1}{i}$$

for every $n \geq n_0(i)$. Hence, using induction, we may find strictly increasing sequences of positive integers $\{n_i\}$ and $\{k_i\}$ such that

$$d(f_{n_i} p_{k_i} x, z_{k_i}) < \frac{1}{i} \tag{5-2}$$

for every positive integer i. Since $\{z_{k_i}\}$ is a Cauchy sequence then by (5-2), $\{f_{n_i} p_{k_i} x\}$ is a Cauchy sequence in X.

Now, either $\{f_{n_i} p_{k_i}\}$ has a convergent subsequence in $\mathrm{Iso}(X)$ (without loss of generality and for the economy of the proof we may assume that $\{f_{n_i} p_{k_i}\}$ converges in $\mathrm{Iso}(X)$) or $f_{n_i} p_{k_i} \to \infty$ in $\mathrm{Iso}(X)$. In both cases, by Corollary 5.6 and Proposition 5.8, there is $h_2 \in C(\widehat{X}, \widehat{X})$ such that $\hat{f}_{n_i} \hat{p}_{k_i} = \widehat{f_{n_i} p_{k_i}} \to h_2$ pointwise on $\widehat{X}$. We will show that $\hat{f}_{n_i} \to h_2 h_1^{-1}$ pointwise on $\widehat{X}$. Take $w \in \widehat{X}$. Since $h_1 \in \mathrm{Iso}(\widehat{X})$, there is some $u \in \widehat{X}$ such that $h_1(u) = w$. Hence

$$\begin{aligned}\hat{d}(f_{n_i} w, h_2 h_1^{-1} w) = \hat{d}(f_{n_i} h_1 u, h_2 u) &\leq \hat{d}(f_{n_i} h_1 u, \widehat{f_{n_i} p_{k_i}} u) + \hat{d}(\widehat{f_{n_i} p_{k_i}} u, h_2 u)\\ &= \hat{d}(h_1 u, \hat{p}_{k_i} u) + \hat{d}(\widehat{f_{n_i} p_{k_i}} u, h_2 u),\end{aligned}$$

which converges to 0, since $\hat{p}_{k_i} \to h_1$ and $\widehat{f_{n_i} p_{k_i}} \to h_2$ pointwise on $\widehat{X}$. Hence $\{f_n x\}$ is a Cauchy sequence for every $x \in X$. Since we assumed that $f_n \to \infty$ in $\mathrm{Iso}(X)$ then, by Corollary 5.6, $\{f_n\}$ converges pointwise on $\widehat{X}$ to $h_2 h_1^{-1} \in E$.

To finish the proof of the second case assume that $\{f_n\}$ is in H. Then arguing as in the first case we can show that $\{f_n\}$ has a convergent subsequence to a point of E.

Next, we show that if Y is a subset of $X \cup X_l$ that contains X such that the map

$$\omega : E \times Y \to Y \times \widehat{X}$$

is proper then $Y \subset X \cup X_p$. To see that take a point $[g_k x] \in Y \setminus X$. This means that $\{g_k\}$ is a sequence in $\mathrm{Iso}(X)$ such that $g_k \to \infty$ in $\mathrm{Iso}(X)$ and $\{g_k x\}$ is a Cauchy sequence in X. By Lemma 5.5(ii), $\hat{g}_n x \to [g_k x]$ and, by Corollary 5.6, $\{\hat{g}_n\}$ converges pointwise on $\widehat{X}$ to some $h \in C(\widehat{X}, \widehat{X})$. Note that $x \in X \subset Y$. Since $\hat{g}_n x \to [g_k x]$ and $\hat{g}_n$ preserves the metric $\hat{d}$ then $\hat{g}_n^{-1}[g_k x] \to x$. Hence, by the properness of ω, we may assume that $\{\hat{g}_n^{-1}\}$ has a subsequence $\{\hat{g}_{n_k}^{-1}\}$ that converges pointwise to some $f \in E$. This makes h a surjection, hence $h \in \mathrm{Iso}(\widehat{X})$. Therefore, $[g_k x] \in X_p$, so $Y \setminus X \subset X_p$. □

Note that, as the following example shows, it may happen that $X_p = X_l \neq \varnothing$, $X \cup X_p \neq \widehat{X}$, and *the set* $X \cup X_p$ *is not the maximal subset of* $\widehat{X}$ *such that the action* $(E, X \cup X_p)$ *is proper*.

Example 5.14. Take

$$X := \{(x, y) \in \mathbb{R} \mid x \in \mathbb{Q} + \sqrt{2}\mathbb{N},\ y > 0\},$$

endowed with the Euclidean metric. Its group of isometries is the additive group of the rational numbers acting by horizontal translations. Therefore, $(\mathrm{Iso}(X), X)$ is Cauchy-indivisible. Obviously $\widehat{X}$ is the closed upper half-plane, $X_p = X_l \neq \varnothing$, $X \cup X_p$ is the open upper half-plane, and E is the additive group of the real numbers acting by horizontal translations on $\widehat{X}$. Hence E acts properly on $\widehat{X}$.

Remark 5.15. The sets X_p and X_l constructed in Theorem 5.13 are *optimal* in the sense that if one may think to replace the sets X_p and X_l with the following more general sets

$$X_l^* = \Big\{ y \in \widehat{X} \;\Big|\; \text{there exists a sequence } \{g_n\} \subset \mathrm{Iso}(X) \text{ and some } x \in X \text{ such that } g_n \to \infty \text{ in } \mathrm{Iso}(X), \{g_n x\} \text{ is a Cauchy sequence, and } y = [g_k x_k], \text{ for some } [x_k] \in \widehat{X} \Big\}$$

and

$$X_p^* = \Big\{ y \in \widehat{X} \;\Big|\; \text{there exists a sequence } \{g_n\} \subset \mathrm{Iso}(X) \text{ and some } x \in X \text{ such that } g_n \to \infty \text{ in } \mathrm{Iso}(X), \{g_n x\} \text{ and } \{g_n^{-1} x\} \text{ are Cauchy sequences, and } y = [g_k x_k], \text{ for some } [x_k] \in \widehat{X} \Big\},$$

and ask if the set $X \cup X_p^*$ is the maximal subset of the completion $\widehat{X}$ such that the map $\omega^* : E \times (X \cup X_p^*) \to (X \cup X_p^*) \times \widehat{X}$ with $\omega^*(f, y) = (y, fy)$, $f \in E$, and $y \in X \cup X_p^*$ is proper *this is not true in general.* This follows from the following assertion and Example 5.17, which shows that there is a metric space X such that $(\mathrm{Iso}(X), X)$ is Cauchy-indivisible, $X_p \neq \varnothing$, and the map ω^* as above is not proper.

Assertion 5.16. If $X_p^* \neq \varnothing$ (equivalently $X_p \neq \varnothing$) then $X_p^* = \widehat{X}$.

Proof. Let $y = [x_k]$ in $\widehat{X}$. By assumption, there exists a sequence $\{g_n\} \subset \mathrm{Iso}(X)$ and a point $x \in X$ such that $g_n \to \infty$ in $\mathrm{Iso}(X)$ and the sequences $\{g_n x\}$ and $\{g_n^{-1} x\}$ are Cauchy. By Lemma 5.5, $\{g_n x_n\}$ is a Cauchy sequence in X and $g_k^{-1}[g_k x_k] \to [g_k^{-1} g_k x_k] = [x_k] = y$. Hence, $y \in X_p^*$. □

Example 5.17. This is a combination of Example 4.3 and of a 3-dimensional variation of the “river metric” [Engelking 1989, Example 4.1.6]. Let

$$X = \{(x, y, z) \mid x \in \mathbb{Q} + \sqrt{2}\mathbb{N},\ y \in \mathbb{Q} + \sqrt{2}\mathbb{N},\ z > 0\}.$$

For every pair of points $w_1 = (x_1, y_1, z_1)$, $w_2 = (x_2, y_2, z_2) \in X$ define

$$d(w_1, w_2) := \begin{cases} |y_1 - y_2| + |z_1 - z_2|, & \text{if } x_1 = x_2, \\ |y_1| + |y_2| + |x_1 - x_2| + |z_1 - z_2|, & \text{if } x_1 \neq x_2. \end{cases}$$

We can easily verify that d is a metric on X. The group of isometries $\mathrm{Iso}(X, d)$ consists of all the maps $g : X \to X$ of the form

$$g(x, y, z) = (x + p, y + q, z), \quad p, q \in \mathbb{Q}.$$

The action $(\mathrm{Iso}(X), X)$ is Cauchy-indivisible since X does not contain the xy-plane (the last coordinate of the points of X is positive). Then

$$X_p = \{(x, y, z) \mid x \in \mathbb{Q} + \sqrt{2}\mathbb{N},\ y \in \mathbb{R},\ z > 0\}.$$

To see that take $x \in \mathbb{Q} + \sqrt{2}\mathbb{N}$, $y \in \mathbb{R}$, and $z > 0$ and choose $k \in \mathbb{N}$ such that $y - \sqrt{2}k \notin \mathbb{Q}$. Let $\{q_n\}$ be a sequence of rational numbers such that $q_n \to y - \sqrt{2}k$. Hence, if we let $\{g_n\} \subset \mathrm{Iso}(X)$ with $g_n(x, y, z) := (x, y + q_n, z)$ then $g_n(x, \sqrt{2}k, z) = (x, q_n + \sqrt{2}k, z) \to (x, y, z)$. Hence $(x, y, z) \in X_p$. Observe that

$$\widehat{X} = \{(x, y, z) \mid x \in \mathbb{Q} + \sqrt{2}\mathbb{N},\ y \in \mathbb{R},\ z \geq 0\},$$

and E consists of all the maps $g : \widehat{X} \to \widehat{X}$ with

$$g(x, y, z) = (x + p, y + r, z), \quad p \in \mathbb{Q}, \quad r \in \mathbb{R}.$$

However, the map $\widehat{\omega} : E \times \widehat{X} \to \widehat{X} \times \widehat{X}$ with $\widehat{\omega}(f, w) = (w, f w)$, $f \in E$, and $w \in \widehat{X}$ is not proper since if we take a sequence of rational numbers $\{p_n\}$ such that $p_n \to \sqrt{2}$ and let $\{g_n\} \subset E$ with $g_n(x, y, z) = (x + p_n, y, z)$ then $g_n(x, 0, 0) \to (x + \sqrt{2}, 0, 0)$ for each $x \in \mathbb{Q} + \sqrt{2}\mathbb{N}$. The sequence $\{g_n\}$ diverges in E since, for instance, the distance of the points $g_n(\sqrt{2}, \sqrt{2}, 1) = (q_n + \sqrt{2}, \sqrt{2}, 1)$ from any point of X is eventually at least $\sqrt{2}$. Hence the limit set $L((x, 0, 0))$ is not empty.

A question that arises naturally from Theorem 5.13 is if the action of the Ellis semigroup E on $X \cup X_l$ is proper. Surprisingly, as the following proposition shows, this is equivalent to the existence of a Weil completion (with respect to the uniformity of pointwise convergence) for the group $\mathrm{Iso}(X)$. But first let us recall a few things about the Weil completion of $\mathrm{Iso}(X)$, defined in the introduction. The uniformity of pointwise convergence on X coincides with the left uniformity of $\mathrm{Iso}(X)$ (see [Bourbaki 1966a, Chapter III, §3.1 and Chapter X, §3, Exercise 19(a), p. 332]) and $\mathrm{Iso}(X)$ has Weil completion with respect to this uniformity if the left and the right uniformities coincide; see [Bourbaki 1966a, Chapter III, §3.4 and §3, Exercise 3, p. 306]. Note that the left completion of $\mathrm{Iso}(X)$ does not depend on the choice of a left-invariant metric on $\mathrm{Iso}(X)$; see [Hjorth 2008, Lemma 2.9].

Proposition 5.18. *The following are equivalent*:

(i) *The map* $\omega : E \times (X \cup X_l) \to (X \cup X_l) \times \widehat{X}$ *is proper.*

(ii) E *is a group* (*precisely a closed subgroup of* $\mathrm{Iso}(\widehat{X})$).

(iii) $\mathrm{Iso}(X)$ *has a Weil completion with respect to the uniformity of pointwise convergence* (*in this case* E *is the Weil completion of* $\mathrm{Iso}(X)$).

Proof. We show that (i) implies (ii) and vice versa. Suppose that ω is proper. Take some $h \in E$. Since E is a semigroup, see Proposition 5.9(iii), we have only to show that h has an inverse in E. If $h = \hat{g} \in \widehat{\mathrm{Iso}(X)}$ for some $g \in \mathrm{Iso}(X)$, then $\hat{g}^{-1}$ is the inverse of h in $\widehat{\mathrm{Iso}(X)} \subset E$. If $h \in H$ there is a sequence $\{g_n\}$ in $\mathrm{Iso}(X)$ such that $g_n \to \infty$ in $\mathrm{Iso}(X)$ and $\hat{g}_n \to h$ pointwise on $\widehat{X}$. Hence, if $x \in X$ then $\hat{g}_n x \to hx$. Since $\hat{g}_n$ preserves the metric $\hat{d}$ then $\hat{g}_n^{-1} hx = \hat{g}_n^{-1} hx \to x$. By Lemma 5.12, $hx \in X \cup X_l$, hence, by the properness of ω, $\{\hat{g}_n^{-1}\}$ has a convergent subsequence $\{\hat{g}_{n_k}^{-1}\}$ to some $f \in E$. This makes h a surjection, hence $h \in \mathrm{Iso}(\widehat{X})$ and h has an inverse in E. To show the converse implication note that if E is a group then $X_l = X_p$; see Proposition 5.11. Hence, by Theorem 5.13, the map ω is proper.

To finish the proof of the proposition let us show that (iii) implies (ii) and vice versa. Note that $\mathrm{Iso}(X)$ has a Weil completion if and only if the map with $g \mapsto g^{-1}$ for every $g \in \mathrm{Iso}(X)$ maps Cauchy sequences of $\mathrm{Iso}(X)$ to Cauchy sequences; see

[Bourbaki 1966a, Chapter III, §3.4, Theorem 1]. It is easy to check that in the case when $\mathrm{Iso}(X)$ is Cauchy-indivisible this is equivalent to $X_l = X_p$. Equivalently, by Proposition 5.11, E is a group. □

Remark 5.19. In the case when $\mathrm{Iso}(X)$ is a locally compact group, for example, if X is a locally compact space and $\mathrm{Iso}(X)$ acts properly on it (as it is known in the case X is connected), then by [Bourbaki 1966a, Chapter III, §3, Exercise 8, p. 307], $\mathrm{Iso}(X)$ has a locally compact completion hence E is a locally compact group.

We summarize with the following.

Corollary 5.20. *If E is a group the action* $(\mathrm{Iso}(X), X)$ *is embedded densely in the proper action* $(E, X \cup X_l)$ *such that the following equivariant diagram commutes*:

$$\begin{array}{ccc} (\mathrm{Iso}(X), X) & \longrightarrow & X \\ \downarrow & & \downarrow \\ (E, X \cup X_l) & \longrightarrow & \widehat{X}, \end{array}$$

where $X \to X \cup X_l$ is the inclusion map and the map $\mathrm{Iso}(X) \to E$ *is defined by $g \mapsto \hat{g}$ for every $g \in \mathrm{Iso}(X)$. By "densely" we mean that X is dense in $X \cup X_l$ and* $\widehat{\mathrm{Iso}(X)}$ *is dense in E.*

Question 5.21. The above embedding of a Cauchy-indivisible action as a dense subaction of a proper one establishes a remarkable connection between Cauchy-indivisible and proper actions, and at the same time proposes an interesting question: Is there any analogy with the situation of embedding of a proper action (on a locally compact and connected space) in an appropriate zero-dimensional compactification, like in [Abels 1972; Manoussos and Strantzalos 2007]? Namely, can we obtain any structurally informative correspondence between divergent nets in $\mathrm{Iso}(X)$ and suitable subsets of X_l?

Remark 5.22. As we will see in the example described in Section 6 it may happen that $X_p \neq X_l$ and $X \cup X_l = \widehat{X}$.

In view of possible questions for refinements of Corollary 5.20 we note that it may happen that $X \cup X_p = \widehat{X}$ and E is not dense in $\mathrm{Iso}(\widehat{X})$, as the following example shows:

Example 5.23. There is a separable metric space (X, d) such that $(\mathrm{Iso}(X), X)$ is Cauchy-indivisible, proper, $X \cup X_p = \widehat{X}$, and $\mathrm{Iso}(X)$ has a Weil completion which does not coincide with the group $\mathrm{Iso}(\widehat{X})$.

Proof. We let X be the set $\mathbb{Q} + \sqrt{2}\mathbb{N}$ endowed with the Euclidean metric; see Example 4.3. It is easy to check that $X \cup X_p = X \cup X_l = \mathbb{R}$, see also Example 5.17, hence by Propositions 5.11 and 5.18, $\mathrm{Iso}(X)$ has a Weil completion (or just observe that $\mathrm{Iso}(X)$ is an abelian group and use [Bourbaki 1966a, Chapter III, §3.5, Theorem 2]). But all the reflections of the space are excluded, hence the pointwise closure E of $\widehat{\mathrm{Iso}(X)}$ does not coincide with $\mathrm{Iso}(\mathbb{R})$. □

6. An example of a proper Cauchy-indivisible action of a group which has no Weil completion

In this section we show that there is a separable metric space X such that the action $(\mathrm{Iso}(X), X)$ is proper and Cauchy-indivisible, and the Ellis semigroup E is not a group. Equivalently, in view of Proposition 5.18, $\mathrm{Iso}(X)$ has no Weil completion. Consider the space of the integers $\mathbb{Z}$ with the discrete metric d, that is, if $m, n \in \mathbb{Z}$ then $d(m, n) = 0$ if $m = n$ and $d(m, n) = 1$ otherwise. The group of isometries $\mathrm{Iso}(\mathbb{Z})$ consists of all the self bijections of $\mathbb{Z}$ and is an example of a topological group that has no Weil completion. To see that take $f_n : \mathbb{Z} \to \mathbb{Z}$ with $f_n z = z$ for $-n < z < 0$, $f_n(-n) = 0$, and $f_n z = z + 1$ otherwise. Then it is easy to verify that $f_n \to f$, where $f z = z$ for $z < 0$, and $f z = z+1$ for $z \geq 0$. Hence $\{f_n z\}$ is a Cauchy sequence in $\mathbb{Z}$ for every $z \in \mathbb{Z}$, therefore $\{f_n\}$ is a Cauchy sequence in $\mathrm{Iso}(\mathbb{Z})$ with respect to the uniformity of pointwise convergence on $\mathbb{Z}$. But $\{f_n^{-1} 0\} = \{-n\}$ is not a Cauchy sequence, so neither is $\{f_n^{-1}\}$. Thus, by [Bourbaki 1966a, Chapter III, §3.4, Theorem 1], $\mathrm{Iso}(\mathbb{Z})$ has no Weil completion. The problem is that the action $(\mathrm{Iso}(\mathbb{Z}), \mathbb{Z})$ is not Cauchy-indivisible. To see that notice that $\{f_n^{-1} 1\} = \{0\}$ but $\{f_n^{-1} 0\} = \{-n\}$ is not a Cauchy sequence. Nevertheless, the group $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ is Cauchy-indivisible and has no Weil completion as we show in the following.

Take an enumeration $A = \{z_i\}$ of $\mathbb{Z}$ and equip $\mathrm{Iso}(\mathbb{Z})$ with the metric

$$\varrho(f, g) = \sum_{i=1}^{\infty} \frac{1}{3^i} d(f z_i, g z_i)$$

for $f, g \in \mathrm{Iso}(\mathbb{Z})$. In view of Proposition 5.3 the uniformity of pointwise convergence, the left uniformity, and the uniformity induced by ϱ on $\mathrm{Iso}(\mathbb{Z})$ coincide (the choice of $\frac{1}{3}$ instead of $\frac{1}{2}$ in Definition 5.2 will be clarified in the proof of Lemma 6.1). Note that $(\mathrm{Iso}(\mathbb{Z}), \varrho)$ is a separable metric space. We will show that $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ is Cauchy-indivisible but has no Weil completion.

Lemma 6.1. *If* $T \in \mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ *and* $f, g \in \mathrm{Iso}(\mathbb{Z})$ *then*

$$d(T(f) z, T(g) z) = d(f z, g z)$$

for every $z \in \mathbb{Z}$.

Proof. Since $\varrho(T(f), T(g)) = \varrho(f, g)$ then

$$\sum_{i=1}^{\infty} \frac{1}{3^i} d(T(f) z_i, T(g) z_i) = \sum_{i=1}^{\infty} \frac{1}{3^i} d(f z_i, g z_i).$$

Since the values of d are 0 or 1 then $d(T(f)z_n, T(g)z_n) = d(fz_n, gz_n)$, for every $z_n \in A = \mathbb{Z}$ (here is the role of the choice of $\frac{1}{3}$ instead of $\frac{1}{2}$). ☐

Proposition 6.2. *If $T \in \mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ and $f \in \mathrm{Iso}(\mathbb{Z})$ then $T(f) = T(e) \circ f$, where e is the unit element of $\mathrm{Iso}(\mathbb{Z})$.*

Proof. Let z_k and z_l be two distinct integers and let $g \in \mathrm{Iso}(\mathbb{Z})$ be such that $gz_k = z_l$, $gz_l = z_k$, and $gz = z$ elsewhere. We show that $T(g) = T(e) \circ g$. If $z \neq z_k, z_l$ then, by Lemma 6.1, $d(T(g)z, T(e)z) = d(gz, z) = 0$. Hence $T(g)z = T(e)z = T(e) \circ gz$. Moreover

$$d(T(g)z_k, T(e)z_k) = d(gz_k, z_k) = d(z_l, z_k) = 1$$

and, similarly, $d(T(g)z_l, T(e)z_l) = 1$. Since $T(g)z_k \neq T(g)z = T(e)z$ for $z \neq z_k, z_l$ and $T(e)$ is surjective then $T(g)z_k = T(e)z_l = T(e) \circ gz_k$ and, similarly, $T(g)z_l = T(e) \circ gz_l$. Therefore $T(g) = T(e) \circ g$.

Fix $f \in \mathrm{Iso}(\mathbb{Z})$ and some $z \in \mathbb{Z}$. If $fz = z$ then $T(f)z = T(e)z = T(e) \circ fz$ since $d(T(f)z, T(e)z) = d(fz, z) = 0$. If $fz \neq z$, let $g \in \mathrm{Iso}(\mathbb{Z})$ with $gz = fz$, $gfz = z$, and $gw = w$ elsewhere. Since $d(T(f)z, T(g)z) = d(fz, gz) = 0$ then $T(f)z = T(g)z$. Using the result of the previous paragraph, $T(f)z = T(g)z = T(e) \circ gz = T(e) \circ fz$. Since z was arbitrary then $T(f) = T(e) \circ f$. ☐

Corollary 6.3. *Let $L, T \in \mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$. Then $L \circ T(e) = L(e) \circ T(e)$ and $T^{-1}(e) = (T(e))^{-1}$.*

Proof. Since $T(f) = T(e) \circ f$ for every $T \in \mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ and $f \in \mathrm{Iso}(\mathbb{Z})$, then

$$L \circ T(f) = L(T(f)) = L(e) \circ T(f) = L(e) \circ T(e) \circ f.$$

Hence, $L(e) \circ T(e) = L \circ T(e)$. If I denotes the identity on $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$, then $f = I(f) = I(e) \circ f$. Hence $I(e) = e$ and $T^{-1}(e) = (T(e))^{-1}$. ☐

Proposition 6.4. *The map $\mathcal{B} : \mathrm{Iso}(\mathrm{Iso}(\mathbb{Z})) \to \mathrm{Iso}(\mathbb{Z})$ with $\mathcal{B}(T) = T(e)$ is a uniform group isomorphism with respect to the uniformities of pointwise convergence on the underlying spaces $\mathrm{Iso}(\mathbb{Z})$ and $\mathbb{Z}$, respectively.*

Proof. By Proposition 5.3 we can equip $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ with a left-invariant metric σ such that the uniformity of pointwise convergence, the left uniformity, and the uniformity induced by σ on $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ coincide. Let $L_n, T_n \in \mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ such that $\sigma(L_n, T_n) \to 0$, hence $\sigma(T_n^{-1}L_n, I) \to 0$. Therefore $T_n^{-1}L_n \to I$ pointwise on $\mathrm{Iso}(\mathbb{Z})$ so $T_n^{-1}L_n(e) \to e$, thus $\varrho(L_n(e), T_n(e)) \to 0$. For the converse, note that if $\varrho(T_n^{-1}L_n(e), e) \to 0$ then $\varrho(T_n^{-1}L_n(e) \circ f, f) \to 0$ for every $f \in \mathrm{Iso}(\mathbb{Z})$ since the map $\mathrm{Iso}(\mathbb{Z}) \to \mathrm{Iso}(\mathbb{Z})$ with $g \mapsto gf$ is continuous. Hence $T_n^{-1}L_n \to I$ pointwise on $\mathrm{Iso}(\mathbb{Z})$. Corollary 6.3 implies that $\mathcal{B}$ is also group isomorphism. ☐

Proposition 6.5. *The group $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ is Cauchy-indivisible and has no Weil completion.*

Proof. Let us show firstly that $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ is Cauchy-indivisible. Let $\{T_n\} \subset \mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ and $f \in \mathrm{Iso}(\mathbb{Z})$ such that $\{T_n(f)\}$ is a Cauchy sequence in $\mathrm{Iso}(\mathbb{Z})$. Take some $g \in \mathrm{Iso}(\mathbb{Z})$. Since $\{T_n(f)\}$ is a Cauchy sequence in $\mathrm{Iso}(\mathbb{Z})$ then it is easy to see that $\{T_n(f)z\}$ is a Cauchy sequence for every $z \in \mathbb{Z}$. Equivalently, $\{T_n(f)f^{-1}gz\}$ is a Cauchy sequence for every $z \in \mathbb{Z}$. By Proposition 6.2,

$$T_n(f)f^{-1}gz = T_n(e) \circ ff^{-1}gz = T_n(e) \circ gz = T_n(g)z.$$

Therefore $\{T_n(g)\}$ is a Cauchy sequence in $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ for every $g \in \mathrm{Iso}(\mathbb{Z})$, hence $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ is Cauchy-indivisible.

Since by the previous proposition the groups $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ and $\mathrm{Iso}(\mathbb{Z})$ are uniformly isomorphic and the group $\mathrm{Iso}(\mathbb{Z})$ has no Weil completion then the same also holds for $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$. □

Proposition 6.6. *The action* $(\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z})), \mathrm{Iso}(\mathbb{Z}))$ *is proper.*

Proof. Let $f, g \in \mathrm{Iso}(\mathbb{Z})$ and $\{T_n\} \subset \mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ be a sequence such that $T_n(f) \to g$. Hence, by Proposition 6.2, $T_n(e) \circ f \to g$ thus $T_n(e) \to gf^{-1}$. Therefore $\{T_n(h)\}$ converges for every $h \in \mathrm{Iso}(\mathbb{Z})$. Since $(T_n(e))^{-1} \to fg^{-1}$ it is easy to verify that $\{T_n\}$ converges in $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ hence the action $(\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z})), \mathrm{Iso}(\mathbb{Z}))$ is proper. □

Remark 6.7. Notice that $\mathrm{Iso}(\mathrm{Iso}(\mathbb{Z}))$ is not locally compact since it has no Weil completion ($\mathrm{Iso}(\mathbb{Z})$ is, of course, not locally compact).

7. Borel sections, fundamental sets, and Cauchy indivisibility

As it is indicated in the introduction a section of an action (G, X) is a subset of X which contains only one point from each orbit. If a section is a Borel subset of X it called a Borel section. Concerning the existence of Borel sections, if (Y, d) is a separable metric space and $\mathcal{R}$ is an equivalence relation on Y such that the $\mathcal{R}$-saturation of each open set is Borel, then there is a Borel set S whose intersection with each $\mathcal{R}$-equivalence class which is complete with respect to d is nonempty, and whose intersection with each $\mathcal{R}$-equivalence class is at most one point; see [Kallman and Mauldin 1978, Lemma 2]. The problem of the existence of a Borel section for a continuous Polish action is of remarkable significance because the existence of a Borel section is equivalent to many interesting facts, like that the underlying space has only trivial ergodic measures and that the orbit space has a standard Borel structure and has no nontrivial atoms. Recall that an action (G, X) is called *Polish* if both G and X are Polish spaces, that is, they are separable and metrizable by a complete metric. Keeping the previous in mind we have the following:

Proposition 7.1. *If the Ellis semigroup E is a group then the action $(E, X \cup X_l)$ has a Borel section.*

Proof. Assume that the Ellis semigroup E is a group. Since by Proposition 5.11 we have $X_l = X_p$ and by Proposition 5.18 the map $\omega : E \times (X \cup X_l) \to (X \cup X_l) \times \widetilde{X}$ is proper then each orbit Ex, $x \in X \cup X_l$, is closed in $\widetilde{X}$. Hence, by [Kallman and Mauldin 1978, Lemma 2] there exists a Borel set $S \subset \widetilde{X}$ such that $S \cap (X \cup X_l)$ is a Borel section (with respect to the relative topology of $X \cup X_l$) for the action $(E, X \cup X_l)$. □

A very useful notion in the theory of proper actions on locally compact spaces with paracompact orbit spaces is the notion of a fundamental set.

Let G be a topological group which acts continuously on a topological space X and $A, B \subset X$. Let us call the set $G_{AB} := \{g \in G : gA \cap B \neq \varnothing\}$ the *transporter* from A to B.

Definition 7.2. A subset F of X is called a fundamental set for the action (G, X) if the following holds.

(i) $GF = X$.

(ii) For every $x \in X$ there exists a neighborhood $V \subset X$ of x such that the transporter G_{VF} of V to F has compact closure in G.

For locally compact spaces we can replace condition (ii) with the following equivalent condition.

(iia) The transporter G_{KF} from K to F has compact closure in G for every nonempty compact subset K of X.

Note that the existence of a fundamental set implies that the action group G is locally compact and the action (G, X) is proper.

The notion of a fundamental set is relative to the notion of a section but it is different in general, in the sense that there are cases where a section is a fundamental set, cases where a fundamental set fails to be a section, and cases where a section fails to be a fundamental set. A section may not be Borel or even if it is Borel may not be contained in any fundamental set, as the following example shows.

Example 7.3. The action $(\mathbb{Z}, \mathbb{R})$ with $(z, r) \mapsto r + z$, $z \in \mathbb{Z}$, $r \in \mathbb{R}$, is proper and it has a Borel section which is not contained in any fundamental set. Indeed, it is easy to see that the set

$$S := \Big([0, 1) \setminus \bigcup_{n \in \mathbb{N}} \Big\{\frac{1}{n}\Big\}\Big) \cup \bigcup_{n \in \mathbb{N}} \Big\{n + \frac{1}{n}\Big\}$$

is a section because the interval $[0, 1)$ is a section (and a fundamental set) for the action $(\mathbb{Z}, \mathbb{R})$. Take an open ball B centered at 0 with radius $\varepsilon > 0$. Then there exists $n_0 \in \mathbb{N}$ such that $1/n < \varepsilon$ for every $n \geq n_0$. Let A be a subset of $\mathbb{R}$ that contains S. Hence $\{n \mid n \geq n_0\}$ is a subset of the transporter $\mathbb{Z}_{BS} \subset \mathbb{Z}_{BA}$, so A cannot be a fundamental set.

It is also possible to construct a section which is not Borel. Take a set $D \subset [0, 1)$ which is not a Borel set and consider the set $S_1 := D \cup \{x+2 \mid x \in \mathbb{R} \setminus D\}$. Obviously S_1 is a section which is not a Borel subset of the reals.

Nevertheless sections, Borel sections, and fundamental sets have a very strong connection as the following theorem shows.

Theorem 7.4. *Let G be a group which acts properly on a locally compact space X, and suppose that the orbit space $G\backslash X$ is paracompact. Let S be a section for the action (G, X). Then*

(i) *For every open neighborhood U of S we can construct a closed fundamental set F_c and an open fundamental set F_o such that $F_c \subset F_o \subset U$.*

(ii) *If, in addition, (X, d) is a separable metric space, in which case the action (G, X) is Cauchy-indivisible, by Theorem 3.3, then there exists a Borel section S_B, which is also a fundamental set, such that $S_B \subset F_c \subset F_o \subset U$.*

Proof. (i) Since U is open it is a union of open balls, let us say S_i, $i \in I$. Let $p : X \to G\backslash X$ be the natural map $x \mapsto Gx$. Then $p(S_i)$, $i \in I$, is an open covering of the locally compact and paracompact space $G\backslash X$. Hence, there is a locally finite refinement $\{W_j\}$, $j \in J$, which consists of open subsets of $G\backslash X$ with compact closures such that $W_j \subset p(S_{i_j})$, for some $i_j \in I$. Now we can follow the classical proof for the existence of fundamental sets; see [Koszul 1965, Lemma 2, p. 8]. Let $\{V_j\}$ be an open covering of $G\backslash X$ such that $\overline{V}_j \subset W_j$ for every $j \in J$. Fix an index $j \in J$ and consider the restriction of the natural map $p : X \to G\backslash X$ on the open ball S_{i_j}. Since S_{i_j} is locally compact then there exists an open set $U_{i_j} \subset S_{i_j}$ with compact closure and a compact set $K_{i_j} \subset U_{i_j} \subset S_{i_j}$ such that $p(U_{i_j}) = W_j$ and $p(K_{i_j}) = \overline{V}_j$. Let $F_c := \bigcup_j K_{i_j}$ and $F_o := \bigcup_j U_{i_j}$. The family $\{U_{i_j}\}_{j \in J}$ is locally finite in X hence the set F_c is closed; see [Bourbaki 1966a, Chapter I, §1.5, Proposition 4]. Moreover, $GF_c = X$. Take a point $x \in X$ and neighborhood A of x with compact closure. Since the covering $\{W_j\}_{j \in J}$ is locally finite, then the transporters $G_{AU_{i_j}}$ from A to U_{i_j} are nonempty for only finitely many $j \in J$. Since the sets A and U_{i_j} have compact closure and the action (G, X) is proper, then the transporter G_{AF_o} of A to F_o has compact closure in G. Thus, F_c and F_o are fundamental sets and by construction $F_c \subset F_o \subset U$.

(ii) Let F_c be a closed fundamental set for the action (G, X) like in (i). Define a relation $\mathcal{R}$ on F_c with $x\mathcal{R}y$, $x, y \in F_c$ if and only if $y \in Gx$. We will find a Borel section for the closed fundamental set F_c with respect to the previous natural relation on F_c and then we will show that it is, also, a Borel section for the action (G, X). Obviously $\mathcal{R}$ is an equivalence relation on the separable metric space (F_c, d). Since the action (G, X) is proper each orbit Gx is closed in X, for every $x \in X$. The $\mathcal{R}$-equivalence class of a point $x \in F_c$ is $Gx \cap F_c$, hence it is a closed subset of X, thus it is a complete space with respect to the metric d. If U is an open subset of F_c with respect to the relative topology of F_c then the $\mathcal{R}$-saturation of U is the set $GU \cap F_c$ which is open in F_c hence it is a Borel set. Therefore we can apply [Kallman and Mauldin 1978, Lemma 2] to find a Borel section $S_B \subset F_c$ for the equivalence relation $\mathcal{R}$. Moreover, S_B is a Borel section (and a fundamental set) for the action (G, X), since it is contained in the closed fundamental set F_c. □

Remark 7.5. Note that the assumption that the orbit space $G\backslash X$ is paracompact is automatically satisfied for proper isometric actions. So we can apply Theorem 7.4 in both cases.

Remark 7.6. The statement of Theorem 7.4 cannot be improved by asserting that "There always exists a section S homeomorphic to the orbit space," even if we omit the requirement that a neighborhood U of S is given, as the following simple example shows: Let φ be the rotation by $\pi/2$ on the unit circle and G be the group with two elements generated by φ. The orbit space is homeomorphic to the half-open interval $(0, 1]$ endowed with a non-Euclidean topology (that is, a sequence tending to 0 converges to 1), therefore it cannot be embedded in S^1.

An answer to the question of whether S can be chosen to be homeomorphic to the orbit space may lead to interesting structure-theorems.

Question 7.7. As Theorem 7.4(ii) indicates the notion of a Borel section is remarkably related to that of a fundamental set in the locally compact case and may be, similarly, used for structural theorems. Note that the Borel section S_B, because of its construction, is a *minimal* fundamental set for the action (G, X), that is, for each point $x \in X$ the transporter $G_{\{x\}S_B} = gG_x$ for some $g \in G$. So, it is interesting to ask whether the existing Borel section for the action $(E, X \cup X_l)$ can be reduced, or lead, to a Borel section for the initial action $(\mathrm{Iso}(X), X)$.

References

[Abels 1972] H. Abels, "Enden von Räumen mit eigentlichen Transformationsgruppen", *Comment. Math. Helv.* **47** (1972), 457–473. MR 47 #5852 Zbl 0253.54039

[Bourbaki 1966a] N. Bourbaki, *Elements of mathematics: General topology*, vol. 1, Hermann, Paris, 1966. MR 34 #5044a Zbl 0301.54001

[Bourbaki 1966b] N. Bourbaki, *Elements of mathematics: General topology*, vol. 2, Hermann, Paris, 1966. MR 34 #5044b Zbl 0301.54002

[Engelking 1989] R. Engelking, *General topology*, 2nd ed., Sigma Series in Pure Mathematics **6**, Heldermann, Berlin, 1989. MR 91c:54001 Zbl 0684.54001

[Hjorth 2008] G. Hjorth, "An oscillation theorem for groups of isometries", *Geom. Funct. Anal.* **18**:2 (2008), 489–521. MR 2009f:03059 Zbl 1153.03028

[Kallman and Mauldin 1978] R. R. Kallman and R. D. Mauldin, "A cross section theorem and an application to C^*-algebras", *Proceedings Amer. Math. Soc.* **69**:1 (1978), 57–61. MR 57 #3342 Zbl 0381.28001

[Koszul 1965] J.-L. Koszul, *Lectures on groups of transformations*, Tata Inst. Fund. Res. Studies in Math. **32**, Tata Institute of Fundamental Research, Bombay, 1965. MR 36 #1571 Zbl 0195.04605

[Manoussos and Strantzalos 2007] A. Manoussos and P. Strantzalos, "On embeddings of proper and equicontinuous actions in zero-dimensional compactifications", *Trans. Amer. Math. Soc.* **359**:11 (2007), 5593–5609. MR 2008f:37022 Zbl 1124.37004

Received September 6, 2011.

ANTONIOS MANOUSSOS
FAKULTÄT FÜR MATHEMATIK, SFB 701
UNIVERSITÄT BIELEFELD
POSTFACH 100131
D-33501 BIELEFELD
GERMANY
amanouss@math.uni-bielefeld.de

POLYCHRONIS STRANTZALOS
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ATHENS
PANEPISTIMIOUPOLIS
GR-157 84 ATHENS
GREECE
pstrantz@math.uoa.gr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 259, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.259.445

THETA LIFTS OF STRONGLY POSITIVE DISCRETE SERIES: THE CASE OF $(\widetilde{\mathrm{Sp}}(n), O(V))$

IVAN MATIĆ

Let F denote a nonarchimedean local field of characteristic zero with odd residual characteristic. Using the results of Gan and Savin, in this paper we determine the first occurrence indices and theta lifts of strongly positive discrete series representations of metaplectic groups over F in terms of our recent classification of this class of representations. Also, we determine the first occurrence indices of some strongly positive representations of odd orthogonal groups.

1. Introduction

One of the main issues in the local theta correspondence is a precise determination of the theta lifts of irreducible representations. This problem is by now completely solved for cuspidal representations [Mœglin et al. 1987, Théorème principal] and for discrete series for the dual pair $(\mathrm{Sp}(n), O(V))$ [Muić 2004, Theorems 4.2 and 4.3]. Muić used an inductive procedure to investigate certain embeddings of theta lifts of discrete series representations so as to obtain explicit information about the structure of these lifts and to derive the first occurrence indices.

The description given there is based on the classification of discrete series of the classical groups given in [Mœglin 2002; Mœglin and Tadić 2002], which relies on certain conjectures called the *basic assumption* (we emphasize that Arthur has recently announced a proof of his conjectures about the stable transfer coming from the twisted endoscopy, which should imply the basic assumption). On the other hand, we have recently classified the strongly positive discrete series of metaplectic groups, and our classification uses no hypothesis and can be applied much more generally. It is natural to try to relate this classification to the determination of the lifts of those representations. Thus, it is the purpose of this paper to determine the first occurrence indices of the strongly positive discrete series for the dual pair $(\widetilde{\mathrm{Sp}}(n), O(V))$, where $\widetilde{\mathrm{Sp}}(n)$ is the universal cover of $\mathrm{Sp}(n)$, and to obtain as much information about the structure of theta lifts of such representations as possible.

Muić [2008] has obtained some fundamental results on the structure of theta

MSC2010: primary 22E35, 22E50; secondary 11F27.

Keywords: theta correspondence, metaplectic groups, strongly positive representations.

lifts of discrete series without using the Mœglin–Tadić classification. Although very powerful, the methods used there could not provide an explicit description of the first occurrence indices. Nevertheless, his results have recently been rewritten by Gan and Savin [2012] for the dual pair $(\widetilde{\mathrm{Sp}}(n), O(V))$ over a nonarchimedean field of characteristic zero with odd residual characteristic. Another crucial result of their paper is a natural correspondence between irreducible representations on a certain level of metaplectic and odd orthogonal towers, which partially generalizes results of Waldspurger [1984; 1991].

These results are of much importance for us, because they allow us to start our investigation of the first occurrence index with the lift that is a discrete series representation at a quite low level of the tower. The disadvantage of this approach is that it prevents us from determining both first occurrence indices when lifting from the metaplectic tower. So we determine just the lower one.

We do not adopt the methods used in [Muić 2004], choosing rather to describe theta lifts of strongly positive discrete series directly from their cuspidal supports. The advantage of using this method lies in the fact that the structure of the obtained theta lifts can be explicitly described in a purely combinatorial way.

We now describe the contents of this paper. The next section presents some preliminaries, while in Section 3, we summarize without proofs the relevant material on the strongly positive discrete series. In that section we also obtain some useful embeddings of the general discrete series representations. Section 4 provides a detailed exposition of the results about Howe correspondence, which will be used through the paper. Section 5 is the technical heart of the paper, containing several results regarding the theta lifts of irreducible representations.

In Section 6, we state and prove our main results about the lifts of strongly positive irreducible representations of the metaplectic groups, using case-by-case consideration. In Section 7, we determine the first occurrence indices of certain strongly positive representations of the odd orthogonal groups. The observed cases happen to be quite similar in both directions, so the proofs made in the sixth section help us shorten those in the seventh one.

However, for the sake of completeness and to avoid possible confusion, we discuss the details of the lifts of representations of the metaplectic groups and those of the orthogonal ones in separate sections.

2. Notations and preliminaries

Let F be a nonarchimedean local field of characteristic zero with odd residual characteristic.

For a reductive group G, let $\mathrm{Irr}(G)$ stand for the set of isomorphism classes of irreducible admissible (genuine) representations of G.

First we discuss the groups that we consider.

Let V_0 be an anisotropic quadratic space over F of odd dimension. Then its dimension can only be 1 or 3. For more details about the invariants of this space, such as the quadratic character χ_{V_0} related to the quadratic form on V_0, we refer the reader to [Kudla 1986] and [Kudla and Rallis 2005]. In each step we add a hyperbolic plane and obtain an enlarged quadratic space, a tower of quadratic spaces, and a tower of corresponding orthogonal groups. In the case when r hyperbolic planes are added to the anisotropic space, the enlarged quadratic space will be denoted by V_r, while a corresponding orthogonal group will be denoted by $O(V_r)$. Set $m_r = (1/2)\dim V_r$.

To a fixed quadratic character χ_{V_0}, one can attach two odd orthogonal towers, one with $\dim V_0 = 1$ (+-tower) and the other with $\dim V_0 = 3$ (−-tower), as in Chapter V of [Kudla 1996]. In that case, for corresponding orthogonal groups of the spaces obtained by adding r hyperbolic planes, we write $O(V_r^+)$ and $O(V_r^-)$.

Let $S_1(n)$ be the Grothendieck group of the category of all admissible representations of finite length of $O(V_n)$ (that is, a free abelian group over the set of all irreducible representations of $O(V_n)$), and define $S_1 = \bigoplus_{n\geq 0} S_1(n)$.

Let $\widetilde{\mathrm{Sp}}(n)$ be the metaplectic group of rank n, the unique nontrivial two-fold central extension of symplectic group $\mathrm{Sp}(n, F)$. In other words, the following holds:

$$1 \to \mu_2 \to \widetilde{\mathrm{Sp}}(n) \to \mathrm{Sp}(n, F) \to 1,$$

where $\mu_2 = \{1, -1\}$. The multiplication in $\widetilde{\mathrm{Sp}}(n)$ (which is as a set given by $\mathrm{Sp}(n, F) \times \mu_2$) is given by Rao's cocycle [Ranga Rao 1993]. More details on the structural theory of metaplectic groups can be found in [Hanzer and Muić 2010], [Kudla 1996], and [Ranga Rao 1993].

In this paper we are interested only in genuine representations of $\widetilde{\mathrm{Sp}}(n)$ (that is, those that do not factor through μ_2). So, let $S_2(n)$ be the Grothendieck group of the category of all admissible genuine representations of finite length of $\widetilde{\mathrm{Sp}}(n)$ and define $S_2 = \bigoplus_{n\geq 0} S_2(n)$.

Let $\widetilde{\mathrm{GL}}(n, F)$ be a double cover of $\mathrm{GL}(n, F)$, where the multiplication is given by

$$(g_1, \epsilon_1)(g_2, \epsilon_2) = (g_1 g_2, \epsilon_1\epsilon_2(\det g_1, \det g_2)_F).$$

Here $\epsilon_i \in \mu_2$, $i = 1, 2$, and $(\cdot\,, \cdot)_F$ denotes the Hilbert symbol of the field F.

The pair $(\mathrm{Sp}(n), O(V_r))$ is a reductive dual pair in $\mathrm{Sp}(n \cdot \dim V_r)$. Since the dimension of the space V_r is odd, the theta correspondence relates the representations of the metaplectic group $\widetilde{\mathrm{Sp}}(n)$ and those of the orthogonal group $O(V_r)$. We use the abbreviation $n_1 = n \cdot \dim V_r$. Let $\omega_{n_1,\psi}$ be the Weil representation of $\widetilde{\mathrm{Sp}}(n_1)$ depending on the nontrivial additive character ψ, and let $\omega_{n,r}$ denote the pull-back of that representation to the pair $(\widetilde{\mathrm{Sp}}(n), O(V_r))$.

Here and subsequently, ψ denotes a nontrivial additive character of F. Further, we fix a character $\chi_{V,\psi}$ of $\widetilde{\mathrm{GL}}(n, F)$ given by

$$\chi_{V,\psi}(g, \epsilon) = \chi_V(\det g)\epsilon\gamma(\det g, \tfrac{1}{2}\psi)^{-1}.$$

Here γ denotes the Weil invariant, while χ_V is a character related to the quadratic form on $O(V_r)$. We write $\alpha = \chi_{V,\psi}^2$ and observe that α is a quadratic character on $\mathrm{GL}(n, F)$.

Let

$$\mathcal{R}^{\mathrm{gen}} = \bigoplus_{n\geq 0} \mathcal{R}^{\mathrm{gen}}(n),$$

where $\mathcal{R}^{\mathrm{gen}}(n)$ denotes the Grothendieck group of smooth genuine representations of finite length of $\widetilde{\mathrm{GL}}(n, F)$. Similarly, define

$$\mathcal{R} = \bigoplus_{n\geq 0} \mathcal{R}(n),$$

where $\mathcal{R}(n)$ denotes the Grothendieck group of smooth genuine representations of finite length of $\mathrm{GL}(n, F)$.

To simplify the notation, in the sequel we write

$$\mathcal{R}' = \begin{cases} \mathcal{R} & \text{in the orthogonal case,} \\ \mathcal{R}^{\mathrm{gen}} & \text{in the metaplectic case,} \end{cases}$$

and

$$S' = \begin{cases} S_1 & \text{in the orthogonal case,} \\ S_2 & \text{in the metaplectic case.} \end{cases}$$

By ν we denote the character of $\mathrm{GL}(n, F)$ defined by $|\det|_F$.

An irreducible representation $\sigma \in S'$ is called strongly positive if for each representation $\nu^{s_1}\rho_1 \times \nu^{s_2}\rho_2 \times \cdots \times \nu^{s_k}\rho_k \rtimes \sigma_{\mathrm{cusp}}$, where $\rho_i \in \mathcal{R}'$, $i = 1, 2, \ldots, k$ are irreducible cuspidal unitary representations, $\sigma_{\mathrm{cusp}} \in S'$ is an irreducible cuspidal representation, and $s_i \in \mathbb{R}$, $i = 1, 2, \ldots, k$ such that

$$\sigma \hookrightarrow \nu^{s_1}\rho_1 \times \nu^{s_2}\rho_2 \times \cdots \times \nu^{s_k}\rho_k \rtimes \sigma_{\mathrm{cusp}},$$

we have $s_i > 0$ for each i.

Irreducible strongly positive representations are called strongly positive discrete series.

If $\rho \in \mathcal{R}'(m)$ is an irreducible unitary cuspidal representation, we say that $\Delta = \{\nu^a\rho, \nu^{a+1}\rho, \ldots, \nu^{a+k}\rho\}$ is a segment, where $a \in \mathbb{R}$ and $k \in \mathbb{Z}_{\geq 0}$; and we abbreviate $\{\nu^a\rho, \nu^{a+1}\rho, \ldots, \nu^{a+k}\rho\}$ as $[\nu^a\rho, \nu^{a+k}\rho]$. We denote by $\delta(\Delta)$ the unique irreducible subrepresentation of $\nu^{a+k}\rho \times \nu^{a+k-1}\rho \times \cdots \times \nu^a\rho$. This $\delta(\Delta)$ is an essentially square-integrable representation attached to the segment Δ.

For every irreducible cuspidal representation $\rho \in \mathcal{R}'(m)$, there exists a unique $e(\rho) \in \mathbb{R}$ such that the representation $\nu^{-e(\rho)}\rho$ is a unitary cuspidal representation. From now on, let $e\big([\nu^a\rho, \nu^b\rho]\big) = (a+b)/2$.

For an ordered partition $s = (n_1, n_2, \dots, n_j)$ of some $m \le n$, we denote by P_s a standard parabolic subgroup of $\mathrm{Sp}(n, F)$ (consisting of block upper-triangular matrices) whose Levi factor equals

$$\mathrm{GL}(n_1) \times \mathrm{GL}(n_2) \times \cdots \times \mathrm{GL}(n_j) \times \mathrm{Sp}(n - |s|, F),$$

where $|s| = m = \sum_{i=1}^{j} n_i$. Then the standard parabolic subgroup $\widetilde{P_s}$ of $\widetilde{\mathrm{Sp}}(n)$ is the preimage of P_s in $\widetilde{\mathrm{Sp}}(n)$. We have the analogous notation for the Levi subgroups of the metaplectic groups, which are described in more detail in Section 2.2 of [Hanzer and Muić 2010]. The standard parabolic subgroups (containing the upper triangular Borel subgroup) of $O(V_r)$ have an analogous description to the standard parabolic subgroups of $\mathrm{Sp}(n, F)$. If $\widetilde{P_s}$ is a standard parabolic subgroup of $\widetilde{\mathrm{Sp}}(n)$ described above, or P_s is a similar standard parabolic subgroup of $O(V_r)$, the normalized Jacquet module of a smooth representation σ of $\widetilde{\mathrm{Sp}}(n)$ (resp. $O(V_r)$) with respect to $\widetilde{P_s}$ (resp. P_s) is denoted by $R_{\widetilde{P_s}}(\sigma)$ (resp. $R_{P_s}(\sigma)$). From now on, $R_{P_1}(\pi)(\chi)$ (or $R_{\widetilde{P_1}}(\pi)(\chi)$) stands for the isotypic component of $R_{P_1}(\pi)$ along the generalized character χ.

Also, in dealing with Jacquet modules of $\omega_{n,r}$, we use the shorthand $R_{P_1}(\omega_{n,r})$ (resp. $R_{\widetilde{P_1}}(\omega_{n,r})$) for $R_{\widetilde{\mathrm{Sp}}(n)\times P_1}(\omega_{n,r})$ (resp. $R_{\widetilde{P_1}\times O(V_m)}(\omega_{n,r})$), following the notation of [Hanzer and Muić 2011].

For any irreducible representation $\pi \in S'(n)$, there exist an ordered partition $s = (n_1, n_2, \dots, n_j)$ of some $m \le n$, cuspidal representations $\rho_i \in \mathrm{Irr}(\mathcal{R}'(n_i))$, and $\pi_{\mathrm{cusp}} \in S'(n-|s|)$ such that π is an irreducible subquotient of the induced representation $\rho_1 \times \rho_2 \times \cdots \times \rho_j \rtimes \pi_{\mathrm{cusp}}$. In this situation, we write $[\pi] = [\rho_1, \rho_2, \dots, \rho_j; \pi_{\mathrm{cusp}}]$, following the notation used in [Kudla 1996].

Let $\sigma \in S'(n)$ denote an irreducible representation. To simplify notation, set $P'_s = P_s$ in the orthogonal case and $P'_s = \widetilde{P_s}$ in the metaplectic one. We define $\mu^*(\sigma) \in \mathcal{R}' \otimes S'$ by

$$\mu^*(\sigma) = \sum_{k=0}^{n} \mathrm{s.s.}(P'_{(k)}(\sigma)),$$

where s.s. denotes the semisimplification. We extend μ^* linearly to the whole of S'.

In the following lemma, we recall a useful formula for calculations with Jacquet modules, which is valid in both the orthogonal and metaplectic cases [Tadić 1995; Hanzer and Muić 2010]. Set $\alpha' = \alpha$ in the metaplectic case, while in the orthogonal case α' denotes a trivial character.

Lemma 2.1. *Let $\rho \in \mathcal{R}'$ be an irreducible cuspidal representation and let $a, b \in \mathbb{R}$ be such that $a + b \in \mathbb{Z}_{\ge 0}$. Let $\sigma \in S'$ be an admissible representation of finite length.*

Write $\mu^*(\sigma) = \sum_{\pi,\sigma'} \pi \otimes \sigma'$. *Then the following holds*:

$$(1)\quad \mu^*\big(\delta([\nu^{-a}\rho, \nu^b\rho]) \rtimes \sigma\big) = \sum_{i=-a-1}^{b} \sum_{j=i}^{b} \sum_{\pi,\sigma'} \delta\big([\nu^{-i}\alpha'\tilde{\rho}, \nu^{a}\alpha'\tilde{\rho}]\big) \times\delta\big([\nu^{j+1}\rho, \nu^b\rho]\big) \times \pi \otimes \delta\big([\nu^{i+1}\rho, \nu^j\rho]\big) \rtimes \sigma'.$$

We omit $\delta\big([\nu^x\rho, \nu^y\rho]\big)$ *if* $x > y$.

We take a moment to recall the formulation of the second Frobenius isomorphism.

Generally, for some reductive group G', its parabolic subgroup P' with the Levi subgroup M', and its opposite parabolic subgroup $\bar{P}'$, the second Frobenius isomorphism is

$$\mathrm{Hom}_{G'}\big(\mathrm{Ind}_{M'}^{G'}(\pi), \Pi\big) \cong \mathrm{Hom}_{M'}\big(\pi, R_{\bar{P}'}(\Pi)\big),$$

for some smooth representation π (resp. Π) of the group M' (resp. G'). We denote the space of the representation π by V_π.

The above isomorphism can be explicitly described in the following way: let Ψ denote the embedding

$$\Psi : V_\pi \hookrightarrow R_{\bar{P}'}\big(\mathrm{Ind}_{M'}^{G'}(V_\pi)\big),$$

which corresponds to the open cell $P'\bar{P}'$ in G' [Bernstein 1987]. Now, for some $T \in \mathrm{Hom}_{G'}\big(\mathrm{Ind}_{M'}^{G'}(\pi), \Pi\big)$, compose Ψ with the corresponding mapping

$$T_{\bar{P}'} : R_{\bar{P}'}\big(\mathrm{Ind}_{M'}^{G'}(\pi)\big) \to R_{\bar{P}'}(\Pi).$$

3. Embeddings of discrete series

In this section we recall the classification of strongly positive discrete series and obtain further embeddings of general discrete series that will be used later.

In the following theorem, we gather the results obtained in Section 5 of [Matić 2011]. The arguments used there rely on Jacquet module methods, and build up in an essentially combinatorial way from the cuspidal reducibility values. Moreover, the underlying combinatorics are essentially the same for classical groups. Thus, our classification is valid for both metaplectic and orthogonal groups.

Theorem 3.1. *We define a collection of pairs* (Jord, σ'), *where* σ' *is an irreducible cuspidal representation of some* $S'(n_{\sigma'})$ *and* Jord *has the following form*: Jord $= \bigcup_{i=1}^{k} \bigcup_{j=1}^{k_i} \{(\rho_i, b_j^{(i)})\}$, *where*:

- $\{\rho_1, \rho_2, \dots, \rho_k\}$ *is a* (*possibly empty*) *set of mutually nonisomorphic irreducible self-dual cuspidal representations of some* $\mathcal{R}'(m_1), \mathcal{R}'(m_2), \dots, \mathcal{R}'(m_k)$ *such that* $\nu^{a_{\rho_i}}\rho_i \rtimes \sigma'$ *reduces for* $a_{\rho_i} > 0$ (*this defines* a_{ρ_i}).
- $k_i = \lceil a_{\rho_i} \rceil$, *the smallest integer that is not smaller than* a_{ρ_i}.

- *For each $i = 1, \dots, k$, the sequence $b_1^{(i)}, \dots, b_{k_i}^{(i)}$ consists of real numbers such that $a_{\rho_i} - b_j^{(i)}$ is an integer, for $j = 1, 2, \dots, k_i$ and $-1 < b_1^{(i)} < b_2^{(i)} < \cdots < b_{k_i}^{(i)}$.*

There is a bijective correspondence between the set of all irreducible strongly positive representations in S' and the set of all pairs (Jord, σ').

We describe this correspondence more precisely. The pair corresponding to an irreducible strongly positive representation $\sigma \in S'$ is denoted by $\big(\mathrm{Jord}(\sigma), \sigma'(\sigma)\big)$.

Suppose that cuspidal support of σ is contained in the set

$$\{\nu^x \rho_1, \dots, \nu^x \rho_k, \sigma_{\mathrm{cusp}} : x \in \mathbb{R}\},$$

with k minimal (here ρ_i denotes an irreducible cuspidal self-dual representation of some $\mathcal{R}'(n_{\rho_i})$).

Let $a_{\rho_i} > 0$, $i = 1, 2, \dots, k$ denote the unique positive $s \in \mathbb{R}$ such that the representation $\nu^s \rho_i \rtimes \sigma_{\mathrm{cusp}}$ reduces. Set $k_i = \lceil a_{\rho_i} \rceil$. For each $i = 1, 2, \dots, k$, there exists a unique increasing sequence of real numbers $b_1^{(i)}, b_2^{(i)}, \dots, b_{k_i}^{(i)}$, where $a_{\rho_i} - b_j^{(i)}$ is an integer, for $j = 1, 2, \dots, k_i$ and $b_1^{(i)} > -1$, such that σ is the unique irreducible subrepresentation of the induced representation

$$\Big(\prod_{i=1}^{k} \prod_{j=1}^{k_i} \delta\big([\nu^{a_{\rho_i} - k_i + j} \rho_i, \nu^{b_j^{(i)}} \rho_i]\big)\Big) \rtimes \sigma_{\mathrm{cusp}}.$$

Now, $\mathrm{Jord}(\sigma) = \bigcup_{i=1}^{k} \bigcup_{j=1}^{k_i} \{(\rho_i, b_j^{(i)})\}$ *and* $\sigma'(\sigma) = \sigma_{\mathrm{cusp}}$.

We note that results of [Arthur 2011] should imply that every a_{ρ_i} in the previous theorem is half integral.

This classification implies some interesting properties of strongly positive discrete series, which are listed in the next two lemmas.

Lemma 3.2 [Matić 2012, Lemma 3.5]. *Let $\sigma \in S'$ be a strongly positive discrete series. Then σ is uniquely determined by* $[\sigma]$.

The next result follows rather straightforwardly from the classification above:

Lemma 3.3. *Let $\sigma \in S'$ denote a strongly positive discrete series and suppose that $\nu^x \rho$ appears in $[\sigma]$, where $\rho \in \mathcal{R}'$ is an irreducible unitarizable cuspidal representation and $|x| \le 1$. Then the representation $\nu^x \rho$ appears in $[\sigma]$ with multiplicity one. Also, if $\nu^y \rho$ appears in $[\sigma]$ for some $y \neq x$, then $|y| > 1$.*

Proof. It is enough to prove the lemma for $x \ge 0$, since otherwise the same conclusion can be drawn for $|x|$.

We write σ as the unique irreducible subrepresentation of the induced representation of the form

$$\Big(\prod_{i=1}^{k} \prod_{j=1}^{k_i} \delta\big([\nu^{a_{\rho_i} - k_i + j} \rho_i, \nu^{b_j^{(i)}} \rho_i]\big)\Big) \rtimes \sigma_{\mathrm{cusp}}.$$

Obviously, ρ is isomorphic to ρ_l for some $l \in \{1, 2, \dots, k\}$.

By the assumption of the lemma, there is some $j \in \{1, 2, \dots, k_l\}$ such that $a_{\rho_l} - k_l + j \leq x \leq b_j^{(l)}$. Strong positivity of σ implies $x > 0$. Since $a_{\rho_l} - k_l + j > 1$ for $j \geq 2$, it follows that $\nu^x \rho$ appears in the segment

$$\left[\nu^{a_{\rho_l} - k_l + 1} \rho_l, \nu^{b_1^{(l)}} \rho_l\right]$$

and $\nu^x \rho$ does not appear in $\left[\nu^{a_{\rho_l} - k_l + j} \rho_l, \nu^{b_j^{(l)}} \rho_l\right]$, for $j \geq 2$. Further, using $x - 1 \leq 0$, we obtain $x = a_{\rho_l} - k_l + 1$.

Consequently, $\nu^x \rho$ appears in $[\sigma]$ with multiplicity one.

The inequality $|y| > 1$ for $y \neq x$ such that $\nu^y \rho$ appears in $[\sigma]$ is a consequence of the fact that $|y| - x$ is a positive integer and $x > 0$. □

The principal significance of the following lemma is that it allows us to obtain certain embeddings of general discrete series.

Lemma 3.4. *Suppose that $\pi \in S'(n)$ is an irreducible representation that is not in the discrete series. Then there exists an embedding of the form*

$$\pi \hookrightarrow \delta\big([\nu^a \rho, \nu^b \rho]\big) \rtimes \pi',$$

where $a + b \leq 0$ and $\rho \in \mathcal{R}'$ and $\pi' \in S'$ are irreducible representations.

Proof. We adopt the approach from Section 3 of [Matić 2011], which was motivated by [Muić 2006]. Suppose that

$$\pi \hookrightarrow \rho_1 \times \rho_2 \times \cdots \times \rho_k \rtimes \pi_{\text{cusp}}$$

is an embedding of the representation π contradicting Casselman's square-integrability criterion (whose metaplectic version is written in [Ban and Jantzen 2009]), $\rho_i \in \mathcal{R}'$ is an irreducible cuspidal representation for $i \in \{1, 2, \dots, k\}$, and $\pi_{\text{cusp}} \in S'(n')$ is an irreducible cuspidal representation. Further, we consider all possible embeddings of the form

$$\pi \hookrightarrow \delta(\Delta_1) \times \delta(\Delta_2) \times \cdots \times \delta(\Delta_m) \rtimes \pi_{\text{cusp}},$$

contradicting the square-integrability criterion, where $\Delta_1 + \Delta_2 + \cdots + \Delta_m = \{\rho_1, \rho_2, \dots, \rho_k\}$, viewed as the equality of multisets. Clearly, $e(\Delta_i) \leq 0$ for some $i \in \{1, 2, \dots, m\}$. The set of all such embeddings is obviously finite and nonempty.

Each $\delta(\Delta_i)$ is an irreducible representation of some $\mathcal{R}'(n_i)$ (this defines n_i), for $i = 1, 2, \dots, m$. To every such embedding we attach an $(n - n')$-tuple

$$\big(e(\Delta_1), \dots, e(\Delta_1), e(\Delta_2), \dots, e(\Delta_2), \dots, e(\Delta_m), \dots, e(\Delta_m)\big) \in \mathbb{R}^{n-n'},$$

where $e(\Delta_i)$ appears n_i times.

Denote by

$$\pi \hookrightarrow \delta(\Delta'_1) \times \delta(\Delta'_2) \times \cdots \times \delta(\Delta'_{m'}) \rtimes \pi_{\text{cusp}}$$

the minimal such embedding with respect to the lexicographic ordering on $\mathbb{R}^{n-n'}$. In the same way as in the proof of Theorem 3.3 of [Matić 2011], we conclude $e(\Delta'_1) \le e(\Delta'_2) \le \cdots \le e(\Delta'_{m'})$. This gives $e(\Delta'_1) \le 0$. Now Lemma 3.2 of [Mœglin and Tadić 2002] finishes the proof. □

We are ready to describe useful embeddings of general discrete series (this parallels the result of Lemma 3.1 of [Mœglin 2002]).

Theorem 3.5. *Let $\sigma \in S'(n)$ denote a discrete series representation. Then there exists an embedding of the form*

$$\sigma \hookrightarrow \delta\big([\nu^{a_1}\rho_1, \nu^{b_1}\rho_1]\big) \times \delta\big([\nu^{a_2}\rho_2, \nu^{b_2}\rho_2]\big) \times \cdots \times \delta\big([\nu^{a_k}\rho_k, \nu^{b_k}\rho_k]\big) \rtimes \sigma_{\mathrm{sp}},$$

where $a_i \le 0$ and $a_i + b_i > 0$, and $\rho_i \in \mathcal{R}'$ is an irreducible representation for $i = 1, 2, \ldots, k$, while $\sigma_{\mathrm{sp}} \in S'$ is a strongly positive discrete series (*we allow $k = 0$*).

Proof. If σ is a strongly positive discrete series, then $k = 0$ and $\sigma \simeq \sigma_{\mathrm{sp}}$. Thus, we may suppose that σ is not strongly positive.

Again, we start with an embedding of the representation σ of the form

$$\sigma \hookrightarrow \rho_1 \times \rho_2 \times \cdots \times \rho_k \rtimes \sigma_{\mathrm{cusp}},$$

where each $\rho_i \in \mathcal{R}'$ is an irreducible cuspidal representation and $\sigma_{\mathrm{cusp}} \in S'(n')$ is a partial cuspidal support of σ, and consider all possible embeddings of the form

$$\sigma \hookrightarrow \delta(\Delta_1) \times \delta(\Delta_2) \times \cdots \times \delta(\Delta_m) \rtimes \sigma_{\mathrm{cusp}},$$

where $\Delta_1 + \Delta_2 + \cdots + \Delta_m = \{\rho_1, \rho_2, \ldots, \rho_l\}$, viewed as the equality of multisets. In the same way as in the proof of the previous lemma, to every such embedding we attach an element of $\mathbb{R}^{n-n'}$ and denote by

$$\sigma \hookrightarrow \delta(\Delta'_1) \times \delta(\Delta'_2) \times \cdots \times \delta(\Delta'_{m'}) \rtimes \sigma_{\mathrm{cusp}} \tag{2}$$

the minimal such embedding with respect to the lexicographic ordering on $\mathbb{R}^{n-n'}$. Analysis similar to that in the proof of Theorem 3.3 of [Matić 2011] shows $e(\Delta'_1) \le e(\Delta'_2) \le \cdots \le e(\Delta'_{m'})$.

Write each element of the multiset $\{\rho_1, \rho_2, \ldots, \rho_l\}$ in form $\rho_i = \nu^{a_i}\rho_{i,u}$, where $\rho_{i,u}$ is an irreducible unitary cuspidal representation. Define

$$a = \min\{a_i : 1 \le i \le l\}.$$

The assumption that σ is not strongly positive yields $a \le 0$. Suppose that $\nu^a\rho$ appears in the segment Δ'_i, with i minimal (for appropriate ρ). Then $\Delta'_i = [\nu^a\rho, \nu^b\rho]$, for some b.

If the segment Δ'_i is not connected in the sense of Zelevinsky with any of the segments $\Delta'_1, \ldots, \Delta'_{i-1}$, we obtain the embedding

$$\sigma \hookrightarrow \delta(\Delta'_i) \times \delta(\Delta'_1) \times \cdots \times \delta(\Delta'_{m'}) \rtimes \sigma_{\mathrm{cusp}}.$$

Suppose that there is some segment Δ'_j, $1 \le j \le i-1$, such that the segments Δ'_i and Δ'_j are connected in the sense of Zelevinsky. We choose the largest such j and denote it by j again. Also, we write $\Delta'_j = [\nu^{a'}\rho, \nu^{b'}\rho]$. The intertwining operator $\delta(\Delta'_j) \times \delta(\Delta'_i) \to \delta(\Delta'_i) \times \delta(\Delta'_j)$ gives the maps

$$\begin{aligned}\sigma &\hookrightarrow \delta(\Delta'_1) \times \cdots \times \delta(\Delta'_j) \times \delta(\Delta'_i) \times \cdots \times \delta(\Delta'_{m'}) \rtimes \sigma_{\text{cusp}} \\ &\to \delta(\Delta'_1) \times \cdots \times \delta(\Delta'_i) \times \delta(\Delta'_j) \times \cdots \times \delta(\Delta'_{m'}) \rtimes \sigma_{\text{cusp}}.\end{aligned}$$

Observe that the kernel of the previous intertwining operator equals

$$\delta(\Delta'_1) \times \cdots \times \delta\big([\nu^a\rho, \nu^{b'}\rho]\big) \times \delta\big([\nu^{a'}\rho, \nu^b\rho]\big) \times \cdots \times \delta(\Delta'_{m'}) \rtimes \sigma_{\text{cusp}}.$$

Since $e(\Delta'_j) \le e(\Delta'_i)$, the inequality $a < a'$ implies $e\big([\nu^a\rho, \nu^{b'}\rho]\big) < e(\Delta'_j)$. Thus, the minimality of the embedding (2) shows that σ is not contained in the kernel of the observed intertwining operator, which gives

$$\sigma \hookrightarrow \delta(\Delta'_1) \times \cdots \times \delta(\Delta'_i) \times \delta(\Delta'_j) \times \cdots \times \delta(\Delta'_{m'}) \rtimes \sigma_{\text{cusp}}.$$

Repeated application of the above procedure enables us to obtain the embedding

$$\sigma \hookrightarrow \delta(\Delta'_i) \times \delta(\Delta'_1) \times \cdots \times \delta(\Delta'_{m'}) \rtimes \sigma_{\text{cusp}}.$$

Lemma 3.2 of [Mœglin and Tadić 2002] implies that there is some irreducible representation σ_1 such that $\sigma \hookrightarrow \delta\big([\nu^a\rho, \nu^b\rho]\big) \rtimes \sigma_1$. Square-integrability of σ shows $a+b>0$. We claim that σ_1 is a discrete series representation.

Suppose, on the contrary, that σ_1 is not in the discrete series. Then the previous lemma shows that it can be written as a subrepresentation of the induced representation of the form $\delta\big([\nu^x\rho', \nu^y\rho']\big) \rtimes \sigma'_1$, where $x+y \le 0$. Thus, $\sigma \hookrightarrow \delta\big([\nu^a\rho, \nu^b\rho]\big) \times \delta\big([\nu^x\rho', \nu^y\rho']\big) \rtimes \sigma'_1$. Square-integrability of the representations σ shows that the segments $[\nu^a\rho, \nu^b\rho]$ and $[\nu^x\rho', \nu^y\rho']$ are connected in the sense of Zelevinsky, and consequently $\sigma \hookrightarrow \delta\big([\nu^a\rho, \nu^y\rho]\big) \times \delta\big([\nu^x\rho', \nu^b\rho']\big) \rtimes \sigma'_1$.

The choice of a shows that $a \le x$, which leads to $a+y \le x+y \le 0$; that is, $e\big([\nu^a\rho, \nu^y\rho]\big) \le 0$, contradicting the square-integrability of σ. In this way we have proved that σ_1 is also a discrete series representation.

We continue in this fashion to obtain that either σ_1 is strongly positive or it can be written as a subrepresentation of the induced representation of the form $\delta\big([\nu^{a'}\rho', \nu^{b'}\rho']\big) \rtimes \sigma_2$, where $a' \le 0$ and $\sigma_2 \in S'$ is a discrete series representation. Repeating this procedure, after a finite number of steps we obtain the claim of the theorem. □

4. Howe's correspondence and results of Gan and Savin and of Kudla

In this section we review some results about Howe correspondence.

For an irreducible genuine smooth representation $\sigma \in S_2(n)$, let $\Theta(\sigma, r)$ be a smooth representation of $O(V_r)$, given as the full lift of σ to the r-level of the orthogonal tower, that is, the biggest quotient of $\omega_{n,r}$ on which $\widetilde{\mathrm{Sp}}(n)$ acts as a multiple of σ. As a representation of $\widetilde{\mathrm{Sp}}(n) \times O(V_r)$ it has a form $\sigma \otimes \Theta(\sigma, r)$. We write $\Theta^+(\sigma, r)$ (resp. $\Theta^-(\sigma, r)$) for the lift on the $+$-tower (resp. $-$-tower), when emphasizing the tower.

Similarly, if τ is an irreducible representation of $O(V_r)$, then one has its full lift $\Theta(\tau, n)$, which is a smooth representation of $\widetilde{\mathrm{Sp}}(n)$.

In the following theorem we summarize some basic results about the theta correspondence.

Theorem 4.1 [Kudla 1996; Mœglin et al. 1987]. *Let σ denote an irreducible genuine representation of $\widetilde{\mathrm{Sp}}(n)$. Then there exists an integer $r \geq 0$ such that $\Theta(\sigma, r) \neq 0$. The smallest such r is called the first occurrence index of σ in the orthogonal tower. Also, $\Theta(\sigma, r') \neq 0$ for $r' \geq r$.*

The representation $\Theta(\sigma, r)$ is either zero or it has finite length. If the residual characteristic of field F is other than 2, then $\Theta(\sigma, r)$ is either zero or it has a unique irreducible quotient. Following [Muić 2004], we write $\sigma(r)$ for this unique irreducible quotient.

The analogous statements hold for $\Theta(\tau, n)$ if τ is an irreducible representation of $O(V_r)$.

Now we state the results of Gan and Savin [2012, Section 6 and Theorem 8.1] that serve as a cornerstone for our determination of lifts of the strongly positive discrete series.

Theorem 4.2. *Let F be a nonarchimedean local field of characteristic 0 with odd residual characteristic. For each nontrivial additive character ψ of F, there is an injection*

$$\Theta_\psi : \mathrm{Irr}(\widetilde{\mathrm{Sp}}(n)) \to \mathrm{Irr}(O(V_n^+)) \sqcup \mathrm{Irr}(O(V_{n-1}^-))$$

given by the theta correspondence (with respect to ψ). Suppose that $\sigma \in \mathrm{Irr}(\widetilde{\mathrm{Sp}}(n))$ and $\tau \in \mathrm{Irr}(O(V))$ correspond under Θ_ψ. Then σ is a discrete series representation if and only if τ is a discrete series representation.

Let σ_{cusp} denote an irreducible cuspidal genuine representation of $\widetilde{\mathrm{Sp}}(n')$. We write $\Theta(\sigma, r)$ for the smooth isotypic component of σ in $\omega_{n,r}$. Since σ_{cusp} is cuspidal, for the smallest r' such that $\Theta(\sigma_{\mathrm{cusp}}, r') \neq 0$, we have that $\Theta(\sigma_{\mathrm{cusp}}, r')$ is an irreducible cuspidal representation of $O(V_{r'})$; we denote it by τ_{cusp}.

Let $\rho \in \mathscr{R}$ be an irreducible cuspidal self-contragredient representation. Results of Silberger [1980] (in the orthogonal case) and of Hanzer and Muić [2011] (in the metaplectic case) show that there exist unique nonnegative real numbers s_1 and s_2 such that the induced representations $\nu^{s_1}\rho \rtimes \tau_{\mathrm{cusp}}$ and $\nu^{s_2}\chi_{V,\psi}\rho \rtimes \sigma_{\mathrm{cusp}}$ reduce.

If ρ is not a trivial character of $F^\times$, then $s_1 = s_2$. Otherwise, the representation $\nu^{s_1} \rtimes \tau_{\text{cusp}}$ reduces for $s_1 = |n' - m_{r'}|$, while the representation $\nu^{s_2}\chi_{V,\psi} \rtimes \sigma_{\text{cusp}}$ reduces for $s_2 = |m_{r'} - n' - 1|$, where $m_{r'} = (1/2)\dim V_{r'}$.

We take a moment to state the results from Section 2 of [Kudla 1986], which happen to be crucial for our investigation.

Theorem 4.3. *Let $\tau \in S_1(r)$ denote an irreducible representation and suppose $[\tau] = [\rho_1, \rho_2, \ldots, \rho_k; \tau_{\text{cusp}}]$, with $\tau_{\text{cusp}} \in S_1(r')$ being an irreducible cuspidal representation. Let $\sigma_{\text{cusp}} = \tau(n')$ be the first nonzero lift of the representation τ_{cusp} and observe that $\sigma_{\text{cusp}} \in S_2(n')$ is an irreducible cuspidal representation. Let σ denote an irreducible quotient of $\Theta(\tau, n)$. We have the following possibilities*:

- *If $n \geq n' + r - r'$, then*

$$[\sigma] = \big[\chi_{V,\psi}\nu^{m_r-n}, \chi_{V,\psi}\nu^{m_r-n+1}, \ldots, \chi_{V,\psi}\nu^{m_{r'}-n'-1}, \\ \chi_{V,\psi}\rho_1, \chi_{V,\psi}\rho_2, \ldots, \chi_{V,\psi}\rho_k; \sigma_{\text{cusp}}\big].$$

- *If $n < n' + r - r'$, set $t = r - r' - n + n'$. Then there exist $i_1, i_2, \ldots, i_t \in \{1, 2, \ldots, k\}$ such that $\rho_{i_j} = \nu^{m_r-n-j}$ for $j = 1, 2, \ldots, t$ and*

  $$[\sigma] = \big[\chi_{V,\psi}\rho_1, \ldots, \widehat{\chi_{V,\psi}\rho_{i_1}}, \ldots, \widehat{\chi_{V,\psi}\rho_{i_t}}, \ldots, \chi_{V,\psi}\rho_k; \sigma_{\text{cusp}}\big],$$

  where $\widehat{\chi_{V,\psi}\rho_i}$ means that we omit $\chi_{V,\psi}\rho_i$.

Similarly, let $\sigma \in S_2(n)$ denote an irreducible representation and suppose $[\sigma] = [\chi_{V,\psi}\rho_1, \chi_{V,\psi}\rho_2, \ldots, \chi_{V,\psi}\rho_k; \sigma_{\text{cusp}}]$, with $\sigma_{\text{cusp}} \in S_2(n')$ being an irreducible cuspidal representation. Let $\tau_{\text{cusp}} = \sigma(r')$ be the first nonzero lift of the representation σ_{cusp}, and observe that $\tau_{\text{cusp}} \in S_1(r')$ is an irreducible cuspidal representation. Let τ denote an irreducible quotient of $\Theta(\sigma, r)$. We have the following possibilities:

- *If $r \geq r' + n - n'$, then*

  $$[\tau] = [\nu^{m_r-n-1}, \nu^{m_r-n-2}, \ldots, \nu^{m_{r'}-n'}, \rho_1, \rho_2, \ldots, \rho_k; \tau_{\text{cusp}}].$$

- *If $r < r' + n - n'$, set $t = r' - n' + n - r$. Then there exist $i_1, i_2, \ldots, i_t \in \{1, 2, \ldots, k\}$ such that $\rho_{i_j} = \nu^{m_r-n+j-1}$ for $j = 1, 2, \ldots, t$ and*

  $$[\tau] = [\rho_1, \ldots, \widehat{\rho_{i_1}}, \ldots, \widehat{\rho_{i_t}}, \ldots, \rho_k; \tau_{\text{cusp}}],$$

  where $\widehat{\rho_i}$ means that we omit ρ_i.

The next theorem that we need is Kudla's filtration of Jacquet modules of the oscillatory representation:

Theorem 4.4 [Kudla 1986, Theorem 2.8]. *Let $\omega_{n,r}$ denote the oscillatory representation of the group $\widetilde{\text{Sp}}(n) \times O(V_r)$ corresponding to the nontrivial additive character ψ.*

- *Let P_j denote the standard maximal parabolic subgroup of $O(V_r)$. Then Jacquet module $R_{P_j}(\omega_{n,r})$ has $\widetilde{\mathrm{Sp}}(n) \times M_j$-invariant filtration given by I_{jk}, $0 \le k \le j$, where*

$$I_{jk} \simeq \mathrm{Ind}_{P_{jk} \times \widetilde{P}_k \times O(V_{r-j})}^{\widetilde{\mathrm{Sp}}(n) \times M_j} (\gamma_{jk} \otimes \Sigma'_k \otimes \omega_{n-k,r-j}).$$

Here, P_{jk} is a standard parabolic subgroup of $\mathrm{GL}(j, F)$ corresponding to the partition $(j-k, k)$, γ_{jk} is a character of $\mathrm{GL}(j-k, F) \times \widetilde{\mathrm{GL}}(k, F)$ given by

$$\gamma_{jk}(g_1, g_2) = \nu^{-(m_r - n - (j-k+1)/2)}(g_1)\chi_{V,\psi}(g_2),$$

and Σ'_k is a twist of the standard representation of $\mathrm{GL}(k, F) \times \mathrm{GL}(k, F)$ on the space of smooth locally constant compactly supported complex-valued functions $C_c^\infty(\mathrm{GL}(k, F))$:

$$\Sigma'_k(g_1, g_2) f(g) = \nu^{-(m_{r-j}+(k-1)/2)} \nu^{m_{r-j}+(k-1)/2} f(g_1^{-1} g g_2).$$

In particular, a quotient I_{j0} equals $\nu^{-(m_r-n-(j+1)/2)} \otimes \omega_{n,r-j}$ and a subrepresentation I_{jj} equals

$$\mathrm{Ind}_{\mathrm{GL}(j,F) \times \widetilde{P}_j \times O(V_{r-j})}^{\widetilde{\mathrm{Sp}}(n) \times M_j} (\chi_{V,\psi} \otimes \Sigma'_j \otimes \omega_{n-j,r-j}).$$

- *Let $\widetilde{P}_j$ denote the standard maximal parabolic subgroup of $\widetilde{\mathrm{Sp}}(n)$. Then Jacquet module $R_{\widetilde{P}_j}(\omega_{n,r})$ has $\widetilde{M}_j \times O(V_r)$-invariant filtration given by J_{jk}, $0 \le k \le j$, where*

$$J_{jk} \simeq \mathrm{Ind}_{\widetilde{P}_{jk} \times P_k \times \widetilde{\mathrm{Sp}}(n-j)}^{\widetilde{M}_j \times O(V_r)} (\beta_{jk} \otimes \Sigma'_k \otimes \omega_{n-j,r-k}).$$

Here, $\widetilde{P}_{jk}$ is a standard parabolic subgroup of $\widetilde{\mathrm{GL}}(j, F)$ corresponding to the partition $(j-k, k)$, β_{jk} is a character of $\widetilde{\mathrm{GL}}(j-k, F) \times \widetilde{\mathrm{GL}}(k, F)$ given by

$$\beta_{jk}(g_1, g_2) = \left(\chi_{V,\psi} \nu^{m_r-n-(j-k-1)/2}\right)(g_1)\chi_{V,\psi}(g_2),$$

and Σ'_k is a twist of the standard representation of $\mathrm{GL}(k, F) \times \mathrm{GL}(k, F)$ on the space of smooth locally constant compactly supported complex-valued functions $C_c^\infty(\mathrm{GL}(k, F))$:

$$\Sigma'_k(g_1, g_2) f(g) = \nu^{m_r+(k+1)/2} \nu^{-(m_r+(k+1)/2)} f(g_1^{-1} g g_2).$$

In particular, a quotient J_{j0} equals $\chi_{V,\psi} \nu^{m_r-n+(j-1)/2} \otimes \omega_{n-j,r}$ and a subrepresentation J_{jj} equals $\mathrm{Ind}_{\widetilde{\mathrm{GL}}(j,F) \times P_j \times \widetilde{\mathrm{Sp}}(n-j)}^{\widetilde{M}_j \times O(V_r)} (\chi_{V,\psi} \otimes \Sigma'_j \otimes \omega_{n-j,r-j})$.

5. Some technical results on lifts

The purpose of this section is to state and prove many technical results that will be of particular importance in the following sections.

An elementary but useful criterion for pushing down the lifts of irreducible representations is established by the following two propositions.

Proposition 5.1. *Let $\tau \in S_1(r)$ be an irreducible representation.*

(1) *Suppose that $\Theta(\tau, n) \neq 0$. Then $R_{\widetilde{P}_1}(\Theta(\tau, n+1))(\chi_{V,\psi}\nu^{m_r-(n+1)}) \neq 0$.*

(2) *Suppose that $R_{P_1}(\tau)(\nu^{m_r-(n+1)}) = 0$. Then $\Theta(\tau, n) \neq 0$ if and only if*

$$R_{\widetilde{P}_1}(\Theta(\tau, n+1))(\chi_{V,\psi}\nu^{m_r-(n+1)}) \neq 0.$$

Proof. The proof follows the same lines as that of Theorem 4.5 of [Hanzer and Muić 2011].

Assume that $\Theta(\tau, n) \neq 0$. Then there exists an epimorphism $\omega_{n,r} \to \tau \otimes \Theta(\tau, n)$. Kudla's filtration gives the epimorphisms

$$R_{\widetilde{P}_1}(\omega_{n+1,r}) \to \chi_{V,\psi}\nu^{m_r-(n+1)} \otimes \omega_{n,r} \to \chi_{V,\psi}\nu^{m_r-(n+1)} \otimes \tau \otimes \Theta(\tau, n).$$

Using Frobenius reciprocity, we get a nontrivial intertwining

$$\Theta(\tau, n+1) \to \chi_{V,\psi}\nu^{m_r-(n+1)} \rtimes \Theta(\tau, n).$$

This obviously proves the first statement of the proposition.

It remains to prove sufficiency in the second statement. The condition

$$R_{\widetilde{P}_1}(\Theta(\tau, n+1))(\chi_{V,\psi}\nu^{m_r-(n+1)}) \neq 0$$

gives $\Theta(\tau, n+1) \neq 0$, which gives an epimorphism $\omega_{n+1,r} \to \tau \otimes \Theta(\tau, n+1)$. Applying Jacquet modules, we get an epimorphism

$$R_{\widetilde{P}_1}(\omega_{n+1,r}) \to \tau \otimes \chi_{V,\psi}\nu^{m_r-(n+1)} \otimes \sigma'$$

for some irreducible representation $\sigma' \in S_1(n)$. If we suppose that the restriction of this epimorphism to a subrepresentation J_{11} is nonzero, second Frobenius reciprocity gives a nonzero intertwining map

$$\chi_{V,\psi} \otimes \Sigma_1' \otimes \omega_{n,r-1} \to \widetilde{R_{P_1}(\widetilde{\tau})} \otimes \chi_{V,\psi}\nu^{m_r-(n+1)} \otimes \sigma'.$$

From this intertwining, we deduce $\tau \hookrightarrow \nu^{m_r-(n+1)} \rtimes \tau'$ for some irreducible representation $\tau' \in S_2(r-1)$, contradicting the assumption of the proposition. Consequently, there exists a nonzero intertwining $J_{10} \to \tau \otimes \chi_{V,\psi}\nu^{m_r-(n+1)} \otimes \sigma'$, which gives $\Theta(\tau, n) \neq 0$. □

We omit the proof of the next proposition, since it is completely analogous to the proof of the previous one.

Proposition 5.2. *Let $\sigma \in S_2(n)$ be an irreducible representation.*

(1) *Suppose that $\Theta(\sigma, r) \neq 0$. Then $R_{P_1}(\Theta(\sigma, r+1))(\nu^{-(m_{r+1}-n-1)}) \neq 0$.*

(2) *Suppose that* $R_{\widetilde{P}_1}(\sigma)(\chi_{V,\psi}\nu^{-(m_{r+1}-n-1)}) = 0$. *Then* $\Theta(\sigma, r) \neq 0$ *if and only if* $R_{P_1}(\Theta(\sigma, r+1))(\nu^{-(m_{r+1}-n-1)}) \neq 0$.

Now we prove an important result regarding the square-integrability of the lifts of strongly positive discrete series. In particular, this result gives an alternative and essentially combinatorial proof of a special case of the results of [Muić 2008].

Proposition 5.3. *Let* $\sigma \in S_2(n)$ *denote a strongly positive discrete series. Suppose that* $\Theta(\sigma, r) \neq 0$, *for some* r *such that* $m_r \leq n + \frac{1}{2}$, *and that*

$$R_{\widetilde{P}_1}(\sigma)(\chi_{V,\psi}\nu^{-(m_k-n-1)}) = 0$$

for $k \geq r+1$. *Then* $\sigma(r)$ *is a discrete series representation.*

Proof. We prove this proposition by downwards induction on r, starting with an r such that $m_r = n + \frac{1}{2}$. If $m_r = n + \frac{1}{2}$, Theorem 4.2 shows our claim. Thus, suppose that the claim holds for some $r+1$ such that $m_{r+1} \leq n + \frac{1}{2}$. We prove it for r.

It may be easily concluded from the proof of Proposition 5.1 (in the same way as in the proof of Lemma 5.1 of [Muić 2004]) that there is a nonzero intertwining $\sigma(r) \hookrightarrow \nu^{-(m_r-n-1)} \rtimes \sigma(r-1)$.

Note that in our case, $m_r < n+\frac{1}{2}$, which implies $-(m_r-n-1) \geq \frac{3}{2}$. Now, suppose that $\sigma(r-1)$ is not a discrete series representation. According to Lemma 3.4, there is an embedding $\sigma(r-1) \hookrightarrow \delta([\nu^a\rho, \nu^b\rho]) \rtimes \sigma'$, where $a+b \leq 0$. Obviously, $a \leq 0$.

Since $m_r-n-1 \leq -\frac{3}{2}$, the strong positivity of the representation σ and Lemma 3.3 together with Theorem 4.3 imply there is at most one $x \in \mathbb{R}$, $0 < |x| \leq 1$ such that $\nu^x\rho$ appears in $[\sigma(r-1)]$. Therefore, $b \leq 0$ and the representation $\nu^{-(m_r-n-1)} \times \nu^b\rho$ is irreducible and isomorphic to $\nu^b\rho \times \nu^{-(m_r-n-1)}$.

We thus get the embeddings and isomorphisms

$$\begin{aligned}\sigma(r) &\hookrightarrow \nu^{-(m_r-n-1)} \rtimes \sigma(r-1) \hookrightarrow \nu^{-(m_r-n-1)} \times \delta([\nu^a\rho, \nu^b\rho]) \rtimes \sigma' \\ &\hookrightarrow \nu^{-(m_r-n-1)} \times \nu^b\rho \times \delta([\nu^a\rho, \nu^{b-1}\rho]) \rtimes \sigma' \\ &\simeq \nu^b\rho \times \nu^{-(m_r-n-1)} \times \delta([\nu^a\rho, \nu^{b-1}\rho]) \rtimes \sigma',\end{aligned}$$

contradicting square-integrability of $\sigma(r)$. This proves the proposition. □

In pretty much the same way one can also prove:

Corollary 5.4. *Let* $\tau \in S_1(r)$ *denote a strongly positive discrete series. Suppose that* $\Theta(\tau, n) \neq 0$, *for some* n *such that* $m_r \geq n + \frac{1}{2}$. *Then* $\tau(n)$ *is a discrete series representation.*

The last two propositions of this section contain rather important results on the transfer of certain embeddings by the theta lifts. We omit the proofs, since these results can be obtained in a completely analogous way as in [Muić 2004, Remark 5.2], that is, by precise examination of the filtration of Jacquet modules quoted in Theorem 4.4.

Proposition 5.5. *Suppose that the representation* $\sigma \in \mathrm{Irr}(\widetilde{\mathrm{Sp}}(n))$ *may be written as an irreducible subrepresentation of the induced representation of the form* $\delta([\nu^a \rho, \nu^b \rho]) \rtimes \sigma'$, *where* ρ *is an irreducible cuspidal genuine representation,* $\sigma' \in \mathrm{Irr}(\widetilde{\mathrm{Sp}}(n'))$, *and* $b - a \geq 0$. *Let* $\Theta(\sigma, r) \neq 0$. *Then one of the following holds*:

- *There is an irreducible representation* τ *of some* $O(V_{r'})$ *such that* $\sigma(r)$ *is a subrepresentation of* $\delta([\nu^a \chi_{V,\psi}^{-1} \rho, \nu^b \chi_{V,\psi}^{-1} \rho]) \rtimes \tau$.
- *There is an irreducible representation* τ *of some* $O(V_{r'})$ *such that* $\sigma(r)$ *is a subrepresentation of* $\delta([\nu^{a+1} \chi_{V,\psi}^{-1} \rho, \nu^b \chi_{V,\psi}^{-1} \rho]) \rtimes \tau$.

The latter situation is impossible unless $(a, \rho) = (m_r - n, \chi_{V,\psi})$.

Proposition 5.6. *Suppose that the representation* $\tau \in \mathrm{Irr}(O(V_r))$ *may be written as an irreducible subrepresentation of the induced representation of the form* $\delta([\nu^a \rho, \nu^b \rho]) \rtimes \tau'$, *where* ρ *is an irreducible cuspidal representation and* $\tau' \in \mathrm{Irr}(O(V_{r'}))$ *and* $b - a \geq 0$. *Let* $\Theta(\tau, n) \neq 0$. *Then one of the following holds*:

- *There is an irreducible representation* σ *of some* $\widetilde{\mathrm{Sp}}(n')$ *such that* $\tau(n)$ *is a subrepresentation of* $\delta([\nu^a \chi_{V,\psi} \rho, \nu^b \chi_{V,\psi} \rho]) \rtimes \sigma$.
- *There is an irreducible representation* σ *of some* $\widetilde{\mathrm{Sp}}(n')$ *such that* $\tau(n)$ *is a subrepresentation of* $\delta([\nu^{a+1} \chi_{V,\psi} \rho, \nu^b \chi_{V,\psi} \rho]) \rtimes \sigma$.

The latter situation is impossible unless $(a, \rho) = (n - m_r + 1, 1_{F^\times})$.

6. Lifts of strongly positive discrete series of the metaplectic groups

In this section we determine the structure of certain lifts of the strongly positive discrete series of the metaplectic groups. We also obtain precise information about the first occurrence of strongly positive discrete series in the orthogonal tower, depending on its cuspidal support.

Let $\sigma \in \mathrm{Irr}(\widetilde{\mathrm{Sp}}(n))$ denote a strongly positive discrete series. According to the classification given in Theorem 3.1, we may write σ as a unique irreducible subrepresentation of the induced representation

$$\left(\prod_{i=1}^{k} \prod_{j=1}^{k_i} \delta([\chi_{V,\psi} \nu^{a_{\rho_i} - k_i + j} \rho_i, \chi_{V,\psi} \nu^{b_j^{(i)}} \rho_i]) \right) \rtimes \sigma_{\mathrm{cusp}}, \tag{3}$$

with k minimal and k_i minimal for $i = 1, 2, \ldots, k$, where

$$\sigma_{\mathrm{cusp}} \in \mathrm{Irr}(\widetilde{\mathrm{Sp}}(n'))$$

is an irreducible cuspidal representation and ρ_i an irreducible cuspidal representation of $\mathrm{GL}(n_{\rho_i}, F)$ (this defines n_{ρ_i}) for $i = 1, 2, \dots, k$. We note that the minimality of k and k_i for $i = 1, 2, \dots, k$ implies that there are no empty segments in (3).

Theorem 4.2 shows that either $\Theta^+(\sigma, n) \neq 0$ or $\Theta^-(\sigma, n-1) \neq 0$.

The following theorem describes the first occurrence indices of the strongly positive discrete series of the metaplectic group.

Theorem 6.1. *Let $\sigma \in \mathrm{Irr}(\widetilde{\mathrm{Sp}}(n))$ be a strongly positive discrete series. If*

$$\Theta^+(\sigma, n) \neq 0,$$

let $(\epsilon, r) = (+, n)$; otherwise let $(\epsilon, r) = (-, n-1)$. Suppose that $\sigma_{\mathrm{cusp}} \in \mathrm{Irr}(\widetilde{\mathrm{Sp}}(n'))$ is a partial cuspidal support of σ and $\tau_{\mathrm{cusp}} \in \mathrm{Irr}(O(V^\epsilon_{r'}))$ is the first nonzero lift of σ_{cusp}. Further, set $M = \{|x| : \chi_{V,\psi}\nu^x$ appears in $[\sigma]\}$ and denote by $a_{\min}$ the minimal element of M. If $M = \varnothing$, let $a_{\min} = n' - \frac{1}{2}\dim V^\epsilon_{r'} + 2$.

If $a_{\min} = \frac{1}{2}$ or $n' = r' + \frac{1}{2}(\dim V^\epsilon_0 - 1)$, then the first occurrence index of σ is r. Otherwise, the first occurrence index of σ is $r - a_{\min} + \frac{3}{2}$.

The rest of this section is devoted to the proof of Theorem 6.1. The proof is divided into several cases depending on the structure of the cuspidal support of σ and on the first nonzero lift of σ_{cusp}.

In this section, m_r denotes $\frac{1}{2}\dim V^\epsilon_r = n + \frac{1}{2}$ and $\sigma(l)$ denotes the unique irreducible quotient of the representation $\Theta^\epsilon(\sigma, l)$.

Observe that Proposition 5.2 implies that the representation $\sigma(l)$ is not a discrete series representation for $l > r$.

There are two main cases that we consider.

Suppose the representation $\chi_{V,\psi}\nu^{1/2}$ does not appear in $[\sigma]$. Since $m_r - n = \frac{1}{2}$, Theorem 4.3 yields $n' \geq r' + \frac{1}{2}(\dim(V^\epsilon_0) - 1)$. We have two possibilities:

- $n' = r' + \frac{1}{2}(\dim(V^\epsilon_0) - 1)$:

In this case, both representations $\chi_{V,\psi}\nu^s \rtimes \sigma_{\mathrm{cusp}}$ and $\nu^s \rtimes \tau_{\mathrm{cusp}}$ reduce for $s = \frac{1}{2}$. Therefore, by Theorem 3.1, there is no representation of the form $\chi_{,\psi}\nu^s$ appearing in $[\sigma]$. Further, Theorem 3.5 of [Hanzer and Muić 2011] implies that the representation $\chi_{V,\psi}\nu^s\rho_i \rtimes \sigma_{\mathrm{cusp}}$ reduces if and only if the representation $\nu^s\rho_i \rtimes \tau_{\mathrm{cusp}}$ reduces.

One of the main results of [Gan and Savin 2012] states that $\sigma(r)$ is a discrete series representation. Applying Equation (2), we obtain the embedding

$$\sigma(r) \hookrightarrow \delta\big([\nu^{a_1}\rho'_1, \nu^{b_1}\rho'_1]\big) \times \delta\big([\nu^{a_2}\rho'_2, \nu^{b_2}\rho'_2]\big) \times \cdots \times \delta\big([\nu^{a_l}\rho'_l, \nu^{b_l}\rho'_l]\big) \rtimes \tau_{\mathrm{sp}},$$

where $a_i \leq 0$ and $\rho'_i \in \{\rho_1, \rho_2, \dots, \rho_k\}$ for $i = 1, 2, \dots, l$, and $\tau_{\mathrm{sp}} \in \mathrm{Irr}(O(V^\epsilon_{r'}))$ is a strongly positive discrete series for some r'.

Since the representation σ is strongly positive, Theorem 4.3 and Lemma 3.3 show that for every $i \in \{1, 2, \dots, k\}$, there is at most one representation of the form $\nu^x \rho_i$ that appears in $[\sigma(r)]$ with $0 \leq |x| < 1$. In the same way as in the proof of Proposition 5.3, we deduce $\sigma(r) \simeq \tau_{\mathrm{sp}}$, that is, $\sigma(r)$ is a strongly positive representation.

It is now easy to see, using Lemma 3.2, that $\sigma(r)$ is a unique irreducible subrepresentation of the induced representation

$$\left(\prod_{i=1}^{k} \prod_{j=1}^{k_i} \delta\big([\nu^{a_{\rho_i}-k_i+j}\rho_i, \nu^{b_j^{(i)}}\rho_i]\big)\right) \rtimes \tau_{\mathrm{cusp}}.$$

Suppose that $\Theta^{\epsilon}(\sigma, r-1) \neq 0$. Then Proposition 5.2 implies $R_{P_1}(\Theta^{\epsilon}(\sigma, r))(\nu^{1/2}) \neq 0$, which is impossible. Thus, r is the first occurrence index of σ.

- $n' > r' + \frac{1}{2}(\dim(V_0^{\epsilon}) - 1)$.

In this case, the representation $\chi_{V,\psi}\nu^s \rtimes \sigma_{\mathrm{cusp}}$ reduces for $s = n' - m_{r'} + 1$, and the representation $\nu^s \rtimes \tau_{\mathrm{cusp}}$ reduces for $s = n' - m_{r'}$.

Observe that $[\sigma(r)]$ is obtained from $[\sigma]$ by multiplying by $\chi_{V,\psi}^{-1}$ all representations of the form $\chi_{V,\psi}\nu^x\rho_i$ appearing in $[\sigma]$, adding the representations $\nu^{-1/2}$, $\nu^{-3/2}, \ldots, \nu^{m_{r'}-n'}$, and replacing σ_{cusp} with τ_{cusp}.

There are two possible cases that we consider:

(1) Some representation of the form $\chi_{V,\psi}\nu^s$, $s \in \mathbb{R}$ appears in $[\sigma]$: We may suppose that ρ_1 is a trivial representation. Note that $a_{\rho_1} - k_1 + 1$ is strictly greater than $\frac{1}{2}$ and that a_{ρ_1} equals $n' - m_{r'} + 1$.

For simplicity of notation, let a_j stand for $a_{\rho_1} - k_1 + j$, for $j = 1, 2, \ldots, k_1$. Again, we know that $\sigma(r)$ is a discrete series representation. Inspecting its cuspidal support more precisely, it is not hard to see that it has to be strongly positive. Using Lemma 3.2, we get that $\sigma(r)$ can be obtained as the unique irreducible subrepresentation of

$$\nu^{1/2} \times \nu^{3/2} \times \cdots \times \nu^{a_1-2} \times \left(\prod_{j=1}^{k_1} \delta\big([\nu^{a_j-1}, \nu^{b_j^{(1)}}]\big)\right) \\ \times \left(\prod_{i=2}^{k} \prod_{j=1}^{k_i} \delta\big([\nu^{a_{\rho_i}-k_i+j}\rho_i, \nu^{b_j^{(i)}}\rho_i]\big)\right) \rtimes \tau_{\mathrm{cusp}}.$$

If $a_1 \geq \frac{5}{2}$, Theorem 5.3 of [Matić 2012] implies $R_{P_1}(\sigma(r))(\nu^{1/2}) \neq 0$. If $a_1 = \frac{3}{2}$, the same result shows that $R_{P_1}(\sigma(r))(\nu^{1/2}) = 0$ (since $b_1^{(1)} \geq a_1 > \frac{1}{2}$). Using Proposition 5.2, we conclude that $\Theta^{\epsilon}(\sigma, r-1) \neq 0$ if $a_1 \geq \frac{5}{2}$, and $\Theta^{\epsilon}(\sigma, r-1) = 0$ otherwise.

If $a_1 \geq \frac{5}{2}$, combining the square-integrability of $\sigma(r-1)$ (by Proposition 5.3) with the fact that $[\sigma(r-1)]$ is obtained from $[\sigma(r)]$ by subtracting $\nu^{1/2}$, we get that $\sigma(r-1)$ is a strongly positive discrete series that can be realized as a unique irreducible subrepresentation of

$$\nu^{3/2} \times \nu^{5/2} \times \cdots \times \nu^{a_1-2} \times \left(\prod_{j=1}^{k_1} \delta\big([\nu^{a_j-1}, \nu^{b_j^{(1)}}]\big) \right) \times \left(\prod_{i=2}^{k} \prod_{j=1}^{k_i} \delta\big([\nu^{a_{\rho_i}-k_i+j}\rho_i, \nu^{b_j^{(i)}}\rho_i]\big) \right) \rtimes \tau_{\mathrm{cusp}}.$$

Proceeding with the same analysis as above, we obtain that $\Theta^{\epsilon}(\sigma, r-l) \neq 0$ for $l = 1, 2, \ldots, r - a_1 + \frac{3}{2}$ and that $\sigma(r-l)$ is a strongly positive discrete series that can be realized as a unique irreducible subrepresentation of

$$\nu^{l+1/2} \times \nu^{l+3/2} \times \cdots \times \nu^{a_1-2} \times \left(\prod_{j=1}^{k_1} \delta\big([\nu^{a_j-1}, \nu^{b_j^{(1)}}]\big) \right) \times \left(\prod_{i=2}^{k} \prod_{j=1}^{k_i} \delta\big([\nu^{a_{\rho_i}-k_i+j}\rho_i, \nu^{b_j^{(i)}}\rho_i]\big) \right) \rtimes \tau_{\mathrm{cusp}}.$$

Further, it is easy to check that the first occurrence index of σ equals $r - a_1 + \frac{3}{2}$.

(2) There is no representation of the form $\chi_{V,\psi}\nu^s$, $s \in \mathbb{R}$ appearing in $[\sigma]$: As in the previous case, we conclude that $\sigma(r)$ is a strongly positive discrete series. An easy computation shows that $\sigma(r)$ is a unique irreducible subrepresentation of the induced representation

$$\nu^{1/2} \times \nu^{3/2} \times \cdots \times \nu^{n'-m_{r'}} \times \left(\prod_{i=1}^{k} \prod_{j=1}^{k_i} \delta\big([\nu^{a_{\rho_i}-k_i+j}\rho_i, \nu^{b_j^{(i)}}\rho_i]\big) \right) \rtimes \tau_{\mathrm{cusp}}.$$

Now Theorem 5.3 of [Matić 2012] shows that $R_{P_1}(\sigma(r))(\nu^{1/2}) \neq 0$. Because $R_{\widetilde{P}_1}(\sigma)(\chi_{V,\psi}\nu^{1/2}) = 0$, part (2) of Proposition 5.2 implies $\Theta^{\epsilon}(\sigma, r-1) \neq 0$.

Note that $[\sigma(r-1)]$ and $[\sigma(r)]$ differ by $\nu^{1/2}$. Proposition 5.3 now shows that $\sigma(r-1)$ is a discrete series representation, and we again conclude that it must be strongly positive. Thus, $\sigma(r-1)$ is a unique irreducible subrepresentation of the induced representation

$$\nu^{3/2} \times \nu^{5/2} \times \cdots \times \nu^{n'-m_{r'}} \times \left(\prod_{i=1}^{k} \prod_{j=1}^{k_i} \delta\big([\nu^{a_{\rho_i}-k_i+j}\rho_i, \nu^{b_j^{(i)}}\rho_i]\big) \right) \rtimes \tau_{\mathrm{cusp}}.$$

If $n' - m_{r'} > \frac{1}{2}$, in the same way as above we deduce $\Theta^{\epsilon}(\sigma, r-2) \neq 0$. We continue in this fashion, obtaining that $\Theta^{\epsilon}(\sigma, r-j) \neq 0$ for $j = 1, 2, \ldots, n' - m_{r'} + \frac{1}{2}$, and that $\sigma(r-j)$ is a strongly positive discrete series that can be characterized as the unique irreducible subrepresentation of

$$\nu^{j+(1/2)} \times \nu^{j+(3/2)} \times \cdots \times \nu^{n'-m_{r'}} \times \Big(\prod_{i=1}^{k} \prod_{j=1}^{k_i} \delta\big([\nu^{a_{\rho_i}-k_i+j}\rho_i, \nu^{b_j^{(i)}}\rho_i]\big)\Big) \rtimes \tau_{\text{cusp}}.$$

From Proposition 5.2, we conclude that the first occurrence index of σ equals

$$r - n' + m_{r'} - \tfrac{1}{2} = r - (n' - \tfrac{1}{2}\dim V_{r'}^{\epsilon} + 2) + \tfrac{3}{2}.$$

Second, suppose that the representation $\chi_{V,\psi}\nu^{1/2}$ appears in $[\sigma]$. There is no loss of generality in assuming that ρ_1 is a trivial representation. We have to examine the following three possibilities:

- $n' = r' + \frac{1}{2}(\dim(V_0^{\epsilon}) - 1)$:

Observe that in this case both representations $\chi_{V,\psi}\nu^{s} \rtimes \sigma_{\text{cusp}}$ and $\nu^{s} \rtimes \tau_{\text{cusp}}$ reduce for $s = \frac{1}{2}$. Obviously, Theorem 3.1 implies $k_1 = 1$.

Observe that $[\sigma(r)]$ is obtained from $[\sigma]$ simply by replacing σ_{cusp} with τ_{cusp} and multiplying all $\widetilde{\text{GL}}$-members of $[\sigma]$ by $\chi_{V,\psi}^{-1}$; consequently, the discrete series $\sigma(r)$ may be realized as the unique irreducible subrepresentation of

$$\delta\big([\nu^{1/2}, \nu^{b_1^{(1)}}]\big) \times \Big(\prod_{i=2}^{k} \prod_{j=1}^{k_i} \delta\big([\nu^{a_{\rho_i}-k_i+j}\rho_i, \nu^{b_j^{(i)}}\rho_i]\big)\Big) \rtimes \tau_{\text{cusp}}.$$

We note that for each $i \in \{1, 2, \ldots, k\}$, there is at most one $x \in \mathbb{R}$, $0 \le |x| \le 1$ such that $\nu^{x}\rho_i$ appears in $[\sigma(r)]$, and thus τ has to be strongly positive.

Obviously, $R_{P_1}(\sigma(r))(\nu^{1/2}) \neq 0$ if and only if $b_1^{(1)} = \frac{1}{2}$.

If $b_1^{(1)} > \frac{1}{2}$, using Proposition 5.2, we directly conclude that $\Theta^{\epsilon}(\sigma, r-1) = 0$. Suppose that $b_1^{(1)} = \frac{1}{2}$. If $\Theta^{\epsilon}(\sigma, r-1) \neq 0$, we get that $\nu^{1/2}$ does not appear in $[\sigma(r-1)]$, contradicting Proposition 5.5 (we are in the first case there). Thus, r is the first occurrence index of σ.

- $n' < r' + \frac{1}{2}(\dim(V_0^{\epsilon}) - 1)$:

In this case, the representation $\chi_{V,\psi}\nu^{s} \rtimes \sigma_{\text{cusp}}$ reduces for $s = m_{r'} - n' - 1$ and the representation $\nu^{s} \rtimes \tau_{\text{cusp}}$ reduces for $s = m_{r'} - n'$.

According to Theorem 4.3, $[\sigma(r)]$ is obtained from $[\sigma]$ by multiplying all $\widetilde{\text{GL}}$-members of $[\sigma]$ by $\chi_{V,\psi}^{-1}$, subtracting the representations $\nu^{1/2}, \nu^{3/2}, \ldots, \nu^{m_{r'}-n'-1}$, and replacing σ_{cusp} with τ_{cusp}. In the same way as before, we conclude that $\sigma(r)$ is a strongly positive discrete series that is characterized as a unique irreducible subrepresentation of

$$\delta([\nu^{3/2}, \nu^{b_1^{(1)}}]) \times \delta([\nu^{5/2}, \nu^{b_2^{(1)}}]) \times \cdots \times \delta([\nu^{m_{r'}-n'}, \nu^{b_{k_1}^{(1)}}]) \times \Big(\prod_{i=2}^{k} \prod_{j=1}^{k_i} \delta([\nu^{a_{\rho_i}-k_i+j}\rho_i, \nu^{b_j^{(i)}}\rho_i]) \Big) \rtimes \tau_{\mathrm{cusp}}.$$

Since $\nu^{1/2}$ does not appear in $[\sigma(r)]$, it follows that r is the first occurrence index of σ.

- $n' > r' + \frac{1}{2}(\dim(V_0^\epsilon) - 1)$:

Now the representation $\chi_{V,\psi}\nu^s \rtimes \sigma_{\mathrm{cusp}}$ reduces for $s = n' - m_{r'} + 1$, and the representation $\nu^s \rtimes \tau_{\mathrm{cusp}}$ reduces for $s = n' - m_{r'}$.

Theorem 4.3 now shows that $[\sigma(r)]$ is obtained from $[\sigma]$ by multiplying all $\widetilde{\mathrm{GL}}$-members of $[\sigma]$ by $\chi_{V,\psi}^{-1}$, adding the representations $\nu^{-1/2}, \nu^{-3/2}, \ldots, \nu^{m_{r'}-n'}$, and replacing σ_{cusp} with τ_{cusp}.

From Theorem 4.2, we know that the representation $\sigma(r)$ is in the discrete series. But $\nu^{1/2}$ appears in $[\sigma(r)]$ with multiplicity two, and consequently $\sigma(r)$ can't be a strongly positive representation (by Lemma 3.3).

In what follows, we use Theorem 3.5 to describe discrete series $\sigma(r)$ as precisely as we can. So, we write $\sigma(r)$ as a subrepresentation of the induced representation of the form

$$\delta([\nu^{a'_1}\rho'_1, \nu^{b'_1}\rho'_1]) \times \delta([\nu^{a'_2}\rho'_2, \nu^{b'_2}\rho'_2]) \times \cdots \times \delta([\nu^{a'_l}\rho'_l, \nu^{b'_l}\rho'_l]) \rtimes \tau_{\mathrm{sp}},$$

where $\rho'_i \in \{\rho_1, \rho_2, \ldots, \rho_k\}$, $a'_i \le 0$, and $a'_i + b'_i > 0$ for $i = 1, 2, \ldots, l$. Further, τ_{sp} is an irreducible strongly positive representation such that $[\tau_{\mathrm{sp}}]$ is contained in $[\sigma(r)]$. Hence, at least one of the representations $\nu^{1/2}$ and $\nu^{-1/2}$ has to appear in some segment $[\nu^{a'_i}\rho'_i, \nu^{b'_i}\rho'_i]$, $i \in \{1, 2, \ldots, l\}$. Since $a'_i \le 0$ and $b'_i > 0$, both these representations appear in this segment.

Our next claim is that $l = 1$. Suppose, on the contrary, that $l > 1$.

Then there is some $j \in \{1, 2, \ldots, l\}$, $j \ne i$ such that $\nu^{1/2} \notin [\nu^{a'_j}\rho'_j, \nu^{b'_j}\rho'_j]$. But the union of the segments $[\nu^{a'_i}, \nu^{b'_i}]$ and $[\nu^{a'_j}\rho'_j, \nu^{b'_j}\rho'_j]$ is contained in $[\sigma(r)]$, so there is at most one x, $0 \le |x| \le 1$ such that $\nu^x\rho'_j$ appears in $[\nu^{a'_j}\rho'_j, \nu^{b'_j}\rho'_j]$. This contradicts the fact that the ends of segment $[\nu^{a'_j}\rho'_j, \nu^{b'_j}\rho'_j]$ satisfy $a'_j \le 0$ and $b'_j > 0$. Thus, $l = 1$ and $\rho'_1 \cong 1_{F^\times}$.

In this way we obtain the following embedding:

$$\sigma(r) \hookrightarrow \delta([\nu^{a'_1}, \nu^{b'_1}]) \rtimes \tau_{\mathrm{sp}}.$$

Since $a'_1 \le 0$, using Proposition 5.6 we obtain a contradiction with the strong positivity of σ. Therefore, this case is impossible and Theorem 6.1 is proved.

The results obtained closely parallel those contained in Theorem 4.2 of [Muić 2004] for the dual pair $(\mathrm{Sp}(n), O(V))$.

7. Lifts of strongly positive discrete series of the orthogonal groups

The purpose of this section is to determine the first occurrence indices of strongly positive discrete series of the odd orthogonal groups that appear in the correspondence given by Theorem 4.2, and to provide a description of the lifts of such representations in the metaplectic tower.

Thus, we let $\tau \in \mathrm{Irr}(O(V_r))$ denote a strongly positive discrete series such that $\Theta(\tau, m_r - \frac{1}{2}) \neq 0$, and realize it as a unique irreducible subrepresentation of the induced representation of the form

$$\left(\prod_{i=1}^{k}\prod_{j=1}^{k_i}\delta\big([\nu^{a_{\rho_i}-k_i+j}\rho_i, \nu^{b_j^{(i)}}\rho_i]\big)\right) \rtimes \tau_{\mathrm{cusp}},$$

with k minimal and k_i minimal for $i = 1, 2, \ldots, k$, where $\tau_{\mathrm{cusp}} \in \mathrm{Irr}(O(V_{r'}))$ is a cuspidal representation and ρ_i an irreducible cuspidal representation of $\mathrm{GL}(n_{\rho_i}, F)$ (this defines n_{ρ_i}) for $i = 1, 2, \ldots, k$.

Note that Proposition 5.1 yields that the representation $\tau(l)$ is not a discrete series representation for $l > m_r - \frac{1}{2}$.

In the following theorem, we describe the first occurrence indices of certain strongly positive discrete series of the odd orthogonal groups.

Theorem 7.1. *Let $\tau \in \mathrm{Irr}(O(V_r))$ be a strongly positive discrete series with a nonzero lift on the $(m_r - \frac{1}{2})$-th level of the metaplectic tower. Suppose that $\tau_{\mathrm{cusp}} \in \mathrm{Irr}(O(V_{r'}))$ is a partial cuspidal support of τ and that $\sigma_{\mathrm{cusp}} \in \mathrm{Irr}(\widetilde{\mathrm{Sp}}(n'))$ is the first nonzero lift of τ_{cusp}. Let $n = m_r - \frac{1}{2}$. Further, define $M = \{|x| : \nu^x$ appears in $[\tau]\}$ and denote by $a_{\min}$ the minimal element of M. If $M = \varnothing$, let $a_{\min} = m_{r'} - n' + 1$.*

If $a_{\min} = \frac{1}{2}$ or $r' = n' - \frac{1}{2}(\dim(V_0) - 1)$, then the first occurrence index of τ is n. Otherwise, the first occurrence index of τ is $n - a_{\min} + \frac{3}{2}$.

The remaining part of this section is devoted to the proof this theorem.

Again, we have two main cases to discuss.

First, assume that $\nu^{1/2}$ does not appear in $[\tau]$. This implies

$$r' \geq n' - \tfrac{1}{2}(\dim(V_0) - 1).$$

This leaves us two possibilities:

- $r' = n' - \frac{1}{2}(\dim(V_0) - 1)$:

In this case, both representations $\chi_{V,\psi}\nu^s \rtimes \sigma_{\mathrm{cusp}}$ and $\nu^s \rtimes \tau_{\mathrm{cusp}}$ reduce for $s = \frac{1}{2}$. From the classification of strongly positive discrete series, elaborated in Section 2, we deduce that there are no representations of the form ν^s appearing in $[\tau]$.

Applying Theorem 4.2, we obtain that $\tau(n)$ is a discrete series representation, and in the same way as before, we may conclude that it is strongly positive. This yields the embedding

$$\tau(n) \hookrightarrow \left(\prod_{i=1}^{k} \prod_{j=1}^{k_i} \delta\big([\chi_{V,\psi}\nu^{a_{\rho_i}-k_i+j}\rho_i, \chi_{V,\psi}\nu^{b_j^{(i)}}\rho_i]\big) \right) \rtimes \sigma_{\mathrm{cusp}}.$$

Proposition 5.1 implies $\Theta(\tau, n-1) = 0$, so n is the first occurrence index of τ.

- $r' > n' - \frac{1}{2}(\dim(V_0) - 1)$:

In this case, the representation $\nu^s \rtimes \tau_{\mathrm{cusp}}$ reduces for $s = m_{r'} - n'$ and the representation $\chi_{V,\psi}\nu^s \rtimes \sigma_{\mathrm{cusp}}$ reduces for $s = m_{r'} - n' - 1$.

Theorem 4.3 shows that $[\tau(n)]$ is obtained from $[\tau]$ by multiplying all elements of $\mathcal{R}$ appearing in $[\tau]$ by $\chi_{V,\psi}$, adding the representations $\chi_{V,\psi}\nu^{1/2}$, $\chi_{V,\psi}\nu^{3/2}, \ldots,$ $\chi_{V,\psi}\nu^{m_{r'}-n'-1}$, and replacing τ_{cusp} with σ_{cusp}.

There are two main cases to consider:

(1) There is no representation of the form ν^s appearing in $[\tau]$, for $s \in \mathbb{R}$: As before, we conclude that $\tau(n)$ is a strongly positive discrete series that is a unique irreducible subrepresentation of

$$\chi_{V,\psi}\nu^{1/2} \times \chi_{V,\psi}\nu^{3/2} \times \cdots \times \chi_{V,\psi}\nu^{m_{r'}-n'-1} \times \left(\prod_{i=1}^{k} \prod_{j=1}^{k_i} \delta\big([\chi_{V,\psi}\nu^{a_{\rho_i}-k_i+j}\rho_i, \chi_{V,\psi}\nu^{b_j^{(i)}}\rho_i]\big) \right) \rtimes \sigma_{\mathrm{cusp}}.$$

Theorem 5.3 of [Matić 2012] implies $R_{P_1}(\tau(n))(\chi_{V,\psi}\nu^{1/2}) \neq 0$. Since

$$R_{P_1}(\tau)(\nu^{1/2}) = 0,$$

part (2) of Proposition 5.1 shows $\Theta(\sigma, n-1) \neq 0$.

From Corollary 5.4, we obtain that $\tau(n-l)$ is a discrete series representation for each $l > 0$ such that $\Theta(\tau, n-l) \neq 0$. In the same way as above, we see that it must be strongly positive.

Since $[\tau(n-l)]$ is obtained from $[\tau(n)]$ by subtraction of the representations $\chi_{V,\psi}\nu^{1/2}, \chi_{V,\psi}\nu^{3/2}, \ldots, \chi_{V,\psi}\nu^{(2l-1)/2}$, for $l \in \{1, 2, \ldots, m_{r'} - n' - \frac{1}{2}\}$, it is not hard to see, using Proposition 5.1, that $\Theta(\tau, n-l) \neq 0$ for $l \in \{1, 2, \ldots, m_{r'} - n' - \frac{1}{2}\}$. Furthermore, $\tau(n-l)$ is a unique irreducible subrepresentation of the induced representation

$$\chi_{V,\psi}\nu^{(2l+1)/2} \times \chi_{V,\psi}\nu^{(2l+3)/2} \times \cdots \times \chi_{V,\psi}\nu^{m_{r'}-n'-1} \times \left(\prod_{i=1}^{k} \prod_{j=1}^{k_i} \delta\big([\chi_{V,\psi}\nu^{a_{\rho_i}-k_i+j}\rho_i, \chi_{V,\psi}\nu^{b_j^{(i)}}\rho_i]\big) \right) \rtimes \sigma_{\mathrm{cusp}},$$

for $l \in \{1, 2, \ldots, m_{r'} - n' - \frac{1}{2}\}$.

There is no representation of the form $\chi_{V,\psi}\nu^s$ appearing in $[\tau(n-m_{r'}+n'+\frac{1}{2})]$, so Proposition 5.1 shows that the first occurrence index of τ equals $n-m_{r'}+n'+\frac{1}{2}$.

(2) There is some representation of the form ν^s appearing in $[\tau]$: We may suppose that ρ_1 is a trivial representation. Obviously, $a_{\rho_1}-k_1+1$ is strictly greater than $\frac{1}{2}$ and a_{ρ_1} equals $m_{r'}-n'$.

For brevity, let a_j stand for $a_{\rho_1}-k_1+j$, for $j=1,2,\dots,k_1$. Since $\chi_{V,\psi}\nu^{1/2}$ appears in $[\tau(n)]$ with multiplicity one, it follows that $\tau(n_1)$ is a strongly positive representation for each $n_1\le n$ such that $\Theta(\tau,n_1)\neq 0$.

Also, $\tau(n)$ is the unique irreducible subrepresentation of

$$\chi_{V,\psi}\nu^{1/2}\times\chi_{V,\psi}\nu^{3/2}\times\cdots\times\chi_{V,\psi}\nu^{a_1-2}\times\Bigl(\prod_{j=1}^{k_1}\delta\bigl([\chi_{V,\psi}\nu^{a_j-1},\chi_{V,\psi}\nu^{b_j^{(1)}}]\bigr)\Bigr)\\ \times\Bigl(\prod_{i=2}^{k}\prod_{j=1}^{k_i}\delta\bigl([\chi_{V,\psi}\nu^{a_{\rho_i}-k_i+j}\rho_i,\chi_{V,\psi}\nu^{b_j^{(i)}}\rho_i]\bigr)\Bigr)\rtimes\sigma_{\text{cusp}}.$$

Arguing in the same way as in the analogous situation in the metaplectic case, we deduce that $\Theta(\tau,n-l)\neq 0$ for $l\in\{1,2,\dots,a_1-\frac{3}{2}\}$ and that $n-a_1+\frac{3}{2}$ is the first occurrence index of τ. Further, $\tau(n-l)$ is a unique irreducible representation of the induced representation

$$\chi_{V,\psi}\nu^{l+1/2}\times\chi_{V,\psi}\nu^{l+3/2}\times\cdots\times\chi_{V,\psi}\nu^{a_1-2}\times\Bigl(\prod_{j=1}^{k_1}\delta\bigl([\chi_{V,\psi}\nu^{a_j-1},\chi_{V,\psi}\nu^{b_j^{(1)}}]\bigr)\Bigr)\\ \times\Bigl(\prod_{i=2}^{k}\prod_{j=1}^{k_i}\delta\bigl([\chi_{V,\psi}\nu^{a_{\rho_i}-k_i+j}\rho_i,\chi_{V,\psi}\nu^{b_j^{(i)}}\rho_i]\bigr)\Bigr)\rtimes\sigma_{\text{cusp}},$$

for $l\in\{1,2,\dots,a_1-\frac{3}{2}\}$.

It remains to consider the case when the representation $\nu^{1/2}$ appears in $[\tau]$. Without loss of generality, we may suppose that ρ_1 is a trivial character. Similarly to the previous section, we have to examine three possibilities.

- $r'=n'-\frac{1}{2}(\dim(V_0)-1)$:

The specificity of this case is that both induced representations $\nu^s\rtimes\tau_{\text{cusp}}$ and $\chi_{V,\psi}\nu^s\rtimes\sigma_{\text{cusp}}$ reduce for $s=\frac{1}{2}$. On account of Theorem 3.1, we have $k_1=1$ and $a_{\rho_1}=\frac{1}{2}$.

Furthermore, $[\tau(n)]$ is obtained from $[\tau]$ by replacing τ_{cusp} with σ_{cusp} and multiplying all other members of $[\tau]$ by $\chi_{V,\psi}$.

From the equality of cuspidal reducibilities for τ_{cusp} and σ_{cusp}, it may be concluded that $\tau(n)$ is the strongly positive discrete series that is a unique irreducible subrepresentation of

$$\delta([\chi_{V,\psi}\nu^{1/2}, \chi_{V,\psi}\nu^{b_1^{(1)}}]) \times \left(\prod_{i=2}^{k}\prod_{j=1}^{k_i} \delta([\chi_{V,\psi}\nu^{a_{\rho_i}-k_i+j}\rho_i, \chi_{V,\psi}\nu^{b_j^{(i)}}\rho_i])\right) \rtimes \sigma_{\mathrm{cusp}}.$$

Suppose that the lift $\Theta(\tau, n-1)$ is nonzero. Then Proposition 5.1, enhanced by Theorem 5.3 of [Matić 2012], implies $b_1^{(1)} = \frac{1}{2}$. From Theorem 4.3, it follows that there is no representation $\chi_{V,\psi}\nu^{1/2}$ appearing in $[\tau(n-1)]$, contrary to Proposition 5.6.

It follows that n is the first occurrence index of τ.

- $r' < n' - \frac{1}{2}(\dim(V_0) - 1)$:

The induced representation $\nu^s \rtimes \tau_{\mathrm{cusp}}$ reduces for $s = n' - m_{r'}$ and the induced representation $\chi_{V,\psi}\nu^s \rtimes \sigma_{\mathrm{cusp}}$ reduces for $s = n' - m_{r'} + 1$. According to Theorem 4.3, $[\tau(n)]$ is obtained from $[\tau]$ by replacing τ_{cusp} with σ_{cusp}, multiplying GL-members of $[\tau]$ by $\chi_{V,\psi}$, and then subtracting the representations $\chi_{V,\psi}\nu^{1/2}, \chi_{V,\psi}\nu^{3/2}, \dots, \chi_{V,\psi}\nu^{n'-m_{r'}}$.

The strong positivity of the representation τ and the above discussion show that for each $i \in \{1, 2, \dots, k\}$, there is at most one x, $|x| \le 1$ such that $\chi_{V,\psi}\nu^x$ appears in $[\tau(n)]$. Since $\tau(n)$ is in the discrete series, from Theorem 3.5 we see that it is strongly positive.

An easy computation shows that $\tau(n)$ is a unique irreducible subrepresentation of the induced representation

$$\begin{aligned}\delta([\chi_{V,\psi}\nu^{3/2}, \chi_{V,\psi}\nu^{b_1^{(1)}}]) \times \delta([\chi_{V,\psi}\nu^{5/2}, \chi_{V,\psi}\nu^{b_2^{(1)}}]) \times \cdots \\ \times \delta([\chi_{V,\psi}\nu^{n'-m_{r'}+1}, \chi_{V,\psi}\nu^{b_{k_1}^{(1)}}]) \\ \times \left(\prod_{i=2}^{k}\prod_{j=1}^{k_i} \delta([\chi_{V,\psi}\nu^{a_{\rho_i}-k_i+j}\rho_i, \chi_{V,\psi}\nu^{b_j^{(i)}}\chi_{V,\psi}\rho_i])\right) \rtimes \sigma_{\mathrm{cusp}}.\end{aligned}$$

That n is the first occurrence index of τ follows directly from Proposition 5.1.

- $r' > n' - \frac{1}{2}(\dim(V_0) - 1)$:

The induced representation $\nu^s \rtimes \tau_{\mathrm{cusp}}$ reduces for $s = m_{r'} - n'$, and the representation $\chi_{V,\psi}\nu^s \rtimes \sigma_{\mathrm{cusp}}$ reduces for $s = m_{r'} - n' - 1$. The representation $\chi_{V,\psi}\nu^{1/2}$ appears in $[\tau(n)]$ with multiplicity two, since $[\tau(n)]$ is obtained from $[\tau]$ by replacing τ_{cusp} with σ_{cusp}, multiplying other members of $[\tau]$ by $\chi_{V,\psi}$, and adding $\chi_{V,\psi}\nu^{1/2}, \chi_{V,\psi}\nu^{3/2}, \dots, \chi_{V,\psi}\nu^{m_{r'}-n'-1}$.

According to Lemma 3.3, $\tau(n)$ is not a strongly positive discrete series, but the results in [Gan and Savin 2012] show that it is a discrete series representation.

Applying Theorem 3.5 and analysis similar to that in the last case considered in the previous section, we write $\tau(n)$ as an irreducible subrepresentation of the induced representation of the form

$$\delta([\chi_{V,\psi}\nu^a, \chi_{V,\psi}\nu^b]) \rtimes \sigma_{\mathrm{sp}},$$

where $a \leq 0$, $a+b>0$, and $\sigma_{\mathrm{sp}} \in S_2$ is a strongly positive discrete series.

Using Proposition 5.5, we obtain an embedding that contradicts the strong positivity of τ. Consequently, this case is not possible.

This completes the proof of Theorem 7.1.

Acknowledgements

The author would like to thank Goran Muić for suggesting this problem. The author would also like to thank the referee for helping to improve the style of the presentation.

References

[Arthur 2011] J. Arthur, "The endoscopic classification of representations: orthogonal and symplectic groups", preprint, 2011, available at http://www.claymath.org/cw/arthur/pdf/Book.pdf.

[Ban and Jantzen 2009] D. Ban and C. Jantzen, "Langlands quotient theorem for covering groups", preprint, 2009.

[Bernstein 1987] J. Bernstein, "Second adjointness for representations of p-adic reductive groups", preprint, 1987, available at http://tinyurl.com/9pmgpmz.

[Gan and Savin 2012] W. T. Gan and G. Savin, "Representations of metaplectic groups, I: epsilon dichotomy and local Langlands correspondence", to appear in *Compos. Math.*, 2012, available at http://tinyurl.com/8ozfjyk.

[Hanzer and Muić 2010] M. Hanzer and G. Muić, "Parabolic induction and Jacquet functors for metaplectic groups", *J. Algebra* **323**:1 (2010), 241–260. MR 2564837 Zbl 1185.22013

[Hanzer and Muić 2011] M. Hanzer and G. Muić, "Rank one reducibility for metaplectic groups via theta correspondence", *Canad. J. Math.* **63**:3 (2011), 591–615. Zbl 1217.22014

[Kudla 1986] S. S. Kudla, "On the local theta-correspondence", *Invent. Math.* **83**:2 (1986), 229–255. MR 87e:22037 Zbl 0583.22010

[Kudla 1996] S. S. Kudla, "Notes on the local theta correspondence", lecture notes, 1996, available at http://www.math.toronto.edu/~skudla/castle.pdf.

[Kudla and Rallis 2005] S. S. Kudla and S. Rallis, "On first occurrence in the local theta correspondence", pp. 273–308 in *Automorphic representations, L-functions and applications: progress and prospects*, Ohio State Univ. Math. Res. Inst. Publ. **11**, de Gruyter, Berlin, 2005. MR 2007d:22028 Zbl 1109.22012

[Matić 2011] I. Matić, "Strongly positive representations of metaplectic groups", *J. Algebra* **334**:1 (2011), 255–274. Zbl 05990155

[Matić 2012] I. Matić, "Jacquet modules of strongly positive representations of the metaplectic group $\widetilde{Sp(n)}$", to appear in *Trans. Amer. Math. Soc.*, 2012.

[Mœglin 2002] C. Mœglin, "Sur la classification des séries discrètes des groupes classiques p-adiques: paramètres de Langlands et exhaustivité", *J. Eur. Math. Soc.* (*JEMS*) **4**:2 (2002), 143–200. MR 2003g:22021 Zbl 1002.22009

[Mœglin and Tadić 2002] C. Mœglin and M. Tadić, "Construction of discrete series for classical p-adic groups", *J. Amer. Math. Soc.* **15**:3 (2002), 715–786. MR 2003g:22020 Zbl 0992.22015

[Mœglin et al. 1987] C. Mœglin, M.-F. Vignéras, and J.-L. Waldspurger, *Correspondances de Howe sur un corps p-adique*, Lecture Notes in Mathematics **1291**, Springer, Berlin, 1987. MR 91f:11040 Zbl 0642.22002

[Muić 2004] G. Muić, "Howe correspondence for discrete series representations; the case of $(Sp(n), O(V))$", *J. Reine Angew. Math.* **567** (2004), 99–150. MR 2005a:22013 Zbl 1037.22037

[Muić 2006] G. Muić, "On the non-unitary unramified dual for classical p-adic groups", *Trans. Amer. Math. Soc.* **358**:10 (2006), 4653–4687. MR 2007j:22029 Zbl 1102.22014

[Muić 2008] G. Muić, "On the structure of theta lifts of discrete series for dual pairs $(Sp(n), O(V))$", *Israel J. Math.* **164** (2008), 87–124. MR 2009c:22017 Zbl 1153.22020

[Ranga Rao 1993] R. Ranga Rao, "On some explicit formulas in the theory of Weil representation", *Pacific J. Math.* **157**:2 (1993), 335–371. MR 94a:22037 Zbl 0794.58017

[Silberger 1980] A. J. Silberger, "Special representations of reductive p-adic groups are not integrable", *Ann. of Math.* (2) **111**:3 (1980), 571–587. MR 82k:22015 Zbl 0437.22015

[Tadić 1995] M. Tadić, "Structure arising from induction and Jacquet modules of representations of classical p-adic groups", *J. Algebra* **177**:1 (1995), 1–33. MR 97b:22023 Zbl 0874.22014

[Waldspurger 1984] J.-L. Waldspurger, "Correspondances de Shimura", pp. 525–531 in *Proceedings of the International Congress of Mathematicians* (Warsaw, 1983), vol. 1, PWN, Warsaw, 1984. MR 86m:11036 Zbl 0567.10020

[Waldspurger 1991] J.-L. Waldspurger, "Correspondances de Shimura et quaternions", *Forum Math.* **3**:3 (1991), 219–307. MR 92g:11054 Zbl 0724.11026

Received March 8, 2011. Revised July 14, 2012.

IVAN MATIĆ
DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OSIJEK
TRG LJUDEVITA GAJA 6
31000 OSIJEK
CROATIA
imatic@mathos.hr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 259, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.259.473

TUNNEL ONE, FIBERED LINKS

MATT RATHBUN

For a fibered link of tunnel number one in S^3, with fiber F and unknotting tunnel τ, we show that τ can be isotoped to lie in F.

1. Introduction and motivation

The study of fibered knots and links is as important today as ever. Giroux's correspondence [2002] between open book decompositions and contact structures mingles classical fibered links with more modern contact geometry. Sutured manifold theory continues to reveal information about fibrations (see, for instance, [Ni 2009; Scharlemann and Thompson 2009]). And fibered links are related to the newest advances in Floer homology, as knot Floer homology detects fibered links [Ni 2007] and sutured Floer homology intersects both contact geometry and sutured manifold theory.

Tunnel number one links are among the most studied links. Much of the work on tunnel number one links revolves around trying to isotope the tunnel to sit nicely with respect to some additional structure in the 3-manifold, including a hyperbolic metric [Adams 1995; Adams and Reid 1996; Akiyoshi et al. 1997; Cooper et al. 2010], polyhedral decompositions [Sakuma and Weeks 1995; Heath and Song 2005], bridge decompositions [Goda et al. 2000; Lackenby 2005], Seifert surfaces [Scharlemann and Thompson 2003], and fibrations [Sakuma 1996]. These studies, and others, have led to the classification of tunnels for many classes of knots and links, including torus knots [Boileau et al. 1988], satellite knots [Morimoto and Sakuma 1991], nonsimple links [Eudave Muñoz and Uchida 1996], 2-bridge knots [Morimoto and Sakuma 1991; Kobayashi 1999], and 2-bridge links [Morimoto 1994; Jones 1995].

Further, the Berge conjecture states that if a knot admits a lens space Dehn surgery, then it is in one of the families of knots classified by John Berge. Many are working on this long-standing conjecture, with recent progress contributed by Ozsváth and Szabó [2005], Hedden [2007], Baker, Grigsby, and Hedden [2008],

The author would like to thank Abigail Thompson. The author was supported in part by NSF VIGRE Grants DMS-0135345 and DMS-0636297, and the RTG Grant DMS-0739208.

MSC2010: primary 57M25; secondary 57M27.

Keywords: tunnel, fibered, knot, link, monodromy.

Saito [2007], and Williams [2007], among others. Yi Ni [2007] recently proved that if a knot admits a lens space surgery, then it is a fibered knot. Additionally, all Berge knots are both fibered and tunnel number one, so further understanding of tunnel one, fibered knots could have profound impacts on the conjecture.

Jesse Johnson [2008] investigated genus-2 Heegaard splittings of closed surface bundles over the circle. This paper looks at the relationship between unknotting tunnels and fibrations for link complements.

Theorem 1.1. *Let K be an oriented, fibered, tunnel number one link in S^3, with fiber F, and unknotting tunnel τ. Then τ can be isotoped to lie in F.*

2. Background and definitions

3-*manifolds.*

Notation 1. Let A be subset of a 3-manifold M. We fix some notation. Let $n(A)$ denote a small open neighborhood of A in M. If F is a properly embedded surface in M, let $M|F = M \setminus n(F)$. If S is the boundary of M, we will refer to $S|\partial F = S \setminus n(\partial F)$. For convenience, we will also sometimes refer to this as $S|F$.

Definition 2.1. Let F be a surface properly embedded in a 3-manifold M. Then F is said to be *compressible* if there exists a disk D embedded in M with $\partial D = D \cap F$ an essential curve in F, and D is called a *compressing disk* for F. If F is not compressible, and is not a 2-sphere, then it is called *incompressible*. The surface F is said to be *boundary compressible* if there exists a disk D embedded in M with $D \cap F = \alpha \subset \partial D$, $D \cap \partial M = \beta \subset \partial D$, where α is an essential arc in F, $\alpha \cap \beta = \partial\alpha = \partial\beta$, and $\alpha \cup \beta = \partial D$. In this case, D is called a *boundary compression disk*. If F is not boundary compressible, it is called *boundary incompressible*.

Definition 2.2. A *compression body* V is the result of taking the product of a surface with [0, 1], attaching 2-handles along $S \times \{0\}$, and then attaching 3-handles along any resulting 2-sphere components. The surface $S \times \{1\}$ is called $\partial_+ V$, and $\partial V \setminus \partial_+ V$ is called $\partial_- V$. A *handlebody* is a compression body where $\partial_- V = \varnothing$. A *Heegaard splitting* is a triple (S, V, W), where S is a surface, V and W are compression bodies, $\partial_+ V = \partial_+ W = S$, and $M = V \cup_S W$.

Definition 2.3. Let K be a knot in a 3-manifold M, and let λ be an essential closed curve in $\partial n(K)$. Let M' be the manifold obtained from M by removing $n(K)$, and attaching a solid torus $S^1 \times D^2$ to $M \setminus n(K)$ via a homeomorphism of the boundaries such that $\{\text{pt.}\} \times \partial D^2$ is identified with the curve λ. Then M' is said to be the result of *λ-sloped Dehn surgery* on M.

Tunnels.

Definition 2.4. A link L in S^3 is called a *tunnel number one* link if there exists an arc τ properly embedded in $S^3 \setminus n(L)$ such that $S^3 \setminus n(L \cup \tau)$ is a handlebody. Then τ is called a *tunnel* for L.

Observe that the complement of a tunnel number one link has a genus-2 Heegaard splitting. Also, note that a tunnel one link has at most two components, and if it has two components, then any tunnel must have one endpoint on each component.

More generally, a knot is *tunnel number n* if n is the smallest number such that there exists a collection of arcs $\{\tau_1, \dots, \tau_n\}$ such that $S^3 \setminus n(L \cup \tau_1 \cup \cdots \cup \tau_n)$ is a handlebody.

Fibered links.

Definition 2.5. Let $L \subset S^3$ be a link. A *Seifert surface* for L is a compact, orientable surface F embedded in S^3 with no closed components such that $\partial F = L$.

Definition 2.6. A map $f : E \to B$ is a *fibration* with *fiber* F if for every point $p \in B$, there is a neighborhood U of p and a homeomorphism $h : f^{-1}(U) \to U \times F$ such that $f = \pi_1 \circ h$, where $\pi_1 : U \times F \to U$ is projection to the first factor. The space E is called the *total* space, and B is called the *base* space. Each set $f^{-1}(b)$ is called a *fiber*, and is homeomorphic to F.

Definition 2.7. A link $L \subset S^3$ is said to be *fibered* if there is a fibration of $S^3 \setminus n(L)$ over S^1, and the fibration is well-behaved near L. That is, each component L_i of L has a neighborhood $S^1 \times D^2$, with $L_i \cong S^1 \times \{0\}$ such that $f\big|_{S^1 \times (D^2 \setminus \{0\})}$ is given by $(x, y) \to y/|y|$.

Each fiber of a fibered link is a Seifert surface for the link. The complement of a fibered link is foliated by copies of this Seifert surface. Cutting along one of these Seifert surfaces produces a surface cross the interval.

Definition 2.8. Let K be a fibered link in S^3. Then $S^3 \setminus n(K)$ can be obtained from $F \times I$, with F a fiber, by identification $(x, 0) \sim (h(x), 1)$, for $x \in F$, where $h : F \to F$ is an orientation-preserving homeomorphism which is the identity on ∂F. We call h a *monodromy* map.

Theorems. Our starting point is the following theorem.

Theorem 2.9 [Scharlemann and Thompson 2003]. *Suppose K is a knot in S^3, and τ an unknotting tunnel for K. Then τ may be slid and isotoped until it is disjoint from some minimal-genus Seifert surface for K.*

The proof consists of arranging K, τ, and a compressing disk for $S^3 \setminus n(K \cup \tau)$ in some minimal fashion, and showing that if $K \cap \tau \neq \varnothing$, this would lead to a contradiction with those minimality assumptions. The result still holds for two-component fibered links.

Theorem 2.10. *Suppose K is an oriented, fibered link, and τ is an unknotting tunnel for K. Then τ may be slid and isotoped until it is disjoint from a fiber of K.*

Our proof will largely mimic [Scharlemann and Thompson 2003].

Proof. By Theorem 2.9, if K has just one component, then an unknotting tunnel can be isotoped and slid to be disjoint from a minimal-genus Seifert surface. But in a fibered knot complement, a fiber is the unique minimal-genus Seifert surface, so the result follows. Henceforth, let us assume that K is a two-component link, and let the two components of K be K_1 and K_2. Observe that τ has one endpoint on each of the components of K. Choose a fiber F, and slide and isotope τ, so as to minimize the number of intersections between τ and F. Our goal will be to prove that $\tau \cap F = \varnothing$.

Suppose, to the contrary, that after the slides and isotopies above, $\tau \cap F$ is nonempty. Let E be an essential disk in the handlebody $S^3 \setminus n(K \cup \tau)$, chosen to minimize the number $|E \cap F|$ of components in $E \cap F$. If $|E \cap F| = 0$, then the incompressible F would lie in a solid torus, namely (a component of) $S^3 \setminus n(K \cup \tau \cup E)$, and so be an annulus. The only fibered link with fiber an annulus is the Hopf link, in which case the result holds. So we may assume that $|E \cap F| > 0$. Furthermore, since F is incompressible, we may assume that $E \cap F$ consists entirely of arcs.

Let e be an outermost arc of $E \cap F$ in E, cutting off a subdisk E_0 from E. If e were inessential in $F \setminus \tau$, then we could surger E along the trivial subdisk cut off by e. The result would be two disks, at least one of which is also essential in $S^3 \setminus n(K \cup \tau)$, but with one fewer intersection with F, contradicting our assumption of minimality. Thus, the arc e is essential in $F \setminus \tau$. Let $f = \partial(E_0) \setminus e$, an arc in $\partial n(K \cup \tau)$ with each end either on a longitude $\partial F \subset \partial n(K)$ or a meridian disk of τ corresponding to a point of $\tau \cap F$.

Now, either no meridian of τ is incident to an end of f, a meridian of τ is incident to exactly one end of f, or there is a meridian which is incident to both ends of f.

(1) If no meridian of τ is incident to an end of f, then both ends of f lie on $\partial F \subset \partial n(K)$. If the interior of f runs over τ, we have finished, for f is disjoint from F. Otherwise, the interior of f lies entirely in $\partial n(K)$, and e is either essential in F, or it is inessential.

   (a) If e is essential in F, then E_0 would be a boundary compression disk for F, contradicting the minimality of the genus of F.

(b) If e is inessential in F, then the disk D_0 that it cuts off from F necessarily contains points of τ (since e is essential in $F \setminus \tau$). But then we could replace D_0 by E_0, and the loop formed by f and $\partial D_0 \setminus e$ is either a trivial loop on one of the torus components of $\partial n(K)$, or it is an essential loop.

(i) If the loop formed by f and $\partial D_0 \setminus e$ is a trivial loop on the torus, say, $\partial n(K_1)$, then the new surface would, again, be a Seifert surface for K, consistent with the orientation of K (and so be a fiber), but with fewer points of intersection $F \cap \tau$.

(ii) If the loop formed by f and $\partial D_0 \setminus e$ is essential in $\partial n(K_1)$, then the original disk E_0 could be slid across D_0 to show that K_1 is unknotted. But the interior of the disk D_0 is disjoint from K_2, so K must be a split link, and split links do not fiber.

(2) If a meridian of τ is incident to exactly one end of f, then we can use E_0 to describe a simple isotopy of τ by sliding τ along E_0 which reduces the number of intersections between τ and F.

(3) If both ends of f lie on the same meridian of τ, then e forms a loop in F, and the ends of f adjacent to e both run along the same subarc τ_0 of τ. Since f is disjoint from F, τ_0 terminates on, say, $\partial n(K_1)$.

Then since the interior of f is disjoint from F, f must intersect $\partial n(K_1)$ either in an inessential arc in the torus or in a longitudinal arc. That is, if $\tau_0 \cap \partial n(K_1)$ were collapsed to a point p, then f would either represent a trivial loop in $\pi_1(\partial n(K_1), p)$, or a nontrivial element. The former case is impossible, because the trivial disk cut off by f cannot contain the other end of τ (since the other end of τ is on $\partial n(K_2)$). Thus, the disk could be isotoped away, reducing $|E \cap F|$. It follows that f intersects the torus $\partial n(K_1)$ in a longitudinal arc. Then, $n(\tau_0 \cup E_0)$ is a thickened annulus A, defining a parallelism in S^3 between K_1 and the loop e on F. Now, the boundary component of A on $\partial n(K_1)$ can be slid across $\partial n(K_1)$, away from e, onto F, parallel to ∂F in F. Since K is a fibered link, the image of A, call it A', is a product annulus in $S^3 \setminus n(K \cup F) \cong F \times I$. But then this demonstrates that e itself is parallel to ∂F in F. Then, substituting A for the annulus between e and ∂F in F would create a Seifert surface of the same genus, still consistent with the orientation of K, and thus a fiber, but with fewer intersections with τ.

In all cases, we obtain contradictions, and conclude that τ and F can be arranged to be disjoint. □

Another theorem that we will find useful is also given by Scharlemann and Thompson. Ni [2009] proves a more general result, though we will not need it here.

Theorem 2.11 [Scharlemann and Thompson 2009]. *Suppose F is a compact orientable surface, L is a knot in $F \times I$, and $(F \times I)_{\text{surg}}$ is the 3-manifold obtained by some nontrivial surgery on L. If $F \times \{0\}$ compresses in $(F \times I)_{\text{surg}}$, then L is parallel to an essential simple closed curve in $F \times \{0\}$. Moreover, the annulus that describes the parallelism determines the slope of the surgery.*

The proof relies on sutured manifold theory, and a theorem of Gabai [1989]. Gabai proves the result for an annulus cross the interval. The idea of Scharlemann and Thompson's proof is to find product disks or annuli in $(F \times I)$ disjoint from the knot, and cut along these product pieces to reduce the complexity of the surface in question. This, with some additional work, allows them to apply the results of Gabai.

3. Pushing a tunnel into a fiber

Proof of Theorem 1.1. By Theorem 2.10, τ can be isotoped and slid to be disjoint from a fiber. Let $F = F' \setminus n(K)$. Cut $S^3 \setminus n(K)$ along F, to obtain $N \cong F \times I$, a handlebody. Then $\tau \subset N$.

Now, as τ is an unknotting tunnel, there exists a compressing disk for $\partial n(K \cup \tau)$ in $S^3 \setminus n(K \cup \tau)$, say D'. Note that $D' \cap F \neq \varnothing$, for otherwise F would be an essential surface in the solid torus $(S^3 \setminus n(K \cup \tau))|D'$, and thus a disk.

Consider $D' \cap F$. Since F is incompressible and N is irreducible, by standard innermost disk arguments we may assume there are no simple closed curves of intersection. Let α be an arc of intersection which is outermost in D', cutting off a subdisk D. Then, D is a disk in N with boundary consisting of three types of arcs: a single essential arc in $F = F \times \{0\}$, α; (several) arcs in $\partial n(K)$, call them ν_i; and (several) arcs in $\partial n(\tau)$, λ_j. We may assume that every arc of $D \cap \partial n(\tau)$ is an essential spanning arc of the annulus $\partial n(\tau)$, for trivial arcs can be removed by isotopy.

Now, consider the double of N, along the vertical boundary $\partial F \times I$. In other words, let $\widehat{N}$ be the result of gluing two copies of N together by the identity along $\partial F \times I$. Similarly, let $\widehat{\tau}$ be the result of gluing two copies of τ, one in each copy of N, along the boundary points; let $\widehat{D}$ come from two copies of D, one in each copy of N, glued along the ν_i; and let $\widehat{\alpha}$ come from two copies of α in the same way.

Then $\widehat{D}$ is a planar surface with one boundary component corresponding to $\widehat{\alpha}$, and several components coming from $\widehat{\lambda}_j$, the doubles of λ_j (see Figure 1).

Then, $\bigcup_j \widehat{\lambda}_j$ is a collection of (parallel) simple closed curves on the torus $\partial n(\widehat{\tau})$. Call the slope determined by these curves λ. If we perform λ-surgery on $\widehat{\tau}$, the result is to cap off $\widehat{D}$ with disks. Since α was essential in F, $\widehat{\alpha}$ is essential in $\widehat{F}$, so our capped-off surface is a compression disk for $\widehat{F}$ in $\widehat{N} \cong \widehat{F} \times I$.

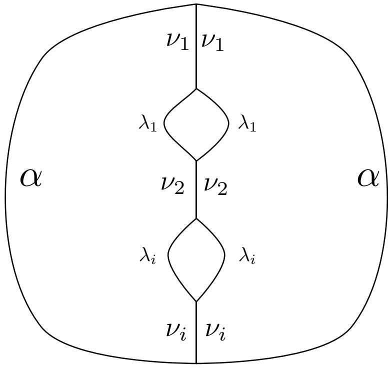


Figure 1. $\widehat{D}$.

By Theorem 2.11, $\widehat{\tau}$ is parallel to an essential closed curve in $\widehat{F} \times \{0\}$. That is, there exists an annulus A properly embedded in $\widehat{F} \times I$ with one boundary component on $\widehat{F} \times \{0\}$, say ψ, and the other boundary component on $\partial n(\widehat{\tau})$, parallel to $\widehat{\tau}$, say ϕ.

Since ϕ is parallel to $\widehat{\tau}$, it must be a longitude of $\partial n(\widehat{\tau})$, and in particular, $|\phi \cap (\partial F \times I)| = 2$. So there are only two possibilities for arcs of intersection between A and $\partial F \times I$ incident to ϕ. Either there is one arc of intersection which is trivial in A, or there are two arcs of intersection, both of which are essential in A (see Figure 2). The former case is impossible, because then the subdisk of A cut off by the arc would show that τ was parallel into $\partial n(K)$, which would imply that K was trivial. Therefore, there are exactly two arcs of $A \cap (\partial F \times I)$, both of which are essential in A.

If there were trivial arcs incident to ψ, then an outermost such arc in A would give rise to a boundary compression for $F \times \{0\}$ in $S^3 \setminus n(K)$. This is impossible as well, so $\partial F \times I$ intersects A in precisely two essential arcs, with no trivial arcs. Cutting A along these arcs provides a parallelism between τ and the arc $\psi \cap (F \times \{0\}) \subset \widehat{F} \times \{0\}$. Thus, τ can be isotoped to lie in the fiber. $\square$

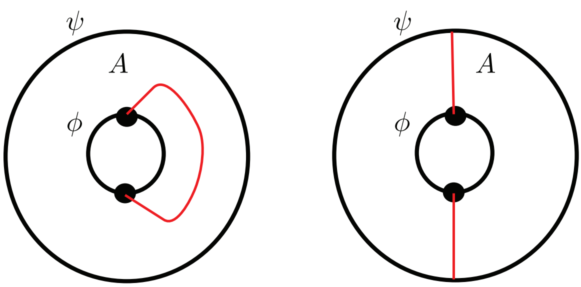


Figure 2. Arcs of $A \cap \partial F \times I$ incident to ϕ.

References

[Adams 1995] C. C. Adams, "Unknotting tunnels in hyperbolic 3-manifolds", *Math. Ann.* **302**:1 (1995), 177–195. MR 96c:57004 Zbl 0830.57009

[Adams and Reid 1996] C. C. Adams and A. W. Reid, "Unknotting tunnels in two-bridge knot and link complements", *Comment. Math. Helv.* **71**:4 (1996), 617–627. MR 98h:57009 Zbl 0873.57006

[Akiyoshi et al. 1997] H. Akiyoshi, Y. Nakagawa, and M. Sakuma, "Shortest vertical geodesics of manifolds obtained by hyperbolic Dehn surgery on the Whitehead link", pp. 433–448 in *KNOTS* (Tokyo, 1996), edited by S. Suzuki, World Scientific, River Edge, NJ, 1997. MR 99k:57034 Zbl 0960.57013

[Baker et al. 2008] K. L. Baker, J. E. Grigsby, and M. Hedden, "Grid diagrams for lens spaces and combinatorial knot Floer homology", *Int. Math. Res. Not.* **2008**:10 (2008), Art. ID rnm024. MR 2009h:57012 Zbl 1168.57009

[Boileau et al. 1988] M. Boileau, M. Rost, and H. Zieschang, "On Heegaard decompositions of torus knot exteriors and related Seifert fibre spaces", *Math. Ann.* **279**:3 (1988), 553–581. MR 89a:57013 Zbl 0616.57008

[Cooper et al. 2010] D. Cooper, M. Lackenby, and J. S. Purcell, "The length of unknotting tunnels", *Algebr. Geom. Topol.* **10**:2 (2010), 637–661. MR 2011c:57014 Zbl 1194.57021

[Eudave Muñoz and Uchida 1996] M. Eudave Muñoz and Y. Uchida, "Non-simple links with tunnel number one", *Proc. Amer. Math. Soc.* **124**:5 (1996), 1567–1575. MR 96g:57007 Zbl 0858.57008

[Gabai 1989] D. Gabai, "Surgery on knots in solid tori", *Topology* **28**:1 (1989), 1–6. MR 90h:57005 Zbl 0678.57004

[Giroux 2002] E. Giroux, "Géométrie de contact: de la dimension trois vers les dimensions supérieures", pp. 405–414 in *Proceedings of the International Congress of Mathematicians* (Beijing, 2002), vol. 2, edited by T. T. Li et al., Higher Ed. Press, Beijing, 2002. MR 2004c:53144 Zbl 1015.53049 arXiv math/0305129

[Goda et al. 2000] H. Goda, M. Scharlemann, and A. Thompson, "Levelling an unknotting tunnel", *Geom. Topol.* **4** (2000), 243–275. MR 2002h:57011 Zbl 0958.57007

[Heath and Song 2005] D. J. Heath and H.-J. Song, "Unknotting tunnels for $P(-2, 3, 7)$", *J. Knot Theory Ramifications* **14**:8 (2005), 1077–1085. MR 2007e:57008 Zbl 1115.57003

[Hedden 2007] M. Hedden, "Knot Floer homology of Whitehead doubles", *Geom. Topol.* **11** (2007), 2277–2338. MR 2008m:57030 Zbl 1187.57015

[Johnson 2008] J. Johnson, "Surface bundles with genus two Heegaard splittings", *J. Topol.* **1**:3 (2008), 671–692. MR 2009f:57025 Zbl 1148.57029

[Jones 1995] A. C. Jones, "Composite two-generator links", *Topology Appl.* **67**:3 (1995), 165–178. MR 97f:57008 Zbl 0854.57005

[Kobayashi 1999] T. Kobayashi, "Classification of unknotting tunnels for two bridge knots", pp. 259–290 in *Proceedings of the Kirbyfest* (Berkeley, CA, 1998), edited by J. Hass and M. Scharlemann, Geom. Topol. Monogr. **2**, Geom. Topol. Publ., Coventry, 1999. MR 2000j:57013 Zbl 0962.57003

[Lackenby 2005] M. Lackenby, "Classification of alternating knots with tunnel number one", *Comm. Anal. Geom.* **13**:1 (2005), 151–185. MR 2006e:57007 Zbl 1084.57007

[Morimoto 1994] K. Morimoto, "On composite tunnel number one links", *Topology Appl.* **59**:1 (1994), 59–71. MR 95f:57016 Zbl 0821.57006

[Morimoto and Sakuma 1991] K. Morimoto and M. Sakuma, "On unknotting tunnels for knots", *Math. Ann.* **289**:1 (1991), 143–167. MR 92e:57015 Zbl 0697.57002

[Ni 2007] Y. Ni, "Knot Floer homology detects fibred knots", *Invent. Math.* **170**:3 (2007), 577–608. MR 2008j:57053 Zbl 1138.57031

[Ni 2009] Y. Ni, "Dehn surgeries that yield fibred 3-manifolds", *Math. Ann.* **344**:4 (2009), 863–876. MR 2010m:57028 Zbl 1227.57012

[Ozsváth and Szabó 2005] P. Ozsváth and Z. Szabó, "On knot Floer homology and lens space surgeries", *Topology* **44**:6 (2005), 1281–1300. MR 2006f:57034 Zbl 1077.57012

[Saito 2007] T. Saito, "Dehn surgery and (1, 1)-knots in lens spaces", *Topology Appl.* **154**:7 (2007), 1502–1515. MR 2008f:57011 Zbl 1115.57005

[Sakuma 1996] M. Sakuma, "Unknotting tunnels and canonical decompositions of punctured torus bundles over a circle", *Sūrikaisekikenkyūsho Kōkyūroku* **967** (1996), 58–70. MR 98k:57021 Zbl 0924.57020

[Sakuma and Weeks 1995] M. Sakuma and J. Weeks, "Examples of canonical decompositions of hyperbolic link complements", *Japan. J. Math.* (*N.S.*) **21**:2 (1995), 393–439. MR 96j:57013 Zbl 0858.57021

[Scharlemann and Thompson 2003] M. Scharlemann and A. Thompson, "Unknotting tunnels and Seifert surfaces", *Proc. London Math. Soc.* (3) **87**:2 (2003), 523–544. MR 2004e:57015 Zbl 1047.57008

[Scharlemann and Thompson 2009] M. Scharlemann and A. A. Thompson, "Surgery on a knot in (surface $\times$ I)", *Algebr. Geom. Topol.* **9**:3 (2009), 1825–1835. MR 2010m:57018 Zbl 1197.57011

[Williams 2007] M. J. Williams, "Lens space surgeries & primitive/Seifert type constructions", preprint, 2007. arXiv 0712.1797

Received March 17, 2012. Revised June 4, 2012.

MATT RATHBUN
DEPARTMENT OF MATHEMATICS
HUXLEY BUILDING 6M33
IMPERIAL COLLEGE LONDON, SOUTH KENSINGTON CAMPUS
LONDON, SW7 2AZ
UNITED KINGDOM

m.rathbun@imperial.ac.uk
http://www2.imperial.ac.uk/~mrathbun/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 259, No. 2, 2012
dx.doi.org/10.2140/pjm.2012.259.483

FUSION SYMMETRIC SPACES AND SUBFACTORS

HANS WENZL

We construct analogs of the embedding of orthogonal and symplectic groups into unitary groups in the context of fusion categories. At least some of the resulting module categories also appear in boundary conformal field theory. We determine when these categories are unitarizable, and explicitly calculate the index and principal graph of the resulting subfactors.

This paper is a sequel of our previous paper [Wenzl 2012], where we introduced a q-deformation of Brauer's centralizer algebra for orthogonal and symplectic groups; this algebra had already appeared more or less before in [Molev 2003]; see also the discussion in [Wenzl 2012]. It is motivated by finding a deformation of orthogonal or symplectic subgroups of a unitary group that is compatible with the standard quantum deformation of the big group. This has been done before on the level of coideal subalgebras of Hopf algebras by Letzter. However, our categorical approach also allows us to extend this to the level of fusion tensor categories, where we find finite analogs of symmetric spaces related to the already mentioned groups. Moreover, we can establish C^* structures, necessary for the construction of subfactors, in this categorical setting; this is not so obvious to see in the setting of coideal algebras.

It is well-known how one can use a subgroup H of a (for simplicity here) finite group G to construct a module category of the representation category $\operatorname{Rep} G$ of G. This module category also appears in the context of subfactors of II_1 von Neumann factors as follows: Let R be the hyperfinite II_1 factor, and let $\mathcal{N} = R^G \subset \mathcal{M} = R^H$ be the fixed points under outer actions of G and H. Then the category of $\mathcal{N} - \mathcal{N}$ bimodules is equivalent to $\operatorname{Rep} G$, and the module category is given via the $\mathcal{M} - \mathcal{N}$ bimodules of the inclusion $\mathcal{N} \subset \mathcal{M}$; its simple objects are labeled by the irreducible representations of H. In particular, an important invariant called the principal graph of the subfactor is determined by the restriction rules for representations from G to H. Important examples of subfactors were constructed from fusion categories whose Grothendieck semirings are quotients of the ones of semisimple Lie groups. So a natural question to ask is whether one can perform a similar construction in this context. More precisely, can we find restriction rules for type A fusion categories

MSC2010: 46L10.
Keywords: module tensor categories, subfactors, q-Brauer algebra.

which describe a subfactor as before, and which will approach in the classical limit the usual restriction rules from $\mathrm{U}(N)$ to $\mathrm{O}(N)$.

We answer this question in the positive in this paper via a fairly elementary construction. We show that certain semisimple quotients of the q-Brauer algebras have a C^* structure and contain C^*-quotients of Hecke algebras of type A. The subfactor is then obtained as the closure of inductive limits of such algebras. Due to its close connection to Lie groups, we can give very explicit general formulas for its index and its first principal graph. Observe that the Lie algebra $\mathfrak{sl}_N$ decomposes as an $\mathfrak{so}_N$ module into the direct sum $\mathfrak{so}_N \oplus \mathfrak{p}$, where $\mathfrak{p}$ is a simple $\mathfrak{so}_N$-module. Then, the index can be expressed explicitly in terms of the weights of $\mathfrak{p}$; see Theorem 3.4. As before in the group case, it can be interpreted as the quotient of the dimension of the given fusion category by the sum of the squares of q-dimensions of representations of orthogonal or symplectic subgroups whose labels are in the alcove of a certain affine reflection group; however in our case, there is no corresponding tensor category for the denominator, and the q-dimensions differ from the ones of the corresponding quantum groups. Also, the restriction rules for the corresponding bimodules of this subfactor, the first principal graph, can be derived from the classical restriction rules via an action of the already mentioned affine reflection group, similarly as it was done before for tensor product rules for fusion categories. However, in our case, the affine reflection comes from the highest short root of the corresponding Lie algebra in the nonsimply laced case; it is also different from the one for fusion categories in the even-dimensional orthogonal case.

Not surprisingly for such a basic question, many related results have been obtained before in the study of subfactors, tensor categories and boundary conformal field theory. For $N = 2$, we obtain the Goodman–de la Harpe–Jones subfactors for Dynkin diagrams D_n. Subgroups and module categories in connection with SU(3) and SU(4) fusion categories have been studied by Ocneanu [2002] and by Evans and Pugh [2011]; our examples for $N = 3$ and $N = 4$ appear among the series in these works. The research in this and the just-mentioned papers has also been influenced by closely related results in boundary conformal field theory, which will be discussed in more detail at the end of the paper. Our examples for the odd-dimensional orthogonal group and for symplectic groups also seem to be closely related to type III_1 subfactors constructed by Feng Xu [2009] and Antony Wassermann [2010] by completely different methods.

The first chapter mostly contains basic material from subfactor theory which will be needed later. In the second chapter we review and expand material on the q-Brauer algebra as defined in [Wenzl 2012]; see also [Molev 2003]. In particular, we define C^*-structures for certain quotients and use that to construct subfactors. The third chapter is mainly concerned with the finer structure of these subfactors, such as explicit closed formulas for the index and calculation of the first principal graph. The same techniques would also extend to other examples, such as the ones in [Xu 2009].

1. II_1 factors

1A. ***Periodic commuting squares.*** We will construct subfactors using the setup of periodic commuting squares (going back to work of Jones and Popa) as in [Wenzl 1988a]. More precisely, we assume that we have increasing sequences of finite-dimensional C^* algebras $A_1 \subset A_2 \subset \cdots$ and $B_1 \subset B_2 \subset \cdots$ such that $A_n \subset B_n$ for all $n \in \mathbb{N}$. Let Λ_n and $\tilde{\Lambda}_n$ be labeling sets for the simple components of B_n and A_n, respectively. Let G_n be the inclusion matrix for $A_n \subset B_n$. If we write a minimal idempotent $p_\mu \in A_{n,\mu}$ as a sum of minimal mutually commuting idempotents of B_n, then the entry $g_{\lambda\mu}$ of G_n denotes the number of those idempotents which are in $B_{n,\lambda}$. We say that our sequences of algebras are periodic with period d if there exists an $n_0 \in \mathbb{N}$ such that for any $n > n_0$ we have bijections j between Λ_n and Λ_{n+d} as well as between $\tilde{\Lambda}_n$ and $\tilde{\Lambda}_{n+d}$ that do not change the inclusion matrices for $A_n \subset B_n$ as well as for $A_n \subset A_{n+1}$ and $B_n \subset B_{n+1}$. This means, in particular, that $g_{j(\lambda)j(\mu)} = g_{\lambda\mu}$ for all $\lambda \in \Lambda_n$, $\mu \in \tilde{\Lambda}_n$, $n > n_0$.

The trace functional defines inner products on the algebras A_n and B_n by

$$(b_1, b_2) = \mathrm{tr}(b_1^* b_2).$$

Let $e_{A_{n+1}}$ and e_{B_n} be the orthogonal projections onto the subspaces A_{n+1} and B_n of B_{n+1}. Then, the *commuting square condition* says that $e_{A_{n+1}} e_{B_n} = e_{A_n} = e_{B_n} e_{A_{n+1}}$ for all $n \in \mathbb{N}$. Finally, we also note that the trace tr is uniquely determined on A_n and B_n by its weight vectors $\boldsymbol{a}_n$ and $\boldsymbol{b}_n$ which are defined as follows: Let p_μ be a minimal idempotent in the simple component of A_n labeled by μ. Then we define $a_{n,\mu} = \mathrm{tr}(p_\mu)$, and $\boldsymbol{a}_n = (a_{n,\mu})_\mu$, where μ runs through a labeling set of the simple components of A_n. The weight vector $\boldsymbol{b}_n$ for B_n is defined similarly. The following proposition follows from [ibid.], Theorem 1.5 (where the matrix $G = (g_{\lambda\mu})$ defined here would correspond to the matrix G^t in [ibid.]).

Proposition 1.1. *Under the given conditions, we get a subfactor $\mathcal{N} \subset \mathcal{M}$ whose index $[\mathcal{M} : \mathcal{N}]$ is equal to $\|\boldsymbol{a}_n\|^2 / \|\boldsymbol{b}_n\|^2$ for any sufficiently large n. Moreover, we have $\sum g_{\lambda\mu} a_{n,\lambda} = [\mathcal{M} : \mathcal{N}]\, b_{n,\mu}$.*

1B. ***Special periodic algebras.*** In general, it can be quite hard to determine finer invariants of the subfactors, the so-called higher relative commutants (or centralizers) from the generating sequence of algebras. However, under certain circumstances, this can become quite easy. We describe such a setup. It is a moderate abstraction of an approach which has already been used before by a number of authors. The reader familiar with tensor categories and module categories should think of the algebras $A_n = \mathrm{End}_{\mathscr{C}}(X^{\otimes n})$ and $B_n = \mathrm{End}_{\mathscr{D}}(Y \otimes X^{\otimes n})$ for X an object in a C^* tensor category $\mathcal{C}$ and Y an object in a module category $\mathcal{D}$ over $\mathcal{C}$. In the following, we will make the following assumptions beyond the ones in the previous subsection:

1. The algebras A_n will be monoidal C^*-algebras. This means we have canonical embeddings of C^* algebras $A_m \otimes A_n \to A_{n+m}$ with multiplicativity of the trace, that is, $\mathrm{tr}(a_1 \otimes a_2) = \mathrm{tr}(a_1)\,\mathrm{tr}(a_2)$.
2. We have canonical embeddings $B_m \otimes A_n \to B_{n+m}$, again with multiplicativity of the trace.
3. We have the commuting square condition for the sequences of algebras $A_n \subset B_n$ and $1 \otimes A_{n-1} \subset A_n$.
4. There exists $d \in \mathbb{N}$ and a projection $p \in A_d$ such that $(1_m \otimes p)A_{m+d}(1 \otimes p) \cong A_m$ and $(1_m \otimes p)B_{m+d}(1 \otimes p) \cong B_m$ for all $m \in \mathbb{N}$.

Examples for this setup will be given at the end of this section and in Section 2. Moreover, any finite depth subfactor $\mathcal{N} \subset \mathcal{M}$ (see [Goodman et al. 1989; Evans and Kawahigashi 1998] for definitions) produces algebras for such a setup as follows: Let $\mathcal{M}^{\otimes n} = \mathcal{M} \otimes_{\mathcal{N}} \mathcal{M} \otimes_{\mathcal{N}} \cdots \otimes_{\mathcal{N}} \mathcal{M}$ (n factors). Obviously, $\mathcal{M}^{\otimes n}$ is an $\mathcal{N} - \mathcal{N}$ as well as an $\mathcal{M} - \mathcal{N}$ bimodule. One can check that for

$$A_n = \mathrm{End}_{\mathcal{N}-\mathcal{N}}\, \mathcal{M}^{\otimes n} \subset B_n = \mathrm{End}_{\mathcal{M}-\mathcal{N}}\, \mathcal{M}^{\otimes n+1}$$

the axioms above are satisfied; here the embedding is defined by letting the elements of A_n act on the second to $(n+1)-st$ factor of $\mathcal{M}^{\otimes n+1}$. It is also possible to define these algebras in connection of relative commutants in the Jones tower of relative commutants (see [Bisch 1997] for details). Recall that for factors $\mathcal{N} \subset \mathcal{M}$ the relative commutant (or centralizer) $\mathcal{N}' \cap \mathcal{M}$ is defined to be the set $\{b \in \mathcal{M} \mid ab = ba \text{ for all } a \in \mathcal{N}\}$.

Lemma 1.2. *The subfactor $\mathcal{N} \subset \mathcal{M}$ generated from the sequences of algebras $1_m \otimes A_n \subset B_{n+m}$ has relative commutant B_m. The same statement also holds with B_{n+m} and B_m in the last sentence replaced by A_{n+m} and A_m.*

Proof. This is essentially the proof used for Theorem 3.7 in [Wenzl 1988a]. Observe that by induction on r and assumption 4 above, we also have

$$(1_m \otimes p^{\otimes r})X_{m+rd}(1_m \otimes p^{\otimes r}) \cong X_m$$

for $X = A, B$. It follows from Theorem 1.6 of [ibid.] that the dimension of the relative commutant $\mathcal{N}' \cap \mathcal{M}$ is at most equal to the dimension of B_m. The claim follows from the fact that $B_m \otimes 1_n$ commutes with $1_m \otimes A_n$ for all n. □

1C. ***Bimodules and principal graphs.*** We calculate the first principal graph for subfactors constructed in our setup, using fairly elementary methods from [ibid.] as well as the bimodule approach. The latter was first used in the subfactor context by Ocneanu; see [Evans and Kawahigashi 1998]. For the connection between bimodules and principal graphs, see [Bisch 1997] and for more details compatible with our notation, see also [Erlijman and Wenzl 2007]. While most of this section has already appeared before implicitly or explicitly, the presentation in our setup might be useful also in other contexts.

Pick k large enough so that $m = kd > n_0$. Hence, the inclusion matrices for $A_{rd} \subset B_{rd}$ coincide for all $r \geq k$ using the bijection of simple components as described in Section 1A. Let Λ_m and $\tilde{\Lambda}_m$ be labeling sets for the simple components of B_m and A_m respectively. Let $\mathcal{N}$ and $\mathcal{M}$ be the factors generated by the increasing sequences of algebras A_n and B_n respectively; see Proposition 1.1 or Lemma 1.2, with the m there equal 0. Both of these factors have a subfactor $\tilde{\mathcal{N}}$ generated by the subalgebras

$$1_m \otimes A_n \subset A_{n+m} \subset B_{n+m}.$$

We now define for each $\lambda \in \tilde{\Lambda}_m$ an $\mathcal{N} - \tilde{\mathcal{N}}$ bimodule N_λ as follows: It is the Hilbert space completion of $\mathcal{N} p_\lambda$ with respect to the inner product induced by tr, where p_λ is a minimal idempotent in $A_{m,\lambda}$, the simple component of A_m labeled by λ, with obvious left and right actions by $\mathcal{N}$ and $\tilde{N}$. To ease notation, we shall often refer to it as an $\mathcal{N} - \mathcal{N}$ bimodule, using the isomorphism between $\tilde{\mathcal{N}}$ and $\mathcal{N}$ given by the trace preserving maps $a \in A_n \mapsto 1_m \otimes a \in A_{n+m}$.

Similarly, we define $\mathcal{M} - \tilde{\mathcal{N}}$ bimodules M_μ for any $\mu \in \Lambda_m$ which are Hilbert space completions of $\mathcal{M} p_\mu$, where p_μ is a minimal idempotent in the simple component $B_{m,\mu}$ of B_m. Finally, we define the inclusion numbers b_μ^λ for elements $\lambda \in \tilde{\Lambda}$ and $\mu \in \Lambda_m$ as usual (see Section 1A).

Lemma 1.3. *The bimodules N_λ and M_μ are irreducible $\mathcal{N} - \tilde{\mathcal{N}}$ and $\mathcal{M} - \tilde{\mathcal{N}}$ bimodules, respectively. We have the decomposition $M_\mu \cong \bigoplus_\lambda b_\mu^\lambda N_\lambda$ as $\mathcal{N} - \tilde{\mathcal{N}}$ modules.*

Proof. This is well-known (see [Erlijman and Wenzl 2007] for more details). It follows from Lemma 1.2 that the endomorphism ring of the $\mathcal{M} - \tilde{\mathcal{N}}$ bimodule $\mathcal{M}$ is given by B_m. Hence the $\mathcal{M} - \tilde{\mathcal{N}}$ bimodules M_μ are simple, as p_μ was chosen to be a minimal idempotent in B_m. One shows similarly that also the N_λ are simple $\mathcal{N} - \tilde{\mathcal{N}}$ bimodules.

Observe that $\dim_{\mathcal{N}} N_\lambda = \mathrm{tr}(p_\lambda)$ and $\dim_{\mathcal{M}} M_\mu = \mathrm{tr}(p_\mu)$; see [Jones 1983]. Now if p_λ is a minimal idempotent in A_m, it follows from the definitions that $\mathrm{Ind}_{\mathcal{N}}^{\mathcal{M}} N_\lambda := \mathcal{M} p_\lambda$ is isomorphic as an $\mathcal{M} - \tilde{\mathcal{N}}$ bimodule to the direct sum $\oplus b_\mu^\lambda M_\mu$. By Frobenius reciprocity (see [Evans and Kawahigashi 1998; Bisch 1997]) it follows that the module N_λ appears with multiplicity b_μ^λ in $\mathcal{M}_\mu$, viewed as an $\mathcal{N} - \tilde{\mathcal{N}}$ bimodule. Hence the $\mathcal{N} - \tilde{\mathcal{N}}$ bimodule M_μ has a submodule which is isomorphic to $\bigoplus_\lambda b_\mu^\lambda N_\lambda$. But as M_μ has $\mathcal{N}$-dimension $[\mathcal{M} : \mathcal{N}]\,\mathrm{tr}(p_\mu)$, it coincides with this submodule, by Proposition 1.1. □

Theorem 1.4. *Let $\mathcal{N} \subset \mathcal{M}$ be the subfactor generated by sequences of algebras $A_n \subset B_n$ satisfying the conditions in Section 1B. Then its first principal graph is given by the inclusion graph for $A_{kd} \subset B_{kd}$ for sufficiently large k.*

Proof. It is well-known that the first principal graph is given by the induction-restriction graph of $\mathcal{M} - \mathcal{N}$ and $\mathcal{N} - \mathcal{N}$ bimodules appearing in the tensor products $\mathcal{M}^{\otimes n}$, $n \in \mathbb{N}$, where $\mathcal{M}^{\otimes n} = \mathcal{M} \otimes_{\mathcal{N}} \mathcal{M} \otimes_{\mathcal{N}} \cdots \otimes_{\mathcal{N}} \mathcal{M}$ (n factors); see [Evans and Kawahigashi 1998; Bisch 1997]. Obviously, this graph does not change if we replace all $X - \mathcal{N}$ bimodules H in this setting by $X - q\mathcal{N}q$ bimodules Hq, for $X = \mathcal{M}, \mathcal{N}$ and q a nonzero projection in $\mathcal{N}$. The claim can now be shown for $q = p^{\otimes k}$ where k is chosen large enough so that $kd > n_0$, using Lemma 1.3. □

Recall that many examples come from module tensor categories, where $A_n = \mathrm{End}_{\mathbb{C}}(X^{\otimes n})$ and $B_n = \mathrm{End}(Y \otimes (X^{\otimes n}))$ for an object X in a tensor category $\mathscr{C}$ and an object Y in the module category $\mathscr{D}$ over $\mathscr{C}$. In this setting, the weight vectors of our trace are given by $a_{n,\lambda} = \tilde{d}_\lambda / x^n$ and $b_{n,\mu} = d_\mu / yx^n$ for positive quantities $d_\mu, \tilde{d}_\lambda, x$ and y. Then we have:

Corollary 1.5. *Assuming the conditions for the trace weights as just given, we have subfactors $\mathcal{N} \subset \mathcal{M}_\mu$ with index $[\mathcal{M}_\mu : \mathcal{N}] = d_\mu^2 [\mathcal{M} : \mathcal{N}]$, with $\mathcal{N} \subset \mathcal{M}$ as in Theorem 1.4.*

Remark 1.6. There is also a second important invariant for $\mathcal{N} \subset \mathcal{M}$, the dual principal graph. It can be analogously defined as an induction-restriction graph between irreducible $\mathcal{M} - \mathcal{M}$ and $\mathcal{M} - \mathcal{N}$ bimodules appearing in the tensor powers $\mathcal{M}^{\otimes n}$. Its calculation is more difficult than that of the first principal graph. This is quite similar to the corresponding problem for subfactors coming from conformal inclusions and related constructions; see [Xu 1998; Böckenhauer et al. 1999; Erlijman and Wenzl 2007]. We plan to study this problem in a future publication via suitable adaptions of techniques in those papers.

1D. ***The GHJ-construction.*** We give a well-known and well-studied example for our current setup, which was first constructed in [Goodman et al. 1989]. Let G be a matrix with nonnegative integer entries and norm less than 2. It is well-known that such matrices are classified by Coxeter graphs of type ADE. We assume that the columns of G are indexed by the even vertices, and the rows by the odd vertices. We define C^*-algebras B_n by $B_0 = \mathbb{C}^{v_e}$, and $B_1 = \oplus M_{d_j}$, where v_e is the number of even vertices, and the summands of B_1 are labeled by the odd vertices j, whose dimension d_j is equal to the number of even vertices to which j is connected. The embedding $B_o \subset B_1$ is given by the inclusion matrix G. Then we define recursively B_{n+1} via Jones' basic construction [1983] for $B_{n-1} \subset B_n$. Here the trace on B_n is the unique normalized trace whose values on minimal idempotents are given by the Perron–Frobenius vector of G^tG or GG^t, depending on whether n is even or odd, and the vector is normalized such that $\mathrm{tr}(1) = 1$. Then the algebra B_{n+1} is generated by B_n, acting on itself via left multiplication, and the orthogonal projection e_n onto the subspace B_{n-1} of B_n, with respect to the inner product coming from the trace. The algebra A_n is defined to be the subalgebra generated by the identity 1 and the projections e_i, $1 \le i < 1$. It is well-known that these algebras satisfy the commuting square condition, that they are periodic with periodicity 2, and that the Jones projections e_i satisfy the conditions of the projection p in Section 1B. This has already been shown in [Goodman et al. 1989].

2. q-Brauer algebras

2A. *Definitions.* Fix $N \in \mathbb{Z}$ and let $[N] = (q^N - q^{-N})/(q - q^{-1})$, where q is considered to be a complex number. We denote by $H_n(q^2)$ the Hecke algebra of type A_{n-1}. It is given by generators $g_1, g_2, \ldots, g_{n-1}$ which satisfy the usual braid relations and the quadratic relation $g_i^2 = (q^2 - 1)g_i + q^2$. The q-Brauer algebra $\mathrm{Br}_n(N)$ is the complex algebra defined via generators $g_1, g_2, \ldots, g_{n-1}$ and e and the following relations:

(H) $g_1, g_2, \ldots, g_{n-1}$ satisfy the relations of the Hecke algebra $H_n(q^2)$.

(E1) $e^2 = [N]e$.

(E2) $eg_i = g_ie$ for $i > 2$, $eg_1 = q^2e$, $eg_2e = q^{N+1}e$ and $eg_2^{-1}e = q^{-1-N}e$.

(E3) $g_2g_3g_1^{-1}g_2^{-1}e_{(2)} = e_{(2)} = e_{(2)}g_2g_3g_1^{-1}g_2^{-1}$, where $e_{(2)} = e(g_2g_3g_1^{-1}g_2^{-1})e$.

Remark 2.1. (a) Relation (E3) can be replaced by the perhaps slightly less mysterious relation $e_{(2)}g_2g_3 = e_{(2)}g_2g_1$ and $g_1^{-1}g_2^{-1}e_{(2)} = g_3^{-1}g_2^{-1}e_{(2)}$.

(b) It is easy to see that this algebra coincides with the algebra defined in [Wenzl 2012] after substituting q there by q^2, and e there by $q^{1-N}e$ (with the q of this paper); this is also compatible with the different definition of $[N]$ in [ibid.]. We have chosen this parametrization as it will make it easier to define a $*$-structure on it. More precisely, if $|q| = 1$, there exists a complex conjugate antiautomorphism $b \mapsto b^*$ on $\mathrm{Br}_n(N)$ defined by

$$e^* = e, \quad g_i^* = g_i^{-1}, \quad \text{where } 1 \le i < n. \tag{2-1}$$

It is easy to check the relations to show this operation is well-defined.

2B. *Molev representation.* We give a representation of our algebra $\mathrm{Br}_n(N)$ in $\mathrm{End}(V^{\otimes n})$, where $V = \mathbb{C}^N$. For this we use the matrices used by Molev [2003] for the definition of his q-deformation of Brauer's centralizer algebra. His defining relations are slightly different from ours, but Molev has informed the author that our algebra satisfies the relations of his algebra. It turns out that also his matrices satisfy the relations of our algebras, which we will outline here. Let R be the well-known solution of the quantum Yang–Baxter equation for type A. For simplicity we will use this notation for what is often denoted as $\check{R}$. If E_{ij} are the matrix units for $n\times n$ matrices, we define the following elements in $\mathrm{End}(V^{\otimes 2})$:

$$R=\sum_i qE_{ii}\otimes E_{ii}+\sum_{i\neq j}E_{ij}\otimes E_{ji}+\sum_{i<j}(q-q^{-1})E_{ii}\otimes E_{jj},$$

and

$$Q=\sum_{i,j}q^{N+1-2i}E_{ij}\otimes E_{ij}.$$

Moreover, if $A\in\mathrm{End}(V^{\otimes 2})$, we define the operator $A_i\in\mathrm{End}(V^{\otimes n})$ by

$$A_i\;=\;1_{i-1}\otimes A\otimes 1_{n-1-i},$$

where 1_k is the identity on $V^{\otimes k}$. Then we have the following proposition, all of whose essential parts were already proved in [Molev 2003]. However, the relations for our algebras are slightly different, so we spell out some of the adjustments of the work in [ibid.] in our context below.

Proposition 2.2. *The map $g_i\mapsto qR_{n-i}$ and $e\mapsto Q_{n-1}$ defines a representation Φ of* $\mathrm{Br}_n(N)$. *It specializes to the usual representation of Brauer's centralizer algebra in* $\mathrm{End}(V^{\otimes n})$ *for* $q=1$.

Proof. Most of the relations are already known or are easy to check. For example, it is well-known that the matrices qR_i satisfy the relations of the Hecke algebra $H_n(q^2)$. Relation (E1) is checked easily, and also the relations in (E2) are fairly straightforward to check. It suffices to check (E3) for $n=4$. For this observe that by [ibid., (4.16)], we have

$$Q_3R_2R_3R_1R_2Q_3=Q_1Q_3+q^{N+1}(q-q^{-1})Q_3(R_1+q^{-1}1),$$

in our notation. Using the relation $R_i=R_i^{-1}+(q-q^{-1})1$ for the second and third factor of the left hand side, one derives from this

$$Q_3R_2^{-1}R_3^{-1}R_1R_2Q_3=Q_1Q_3.$$

To check relation (E3), observe that

$$R_1R_2(v_i\otimes v_i\otimes v_j\otimes v_j)=R_3R_2(v_i\otimes v_i\otimes v_j\otimes v_j),$$

where (v_i) is the standard basis for $\mathbb{C}^N=V$. One derives from this that

$$R_2^{-1}R_1^{-1}R_3R_2Q_1Q_3=Q_1Q_3.$$

Moreover, the same calculations above also work with R_i replaced by R_i^{-1} and Q_j replaced by its transpose Q_j^T. Hence one can show as before that

$$R_2R_3R_1^{-1}R_2^{-1}Q_1^TQ_3^T=Q_1^TQ_3^T.$$

Transposing this, using $R_i^T = R_i$ shows the last part of the claim. □

2C. *Quotients.* We can now rephrase the main results of [Wenzl 2012] in our notation as follows:

Theorem 2.3. (a) *There exists a well-defined functional* tr *on* $\mathrm{Br}_n(N)$ *defined inductively by* $\mathrm{tr}(g_1) = q^{N+1}/[N]$, $\mathrm{tr}(e) = 1/[N]$ *and* $\mathrm{tr}(bg_n) = \mathrm{tr}(b)\,\mathrm{tr}(g_n)$ *for all* $b \in \mathrm{Br}_n(N)$.

(b) *Let* $\overline{\mathrm{Br}}_n(N) = \mathrm{Br}_n(N)/I_n$, *where* I_n *is the annihilator ideal of* tr. *Then* $\overline{\mathrm{Br}}_n(N)$ *is semisimple and the inclusion* $\overline{\mathrm{Br}}_n(N) \subset \overline{\mathrm{Br}}_{n+1}(N)$ *is well-defined for all* n.

It is possible to explicitly describe the structure of the quotients $\overline{\mathrm{Br}}_n = \overline{\mathrm{Br}}_n(N)$. To do so, we need the following definitions for the labeling sets of simple representations. More conceptually, the labeling sets $\Lambda(N, \ell)$ consist of all such diagrams λ for which the quantities $d_\mu(q) \neq 0$ for any subdiagrams $\mu \subset \lambda$ including λ itself, where q^2 is a primitive ℓ-th root of unity and the d_μ are defined in Section 2E.

Definition 2.4. Fix integers N and ℓ satisfying $1 < |N| < \ell$.

(i) The set $\tilde{\Lambda}(N, \ell)$ consists of all Young diagrams with at most N rows such that the first and N-th row differ by at most $\ell - N$ boxes for $N > 0$. If $N < 0$, the Young diagrams have at most $|N|$ columns, where the first and $|N|$-th column differ by at most $\ell - |N|$ boxes.

(ii) The set $\Lambda(N, \ell)$ consists of all Young diagrams λ with λ_i boxes in the i-th row and λ'_j boxes in the j-th column which satisfy

(a) $\lambda'_1 + \lambda'_2 \leq N$ and $\lambda_1 \leq (\ell - N)/2$ if $N > 0$ and $\ell - N$ even,
(b) $\lambda'_1 + \lambda'_2 \leq N$ and $\lambda_1 + \lambda_2 \leq \ell - N$ if $N > 0$ and $\ell - N$ odd,
(c) $\lambda_1 \leq |N|/2$ and $\lambda'_1 + \lambda'_2 \leq \ell - |N|$ if $N < 0$ is even,
(d) $\lambda_1 + \lambda_2 \leq |N|$ and $\lambda'_1 + \lambda'_2 \leq \ell - |N|$ if $N < 0$ is odd.

Diagrams which miss one of these inequalities only by the quantity one are called boundary diagrams of $\Lambda(N, \ell)$; for example in case (a) if $\lambda'_1 + \lambda'_2 = N + 1$.

Theorem 2.5 [Wenzl 2012, Section 5]. *Let* q^2 *be a primitive* ℓ*-th root of unity, and let* N *be an integer satisfying* $1 < |N| < \ell$. *Then the simple components of* $\overline{\mathrm{Br}}_n = \overline{\mathrm{Br}}_n(N)$ *are labeled by the Young diagrams in* $\Lambda(N, \ell)$ *with* $n, n-2, n-4, \ldots$ *boxes. If* $V_{n,\lambda}$ *is a simple* $\overline{\mathrm{Br}}_n$*-module for such a diagram* λ, *it decomposes as a* Br_{n-1} *module as*

$$V_{n,\lambda} \cong \bigoplus_\mu V_{n-1,\mu}, \tag{2-2}$$

where μ *runs through diagrams in* $\Lambda(N, \ell)$ *obtained by removing or, if* $|\lambda| < n$, *also by adding a box to* λ.

2D. ***Path idempotents and matrix units.*** We will give some details about the proof of Theorem 2.5 which will also be needed for further results. Observe that the restriction rule (2-2) implies that a minimal idempotent p_μ in $\overline{\mathrm{Br}}_{n-1,\mu}$ can be written as a sum of minimal idempotents with exactly one in $\overline{\mathrm{Br}}_{n,\lambda}$ for each diagram λ in $\Lambda(N,\ell)$ that can be obtained by adding or subtracting a box from μ. This inductively determines a system of minimal idempotents and matrix units of $\overline{\mathrm{Br}}_n(q^N,q)$ labeled by paths and pairs of paths, respectively, in $\Lambda(N,\ell)$ of length n. Such a path is defined to be a sequence of Young diagrams $(\lambda^{(i)})_{i=0}^n$ where $\lambda^{(0)}$ is the empty Young diagram, and $\lambda^{(i+1)}$ is obtained from $\lambda^{(i)}$ by adding or removing a box. It follows from the restriction rule above that the dimension of $V_{n,\lambda}$ is equal to the number of paths of length n with $\lambda^{(n)}=\lambda$, and that we can label a complete system of matrix units for the simple component $\overline{\mathrm{Br}}_{n,\lambda}$ by pairs of such paths. We then have the following lemma:

Lemma 2.6. *For each pair of paths t_1, t_2 in $\Lambda(N,\infty)$ with the same endpoint we can define the matrix unit E_{t_1,t_2} as a linear combination of products of generators over algebraic functions (rational for path idempotents) in q with poles only at roots of unity. More precisely, the formula for E_{t_1,t_2} is well-defined for q^2 a primitive ℓ-th root of unity if both t_1 and t_2 are paths in $\Lambda(N,\ell)$.*

Proof. This was proved in [Wenzl 2012], Section 5. As the result is not explicitly stated as such, we give some details here. One observes that the two-sided ideal generated by the element $\bar e\in\overline{\mathrm{Br}}_{n+1}$ is isomorphic to Jones' basic construction for the algebras $\overline{\mathrm{Br}}_{n-1}\subset\overline{\mathrm{Br}}_n$ (or, strictly speaking, by certain conjugated subalgebras which are denoted by $i_1(\overline{\mathrm{Br}}_n)$ and $i_2(\overline{\mathrm{Br}}_{n-1})$; see Section 5.2 in [Wenzl 2012]). One can then define path idempotents and matrix units inductively as it was done in [Ram and Wenzl 1992, Theorem 1.4], using the formulas for the weights of traces, which will also be reviewed in Section 2E; this is closely related to what is also known in subfactor theory as the Ocneanu–Sunder path model [Sunder 1987]. The complement of this ideal is a quotient of the Hecke algebra $\bar H_{n+1}$, for which matrix units already were more or less defined in [Wenzl 1988a, p. 366]. □

Lemma 2.7. *Let $p_{[N]}$ and $p_{[1^N]}$ be the minimal idempotents in H_N corresponding to its one-dimensional trivial and sign representations. Then for all $m>0$ we have*

$$\bar p_{[1^N]}^{\otimes 2}\overline{\mathrm{Br}}_{m+2N}\,\bar p_{[1^N]}^{\otimes 2}\cong\overline{\mathrm{Br}}_m$$

for $N>0$, and

$$p_{[-N]}\overline{\mathrm{Br}}_{-N+m}\,p_{[-N]}\cong\overline{\mathrm{Br}}_m$$

for $N<0$ even.

Proof. Let us assume $N>0$ first. Observe that if $p\in\overline{\mathrm{Br}}_{2N,\varnothing}(N)$, the simple component labeled by the empty Young diagram $\varnothing$, then it follows from the restriction rule (2-2) (see also the equivalent version below Theorem 2.5) by induction on m that $p\overline{\mathrm{Br}}_{m+2N}(N)p \cong \overline{\mathrm{Br}}_m(N)$ for all $m \geq 0$. Hence it suffices to show that $p_{[1^N]}^{\otimes 2}$ is such an idempotent.

If $q = 1$ and $N > 0$, $\Phi(\mathrm{Br}_n(N))$ coincides with the commutant of the action of the orthogonal group $\mathrm{O}(N)$ on $V^{\otimes n}$, which is semisimple. Moreover, the trace tr is just a multiple of the pull-back of the natural trace on $\mathrm{End}(V^{\otimes n})$, so $\Phi(\mathrm{Br}_n(N)) \cong \overline{\mathrm{Br}}_n(N)$ at $q = 1$. As $\Phi(p_{[1^N]})$ projects onto the one-dimensional determinant representation in $V^{\otimes N}$, the claim follows easily in that case, using Brauer duality, that is, the fact that $\Phi(\mathrm{Br}_n(N))$ is equal to the commutant of $\mathrm{O}(N)$ on $V^{\otimes n}$ for all n.

We will now use the fact that we can also define $\mathrm{Br}_n(N)$ over the field of rational functions $\mathbb{C}(q)$; see [Wenzl 2012]. It follows from Lemma 2.6 that we can also define the path idempotents for $\overline{\mathrm{Br}}_n(N)$ over that field for paths of length n in $\Lambda(N, \infty)$. As the rank of an idempotent is an integer, the claim follows as well for q a variable, and for $q \in \mathbb{C}$ not a root of unity. But as $p_t \bar{p}_{[1^N]}^{\otimes 2} p_t = 0$ for any path t of length $2N$ in $\Lambda(N, \ell)$ which ends in $\lambda \neq \varnothing$, we also get rank 0 for $\bar{p}_{[1^N]}^{\otimes 2}$ at q^2 a primitive ℓ-th root of unity in $\overline{\mathrm{Br}}_{2N,\lambda}(N)$. One also shows by a similar continuity and path idempotent argument that the rank of $p_{[1^N]}^{\otimes 2}$ is equal to 1 in $\overline{\mathrm{Br}}_{2N,\varnothing}(2n)$. This finishes the proof for $N > 0$.

If $N < 0$ even, we would map the permutation $(i, i+1)$ to the negative of the linear map permuting the i-th and $(i+1)$-st factor of $V^{\otimes n}$, where $V = \mathbb{C}^{|N|}$. Hence $p_{[-N]}$ would map onto the antisymmetrization of $V^{\otimes -N}$, on which $\mathrm{Sp}(|N|)$ acts trivially. The map above extends to a map of $\mathrm{Br}_n(N)$ onto $\mathrm{End}_{\mathrm{Sp}(|N|)}(V^{\otimes n})$ (see [Wenzl 1988b]), and we can now duplicate the proof for the orthogonal case. □

2E. *Weights of the trace.* Using the character formulas of orthogonal groups, one can calculate the weights of tr for the algebras $\mathrm{Br}_n(N)$, that is, its values at minimal idempotents of $\mathrm{Br}_n(N)$. We will need the following quantities for a given Young diagram λ:

$$d(i, j) = \begin{cases} \lambda_i + \lambda_j - i - j & \text{if } i \leq j, \\ -\lambda'_i - \lambda'_j + i + j - 2 & \text{if } i > j. \end{cases} \tag{2-3}$$

Moreover, we define $h(i, j)$ to be the length of the hook in the Young diagram λ whose corner is the box in the i-th row and j-th column. We can now restate [Wenzl 2012], Theorem 4.6 in the notations of this paper as follows:

Theorem 2.8. *The weights of the Markov trace* tr *for the Hecke algebra* $\bar{H}_n(q^2)$ *are given by* $\tilde{\omega}_\lambda = \tilde{d}_\lambda/[N]^n$, *where* $|\lambda| = n$, *and for* $\overline{\mathrm{Br}}_n(N)$ *they are given by* $\omega_{\lambda,n} = d_\lambda/[N]^n$, *where*

$$\tilde{d}_\lambda = \prod_{(i,j)\in\lambda} \frac{[N+j-i]}{[h(i, j)]}, \qquad d_\lambda = \prod_{(i,j)\in\lambda} \frac{[N+d(i, j)]}{[h(i, j)]},$$

where λ runs through all the Young diagrams in $\tilde\Lambda(N,\ell)$ with n boxes for $\bar H_n(q^2)$, and through all Young diagrams in $\Lambda(N,\ell)$ with $n, n-2, n-4, \dots$ boxes for $\overline{\mathrm{Br}}_n$.

Lemma 2.9. *The weights $\omega_{\lambda,n}$ are positive for all $\lambda \in \Lambda(N,\ell)$ if and only if $q^2 = e^{\pm 2\pi i/\ell}$ with $\ell > N$ and*

(a) *$N > 0$ and $\ell - N$ even, or*

(b) *$N < 0$ odd.*

Proof. The weights can be rewritten for our choice of q as

$$\omega_{\lambda,n} = \frac{\sin^n(\pi/\ell)}{\sin^n(N\pi/\ell)^n} \prod_{(i,j)\in\lambda} \frac{\sin(N+d(i,j))\pi/\ell}{\sin(h(i,j)\pi/\ell)}.$$

As $h(i,j) \le h(1,1) = \lambda_1 + \lambda_1' - 1 < \ell$ for all boxes (i,j) of λ, it follows that all factors in the formula above are positive for $N > 0$ (negative for $N < 0$) except possibly the ones in the numerator under the product. If $N > 0$ and $\ell - N$ odd, one checks that for the diagram $[\ell - N + 1)/2]$ we have $\omega_{\lambda,|\lambda|} < 0$. By the same argument, one shows that $\omega_{\lambda,|\lambda|} < 0$ for $\lambda = [(|N|+1)/2]$ and $N < 0$. In the other two cases, one checks that $0 < |d(i,j)| < \ell$ for all boxes (i,j) of a diagram $\lambda \in \Lambda(N,\ell)$. □

2F. *C^*-quotients.*

Proposition 2.10. *If the weights $\omega_{\lambda,n}$ are positive for all $\lambda \in \Lambda(N,\ell)$, the star operation defined by $e^* = e$ and by $g_i^* = g_i^{-1}$ makes the quotients $\overline{\mathrm{Br}}_n$ into C^*-algebras.*

Proof. The proof goes by induction on n, with the claims for $n = 1$ and $n = 2$ easy to check. By [Wenzl 2012], the two-sided ideal I_{n+1} generated by e in $\overline{\mathrm{Br}}_{n+1}$ is isomorphic to Jones' basic construction for $\overline{\mathrm{Br}}_{n-1} \subset \overline{\mathrm{Br}}_n$; see also the remarks before Lemma 2.7. In particular, I_{n+1} is spanned by elements $b_1 e b_2$, with $b_1, b_2 \in i_1(\overline{\mathrm{Br}}_n)$, where $i_1(a) = \Delta_{n+1} a \Delta_{n+1}^{-1}$, with $\Delta = (g_1 g_2 \cdots g_{n-1})(g_1 \cdots g_{n-2}) \cdots g_1$. By induction assumption and properties of Jones' basic construction, this ideal has a C^*-structure given by $(b_1 e b_2)^* = b_2^* e b_1^*$. This coincides with the $*$ operation defined before algebraically. It was shown in [Wenzl 2012] that $\overline{\mathrm{Br}}_{n+1} \cong I_{n+1} \oplus \bar H_{n+1}$, where $\bar H_{n+1}$ is a semisimple quotient of the Hecke algebra H_{n+1} whose simple components are labeled by the Young diagrams $\lambda \in \Lambda(N,\ell)$ with $n+1$ boxes. All these simple representations satisfy the (k,ℓ) condition in [Wenzl 1988a]. It follows from that paper that the map $g_i^* = g_i^{-1}$ induces a C^* structure for any such representation. This finishes the proof. □

Theorem 2.11. *For each choice of N and ℓ with $q^2 = e^{\pm 2\pi i/\ell}$, and for each nonnegative integer m, we obtain a subfactor $\mathcal{N} \subset \mathcal{M}$ with $\mathcal{N}' \cap \mathcal{M} = \overline{\mathrm{Br}}_m$ and with index*

$$[\mathcal{M} : \mathcal{N}] = [N]^m \frac{\sum_{\mu \in \tilde\Lambda(N,\ell), |\mu| = k|N|} \tilde d_\mu^2}{\sum_{\lambda \in \Lambda(N,\ell), 2||\lambda|} d_\lambda^2},$$

with notations as in Definition 2.4, and k fixed and sufficiently large. Moreover, its first principal graph is given by the inclusion graph for $\bar{H}_{2|N|k} \subset \overline{\mathrm{Br}}_{2|N|k+m}$ *for any sufficiently large k.*

Proof. We first check conditions 1–4 of Section 1B with $A_n = \bar{H}_n$ and $B_n = \overline{\mathrm{Br}}_n(N)$ for $q = e^{\pi i/\ell}$ and $1 < |N| < \ell$. Condition 1 is well-known and was checked, for instance, in [Wenzl 1988a]. Similarly, Condition 2 follows from the results in [Wenzl 2012], using the map $b \otimes g_i \in \overline{\mathrm{Br}}_m \otimes \bar{H}_n \mapsto bg_{m+i}$. Condition 3 means that the conditional expectation from $\overline{\mathrm{Br}}_{n+1}$ to $\overline{\mathrm{Br}}_n$ maps $\bar{H}_{n+1}$ onto $\bar{H}_n$. But as any element of $\bar{H}_{n+1}$ can be written as a linear combination of elements of the form ag_nb, with $a, b \in \bar{H}_n$, we have for any $c \in \overline{\mathrm{Br}}_n$ that

$$\mathrm{tr}(ag_nbc) = \mathrm{tr}(g_n)\,\mathrm{tr}(abc) = \mathrm{tr}(E_{\bar{H}_n}(ag_nb)c).$$

Hence the commuting square condition is satisfied for any four algebras of the type above. Finally, Condition 4 follows for $d = 2N$ and the projection $p = p_{[1^N]}^{\otimes 2r}$ from Lemma 2.7.

The periodicity condition for $\bar{H}_n$ was shown in [Wenzl 1988a] by proving that $\bar{p}_{[1^N]}\bar{H}_{m+N}\bar{p}_{[1^N]} \cong \bar{H}_m$, for $N > 0$. This induces an injective map

$$\tilde{\Lambda}(N, \ell)_m \to \tilde{\Lambda}(N, \ell)_{m+N}$$

by adding a column of N boxes to the given Young diagram which has to become surjective for sufficiently large m by definition of $\tilde{\Lambda}(N, \ell)$. The $2N$ periodicity for the algebras $\overline{\mathrm{Br}}_n(N)$ follows similarly using Lemma 2.7; or, see [Wenzl 2012]. The reader should have no problem adjusting this proof to the case $N < 0$ even, using Lemma 2.7. □

3. *S*-matrix

We will need certain well-known identities, which can be found in [Kac 1990], except for one case, which is a variation of the other ones. Because of this, we review the material in more detail. This might also be useful to the nonexpert reader, as the identities needed here can be derived by completely elementary methods.

3A. *Lattices.* Let $M \subset L \subset \mathbb{R}^k$ be two lattices of full rank. This means that they are isomorphic to $\mathbb{Z}^k$ as abelian groups, and each of them spans $\mathbb{R}^k$ over $\mathbb{R}$. Moreover, we assume that we have an inner product on $\mathbb{R}^k$ such that $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{Z}$ for all $\boldsymbol{x}, \boldsymbol{y} \in M$. We define the dual lattice M^* to be the set of all $\boldsymbol{y} \in \mathbb{R}^k$ such that $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{Z}$ for all $\boldsymbol{x} \in M$; the dual lattice L^* is defined similarly. Obviously $M \subset L$ implies $L^* \subset M^*$. Finally, we also assume that $A = L/M$ is a finite abelian group. Then each $\gamma \in M^*$ defines a character of A via the map $e^{\gamma} : \boldsymbol{x} \in L \mapsto e^{2\pi i(\gamma, \boldsymbol{x})}$. In particular, one can identify the group dual of A with M^*/L^*. Define the matrix

$$\tilde{S} = \frac{1}{|L:M|^{1/2}}(e^\gamma(\boldsymbol{x})),$$

where γ and $\boldsymbol{x}$ are representatives for the cosets M^*/L^* and L/M. Then $\tilde{S}$ is the character matrix of A up to a multiple and one easily concludes that it is unitary. More precisely, we can view it as a unitary operator between Hilbert spaces V and V^* with orthonormal bases labeled by the elements of L/M and M^*/L^* respectively.

3B. ***Weights of traces.*** We will primarily be interested in lattices related to root, coroot and weight lattices of orthogonal and symplectic groups. We define the lattices

$$Q = \left\{\boldsymbol{x} \in \mathbb{Z}^k, 2 \,\middle|\, \sum x_i\right\} \quad \text{and} \quad P = \mathbb{Z}^k \cup (\varepsilon + \mathbb{Z}^k), \tag{3-1}$$

where ε is the element in $\mathbb{R}^k$ with all its coordinates equal to $1/2$. Observe that $P^* = Q$ with respect to the usual scalar product of $\mathbb{R}^k$. Moreover, one can identify coroot and weight lattices of $\mathfrak{so}_{2k}$ or $\mathfrak{so}_{2k+1}$ with Q and P respectively. In particular, we define for any $\gamma \in P$ the functional $e^\gamma : \mathbb{R}^k \to \mathbb{C}$ by $e^\gamma(\boldsymbol{x}) = e^{2\pi i(\gamma,\boldsymbol{x})}$. The Weyl group of type B_k acts as usual via permutations and sign changes on the coordinates. Let $a_W = \sum_w \varepsilon(w)w$, where $\varepsilon(w)$ is the sign of the element w. Then the characters χ_λ for $\mathfrak{so}_{2k+1}$ and $\mathfrak{sp}_{2k}$ are given by $\chi_\lambda = a_w(e^{\lambda+\rho})/a_w(e^\rho)$, where $\rho = (k+1/2-i)$ for $\mathfrak{so}_{2k+1}$ and $\rho = (k+1-i)$ for $\mathfrak{sp}_{2k}$, and W is the Weyl group of type B_k.

We will also need the somewhat less familiar character formulas for the full orthogonal group $\mathrm{O}(N)$: Recall that the irreducible representations of $\mathrm{O}(N)$ are labeled by Young diagrams λ with at most N boxes in the first two columns. $\mathrm{O}(N)$-modules labeled by Young diagrams $\lambda \neq \lambda^\dagger$ restrict to isomorphic $\mathrm{SO}(N)$-modules if and only if $\lambda'_1 = N - (\lambda^\dagger)'_1$ and $\lambda'_i = (\lambda^\dagger)'_i$ for $i > 1$. Hence if $g = \exp(\boldsymbol{x})$ is an element in $\mathrm{SO}(N)$, it suffices to consider the quantities $\chi_\lambda(g) = \chi_\lambda(\boldsymbol{x})$ for λ with at most k rows for $N = 2k$ or $N = 2k+1$. We can now express the weights of Theorem 2.8 in terms of these characters; in fact the formulas in Theorem 2.8 were derived from these characters; see [Koike 1997; Wenzl 2012].

Lemma 3.1. *Let d_λ, $\tilde{d}_\lambda$ be as in Theorem 2.8 for $q = e^{\pi i/\ell}$. Moreover, we define for $|N| = 2k$ or $N = 2k+1$ the vector $\check{\rho} \in \mathbb{R}^k$ by $\check{\rho} = ((|N|+1)/2 - i)_i$. By the discussion above, it suffices to evaluate $\chi^{\mathrm{O}(N)}(\check{\rho}/\ell)$ for Young diagrams λ with $\lambda'_1 \leq N/2$, which will be assumed in the following*:

(a) *If $N = 2k+1 > 0$, then $d_\lambda = \chi_\lambda^{\mathrm{O}(N)}(\check{\rho}/\ell) = \chi_\lambda^{\mathrm{SO}(N)}(\check{\rho}/\ell)$.*

(b) *If $N = 2k > 0$ and $\lambda'_1 \leq k$, then $d_\lambda = m(\lambda)\det(\cos(l_j\check{\rho}_i)/\det(\cos(k-j)\check{\rho}_i)$, where $l_j = (\lambda+\rho)_j = \lambda_j + k - j$ and where $m(\lambda) = 2$ or 1, depending on whether λ has exactly k rows or not.*

(c) *If* $N = -2k$, *then* $d_\lambda = (-1)^{|\lambda|}\chi_{\lambda^t}^{\mathrm{Sp}(|N|)}(\check{\rho}/\ell)$ *for the symplectic character labeled by the transposed diagram* λ^T.

(d) *We have* $\tilde{d}_\lambda = \chi_\lambda^{\mathrm{SU}(N)}(\rho/\ell)$ *for* $N > 0$ *and* $\tilde{d}_\lambda = (-1)^{|\lambda|}\chi_{\lambda^T}^{\mathrm{SU}(N)}(\rho/\ell)$ *for* $N < 0$, *where* $\rho = ((|N|+1)/2 - i) \in \mathbb{R}^{|N|}$.

Proof. Observe that $\check{\rho}$ is the element ρ of the Cartan subalgebra of $\mathfrak{sl}_N$, viewed as an element of the Cartan subalgebra of the Lie subalgebra $\mathfrak{so}_N$ or $\mathfrak{sp}_N$, depending on the case. The proof now goes as the proof of Theorem 4.6 in [Wenzl 2012], which is essentially the one of [Koike 1997]. The fact that these arguments also work for the special quotients $\overline{\mathrm{Br}}_n$ follows from the proof of [Wenzl 2012, Theorem 5.5]. ☐

Remark 3.2. Let Δ_+ be the set of positive roots of a semisimple Lie algebra and $|\Delta_+|$ be its cardinality. As usual, we can express the Weyl denominator in $\chi_\lambda(\check{\rho}/\ell)$ in product form as

$$\Delta(\check{\rho}/\ell) = \prod_{\alpha>0}\left(e^{(\alpha,\check{\rho})\pi i/\ell} - e^{-(\alpha,\check{\rho})\pi i/\ell}\right) = (-i)^{|\Delta_+|}\prod_{\alpha>0} 2\sin((\alpha,\check{\rho})\pi/\ell). \tag{3-2}$$

3C. ***Usual S-matrices.*** As usual, we pick as dominant chamber C_+ the regions given by $x_1 > x_2 > \cdots > x_k > 0$ for Lie types B_k and C_k. We also choose the fundamental domains D with respect to the translation actions of M, M^*, L, L^* such that it has 0 in its center; here the lattices M and L will be certain multiples of the lattices P, Q or $\mathbb{Z}^k$ to be specified later. Let $\bar{P}_+$ be the intersection of M^* with the fundamental alcove $D \cap C_+$.

Observe that we also obtain a representation of the Weyl group W on the vector spaces V and V^*. Then it is easy to check that $a_W(V^*)$ has an orthonormal basis $|W|^{-1/2}a_w(e^\gamma)$, with $\gamma \in \bar{P}_+$, and we can define a similar basis $a_W(\boldsymbol{x})$ for $a_W(V)$. Let S be the matrix which describes the action of $\tilde{S}_{|a_W(V)}$ with respect to that basis. Then it is not hard to check (and we will do a slightly more complicated case below) that its coefficients are given by

$$s_{\gamma,\boldsymbol{x}} = \frac{1}{|L:M|^{1/2}}\sum_w \varepsilon(w)e^{2\pi i(w.\gamma,\boldsymbol{x})}. \tag{3-3}$$

If L is the weight lattice of a simple Lie algebra, the entry $s_{\gamma,\boldsymbol{x}}$ is the numerator of Weyl's character formula for the dominant weight $\lambda = \gamma - \rho$, up to the factor $|L:M|^{-1/2}$. As the columns of the unitary matrix S have norm one, it follows that

$$\sum_\lambda \chi_\lambda^2(\boldsymbol{x}) = \frac{|L:M]}{\Delta^2(\boldsymbol{x})}, \tag{3-4}$$

where Δ is the Weyl denominator, and the summation goes over the dominant weights λ such that $\lambda + \rho \in \bar{P}_+$. We are now in the position to prove some cases of the following proposition:

Proposition 3.3. *Let $\Lambda(N,\ell)_{ev}$ be the subset of $\Lambda(N,\ell)$ consisting of Young diagrams with an even number of boxes. Then we have*

$$\sum_{\lambda\in\Lambda(N,\ell)_{ev}} d_\lambda^2 = \frac{\ell^k}{b(N)} \prod_{\alpha>0} \frac{1}{4\sin^2(\alpha,\check{\rho})\pi/\ell},$$

where $\check{\rho}=((|N|+1)/2-i)$ and $\alpha>0$ runs through the positive roots of $\mathfrak{so}_N$ for $N>0$ and of $\mathfrak{sp}_{|N|}$ for $N<0$ even, and $b(N)=2$ for $N=2k>0$, and $b(N)=1$ otherwise.

Proof. Let us consider the case $N=2k+1>0$, with P and Q as in (3-1). Let $L=\ell^{-1}\mathbb{Z}^k$ and let $M_1=Q$ and $M_2=\mathbb{Z}^k$. Then we have $M_1^*=P$, and $M_2^*=\mathbb{Z}^k$. Now observe that M_1^* is the weight lattice of $\mathfrak{so}_N$, and the elements $\gamma\in\bar{P}_+$ are in 1-1 correspondence with the dominant weights λ of $\mathfrak{so}_N$ satisfying $\lambda_1\le(\ell-N)/2$, via the correspondence $\gamma=\lambda+\rho$. Moreover, $|L:M_1|=2\ell^k$. Hence it follows from (3-4) that $\sum\chi_\lambda^2(\check{\rho})=2\ell^k/\Delta^2(\check{\rho})$. Playing the same game for the lattice M_2, we now only get the sum over the characters χ_λ^2 for which $\lambda+\rho$ is in $\mathbb{Z}^k$, which is only half as large as before. Hence also the sum over the characters χ_λ^2 for which $\lambda\in\mathbb{Z}^k$ has to have the same value. This sum coincides with the right hand side of the statement for $N>0$ odd, by the restriction rules for $\mathrm{O}(N)$ to $\mathrm{SO}(N)$ (see Lemma 3.1 and its preceding discussion).

The symplectic case $N=-2k<0$ goes similarly. Here we define $M\subset L=\ell^{-1}P$, and with $L^*=\ell Q\subset M^*=\mathbb{Z}^k$. Then it follows that $\sum d_\lambda^2=2\ell^k/\Delta^2(\check{\rho}/\ell)$, where the summation goes over all diagrams λ such that $\lambda^T\in\Lambda(N,\ell)$. Playing the same game for $M=P$ and $M^*=Q$, we get $\sum d_\lambda^2=\ell^k/\Delta(\check{\rho}/\ell)$, where now the summation goes over all even, or over all odd diagrams in $\Lambda(N,\ell)$, depending on whether the sum of coordinates of $\rho=(k+1-i)$ is odd or even. In each case, we obtain that $\sum_{ev} d_\lambda^2=\ell^k/\Delta(\check{\rho}/\ell)$. We have proved the proposition except for the case $N=2k>0$, for which we need a little more preparation. □

3D. ***Another S-matrix.*** We now consider a slight generalization of the above. Observe that we can define a second sign function $\tilde{\varepsilon}$ for $W=W(B_k)$ which coincides with the usual sign function on its normal subgroup $W(D_k)$, while we have $\tilde{\varepsilon}(w)=-\varepsilon(w)$ for $w\notin W(D_k)$. It is easy to see that also in this case we have $\tilde{\varepsilon}(vw)=\tilde{\varepsilon}(v)\tilde{\varepsilon}(w)$ for all $v,w\in W$. We define $\tilde{a}_W=\sum\tilde{\varepsilon}(w)w$, and also denote the corresponding operators on the various (quotient) lattices and on the vector spaces V and V^* by the same symbol. One observes that now we get an orthonormal basis for $\tilde{a}_W(V^*)$ of the form $\boldsymbol{b}_\gamma=|\mathrm{Stab}(\gamma)|^{-1/2}|W|^{-1/2}\tilde{a}_W(e^\gamma)$, labeled by the elements of $\bar{P}_+$ which now consist of the $\gamma\in D$ such that $\gamma_1>\gamma_2>\cdots>\gamma_k\ge 0$. Observe that $|\mathrm{Stab}(\gamma)|$ is equal to 1 or 2, depending on whether $\gamma_k>0$ or $\gamma_k=0$. One similarly defines a basis for $\tilde{a}_W(V)$. Let $\boldsymbol{x}$ be such that $\mathrm{Stab}(\boldsymbol{x})=1$, that is,

$x_k > 0$, and let $b_{\boldsymbol{x}} = |W|^{-1/2}\tilde{a}_W(\boldsymbol{x})$. Then, writing $M^*/\ell L^*$ as a collection of W orbits, we obtain

$$\tilde{S}\boldsymbol{b}_{\boldsymbol{x}} = |W|^{-1/2} \sum_{\lambda\in\bar{P}_+} \sum_{v,w\in W} \frac{1}{|\mathrm{Stab}_W(\gamma)|}\,\tilde{\varepsilon}(w)\,\tilde{s}_{v.\gamma,w.\boldsymbol{x}}\,v.\gamma$$
$$= \sum_{\lambda\in\bar{P}_+} \sum_{v} \Bigl(\sum_{w} \tilde{\varepsilon}(w)\,\tilde{s}_{w.\gamma,\boldsymbol{x}} \frac{1}{|\mathrm{Stab}_W(\lambda)|}\Bigr)\tilde{\varepsilon}(v)\,v.\gamma,$$

where we replaced $\tilde{\varepsilon}(w)$ by $\tilde{\varepsilon}(v)\tilde{\varepsilon}(w^{-1}v)$, $\tilde{s}_{v.\lambda,w.\boldsymbol{x}}$ by $\tilde{s}_{w^{-1}v.\gamma,\boldsymbol{x}}$ and finally also substituted $w^{-1}v$ by w. We see from this that the coefficient of $v.\gamma$ is equal to 0 if γ has a nontrivial stabilizer except in the case when $\gamma_k = 0$. Hence it follows that $\tilde{S}$ maps $a_W(V)$ into $a_W(V^*)$. Taking bases $(\tilde{a}_W(\gamma))_{\gamma\in P_+}$ and $(\tilde{a}_W(\boldsymbol{x}))$, we see that $\tilde{S}_{|a_W(V)}$ can be described by the matrix $S = (s_{\gamma,\boldsymbol{x}})$ whose coefficients are given for $\boldsymbol{x}$ with trivial stabilizer by

$$s_{\gamma,\boldsymbol{x}} = |\mathrm{Stab}(\gamma)|^{-1/2}|L:M|^{-1/2} \sum_{w} \varepsilon(w) e^{2\pi i(w.\gamma,\boldsymbol{x})}. \tag{3-5}$$

3E. ***Squares of characters.*** Using the discussion from before and the formulas of Lemma 3.1 it is not hard to see that for N even and $\lambda_1' \le N/2$ we can write

$$\chi_\lambda^{\mathrm{O}(N)} = m(\lambda)\tilde{a}_W(e^{\lambda+\rho})/\tilde{a}_W(e^{\rho}),$$

where $m(\lambda) = 2$ or 1 depending on whether λ has exactly k rows or not. In particular, applying this to the trivial representation, we obtain $2\Delta(\rho) = \tilde{a}_W(e^{\rho})$.

Let P and Q be as in (3-1), and set $L = \ell^{-1}P$ and $M = \mathbb{Z}^k$. Then

$$L^* = \ell Q \subset M^* = \mathbb{Z}^k,$$

and it is easy to see that all of these lattices are $W = W(B_k)$-invariant. Moreover, let $\check{\rho}/\ell = (k+1/2-i)/\ell \in \ell^{-1}P = M^*$. Then it follows for $N = 2k$ and ℓ even that

$$\sum_{\lambda\in\Lambda(N,\ell)} \chi_\lambda^2(\check{\rho}\ell) = \frac{1}{\Delta^2(\check{\rho}\ell)} \sum_{\substack{\lambda_{k+1}=0\\ \lambda_1\le(\ell-N)/2}} (\tilde{a}_W(e^{\lambda+\rho})(\check{\rho})\ell)^2 = \frac{|L:M]}{2\Delta^2(\check{\rho})} \sum_{\lambda} s^2_{\lambda,\check{\rho}/\ell}\,.$$

Now observe that the matrix S is unitary and that $[L:M] = 2\ell^k$. Moreover, by Proposition 1.1 and Theorem 2.11, the square sum over odd diagrams must be equal to the square sum over even diagrams. Hence we obtain for $N > 0$ even, and ℓ even that

$$\sum_{\lambda\in\Lambda(N,\ell)_{ev}} d_\lambda^2 = \frac{\ell^k}{2\Delta^2(\check{\rho})}, \tag{3-6}$$

where $\Lambda(N,\ell)_{ev}$ denotes the set of diagrams in $\Lambda(N,\ell)$ with an even number of boxes. This finishes the last case of the proof of Proposition 3.3

3F. *Calculation of index.* As usual, identify the Cartan algebra of $\mathfrak{sl}_N$ with the diagonal $N\times N$ matrices with zero trace. The embedding of the Cartan algebras of an orthogonal or symplectic subalgebra is given via diagonal matrices for which the $(N+1-i)$-th entry is the negative of the i-th entry, for $1\le i\le N/2$. Hence, if ϵ_i is the $\mathfrak{sl}_N$ weight given by the projection onto the i-th diagonal entry, we have $(\epsilon_{N+1-i})_{|\mathfrak{so}_N}=(-\epsilon_i)_{|\mathfrak{so}_N}$, with a similar identity also holding for symplectic subalgebras. Using our description of coroot and weight lattices of orthogonal and symplectic Lie algebras as sublattices of $\mathbb{R}^k$, and defining ϕ_i to be the projection onto the i-coordinate, we see that $(\epsilon_{N+1-i})_{|\mathfrak{so}_N}=-\phi_i=(-\epsilon_i)_{|\mathfrak{so}_N}$. This allows us to describe the decomposition of $\mathfrak{sl}_N$ as both an $\mathfrak{so}_N$- and $\mathfrak{sp}_N$-module:

$$\mathfrak{sl}_N=\mathfrak{so}_N\oplus\mathfrak{p}\quad\text{and}\quad\mathfrak{sl}_N=\mathfrak{sp}_N\oplus\mathfrak{p},\tag{3-7}$$

where $\mathfrak{p}$ denotes, respectively, the nontrivial irreducible submodule in the symmetrization of the vector representation of $\mathfrak{so}_N$, and the nontrivial irreducible submodule in the antisymmetrization of the vector representation of $\mathfrak{sp}_N$. The nonzero weights $\omega>0$ of $\mathfrak{p}$ coming from positive roots of $\mathfrak{sl}_N$ and the multiplicity $n(\mathfrak{p})$ of the weight 0 in $\mathfrak{p}$ are given by

(a) $2\phi_i$, ϕ_i and $\phi_i\pm\phi_j$ for $1\le i<j\le k$ with $n(\mathfrak{p})=k$ for $\mathfrak{so}_N$ with $N=2k+1$ odd,

(b) $2\phi_i$ and $\phi_i\pm\phi_j$ for $1\le i<j\le k$ with $n(\mathfrak{p})=k-1$ for $\mathfrak{so}_N$ with $N=2k$ even,

(c) $\phi_i\pm\phi_j$ for $1\le i<j\le k$ with $n(\mathfrak{p})=k-1$ for $\mathfrak{sp}_{|N|}$ with $N=-2k<0$ even.

Theorem 3.4. *The index of the subfactor $\mathcal{N}\subset\mathcal{M}$ obtained from the inclusions of algebras $\bar{H}_n(q)\subset\overline{\mathrm{Br}}_n(q^N,q)$ is given by*

$$[\mathcal{M}:\mathcal{N}]=b(N)\ell^{n(\mathfrak{p})}\prod_{\omega>0}\frac{1}{4\sin^2(\omega,\check{\rho})\pi/\ell},$$

where the product goes over the weights $\omega>0$ of $\mathfrak{p}$ coming from positive roots of $\mathfrak{sl}_N$, as listed above, $n(\mathfrak{p})$ is the multiplicity of the zero weight in $\mathfrak{p}$, and $b(N)$ and $\check{\rho}$ are as in Proposition 3.3.

Corollary 3.5. *If $q=e^{\pi i/\ell}\to 1$, the index $[\mathcal{M}:\mathcal{N}]$ goes to ∞ with asymptotics $\ell^{\dim\mathfrak{p}}$.*

Proof. We use Theorem 2.11, where the denominator has been calculated in Proposition 3.3. The numerator follows from a standard argument for S-matrices for Lie type A; see [Kac 1990], versions of which have also been used in this section. For an elementary calculation, see [Erlijman 1998]. □

Remark 3.6. It is straightforward to adapt our index formula to subfactors related to other fixed points $H=G^\alpha$ of an order two automorphism α of a compact Lie group G, up to some integer (or perhaps rational) constant $b(H, G)$. Again, $\mathfrak{p}$ would be the -1 eigenspace of the induced action of α on the Lie algebra $\mathfrak{g}$, and the same S-matrix techniques applied in this section would go through. For example, our formulas for $N = 3$ and ℓ odd coincide with the ones at the end of [Xu 2009] for even level of SU(3), up to a factor 3 (and missing squares, a misprint according to the author). This is to be expected as in our case only those diagrams appear in the principal graph (see next section) which also label representations of the projective group PSU(3).

3G. ***Restriction rules and principal graph.*** It follows from Theorem 1.4 that the principal graph of $\mathcal{N} \subset \mathcal{M}$ is given by the inclusion matrix for $\bar{H}_{2k} \subset \overline{\mathrm{Br}}_{2k}$ for k sufficiently large. This still leaves the question of how to explicitly calculate these graphs. Observe that in the classical case $q = 1$ these would be given by the restriction rules from the unitary group $\mathrm{U}(N)$ to $\mathrm{O}(N)$, for $N > 0$. Formulas for these restriction coefficients are well-known; see for instance [Weyl 1997, Theorems 7.8F and 7.9C] and Littlewood's formula (see [Koike and Terada 1987, Section 1.5], and the whole paper for additional results). Another approach closely related to the setting of fusion categories can also be found in [Wenzl 2011].

Let $b_\mu^\lambda(N)$ be the multiplicity of the simple $\mathrm{O}(N)$-module V_μ in the $\mathrm{U}(N)$ module F_λ, for $N > 0$, where λ, μ are Young diagrams. It is well-known that for fixed Young diagrams λ and μ, the number $b_\mu^\lambda(N)$ will become a constant b_μ^λ for N large enough. Fix now also $\ell > |N|$. We define similar coefficients in our setting as follows: Recall that the simple components of $\bar{H}_n$ are labeled by the diagrams in $\tilde{\Lambda}(N, \ell)_n$ and the ones of $\overline{\mathrm{Br}}_n$ by the diagrams in $\Lambda(N, \ell)$. We then define for $\lambda \in \tilde{\Lambda}(N, \ell)$ and $\mu \in \Lambda(N, \ell)$ the number $b_\mu^\lambda(N, \ell)$ to be the multiplicity of a simple $\bar{H}_{n,\lambda}$ module in a simple $\overline{\mathrm{Br}}_{n,\mu}$ module.

In the following lemma the symbol χ_μ will also be used for the $\mathrm{O}(N)$ character corresponding to the simple representation labeled by the Young diagram μ. Moreover, we also denote by $\overline{\mathrm{Br}}_\infty$ the inductive limit of the finite dimensional algebras $\overline{\mathrm{Br}}_n$ under their standard inclusions, for fixed N and ℓ.

Lemma 3.7. (a) *Each $g \in O(N)$ for which $\chi_\mu(g) = 0$ for all boundary diagrams μ of $\Lambda(N, \ell)$ defines a trace on $\overline{\mathrm{Br}}_\infty$ determined by $\mathrm{tr}(p_\mu) = \chi_\mu(g)/\chi_{[1]}(g)^n$, where p_μ is a minimal projection of $\overline{\mathrm{Br}}_{n,\mu}$.*

(b) *For given $\lambda \in \tilde{\Lambda}(N, \ell)_n$ the coefficients $b_\mu^\lambda(N, \ell)$ are uniquely determined by the equations*

$$\chi_\lambda^{\mathrm{U}(N)}(g) = \sum_\mu b_\mu^\lambda(N, \ell)\chi_\mu(g)$$

for all g as in (a), where the summation goes over all diagrams μ in $\Lambda(N, \ell)$ with $n, n-2, \ldots$ boxes.

Proof. The formula in statement (a) determines a trace on $\overline{\mathrm{Br}}_n$ for each n. To show that these formulas are compatible with the standard embeddings we observe that a minimal idempotent $p_\mu \in \overline{\mathrm{Br}}_{n,\mu}$ is the sum of minimal idempotents $e_\lambda \in \overline{\mathrm{Br}}_{n+1,\lambda}$ where λ runs through all diagrams in $\Lambda(N,\ell)$ obtained by adding or removing a box to/from λ; see (2-2) and the remarks below that theorem. Evaluating the traces of these idempotents and multiplying everything by $\chi_\lambda(g)^{n+1}$, equality of the traces is equivalent to

$$\chi_\mu(g)\chi_{[1]}(g) = \sum_\lambda \chi_\lambda(g).$$

By the usual tensor product rule for orthogonal groups, the left hand side would be equal to the sum of characters corresponding to *all* diagrams λ which differ from μ by only one box. It is easy to check that this differs from the sum above only by boundary diagrams, for which the characters at g are equal to 0. This shows (a).

For (b), we first show that

$$\mathrm{tr}(p_\lambda) = \frac{\chi_\lambda^{\mathrm{U}(N)}(g)}{\chi_{[1]}^{\mathrm{U}(N)}(g)^n}$$

for $p_\lambda \in \bar{H}_{n,\lambda}$ a minimal idempotent and tr a trace as in (a). As the weight vector for $\overline{\mathrm{Br}}_{n+2N}$ is a multiple of that of $\overline{\mathrm{Br}}_n$, for n large enough, the same must also hold for the weight vectors of $\bar{H}_n$ and $\bar{H}_{n+2N}$, by periodicity of the inclusions. Hence, these weight vectors must be eigenvectors of the inclusion matrix for $\bar{H}_n \subset \bar{H}_{n+2N}$. As this inclusion matrix is just a block of the $2N$-th power of the fusion matrix of the vector representation for the corresponding type A fusion category, its entries must be given by $\mathrm{U}(N)$ characters of a suitable group element. To identify these elements, it suffices to observe that the antisymmetrizations of the vector representation, labeled by the Young diagrams $\lambda = [1^j]$, $1 \le j \le N$, remain irreducible as $\mathrm{O}(N)$ modules. This means the corresponding Hecke algebra idempotent remains a minimal idempotent also in $\overline{\mathrm{Br}}_j$. Hence $\mathrm{tr}(p_\lambda) = \chi_\lambda^{\mathrm{U}(N)}(g)$ for $\lambda = [1^j]$ and $1 \le j \le N$. But as the antisymmetrizations generate the representation ring of $\mathrm{U}(N)$, and also of the corresponding fusion ring, the claim follows for general λ. For more details, see [Goodman and Wenzl 1990].

Recall that the coefficient $b_\mu^\lambda(N,\ell)$ can be defined as the rank of p_λ in an irreducible $\overline{\mathrm{Br}}_{n,\mu}$ representation. So obviously the formula in the statement holds for any g as in (a). Examples for such g come from $\exp(\boldsymbol{x})$ with $\boldsymbol{x} \in M^* = \ell^{-1}Q$ for which the character is given by the expression $\chi_\lambda(\boldsymbol{x})$ as in Section 2E. As the columns of the orthogonal S-matrix are linearly independent, this would identify $\mathrm{SO}(N)$ representations. If N is odd, the two $\mathrm{O}(N)$ representations which reduce to the same $\mathrm{SO}(N)$ representation are labeled by Young diagrams with opposite parities. Hence only one of them can occur in the decomposition of a given U(N) representation. A similar argument also works in the symplectic case.

For N even, we can have two diagrams λ and $\lambda^\dagger$ with the same SO(N) character, where one of them, say λ, has less than k rows. They can be distinguished by elements $g \in O(N)\setminus \mathrm{SO}(N)$ for which $\chi_{\lambda^\dagger}(g) = -\chi_\lambda(g)$. It is well-known that such elements g must have eigenvalues ± 1, and $\chi_\lambda(g)$ is given by the character formula for Sp($2k-2$) in the remaining $2k-2$ eigenvalues; see [Weyl 1997]. It follows from the invertibility of the S-matrix for Sp($2k-2$) at level $\ell/2-k$ (see [Kac 1990]) that we can identify those diagrams λ by evaluating $\chi_\lambda^{\mathrm{Sp}(2k-2)}(\boldsymbol{x}/\ell)$ for $\boldsymbol{x} \in \mathbb{Z}^{k-1}$ with $\ell/2 > x_1 > x_2 > \cdots > x_{k-1} > 0$, and that those elements satisfy the boundary condition $\chi_\lambda^{\mathrm{Sp}(2k-2)}(\boldsymbol{x}) = 0$ for any boundary diagram λ. □

The lemma above is illustrated in the following section for a number of explicit examples. We can also give a closed formula for the restriction coefficients, using a well-known quotient map for fusion rings (even though in our case, the quotient ring does not correspond to a tensor category as far as we know). In the context of fusion rings, this is known as the Kac–Walton formula; for type A see also [Goodman and Wenzl 1990]. In our case, we need to use a slightly different affine reflection group $\mathcal{W}$. In the orthogonal case $N = 2k$ and $N = 2k+1$ it is given by the semidirect product of $\ell\mathbb{Z}^k$ with the Weyl group of type B_k. In the symplectic case, it is given by the semidirect product of ℓQ with the Weyl group of type B_k. As usual, we define the dot action of $\mathcal{W}$ on $\mathbb{R}^k$ by $w.\boldsymbol{x} = w(\boldsymbol{x}+\rho)-\rho$, where ρ is half the sum of the positive roots of the corresponding Lie algebra, with the roots embedded into $\mathbb{R}^k$ as described above, and ε is the usual sign function for reflection groups. This can be extended to an action on the labeling set of O(N) representations by identifying a Young diagrams with at most k rows with the corresponding vector in $\mathbb{Z}^k$, and by using the restriction rule from O(N) to SO(N) in the other cases. See also [Wenzl 2011, Lemma 1.7] for more details.

Theorem 3.8. *With notations as above, the restriction multiplicity* $b_\mu^\lambda(N,\ell)$ *for* $N = 2k+1 > 0$ *and* $N = -2k$ *is given by*

$$b_\mu^\lambda(N,\ell) = \sum_{w\in\mathcal{W}} \varepsilon(w) b_{w.\mu}^\lambda(N).$$

If $N = 2k > 0$, *we have to replace* ε *by* $\tilde{\varepsilon}$ (*see Section 3D*) *in the formula above.*

Proof. Looking at the character formulas, we see that an action of an element w of the finite reflection group on λ just changes the character by the sign of w. Moreover, by definition of the elements $\boldsymbol{x}$ we have that $\chi_\lambda(\boldsymbol{x}) = \chi_{\lambda+\mu}(\boldsymbol{x})$ for any $\mu \in M$. It follows that $\chi_{w.\lambda}(\boldsymbol{x}) = \varepsilon(w)\chi_\lambda(\boldsymbol{x})$ for all $\boldsymbol{x} \in M^*$ and $w \in \mathcal{W}$. Hence summing over the $\mathcal{W}$-orbits, we obtain for any $\boldsymbol{x} \in M^*$, $\lambda \in \tilde{\Lambda}(N, \ell)$ and $\mu \in \Lambda(N, \ell)$ that

$$\chi_\lambda^{\mathrm{U}(N)}(\boldsymbol{x}) = \sum_\gamma b_\gamma^\lambda(N)\chi_\gamma = \sum_\mu \Big(\sum_w b_{w.\mu}^\lambda(N) \Big) \chi_\mu .$$

The claim now follows from this and Lemma 3.7. □

4. Examples and other approaches

4A. ***The case $N = 2$.*** This corresponds to the Goodman–de la Harpe–Jones subfactors for type $D_{\ell/2+1}$, where $\ell > 2$ has to be even. It follows from our theorem that the even vertices of the principal graph are labeled by the Young diagrams λ with an even number n of boxes, at most two rows and with $\lambda_1 - \lambda_2 \le \ell - 2$; there are $(\ell - 2)/2$ such diagrams. Their dimensions are given by $\tilde{d}_k = [2k + 1]$, for $0 \le k < (\ell - 2)/2$.

Moreover, one checks that $\Lambda(2, \ell)$ consists of Young diagrams $[j]$ such that $0 \le j \le (\ell/2) - 1$ and of $[1^2]$, one column with 2 boxes, with dimensions $d_{[j]} = 2 \cos j\pi/\ell$ for $j > 0$ and dimension equal to 1 for the remaining cases (that is, for $\varnothing$ and for $[1^2]$). The restriction rule (that is, principal graph) follows from writing the dimensions as

$$\tilde{d}_k = 2 \cos \tilde{k}\pi/\ell + 2 \cos(\tilde{k} - 2)\pi/\ell + \cdots + 1,$$

where $\tilde{k} = \min\{k, (\ell/2) - k\}$. Indeed, this determines the graph completely except for whether to pick the diagram $\varnothing$ or $[1^2]$ for the object with dimension 1. It follows from the restriction rule $\mathrm{O}(2) \subset U(2)$ that we take $\varnothing$ for j even, and $[1^2]$ for j odd. To calculate the index one can check by elementary means that $\sum_{\lambda \text{ even}} d_\lambda^2 = \ell/2$. Moreover, it is well-known that the sum $\sum_{\lambda \text{ even}} \tilde{d}_\lambda^2$ over even partitions for $\mathfrak{sl}_2$ is equal to $\ell/(4 \sin^2 \pi/\ell)$. Hence we obtain as index $[\mathcal{M} : \mathcal{N}] = 1/(2 \sin^2 \pi/\ell)$.

4B. ***The case $N = 3$.*** It is also fairly elementary to work out this case in detail. A detailed discussion of SU(3) fusion modular categories has already been given in [Ocneanu 2002] (without proofs) and in [Evans and Pugh 2011] and references therein. These include our examples here. Recall that by Weyl's dimension formula we have

$$\tilde{d}_\lambda = \frac{[\lambda_1 - \lambda_2 + 1][\lambda_2 - \lambda_3 + 1][\lambda_1 - \lambda_3 + 2]}{[1]^2[2]} .$$

Now observe that the product of two q-numbers is given by the tensor product rules for $\mathfrak{sl}_2$, that is, we have for $n \ge m$ that

$$[n][m] = [n + m - 1] + [n + m - 3] + \cdots + [n - m + 1].$$

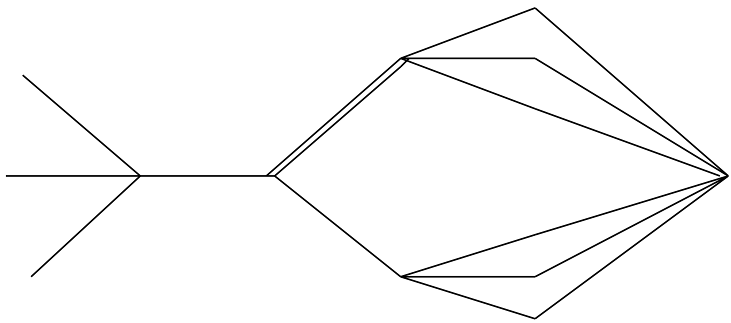


Figure 1. The principal graph of SO(3) for $\ell = 9$.

As an example, we have

$$\tilde{d}_{[4]} = \frac{[6][5]}{[2]} = \frac{[10]+[8]+[6]+[4]+[2]}{[2]} = [9]+[5]+[1],$$

that is, the fourth antisymmetrization of the vector representation of U(3) decomposes as a direct sum of the one-, five- and nine-dimensional representation of SO(3). One similarly can show the well-known result that the adjoint representation of SU(3), labeled by the Young diagram [2, 1] decomposes into the direct sum of the three- and the five-dimensional representation of SO(3), that is, $\mathfrak{p}$ is the five dimensional representation of SO(3). Hence we get from Theorem 3.4 that the index is equal to

$$[\mathcal{M} : \mathcal{N}] = \frac{\ell}{4^2 \sin^2(2\pi/\ell) \sin^2(\pi/\ell)} .$$

We note that here as well as in the other examples, the dimensions (that is, entries of the Perron–Frobenius vectors) are given by $|\tilde{d}_\lambda|$ for even vertices, and by $\sqrt{[\mathcal{M} : \mathcal{N}]}\,|d_\mu|$ for odd vertices, with $\tilde{d}_\lambda$ and d_μ as in Lemma 3.1. To consider explicit examples, the first nontrivial case for $N = 3$ occurs for $\ell = 7$. We leave it to the reader to check that in this case the first principal graph is given by the Dynkin graph D_8. A more interesting graph is obtained for $\ell = 9$; see Figure 1. Here we have the three invertible objects of the $SU(3)_6$ fusion category, including the trivial object (often denoted as $*$) on the left; they generate a group isomorphic to $\mathbb{Z}/3$. The vertices with the double edge are labeled by the object corresponding to the 5-dimensional representation of SO(3) and the diagram [4, 2] for $SU(3)_6$. This is the only fixed point under the $\mathbb{Z}/3$ action given by the invertible objects (or, in physics language, the currents). It would be interesting to see whether one can carry out an orbifold construction in this context related to the one in [Evans and Kawahigashi 1994].

4C. ***The case $N = 4$.*** The combinatorics of these subfactors has already appeared in [Ocneanu 2002; Evans and Gannon 2010] and also in the mathematical physics literature (see below), but the author is not aware of a rigorous general construction of the subfactors in the literature (but see the remarks below about the work in [Wassermann 2010; Xu 2009]). As we shall see, somewhat surprisingly, the corresponding construction for SO(4) does not seem to work. We do the case with $\ell = 8$ in explicit detail. It is not hard to check that we already get the periodic inclusion matrix for $n = 12$. As we consider an analog of the restriction to O(4) for which the determinant can be ± 1, we should, strictly speaking, consider a fusion category for $SU(4) \times \{\pm 1\}$. We shall actually use the Young diagram notation for representations of U(4). For $n = 12$ we have the invertible objects labeled by $[3^4]$, $[4^3]$, $[5^2 1^2]$ and $[62^3]$ (that is, the last diagram, for instance, has six boxes in the first and two boxes each in the second, third and fourth rows). They generate a subgroup isomorphic to $\mathbb{Z}/4$. It follows from the O(4) restriction rules that $[3^4]$ and $[5^2 1^2]$ contain the determinant representation, and $[4^3]$ and $[62^3]$ contain the trivial representation as one-dimensional O(4) subrepresentations. This allows us to calculate the restrictions for representations of each $\mathbb{Z}/4$ orbit simultaneously. As usually for at least one element of each orbit the ordinary restriction rules still hold, it makes the general calculations easier. The principal graph can be seen in Figure 2. As in the $N = 3$ example, the one-dimensional currents, including the trivial object $*$ appear as the left- and right-most vertices in the graph. The lowest vertex corresponds to the O(4)-object [2] which is connected to the objects in the $\mathbb{Z}/4$-orbit $\{[2, 1^2], [3, 1], [4, 3, 1], [3, 3, 2]\}$. We also note that we get the same graph for the Sp(4) case $N = -4$ for $\ell = 8$. However, for other roots of unity, already the indices of the subfactors differ, being given by

$$O(4) : \frac{2\ell}{4\sin^2(3\pi/\ell)4\sin^2(2\pi/\ell)16\sin^4(\pi/\ell)} \qquad Sp(4) : \frac{\ell}{4\sin^2(2\pi/\ell)4\sin^2(\pi/\ell)}.$$

It was originally thought that we should also be able to get fusion category analogs for the restriction from $SU(N)$ to $SO(N)$ for N even. It is easy to check that this is not possible for O(2). Some initial checks also seem to suggest a similar phenomenon for higher ranks. For example, using the same element $\check{\rho}$ in the $SO(N)$ character formula would give dimension functions which are not invariant under the D_N diagram automorphism.

4D. ***Related results.*** We discuss several results related to our findings. Our original motivation was to construct subfactors related to twisted loop groups. It was shown in R. Verrill's PhD thesis [2001] that it is not possible to construct a fusion tensor product for representations of twisted loop groups. However, it seemed reasonable to expect that representations of twisted loop groups could become a module category over representations of their untwisted counterparts. Many results, in

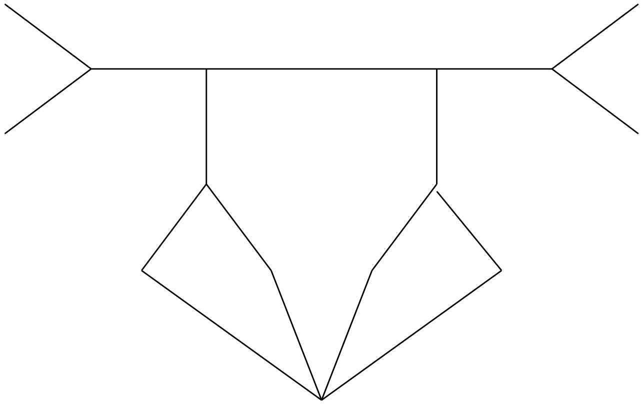


Figure 2. The principal graph of O(4) and Sp(4) for $\ell = 8$.

particular about the combinatorics of such categories, can be found in the context of boundary conformal field theory in papers by Evans, Gaberdiel, Gannon, Fuchs, Pugh, Schweigert, Di Francesco, Petkova, Zuber and others; see for instance [Evans and Gannon 2010; Gaberdiel and Gannon 2002; Fuchs and Schweigert 2000; Petkova and Zuber 2002] and the papers cited therein. Understanding results in these papers in mathematical terms was one of the motivations for this author. Similarly, mathematical results in [Ocneanu 2002; Evans and Pugh 2011] for the cases $N = 3$ and $N = 4$ (see the introduction) were influenced by these papers, in particular by work of Zuber and his coauthors.

In the mathematics literature, one can find closely related results in [Xu 2009; Wassermann 2010]. Here the authors construct module categories via a completely different approach in the context of type III_1 factors, using loop groups. For instance, the formulas at the end of [Xu 2009] for the special case $N = 3$ differ only by a factor 3 (which can be explained; see Remark 3.6), by our formulas for $N = 3$ for even level (together with Corollary 1.5), modulo misprints. Similar formulas for the symplectic case as well as restriction coefficients also appear at the end of [Wassermann 2010]. We cannot get results corresponding to the odd level cases in [Xu 2009]. The combinatorics there suggests that this would require considering an embedding of $\mathrm{Sp}(N-1)$ into $\mathrm{SU}(N)$ under which the vector representation would not remain irreducible. In contrast, we can also construct module categories for $\ell - N$ odd, which would correspond to odd level; however, these categories are not unitarizable (which follows from Lemma 2.9) and they have different fusion rules. However, we do get fairly general formulas for the index and principal graphs of this type of subfactors in the unitary case, which was one of the problems posed in [Xu 2009]. These formulas were known to this author as well as to Antony

Wassermann at least back in 2008 when they had discussions about their respective works in Oberwolfach and at the Schrödinger Institute.

We close this section by mentioning that while our results for $N > 0$ odd and $N < 0$ even are in many ways parallel to results obtained via other approaches in connection with twisted loop groups, there does not seem to be an obvious analog for our results for $N > 0$ even. For instance, the combinatorial results in [Gaberdiel and Gannon 2002] for that case seem to be different to ours.

4E. ***Conclusions and further explorations.*** We have constructed module categories of fusion categories of type A via deformations of centralizer algebras of certain subgroups of unitary groups. We have also classified when they are unitarizable, we have constructed the corresponding subfactors, and we have explicitly calculated their indices and first principal graphs. These deformations are compatible with the Drinfeld–Jimbo deformation of the unitary group but not with the Drinfeld–Jimbo deformation of the subgroup. Most of the deformation was already done in [Wenzl 2012] via elementary methods. In principle, at least, it should be possible to use this elementary approach also for other inclusions. However, this might become increasingly tedious.

As we have seen already in Section 2B, it should be possible to get a somewhat more conceptual approach using different deformations of the subgroup; see [Noumi 1996; Molev 2003; Letzter 1997; Letzter 2002; Iorgov and Klimyk 2005] and references therein. In particular in the work of Letzter, such deformations via coideal algebras have been defined for a large class of embeddings of a semisimple Lie algebra into another one. At this point, it does not seem obvious how to define C^*-structures in this setting, and additional complications arise as these coideal algebras are not expected to be semisimple at roots of unity. Nevertheless, the results in this and other papers such as [Xu 2009; Wassermann 2010] would seem to suggest that similar constructions might be possible also in a more general setting.

Acknowledgments

It is a pleasure to thank Antony Wassermann, David Jordan, Viktor Ostrik, Feng Xu and David Evans for useful references, and Fred Goodman also for technical advice.

References

[Bisch 1997] D. Bisch, "Bimodules, higher relative commutants and the fusion algebra associated to a subfactor", pp. 13–63 in *Operator algebras and their applications* (Waterloo, ON, 1994/1995)), edited by P. A. Fillmore and J. A. Mingo, Fields Inst. Commun. **13**, Amer. Math. Soc., Providence, RI, 1997. MR 97i:46109 Zbl 0894.46046

[Böckenhauer et al. 1999] J. Böckenhauer, D. E. Evans, and Y. Kawahigashi, "On α-induction, chiral generators and modular invariants for subfactors", *Comm. Math. Phys.* **208**:2 (1999), 429–487. MR 2001c:81180 Zbl 0948.46048

[Erlijman 1998] J. Erlijman, "New subfactors from braid group representations", *Trans. Amer. Math. Soc.* **350**:1 (1998), 185–211. MR 98k:46100 Zbl 0887.46035

[Erlijman and Wenzl 2007] J. Erlijman and H. Wenzl, "Subfactors from braided C^* tensor categories", *Pacific J. Math.* **231**:2 (2007), 361–399. MR 2008g:46111 Zbl 1152.46051

[Evans and Gannon 2010] D. E. Evans and T. Gannon, "Modular invariants and twisted equivariant K-theory II: Dynkin diagram symmetries", preprint, 2010. arXiv 1012.1634

[Evans and Kawahigashi 1994] D. E. Evans and Y. Kawahigashi, "Orbifold subfactors from Hecke algebras", *Comm. Math. Phys.* **165**:3 (1994), 445–484. MR 96c:46059 Zbl 0805.46077

[Evans and Kawahigashi 1998] D. E. Evans and Y. Kawahigashi, *Quantum symmetries on operator algebras*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 1998. MR 99m:46148 Zbl 0924.46054

[Evans and Pugh 2011] D. E. Evans and M. Pugh, "Pugh, Mathew, Braided subfactors, spectral measures, planar algebras and Calabi–Yau algebras associated to SU(3) modular invariants", preprint, 2011. arXiv 1110.4547

[Fuchs and Schweigert 2000] J. Fuchs and C. Schweigert, "Solitonic sectors, α-induction and symmetry breaking boundaries", *Phys. Lett. B* **490** (2000), 163–172. MR 2001k:81222 Zbl 1031.81605

[Gaberdiel and Gannon 2002] M. R. Gaberdiel and T. Gannon, "Boundary states for WZW models", *Nuclear Phys. B* **639**:3 (2002), 471–501. MR 2003j:81192 Zbl 0997.81052

[Goodman and Wenzl 1990] F. M. Goodman and H. Wenzl, "Littlewood–Richardson coefficients for Hecke algebras at roots of unity", *Adv. Math.* **82**:2 (1990), 244–265. MR 91i:20013 Zbl 0714.20004

[Goodman et al. 1989] F. M. Goodman, P. de la Harpe, and V. F. R. Jones, *Coxeter graphs and towers of algebras*, Mathematical Sciences Research Institute Publications **14**, Springer, New York, 1989. MR 91c:46082 Zbl 0698.46050

[Iorgov and Klimyk 2005] N. Z. Iorgov and A. U. Klimyk, "Classification theorem on irreducible representations of the q-deformed algebra $U'_q(\mathrm{so}_n)$", *Int. J. Math. Math. Sci.* 2 (2005), 225–262. MR 2006c:17021 Zbl 1127.17016

[Jones 1983] V. F. R. Jones, "Index for subfactors", *Invent. Math.* **72**:1 (1983), 1–25. MR 84d:46097 Zbl 0508.46040

[Kac 1990] V. G. Kac, *Infinite-dimensional Lie algebras*, 3rd ed., Cambridge University Press, 1990. MR 92k:17038 Zbl 0716.17022

[Koike 1997] K. Koike, "Principal specializations of the classical groups and q-analogs of the dimension formulas", *Adv. Math.* **125**:2 (1997), 236–274. MR 98f:20027 Zbl 0867.22013

[Koike and Terada 1987] K. Koike and I. Terada, "Young-diagrammatic methods for the representation theory of the classical groups of type B_n, C_n, D_n", *J. Algebra* **107**:2 (1987), 466–511. MR 88i:22035 Zbl 0622.20033

[Letzter 1997] G. Letzter, "Subalgebras which appear in quantum Iwasawa decompositions", *Canad. J. Math.* **49**:6 (1997), 1206–1223. MR 99g:17022 Zbl 0898.17005

[Letzter 2002] G. Letzter, "Coideal subalgebras and quantum symmetric pairs", pp. 117–165 in *New directions in Hopf algebras*, edited by S. Montgomery and H.-J. Schneider, Math. Sci. Res. Inst. Publ. **43**, Cambridge Univ. Press, 2002. MR 2003g:17025 Zbl 1025.17005

[Molev 2003] A. I. Molev, "A new quantum analog of the Brauer algebra: Quantum groups and integrable systems", *Czechoslovak J. Phys.* **53**:11 (2003), 1073–1078. MR 2005e:17024

[Noumi 1996] M. Noumi, "Macdonald's symmetric polynomials as zonal spherical functions on some quantum homogeneous spaces", *Adv. Math.* **123**:1 (1996), 16–77. MR 98a:33004 Zbl 0874.33011

[Ocneanu 2002] A. Ocneanu, "The classification of subgroups of quantum SU(N)", pp. 133–159 in *Quantum symmetries in theoretical physics and mathematics* (Bariloche, 2000), edited by R. Coquereaux et al., Contemp. Math. **294**, Amer. Math. Soc., Providence, RI, 2002. MR 2003h:81101 Zbl 1193.81055

[Petkova and Zuber 2002] V. Petkova and J.-B. Zuber, "Conformal field theories, graphs and quantum algebras", pp. 415–435 in *MathPhys odyssey, 2001*, edited by M. Kashiwara and T. Miwa, Prog. Math. Phys. **23**, Birkhäuser, Boston, MA, 2002. MR 1903984 Zbl 1026.81053

[Ram and Wenzl 1992] A. Ram and H. Wenzl, "Matrix units for centralizer algebras", *J. Algebra* **145**:2 (1992), 378–395. MR 93g:16024 Zbl 0761.16008

[Sunder 1987] V. S. Sunder, "A model for AF algebras and a representation of the Jones projections", *J. Operator Theory* **18**:2 (1987), 289–301. MR 89e:46079 Zbl 0693.46054

[Verrill 2001] R. Verrill, *Positive energy representations of L^σ $SU(2r)$ and orbifold fusion*, thesis, University of Cambridge, 2001.

[Wassermann 2010] A. Wassermann, "Subfactors and Connes fusion for twisted loop groups", preprint, 2010. arXiv 1003.2292

[Wenzl 1988a] H. Wenzl, "Hecke algebras of type A_n and subfactors", *Invent. Math.* **92**:2 (1988), 349–383. MR 90b:46118 Zbl 0663.46055

[Wenzl 1988b] H. Wenzl, "On the structure of Brauer's centralizer algebras", *Ann. of Math.* (2) **128**:1 (1988), 173–193. MR 89h:20059 Zbl 0656.20040

[Wenzl 2011] H. Wenzl, "Quotients of representation rings", *Represent. Theory* **15** (2011), 385–406. MR 2012i:22019 Zbl 06043183

[Wenzl 2012] H. Wenzl, "A q-Brauer algebra", *J. Algebra* **358** (2012), 102–127. MR 2905021

[Weyl 1997] H. Weyl, *The classical groups: Their invariants and representations*, Princeton Landmarks in Mathematics, Princeton University Press, 1997. MR 98k:01049 Zbl 1024.20501

[Xu 1998] F. Xu, "New braided endomorphisms from conformal inclusions", *Comm. Math. Phys.* **192**:2 (1998), 349–403. MR 99b:81139 Zbl 0908.46044

[Xu 2009] F. Xu, "On affine orbifold nets associated with outer automorphisms", *Comm. Math. Phys.* **291**:3 (2009), 845–861. MR 2011d:46141 Zbl 1186.81108

Received July 19, 2011.

HANS WENZL
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, SAN DIEGO
9500 GILMAN DRIVE
DEPT 0112
LA JOLLA CA 92093-0112
UNITED STATES

hwenzl@ucsd.edu

CONTENTS

Volume 259, no. 1 and no. 2

Gabriel **Acosta** and Ignacio Ojea: *Extension theorems for external cusps with minimal regularity* 1

Christos A. **Athanasiadis**: *Flag subdivisions and γ-vectors* 257

Maria **Athanassenas** and Sevvandi Kandanaarachchi: *Convergence of axially symmetric volume-preserving mean curvature flow* 41

Igor **Belegradek**, Eric Choi and Nobuhiro Innami: *Rays and souls in von Mangoldt planes* 279

Eric **Choi** with Igor Belegradek and Nobuhiro Innami 279

Françoise **Dal'bo**, Marc Peigné and Andrea Sambusetti: *On the horoboundary and the geometry of rays of negatively curved manifolds* 55

Abraham **Frandsen**, Donald Sampson and Neil Steinburg: *Isoperimetric surfaces with boundary, II* 307

Kei **Funano**: *Two infinite versions of the nonlinear Dvoretzky theorem* 101

Nobuhiro **Innami** with Igor Belegradek and Eric Choi 279

Alexander J. **Izzo**: *Nonlocal uniform algebras on three-manifolds* 109

Yuichi **Kabaya**: *Cyclic branched coverings of knots and quandle homology* 315

Sevvandi **Kandanaarachchi** with Maria Athanassenas 41

John S. **Kauta**: *On a class of semihereditary crossed-product orders* 349

Demetre **Kazaras** and Ivan Sterling: *An explicit formula for spherical curves with constant torsion* 361

Patrick W. **Keef**: *Mahlo cardinals and the torsion product of primary abelian groups* 117

Jean-François **Lafont** and Christophe Pittet: *Comparing seminorms on homology* 373

Nan **Li** and Xiaochun Rong: *Relatively maximum volume rigidity in Alexandrov geometry* 387

Antonios **Manoussos** and Polychronis Strantzalos: *Properness, Cauchy indivisibility and the Weil completion of a group of isometries* 421

Ivan **Matić**: *Theta lifts of strongly positive discrete series: the case of* $(\widetilde{\mathrm{Sp}}(n), O(V))$ 445

Aaron **Melman**: *Geometry of trinomials* 141

Ignacio **Ojea** with Gabriel Acosta 1

Marc **Peigné** with Françoise Dal'bo and Andrea Sambusetti 55

Christophe **Pittet** with Jean-François Lafont 373

Matt **Rathbun**: *Tunnel one, fibered links* 473

Xiaochun **Rong** with Nan Li 387

Andrea **Sambusetti** with Françoise Dal'bo and Marc Peigné 55

Donald **Sampson** with Abraham Frandsen and Neil Steinburg 307

Anne V. **Shepler** and Sarah Witherspoon: *Drinfeld orbifold algebras* 161

Neil **Steinburg** with Abraham Frandsen and Donald Sampson 307

Ivan **Sterling** with Demetre Kazaras 361

Polychronis **Strantzalos** with Antonios Manoussos 421

Jyh-Haur **Teh**: *Semi-topological cycle theory I* 195

Hans **Wenzl**: *Fusion symmetric spaces and subfactors* 483

Sarah **Witherspoon** with Anne V. Shepler 161

Tiehong **Zhao**: *New construction of fundamental domains for certain Mostow groups* 209

Guidelines for Authors

Authors may submit manuscripts at msp.berkeley.edu/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 259 No. 2 October 2012

Flag subdivisions and γ-vectors 257
CHRISTOS A. ATHANASIADIS

Rays and souls in von Mangoldt planes 279
IGOR BELEGRADEK, ERIC CHOI and NOBUHIRO INNAMI

Isoperimetric surfaces with boundary, II 307
ABRAHAM FRANDSEN, DONALD SAMPSON and NEIL STEINBURG

Cyclic branched coverings of knots and quandle homology 315
YUICHI KABAYA

On a class of semihereditary crossed-product orders 349
JOHN S. KAUTA

An explicit formula for spherical curves with constant torsion 361
DEMETRE KAZARAS and IVAN STERLING

Comparing seminorms on homology 373
JEAN-FRANÇOIS LAFONT and CHRISTOPHE PITTET

Relatively maximum volume rigidity in Alexandrov geometry 387
NAN LI and XIAOCHUN RONG

Properness, Cauchy indivisibility and the Weil completion of a group of isometries 421
ANTONIOS MANOUSSOS and POLYCHRONIS STRANTZALOS

Theta lifts of strongly positive discrete series: the case of $(\widetilde{\mathrm{Sp}}(n), O(V))$ 445
IVAN MATIĆ

Tunnel one, fibered links 473
MATT RATHBUN

Fusion symmetric spaces and subfactors 483
HANS WENZL