Pacific Journal of Mathematics

Volume 278 No. 1 November 2015

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

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

EDITORS

Don Blasius (Managing Editor)
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
blasius@math.ucla.edu

Paul Balmer
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
balmer@math.ucla.edu

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

Daryl Cooper
Department of Mathematics
University of California
Santa Barbara, CA 93106-3080
cooper@math.ucsb.edu

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

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

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

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

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

Paul Yang
Department of Mathematics
Princeton University
Princeton NJ 08544-1000
yang@math.princeton.edu

PRODUCTION

Silvio Levy, Scientific Editor, production@msp.org

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 msp.org/pjm for submission instructions.

The subscription price for 2015 is US $420/year for the electronic version, and $570/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/
© 2015 Mathematical Sciences Publishers

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.278.1

GROWTH TIGHT ACTIONS

GOULNARA N. ARZHANTSEVA, CHRISTOPHER H. CASHEN AND JING TAO

We introduce and systematically study the concept of a growth tight action. This generalizes growth tightness for word metrics as initiated by Grigorchuk and de la Harpe. Given a finitely generated, nonelementary group G acting on a G-space $\mathcal{X}$, we prove that if G contains a strongly contracting element and if G is not too badly distorted in $\mathcal{X}$, then the action of G on $\mathcal{X}$ is a growth tight action. It follows that if $\mathcal{X}$ is a cocompact, relatively hyperbolic G-space, then the action of G on $\mathcal{X}$ is a growth tight action. This generalizes all previously known results for growth tightness of cocompact actions: every already known example of a group that admits a growth tight action and has some infinite, infinite index normal subgroups is relatively hyperbolic; conversely, relatively hyperbolic groups admit growth tight actions. This also allows us to prove that many CAT(0) groups, including flip-graph manifold groups and many right angled Artin groups, and snowflake groups admit cocompact, growth tight actions. These provide first examples of non relatively hyperbolic groups admitting interesting growth tight actions. Our main result applies as well to cusp uniform actions on hyperbolic spaces and to the action of the mapping class group on Teichmüller space with the Teichmüller metric. Towards the proof of our main result, we give equivalent characterizations of strongly contracting elements and produce new examples of group actions with strongly contracting elements.

0. Introduction 2

Part I. Growth tight actions 8
1. Preliminaries 8
2. Contraction and constriction 12
3. Abundance of strongly contracting elements 20
4. A minimal section 25
5. Embedding a free product set 26
6. The growth gap 27
7. The growth of conjugacy classes 29

MSC2010: 20E06, 20F65, 20F67, 37C35, 57MXX.

Keywords: growth tight, Hopf property, entropy, relatively hyperbolic groups, mapping class groups, snowflake groups, rank 1 isometries, CAT(0) groups.

Part II. Examples with strongly contracting elements 31
8. Actions on relatively hyperbolic spaces 31
9. Rank 1 actions on CAT(0) spaces 36
10. Mapping class groups 38
11. Snowflake groups 40
Acknowledgments 45
References 45

0. Introduction

The growth exponent of a set $\mathcal{A} \subset \mathcal{X}$ with respect to a pseudometric d is

$$\delta_{\mathcal{A},d} := \limsup_{r\to\infty} \frac{\log \#\{a \in \mathcal{A} \mid d(o,a) \leqslant r\}}{r}$$

where # denotes cardinality and $o \in \mathcal{X}$ is some basepoint. The limit is independent of the choice of basepoint.

Let G be a finitely generated group. A left invariant pseudometric d on G induces a left invariant pseudometric $\bar{d}$ on any quotient G/Γ of G by the formula $\bar{d}(g\Gamma, g'\Gamma) := d(g\Gamma, g'\Gamma)$.

Definition 0.1. *G is growth tight with respect to d if $\delta_{G,d} > \delta_{G/\Gamma,\bar{d}}$ for every infinite normal subgroup $\Gamma \trianglelefteq G$.*

One natural way to put a left invariant metric on a finitely generated group is to choose a finite generating set and consider the word metric. More generally, pseudometrics on a group are provided by actions of the group on metric spaces. Let $\mathcal{X}$ be a G-space, that is, a proper, geodesic metric space with a properly discontinuous, isometric G-action $G \curvearrowright \mathcal{X}$. The choice of a basepoint $o \in \mathcal{X}$ induces a left invariant pseudometric on G by $d_G(g, g') := d_{\mathcal{X}}(g\,.\,o, g'.\,o)$.

Define the growth exponent δ_G of G with respect to $\mathcal{X}$ to be the growth exponent of G with respect to an induced pseudometric d_G; this depends only on the G-space $\mathcal{X}$, since a different choice of basepoint in $\mathcal{X}$ defines a pseudometric that differs from d_G by an additive constant. Likewise, let $\delta_{G/\Gamma}$ denote the growth exponent of G/Γ with respect to a pseudometric on G/Γ induced by $d_{\mathcal{X}}$.

Definition 0.2. *$G \curvearrowright \mathcal{X}$ is a growth tight action if $\delta_G > \delta_{G/\Gamma}$ for every infinite normal subgroup $\Gamma \trianglelefteq G$.*

Some groups admit growth tight actions for the simple reason that they lack any infinite, infinite index normal subgroups. For such a group G, every action on a G-space with positive growth exponent will be growth tight. Exponentially growing simple groups are examples, as are irreducible lattices in higher rank semisimple Lie groups, by the Margulis normal subgroup theorem [1991].

Growth tightness[1] for word metrics was introduced and studied by Grigorchuk and de la Harpe [1997], who showed, for example, that a finite rank free group equipped with the word metric from a free generating set is growth tight. On the other hand, they showed that the product of a free group with itself, generated by free generating sets of the factors, is not growth tight. Together with the normal subgroup theorem, these results suggest that for interesting examples of growth tightness we should examine "rank 1" type behavior. Further evidence for this idea comes from the work of Sambusetti and collaborators, who in [Sambusetti 2002b; 2003; 2004; Dal'Bo et al. 2011] proved growth tightness for the action of the fundamental group of a negatively curved Riemannian manifold on its Riemannian universal cover.

In the study of nonpositively curved, or CAT(0), spaces there is a well established idea that a space may be nonpositively curved but have some specific directions that look negatively curved. More precisely:

Definition 0.3 [Ballmann and Brin 1995]. A hyperbolic isometry of a proper CAT(0) space is *rank 1* if it has an axis that does not bound a half-flat.

In Definition 2.17, we introduce the notion for an element of G to be *strongly contracting* with respect to $G \curvearrowright \mathcal{X}$. In the case that $\mathcal{X}$ is a CAT(0) G-space, the strongly contracting elements of G are precisely those that act as rank 1 isometries of $\mathcal{X}$; see Theorem 9.1.

In addition to having a strongly contracting element, we will assume that the orbit of G in $\mathcal{X}$ is not too badly distorted. There are two different ways to make this precise.

We say a G-space is *C-quasiconvex* if there exists a C-quasiconvex G-orbit; see Definitions 1.3 and 1.4. This means that it is possible to travel along geodesics joining points in the orbit of G without leaving a neighborhood of the orbit.

Theorem 6.4. *Let G be a finitely generated, nonelementary group. Let $\mathcal{X}$ be a quasiconvex G-space. If G contains a strongly contracting element, then $G \curvearrowright \mathcal{X}$ is a growth tight action.*

Alternatively, we can assume that the growth rate of the number of orbit points that can be reached by geodesics lying entirely, except near the endpoints, outside a neighborhood of the orbit is strictly smaller than the growth rate of the group:

Theorem 6.3. *Let G be a finitely generated, nonelementary group. Let $\mathcal{X}$ be a G-space. If G contains a strongly contracting element and there exists a $C \geqslant 0$ such*

[1] Grigorchuk and de la Harpe define growth tightness in terms of "growth rate", which is just the exponentiation of our growth exponent. The growth exponent definition is analogous to the notion of "volume entropy" familiar in Riemannian geometry, and is more compatible with the Poincaré series in Section 1B.

that the C-complementary growth exponent of G is strictly less than the growth exponent of G, then $G \curvearrowright \mathcal{X}$ is a growth tight action.

See Definition 6.2 for a precise definition of the C-complementary growth exponent. The proof of Theorem 6.4 is a special case of the proof of Theorem 6.3. Using Theorem 6.4, we prove:

Theorem 8.6. *If $\mathcal{X}$ is a quasiconvex, relatively hyperbolic G-space and G does not coarsely fix a peripheral subspace, then $G \curvearrowright \mathcal{X}$ is a growth tight action.*

This generalizes all previously known results for growth tightness of cocompact actions: every example already known of a group that admits a growth tight action and has some infinite, infinite index normal subgroups is relatively hyperbolic; conversely, relatively hyperbolic groups admit growth tight actions [Arzhantseva and Lysenok 2002; Sambusetti 2002a; Yang 2013; Sambusetti 2003; Sabourau 2013; Dal'Bo et al. 2011].

We also use Theorem 6.4 to prove growth tightness for actions on non-relatively hyperbolic spaces. For instance, we prove that a group action on a proper CAT(0) space with a rank 1 isometry is growth tight:

Theorem 9.2. *If G is a finitely generated, nonelementary group and $\mathcal{X}$ is a quasiconvex, CAT(0) G-space such that G contains an element that acts as a rank 1 isometry on $\mathcal{X}$, then $G \curvearrowright \mathcal{X}$ is a growth tight action.*

Two interesting classes of non-relatively hyperbolic groups to which Theorem 9.2 applies are nonelementary right angled Artin groups, which are non-relatively hyperbolic when the defining graph is connected, and flip-graph manifolds. These are the first examples of non-relatively hyperbolic groups that admit nontrivial growth tight actions.

Theorem 9.3. *Let Θ be a finite graph that is not a join and has more than one vertex. The action of the right angled Artin group G defined by Θ on the universal cover $\mathcal{X}$ of the Salvetti complex associated to Θ is a growth tight action.*

Theorem 9.4. *Let M be a flip-graph manifold. Let G and $\mathcal{X}$ be the fundamental group and universal cover, respectively, of M. Then the action of G on $\mathcal{X}$ by deck transformations is a growth tight action.*

We even exhibit an infinite family of non-relatively hyperbolic, non-CAT(0) groups that admit cocompact, growth tight actions:

Theorem 11.1. *The Brady–Bridson snowflake groups $BB(1,r)$ for $r \geqslant 3$ admit cocompact, growth tight actions.*

Using Theorem 6.3, we prove growth tightness for interesting nonquasiconvex actions. We generalize a theorem of Dal'bo, Peigné, Picaud, and Sambusetti [Dal'Bo et al. 2011] for Kleinian groups satisfying an additional parabolic gap condition — see Definition 8.9 — to cusp-uniform actions on arbitrary hyperbolic spaces satisfying the parabolic gap condition:

Theorem 8.10. *Let G be a finitely generated, nonelementary group. Let $G \curvearrowright \mathcal{X}$ be a cusp uniform action on a hyperbolic space. Suppose that G satisfies the parabolic gap condition. Then $G \curvearrowright \mathcal{X}$ is a growth tight action.*

Once again, our theorems extend beyond actions on relatively hyperbolic spaces, as we use Theorem 6.3 to prove:

Theorem 10.2. *The action of the mapping class group of a hyperbolic surface on its Teichmüller space with the Teichmüller metric is a growth tight action.*

Mapping class groups, barring exceptional low complexity cases, are neither relatively hyperbolic nor CAT(0).

In Part I of this paper we prove our main results, Theorem 6.3 and Theorem 6.4. We show in Proposition 3.1 that if there exists a strongly contracting element for $G \curvearrowright \mathcal{X}$ then every infinite normal subgroup Γ contains a strongly contracting element h. We prove growth tightness by bounding the growth exponent of a subset that is orthogonal, in a coarse sense, to every translate of an axis for h.

A dual problem, which is of independent interest, is to find the growth exponent of the conjugacy class of h. In Section 7 we show that the growth exponent of the conjugacy class of a strongly contracting element is exactly half the growth exponent of the group, provided the strongly contracting element moves the base point far enough.

In Part II we produce new examples of group actions with strongly contracting elements. These include groups acting on relatively hyperbolic metric spaces (Section 8), certain CAT(0) groups (Section 9), mapping class groups (Section 10), and snowflake groups (Section 11). Our main theorems imply that all these groups admit growth tight actions. These are first examples of growth tight actions and groups which do not come from and are not relatively hyperbolic groups.

0A. *Invariance.* Growth tightness is a delicate condition. A construction of Dal'bo, Otal, and Peigné [Dal'bo et al. 2000] — see Observation 8.8 — shows that there exist groups G and noncocompact, hyperbolic, equivariantly quasi-isometric G-spaces $\mathcal{X}$ and $\mathcal{X}'$ such that $G \curvearrowright \mathcal{X}$ is growth tight and $G \curvearrowright \mathcal{X}'$ is not.

In [Cashen and Tao 2014], we extend the techniques of this paper to produce the first examples of groups that admit a growth tight action on one of their Cayley graphs and a non-growth tight action on another. This answers in the affirmative the following question of Grigorchuk and de la Harpe [1997]:

Question 1. Does there exist a word metric for which $F_2 \times F_2$ is growth tight?

Recall that $F_2 \times F_2$ is not growth tight with respect to a generating set that is a union of free generating sets of the two factors.

More generally, a product of infinite groups acting on the l^1 product of their Cayley graphs is not growth tight. Such l^1 products and the Dal'bo, Otal, Peigné examples are the only known general constructions of non-growth tight examples. It would be interesting to have a condition to exclude growth tightness. One can not hope to bound the growth exponents of quotients away from that of the group, as Shukhov [1999] and Coulon [2013] have given examples of hyperbolic groups and sequences of quotients whose growth exponents limit to that of the group. At present, growth tightness can only be excluded for a particular action by exhibiting a quotient of the group by an infinite normal subgroup whose growth exponent is equal to that of the group.

0B. ***The Hopf property.*** A group G is *Hopfian* if there is no proper quotient of G isomorphic to G.

Let $\mathfrak{D}$ be a set of pseudometrics on G that is *quotient-closed*, in the sense that if Γ is a normal subgroup of G such that there exists an isomorphism $\phi : G \to G/\Gamma$, then for every $d \in \mathfrak{D}$, the pseudometric on G obtained by pulling back via ϕ the pseudometric on G/Γ induced by d is also in $\mathfrak{D}$. For example, the set of word metrics on G coming from finite generating sets is quotient-closed.

Suppose further that $\mathfrak{D}$ contains a minimal growth pseudometric d_0, i.e., that $\delta_{G,d_0} = \inf_{d \in \mathfrak{D}} \delta_{G,d}$, and that G is growth tight with respect to d_0.

Proposition 0.4. *Let G be a finitely generated group with a bound on the cardinalities of its finite normal subgroups. Suppose that there exists a quotient-closed set $\mathfrak{D}$ of pseudometrics on G that contains a growth tight, minimal growth element d_0, as above. Then, G is Hopfian.*

The hypothesis on bounded cardinalities of finite normal subgroups holds for all groups of interest in this paper; see Theorem 1.12.

Proof. Suppose that Γ is a normal subgroup of G such that $G \cong G/\Gamma$. Let d be the pseudometric on G obtained from pulling back the pseudometric on G/Γ induced by d_0. Since $\mathfrak{D}$ is quotient-closed, $d \in \mathfrak{D}$. By minimality, $\delta_{G,d_0} \leqslant \delta_{G,d}$, but by growth tightness, $\delta_{G,d} \leqslant \delta_{G,d_0}$, with equality only if Γ is finite. Thus, the only normal subgroups Γ for which we could have $G \cong G/\Gamma$ are finite. However, if $G \cong G/\Gamma$ for some finite Γ then G has arbitrarily large finite normal subgroups, contrary to hypothesis. □

Grigorchuk and de la Harpe [1997] suggested this as a possible approach to the question of whether a nonelementary Gromov hyperbolic group is Hopfian, in the particular case that $\mathfrak{D}$ is the set of word metrics on G. Arzhantseva and Lysenok [2002] proved that every word metric on a nonelementary hyperbolic group is growth tight. They conjectured that the growth exponent of such a group achieves its infimum on some finite generating set and proved a step towards this conjecture [Arzhantseva and Lysenok 2006]. Sambusetti [2002a] gave an examples of a (nonhyperbolic) group for which the set of word metrics does not realize its infimal growth exponent. In general, it is difficult to determine whether a given group has a generating set that realizes the infimal growth exponent among word metrics. Part of our motivation for studying growth tight actions is to open new possibilities for the set $\mathfrak{D}$ of pseudometrics considered above.

Torsion free hyperbolic groups are Hopfian [Sela 1999]. Reinfeldt and Weidmann [2014] have announced a generalization of Sela's techniques to hyperbolic groups with torsion, and concluded that all hyperbolic groups are Hopfian.

0C. ***The rank rigidity conjecture.*** The rank rigidity conjecture (see [Caprace and Sageev 2011; Ballmann and Buyalo 2008]) asserts that if $\mathcal{X}$ is a locally compact, irreducible, geodesically complete CAT(0) space, and G is an infinite discrete group acting properly and cocompactly on $\mathcal{X}$, then one of the following holds:

(1) $\mathcal{X}$ is a higher rank symmetric space.

(2) $\mathcal{X}$ is a Euclidean building of dimension at least 2.

(3) G contains a rank 1 isometry.

In case (1), the Margulis normal subgroup theorem implies that G is trivially growth tight, since it has no infinite, infinite index normal subgroups. Conjecturally, the Margulis normal subgroup theorem also holds in case (2). Our Theorem 9.2 says that if $\mathcal{X}$ is proper then $G \curvearrowright \mathcal{X}$ is a growth tight action in case (3). Thus, a non-growth tight action of a nonelementary group on a proper, irreducible CAT(0) space as above would provide a counterexample either to the rank rigidity conjecture or to the conjecture that the Margulis normal subgroup theorem applies to Euclidean buildings.

The rank rigidity conjecture is known to be true for many interesting classes of spaces, such as Hadamard manifolds [Ballmann 1995], 2-dimensional, piecewise Euclidean cell complexes [Ballmann and Brin 1995], Davis complexes of Coxeter groups [Caprace and Fujiwara 2010], universal covers of Salvetti complexes of right angled Artin groups [Behrstock and Charney 2012], and finite dimensional CAT(0) cube complexes [Caprace and Sageev 2011]; so Theorem 9.2 provides many new examples of growth tight actions.

It is unclear when growth tightness holds if $\mathcal{X}$ is reducible. A direct product of infinite groups acting via a product action on a product space with the l^1 metric fails to be growth tight. However, there are also examples [Burger and Mozes 1997] of infinite simple groups acting cocompactly on products of trees. In [Cashen and Tao 2014], we find partial results in the case that the group action is a product action.

0D. ***Outline of the proof of the main theorems.*** Sambusetti [2002a] proved that a nonelementary free product of nontrivial groups has a greater growth exponent than that of either factor. Thus, a strategy to prove growth tightness is to find a subset of G that looks like a free product, with one factor that grows like the quotient group we are interested in. Specifically:

(1) Find a subset $A \subset G \subset \mathcal{X}$ such that $\delta_A = \delta_{G/\Gamma}$. We will obtain A as a coarsely dense subset of a minimal section of the quotient map $G \to G/\Gamma$; see Definition 4.4.

(2) Construct an embedding of a free product set $A * \mathbb{Z}_2$ into $\mathcal{X}$. The existence of a strongly contracting element $h \in \Gamma$ is used in the construction of this embedding; see Proposition 5.1.

(3) Show that $\delta_{G/\Gamma} = \delta_{A,d_{\mathcal{X}}} < \delta_{A*\mathbb{Z}/2\mathbb{Z},d_{\mathcal{X}}} \leqslant \delta_G$. In this step it is crucial that A is divergent; see Definition 1.7 and Lemma 6.1. We use the quasiconvexity and complementary growth exponent to establish divergence.

This outline, due to Sambusetti, is nowadays standard. Typically, step (2) is accomplished by a ping-pong argument, making use of fine control on the geometry of the space $\mathcal{X}$. Our methods are coarser than such a standard approach, and therefore can be applied to a wider variety of spaces. We use, in particular, a technique of Bestvina, Bromberg, and Fujiwara [Bestvina et al. 2014] to construct an action of G on a quasitree. Verifying that the map from the free product set into $\mathcal{X}$ is an embedding amounts to showing that elements in A do not cross certain coarse edges of the quasitree.

Part I. Growth tight actions

1. Preliminaries

Fix a G-space $\mathcal{X}$. From now on, d is used to denote the metric on $\mathcal{X}$ as well as the induced pseudometric on G and G/Γ. Since there will be no possibility of confusion, we suppress d from the growth exponent notation.

We denote by $\mathcal{B}_r(x)$ the open ball of radius r about the point x and by $\mathcal{B}_r(\mathcal{A}) := \bigcup_{x\in\mathcal{A}} \mathcal{B}_r(x)$ the open r-neighborhood about the set $\mathcal{A}$. The closed r-ball and closed r-neighborhood are denoted $\overline{\mathcal{B}}_r(x)$ and $\overline{\mathcal{B}}_r(\mathcal{A})$, respectively.

1A. ***Coarse language.*** All of the following definitions may be written without specifying C to indicate that some such $C \geqslant 0$ exists: Two subsets $\mathcal{A}$ and $\mathcal{A}'$ of $\mathcal{X}$ are C-*coarsely equivalent* if $\mathcal{A} \subset \overline{\mathcal{B}}_C(\mathcal{A}')$ and $\mathcal{A}' \subset \overline{\mathcal{B}}_C(\mathcal{A})$. A subset $\mathcal{A}$ of $\mathcal{X}$ is C-*coarsely dense* if it is C-coarsely equivalent to $\mathcal{X}$. A subset $\mathcal{A}$ of $\mathcal{X}$ is C-*coarsely connected* if for every a and a' in $\mathcal{A}$ there exists a chain $a = a_0, a_1, \dots, a_n = a'$ of points in $\mathcal{A}$ with $d(a_i, a_{i+1}) \leqslant C$.

A *pseudomap* $\phi : \mathcal{X} \to \mathcal{Y}$ assigns to each point in $\mathcal{X}$ a subset $\phi(x)$ of $\mathcal{Y}$. A pseudomap is *C-coarsely well-defined* if for every $x \in \mathcal{X}$ the set $\phi(x)$ of $\mathcal{Y}$ has diameter at most C. Pseudomaps ϕ and ϕ' with the same domain and codomain are *C-coarsely equivalent* or *C-coarsely agree* if $\phi(x)$ is C-coarsely equivalent to $\phi'(x)$ for every x in the domain. A C-coarsely well-defined pseudomap is called a *C-coarse map*. From a C-coarse map we can obtain a C-coarsely equivalent map by selecting one point from every image set. Conversely:

Lemma 1.1. *If $\phi : \mathcal{X} \to \mathcal{Y}$ is coarsely G-equivariant then there is an equivariant coarse map coarsely equivalent to ϕ.*

Proof. Suppose there is a C such that $d(g\,.\,\phi(x), \phi(g\,.\,x)) \leqslant C$ for all $x \in \mathcal{X}$ and $g \in G$. Define $\phi'(x) := \bigcup_{g \in G} g^{-1}\,.\,\phi(g\,.\,x)$. Then, ϕ' is G-equivariant and C-coarsely equivalent to ϕ. □

Notation 1.2. If $\phi : \mathcal{X} \to \mathcal{Y}$ is a pseudomap and $\mathcal{A}$ and $\mathcal{A}'$ are subsets of $\mathcal{X}$, let $d^{\phi}(\mathcal{A}, \mathcal{A}')$ denote the diameter of $\phi(\mathcal{A}) \cup \phi(\mathcal{A}')$.

Definition 1.3. A subset $\mathcal{A} \subset \mathcal{X}$ is *C-quasiconvex* if for every $a_0, a_1 \in \mathcal{A}$, there exists a geodesic γ between a_0 and a_1 such that $\gamma \subset \overline{\mathcal{B}}_C(\mathcal{A})$. It is *$C$-strongly quasiconvex* if every geodesic with endpoints in $\mathcal{A}$ stays in $\overline{\mathcal{B}}_C(\mathcal{A})$.

Definition 1.4. A G-space $\mathcal{X}$ is *C-quasiconvex* if it contains a C-quasiconvex G-orbit.

For convenience, if $\mathcal{X}$ is a quasiconvex G-space we assume we have chosen a basepoint $o \in \mathcal{X}$ such that $G\,.\,o$ is quasiconvex.

A group is *elementary* if it has a finite index cyclic subgroup.

Definition 1.5. Let $g \in G$. The *elementary closure* of g, denoted by $E(g)$, is the largest virtually cyclic subgroup containing g, if such a subgroup exists.

A map $\phi : \mathcal{X} \to \mathcal{Y}$ is an *(M, C)-quasi-isometric embedding*, for some $M \geqslant 1$ and $C \geqslant 0$, if, for all $x_0, x_1 \in \mathcal{X}$:

$$\frac{1}{M} d(x_0, x_1) - C \leqslant d(\phi(x_0), \phi(x_1)) \leqslant M\, d(x_0, x_1) + C$$

A map ϕ is *C-coarsely M-Lipschitz* if the second inequality holds, and is a *quasi-isometry* if it is a quasi-isometric embedding whose image is C-coarsely dense.

An *(M, C)-quasigeodesic* is an (M, C)-quasi-isometric embedding of a coarsely connected subset of $\mathbb{R}$. If $\gamma : I \to \mathcal{X}$ is a quasigeodesic we let γ_t denote the point $\gamma(t)$, and let γ denote the image of γ in $\mathcal{X}$.

Definition 1.6. A quasigeodesic $\mathcal{Q}$ is *Morse* if for every $M \geqslant 1$ there exists a $K \geqslant 0$ such that every (M, M)-quasigeodesic with endpoints on $\mathcal{Q}$ is contained in the K-neighborhood of $\mathcal{Q}$.

We will use notation to simplify some calculations. Let C be a "universal constant". For us this will usually mean a constant that depends on $G \curvearrowright \mathcal{X}$ and a choice of $o \in \mathcal{X}$, but not on the point in $\mathcal{X}$ at which quantities a and b are calculated.

- For $a \leqslant Cb$, we write $a \overset{*}{\prec} b$.
- For $\frac{1}{C}b \leqslant a \leqslant Cb$, we write $a \overset{*}{\asymp} b$.
- For $a \leqslant b + C$, we write $a \overset{+}{\prec} b$.
- For $b - C \leqslant a \leqslant b + C$, we write $a \overset{+}{\asymp} b$.
- For $a \leqslant Cb + C$, we write $a \prec b$.
- For $\frac{1}{C}b - C \leqslant a \leqslant Cb + C$, we write $a \asymp b$.

1B. *The Poincaré series and growth.* Let $(\mathcal{X}, o, d)$ be a pseudometric space with choice of basepoint. Let $|x| := d(o, x)$ be the induced seminorm. Define the *Poincaré series* of $\mathcal{A} \subset \mathcal{X}$ to be

$$\Theta_{\mathcal{A}}(s) := \sum_{a \in \mathcal{A}} \exp(-s|a|)$$

Another related series is:

$$\Theta'_{\mathcal{A}}(s) := \sum_{n=0}^{\infty} \#(\overline{\mathcal{B}}_n(o) \cap \mathcal{A}) \cdot \exp(-sn)$$

The series $\Theta_{\mathcal{A}}$ and $\Theta'_{\mathcal{A}}$ have the same convergence behavior, since $\Theta_{\mathcal{A}}(s) = \Theta'_{\mathcal{A}}(s) \cdot (1 - \exp(-s))$. It follows that the growth exponent of $\mathcal{A}$ is a *critical exponent* for $\Theta'_{\mathcal{A}}$ and $\Theta_{\mathcal{A}}$: the series converge for s greater than the critical exponent and diverge for s less than the critical exponent.

Definition 1.7. $\mathcal{A} \subset \mathcal{X}$ is *divergent* if $\Theta_{\mathcal{A}}$ diverges at its critical exponent.

Since point stabilizers are finite, if $A < G$ and we set $\mathcal{A} := A\,.\,o$ then $\Theta_A \overset{*}{\asymp} \Theta_{\mathcal{A}}$ and $\Theta'_A \overset{*}{\asymp} \Theta'_{\mathcal{A}}$. This implies $\delta_A = \delta_{\mathcal{A}}$, so we can compute the growth exponent of A with respect to the pseudometric on A induced by $G \curvearrowright \mathcal{X}$ by computing the growth exponent of the A-orbit as a subset of $\mathcal{X}$.

1C. *The quasitree construction.* We recall the method of Bestvina, Bromberg, and Fujiwara [Bestvina et al. 2014] for producing group actions on quasitrees. A *quasitree* is a geodesic metric space that is quasi-isometric to a simplicial tree. Manning [2005] gave a characterization of quasitrees as spaces satisfying a "bottleneck" property. We use an equivalent formulation:

Definition 1.8. A geodesic metric space satisfies the *bottleneck property* if there exists a number Δ such that for all x and y in $\mathcal{X}$, and for any point m on a geodesic segment from x to y, every path from x to y passes through $\overline{\mathcal{B}}_{\Delta}(m)$.

Theorem 1.9 [Manning 2005, Theorem 4.6]. *A geodesic metric space is a quasitree if and only if it satisfies the bottleneck property.*

Let $\mathbb{Y}$ be a collection of geodesic metric spaces, and suppose for each $X,\ Y \in \mathbb{Y}$ we have a subset $\pi_Y(X) \subset Y$, which is referred to as the *projection of X to Y*. Let $d_Y^\pi(X, Z) := \operatorname{diam} \pi_Y(X) \cup \pi_Y(Z)$.

Definition 1.10 (projection axioms). A set $\mathbb{Y}$ with projections as above satisfies the *projection axioms* if there exist $\xi \geqslant 0$ such that for all distinct $X, Y, Z \in \mathbb{Y}$:

(P0) $\operatorname{diam} \pi_Y(X) \leqslant \xi$

(P1) At most one of $d_X^\pi(Y, Z)$, $d_Y^\pi(X, Z)$, or $d_Z^\pi(X, Y)$ is strictly greater than ξ.

(P2) $|\{V \in \mathbb{Y} \mid d_V^\pi(X, Y) > \xi\}| < \infty$

For a motivating example, let G be the fundamental group of a closed hyperbolic surface, and let $\mathcal{H}$ be the axis in $\mathbb{H}^2$ of $h \in G$. Let $\mathbb{Y}$ be the distinct G-translates of $\mathcal{H}$, and for each $Y \in \mathbb{Y}$, let π_Y be closest point projection to Y. In this example, projection distances arise as closest point projection in an ambient space containing $\mathbb{Y}$. Bestvina, Bromberg, and Fujiwara consider abstractly the collection $\mathbb{Y}$ and projections satisfying the projection axioms, and build an ambient space containing a copy of $\mathbb{Y}$ such that closest point projection agrees with the given projections, up to bounded error:

Theorem 1.11 [Bestvina et al. 2014, Theorems A and B]. *Consider a set $\mathbb{Y}$ of geodesic metric spaces and projections satisfying the projection axioms. There exists a geodesic metric space $\mathcal{Y}$ containing disjoint, isometrically embedded, totally geodesic copies of each $Y \in \mathbb{Y}$, such that for $X,\ Y \in \mathbb{Y}$, closest point projection of X to Y in $\mathcal{Y}$ is uniformly coarsely equivalent to $\pi_Y(X)$.*

The construction is equivariant with respect to any group action that preserves the projections. Also, if each $Y \in \mathbb{Y}$ is a quasitree, with uniform bottleneck constants, then $\mathcal{Y}$ is a quasitree.

The basic idea is that Z is "between" X and Y in $\mathbb{Y}$ if $d_Z^\pi(X, Y)$ is large, and X and Y are "close" if there is no Z between them. Essentially, $\mathcal{Y}$ is constructed by choosing parameters C and K and connecting every point of $\pi_Y(X)$ to every point of $\pi_X(Y)$ by an edge of length K if there does not exist $Z \in \mathbb{Y}$ with $d_Z^\pi(X, Y) > C$. For technical reasons one actually must perturb the projection distances by a bounded amount first. Then, if C is chosen sufficiently large and K is chosen sufficiently large with respect to C, the resulting space is the space $\mathcal{Y}$ of Theorem 1.11.

1D. ***Hyperbolically embedded subgroups.*** Dahmani, Guirardel, and Osin [Dahmani et al. 2011] define the concept of a *hyperbolically embedded subgroup*. This is a generalization of a peripheral subgroup of a relatively hyperbolic group. We will not state the definition, as it is technical and we will not work with this property directly, but it follows from Theorem 4.42 from that reference that $E(h)$ is hyperbolically embedded in G for any strongly contracting element h. The proof of this theorem proceeds by considering the action of $E(h)$ on a quasitree constructed via the method of Bestvina, Bromberg, and Fujiwara.

We state some results on hyperbolically embedded subgroups that are related to the work in this paper. These are not used in the proofs of the main theorems.

Theorem 1.12 [Dahmani et al. 2011, Theorem 2.23]. *If G has a hyperbolically embedded subgroup, then G has a maximal finite normal subgroup.*

Recall that this theorem guarantees one of the hypotheses of Proposition 0.4.

Theorem 1.13. *If G contains an infinite order element h such that $E(h)$ is hyperbolically embedded, then G has an infinite, infinite index normal subgroup.*

Proof. By [Dahmani et al. 2011, Theorem 5.15], for a sufficiently large n, the normal closure of $\langle h^n \rangle$ in G is the free product of the conjugates of $\langle h^n \rangle$. □

This theorem says that our main results are true for interesting reasons, not simply for lack of normal subgroups.

Minasyan and Osin [2015] produce hyperbolically embedded subgroups in certain graphs of groups. We use these to produce growth tight examples in Theorem 9.5.

Theorem 1.14 [Minasyan and Osin 2015, Theorem 4.17]. *Let G be a finitely generated, nonelementary group that splits nontrivially as a graph of groups and is not an ascending HNN-extension. If there exist two edges of the corresponding Bass–Serre tree whose stabilizers have finite intersection then G contains an infinite order element h such that $E(h)$ is hyperbolically embedded in G.*

2. Contraction and constriction

In this section we introduce properties called "contracting" and "constricting" that generalize properties of closest point projection to a geodesic in hyperbolic space, and verify that the "strong" versions of these properties are sufficient to satisfy the projection axioms of Definition 1.10. These facts are well known to the experts[2], but as there is currently no published general treatment of this material, we provide a detailed account.

2A. ***Contracting and constricting.*** In this section we define contracting and constricting maps and show that the strong versions of these properties are equivalent.

Definition 2.1. A C-coarse map $\pi : \mathcal{X} \to \mathcal{A}$ is *C-coarsely a closest point projection* if for all x there exists an $a \in \mathcal{A}$ with $d(x, \mathcal{A}) = d(x, a)$ such that $\operatorname{diam}\{a\} \cup \pi(x) \leqslant C$.

[2] For example, [Sisto 2011] shows that the projection axioms are satisfied for constricting elements, without assuming that $\mathcal{X}$ is proper.

Recall $d^\pi(x_0, x_1) := \operatorname{diam} \pi(x_0) \cup \pi(x_1)$.

Definition 2.2. $\pi : \mathcal{X} \to \mathcal{A}$ is (M, C)-*contracting* for $C \geqslant 0$ and $M \geqslant 1$ if

(1) π and $\mathrm{Id}_\mathcal{A}$ are C-coarsely equivalent on $\mathcal{A}$, and

(2) $d(x_0, x_1) < \frac{1}{M} d(x_0, \mathcal{A}) - C$ implies $d^\pi(x_0, x_1) \leqslant C$ for all $x_0, x_1 \in \mathcal{X}$.

We say π is *strongly contracting* if it is $(1, C)$-contracting and, for all $x \in \mathcal{X}$, $d(x, \pi(x)) - d(x, \mathcal{A}) \leqslant C$.

Another formulation of strong contraction says that geodesics far from $\mathcal{A}$ have bounded projections to $\mathcal{A}$:

Definition 2.3. A coarse map $\pi : \mathcal{X} \to \mathcal{A}$ has the *bounded geodesic image property* if there is a constant C such that for every geodesic $\mathcal{L}$, if $\mathcal{L} \cap \mathcal{B}_C(\mathcal{A}) = \varnothing$ then $\operatorname{diam}(\pi(\mathcal{L})) \leqslant C$.

Lemma 2.4. *If $d(x, \pi(x)) - d(x, \mathcal{A})$ is uniformly bounded, then π has the bounded geodesic image property if and only if it is strongly contracting.*

Proof. First, assume that π has the bounded geodesic image property, for some constant C. Let x be any point in $\mathcal{X} \setminus \mathcal{B}_C(\mathcal{A})$. For any y such that $d(x, y) < d(x, \mathcal{A}) - C$, every geodesic from x to y remains outside $\mathcal{B}_C(\mathcal{A})$, so its projection has diameter at most C.

For the converse, suppose $\pi : \mathcal{X} \to \mathcal{A}$ is a C-coarse map that is $(1, C)$-contracting and $d(x, \pi(x)) - d(x, \mathcal{A}) \leqslant C$ for all $x \in \mathcal{X}$. If $C = 0$ then balls outside of $\mathcal{B}_C(\mathcal{A})$ project to a single point, and we are done; so assume that $C > 0$. Let $\mathcal{L} : [0, T] \to \mathcal{X}$ be a geodesic that stays outside $\mathcal{B}_{3C}(\mathcal{A})$. Let $t_0 := d(\mathcal{L}_0, \mathcal{A}) - C$, and let $s := T - d(\mathcal{L}_T, \mathcal{A}) + C$. If $s \leqslant t_0$ then $d^\pi(\mathcal{L}_0, \mathcal{L}_T) \leqslant 2C$. Otherwise, define $t_{i+1} := t_i + d(\mathcal{L}_{t_i}, \mathcal{A}) - C$, provided $t_{i+1} < s$. Each $t_{i+1} - t_i \geqslant 2C$, so we have a partition of $[0, T]$ into subintervals $[0, t_0]$, $[t_0, t_1]$, $\dots$, $[t_{k-1}, t_k]$, $[t_k, s]$, $[s, T]$ with $k < (s - t_0)/(2C)$, and if $[a, b]$ is one of these intervals then $d^\pi(\mathcal{L}_a, \mathcal{L}_b) \leqslant C$, by strong contraction. Now,

$$\begin{aligned} d(\mathcal{L}_0, \mathcal{L}_T) &\leqslant d(\mathcal{L}_0, \pi(\mathcal{L}_0)) + d(\pi(\mathcal{L}_0), \pi(\mathcal{L}_{t_0})) + d(\pi(\mathcal{L}_{t_0}), \pi(\mathcal{L}_s)) \\ &\quad + d(\pi(\mathcal{L}_s), \pi(\mathcal{L}_T)) + d(\pi(\mathcal{L}_T), \mathcal{L}_T) \\ &\leqslant d(\mathcal{L}_0, \pi(\mathcal{L}_0)) + d(\pi(\mathcal{L}_T), \mathcal{L}_T) + C\left(2 + \frac{s - t_0}{2C}\right), \end{aligned}$$

and

$$\begin{aligned} d(\mathcal{L}_0, \mathcal{L}_T) &= d(\mathcal{L}_0, \mathcal{L}_{t_0}) + d(\mathcal{L}_{t_0}, \mathcal{L}_s) + d(\mathcal{L}_s, \mathcal{L}_T) \\ &= d(\mathcal{L}_0, \mathcal{A}) - C + s - t_0 + d(\mathcal{L}_T, \mathcal{A}) - C, \end{aligned}$$

so

$$s - t_0 \leqslant 2\big(5C + d(\mathcal{L}_0, \pi(\mathcal{L}_0)) - d(\mathcal{L}_0, \mathcal{A}) + d(\mathcal{L}_T, \pi(\mathcal{L}_T)) - d(\mathcal{L}_T, \mathcal{A})\big) \leqslant 14C.$$

This means that $k < 7$, so $d^\pi(\mathcal{L}_0, \mathcal{L}_T) \leqslant C(3 + k) < 10C$. □

If π is only (M, C)-contracting then a similar argument shows that $d^\pi(\mathcal{L}_0, \mathcal{L}_T)$ is bounded in terms of C and $\log_{(M+1)/(M-1)}(d(\mathcal{L}_0, \mathcal{A})\, d(\mathcal{L}_T, \mathcal{A}))$.

We now introduce the notion of a constricting map. Using constricting maps will simplify some of our proofs, but it turns out that the strong versions of constricting and contracting are equivalent.

Definition 2.5. A *path system* is a transitive collection of quasigeodesics with uniform constants that is closed under taking subpaths.

A path system is *minimizing* if, for some $C \geqslant 0$, it contains a path system consisting of $(1, C)$-quasigeodesics.

Definition 2.6. Let $\mathcal{PS}$ be a path system. For $M \geqslant 1$ and $C \geqslant 0$, a coarse map $\pi : \mathcal{X} \to \mathcal{A}$ is (M, C)-*$\mathcal{PS}$-constricting*[3] if:

(1) $\mathcal{PS}$ contains a path system consisting of (M, C)-quasigeodesics,

(2) π and $\mathrm{Id}_{\mathcal{A}}$ are C-coarsely equivalent on $\mathcal{A}$, and

(3) for every $\mathcal{P} \in \mathcal{PS}$ with endpoints x_0 and x_1, if $d^\pi(x_0, x_1) > C$ then, for both $i \in \{0, 1\}$, we have $d(\pi(x_i), \mathcal{P}) \leqslant C$.

A coarse map is *constricting* if it is (M, C)-$\mathcal{PS}$-constricting for some path system $\mathcal{PS}$ and *strongly constricting* if it is $(1, C)$-constricting for the path system consisting of all geodesics.

Lemma 2.7. *If $\pi : \mathcal{X} \to \mathcal{A}$ is constricting, then it is contracting.*

Proof. Suppose π is (M, C)-$\mathcal{PS}$-constricting C-coarse map for a path system $\mathcal{PS}$ consisting of (M, C)-quasigeodesics. Suppose $\mathcal{P} : [0, T] \to \mathcal{X}$ is a path in $\mathcal{PS}$ with $\mathcal{P}_0 = x$ and $\mathcal{P}_T = y$, and suppose $z = \mathcal{P}_s \in \bar{\mathcal{B}}_C(\mathcal{A})$. Using the fact that $\mathcal{P}$ is an (M, C)-quasigeodesic on the intervals $[0, T]$, $[0, s]$, and $[s, T]$, one sees that $d(x, y) \geqslant (1/M^2)(d(x, \mathcal{A}) + d(y, \mathcal{A}) - 4C)$. Therefore, if

$$d(x, y) < \frac{1}{M^2} d(x, \mathcal{A}) - \frac{4C}{M^2},$$

then $\mathcal{P}$ can not enter $\bar{\mathcal{B}}_C(\mathcal{A})$. This would contradict the constricting property, unless $d^\pi(x, y) \leqslant C$. Therefore, π is $(M^2, \max\{C, 4C/M^2\})$-contracting. $\square$

Lemma 2.8. *Let $\pi : \mathcal{X} \to \mathcal{A}$ be a C-coarse map that is $(1, C)$-$\mathcal{PS}$-constricting. For all $x \in \mathcal{X}$ and all $r \geqslant 0$, we have*

$$\{a \in \mathcal{A} \mid d(x, a) \leqslant d(x, \mathcal{A}) + r\} \subset \{a \in \mathcal{A} \mid d(a, \pi(x)) \leqslant r + 5C\}.$$

In particular, setting $r = 0$ shows that closest point projection to $\mathcal{A}$ is coarsely well defined and coarsely equivalent to π.

[3] Sisto [2011] calls this property "$\mathcal{PS}$-contracting". We change the name to avoid conflict with the better established "contracting" terminology of Definition 2.2.

Proof. For $x \in \mathcal{X}$ and $r \geqslant 0$, let $a \in \mathcal{A}$ be a point such that $d(x, a) \leqslant d(x, \mathcal{A}) + r$. Let $\mathcal{P}$ be a $(1, C)$-quasigeodesic from x to a in $\mathcal{PS}$. If $d(a, \pi(x)) > 2C$, then $d^{\pi}(a, x) > C$, so there is a point $z \in \mathcal{P} \cap \overline{\mathcal{B}}_C(\pi(x))$. Now,

$$d(x, z) + C \geqslant d(x, \pi(x)) \geqslant d(x, \mathcal{A}) \geqslant d(x, a) - r.$$

Since $\mathcal{P}$ is a $(1, C)$-quasigeodesic, $d(x, a) \geqslant d(x, z) + d(z, a) - 3C$. As a result, $d(z, a) \leqslant r + 4C$, and $d(a, \pi(x)) \leqslant r + 5C$. □

Proposition 2.9. *Let $\pi : \mathcal{X} \to \mathcal{A}$. The following are equivalent*:

(1) *π is strongly constricting.*

(2) *π is constricting for some minimizing path system.*

(3) *π is strongly contracting.*

(4) *π has the bounded geodesic image property and $d(x, \pi(x)) - d(x, \mathcal{A})$ is uniformly bounded.*

Proof. The fact that (1) implies (2) is immediate.

Suppose π is $(1, C)$-$\mathcal{PS}$-constricting for a minimizing path system $\mathcal{PS}$ consisting of $(1, C)$-quasigeodesics. Lemma 2.7 shows π is $(1, C')$-contracting. By Lemma 2.8, π is coarsely a closest point projection, so $d(x, \pi(x)) - d(x, \mathcal{A})$ is uniformly bounded. Thus, (2) implies (3).

Now suppose π is $(1, C)$-contracting and $d(x, \pi(x)) - d(x, \mathcal{A}) \leqslant C$ for all $x \in \mathcal{X}$. Take any geodesic $\mathcal{L} : [0, T] \to \mathcal{X}$. If $d^{\pi}(\mathcal{L}_0, \mathcal{L}_T) > 10C$ then $\mathcal{L} \cap \mathcal{B}_{3C}(\mathcal{A}) \neq \varnothing$, as in Lemma 2.4. Let $t = t_0, t_1$ be the first and last times, respectively, such that $d(\mathcal{L}_t, \mathcal{A}) \leqslant 3C$. By Lemma 2.4, $d^{\pi}(\mathcal{L}_0, \mathcal{L}_{t_0}) \leqslant 10C$. Thus,

$$d(\pi(\mathcal{L}_0), \mathcal{L}_{t_0}) \leqslant d^{\pi}(\mathcal{L}_0, \mathcal{L}_{t_0}) + d(\pi(\mathcal{L}_{t_0}), \mathcal{L}_{t_0}) \leqslant 14C.$$

The same argument shows that $d(\pi(\mathcal{L}_T), \mathcal{L}_{t_1}) \leqslant 14C$, so π is $(1, 14C)$-constricting for the path system of all geodesics. Thus, (3) implies (1).

Finally, (3) is equivalent to (4) by Lemma 2.4. □

2B. ***Additional properties of contracting and constricting maps.*** We establish some properties of contracting and constricting maps that will be useful in the sequel.

Lemma 2.10. *If π is a $(1, C)$-strongly constricting C-coarse map and $d^{\pi}(x, y) > C$, then $d(x, y) \geqslant d(x, \pi(x)) + d^{\pi}(x, y) + d(\pi(y), y) - 6C$.*

Proof. Let $\mathcal{L}$ be a geodesic from x to y; by strong constriction, there exist s and t such that $d(\mathcal{L}_s, \pi(x)) \leqslant C$ and $d(\mathcal{L}_t, \pi(y)) \leqslant C$. The lemma follows from the triangle inequality and the fact that $\pi(x)$ and $\pi(y)$ have diameter at most C. □

Lemma 2.11. *If $\pi : \mathcal{X} \to \mathcal{A}$ is strongly constricting, then it is coarsely 1-Lipschitz.*

Proof. Let π be a C-coarse map that is $(1, C)$-constricting on the path system of geodesics. Let x_0 and x_1 be arbitrary points, and let $\mathcal{L}$ be a geodesic from x_0 to x_1. If $d^\pi(x_0, x_1) > 4C$ then $\mathcal{L} \cap \mathcal{B}_C(x_i) \neq \varnothing$ for each i, which implies that $d(x_0, x_1) \geqslant d(x_0, \pi(x_0)) + d^\pi(x_0, x_1) + d(\pi(x_1), x_1) - 8C$. Thus, for all x_0 and x_1, we have $d^\pi(x_0, x_1) \leqslant d(x_0, x_1) + 8C$. □

Lemma 2.12. *Let $\pi : \mathcal{X} \to \mathcal{A}$ be an (M, C)-contracting C-coarse map such that $d(x, \pi(x)) - d(x, \mathcal{A}) \leqslant C$ for all $x \in \mathcal{X}$. Fix $K \geqslant 1$. For all sufficiently large D, there exists a $T_{\max}$ such that if $\mathcal{Q} : [0, T] \to \mathcal{X}$ is a (K, K)-quasigeodesic with $d(\mathcal{Q}_0, A) = D = d(\mathcal{Q}_T, \mathcal{A})$ and $\mathcal{Q} \cap \mathcal{B}_D(\mathcal{A}) = \varnothing$ then $T \leqslant T_{\max}$.*

Proof. Let $D > M(K^2C + C + K)$. Let $t_0 := 0$ and let t_{i+1} be the last time that $d(\mathcal{Q}_{t_i}, \mathcal{Q}_{t_{i+1}}) = (1/M)d(\mathcal{Q}_{t_i}, \mathcal{A}) - C$, provided that $t_{i+1} < T$. This subdivides $[0, T]$ into at most $1 + (TK)/(D/M - C - K)$ intervals $[t_0, t_1], \dots, [t_k, T]$, each of which has endpoints whose π-images are distance at most C apart.

Since $\mathcal{Q}$ is a quasigeodesic, $T \leqslant Kd(\mathcal{Q}_0, \mathcal{Q}_T) + K^2$. On the other hand:

$$d(\mathcal{Q}_0, \mathcal{Q}_T) \leqslant 2D + 2C + d^\pi(\mathcal{Q}_0, \mathcal{Q}_T) \leqslant 2D + 2C + C\left(1 + \frac{TK}{\frac{D}{M} - C - K}\right)$$

Combined with the condition on D, this yields an upper bound on T. □

Corollary 2.13. *If $\pi : \mathcal{X} \to \mathcal{A}$ is contracting and $d(x, \pi(x)) - d(x, \mathcal{A})$ is uniformly bounded, then for all $M \geqslant 1$ and $D \geqslant 0$ there exists a K such that every (M, M)-quasigeodesic with endpoints at distance at most D from $\mathcal{A}$ is contained in $\overline{\mathcal{B}}_K(\mathcal{A})$.*

In particular, if $\mathcal{A}$ is a quasigeodesic, then it is Morse.

Lemma 2.14. *Let $\mathcal{Q} : \mathbb{R} \to \mathcal{X}$ be a quasigeodesic, and let $\pi : \mathcal{X} \to \mathcal{Q}$ be a strongly contracting projection. For all $D \geqslant 0$, there exists a K such that if $\mathcal{P} : [0, T] \to \mathcal{X}$ is a geodesic and t_0 and t_1 are such that $d(\mathcal{P}_0, \mathcal{Q}_{t_0}) \leqslant D$ and $d(\mathcal{P}_T, \mathcal{Q}_{t_1}) \leqslant D$, then $\mathcal{Q}_{[t_0, t_1]} \subset \overline{\mathcal{B}}_K(\mathcal{P})$.*

Proof. By Proposition 2.9, π is strongly constricting, so $\mathcal{P}$ passes close to every point in $\pi(\mathcal{P})$. Let i and j be numbers in the domain of $\mathcal{P}$, with $0 < j - i \leqslant 1$. Let s_i and s_j be such that $\mathcal{Q}_{s_i} \in \pi(\mathcal{P}_i)$ and $\mathcal{Q}_{s_j} \in \pi(\mathcal{P}_j)$. Then s_i and s_j are boundedly far apart, since π is coarsely 1-Lipschitz, by Lemma 2.11, and $\mathcal{Q}$ is a quasigeodesic. Therefore, the diameter of $\mathcal{Q}_{[s_i, s_j]}$ is bounded, and we have already noted that $\mathcal{Q}(s_i)$ and $\mathcal{Q}(s_j)$ are close to $\mathcal{P}$, since they are in the image of π. □

Lemma 2.15. *Let $\mathcal{A}$ and $\mathcal{A}'$ be coarsely equivalent subsets of $\mathcal{X}$. Let $\sigma : \mathcal{A} \to \mathcal{A}'$ and $\bar{\sigma} : \mathcal{A}' \to \mathcal{A}$ be C-coarse maps such that $d(a, \sigma(a)) \leqslant C$ for all $a \in \mathcal{A}$ and $d(a', \bar{\sigma}(a')) \leqslant C$ for all $a' \in \mathcal{A}'$. Then, $\pi_{\mathcal{A}} : \mathcal{X} \to \mathcal{A}$ is strongly contracting if and only if $\pi_{\mathcal{A}'} := \sigma \circ \pi_{\mathcal{A}} : \mathcal{X} \to \mathcal{A}'$ is strongly contracting.*

Proof. Suppose $\pi_{\mathcal{A}}$ is $(1, C)$-contracting and $d(x, \pi(x)) - d(x, \mathcal{A}) \leqslant C$ for all $x \in \mathcal{X}$. If $d(x, y) \leqslant d(x, \mathcal{A}') - 2C \leqslant d(x, \mathcal{A}) - C$, then $d^{\pi}_{\mathcal{A}'}(x, y) \leqslant d^{\pi}_{\mathcal{A}}(x, y) + 2C \leqslant 3C$, so $\pi_{\mathcal{A}'}$ is $(1, 3C)$-contracting.

Take a point x and let $a' \in \mathcal{A}'$ such that $d(x, \mathcal{A}') = d(x, a')$. Then,

$$d(x, \bar{\sigma}(a')) - C \leqslant d(x, a') \leqslant d(x, \pi_{\mathcal{A}'}(x)) \leqslant d(x, \pi_{\mathcal{A}}(x)) + 2C,$$

so $d(x, \bar{\sigma}(a')) \leqslant d(x, \mathcal{A}) + 3C$. By Proposition 2.9, $\pi_{\mathcal{A}}$ is strongly constricting, so by Lemma 2.8, there is a constant D such that $d(\pi_{\mathcal{A}}(x), \bar{\sigma}(a')) \leqslant 3C + D$. Thus, $\pi_{\mathcal{A}'}$ is $(5C + D)$-coarsely a closest point projection, hence, strongly contracting. □

Lemma 2.16. *Let $\pi : \mathcal{X} \to \mathcal{A}$ be strongly constricting. There exists a number K such that if $d(\mathcal{A}, g\mathcal{A}) > K$ then* $\operatorname{diam} \pi(g\mathcal{A})$ *is bounded, independent of g.*

Proof. Let π be $(1, C)$-strongly constricting. By Proposition 2.9, π is strongly contracting, so by Corollary 2.13 there is a constant K such that a geodesic with endpoints in $\mathcal{A}$ stays in the $(K - C)$-neighborhood of $\mathcal{A}$. Therefore, a geodesic with endpoints in $g\mathcal{A}$ stays in $\bar{\mathcal{B}}_{K-C}(g\mathcal{A})$. Choose $x \in g\mathcal{A}$ such that $d(x, \mathcal{A}) = d(g\mathcal{A}, \mathcal{A})$. For all $y \in g\mathcal{A}$, if $d^{\pi}(x, y) > C$, then a geodesic from x to y passes within distance C of $\pi(x)$ and $\pi(y)$. This means $\bar{\mathcal{B}}_C(\mathcal{A}) \cap \bar{\mathcal{B}}_{K-C}(g\mathcal{A}) \neq \varnothing$, so $d(\mathcal{A}, g\mathcal{A}) \leqslant K$. Thus, if $d(\mathcal{A}, g\mathcal{A}) > K$, then $d^{\pi}(x, y) \leqslant C$, so $\operatorname{diam} \pi(g\mathcal{A}) \leqslant 2C$. □

2C. ***Strongly contracting elements.*** We have defined contraction and constriction for maps. We now give definitions for group elements:

Definition 2.17. An element $h \in G$ is called *contracting*, with respect to $G \curvearrowright \mathcal{X}$, if $i \mapsto h^i . o$ is a quasigeodesic and if there exists a subset $\mathcal{A} \subset \mathcal{X}$ on which $\langle h \rangle$ acts cocompactly and a map $\pi : \mathcal{X} \to \mathcal{A}$ that is contracting.

An element $h \in G$ is called *constricting*, with respect to $G \curvearrowright \mathcal{X}$, if $i \mapsto h^i . o$ is a quasigeodesic and if there exists a subset $\mathcal{A} \subset \mathcal{X}$ on which $\langle h \rangle$ acts cocompactly, a G-invariant path system $\mathcal{PS}$, and a map $\pi : \mathcal{X} \to \mathcal{A}$ that is $\mathcal{PS}$-constricting.

An element is *strongly contracting* or *strongly constricting* if the projection π is, respectively, strongly contracting or strongly constricting.

For π and $\mathcal{A}$ as in the definition, Proposition 2.9 says π is strongly contracting if and only if it is strongly constricting. Thus, Lemma 2.8 says closest point projection to $\mathcal{A}$ is coarsely well defined and coarsely equivalent to π. Lemma 2.15 says that the choice of the set $\mathcal{A}$ only affects the constants of strong contraction. It follows that an element h is strongly contracting if and only if $i \mapsto h^i . o$ is a quasigeodesic and closest point projection to $\langle h \rangle . o$ is strongly contracting. In the remainder of this section we produce more finely tailored choices for $\mathcal{A}$ and π. In particular, we would like π to be compatible with the group action; see Remark 2.22.

Proposition 2.18 (compare [Dahmani et al. 2011, Lemma 6.5]). *Let G be a finitely generated group, and let $\mathcal{X}$ be a G-space. Let $h \in G$ be an infinite order element. If there exists a strongly constricting $\pi : \mathcal{X} \to \langle h \rangle . o$, then $E(h) = H$, where*

$$H := \{g \in G \mid g\langle h \rangle . o \text{ is coarsely equivalent to } \langle h \rangle . o\}.$$

Proof. H is a group containing every finite index supergroup of $\langle h \rangle$. Let D be the constant of Lemma 2.16, and let $S := \{g \in \mathcal{G} \mid d(g\langle h \rangle . o, \langle h \rangle . o) \leqslant D\}$. Then, Lemma 2.16 implies $H \subset S$. Since $G \curvearrowright \mathcal{X}$ is properly discontinuous, S is contained in finitely many h-orbits, so $\langle h \rangle < H$ has finite index. Therefore, $E(h)$ exists and is equal to H. □

Definition 2.19. If h is a strongly contracting element, define the *(quasi-)axis* of h, with respect to the basepoint o, to be $\mathcal{H} := E(h) . o$.

Lemma 2.20. *If h is a strongly contracting element, then there exists an $E(h)$-equivariant, strongly contracting coarse map $\pi_{\mathcal{H}} : \mathcal{X} \to \mathcal{H}$.*

Proof. By Proposition 2.9, Lemma 2.8, and Lemma 2.15 any choice of closest point projection map to $\mathcal{H}$ is strongly contracting and coarsely $E(h)$-equivariant, so, by Lemma 1.1, we can replace it by a coarsely equivalent, $E(h)$-equivariant coarse map, which will still be strongly contracting, by Lemma 2.15. □

Definition 2.21. From the projection $\pi_{\mathcal{H}}$ of Lemma 2.20, define strongly contracting projections onto each translate of $\mathcal{H}$ by $\pi_{g\mathcal{H}} : \mathcal{X} \to g\mathcal{H}$, $x \mapsto g . \pi_{\mathcal{H}}(g^{-1} . x)$.

If $g'\mathcal{H} = g\mathcal{H}$ then $g^{-1}g' \in E(h)$, so Lemma 2.20 implies that $\pi_{g'\mathcal{H}}(x) = \pi_{g\mathcal{H}}(x)$ for all $x \in \mathcal{X}$.

Remark 2.22. The projections of Definition 2.21 satisfy $g . \pi_{\mathcal{H}}(x) = \pi_{g\mathcal{H}}(g . x)$ for all $x \in \mathcal{X}$ and $g \in G$.

2D. *Strongly contracting elements and the projection axioms.* Let $h \in G$ be a strongly contracting element with respect to $G \curvearrowright \mathcal{X}$. Let $\mathcal{H}$ be a quasi-axis of h defined in Definition 2.19. We wish to apply Theorem 1.11 to the collection of G-translates of $\mathcal{H}$ with the projections of Definition 2.21. To see that the hypotheses of the theorem are satisfied, we first embed $\mathcal{H}$ into a geodesic metric space and then verify the projection axioms of Definition 1.10.

Choose representatives $1 = g_0, \dots, g_{n-1}$ for $\langle h \rangle \backslash E(h)$, so that for each i we have $d(g_i . o, o) = \min_{g \in \langle h \rangle g_i} d(g . o, o)$. Let $g_n := h$. Let $\hat{H}$ be the Cayley graph of $E(h)$ with respect to the generating set $\{g_1, \dots, g_n\}$. The graph $\hat{H}$ becomes a geodesic metric space by assigning each edge length one, and it is a quasitree since $E(h)$ is virtually cyclic.

Choose representatives $1 = f_0, f_1, \dots$ for $G/E(h)$. Let $\mathbb{Y}$ be a disjoint union of copies of $\hat{H}$, one for each coset $f_i E(h) \in G/E(h)$, denoted $f_i \hat{H}$. The orbit map $f_i\hat{H} \to f_i\mathcal{H}$, defined by $f_ie \mapsto f_ie\,.\,o$, is a quasi-isometric embedding, so its inverse $\phi_{f_i\mathcal{H}} : f_i\mathcal{H} \to f_i\hat{H}$ is a coarse map that is a quasi-isometry. Define $\pi_{f_i\hat{H}}(f_j\hat{H}) := \phi_{f_i}(\pi_{f_i\mathcal{H}}(f_j\mathcal{H}))$. Since ϕ_{f_i} is a quasi-isometry, it suffices to check the projection axioms on translates of $\mathcal{H}$ in $\mathcal{X}$.

Lemma 2.23 (Axiom (P0)). *There is a uniform bound on the diameter of $\pi_{\mathcal{H}}(g\mathcal{H})$ for $g \notin E(h)$.*

Proof. Let $\pi_{\mathcal{H}} : \mathcal{X} \to \mathcal{H}$ be $(1, C')$-strongly constricting. Let $\mathcal{Q} : \mathbb{R} \twoheadrightarrow \mathcal{H}$ be an (M, C'')-quasigeodesic parametrization that agrees with $i \mapsto h^i\,.\,o$ on the integers. Replace C' and C'' by $C := \max\{C', C''\}$.

Let $D := \operatorname{diam}\langle h\rangle\backslash\mathcal{H}$. Let K be large enough such that whenever $\mathcal{P}$ is a geodesic with $d(\mathcal{P}_{s_0}, \mathcal{Q}_{t_0}) \leqslant C$ and $d(\mathcal{P}_{s_1}, \mathcal{Q}_{t_1}) \leqslant C$, we have $\mathcal{P}_{[s_0,s_1]} \subset \overline{\mathcal{B}}_K(\mathcal{Q}_{[t_0,t_1]})$ and $\mathcal{Q}_{[s_0,s_1]} \subset \overline{\mathcal{B}}_K(\mathcal{P}_{[t_0,t_1]})$, as in Corollary 2.13 and Lemma 2.14.

Suppose $g \notin E(h)$. For a pair of points $x_0, x_1 \in g\mathcal{H}$, take t_0 and t_1 such that $\mathcal{Q}_{t_i} \in \pi_{\mathcal{H}}(x_i)$ for each i. Let $\mathcal{P}$ be a geodesic connecting x_0 to x_1. If $d^{\pi}_{\mathcal{H}}(x_0, x_1) > C$, then for each i there exists s_i such that $d(\mathcal{P}_{s_i}, \mathcal{Q}_{t_i}) \leqslant C$.

Now, $\mathcal{Q}_{[t_0,t_1]}$ is K-close to $\mathcal{P}_{[s_0,s_1]}$, which in turn is K-close to a subinterval of $g\mathcal{H}$. Therefore, for each integer $i \in [t_0, t_1]$ there is an integer α_i such that $d(h^i\,.\,o,\ gh^{\alpha_i}g^{-1}\,.\,o) \leqslant 2K + D$.

If, for some $i \neq j$, the equation $h^{-i}gh^{\alpha_i}g^{-1}.\,o = h^{-j}gh^{\alpha_j}g^{-1}.\,o$ is satisfied, then $h^{j-i} = gh^{\alpha_j-\alpha_i}g^{-1}$, which implies $\langle h\rangle$ and $\langle ghg^{-1}\rangle$ are commensurable. However, this would imply $g \in E(h)$, contrary to hypothesis. Therefore, for each integer i in $[t_0, t_1]$ we get a distinct point $h^{-i}gh^{\alpha_i}g^{-1}\,.\,o \in \overline{\mathcal{B}}_{2K+D}(o)$. Since the action of G is properly discontinuous, the number of orbit points in $\overline{\mathcal{B}}_{2K+D}(o)$ is finite, so $\operatorname{diam}\pi_{\mathcal{H}}(g\mathcal{H})$ is bounded, independent of g. $\square$

Lemma 2.24 (Axiom (P1)). *For all sufficiently large ξ and for any $X, Y, Z \in \mathbb{Y}$, at most one of $d^{\pi}_X(Y, Z)$, $d^{\pi}_Y(X, Z)$, and $d^{\pi}_Z(X, Y)$ is greater than ξ.*

Proof. Suppose π_Y is $(1, C)$-strongly constricting. Let ξ' be the constant from Lemma 2.23. Let $\xi \geqslant 2\xi' + 14C$. Suppose that $d^{\pi}_X(Y, Z) > \xi$. We show $d^{\pi}_X(Y, Z) \leqslant \xi$; the inequality $d^{\pi}_Z(X, Y) \leqslant \xi$ follows by a similar argument.

Take any point $z \in Z$, and let $y \in Y$ be a point such that $d(z, y) = d(z, Y)$. Let $\mathcal{L} : [0, T] \to \mathcal{X}$ be a geodesic from z to y. For every point of $\mathcal{L}$, y is the closest point of Y. By Lemma 2.8, $\pi_Y(\mathcal{L}) \subset \overline{\mathcal{B}}_{5C}(y)$. Now, $d^{\pi}_X(Y, Z) > \xi$ implies $d^{\pi}_X(\mathcal{L}_0, \mathcal{L}_T) > C$, so there are $z' \in \mathcal{L}$ and $x \in X$ with $d(x, z') \leqslant D$. By Lemma 2.11, π_Y is $8C$-coarsely 1–Lipschitz, which means $d^{\pi}_Y(x, z') \leqslant 9C$. Thus,

$$d^{\pi}_Y(X, Z) \leqslant 2\xi' + d^{\pi}_Y(x, z) \leqslant 2\xi' + 5C + d^{\pi}_Y(x, z') \leqslant 2\xi' + 14C \leqslant \xi. \qquad \square$$

Lemma 2.25 (Axiom (P2)). *For all sufficiently large ξ and for all $X, Y \in \mathbb{Y}$, the set $\{V \in \mathbb{Y} \mid d^{\pi}_V(X, Y) > \xi\}$ is finite.*

Proof. Let ξ' be the constant of Lemma 2.23. Suppose $\pi_{\mathcal{H}}$ is $(1, C)$-strongly constricting. Let $\xi > C+2\xi'$. Take arbitrary $X, Y \in \mathbb{Y}$, and let $\mathcal{L}$ be a geodesic from some point in $\pi_X(Y)$ to some point in $\pi_Y(X)$. If $d^\pi_V(X, Y) > \xi$, then $d^\pi_V(\mathcal{L}_0, \mathcal{L}_T) > C$, so $\mathcal{L}$ comes within distance C of V. By proper discontinuity of the action, there are only finitely many elements of $\mathbb{Y}$ that come within distance C of the finite geodesic $\mathcal{L}$. □

Notation 2.26. Let $\mathcal{Y}$ be the quasitree produced by Theorem 1.11 from $\mathbb{Y}$, and let $\star \in \mathcal{Y}$ be the vertex corresponding to $o \in \mathcal{X}$. Furthermore, let $\hat{\pi}_{g\hat{H}} : \mathcal{Y} \to g\hat{H}$ be closest point projection to the isometrically embedded copy of $g\hat{H}$ in $\mathcal{Y}$, which the theorem says coarsely agrees with $\pi_{g\hat{H}}$.

Definition 2.27. Define uniform quasi-isometric embeddings $\phi_{g\mathcal{H}} : g\mathcal{H} \to \mathcal{Y}$ for each translate $g\mathcal{H}$ of $\mathcal{H}$ by sending $g\mathcal{H}$ to $f_i\hat{H}$ via ϕ_{f_i}, where $g \in f_iE(h)$, and post-composing by the isometric embedding of $f_i\hat{H}$ into $\mathcal{Y}$ provided by Theorem 1.11.

Proposition 2.28. *If there is a strongly contracting element for $G \curvearrowright \mathcal{X}$, then G has nonzero growth exponent.*

Proof. [Bestvina et al. 2014, Proposition 3.23] says that G contains a free subgroup, so it has exponential growth. □

3. Abundance of strongly contracting elements

In this section we show that strongly contracting elements are abundant.

Proposition 3.1. *If G contains a strongly contracting element for $G \curvearrowright \mathcal{X}$, then so does every infinite normal subgroup.*

In effect, the proposition reduces the problem of growth tightness for arbitrary quotients of G to quotients by the normal closure of a strongly contracting element.

Given a strongly contracting element $h \in \mathcal{G}$ and an infinite normal subgroup Γ of G we find an element $g \in \Gamma$ such that $f := gh^ng^{-1}h^{-n} \in \Gamma$ is strongly contracting for all sufficiently large n. To prove f is strongly contracting we follow a standard strategy by showing that an axis for f has "long" ($\asymp n$) segments in contracting sets, separated by "short" ($= d(o, g\,.\,o)$) hops between such segments. For each $x \in \mathcal{X}$ there is, coarsely, a unique one of these segments such that the projection of x transitions from landing at the end of the segment to landing at the beginning of the segment. We use this transition point to define the projection to the f-axis, and verify that this projection is strongly contracting.

We first prove some preliminary lemmas.

Lemma 3.2. *Let $h \in G$ be an infinite order element and let $\pi : \mathcal{X} \to \langle h\rangle\,.\,o$ be a contracting coarse map such that $d(x, \pi(x)) - d(x, \mathcal{A})$ is uniformly bounded. Then, $i \mapsto h^i\,.\,o$ is a quasigeodesic.*

Proof. Take any $\alpha < \beta$ in $\mathbb{Z}$. By the triangle inequality, $d(h^{\alpha}.o,\ h^{\beta}.o) \stackrel{*}{\prec} (\beta-\alpha)$. We now prove the opposite inequality. Let $\mathcal{L}\colon [0,T] \to \mathcal{X}$ be a geodesic from $h^{\alpha}.o$ to $h^{\beta}.o$. By Corollary 2.13, there exists a D such that for every $i \in [0,T]\cap\mathbb{Z}$ there exists an $\alpha \leqslant \alpha_i \leqslant \beta$ such that $d(\mathcal{L}_i, h^{\alpha_i}.o) \leqslant D$. Since the action of G on $\mathcal{X}$ is properly discontinuous, there exists a maximum γ such that $d(o, h^{\gamma}.o) \leqslant 2D+1$, so $\alpha_{i+1}-\alpha_i \leqslant \gamma$ for all i. Setting $\alpha_0 := \alpha$ and $\alpha_{\lceil T\rceil} := \beta$, we have

$$\beta-\alpha = \sum_{i=0}^{\lceil T\rceil-1} \alpha_{i+1}-\alpha_i \leqslant \gamma\lceil T\rceil \leqslant \gamma\big(d(h^{\alpha}.o,\ h^{\beta}.o)+1\big). \qquad \square$$

Fix a strongly contracting element h, and let $\mathcal{Y}$ be the quasitree of Notation 2.26, with bottleneck constant Δ.

Lemma 3.3. *There exists $K \geqslant 0$ such that $d^{\pi}_{\mathcal{H}}(o, g_1.o) - d^{\pi}_{\mathcal{H}}(g_1.o, g_0.o) \geqslant K$ implies that $g_0.\star$ and $g_1.\star$ are contained in the same component of $\mathcal{Y}\setminus\overline{\mathcal{B}}_{\Delta}(\star)$.*

Proof. Let $D := \operatorname{diam}\langle h\rangle\backslash\hat{H}$ in $\mathcal{Y}$. For each $i \in \{0,1\}$, choose an m_i such that $d(h^{m_i}.\star,\ \hat{\pi}_{\hat{H}}(g_i.\star)) \leqslant D$. Choose a geodesic $\mathcal{L}$ from $\star$ to $h.\star$. Take $M > 0$ such that $h^m.\mathcal{L}\cap\overline{\mathcal{B}}_{\Delta}(\star) = \varnothing$ when $|m| \geq M$.

For each i, $|m_i| \asymp d(h^{m_i}.\star,\ \star) \succ K$, so for sufficiently large K, we have the lower bound $d(h^{m_i}.\star,\ \star) > 2\Delta + D$ and $|m_i| > M$. Furthermore, m_0 and m_1 must have the same sign if K is large enough: by Lemma 2.14, the interval of $\mathcal{H}$ between $h^{m_0}.o$ and $h^{m_1}.o$ stays close to a geodesic between $h^{m_0}.o$ and $h^{m_1}.o$, so if m_0 and m_1 have different signs, then

$$\begin{aligned} d^{\pi}_{\mathcal{H}}(g_0.o,\ g_1.o) &\stackrel{+}{\asymp} d(h^{m_0}.o,\ h^{m_1}.o) \\ &\stackrel{+}{\asymp} d(o,\ h^{m_0}.o) + d(o,\ h^{m_1}.o) \stackrel{+}{\asymp} d^{\pi}_{\mathcal{H}}(o,\ g_0.o) + d^{\pi}_{\mathcal{H}}(o,\ g_1.o). \end{aligned}$$

However, $d^{\pi}_{\mathcal{H}}(g_0.o,\ g_1.o) \leqslant d^{\pi}_{\mathcal{H}}(o,\ g_1.o) - K$, so this would imply

$$K \stackrel{+}{\prec} d^{\pi}_{\mathcal{H}}(o,\ g_0.o) \stackrel{+}{\prec} -K,$$

which is false for sufficiently large K.

No geodesic between $g_i.\star$ and $h^{m_i}.\star$ enters $\overline{\mathcal{B}}_{\Delta}(\star)$, since this would imply:

$$d(h^{m_1}.\star,\ \star) \leqslant 2\Delta + D.$$

For $\min\{m_0, m_1\} \leqslant m \leqslant \max\{m_0, m_1\} - 1$, the geodesic $h^m.\mathcal{L}$ stays outside $\overline{\mathcal{B}}_{\Delta}(\star)$ since m_0 and m_1 have the same sign and magnitude at least M, which implies that $|m| \geqslant M$. By concatenating such geodesics, we construct a path from $g_0.\star$ to $g_1.\star$ in $\mathcal{Y}\setminus\overline{\mathcal{B}}_{\Delta}(\star)$. $\square$

Corollary 3.4. *There exists an $N > 0$ such that for all $n \geqslant N$ the points $h^n.\star$ and $h^N.\star$ are in the same component of $\mathcal{Y}\setminus\overline{\mathcal{B}}_{\Delta}(\star)$.*

Proof. Take N large enough so that $d^{\pi}_{\mathcal{H}}(o,\ h^n.o) \geqslant K + d(o,\ h.o) + 2C$ for all $n \geqslant N$. Then, $d^{\pi}_{\mathcal{H}}(o,\ h^{n+1}.o) - d^{\pi}_{\mathcal{H}}(h^n.o,\ h^{n+1}.o) \geqslant K$. Apply Lemma 3.3. □

Definition 3.5. Call the component of $\mathcal{Y} \setminus \mathcal{B}_{\Delta}(g.\star)$ containing $gh^n.\star$ for all sufficiently large n the gh^{∞} *component* and the component containing $gh^{-n}.\star$ for all sufficiently large n the $gh^{-\infty}$ *component*.

Lemma 3.6. *For some $K \geqslant 0$, suppose g_0 and g_1 are elements of G such that $g_0\mathcal{H} \neq g_1\mathcal{H}$ and $d^{\pi}_{g_0\mathcal{H}}(g_0.o,\ g_1.o) \leqslant K$ and $d^{\pi}_{g_1\mathcal{H}}(g_0.o,\ g_1.o) \leqslant K$. Then, there exists an $N > 0$ such that for all $n \geqslant N$, $\epsilon_0, \epsilon_1 \in \{\pm 1\}$, and $f_0, f_1 \in \{g_0, g_1\}$,*

- *the balls $\overline{\mathcal{B}}_{\Delta}(f_0h^{\epsilon_0 n/2}.\star)$ and $\overline{\mathcal{B}}_{\Delta}(f_1h^{\epsilon_1 n/2}.\star)$ in $\mathcal{Y}$ are disjoint unless $f_0 = f_1$ and $\epsilon_0 = \epsilon_1$,*
- *$f_0.\star$ and $f_1.\star$ are in the $f_0h^{-\epsilon_0\infty}$ component of $\mathcal{Y} \setminus \overline{\mathcal{B}}_{\Delta}(f_0h^{\epsilon_0 n/2}.\star)$, and*
- *$f_0h^{\epsilon_0 n}.\star$ and $f_0h^{\epsilon_0 n}f_1.\star$ are in the $f_0h^{\epsilon_0\infty}$ component of $\mathcal{Y} \setminus \overline{\mathcal{B}}_{\Delta}(f_0h^{\epsilon_0 n/2}.\star)$.*

Proof. $\overline{\mathcal{B}}_{\Delta}(f_0h^{n/2}.\star)$ and $\overline{\mathcal{B}}_{\Delta}(f_0h^{-n/2}.\star)$ are disjoint for all sufficiently large n since $i \mapsto h^i.\star$ is a quasigeodesic. In the other cases, $f_0\mathcal{H}$ and $f_1\mathcal{H}$ are distinct axes, so $f_0\hat{H}$ and $f_1\hat{H}$ are disjoint. For each $i \in \{0, 1\}$, the bounds

$$d^{\pi}_{f_i\mathcal{H}}(f_i.o,\ f_{1-i}.o) \leqslant K$$

imply that the closest point projection $\hat{\pi}_{f_i\hat{H}}(f_{1-i}\hat{H})$ of $f_{1-i}\hat{H}$ to $f_i\hat{H}$ is contained in a bounded neighborhood of $f_i.\star$. For any point $y.\star \in \overline{\mathcal{B}}_{\Delta}(f_1h^{\epsilon_1 n/2}.\star) \setminus f_1\hat{H}$, we have that $\hat{\pi}_{f_1\hat{H}}(y\hat{H})$ is 2Δ-close to $f_1h^{\epsilon_1 n/2}.\star$. Therefore,

$$d^{\pi}_{f_1\hat{H}}(f_0\hat{H}, y\hat{H}) \stackrel{+}{\asymp} d(f_1.\star,\ f_1h^{\epsilon_1 n/2}.\star) \asymp n,$$

so for n sufficiently large we can make $d^{\pi}_{f_1\hat{H}}(f_0\hat{H}, y\hat{H})$ larger then the constant ξ of projection axiom (P1), which implies $d^{\pi}_{f_0\hat{H}}(f_1\hat{H}, y\hat{H}) \leqslant \xi$. On the other hand, $\overline{\mathcal{B}}_{\Delta}(f_0h^{\epsilon_0 n/2}.\star)$ projects close to $f_0h^{\epsilon_0 n/2}.\star$ in $f_0\hat{H}$, so for large enough n the balls have disjoint projections, which means the balls are disjoint.

For the second statement, suppose N is large enough so that for all $n \geqslant N$ we have $d(o, h^{n/2}.o) \geqslant K' + K + 2C$, where K' is the constant of Lemma 3.3. Then,

$$d^{\pi}_{f_0\mathcal{H}}(f_0h^{\epsilon_0 n/2}.o,\ f_0.o) - d^{\pi}_{f_0\mathcal{H}}(f_0.o,\ f_1.o) \geqslant K',$$

so Lemma 3.3 implies $f_0.\star$ and $f_1.\star$ are in the same component of $\mathcal{Y} \setminus \overline{\mathcal{B}}_{\Delta}(f_0h^{\epsilon_0 n/2}.\star)$. If, in addition, N is at least twice the constant of Corollary 3.4, then this is the $f_0h^{-\epsilon_0\infty}$ component.

The proof of the third statement is similar. □

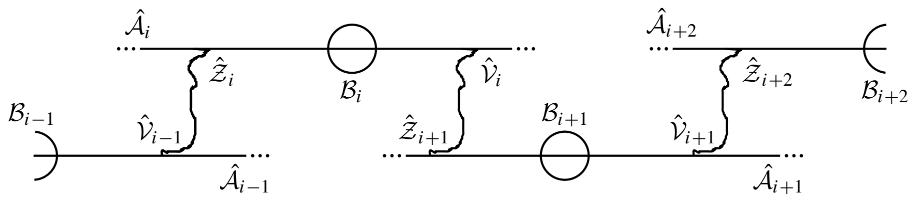


Figure 1. Disjoint balls in $\mathcal{Y}$.

Proof of Proposition 3.1. Strongly constricting is the same as strongly contracting, by Proposition 2.9, so suppose h is a $(1, C)$-strongly constricting element. By Lemma 2.8, there exists a D such that $\pi_{\mathcal{H}}$ is D-coarsely equivalent to closest point projection. Recall that $D > C$. By Lemma 2.11, there exists a D' such that $\pi_{\mathcal{H}}$ is D'-coarsely 1-Lipschitz.

Let Γ be an infinite normal subgroup of G. Every infinite order element of $E(h)$ is strongly contracting, so if Γ contains such an element then we are done. Otherwise, $\Gamma \cap E(h)$ is finite. Since Γ is infinite, there exists an element $g \in \Gamma$ such that $g \notin E(h)$. We claim that for sufficiently large n the element $f := gh^n g^{-1} h^{-n} \in \Gamma$ is strongly constricting.

For brevity, let $f^{i+1/2}$ denote $f^i gh^n$. Let $\hat{\mathcal{A}}_i := f^{i/2}\hat{H}$ and $\mathcal{A}_i := f^{i/2}\mathcal{H}$. Define $\mathcal{B}_0 := \overline{\mathcal{B}}_\Delta(h^{n/2}.\star)$, $\mathcal{B}_1 := \overline{\mathcal{B}}_\Delta(f^{1/2}h^{-n/2}.\star)$, and $\mathcal{B}_{2k+i} := f^k\mathcal{B}_i$ for $k \in \mathbb{Z}$. Let $\hat{\mathcal{Z}}_i := f^{i/2}h^{(-1)^i n}.\star \in \mathcal{Y}$ and $\mathcal{Z}_i := f^{i/2}h^{(-1)^i n}.o \in \mathcal{X}$. Let $\hat{\mathcal{V}}_i := f^{i/2}.\star \in \mathcal{Y}$ and $\mathcal{V}_i := f^{i/2}.o \in \mathcal{X}$. See Figure 1.

By repeated applications of Lemma 3.6, for large enough n, the balls $\mathcal{B}_i$ are pairwise disjoint. There are two orbits of these balls under the f-action, so f is an infinite order element. Furthermore, the balls are linearly ordered by separation, consistent with the subscripts, since for all i we have that $\mathcal{B}_j$ is contained in the $f^{i/2}h^{(-1)^{i+1}\infty}$ component of $\mathcal{Y}\setminus\mathcal{B}_i$ for all $j > i$, and in the $f^{i/2}h^{(-1)^i\infty}$ component for all $j < i$.

For any i and any $j < i-1$ the ball $\mathcal{B}_{i-1}$ separates $\hat{\mathcal{A}}_j$ from $\hat{\mathcal{A}}_i$ in $\mathcal{Y}$, so $\hat{\pi}_{\hat{\mathcal{A}}_i}(\hat{\mathcal{A}}_j)$ is contained in a bounded neighborhood of $\hat{\pi}_{\hat{\mathcal{A}}_i}(\hat{\mathcal{A}}_{i-1})$, which in turn we know is contained in a bounded neighborhood of $\hat{\mathcal{Z}}_i$. Conversely, $\hat{\pi}_{\hat{\mathcal{A}}_i}(\hat{\mathcal{A}}_j)$ is contained in a bounded neighborhood of $\hat{\mathcal{V}}_i$ for $j > i$. Since $\hat{\pi}_{\hat{\mathcal{A}}_i}$ agrees with $\pi_{\mathcal{A}_i}$ up to bounded error, the same statements are true for the axes in $\mathcal{X}$. That is, there exists a K, independent of n, such that for all i,

- $d^\pi_{\mathcal{A}_i}(\mathcal{Z}_i, \mathcal{A}_j) \leqslant K$ if $j < i$, and
- $d^\pi_{\mathcal{A}_i}(\mathcal{V}_i, \mathcal{A}_j) \leqslant K$ if $j > i$.

Define $K' := 2K + C + 2D + D'$.

Suppose that for some $x \in \mathcal{X}$ there exists an i such that $d^\pi_{\mathcal{A}_i}(x, \mathcal{V}_i) > K'$. Then for any $j > i$ we have $d(\pi_{\mathcal{A}_i}(x), \pi_{\mathcal{A}_i}(\mathcal{A}_j)) > D > C$. Let y be a point of $\mathcal{A}_j$ closest

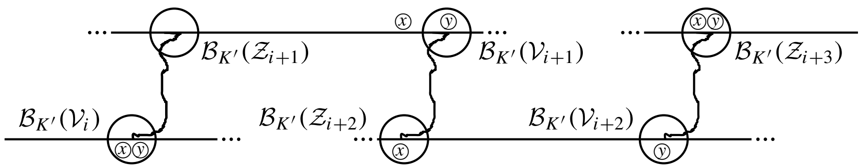


Figure 2. Projections x of x and y of y to each axis.

to x. On any given geodesic from x to y, there is a point $z \in \overline{\mathcal{B}}_{C+K}(\mathcal{V}_i)$, since $d^{\pi}_{\mathcal{A}_i}(x, y) > C$. Now $\pi_{\mathcal{A}_j}$ is D-coarsely equivalent to closest point projection, and y is closest to both x and z, so $d^{\pi}_{\mathcal{A}_j}(x, z) \leqslant 2D$. However, z is $(C+K)$-close to $\mathcal{V}_i$, and $d^{\pi}_{\mathcal{A}_j}(\mathcal{V}_i, \mathcal{Z}_j) \leqslant K$, so $d^{\pi}_{\mathcal{A}_j}(x, \mathcal{Z}_j) \leqslant 2D + C + K + D' + K = K'$.

We have shown that $d^{\pi}_{\mathcal{A}_i}(x, \mathcal{V}_i) > K'$ implies $d^{\pi}_{\mathcal{A}_j}(x, \mathcal{Z}_j) \leqslant K'$ for all $j > i$. A similar argument shows that $d^{\pi}_{\mathcal{A}_i}(x, \mathcal{Z}_i) > K'$ implies $d^{\pi}_{\mathcal{A}_j}(x, \mathcal{V}_j) \leqslant K'$ for all $j < i$.

Assume that n is large enough so that

$$d^{\pi}_{\mathcal{A}_0}(\mathcal{Z}_0, \mathcal{V}_0) = d^{\pi}_{\mathcal{H}}(h^n . o, o) > 2K' + 2C + 2D + d(o,\ g . o).$$

Define $\mathcal{F} := \bigcup_{i \in \mathbb{Z}} \{\mathcal{V}_i\}$. We wish to define $\pi_{\mathcal{F}} : \mathcal{X} \to \mathcal{F}$ by sending a point x to the point $\mathcal{V}_\alpha$ where α is the greatest integer such that $d^{\pi}_{\mathcal{A}_\alpha}(x, \mathcal{V}_\alpha) \leqslant K'$, but we must verify that such an α exists. Fix an $x \in \mathcal{X}$, and suppose that $\iota \in \mathbb{Z}$ is such that $d(x, \mathcal{A}_\iota) = \min_{j \in \mathbb{Z}} d(x, \mathcal{A}_j)$. Such an ι exists since the action is properly discontinuous. Suppose that $d^{\pi}_{\mathcal{A}_\iota}(x, \mathcal{V}_\iota) \leqslant K'$. By the assumption on n, $d^{\pi}_{\mathcal{A}_\iota}(x, \mathcal{Z}_\iota) > K'$, so $d^{\pi}_{\mathcal{A}_j}(x, \mathcal{V}_\iota) \leqslant K'$ for all $j < \iota$. A brief computation shows that

$$d^{\pi}_{\mathcal{A}_{\iota+1}}(x, \mathcal{Z}_{\iota+1}) \leqslant d(x, \mathcal{A}_{\iota+1}) + d(o,\ g . o) + K' + 2C + D.$$

By Lemma 2.8, $d(\mathcal{Z}_{\iota+1}, \pi_{\mathcal{A}_{\iota+1}}(x)) \leqslant d(o, g . o) + K' + 2C + 2D$, which, again by our assumption on n, implies $d^{\pi}_{\mathcal{A}_{\iota+1}}(x, \mathcal{V}_{\iota+1}) > K'$. We conclude that $\alpha \leqslant \iota$. The previous paragraph then tells us that $d^{\pi}_{\mathcal{A}_j}(x, \mathcal{Z}_j) \leqslant K'$ for all $j > \alpha + 1$.

Now suppose x and y are points with $\pi_{\mathcal{F}}(x) = \mathcal{V}_i$ and $\pi_{\mathcal{F}}(y) = \mathcal{V}_j$ for $j > i + 1$. Then for each $i + 2 \leqslant k \leqslant j$, we have $d^{\pi}_{\mathcal{A}_k}(x, y) \geqslant d^{\pi}_{\mathcal{A}_k}(\mathcal{Z}_k, \mathcal{V}_k) - 2K' > C$. Figure 2 depicts a situation with $j = i + 2$ that shows $j > i + 1$ is necessary, since the projections to $\mathcal{A}_{i+1}$ may be close. By the strong constriction property for each $\mathcal{A}_k$, every geodesic from x to y passes $(C + K')$-close to $\mathcal{Z}_k$ and $\mathcal{V}_k$. So every geodesic passes within $C + K'$ of $\pi_{\mathcal{F}}(y) = \mathcal{V}_j$ and within $C + K'$ of $\mathcal{Z}_{i+2}$, which is boundedly close to $\pi_{\mathcal{F}}(x) = \mathcal{V}_i$.

Therefore, $\pi_{\mathcal{F}}$ is $(1, \max\{d(\mathcal{V}_0, \mathcal{V}_2),\ C + K' + d(\mathcal{V}_0, \mathcal{Z}_2)\})$-strongly constricting. Lemma 3.2 says $i \mapsto f^i . o$ is a quasigeodesic, so $f \in \Gamma$ is a strongly contracting element. $\square$

4. A minimal section

Let $\mathcal{X}$ be a G-space with basepoint o. Suppose that there exists a strongly contracting element for $G \curvearrowright \mathcal{X}$. Let Γ be an infinite normal subgroup of G. By Proposition 3.1, there exists a strongly contracting element $h \in \Gamma$. Let $\mathcal{H} = E(h) \,.\, o$ be an axis for h, and define equivariant projections to translates of $\mathcal{H}$ as in Definition 2.21. Suppose $\pi_{\mathcal{H}}$ is a $(1, C)$-strongly constricting C-coarse map.

Definition 4.1. For each element $g\Gamma \in G/\Gamma$ choose an element $\bar{g} \in g\Gamma$ such that $d(o, \bar{g} \,.\, o) = d(o, g\Gamma \,.\, o) = d(\Gamma \,.\, o, g\Gamma \,.\, o)$. Let $\bar{G} := \{\bar{g} \mid g\Gamma \in G/\Gamma\}$. We call $\bar{G}$ a *minimal section*, and let $\bar{\mathcal{G}}$ denote $\bar{G} \,.\, o$.

Observe that $\Theta'_{G/\Gamma}(s) = \Theta'_{\bar{G}}(s)$, so $\delta_{G/\Gamma} = \delta_{\bar{G}}$. The next lemma says, coarsely, that the minimal section is orthogonal to translates of $\mathcal{H}$.

Lemma 4.2. *For every $\bar{g} \in \bar{G}$ and for every $f \in G$ we have $d^{\pi}_{f\mathcal{H}}(o, \bar{g} \,.\, o) \leqslant 8C + D$, where $D := \operatorname{diam}\langle h\rangle\backslash\mathcal{H}$.*

Proof. Suppose not. Then there exists an $n \neq 0$ such that

$$D \geqslant d(\pi_{f\mathcal{H}}(o),\ fh^nf^{-1} \,.\, \pi_{f\mathcal{H}}(\bar{g} \,.\, o)) \geqslant d^{\pi}_{f\mathcal{H}}(o,\ fh^nf^{-1}\bar{g} \,.\, o) - 2C.$$

Thus, $d^{\pi}_{f\mathcal{H}}(o, \bar{g} \,.\, o) - d^{\pi}_{f\mathcal{H}}(o, fh^nf^{-1}\bar{g} \,.\, o) > 6C$. However,

$$\begin{aligned}
&d(o,\ fh^nf^{-1}\bar{g} \,.\, o)\\
&\quad\leqslant d(o,\ \pi_{f\mathcal{H}}(o)) + d^{\pi}_{f\mathcal{H}}(o,\ fh^nf^{-1}\bar{g} \,.\, o) + d(\pi_{f\mathcal{H}}(fh^nf^{-1}\bar{g} \,.\, o),\ fh^nf^{-1}\bar{g} \,.\, o)\\
&\quad< d(o,\ \pi_{f\mathcal{H}}(o)) + d^{\pi}_{f\mathcal{H}}(o,\ \bar{g} \,.\, o) + d(\pi_{f\mathcal{H}}(fh^nf^{-1}\bar{g} \,.\, o),\ fh^nf^{-1}\bar{g} \,.\, o) - 6C\\
&\quad= d(o,\ \pi_{f\mathcal{H}}(o)) + d^{\pi}_{f\mathcal{H}}(o,\ \bar{g} \,.\, o) + d(\pi_{f\mathcal{H}}(\bar{g} \,.\, o),\ \bar{g} \,.\, o) - 6C\\
&\quad\leqslant d(o,\ \bar{g} \,.\, o) \qquad\qquad \text{(by Lemma 2.10)}
\end{aligned}$$

This contradicts minimality of $\bar{G}$, since $fh^nf^{-1}\bar{g} = \bar{g}\bar{g}^{-1}fh^nf^{-1}\bar{g} \in \bar{g}\Gamma$. □

Corollary 4.3. *If $d(\bar{g} \,.\, o,\ \bar{g}' \,.\, o) \geqslant 18C + 2D$ for $\bar{g},\ \bar{g}' \in \bar{G}$, then there is no $f \in G$ such that $\bar{g} \,.\, o \in f\mathcal{H}$ and $\bar{g}' \,.\, o \in f\mathcal{H}$.*

Proof. If there were such an f, we would have $d^{\pi}_{f\mathcal{H}}(\bar{g} \,.\, o,\ \bar{g}' \,.\, o) \geqslant 2(8C + D)$, which means either $\bar{g}$ or $\bar{g}'$ would contradict Lemma 4.2. □

In light of Corollary 4.3, it will be convenient to pass to a coarsely dense subset of $\bar{\mathcal{G}}$ whose elements yield distinct translates of $\mathcal{H}$:

Definition 4.4. Let $K \geqslant 18C + 2D$, and let A be a maximal subset of $\bar{G}$ such that $1 \in A$ and $d(\bar{g} \,.\, o,\ \bar{g}' \,.\, o) \geqslant K$ for all distinct $\bar{g},\ \bar{g}' \in A$. Let $\mathcal{A} := A \,.\, o$.

By maximality, for every $\bar{g} \in \bar{G}$ there is some $a \in A$ such that $d(a\,.\,o,\ \bar{g}\,.\,o) \leqslant K$. There are boundedly many points of $\bar{\mathcal{G}}$ in a ball of radius K, so $\Theta_{\bar{\mathcal{G}}}(s)$ is bounded below by $\Theta_{\mathcal{A}}(s)$ and above by a constant multiple of $\Theta_{\mathcal{A}}(s)$. In particular, $\Theta_{\mathcal{A}}(s)$ has the same convergence behavior as $\Theta_{\bar{\mathcal{G}}}(s)$, so $\delta_A = \delta_{\bar{G}} = \delta_{G/\Gamma}$.

Corollary 4.3 implies $a\mathcal{H} \neq a'\mathcal{H}$ for distinct $a, a' \in A$.

5. Embedding a free product set

Let $A \subset \bar{G}$ as in Definition 4.4, and let $A^* := A \setminus \{1\}$. Consider the free product set $A^* * \mathbb{Z}_2 := \bigcup_{k=1}^{\infty} \{(a_1, \ldots, a_k) \mid a_i \in A^*\}$. For any $n > 0$ we can map the free product set into G by $\phi_n : (a_1, \ldots, a_k) \mapsto a_1 h^n a_2 h^n \cdots a_k h^n$. Our goal is to show $\delta_{\phi_n(A^* * \mathbb{Z}_2)} > \delta_A$. We establish the inequality in the next section. In this section we show ϕ_n is an injection for all sufficiently large n. In fact, we prove something stronger:

Proposition 5.1. *The map $A^* * \mathbb{Z}_2 \to G$, sending $(a_1, \ldots, a_k) \mapsto a_1 h^n \cdots a_n h^n\,.\,o$ is an injection for all sufficiently large n.*

The map is an injection because we have an action of G on the quasitree $\mathcal{Y}$, and for large enough n we have "quasi-edges" of the form $[y, yh^n]$. We have set things up so that the a_i do not backtrack across such edges. See Figure 3. We make this precise:

Proof. Let $\underline{a} = (a_1, \ldots, a_k) \in A^* * \mathbb{Z}_2$. By Lemma 4.2, there is a K such that $d^{\pi}_{f\mathcal{H}}(o,\ \bar{g}\,.\,o) \leqslant K$ for every $f \in G$ and every $\bar{g} \in \bar{G}$. The choice of $A \subset \bar{G}$ in Definition 4.4 guarantees that the axes $a\mathcal{H}$ for $a \in A$ are distinct. Let N be the constant of Lemma 3.6 for this K, and choose $n \geqslant N$.

Note that the proof of Lemma 3.6 includes the fact that

$$d(o,\ h^{n/2}.\,o) \geqslant K' + K + 2C,$$

where K' is the constant of Lemma 3.3. Therefore, if $\phi_n(\underline{a})\,.\,o = \phi_n(\underline{a}')\,.\,o$, then

$$d^{\pi}_{\phi_n(\underline{a})\mathcal{H}}(\phi_n(\underline{a})\,.\,o,\ \phi_n(\underline{a})h^{-n/2}.\,o) - d^{\pi}_{\phi_n(\underline{a})\mathcal{H}}(\phi_n(\underline{a})\,.\,o,\ \phi_n(\underline{a}')\,.\,o) \geq K' + C > K',$$

so Lemma 3.3 implies that $\phi_n(\underline{a})\,.\,\star$ and $\phi_n(\underline{a}')\,.\,\star$, though they might not be equal, are at least contained in the same component of $\mathcal{Y} \setminus \bar{\mathcal{B}}_{\Delta}(\phi_n(\underline{a})h^{-n/2}.\,\star)$.

Define $\mathcal{V}_i(\underline{a})$ to be the $a_1 h^n \cdots a_i h^{\infty}$ component of $\mathcal{Y} \setminus \bar{\mathcal{B}}_{\Delta}(a_1 h^n \cdots a_i h^{n/2}.\,\star)$ for $i \leqslant k$ (recall Definition 3.5). Lemma 3.6 implies that $\mathcal{V}_i(\underline{a}) \supset \mathcal{V}_{i+1}(\underline{a})$ and $\phi_n(\underline{a})\,.\,\star \in \mathcal{V}_k(\underline{a})$. Moreover, for $i \leq \min\{k, k'\}$, $\mathcal{V}_i(\underline{a})$ and $\mathcal{V}_i(\underline{a}')$ are disjoint unless $a_j = a'_j$ for all $j \leqslant i$.

If $\phi_n(\underline{a})\,.\,o = o$, then Lemma 3.3 implies that $\star \in \mathcal{V}_k(\underline{a}) \subset \mathcal{V}_1(\underline{a})$. This contradicts the fact that $\star$ is contained in the $a_1 h^{-\infty}$ component of $\mathcal{Y} \setminus \bar{\mathcal{B}}_{\Delta}(a_1 h^{n/2}.\,\star)$. The same argument shows that if $\underline{a}$ is a proper prefix of $\underline{a}'$, that is, if $\underline{a} = (a_1, \ldots, a_k)$ and $\underline{a}' = (a_1, \ldots, a_k, a'_{k+1}, \ldots, a'_{k'})$ with $k' > k$, then $\phi_n(\underline{a})\,.\,o \neq \phi_n(\underline{a}')\,.\,o$.

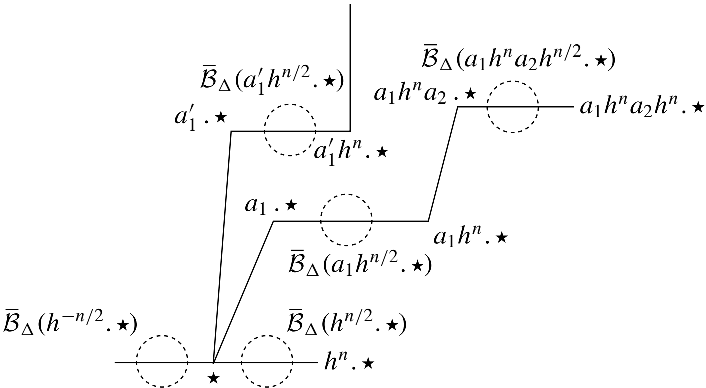


Figure 3. A does not cross h^n quasi-edges.

Suppose $\phi_n(\underline{a}).o = \phi_n(\underline{a}').o$ with $k \leqslant k'$. Lemma 3.3 implies $\phi_n(\underline{a}).\star \in \mathcal{V}_{k'}(\underline{a}')$, so $a_i = a_i'$ for all $i \leqslant k$. Since $\underline{a}$ cannot be a proper prefix of $\underline{a}'$, we have $k = k'$. Hence, $\phi_n(\underline{a}).o = \phi_n(\underline{a}').o$ implies $\underline{a} = \underline{a}'$ for all sufficiently large n. □

6. The growth gap

A free product of groups has greater growth exponent than the factor groups, with respect to a word metric, so we expect that $\phi_n(A^* * \mathbb{Z}_2)$ should have a larger growth exponent than A. To verify this intuition, one must show that the Poincaré series for $\phi_n(A^* * \mathbb{Z}_2)$ diverges at $\delta_A + \epsilon$ for some $\epsilon > 0$. A clever manipulation of Poincaré series yields the following criterion:

Lemma 6.1 [Dal'Bo et al. 2011, Criterion 2.4; Sambusetti 2002a, Proposition 2.3]. *If the map*

$$\phi_n : A^* * \mathbb{Z}_2 \to G, \quad (a_1, \dots, a_k) \mapsto a_1h^n \cdots a_kh^n$$

is an injection and $\exp(|h^n| \cdot \delta_A) < \Theta_A(\delta_A)$, *then* $\delta_{\phi_n(A^* * \mathbb{Z}_2)} > \delta_A$.

Because our methods are coarse, we have passed to a high power h^n of h and therefore do not have control over $|h^n|$. However, the criterion is satisfied automatically if A, or, equivalently, $\overline{G}$, is divergent, which, recalling Definition 1.7, means Θ_A diverges at δ_A. The following definition will be used in a condition to guarantee divergence of $\overline{G}$.

Definition 6.2. Let $\mathrm{Comp}^G_{Q,r} \subset G.o$ be the set of points $g.o$ such that there exists a geodesic $[x, y]$ of length r with $x \in \overline{\mathcal{B}}_Q(o)$ and $y \in \overline{\mathcal{B}}_Q(g.o)$, whose interior is contained in $\mathcal{X} \setminus \overline{\mathcal{B}}_Q(G.o)$. Define the *$Q$-complementary growth exponent* of G to be

$$\delta^c_G := \limsup_{r\to\infty} \frac{\log \# \mathrm{Comp}^G_{Q,r}}{r}$$

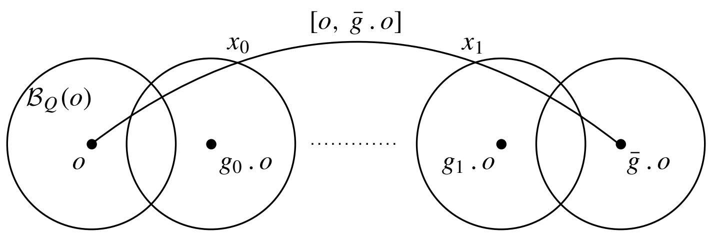


Figure 4. Splitting a geodesic into three subsegments.

Theorem 6.3. *Let G be a finitely generated, nonelementary group. Let $\mathcal{X}$ be a G-space. If G contains a strongly contracting element and there exists a $Q \geqslant 0$ such that the Q-complementary growth exponent of G is strictly less than the growth exponent of G, then $G \curvearrowright \mathcal{X}$ is a growth tight action.*

The proof of Theorem 6.3 follows in part the proof of [Dal'Bo et al. 2011, Theorem 1.4] for geometrically finite Kleinian groups. For the divergence part of the proof, the Kleinian group ingredients of [op. cit., Theorem 1.4] are inessential, and our changes are mostly cosmetic. The real generalization is in the use of Proposition 5.1 instead of a ping-pong argument.

Proof. Let Γ be an infinite, infinite index normal subgroup of G. By Proposition 3.1, there is a strongly contracting element in Γ. Let $\overline{G}$ be a minimal section of G/Γ. If $\delta_{\overline{G}} \leqslant \delta^c_G$ then we are done since $\delta^c_G < \delta_G$, so suppose $\delta_{\overline{G}} > \delta^c_G$.

Claim: $\overline{G}$ is divergent.

Assume the claim, and let A be a maximal separated set in $\overline{G}$ as in Definition 4.4. Then A and $\overline{G}$ have the same critical exponent, and are both divergent. By Proposition 5.1, for sufficiently large n the map $\phi_n : A^* * \mathbb{Z}_2 \to G$ is an injection. By Lemma 6.1, $\delta_A < \delta_{\phi_n(A^* * \mathbb{Z})}$. Thus, $\delta_{G/\Gamma} = \delta_A < \delta_{\phi_n(A^* * \mathbb{Z}_2)} \leqslant \delta_G$.

It remains to prove the claim. Let $r > 0$, and suppose $d(o, \bar{g}\,.\,o) = r$. Let $0 \leqslant M_0 \leqslant r$ and $M_1 = r - M_0$. Choose a geodesic $[o,\ \bar{g}\,.\,o]$ from o to $\bar{g}\,.\,o$, and let $[o,\ \bar{g}\,.\,o](M_0)$ denote the point of $[o,\ \bar{g}\,.\,o]$ at distance M_0 from o.

First, we suppose that $[o,\ \bar{g}\,.\,o](M_0) \in \mathcal{X} \setminus \overline{\mathcal{B}}_Q(G\,.\,o)$. Let $[x_0, x_1] \subset [o,\ \bar{g}\,.\,o]$ be the largest subsegment containing $[o,\ \bar{g}\,.\,o](M_0)$ such that $(x_0, x_1) \subset \mathcal{X} \setminus \overline{\mathcal{B}}_Q(G\,.\,o)$. Let $m_0 = d(o, x_0)$, and let $m_1 = d(x_1,\ \bar{g}\,.\,o)$. There exist group elements $g_i \in G$ such that $d(g_i\,.\,o,\ x_i) \leqslant Q$; see Figure 4. We have $\bar{g}\,.\,o = g_0 \cdot g_0^{-1} g_1 \cdot g_1^{-1} \bar{g}\,.\,o$. Now, $m_0 - Q \leqslant d(o,\ \bar{g}_0\,.\,o) \leqslant d(o,\ g_0\,.\,o) \leqslant m_0 + Q$, and

$$m_1 - Q \leqslant d(o,\ \overline{g_1^{-1}\bar{g}}\,.\,o) \leqslant d(o,\ g_1^{-1}\bar{g}\,.\,o) \leqslant m_1 + Q.$$

Furthermore, $g_0^{-1} g_1 \in \mathrm{Comp}^G_{Q,\,r-(m_0+m_1)}$. Thus, the point $\bar{g}\,.\,o$ can be expressed as the product of an element of $\overline{G}$ of length $m_0 \pm Q$, an element of $\overline{G}$ of length $m_1 \pm Q$, and the quotient of an element of $\mathrm{Comp}^G_{Q,\,r-(m_0+m_1)}$.

(†) The same is also true if $[o,\ \bar{g}\,.\,o](M_0) \in \bar{\mathcal{B}}_Q(G\,.\,o)$, in which case we can take $m_0 = M_0$ and $m_1 = r - m_0$. Then choose $g_0 = g_1$ so that the contribution from $\mathrm{Comp}^G_{Q,\,r-(m_0+m_1)}$ is trivial.

Let $V_{r,Q} := \#\big(\bar{G}\,.\,o \cap \bar{\mathcal{B}}_{r+Q}(o) \setminus \mathcal{B}_{r-Q}(o)\big)$. For every $M_0 + M_1 = r$,

$$V_{r,Q} \overset{*}{\prec} \sum_{m_0=0}^{M_0} \sum_{m_1=0}^{M_1} V_{m_0,Q} \cdot V_{m_1,Q} \cdot \#\,\mathrm{Comp}^G_{Q,\,r-(m_0+m_1)}$$

Choose $\xi > 0$ such that $\delta_{\bar{G}} \geqslant 2\xi + \delta^c_G$. Since

$$\#\,\mathrm{Comp}^G_{Q,\,r-(m_0+m_1)} \overset{*}{\prec} \exp\big((r-(m_0+m_1))(\delta_{\bar{G}} - \xi)\big),$$

whenever $r - (m_0 + m_1)$ is sufficiently large, it follows that

$$(1) \quad V_{r,Q} \cdot \exp(-r(\delta_{\bar{G}} - \xi)) \overset{*}{\prec} \left(\sum_{m_0=0}^{M_0} V_{m_0,Q} \cdot \exp(-m_0(\delta_{\bar{G}} - \xi))\right) \cdot \left(\sum_{m_1=0}^{M_1} V_{m_1,Q} \cdot \exp(-m_1(\delta_{\bar{G}} - \xi))\right).$$

Set $w_i := V_{i,Q} \cdot \exp(-i(\delta_{\bar{G}} - \xi))$ and $W_i := \sum_{j=1}^{i} w_i$. Then, (1) and [Dal'Bo et al. 2011, Lemma 4.3] imply that $\sum_i w_i \cdot \exp(-is)$ diverges at its critical exponent, which is

$$\limsup_i \frac{\log w_i}{i} = \left(\limsup_i \frac{\log V_{i,Q}}{i}\right) - (\delta_{\bar{G}} - \xi) = \xi.$$

So, $\infty = \sum_i w_i \cdot \exp(i\xi) = \sum_i V_{i,Q} \cdot \exp(-i\delta_{\bar{G}}) \overset{*}{\asymp} \Theta_{\bar{G}}(\delta_{\bar{G}})$. □

Theorem 6.4. *Let G be a finitely generated, nonelementary group. Let $\mathcal{X}$ be a quasiconvex G-space. If G contains a strongly contracting element then $G \curvearrowright \mathcal{X}$ is a growth tight action.*

Proof. The proof is an easier special case of the proof of Theorem 6.3. If $\mathcal{X}$ is Q-quasiconvex, then we can always choose to be in case (†) of the proof. □

7. The growth of conjugacy classes

Parkkonen and Paulin [2015] ask: given a finitely generated group G with a word metric and an element $h \in G$, what is the growth rate of the conjugacy class $[h]$ of h? In a hyperbolic group G there is a finite subgroup, the *virtual center*, consisting of elements whose centralizer is finite index in G. The growth exponent of a conjugacy class in the virtual center is clearly zero. Parkkonen and Paulin show that for every element h not in the virtual center, $\delta_{[h]} = \frac{1}{2}\delta_G$. This generalizes an old result of Huber [1956] for the case of G acting cocompactly on the hyperbolic plane and h loxodromic.

Since strongly contracting elements behave much like infinite order elements in hyperbolic groups, it is natural to ask whether the growth exponent of the conjugacy class of a strongly contracting element h also satisfies $\delta_{[h]} = \frac{1}{2}\delta_G$.

We show that the lower bound holds, and the upper bound holds if h moves the basepoint sufficiently far with respect to the contraction constant for the axis.

Theorem 7.1. *Let G be a nonelementary, finitely generated group, and let $\mathcal{X}$ be a G-space. Let h be a strongly contracting element for $G \curvearrowright \mathcal{X}$. Then, $\delta_{[h]} \geqslant \frac{1}{2}\delta_G$.*

Let $D := \operatorname{diam} Z(h)\backslash\mathcal{H}$, where $Z(h)$ is the centralizer of h in G. Suppose $\pi_{\mathcal{H}}$ is a $(1, C)$-strongly constricting, C-coarse map. If $d(o,\ h\,.\,o) > 15C + 2D$, then $\delta_{[h]} = \frac{1}{2}\delta_G$.

Corollary 7.2. *For h strongly contracting, $\delta_{[h^n]} = \frac{1}{2}\delta_G$ for all sufficiently large n.*

Proof. For n nonzero, $E(h^n) = E(h)$ and $Z(h^n) \supset Z(h)$, so the same C and D work for h^n as work for h. On the other hand, $\langle h\rangle$ is quasi-isometrically embedded, so $d(o,\ h^n.\,o) \asymp n$. Thus, $d(o,\ h^n.\,o) > 15C + 2D$ for large enough n. $\square$

It would be interesting to know whether the restriction on $d(o,\ h\,.\,o)$ is really necessary.

Question 2. Does there exist an action $G \curvearrowright \mathcal{X}$ such that h is a strongly contracting element with $\delta_{[h]} > \frac{1}{2}\delta_G$?

Proof of Theorem 7.1. Define $K := 6C + D$ and $F := \{g \in G \mid d^{\pi}_{g\mathcal{H}}(o,\ g\,.\,o) \leqslant K\}$. First, we will show $\delta_F = \delta_G$. Then, we will relate $\delta_{[h]}$ to δ_F.

For any $r \geqslant 0$ consider

$$\phi : \{f \in F \setminus E(h) \mid d(o,\ f\,.\,o) \leqslant r\} \to \{g\mathcal{H} \mid g \in G \setminus E(h) \text{ and } g\mathcal{H} \cap \overline{\mathcal{B}}_r(o) \neq \varnothing\},$$

defined by $\phi(f) := f\mathcal{H}$. For each axis $g\mathcal{H}$ meeting $\overline{\mathcal{B}}_r(o)$ there exists a $g' \in gE(h)$ such that $d(o,\ g'.\,o) = d(o, g\mathcal{H}) \leqslant r$. Since $\pi_{g\mathcal{H}}$ is within $5C$ of closest point projection, by Lemma 2.8, we have $d^{\pi}_{g'\mathcal{H}}(o,\ g'.\,o) \leqslant 6C \leqslant K$. Therefore, $g' \in F$ with $\phi(g') = g\mathcal{H}$, so ϕ is surjective.

We estimate:

$$\#\{\text{axes meeting } \overline{\mathcal{B}}_r(o)\} \geqslant \frac{|G\,.\,o \cap \overline{\mathcal{B}}_r(o)| \times \#\{\text{axes per orbit point}\}}{\text{maximum number of orbit points per axis}}.$$

The basepoint belongs to $[\operatorname{Stab}_G(o) : E(h) \cap \operatorname{Stab}_G(o)]$ distinct translates of $\mathcal{H}$, so the number of axes per orbit point is constant. The maximum number of orbit points in $\overline{\mathcal{B}}_r(o)$ contained in a single axis is proportional to r, since each axis is a quasi-isometrically embedded image of a virtually cyclic group. Combined with surjectivity of ϕ, this gives

$$|F\,.\,o \cap \overline{\mathcal{B}}_r(o)| \stackrel{*}{\succ} \frac{|G\,.\,o \cap \overline{\mathcal{B}}_r(o)|}{r}$$

Thus,

$$\begin{aligned}\delta_F &= \limsup_{r\to\infty}\frac{1}{r}\log|F.o\cap\bar{\mathcal{B}}_r(o)| \geqslant \limsup_{r\to\infty}\frac{1}{r}\log\frac{|G.o\cap\bar{\mathcal{B}}_r(o)|}{r}\\ &= \limsup_{r\to\infty}\frac{1}{r}\log|G.o\cap\bar{\mathcal{B}}_r(o)| = \delta_G.\end{aligned}$$

The reverse inequality is trivial, since $F\subset G$, so $\delta_F=\delta_G$.

Now, consider the map $\psi : F\setminus E(h)\to [h]\setminus E(h)$ defined by $\psi(f) := fhf^{-1}$. Choose minimal length representatives $e_1,\dots,e_m$ of $Z(h)\backslash E(h)$. Associated to each $g\in G\setminus E(h)$, there is a $g'\in gE(h)$ such that $d(o, g\mathcal{H}) = d(o,\ g'.o)$. There exist $z\in Z(h)$ and i such that $g' = gze_i$. By setting $f := g'e_i^{-1}$, we get

$$fhf^{-1} = gze_ie_i{}^{-1}he_ie_i{}^{-1}z^{-1}g^{-1} = ghg^{-1}.$$

Since e_i has length at most D and $\pi_{g\mathcal{H}}$ is $5C$-close to closest point projection, it follows that $f\in F$; hence, ψ is surjective. Furthermore, $d(o,\ fhf^{-1}.o)\leqslant 2d(o,\ f.o)+d(o,\ h.o)$, by the triangle inequality.

On the other hand, ψ is boundedly many-to-one, since if $fhf^{-1} = f'hf'^{-1}$ then $f'\in fE(h)$, so $f\mathcal{H} = f'\mathcal{H}$. By definition of F, we then have $d^{\pi}_{f\mathcal{H}}(o,\ f.o)\leqslant K$ and $d^{\pi}_{f\mathcal{H}}(o,\ f'.o)\leqslant K$, so $d(f.o,\ f'.o)\leqslant 2(C+K)$. There are uniformly boundedly many such f' for each f.

We conclude that ψ is a surjective, boundedly-many-to-one map such that $d(o,\psi(f).o)\overset{+}{\prec} 2d(o, f.o)$ for all f. We excluded $E(h)$ from the domain and range, but its growth exponent is zero, since it embeds quasi-isometrically into $\mathcal{X}$, so $\delta_{[h]} = \delta_{[h]\setminus E(h)}\geqslant \frac{1}{2}\delta_{F\setminus E(h)} = \frac{1}{2}\delta_F = \frac{1}{2}\delta_G$.

Now, $d^{\pi}_{f\mathcal{H}}(fh.o,\ fhf^{-1}.o) = d^{\pi}_{f\mathcal{H}}(o,\ f.o)\leqslant K$ for $f\in F$, so

$$d^{\pi}_{f\mathcal{H}}(o,\ fhf^{-1}.o) > d(f.o,\ fh.o)-2(C+K).$$

If $d(o,\ h.o) > 15C+2D = C+2(C+K)$, then $d^{\pi}_{f\mathcal{H}}(o,\ fhf^{-1}.o) > C$, so by strong constriction,

$$d(o,\ fhf^{-1}.o)\geqslant 2d(o,\ f.o)+d(o,\ h.o)-4(C+K).$$

Thus, $d(o,\ \psi(f).o)\overset{+}{\asymp} 2d(o,\ f.o)$ and $\delta_{[h]} = \frac{1}{2}\delta_G$. □

Part II. Examples of actions with strongly contracting elements

8. Actions on relatively hyperbolic spaces

Yang [2013] proved that the action of a finitely generated group G with a nontrivial Floyd boundary on any of its Cayley graphs is growth tight. Relatively hyperbolic groups have nontrivial Floyd boundaries by a theorem of Gerasimov [2012], so the action of a relatively hyperbolic group on any of its Cayley graphs is growth tight.

It is an open question whether there exists a group with a nontrivial Floyd boundary that is not relatively hyperbolic.

There is also a notion of relative hyperbolicity of metric spaces, which we will review in Section 8A. One motivating example of a relatively hyperbolic metric space is a Cayley graph of a relatively hyperbolic group. Another is the universal cover $\widetilde{M}$ of a complete, finite volume hyperbolic manifold M. The fundamental group $\pi_1(M)$ of such a manifold is a relatively hyperbolic group, so the action of $\pi_1(M)$ on any of its Cayley graphs is growth tight by Yang's theorem. However, this does not tell us whether the action of $\pi_1(M)$ on $\widetilde{M}$ is growth tight. This question was addressed for a more general class of manifolds by Dal'bo, Peigné, Picaud, and Sambusetti [Dal'Bo et al. 2011], who proved growth tightness results for geometrically finite Kleinian groups. Using our main theorems, Theorem 6.3 and Theorem 6.4, we generalize their results to all groups acting on relatively hyperbolic metric spaces.

8A. *Relatively hyperbolic metric spaces.*

Definition 8.1 (compare [Druţu 2009; Sisto 2012]). Let $\mathcal{X}$ be a geodesic metric space and let $\underline{\mathcal{P}}$ be a collection of uniformly coarsely connected subsets of $\mathcal{X}$. We say $\mathcal{X}$ is *hyperbolic relative to the peripheral sets* $\underline{\mathcal{P}}$ if the following conditions are satisfied:

(1) For each A there exists a B such that $\operatorname{diam}(\overline{\mathcal{B}}_A(\mathcal{P}_0) \cap \overline{\mathcal{B}}_A(\mathcal{P}_1)) \leqslant B$ for distinct $\mathcal{P}_0, \mathcal{P}_1 \in \underline{\mathcal{P}}$.

(2) There exists an $\epsilon \in (0, \frac{1}{2})$ and $M \geqslant 0$ such that if $x_0, x_1 \in \mathcal{X}$ are points such that for some $\mathcal{P} \in \underline{\mathcal{P}}$ we have $d(x_i, \mathcal{P}) \leqslant \epsilon \cdot d(x_0, x_1)$ for each i, then every geodesic from x_0 to x_1 intersects $\overline{\mathcal{B}}_M(\mathcal{P})$.

(3) There exist σ and δ so that for every geodesic triangle either:
   (a) there exists a ball of radius σ intersecting all three sides, or
   (b) there exists a $\mathcal{P} \in \underline{\mathcal{P}}$ such that $\overline{\mathcal{B}}_\sigma(\mathcal{P})$ intersects all three sides and for each corner of the triangle, the points of the outgoing geodesics from that corner which first enter $\overline{\mathcal{B}}_\sigma(\mathcal{P})$ are distance at most δ apart.

We say $\mathcal{X}$ is *hyperbolic* if it hyperbolic relative to $\underline{\mathcal{P}} = \varnothing$.

If $\mathcal{X}$ is hyperbolic in the sense of Definition 8.1, then the only nontrivial condition is (a), which is equivalent to the usual definition of hyperbolic metric space.

Definition 8.2. A group G is *hyperbolic relative to a collection of finitely generated peripheral subgroups* if a Cayley graph of G is hyperbolic relative to the cosets of the peripheral subgroups.

Sisto [2012] shows that Definition 8.2 is equivalent to Bowditch's definition [2012] of relatively hyperbolic groups.

Definition 8.3 (compare [Groves and Manning 2007]). Let $\mathcal{X}$ be a connected graph with edges of length bounded below. A *combinatorial horoball* based on X with parameter $a > 0$ is a graph whose vertex set is $\operatorname{Vert}\mathcal{X} \times (\{0\} \cup \mathbb{N})$, contains an edge of length 1 between (v, n) and $(v, n+1)$ for all $v \in \operatorname{Vert}\mathcal{X}$ and all $n \in \{0\} \cup \mathbb{N}$, and for each edge $[v, w] \in \mathcal{X}$ contains an edge $[(v, n), (w, n)]$ of length $e^{-an} \cdot \operatorname{length}([v, w])$.

Let $\mathcal{X}$ be hyperbolic relative to $\underline{\mathcal{P}}$. An *augmented space* is a space obtained from $\mathcal{X}$ as follows. By definition, there exists a constant C such that each $\mathcal{P} \in \underline{\mathcal{P}}$ is C-coarsely connected. For each $\mathcal{P} \in \underline{\mathcal{P}}$ choose a maximal subset of points that pairwise have distance at least C from one another. Let these points be the vertex set of a graph. For edges, choose a geodesic connecting each pair of vertices at distance at most $2C$ from each other. Use this graph as the base of a combinatorial horoball with parameter $a_{\mathcal{P}} > 0$. The augmented space is the space obtained from the union of $\mathcal{X}$ with horoballs $\mathcal{X}_{\mathcal{P}}$ for each $\mathcal{P} \in \underline{\mathcal{P}}$ by identifying the base of $\mathcal{X}_{\mathcal{P}}$ with the graph constructed in $\mathcal{P}$.

Definition 8.4. Let $\mathcal{X}$ be a hyperbolic G-space, and let $\underline{P}$ be the collection of maximal parabolic subgroups of G. Suppose there exists a G-invariant collection of disjoint open horoballs centered at the points fixed by the parabolic subgroups. The *truncated space* is $\mathcal{X}$ minus the union of these open horoballs. We say $G \curvearrowright \mathcal{X}$ is *cusp uniform* if G acts cocompactly on the truncated space.

If G acts cocompactly on a G-space $\mathcal{X}'$ that is hyperbolic relative to a G-invariant peripheral system $\underline{\mathcal{P}}$, then an augmented space $\mathcal{X}$ can be constructed G-equivariantly, and $G \curvearrowright X$ will be a cusp uniform action.

Several different versions of the following theorem occur in the literature on relatively hyperbolic groups:

Theorem 8.5 [Bowditch 2012; Groves and Manning 2008; Sisto 2012]. *If $\mathcal{X}$ is hyperbolic relative to $\underline{\mathcal{P}}$, then any augmented space with horoball parameters bounded below is hyperbolic.*

If $G \curvearrowright \mathcal{X}$ is a cusp uniform action, then G is hyperbolic relative to the maximal parabolic subgroups and the truncated space is hyperbolic relative to boundaries of the deleted horoballs.

8B. *Quasiconvex actions.*

Theorem 8.6. *If $\mathcal{X}$ is a quasiconvex, relatively hyperbolic G-space and G does not coarsely fix a peripheral subspace then $G \curvearrowright \mathcal{X}$ is a growth tight action.*

Proof. It follows from [Sisto 2012, Lemma 5.4] that every infinite order element of G that does not coarsely fix a peripheral subspace is strongly constricting. We conclude by Theorem 6.4. □

Theorem 8.6 unifies the existing proofs of growth tightness for cocompact actions on hyperbolic spaces [Sabourau 2013] and for actions of a relatively hyperbolic group on its Cayley graphs [Yang 2013], and extends to actions on a more general class of spaces.

Corollary 8.7. *The action of a finitely generated group G with infinitely many ends on any one of its Cayley graphs is growth tight.*

Proof. Stallings' theorem [1971] says that G splits nontrivially over a finite subgroup. G is hyperbolic relative to the factor groups of this splitting. Since the splitting is nontrivial, G does not fix any factor group, so Theorem 8.6 gives the result. □

Corollary 8.7 generalizes a result of Sambusetti [2002a, Theorem 1.4], who proved it with additional constraints on the factor groups.

8C. ***Cusp uniform actions.*** Theorems 8.6 and 8.5 show that if $G \curvearrowright \mathcal{X}$ is a cusp uniform action on a hyperbolic space then the action of G on the truncated space is a growth tight action. A natural question is whether $G \curvearrowright \mathcal{X}$ is a growth tight action. This action is not quasiconvex if the parabolic subgroups are infinite, as geodesics in $\mathcal{X}$ will travel deeply into horoballs, and, indeed, an example of Dal'bo, Otal, and Peigné [Dal'bo et al. 2000] shows $G \curvearrowright \mathcal{X}$ need not be growth tight.

To see how growth tightness can fail, consider the combinatorial horoball from Definition 8.3. If $\mathcal{X}$ is, say, the Cayley graph of some group and we build the combinatorial horoball with parameter $a > 0$ based on $\mathcal{X}$, then the r-ball about a basepoint $o \in \mathcal{X}$ in the horoball metric intersected with $\mathcal{X} \times \{0\}$ contains the ball of radius $C \cdot \exp(ar/2)$ in the $\mathcal{X}$-metric, for a constant C depending only on a. Thus, if the number of vertices of balls in $\mathcal{X}$ grows faster than polynomially in the radius, then the growth exponent with respect to the horoball metric will be infinite. Furthermore, even if growth in $\mathcal{X}$ is polynomial we can make the growth exponent in the horoball be as large as we like by taking a to be sufficiently large. Dal'bo, Otal, and Peigné construct non-growth-tight examples of relatively hyperbolic groups with two cusps by taking one of the horoball parameters to be large enough so that the corresponding parabolic subgroup dominates the growth of the group; that is, the growth exponent of the parabolic subgroup is equal to the growth exponent of the whole group. Quotienting by the second parabolic subgroup then does not decrease the growth exponent, so this action is not growth tight.

Not only does this provide an example of a non-growth-tight action on a hyperbolic space, but since augmented spaces with different horoball parameters are equivariantly quasi-isometric to each other, we have:

Observation 8.8. Growth tightness is not invariant among equivariantly quasi-isometric G-spaces.

It is shown in [Dal'Bo et al. 2011, Theorem 1.4] that this is essentially the only way that growth tightness can fail for cusp uniform actions. Their proof is for geometrically finite Kleinian groups, but our Theorem 6.3 generalizes this result.

Definition 8.9. Let $G \curvearrowright \mathcal{X}$ be a cusp uniform action on a hyperbolic space. Let $\underline{P}$ be the collection of maximal parabolic subgroups of G. Then $G \curvearrowright \mathcal{X}$ satisfies the *parabolic gap condition* if $\delta_P < \delta_G$ for all $P \in \underline{P}$.

Theorem 8.10. *Let G be a finitely generated, nonelementary group. Let $G \curvearrowright \mathcal{X}$ be a cusp uniform action on a hyperbolic space. Suppose that $G \curvearrowright \mathcal{X}$ satisfies the parabolic gap condition. Then $G \curvearrowright \mathcal{X}$ is a growth tight action.*

Proof. Let Q be the diameter of the quotient of the truncated space. Then the Q-complementary growth exponent is the maximum of the parabolic growth exponents, which, by the parabolic gap condition, is strictly less than the growth exponent of G. Apply Theorem 6.3. □

Theorem 8.11. *Let G be a finitely generated group hyperbolic relative to a collection $\underline{P}$ of virtually nilpotent subgroups. Then there exists a hyperbolic G-space $\mathcal{X}$ such that $G \curvearrowright \mathcal{X}$ is cusp uniform and growth tight.*

Proof. Construct $\mathcal{X}$ as an augmented space by taking a Cayley graph for G and attaching combinatorial horoballs to the cosets of the peripheral subgroups. Since the peripheral groups are virtually nilpotent, they have polynomial growth in any word metric [Gromov 1981]. It follows that the growth exponent of each parabolic group with respect to the horoball metric is bounded by a multiple of the horoball parameter. By choosing the horoball parameters small enough, we can ensure $G \curvearrowright \mathcal{X}$ satisfies the parabolic gap condition, and apply Theorem 8.10. □

8D. ***Beyond relative hyperbolicity.*** In subsequent sections we provide further examples of growth tight actions. To show these are not redundant we will verify that the groups are not relatively hyperbolic.

In this section we recall a technique for showing that a group is not relatively hyperbolic, due to Anderson, Aramayona, and Shackleton [Anderson et al. 2007]. Another approach to the question, contemporaneous to and more general than the one just cited, and also implying Theorem 8.13, was developed by Behrstock, Druțu, and Mosher [Behrstock et al. 2009].

Theorem 8.12 [Anderson et al. 2007, Theorem 2]. *Let G be a finitely generated, nonelementary group, and let S be a (possibly infinite) generating set consisting of infinite order elements. Consider the "commutativity graph" with one vertex for each element of S and an edge between vertices s and s' if some nontrivial powers of s and s' commute. If this graph is connected and there is at least one pair $s, s' \in S$ such that $\langle s, s' \rangle$ contains a rank 2 free abelian subgroup, then G is not hyperbolic relative to any finite collection of proper finitely generated subgroups.*

To prove this theorem, one shows that the subgroup generated by S is contained in one of the peripheral subgroups. Since S generates G this gives a contradiction, because the peripheral subgroups are proper subgroups of G.

We will actually use a mild generalization of Theorem 8.12 to the case when S generates a proper subgroup of G:

Theorem 8.13. *Let G be a finitely generated, nonelementary group. Let S be a set of infinite order elements whose commutativity graph is connected and such that there is a pair $s, s' \in S$ such that $\langle s, s' \rangle$ contains a rank 2 free abelian subgroup. Consider the "coset graph" whose vertices are cosets of $\langle S \rangle$, with an edge connecting $g\langle S \rangle$ and $h\langle S \rangle$ if $g\langle S \rangle g^{-1} \cap h\langle S \rangle h^{-1}$ is infinite. If this graph is connected, then G is not hyperbolic relative to any finite collection of proper finitely generated subgroups.*

Proof. Suppose G is hyperbolic relative to $\{P_1, \dots, P_k\}$. As in the proof of Theorem 8.12, $\langle S \rangle$ is contained in a conjugate of some P_i. We assume, without loss of generality, that $\langle S \rangle \subset P_1$. Condition (1) of Definition 8.1 implies $P_i \cap g P_i g^{-1}$ is finite for $g \notin P_i$. Thus, for $g\langle S \rangle$ adjacent to $\langle S \rangle$ in the coset graph, $g \in P_1$ and $g\langle S \rangle g^{-1} \subset P_1$. Connectivity of the coset graph implies that every element of G is contained in P_1, contradicting the hypothesis that P_1 is a proper subgroup. □

We also note that Theorem 8.12 and Theorem 8.13 imply the, a priori, stronger result that G has trivial Floyd boundary.

9. Rank 1 actions on CAT(0) spaces

A metric space is *CAT(0)* if every geodesic triangle is at least as thin as a triangle in Euclidean space with the same side lengths. An isometry ϕ of a CAT(0) space $\mathcal{X}$ is *hyperbolic* if $\inf_{x \in \mathcal{X}} d(x, \phi(x))$ is positive and is attained. See, for example, [Bridson and Haefliger 1999] for more background.

Let $\mathcal{X}$ be a CAT(0) G-space. Recall that our definition of "G-space" includes the hypothesis that $\mathcal{X}$ is proper, so an element is strongly contracting if and only if it acts as a rank 1 isometry:

Theorem 9.1 [Bestvina and Fujiwara 2009, Theorem 5.4]. *Let h be a hyperbolic isometry of a proper CAT(0) space $\mathcal{X}$ with axis $\mathcal{A}$. Closest point projection to $\mathcal{A}$ is strongly contracting if and only if $\mathcal{A}$ does not bound an isometrically embedded half-flat in $\mathcal{X}$.*

Theorems 9.1 and 6.4 show:

Theorem 9.2. *If G is a nonelementary, finitely generated group and $\mathcal{X}$ is a quasi-convex, CAT(0) G-space such that G contains an element that acts as a rank 1 isometry on $\mathcal{X}$, then $G \curvearrowright \mathcal{X}$ is a growth tight action.*

Recall from Section 0C that there are many interesting classes of CAT(0) spaces that admit rank 1 isometries. In the remainder of this section we highlight a few examples.

Let Θ be a simple graph. The *right angled Artin group* $G(\Theta)$ *defined by* Θ is the group defined by the presentation

$$\langle g_v \text{ for } v \in \mathrm{Vert}(\Theta) \mid g_v g_w g_v^{-1} g_w^{-1} = 1 \text{ for } [v, w] \in \mathrm{Edge}(\Theta)\rangle.$$

The graph Θ also determines a cube complex constructed by taking a rose with one loop for each vertex of Θ, and then gluing in a k-cube to form a k-torus for each complete k-vertex subgraph of Θ. The resulting complex is called the *Salvetti complex*, and its fundamental group is $G(\Theta)$. The universal cover of the Salvetti complex turns out to be a CAT(0) cube complex. See [Charney 2007] for more background on right angled Artin groups.

If Θ is a single vertex then $G(\Theta) \cong \mathbb{Z}$ is elementary. If Θ is a join, that is, if it is a complete bipartite graph, then $G(\Theta)$ is a direct product of right angled Artin groups defined by the two parts. In all other cases, we find a growth tight action:

Theorem 9.3. *Let Θ be a finite simple graph that is not a join and has more than one vertex. The action of the right angled Artin group $G(\Theta)$ defined by Θ on the universal cover $\mathcal{X}$ of the Salvetti complex associated to Θ is a growth tight action.*

Proof. The universal cover $\mathcal{X}$ of the Salvetti complex of Θ is a cocompact, CAT(0) $G(\Theta)$-space. If Θ is not connected then $\mathcal{X}$ is hyperbolic relative to subcomplexes defined by the components of Θ, so $G(\Theta) \curvearrowright \mathcal{X}$ is growth tight by Theorem 8.6. If Θ is connected then $G(\Theta)$ contains a rank 1 isometry by a theorem of Behrstock and Charney [2012]. The result follows from Theorem 9.2. □

The defining graph of a right angled Artin group is a commutativity graph. If this graph is connected then the group is not relatively hyperbolic by Theorem 8.12.

A *flip-graph manifold* is a compact three dimensional manifold M with boundary obtained from a finite collection of Seifert fibered pieces that are each a product of a circle with a compact oriented hyperbolic surface with boundary. These are glued together along boundary tori by a map exchanging the fiber and base directions. Such manifolds were studied by Kapovich and Leeb [1998], who show that the universal cover of M admits a CAT(0) metric, and that an element of $\pi_1(M)$ that acts hyperbolically is rank 1 if and only if it is not represented by a loop contained in a single Seifert fibered piece. Thus, Theorem 9.2 implies the following:

Theorem 9.4. *The action of the fundamental group of a flip-graph manifold by deck transformations on its universal cover with its natural CAT(0) metric is a growth tight action.*

To see that the fundamental group of a flip-graph manifold is not relatively hyperbolic, apply Theorem 8.13 where S is the set of elliptic elements for the action of G on the Bass–Serre tree of the defining graph of groups decomposition.

Theorems 9.3 and 9.4 give the first nontrivial examples of growth tight actions on spaces that are not relatively hyperbolic.

The idea of the proof for flip-graph manifolds generalizes to other CAT(0) graphs of groups via Theorem 1.14:

Theorem 9.5. *Let G be a nonelementary, finitely generated group that splits nontrivially as a graph of groups and is not an ascending HNN-extension. Suppose that the corresponding action of G on the Bass–Serre tree of the splitting has two edges whose stabilizers have finite intersection. Suppose there exists a cocompact, CAT(0) G-space $\mathcal{X}$. Then, $G \curvearrowright \mathcal{X}$ is a growth tight action.*

Proof. By Theorem 1.14, G contains an infinite order element h such that $E(h)$ is hyperbolically embedded. A theorem of Sisto [2013] implies that any axis of h is a Morse quasigeodesic. An element with an axis that bounds a half-flat is not Morse, so h is rank 1, and the result follows by Theorem 9.2. □

10. Mapping class groups

Let $\mathcal{S} = \mathcal{S}_{g,p}$ be a connected and oriented surface of genus g with p punctures. We require $\mathcal{S}$ to have negative Euler characteristic.

Given two orientation preserving homeomorphisms $\phi, \psi \colon \mathcal{S} \to \mathcal{S}$, we will consider ϕ and ψ to be equivalent if $\phi \circ \psi^{-1}$ is isotopic to the identity map on $\mathcal{S}$. Each equivalence class is called a *mapping class* of $\mathcal{S}$, and the set $\mathrm{Mod}(\mathcal{S})$ of all equivalence classes naturally forms a group called the *mapping class group* of $\mathcal{S}$.

A mapping class $f \in \mathrm{Mod}(\mathcal{S})$ is called *reducible* is there exists an f-invariant curve system on $\mathcal{S}$ and *irreducible* otherwise. By the Nielsen–Thurston classification of elements of $\mathrm{Mod}(\mathcal{S})$, a mapping class is irreducible and infinite order if and only if it is pseudo-Anosov [Thurston 1998].

Let $\mathcal{X}$ be the Teichmüller space of marked hyperbolic structures on $\mathcal{S}$, equipped with the Teichmüller metric. See [Hubbard 2006; Papadopoulos 2007] for more information.

Theorem 10.1 [Minsky 1996]. *Each pseudo-Anosov element is strongly contracting for* $\mathrm{Mod}(\mathcal{S}) \curvearrowright \mathcal{X}$.

For each $\epsilon > 0$ there is a decomposition of $\mathcal{X}$ into a "thick part" $\mathcal{X}^{\geqslant\epsilon}$ and a "thin part" $\mathcal{X}^{<\epsilon}$ according to whether the hyperbolic structure on $\mathcal{S}$ corresponding to the point $x \in \mathcal{X}$ has any closed curves of length $< \epsilon$. This decomposition is $\mathrm{Mod}(\mathcal{S})$-invariant, and $\mathrm{Mod}(\mathcal{S}) \curvearrowright \mathcal{X}^{\geqslant\epsilon}$ is cocompact; see [Mumford 1971; Farb and Margalit 2012]. Geodesics between points in the thick part can travel deeply into the thin part, so the action of $\mathrm{Mod}(\mathcal{S})$ on Teichmüller space is not quasiconvex. To prove growth tightness, we need a bound on the complementary growth exponent. Such a bound is provided by Eskin, Mirzakhani, and Rafi [Eskin et al. 2012, Theorem 1.7].

Theorem 10.2. *The action of the mapping class group* $\mathrm{Mod}(\mathcal{S})$ *of* $\mathcal{S}=\mathcal{S}_{g,p}$ *on its Teichmüller space* $\mathcal{X}$ *with the Teichmüller metric is a growth tight action.*

Proof. Let $\zeta=6g-6+2p\geqslant 2$. The growth exponent of $\mathrm{Mod}(\mathcal{S})$ with respect to its action on $\mathcal{X}$ is ζ [Athreya et al. 2012]. (We remark that the result in that reference is stated for closed surfaces, but their proof works in general. For our interest, it is enough that the growth exponent of $\mathrm{Mod}(\mathcal{S})$ is bounded below by ζ. This can be obtained from [Hamenstädt 2013; Eskin et al. 2012].)

Choose r_0 and a maximal r_0-separated set in the moduli space $\mathrm{Mod}(\mathcal{S})\backslash\mathcal{X}$, and let $\mathcal{A}$ be its full lift to $\mathcal{X}$. Given r_0 as above and $\delta=\frac{1}{2}$, let ϵ be sufficiently small, as in [Eskin et al. 2012, Theorem 1.7]. Let Q be the smallest number such that the ϵ-thick part of $\mathcal{X}$ is contained in $\overline{\mathcal{B}}_Q(\mathrm{Mod}(\mathcal{S})\,.\,o)$. Choose a finite subset $\{a_1,\dots,a_n\}\subset\mathcal{A}$ such that

$$\overline{\mathcal{B}}_Q(o)\setminus\mathcal{B}_Q(\mathrm{Mod}(\mathcal{S})\,.\,o)\subset\bigcup_{i=1}^{n}\mathcal{B}_{r_0}(a_i).$$

Suppose that $g\in\mathrm{Mod}(\mathcal{S})$ is such that there exists a geodesic $[x,y]$ between $\overline{\mathcal{B}}_Q(o)$ and $\overline{\mathcal{B}}_Q(g\,.\,o)$ whose interior stays in $\mathcal{X}\setminus\overline{\mathcal{B}}_Q(\mathrm{Mod}(\mathcal{S})\,.\,o)$. Then there are indices i and j such that $x\in\mathcal{B}_{r_0}(a_i)$ and $y\in\mathcal{B}_{r_0}(g\,.\,a_j)$. This means that every element contributing to $\mathrm{Comp}^{\mathrm{Mod}(\mathcal{S})}_{Q,r}$ of Definition 6.2 also contributes to some $N_1(\mathcal{Q}_{1,\epsilon},a_i,a_j,r)$ of [Eskin et al. 2012, Theorem 1.7]. The conclusion of that theorem is that $N_1(\mathcal{Q}_{1,\epsilon},a_i,a_j,r)\leqslant G(a_i)G(a_j)\exp(r\cdot(\zeta-\frac{1}{2}))$ for all sufficiently large r, where G is a particular function on $\mathcal{X}$. There are finitely many such sets, and the function G is bounded on $\{a_1,\dots,a_n\}$, so there is a constant C such that $\mathrm{Comp}^{\mathrm{Mod}(\mathcal{S})}_{Q,r}\leqslant C\cdot\exp(r\cdot(\zeta-\frac{1}{2}))$ for all sufficiently large r. Thus, the Q-complementary growth exponent is at most $\zeta-\frac{1}{2}<\zeta$. The theorem now follows from Theorems 10.1 and 6.3. □

When the genus of $\mathcal{S}$ is at least 3 then there does not exist a cocompact, CAT(0) $\mathrm{Mod}(\mathcal{S})$-space [Bridson 2010]. The fact that such an $\mathrm{Mod}(\mathcal{S})$ is not relatively hyperbolic (in fact, has trivial Floyd boundary) is an application of Theorem 8.12 appearing in [Anderson et al. 2007]. Therefore, Theorem 10.2 does not follow from the results of the previous sections.

A natural question is whether the action of a mapping class group on its Cayley graphs is growth tight. There is also a combinatorial model for the mapping class group known as the *marking complex*. Finally, a mapping class group acts cocompactly on a thick part of the Teichmüller space. All of these spaces are quasi-isometric, and Duchin and Rafi [2009] show that pseudo-Anosov elements are contracting for the action of a mapping class group on any one of its Cayley graphs, but we do not know whether one of these actions admits a strongly contracting element.

Question 3. Is the action of a mapping class group of a hyperbolic surface on one of its Cayley graphs/marking complex/thick part of Teichmüller space growth tight?

The outer automorphism group of a finite rank nonabelian free group, $\mathrm{Out}(F_n)$ is often studied in analogy with $\mathrm{Mod}(\mathcal{S})$. Algom-Kfir [2011] has proven an analogue of Minsky's theorem that says that a *fully irreducible* outer automorphism class is strongly contracting for the action of $\mathrm{Out}(F_n)$ on its outer space, which is the analogue of the Teichmüller space. However, we lack the analogue of the theorem of Eskin, Mirzakhani, and Rafi that was used to control the complementary growth exponent in the mapping class group case.

There is also an analogue of the thick part of Teichmüller space called the *spine* of the outer space, on which $\mathrm{Out}(F_n)$ acts cocompactly.

Question 4. Is the action of $\mathrm{Out}(F_n)$ on one of its Cayley graphs/outer space/spine of outer space growth tight?

11. Snowflake groups

Let

$$G := BB(1,r) = \langle a, b, s, t \mid aba^{-1}b^{-1} = 1,\ s^{-1}as = a^r b,\ t^{-1}at = a^r b^{-1}\rangle$$

be a Brady–Bridson snowflake group with $r \geqslant 3$. Let $L := 2r$. These groups have an interesting mixture of positive and negative curvature properties. G splits as an amalgam of $\mathbb{Z}^2 = \langle a, b\rangle$ by two cyclic groups $\langle a^r b\rangle$ and $\langle a^r b^{-1}\rangle$, and the action of G on the Bass–Serre tree $\mathcal{T}$ of this splitting satisfies Theorem 1.14, so G has hyperbolically embedded subgroups. However, we can not automatically conclude that such a hyperbolically embedded subgroup gives rise to a strongly contracting element, as there does not exist a cocompact, CAT(0) G-space. If such a space existed, then the Dehn function of G would be at most quadratic, but Brady and Bridson [2000] have shown that the Dehn function of $BB(1,r)$ is $n^{2\log_2 L} > n^2$.

We will fix a G-space $\mathcal{X}$ and demonstrate two different elements of G that act hyperbolically on $\mathcal{T}$ such that the pointwise stabilizer of any length 3 segment of their axes is finite. One of these elements will be strongly contracting for the action on $\mathcal{X}$, and the other will not.

Theorem 11.1. *G admits a cocompact growth tight action.*

Observe that Theorem 8.13 with $S := \{a, b\}$ shows that G is not relatively hyperbolic.

11A. ***The model space*** $\mathcal{X}$**.** Let $\mathcal{X}$ be the Cayley graph for G with respect to the generating set $\{a, a^r b, a^r b^{-1}, s, t\}$, where the edges corresponding to $a^r b$ and $a^r b^{-1}$ have been rescaled to have length $L := 2r$. The point of scaling these edges is that $a^r b$, $a^r b^{-1}$, and a^{2r} form an equilateral triangle of side length L, which will facilitate finding geodesics in this particular model.

It is also useful to consider G as the fundamental group of the topological space obtained from a torus by gluing on two annuli. Choose a basepoint for the torus and for each boundary component of the annuli. For one annulus, the s-annulus, glue the two boundary curves to the curves a and $a^r b$ in the torus, gluing basepoints to the basepoint of the torus. For the other annulus, the t-annulus, glue the two boundary curves to the curves a and $a^r b^{-1}$ of the torus. The resulting space is a graph of spaces that Scott and Wall [1979] associated to the given graph of groups decomposition of G.

The fundamental group of this space is G, which acts freely by deck transformations on the universal cover $\mathcal{X}'$. Choose the basepoint o of $\mathcal{X}'$ to be a lift of the basepoint of the torus. The correspondence between a vertex $g \in \mathcal{X}$ and the point $g\,.\,o \in \mathcal{X}'$ inspires the following terminology: A *plane* is a coset $g\langle a, b\rangle \in G/\langle a, b\rangle$, which corresponds to a lift of the torus at the point $g\,.\,o \in \mathcal{X}'$. An s-*wall* is the set of outgoing s-edges incident to a coset $g\langle a\rangle \in G/\langle a\rangle$. This corresponds to a lift of the s-annulus at the point $g\,.\,o \in \mathcal{X}'$. A t-*wall* is the set of outgoing t-edges incident to a coset $g\langle a\rangle \in G/\langle a\rangle$. This corresponds to a lift of the t-annulus at the point $g\,.\,o \in \mathcal{X}'$. Each wall separates $\mathcal{X}$ (and $\mathcal{X}'$) into two complementary components. Notice that the origins of consecutive edges in an s-wall are connected by a single a-edge of length 1, while the termini of those edges are connected by a single $a^r b$-edge of length L. We say that crossing an s-wall in the positive direction scales distance by a factor of L. The same is true for the t-walls.

11B. ***Geodesics between points in a plane.*** We will define a family of $\mathcal{X}$-geodesics joining 1 to every point of $\langle a, b\rangle$. This is similar to the argument of [Brady and Bridson 2000].

From the fact that $\langle a, b\rangle$ is abelian, for every point $a^x b^y$ there is a geodesic from 1 to $a^x b^y$ of the form $[1, (a^r b)^m] + (a^r b)^m [1, (a^r b^{-1})^n] + (a^r b)^m (a^r b^{-1})^n [1, a^p]$, where $[g, h]$ indicates a geodesic from g to h.

For a point of the form $(a^r b)^m$ there is an $a^r b$-edge path from 1 to $(a^r b)^m$ of length mL. This path is clearly inefficient, as it lies along the boundary of an s^{-1}-wall that scales distance by $1/L$, so we can push the original edge path across the wall to a path $s^{-1} a^m s$ of length $2 + m$. We claim there is a geodesic from 1 to $(a^r b)^m$ of the form $[1, s^{-1}] + s^{-1}[1, a^m] + s^{-1} a^m [s^{-1}, 1]$. We have already exhibited a wall crossing path of length $2 + m$, which is shorter than any path from 1 to $(a^r b)^m$ that stays in the plane $\langle a, b\rangle$. Thus, a geodesic must cross some walls.

Every path from 1 to $(a^r b)^m$ can, by rearranging subsegments and eliminating backtracking, be replaced by a path of at most the same length and having the form $\gamma_s + \gamma_t + \gamma'$ where:

- $\gamma_s = [1, s^{-1}] + s^{-1}[1, a^n] + s^{-1}a^n[s^{-1}, 1]$, if nontrivial;
- $\gamma_t = s^{-1}a^n s[1, t^{-1}] + s^{-1}a^n st^{-1}[1, a^p] + s^{-1}a^n st^{-1}a^p[t^{-1}, 1]$, if nontrivial;
- $\gamma' = s^{-1}a^n st^{-1}a^p t[1, a^q]$, if nontrivial.

The path $\gamma = \gamma_s + \gamma_t + \gamma'$ is a path from 1 to

$$s^{-1}a^n st^{-1}a^p ta^q = (a^r b)^n (a^r b^{-1})^p a^q = a^{r(n+p)+q} b^{n-p} = a^{rm} b^m,$$

so $p = n - m$ and $q = -Lp$. Since p and q are proportional, γ_t and γ' are either both trivial or both nontrivial. Suppose they are nontrivial. There is a symmetry that exchanges γ_t with a path

$$\gamma_t' = s^{-1}a^n s[1, s^{-1}] + s^{-1}a^n ss^{-1}[1, a^{-p}] + s^{-1}a^n ss^{-1}a^{-p}[s^{-1}, 1]$$

of the same length. However, γ_t' and $\gamma_t + \gamma'$ have the same endpoints, and γ_t' is shorter, so γ could not have been geodesic if γ_t and γ' are nontrivial. Thus, if γ is geodesic then $\gamma = \gamma_s$. This reduces the problem of finding a geodesic from 1 to $(a^r b)^m$ to finding a geodesic from 1 to a^n.

A similar argument holds for geodesics from 1 to $(a^r b^{-1})^m$, so we can find geodesics from 1 to any point in $\langle a, b\rangle$ if we know geodesics from 1 to powers of a.

For powers of a the idea is that a^{mL}, $(a^r b)^m$, and $(a^r b^{-1})^m$ form an equilateral triangle in the plane, but the latter two can be shortened by a factor of L by pushing across a wall. Since $L \geq 6$, the savings of a factor of $L/2$ in length outweighs the added overhead from crossing walls.

For small powers of a we can find geodesics by inspection of the Cayley graph. For $0 \leqslant |p| \leqslant L/2 + 3$, the edge path a^p from 1 to a^p is a geodesic of length $|p|$. For $L/2 + 3 \leqslant p \leqslant L$ the edge path $s^{-1}ast^{-1}ata^{p-L}$ is a geodesic from 1 to a^p of length $6 + L - p$. We conclude that for $m > 0$ and $-L/2 + 3 \leq p \leq L/2 + 3$ there is a geodesic from 1 to a^{mL+p} of the form

$$\begin{aligned}[1, s^{-1}] + s^{-1}[1, a^m] + s^{-1}a^m[s^{-1}, 1] + s^{-1}a^m s[1, t^{-1}]\\ + s^{-1}a^m st^{-1}[1, a^m] + s^{-1}a^m st^{-1}a^m[t^{-1}, 1] + s^{-1}a^m st^{-1}a^m t[1, a^p]\end{aligned}$$

We can now find geodesics from 1 to powers of a by induction, and from these we know a geodesic from 1 to any $a^x b^y$. We see an example in Figure 5, where trapezoids are walls and triangles are contained in planes. The top half boundary and bottom half boundary of the figure each give geodesics of length $5 \cdot 2^5 - 4$ between 1 and a^{L^5}. (This form of geodesic loop bears witness to the Dehn function [Brady and Bridson 2000], and inspired the name “snowflake group” [Brady et al. 2009].)

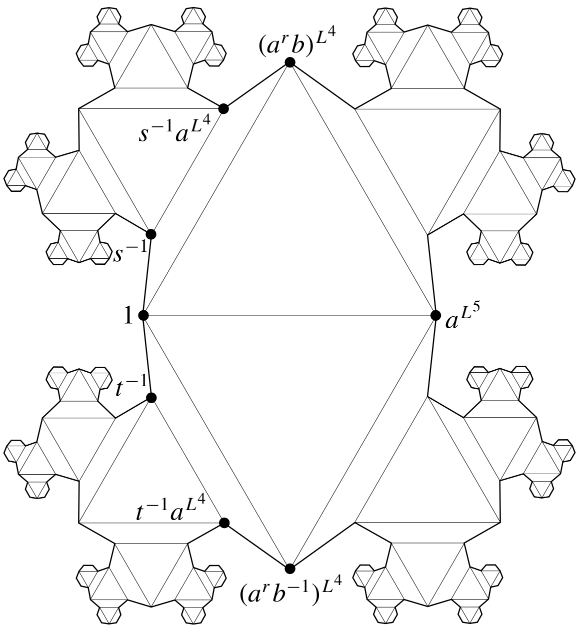


Figure 5. Snowflake: the boundary is a geodesic loop of length $2(5 \cdot 2^5 - 4)$.

11C. *Projections to geodesics in $\mathcal{X}$.* In this section we consider two different geodesics: $\alpha(2n) = (s^{-1}t)^n$ and $\beta(n) = s^{-n}$. These are geodesics since for each of these paths, every edge crosses a distinct wall. Let $\mathcal{T}$ be the Bass–Serre tree of G, and let $o \in \mathcal{T}$ be the vertex fixed by the subgroup $\langle a, b\rangle$. The orbit map $g \mapsto g\,.\,o$ sends each of α and β isometrically to a geodesic in $\mathcal{T}$. We will use π_α to denote closest point projection to α, both in $\mathcal{X}$ and in $\mathcal{T}$, and similarly for β.

Both of these geodesics have the property that for any vertices at distance at least three in the corresponding geodesic of the Bass–Serre tree, the pointwise stabilizers of the pair of vertices is trivial. We might hope, in analogy to Theorem 9.5, that these would be strongly contracting geodesics. As in Theorem 9.5, $\langle s^{-1}t\rangle$ and $\langle s\rangle$ are hyperbolically embedded subgroups in G, but, of the two, we will see only $s^{-1}t$ is strongly contracting.

Geodesic α. We claim that closest point projection $\pi_\alpha : \mathcal{X} \to \alpha$ is coarsely well defined and strongly contracting. First, consider π_α on $\langle a, b\rangle$. The geodesic α enters $\langle a, b\rangle$ through the incoming t-wall V at 1, and exits through the outgoing s^{-1}-wall W at 1.

Lemma 11.2. *For every $v \in V$ and every $w \in W$ there exists a geodesic from v to w that includes the vertex* 1.

Proof. The lemma follows from the discussion of geodesics in Section 11B. $\square$

Lemma 11.3. *The orbit map $\mathcal{X} \to \mathcal{T}$ defined by $g \mapsto g\,.\,o$ coarsely commutes with closest point projection to α. In particular, closest point projection to α in $\mathcal{X}$ is coarsely well defined.*

Proof. Suppose $z \in \mathcal{X}$ is some vertex that is separated from 1 by V, and suppose there is an $n \geqslant 0$ such that $\alpha(n) \in \pi_\alpha(z)$. Let σ be a geodesic from z to $\alpha(n)$. Write $\sigma = \sigma_1 + \sigma_2 + \sigma_3$, where σ_2 is the subsegment of σ from the first time σ crosses V until the first time σ reaches W. By Lemma 11.2, we can replace σ_2 by a geodesic segment $\sigma_2' + \sigma_2''$ where the concatenation point is 1. This means that z is connected to $1 = \alpha(0)$ by a path $\sigma_1 + \sigma_2'$. By hypothesis, the length of this path is at least the length of σ, so σ_2'' and σ_3 are trivial and $n = 0$. It follows immediately that the orbit map $\mathcal{X} \to \mathcal{T}$ commutes with π_α up to an error of 4. (In fact, a little more work will show the error is at most 2.) $\square$

Lemma 11.4 (bounded geodesic image property for π_α). *For any geodesic σ in $\mathcal{X}$, if the diameter of $\pi_\alpha(\sigma\,.\,o)$ is at least 5, then $\sigma \cap \alpha \neq \varnothing$.*

Proof. Suppose $\alpha([-1, 3]).o \subset \pi_\alpha(\sigma.o)$. Then σ crosses the walls V, W, $s^{-1}tV$, and $s^{-1}tW$. Write σ as a concatenation of geodesic subsegments $\sigma_1+\sigma_2+\sigma_3+\sigma_4+\sigma_5$, where σ_1 is all of σ prior to the first V crossing, σ_2 is the part of σ between the first V crossing and the last W crossing, σ_3 is the part between the last W crossing and the first $s^{-1}tV$ crossing, which included edges labeled s^{-1} and t, σ_4 is the part from the first $s^{-1}tV$ crossing until the last $s^{-1}tW$ crossing, and σ_5 is the remainder of σ. We can apply Section 11B to replace σ_2 by a geodesic $\sigma_2' + \sigma_2''$ with the same endpoints and concatenated at 1. Similarly, we can replace σ_4 by a geodesic $\sigma_4' + \sigma_4''$ with the same endpoints and concatenated at $s^{-1}t$. But then we can replace the subsegment $\sigma_2 + \sigma_3 + \sigma_4$ of σ by the path $\sigma_2'' + [1, s^{-1}t] + \sigma_4''$ with the same endpoints. This path is strictly shorter unless σ_2'' and σ_4'' are trivial. This means that $[1, s^{-1}t] \subset \sigma \cap \alpha$. $\square$

By Proposition 2.9, this means the following:

Corollary 11.5. *The element $s^{-1}t$ is strongly contracting for $G \curvearrowright \mathcal{X}$.*

Together with Theorem 6.4, this proves Theorem 11.1.

Geodesic β. Using out knowledge of geodesics from Section 11B, we see that the closest point of the s^{-1}-wall at 1 to the point a^{L^k} is $(a^r b)^{L^{k-1}}$, which is the midpoint of a geodesic from 1 to a^{L^k}. This geodesic coincides with β on the interval from 1 to s^{-k}. It follows that $\pi_\beta(a^{L^j}) = \beta(j)$ for all $j \geqslant 0$.

For $0 < j < k$ there is a geodesic $\sigma_{j,k}$ from a^{L^j} to a^{L^k} such that $d(\sigma_{j,k}, \beta) = d(a^{L^j}, \beta)$; see Figure 6. Letting j and $k - j$ grow large, the geodesics $\sigma_{j,k}$ stay outside large neighborhoods of β but have large projections to β. Therefore, π_β is not strongly contracting, since it does not enjoy the bounded geodesic image property.

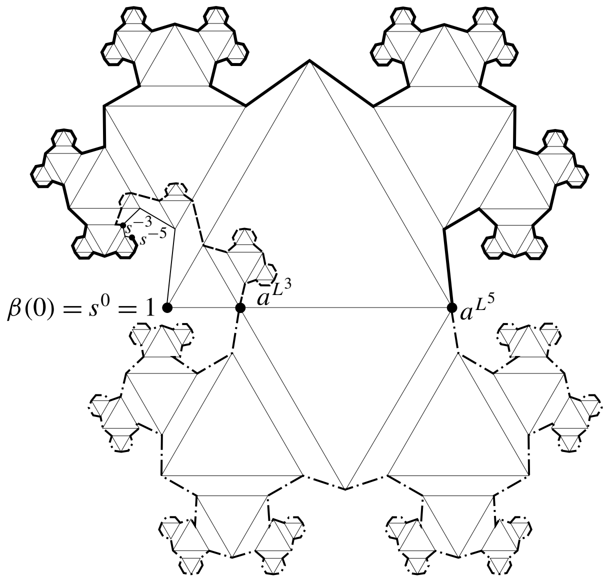


Figure 6. Geodesics $[a^{L^3}, \pi_\beta(a^{L^3})]$ (dashed), $[a^{L^5}, \pi_\beta(a^{L^5})]$ (solid), and $\sigma_{3,5} = [a^{L^3}, a^{L^5}]$ (dash-dot).

Acknowledgments

Arzhantseva and Cashen were supported by the European Research Council (ERC) grant of Goulnara ARZHANTSEVA, grant agreement #259527 and the Erwin Schrödinger Institute workshop “Geometry of Computation in Groups”. Cashen was also partially supported by the Austrian Science Fund (FWF):M1717-N25. Tao was supported by NSF grant DMS-1311834.

References

[Algom-Kfir 2011] Y. Algom-Kfir, “Strongly contracting geodesics in outer space”, *Geom. Topol.* **15**:4 (2011), 2181–2233. MR 2862155 Zbl 1250.20019

[Anderson et al. 2007] J. W. Anderson, J. Aramayona, and K. J. Shackleton, “An obstruction to the strong relative hyperbolicity of a group”, *J. Group Theory* **10**:6 (2007), 749–756. MR 2009c:20075 Zbl 1188.20041

[Arzhantseva and Lysenok 2002] G. N. Arzhantseva and I. G. Lysenok, “Growth tightness for word hyperbolic groups”, *Math. Z.* **241**:3 (2002), 597–611. MR 2003h:20077 Zbl 1045.20035

[Arzhantseva and Lysenok 2006] G. N. Arzhantseva and I. G. Lysenok, “A lower bound on the growth of word hyperbolic groups”, *J. London Math. Soc.* (2) **73**:1 (2006), 109–125. MR 2006k:20088 Zbl 1093.20024

[Athreya et al. 2012] J. Athreya, A. Bufetov, A. Eskin, and M. Mirzakhani, “Lattice point asymptotics and volume growth on Teichmüller space”, *Duke Math. J.* **161**:6 (2012), 1055–1111. MR 2913101 Zbl 1246.37009

[Ballmann 1995] W. Ballmann, *Lectures on spaces of nonpositive curvature*, DMV Seminar **25**, Birkhäuser, Basel, 1995. MR 97a:53053 Zbl 0834.53003

[Ballmann and Brin 1995] W. Ballmann and M. Brin, "Orbihedra of nonpositive curvature", *Inst. Hautes Études Sci. Publ. Math.* **82** (1995), 169–209. MR 97i:53049 Zbl 0866.53029

[Ballmann and Buyalo 2008] W. Ballmann and S. Buyalo, "Periodic rank one geodesics in Hadamard spaces", pp. 19–27 in *Geometric and probabilistic structures in dynamics*, edited by K. Burns et al., Contemp. Math. **469**, Amer. Math. Soc., Providence, 2008. MR 2010c:53062 Zbl 1167.53034

[Behrstock and Charney 2012] J. Behrstock and R. Charney, "Divergence and quasimorphisms of right-angled Artin groups", *Math. Ann.* **352**:2 (2012), 339–356. MR 2874959 Zbl 1251.20036

[Behrstock et al. 2009] J. Behrstock, C. Druţu, and L. Mosher, "Thick metric spaces, relative hyperbolicity, and quasi-isometric rigidity", *Math. Ann.* **344**:3 (2009), 543–595. MR 2010h:20094

[Bestvina and Fujiwara 2009] M. Bestvina and K. Fujiwara, "A characterization of higher rank symmetric spaces via bounded cohomology", *Geom. Funct. Anal.* **19**:1 (2009), 11–40. MR 2010m:53060 Zbl 1203.53041

[Bestvina et al. 2014] M. Bestvina, K. Bromberg, and K. Fujiwara, "Constructing group actions on quasi-trees and applications to mapping class groups", *Publications mathématiques de l'IHÉS* (online publication October 2014), 1–64.

[Bowditch 2012] B. H. Bowditch, "Relatively hyperbolic groups", *Internat. J. Algebra Comput.* **22**:3 (2012), 1250016, 66. MR 2922380 Zbl 1259.20052

[Brady and Bridson 2000] N. Brady and M. R. Bridson, "There is only one gap in the isoperimetric spectrum", *Geom. Funct. Anal.* **10**:5 (2000), 1053–1070. MR 2001j:20046 Zbl 0971.20019

[Brady et al. 2009] N. Brady, M. R. Bridson, M. Forester, and K. Shankar, "Snowflake groups, Perron–Frobenius eigenvalues and isoperimetric spectra", *Geom. Topol.* **13**:1 (2009), 141–187. MR 2010a:20090 Zbl 1228.20031

[Bridson 2010] M. R. Bridson, "Semisimple actions of mapping class groups on CAT(0) spaces", pp. 1–14 in *Geometry of Riemann surfaces*, edited by F. P. Gardiner, London Math. Soc. Lecture Note Ser. **368**, Cambridge Univ. Press, Cambridge, 2010. MR 2011i:30041 Zbl 1204.30030

[Bridson and Haefliger 1999] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften **319**, Springer, Berlin, 1999. MR 2000k:53038 Zbl 0988.53001

[Burger and Mozes 1997] M. Burger and S. Mozes, "Finitely presented simple groups and products of trees", *C. R. Acad. Sci. Paris Sér. I Math.* **324**:7 (1997), 747–752. MR 98g:20041 Zbl 0966.20013

[Caprace and Fujiwara 2010] P.-E. Caprace and K. Fujiwara, "Rank-one isometries of buildings and quasi-morphisms of Kac–Moody groups", *Geom. Funct. Anal.* **19**:5 (2010), 1296–1319. MR 2011e:20061 Zbl 1206.20046

[Caprace and Sageev 2011] P.-E. Caprace and M. Sageev, "Rank rigidity for CAT(0) cube complexes", *Geom. Funct. Anal.* **21**:4 (2011), 851–891. MR 2012i:20049 Zbl 1266.20054

[Cashen and Tao 2014] C. H. Cashen and J. Tao, "Growth tight actions of product groups", preprint, 2014. arXiv 1404.4335v2

[Charney 2007] R. Charney, "An introduction to right-angled Artin groups", *Geom. Dedicata* **125** (2007), 141–158. MR 2008f:20076 Zbl 1152.20031

[Coulon 2013] R. Coulon, "Growth of periodic quotients of hyperbolic groups", *Algebr. Geom. Topol.* **13**:6 (2013), 3111–3133. MR 3248727 Zbl 06213053

[Dahmani et al. 2011] F. Dahmani, V. Guirardel, and D. Osin, "Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces", preprint, 2011. arXiv 1111.7048v5

[Dal'bo et al. 2000] F. Dal'bo, J.-P. Otal, and M. Peigné, "Séries de Poincaré des groupes géométriquement finis", *Israel J. Math.* **118** (2000), 109–124. MR 2001g:37040 Zbl 0968.53023

[Dal'Bo et al. 2011] F. Dal'Bo, M. Peigné, J.-C. Picaud, and A. Sambusetti, "On the growth of quotients of Kleinian groups", *Ergodic Theory Dynam. Systems* **31**:3 (2011), 835–851. MR 2012f:37069 Zbl 1233.37021

[Druţu 2009] C. Druţu, "Relatively hyperbolic groups: geometry and quasi-isometric invariance", *Comment. Math. Helv.* **84**:3 (2009), 503–546. MR 2011a:20111

[Duchin and Rafi 2009] M. Duchin and K. Rafi, "Divergence of geodesics in Teichmüller space and the mapping class group", *Geom. Funct. Anal.* **19**:3 (2009), 722–742. MR 2011c:30115 Zbl 1187.30041

[Eskin et al. 2012] A. Eskin, M. Mirzakhani, and K. Rafi, "Counting closed geodesics in strata", preprint, 2012. arXiv 1206.5574v1

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

[Gerasimov 2012] V. Gerasimov, "Floyd maps for relatively hyperbolic groups", *Geom. Funct. Anal.* **22**:5 (2012), 1361–1399. MR 2989436 Zbl 1276.20050

[Grigorchuk and de la Harpe 1997] R. Grigorchuk and P. de la Harpe, "On problems related to growth, entropy, and spectrum in group theory", *J. Dynam. Control Systems* **3**:1 (1997), 51–89. MR 98d:20039 Zbl 0949.20033

[Gromov 1981] M. Gromov, "Groups of polynomial growth and expanding maps", *Inst. Hautes Études Sci. Publ. Math.* **53**:1 (1981), 53–73. MR 83b:53041 Zbl 0474.20018

[Groves and Manning 2007] D. Groves and J. F. Manning, "Fillings, finite generation and direct limits of relatively hyperbolic groups", *Groups Geom. Dyn.* **1**:3 (2007), 329–342. MR 2008f:20106 Zbl 1169.20023

[Groves and Manning 2008] D. Groves and J. F. Manning, "Dehn filling in relatively hyperbolic groups", *Israel J. Math.* **168** (2008), 317–429. MR 2009h:57030 Zbl 1211.20038

[Hamenstädt 2013] U. Hamenstädt, "Bowen's construction for the Teichmüller flow", *J. Mod. Dyn.* **7**:4 (2013), 489–526. MR 3177770 Zbl 1291.30245

[Hubbard 2006] J. H. Hubbard, *Teichmüller theory and applications to geometry, topology, and dynamics*, vol. 1, Matrix Editions, Ithaca, 2006. MR 2008k:30055 Zbl 1102.30001

[Huber 1956] H. Huber, "Über eine neue Klasse automorpher Funktionen und ein Gitterpunktproblem in der hyperbolischen Ebene. I", *Comment. Math. Helv.* **30** (1956), 20–62. MR 17,603b

[Kapovich and Leeb 1998] M. Kapovich and B. Leeb, "3-manifold groups and nonpositive curvature", *Geom. Funct. Anal.* **8**:5 (1998), 841–852. MR 2000a:57040 Zbl 0915.57009

[Manning 2005] J. F. Manning, "Geometry of pseudocharacters", *Geom. Topol.* **9** (2005), 1147–1185. MR 2006j:57002 Zbl 1083.20038

[Margulis 1991] G. A. Margulis, *Discrete subgroups of semisimple Lie groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **17**, Springer, Berlin, 1991. MR 92h:22021 Zbl 0732.22008

[Minasyan and Osin 2015] A. Minasyan and D. Osin, "Acylindrical hyperbolicity of groups acting on trees", *Math. Ann.* **362**:3-4 (2015), 1055–1105. MR 3368093

[Minsky 1996] Y. N. Minsky, "Quasi-projections in Teichmüller space", *J. Reine Angew. Math.* **473** (1996), 121–136. MR 97b:32020 Zbl 0848.30031

[Mumford 1971] D. Mumford, "A remark on Mahler's compactness theorem", *Proc. Amer. Math. Soc.* **28** (1971), 289–294. MR 43 #2157 Zbl 0215.23202

[Papadopoulos 2007] A. Papadopoulos, "Introduction to Teichmüller theory, old and new", pp. 1–30 in *Handbook of Teichmüller theory*, vol. 1, edited by A. Papadopoulos, IRMA Lect. Math. Theor. Phys. **11**, Eur. Math. Soc., Zürich, 2007. MR 2009c:30115

[Parkkonen and Paulin 2015] J. Parkkonen and F. Paulin, "On the hyperbolic orbital counting problem in conjugacy classes", *Math. Z.* **279**:3-4 (2015), 1175–1196. MR 3318265 Zbl 06422655

[Reinfeldt and Weidmann 2014] C. Reinfeldt and R. Weidmann, "Makanin–Razborov diagrams for hyperbolic groups", preprint, Christian-Albrechts-Universität zu Kiel, 2014, Available at http://tinyurl.com/o7qglq9.

[Sabourau 2013] S. Sabourau, "Growth of quotients of groups acting by isometries on Gromov-hyperbolic spaces", *J. Mod. Dyn.* **7**:2 (2013), 269–290. MR 3106714 Zbl 1286.20057

[Sambusetti 2002a] A. Sambusetti, "Growth tightness of free and amalgamated products", *Ann. Sci. École Norm. Sup.* (4) **35**:4 (2002), 477–488. MR 2004c:20042 Zbl 1018.20036

[Sambusetti 2002b] A. Sambusetti, "Growth tightness of surface groups", *Expo. Math.* **20**:4 (2002), 345–363. MR 2003k:57004 Zbl 1028.53033

[Sambusetti 2003] A. Sambusetti, "Growth tightness of negatively curved manifolds", *C. R. Math. Acad. Sci. Paris* **336**:6 (2003), 487–491. MR 2004a:53048 Zbl 1030.53039

[Sambusetti 2004] A. Sambusetti, "Growth tightness in group theory and Riemannian geometry", pp. 341–352 in *Recent advances in geometry and topology* (Cluj-Napoca, Romania, 2003), edited by D. Andrica and P. A. Blaga, Cluj Univ. Press, 2004. MR 2005m:20078 Zbl 1086.53055

[Scott and Wall 1979] P. Scott and T. Wall, "Topological methods in group theory", pp. 137–203 in *Proceedings of a Symposium on Homological and Combinational Techniques in Group Theory* (Durham, 1977), edited by C. T. C. Wall, London Math. Soc. Lecture Note Ser. **36**, Cambridge Univ. Press, 1979. MR 81m:57002 Zbl 0423.20023

[Sela 1999] Z. Sela, "Endomorphisms of hyperbolic groups, I: The Hopf property", *Topology* **38**:2 (1999), 301–321. MR 99m:20081 Zbl 0929.20033

[Shukhov 1999] A. G. Shukhov, "On the dependence of the growth exponent on the length of the defining relation", *Mat. Zametki* **65**:4 (1999), 612–618. MR 2000i:20052

[Sisto 2011] A. Sisto, "Contracting elements and random walks", preprint, 2011. arXiv 1112.2666v2

[Sisto 2012] A. Sisto, "On metric relative hyperbolicity", preprint, 2012. arXiv 1210.8081v1

[Sisto 2013] A. Sisto, "Quasi-convexity of hyperbolically embedded subgroups", preprint, 2013. arXiv 1310.7753

[Stallings 1971] J. Stallings, *Group theory and three-dimensional manifolds*, Yale Mathematical Monographs **4**, Yale University Press, New Haven-London, 1971. MR 0415622 Zbl 0241.57001

[Thurston 1998] W. P. Thurston, "Hyperbolic structures on 3-manifolds, II: Surface groups and 3-manifolds which fiber over the circle", preprint, 1998. arXiv math/9801045v1

[Yang 2013] W. Yang, "Growth tightness of groups with nontrivial Floyd boundary", preprint, 2013. arXiv 1301.5623v1

Received February 20, 2015.

GOULNARA N. ARZHANTSEVA
FAKULTÄT FÜR MATHEMATIK
UNIVERITÄT WIEN
OSKAR-MORGENSTERN-PLATZ 1
1090 WIEN
AUSTRIA
goulnara.arzhantseva@univie.ac.at

Christopher H. Cashen
Fakultät für Mathematik
Universität Wien
Oskar-Morgenstern-Platz 1
1090 Vienna
Austria
christopher.cashen@univie.ac.at

Jing Tao
Department of Mathematics
University of Oklahoma
601 Elm Avenue Rm 423
Norman, OK 73019
United States
jing@ou.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.278.51

A FLAG STRUCTURE ON A CUSPED HYPERBOLIC 3-MANIFOLD

ELISHA FALBEL AND RAFAEL SANTOS THEBALDI

A flag structure on a 3-manifold is an (X, G) structure where $G = \mathrm{SL}(3, \mathbb{R})$ and X is the space of flags on the 2-dimensional projective space. We construct a flag structure on a cusped hyperbolic manifold with unipotent boundary holonomy. The holonomy representation can be obtained from a punctured torus group representation into $\mathrm{SL}(3, \mathbb{R})$ which is equivariant under a pseudo-Anosov.

1. Introduction

A flag structure on a 3-manifold is an (X, G) structure, where $G = \mathrm{SL}(3, \mathbb{R})$ and X is the space of flags on the 2-dimensional projective space, that is, the space of pairs: point and line containing it. The most direct construction of such structures starts with a real projective surface or orbifold. The projectivization of its tangent bundle is a Seifert manifold and has a natural flag structure. Other constructions on Seifert manifolds are studied in [Barbot 2001]. Note that projective structures on 3-manifolds concern instead the group $\mathrm{SL}(4, \mathbb{R})$; see [Cooper et al. 2015].

Representations of fundamental groups of three manifolds into $\mathrm{SL}(3, \mathbb{R})$ were obtained in [Falbel et al. 2015], following the method described in [Bergeron et al. 2014]; see also the CURVE project [Falbel et al. 2015–]. A fundamental question is whether these representations correspond to holonomies of flag structures on the manifold.

The goal of this paper is to construct a flag structure on a cusped hyperbolic manifold with unipotent boundary holonomy; see Theorem 6.8. We introduce a general method of construction via gluings of tetrahedra which are defined on the flag space. The tetrahedra are canonical up to a finite choice related to an order on the 0-skeleton of an ideal triangulation of the manifold, once one fixes a decoration (that is a choice of a flag at each vertex) satisfying certain compatibility conditions; see [Bergeron et al. 2014]. Definitions of simplices in Grassmannian spaces (although not containing the case of flag space) were also considered in [Gelfand and MacPherson 1982] and inspired our definition of tetrahedron. In

MSC2010: primary 57M50; secondary 57S30.
Keywords: flag structures, hyperbolic structures, SL(3, R) representations.

the Appendix we describe a slicing of the tetrahedra which allows an algorithmic treatment of the many compatibility verifications needed in the paper.

The method presented here can be considered as a flag structure analog of Thurston's construction [1981] of hyperbolic structures on cusped manifolds by gluing ideal hyperbolic tetrahedra and of the construction of CR structures, as in [Falbel 2008].

The holonomy representation of the structure we obtained is not faithful. It turns out that the manifold m009 we analyzed here has holonomy group contained in a triangle group of type $(3, 3, 5)$; see the end of the Appendix. An isomorphic triangle group was obtained in [Deraux 2015], where the holonomy representation has values in $\mathrm{PU}(2, 1)$. These representations are Galois conjugates, as explained in [Falbel et al. 2015]; indeed, they are all parametrized by solutions of a degree four irreducible polynomial in one variable. Two solutions correspond to conjugate representations in $\mathrm{PU}(2, 1)$ and the other two to two dual flag structures.

It is interesting to remark that the manifold m009 is fibered over the circle with fiber a punctured torus. The representation into $\mathrm{SL}(3, \mathbb{R})$ of the fiber surface group is then equivariant with respect to the mapping class group element defining the bundle.

2. Flag structures on 3-manifolds

A flag structure on a 3-manifold is an (X, G) structure, where X is a homogeneous space described in the following paragraph and $G = \mathrm{SL}(3, \mathbb{R}) = \mathrm{PGL}(3, \mathbb{R})$.

The homogeneous space X is the space of flags in $\mathbb{P}(\mathbb{R}^3)$. An affine flag in $V = \mathbb{R}^3$ is a pair (line, plane), the line belonging to the plane. They project to flags in $\mathbb{P}(V)$, that is, pairs (point, line). Using the dual vector space V^* and the projective spaces $\mathbb{P}(V)$ and $\mathbb{P}(V^*)$, define the spaces of *flags* $\mathcal{F}l$ by the following:

$$\mathcal{F}l = \{([p], [l]) \in \mathbb{P}(V) \times \mathbb{P}(V^*) \mid l(p) = 0\}.$$

The action of $\mathrm{SL}(3, \mathbb{R})$ on V induces an action on $\mathbb{P}(V)\times\mathbb{P}(V^*)$. Indeed, identify V and V^* using the canonical scalar product and then, via this identification, the contragredient action (that is $g\,.\,v = (g^{-1})^T v$) on V^*. We denote by π_1 and π_2 the two projections of $\mathcal{F}l$ into $\mathbb{P}(V)$ and $\mathbb{P}(V^*)$, respectively.

Observe that

$$\mathcal{F}l = \mathrm{SL}(3, \mathbb{R})/B,$$

where B is the Borel subgroup of upper triangular matrices in $\mathrm{SL}(3, \mathbb{R})$. The flag space is identified with the projectivization of the tangent bundle to $P(V)$, and the differential action of $\mathrm{SL}(3, \mathbb{R})$ on the tangent bundle induces the above action. Observe that, in fact, $\mathrm{SL}(3, \mathbb{R})$ acts on the unit tangent bundle of $P(V)$ (which has

S^3 as a double cover); therefore, the double cover of $\mathrm{SL}(3, \mathbb{R})$ (which is simply connected) acts on the sphere S^3.

Definition 2.1. A *flag structure* on a 3-manifold M is a $(\mathscr{F}l, \mathrm{SL}(3, \mathbb{R}))$-structure on that manifold.

The involution $\Theta(v, w) = (w, v)$ on $\mathscr{F}l$ and the Cartan involution $\theta(g) = (g^{-1})^T$ on $\mathrm{SL}(3, \mathbb{R})$ satisfy

$$\Theta \circ g = \theta(g) \circ \Theta.$$

Given a flag structure on a 3-manifold, we call a *dual flag structure* the structure obtained by using transition functions composed with θ.

2.2. ***Coordinates in*** $\mathbb{P}(V)$**.** To facilitate visualization of the flags we will choose a chart (called the *preferred chart*) on $\mathbb{P}(V)$. Consider the hyperplane in $\mathbb{R}^3$ defined by the three basis unit vectors, that is

$$x + y + z = 1.$$

The chart is defined by projecting lines passing through the origin in that hyperplane and imposing that

$$[1, 0, 0] \mapsto (0, 0), \quad [0, 1, 0] \mapsto (1, 0), \quad [0, 0, 1] \mapsto (0, 1).$$

The chart is defined on the complement of the projectivization of the plane $x + y + z = 0$ and has the expression

$$[x, y, z] \mapsto \left(\frac{y}{x+y+z}, \frac{z}{x+y+z}\right).$$

Observe that, on the hyperplane, $[x, y, z] \mapsto (y, z)$ and $[1, 1, 1] \mapsto \left(\frac{1}{3}, \frac{1}{3}\right)$.

Given a flag $[[x, y, z], [a, b, c]]$, with $x + y + z = 1$, the point $[x, y, z]$ and the line on $\mathbb{P}(V)$ defined by the image of the plane orthogonal to the vector (a, b, c) are described in the chart above by:

- the point (y, z),
- the line defined by the vector $(a - c, b - a)$ passing through the point (y, z).

Therefore, the line makes an angle θ with the first coordinate axis satisfying

$$\tan\theta = \frac{b-a}{a-c} \tag{2.2.1}$$

with the horizontal direction. Figure 1 shows three flags corresponding to planes passing through the three basis vectors in $\mathbb{R}^3$.

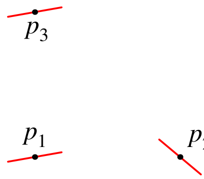


Figure 1. Three flags corresponding to planes passing through the three basis vectors in $\mathbb{R}^3$.

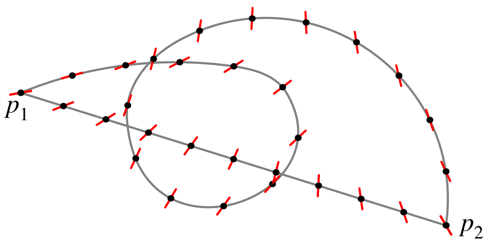


Figure 2. Two simple paths of flags projected into $\mathbb{P}(V)$.

3. Edges

One can join a pair of flags by simple paths (see Figure 2), but there is a canonical construction of a unique line containing two flags.

Consider two flags in generic position, that is, $f_1 = (p_1, l_1)$, $f_2 = (p_2, l_2)$ such that $l_i(p_j) \neq 0$ if $i \neq j$. The action of $\mathrm{SL}(3, \mathbb{R})$ is transitive on these pairs. There exists a unique point p_{12} such that $l_i(p_{12}) = 0$, for $i = 1, 2$. Up to the action of $\mathrm{SL}(3, \mathbb{R})$ we can normalize so that

$$p_1 = (1,0,0),\ l_1 = (0,1,0) \quad \text{and} \quad p_2 = (0,1,0),\ l_2 = (1,0,0).$$

The intersection point of the two lines is $p_{12} = (0, 0, 1)$. Projective transformations fixing the three points are diagonal and they preserve the line $[p_1, p_2]$. For each line l passing through p_{12} we consider its intersection p with the line $[p_1, p_2]$; see Figure 3. This defines a circle of flags (p, l) containing f_1 and f_2. It is divided

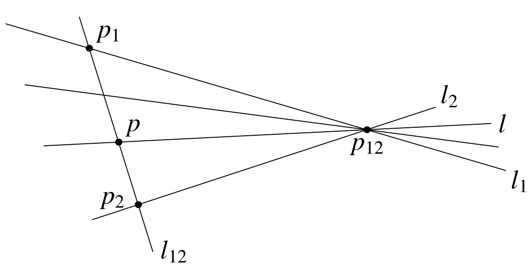


Figure 3. A segment between two flags.

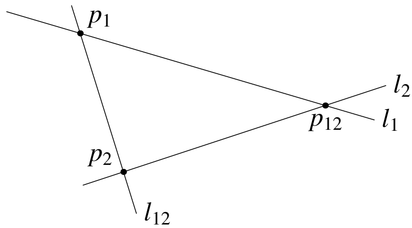


Figure 4. Points and lines fixed by H_{12}^0.

into two segments with two given flags as boundaries. Following [Gelfand and MacPherson 1982], we let H_{12}^0 be the connected component of the identity of the group preserving the points p_1, p_2, p_{12}. It preserves the lines l_1, l_2, l_{12} (see Figure 4), and the two segments are orbits of its action on the space of flags whose closure contains the flags f_1 and f_2. In the normalization above we have

$$H_{12}^0 = \begin{bmatrix} h_1 & 0 & 0 \\ 0 & h_2 & 0 \\ 0 & 0 & h_3 \end{bmatrix}$$

with $h_i > 0$. The circle of flags is given by

$$p = [\lambda_1, \lambda_2, 0], \quad l = [\lambda_2, -\lambda_1, 0].$$

More generally, if $f_1 = (p_1, l_1)$, $f_2 = (p_2, l_2)$ are two flags in generic position, then the line containing the flags is

$$\left(\lambda_1 p_1 + \lambda_2 p_2,\ \frac{\lambda_2}{l_2(p_1)} l_2 - \frac{\lambda_1}{l_1(p_2)} l_1\right).$$

The line is divided into two segments, corresponding to the relative signs of λ_1 and λ_2. Observe that, if the two flags are in the preferred chart, then only one of the segments is contained in the chart. Indeed, one of the flags in the circle is not in the preferred chart as its corresponding point is at infinity for that chart.

The next lemma states a simple property of a segment between two flags. It is the basic technical result we need to construct the tetrahedra of flags and will be repeatedly used in the analysis of the example in the last section.

Lemma 3.1 (monotonicity lemma). *Let $f_1 = (p_1, l_1)$, $f_2 = (p_2, l_2)$ be two flags. Suppose they are contained in the preferred chart and the angles θ_1 and θ_2 of the lines in the chart coordinates satisfy $\theta_1 \le \theta_2$. Then, along the finite segment from f_1 to f_2, the angles of the projected lines are increasing (and satisfy $\theta_1 \le \theta \le \theta_2$).*

If $f_2' = (p_2, l_2')$ is another flag such that $\theta_2 < \theta_2'$ then, along the corresponding segment from f_1 to f_2', the angles of the projected lines satisfy $\theta < \theta'$.

Proof. We use the preferred chart (see Figure 3). The two lines l_1 and l_2 intersect at a point p_{12} (which might be at infinity). The lines belonging to the finite segment of flags between f_1 and f_2 are lines passing through p between l_1 and l_2. The angle of each line is also read at p_{12} and is clearly a monotone function between l_1 and l_2.

For the second assertion, in the preferred chart, suppose that the vertices p_1, p_2, and p_{12} determine a finite triangle (otherwise the lines l_1 and l_2 are parallel and the analysis is simpler). If $f_2' = (p_2, l_2')$ is another flag such that $\theta_2 \leq \theta_2'$, then p_{12}', the intersection of l_2' and l_1, is on the side $[p_1, p_{12}]$. Therefore, the line passing through p_{12} and a point t in $[p_1, p_2]$ has smaller angle than the line passing through p_{12}' and t. □

4. Triangles

By a *generic configuration of flags* $([p_i], [l_i])$, $1 \leq i \leq n+1$, we mean $n+1$ points $[p_i]$ in general position and $n+1$ lines l_i in $\mathbb{P}(V)$ such that $l_j(p_i) \neq 0$ if $i \neq j$. Recall that a configuration of ordered points in $\mathbb{P}(V)$ is said to be in *general position* when no three points are contained in the same line. Notice that we give priority to the points in the above definition and don't impose that the lines are in generic position.

Let (e_1, e_2, e_3) be the canonical basis of V and (e_1^*, e_2^*, e_3^*) its dual basis. Up to the action of $\mathrm{SL}(3, \mathbb{R})$, a generic configuration of three flags $([p_i], [l_i])$, $1 \leq i \leq 3$, can be normalized in these coordinates as

- $p_1 = (1, 0, 0)$, $\quad l_1 = (0, 1, 1)$,
- $p_2 = (0, 1, 0)$, $\quad l_2 = (1, 0, 1)$, and
- $p_3 = (0, 0, 1)$, $\quad l_3 = (z, 1, 0)$, with $z \neq 0$.

Therefore, the only invariant of a generic configuration of three flags, up to $\mathrm{SL}(3, \mathbb{R})$, is the triple ratio, given by

$$z = \frac{l_1(p_2)l_2(p_3)l_3(p_1)}{l_1(p_3)l_2(p_1)l_3(p_2)} \in \mathbb{R}^\times$$

Observe that the three lines of the triple of flags are linearly independent if and only if $z \neq -1$.

Given three flags in general position, $f_1 = (p_1, l_1)$, $f_2 = (p_2, l_2)$, $f_3 = (p_3, l_3)$, we may form a triangle (a 1-skeleton as in Figure 5) by choosing three edges as above. There are 8 possible choices, namely for each pair of flags in a chart one can choose either the bounded segment or the unbounded segment with end points given by the two flags.

Fixing a choice of edges, we define a *face* as an embedded 2-simplex whose boundary is the union of the three edges. Observe that this imposes a restriction on the 1-simplex; it should be null-homotopic. This is equivalent to the condition that

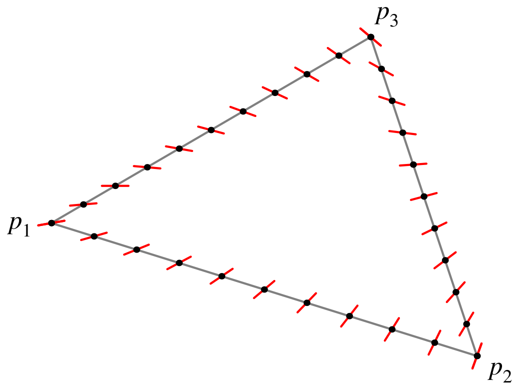


Figure 5. A triangle of flags projected into $\mathbb{P}(V)$.

both projections by π_1 and π_2 of the 1-skeleton should be null-homotopic. If the edges are as in the previous section, then there is a restriction on the triple ratio of a triple of flags:

Lemma 4.1. *A triple of flags defines a null-homotopic canonical* 1*-skeleton if and only if the triple ratio of the three flags is negative. In that case there are precisely four canonical* 1*-skeletons which are null-homotopic.*

The proof of the lemma consists of comparing the two possible situations in Figure 6, corresponding to negative and positive triple ratios respectively. To obtain the sign of the triple ratio one simply counts the number of times the lines separate the points not contained in them.

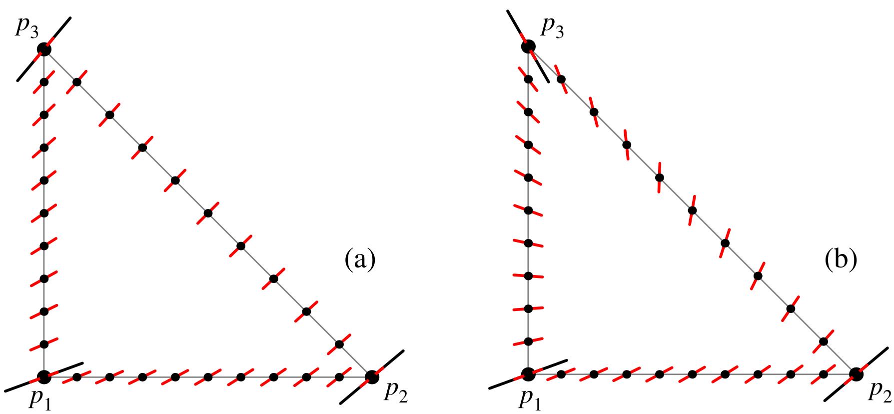


Figure 6. Each diagram shows three flags and the segments joining them, projected in the preferred chart. We only draw the finite triangle. The Euler number of a vector field parallel to the line field along the triangle is 0 in case (a) and has absolute value 1 in case (b). The triple ratio is negative in (a) and positive in (b).

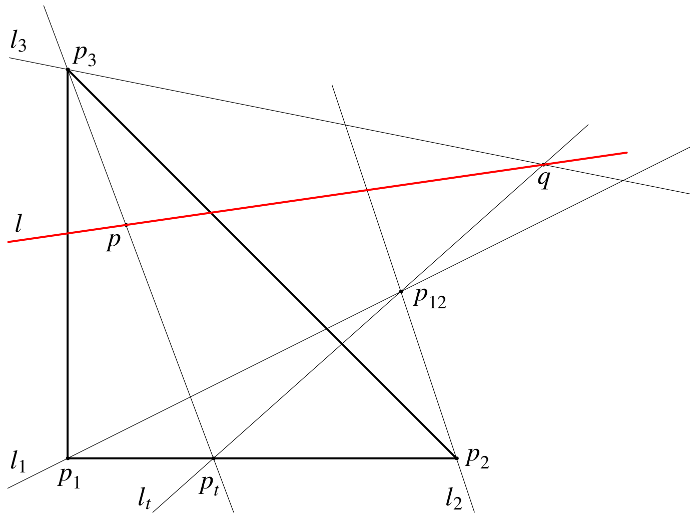


Figure 7. A synthetic construction of the flag (p, l) in the face F_{312}. The triple ratio is negative and f_3 is the source of the face.

Once the 1-skeleton is defined, we should define a 2-simplex whose boundary is the given 1-skeleton. A particular canonical choice is given as a union of segments:

Definition 4.2. A face F_{123} in the flag space, with vertices f_i, $i = 1, 2, 3$ (with negative triple ratio) and a choice of edges $[f_1, f_2]$, $[f_2, f_3]$, $[f_3, f_1]$, is the 2-skeleton, which is the union of segments between f_1 and f_t, where $f_t \in [f_2, f_3]$; that is,

$$F_{123} = \{ f \in \mathscr{F}l \mid f \in [f_1, f_t] \text{ for } f_t \in [f_2, f_3] \}.$$

The flag f_1 is called the *source of the face*. For example, in Figure 7 the edges are bounded segments, f_3 is the source of the face F_{312} and the triple ratio is negative. Notice that with the same vertices and edges we can construct the face F_{213} in the same way, but we can't obtain the face F_{123} because as $f_t \in [f_2, f_3]$, there is a flag $f_0 = (p_0, l_0) \in [f_2, f_3]$ such that $p_1 \in l_0$, so the flags f_0 and f_1 are not in general position. Thus, given a triple of flags with negative triple ratio, the surface obtained is embedded with boundary the union of edges only for two good choices of the source.

If the triple of flags has positive triple ratio it will be impossible to fill up a triangle unless we change the 1-skeleton in the following way: in the configuration represented in Figure 8, there is a flag $f_0 = (p_0, l_0) \in [f_2, f_3]$ such that $p_1 \in l_0$, so the flags f_0 and f_1 are not in general position. In order to define the triangle we should add, along the points $p \in [p_0, p_1)$ the flags (p, l_0) and over the point p_1 the flags $\pi_1^{-1}(p_1)$. In this way the projection of the 1-skeleton is twice the generator

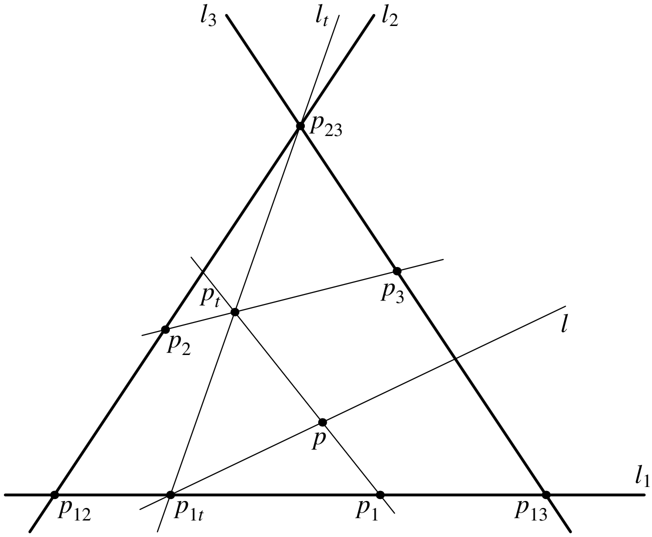


Figure 8. A synthetic construction of the flag (p, l) in the face F_{123}. The triple ratio is positive.

and therefore it is null-homotopic. In this paper, though, we will only use triples with negative ratio and good choices of the source of the face.

The 2-skeleton determines a triangle $T_{123} \subset \mathbb{P}(V)$ when projected by π_1 and $T^*_{123} \subset \mathbb{P}(V^*)$ when projected by π_2. That is,

$$\pi_1(F_{123}) = T_{123}, \quad \pi_2(F_{123}) = T^*_{123}.$$

5. Coordinates on a flag tetrahedron

In this section we recall the coordinates parametrizing configurations of four flags in the projective space $\mathbb{P}(\mathbb{R}^3)$, as in [Bergeron et al. 2014; Falbel et al. 2015]; also see [Fock and Goncharov 2007; Garoufalidis et al. 2011].

5.1. ***Coordinates for a tetrahedron of flags.*** Let $([p_i], [l_i])$, $1 \le i \le 4$, be a generic tetrahedron. Arrange these flags symbolically on a tetrahedron 1234 as in Figure 9. We define a set of 12 coordinates on the edges of the tetrahedron, one for each oriented edge.

To define the coordinate z_{ij} associated to the edge ij, we first define k and l such that the permutation $(1, 2, 3, 4) \mapsto (i, j, k, l)$ is even. The pencil of (projective) lines through the point p_i is a projective line $\mathbb{P}_1(k)$. We have four points on this projective line: the line $\ker(l_i)$ and the each of the lines going through p_i and one of the p_l, for $l \neq i$. We define z_{ij} as the cross-ratio of four flags by

$$z_{ij} := [\ker(l_i),\ (p_i p_j),\ (p_i p_k),\ (p_i p_l)].$$

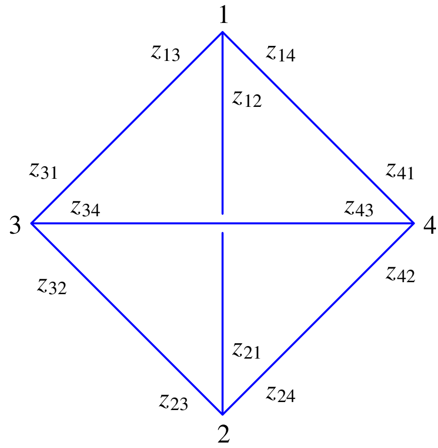


Figure 9. The z-coordinates.

We follow the usual convention that the cross-ratio of four points p_1, p_2, p_3, p_4 on a line is the value at p_4 of a projective coordinate taking value ∞ at p_1, 0 at p_2, and 1 at p_3. Figure 9 displays the coordinates.

At each face (ijk), oriented as the boundary of the tetrahedron (1234), we associate the triple ratio:

$$z_{ijk} = \frac{l_i(p_j)l_j(p_k)l_k(p_i)}{l_i(p_k)l_j(p_i)l_k(p_j)}.$$

Observe that if the same face with opposite orientation (ikj) is common to a second tetrahedron, then

$$z_{ikj} = \frac{1}{z_{ijk}}.$$

Of course there are relations between the whole set of coordinates. Fix an even permutation (i, j, k, l) of $(1, 2, 3, 4)$. First, for each face (ijk), the triple ratio is the opposite of the product of all cross-ratios "leaving" this face:

$$z_{ijk} = -z_{il}z_{jl}z_{kl}. \tag{5.1.1}$$

Second, the three cross-ratios leaving a vertex are algebraically related. For instance, in the vertex 1,

$$z_{13} = \frac{1}{1-z_{12}} \quad \text{and} \quad z_{14} = 1 - \frac{1}{z_{12}}, \tag{5.1.2}$$

and analogously for the other vertices. The next proposition shows that a tetrahedron is uniquely determined, up to the action of $\mathrm{SL}(3, \mathbb{R})$, by four numbers.

Proposition 5.2 [Bergeron et al. 2014, Proposition 2.4.1]. *The space of generic tetrahedra is parametrized by the 4-tuple $(z_{12}, z_{21}, z_{34}, z_{43})$ of elements in $\mathbb{R}\setminus\{0, 1\}$.*

In particular, one can normalize the coordinates of four flags up to the action of $\mathrm{SL}(3,\mathbb{R})$ as

$$\begin{aligned}
f_1:&\quad p_1=(1,0,0),\quad l_1=(0,\,z_{14},\,-1)=\Big(-1,\,1-\frac{1}{z_{12}},\,-1\Big),\\
f_2:&\quad p_2=(0,1,0),\quad l_2=\Big(\frac{1}{z_{24}},\,0,\,-1\Big)=(1-z_{21},\,0,\,-1),\\
f_3:&\quad p_3=(0,0,1),\quad l_3=(z_{34},\,-1,\,0),\\
f_4:&\quad p_4=(1,1,1),\quad l_4=\Big(z_{42},\,\frac{1}{z_{41}},\,-1\Big)=\Big(\frac{1}{1-z_{43}},\,\frac{z_{43}}{z_{43}-1},\,-1\Big).
\end{aligned}$$

6. Example: m009

The manifold m009 is an open manifold which has a complete hyperbolic structure with finite volume. It is obtained by gluing three tetrahedra $T_0(u_{ij})$, $T_1(v_{ij})$, and $T_2(w_{ij})$ as shown in Figure 10.

The face identifications are

$$\begin{aligned}
&(234)^0\leftrightarrow(243)^1,\quad (142)^0\leftrightarrow(314)^1,\quad (134)^0\leftrightarrow(143)^2,\\
&(123)^0\leftrightarrow(213)^2,\quad (142)^1\leftrightarrow(241)^2,\quad (123)^1\leftrightarrow(342)^2.
\end{aligned}$$

In [Falbel et al. 2015], we obtained a particular realization of these tetrahedra by 4-tuples of flags giving rise to representations into $\mathrm{SL}(3,\mathbb{R})$ with unipotent boundary holonomy. The invariants of the 4-tuple of flags all depend on

$$\gamma=-\tfrac{1}{2}+\tfrac{1}{2}\sqrt{5+4\sqrt{5}}.$$

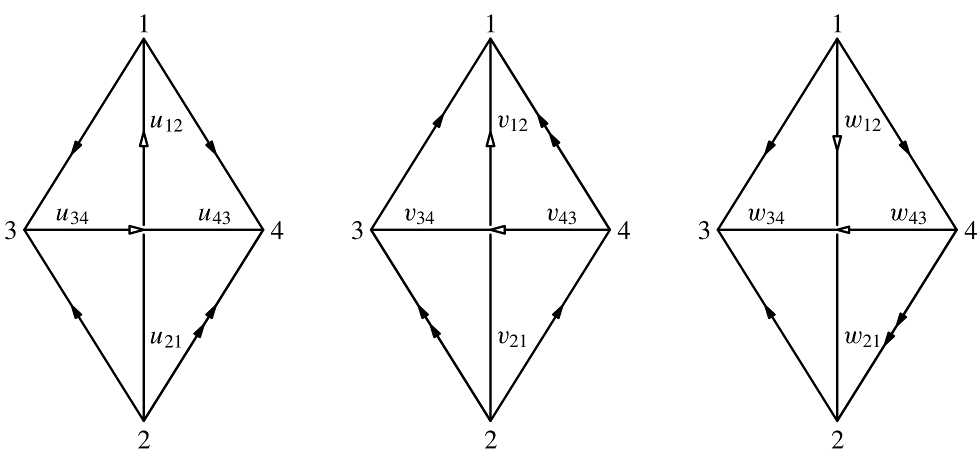


Figure 10. Three tetrahedra glued to obtain the manifold m009. The tetrahedra are numbered from 0 to 2 from left to right.

Explicitly:

$$u_{12} = w_{34} = \frac{\gamma+3}{\gamma+1}, \qquad u_{21} = w_{43} = \gamma, \qquad u_{34} = w_{12} = \frac{\gamma-2}{\gamma},$$
$$u_{43} = w_{21} = -1-\gamma, \quad v_{12} = v_{34} = \frac{1}{\gamma+3}, \quad v_{21} = v_{43} = \frac{1}{2-\gamma}.$$

The group obtained has rank one boundary holonomy, and one can chose generators, called meridian g_M and longitude g_L, satisfying $g_M g_L^2 = 1$.

The realization described above comes paired with another one giving rise to a dual flag structure. It is also related to a representation of the fundamental group in PU(2, 1) with boundary holonomy of rank one, which seems to give rise to a uniformizable CR structure on m009 [Deraux 2015].

6.1. ***The tetrahedron T_0.*** Using the coordinates above, the four flags $f_i = [p_i, l_i]$, $1 \le i \le 4$, defining T_0 can be represented in the preferred chart, as in Figure 11. Setting

$$\kappa = 2\gamma + 1 = \sqrt{5+4\sqrt{5}},$$

the coordinates of the flags are

$$f_1 = \left[[1, 0, 0], \left[0, \frac{4}{5+\kappa}, -1\right]\right] \approx \big[[1, 0, 0], [0, 0.458, -1]\big],$$
$$f_2 = \left[[0, 1, 0], \left[\frac{3-\kappa}{2}, 0, -1\right]\right] \approx \big[[0, 1, 0], [-0.367, 0, -1]\big],$$

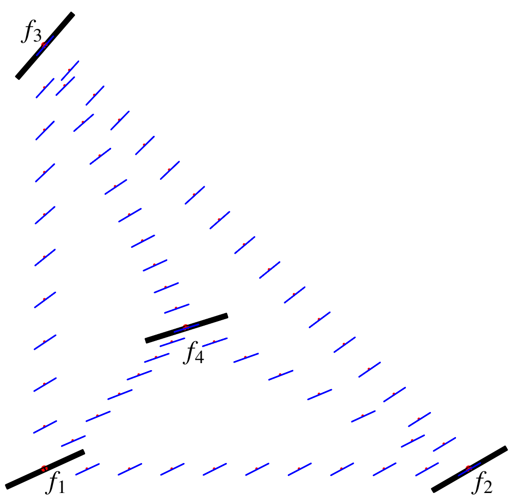


Figure 11. The four flags of tetrahedron T_0 and segments joining them projected in the preferred chart. Here $\theta_4 < \theta_1 < \theta_2 < \theta_3$.

$$f_3 = \left[[0,0,1],\ \left[\frac{-5+\kappa}{-1+\kappa}, -1, 0\right]\right] \approx \left[[0,0,1],\ [-0.463, -1, 0]\right],$$

$$f_4 = \left[[1,1,1],\ \left[\frac{2}{3+\kappa}, \frac{1+\kappa}{3+\kappa}, -1\right]\right] \approx \left[[1,1,1],\ [0.297, 0.703, -1]\right].$$

The angles at each flag in the preferred chart are computed using formula (2.2.1):

$$\tan\theta_1 = \frac{4}{5+\kappa} \qquad \Longrightarrow \quad \theta_1 \approx 0.43,$$

$$\tan\theta_2 = -\frac{-3+\kappa}{-5+\kappa} \qquad \Longrightarrow \quad \theta_2 \approx 0.53,$$

$$\tan\theta_3 = -2\,\frac{-3+\kappa}{-5+\kappa} \qquad \Longrightarrow \quad \theta_3 \approx 0.86,$$

$$\tan\theta_4 = \frac{-1+\kappa}{5+\kappa} \qquad \Longrightarrow \quad \theta_4 \approx 0.30.$$

We choose the segments between the flags so that all of them are finite and contained in the preferred chart.

Proposition 6.2. *The four flags defining T_0 and the 1-skeleton E_{ij} (defined by the finite segments joining the flags i and j in the preferred chart) can be extended to a simplex with faces F^0_{314}, F^0_{342}, F^0_{412}, F^0_{312}.*

Proof. We first define an embedded 2-skeleton. In the last paragraph of the proof below we fill it up to a 3-simplex.

We need to construct the four faces of the tetrahedron and verify that their intersections are precisely their common edges. They are (writing $F^0_{ijk} = F_{ijk}$, etc., to simplify the notation):

$$F_{314}, \quad F_{342}, \quad F_{412}, \quad F_{312}.$$

Clearly, the first three faces only intersect in their common edges. The only verification to be done is on the intersection of these faces with F_{312}. We need to prove that

$$F_{412} \cap F_{312} = E_{12}, \quad F_{314} \cap F_{312} = E_{31}, \quad F_{342} \cap F_{312} = E_{32}.$$

The argument uses Lemma 3.1 in a simple way. We choose the preferred chart. Observe first, because $\theta_4 < \theta_1 < \theta_2$, that the segment E_{12} has all flags with angles greater than the flags at the edges E_{14}. By the lemma, $F_{314} \cap F_{312} = E_{31}$.

Observe that the line from p_3 to p_4 intersects the edge E_{12} at a point, say p, whose flag has angle $\theta > \theta_4$. Moreover, a simple drawing (see Figure 12) or computation shows that the intersection point of l_4 with the line l_2 is between p_2 and the intersection point between l_1 and l_2. This is sufficient to prove that the angle of a flag along the segment E_{24} is smaller than the corresponding flag (along

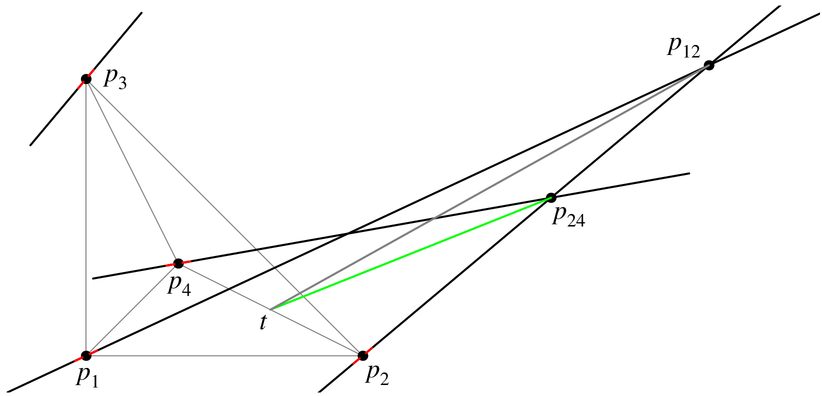


Figure 12. Comparison of two flags over a point $t \in p_2 p_4$. At the point t the flag of the face F_{312} has greater angle than the one of the face F_{342}.

the segment whose projection contains p_3 and the projection of the flag in E_{24}) passing at the edge E_{12}. This implies, again by the lemma, that $F_{342} \cap F_{312} = E_{32}$.

To analyze $F_{412} \cap F_{312}$, observe that if x belongs to the triangle $p_1 p_2 p_4$ and is to the left of the line $p_3 p_4$, then, because $\theta_1 < \theta_2$, the angle at x along the line from p_3 is greater than the angle along the line from p_4. For a point to the right of the line $p_3 p_4$, we conclude with an argument analogous to the previous paragraph. This implies again that $F_{412} \cap F_{312} = E_{12}$.

The last part of the proof consists in completing the 2-skeleton to a 3-simplex. We do it explicitly in the following way. For each point p in the preferred coordinate chart in the triangle $p_1 p_2 p_3$, there are two angles $\theta \geq \theta'$, the first one corresponding to the face F_{312} and the other to one of the other three faces. We define a segment of flags for each of these points by varying the angle from the first angle (at face F_{312}) to the one on the other face. That is, we consider all flags passing through p with angles ϕ, with $\theta \geq \phi \geq \theta'$, where we have strict inequality outside the edges of the triangle $p_1 p_2 p_3$. □

6.3. ***The tetrahedra T_1 and T_2.*** In Figure 13 we show the three tetrahedra glued according to $g_1 : (243)^1 \to (234)^0$ and $g_2 : (142)^2 \to (241)^1$. The three tetrahedra are T_0, $g_1(T_1)$, and $g_1 g_2(T_2)$.

The points in the figure are projections of the flags

$$f_5 = [p_5, l_5] = g_1[p_1, l_1] \quad \text{and} \quad f_6 = [p_6, l_6] = g_1 g_2[p_3, l_3].$$

Due to the face pairings, the faces of T_1 and T_2 are in part determined by the choice of the faces of T_0, namely, for T_1, F^1_{432}, and F^1_{134} and for T_2, F^2_{413}, and F^2_{312} are determined. The remaining two pairs of faces can be chosen arbitrarily.

Observe that F^1_{432} and F^1_{134} are represented, in the glued configuration, by F_{342} and F_{543}, respectively. Also, F^2_{413} and F^2_{312} are represented by F_{326} and F_{625}.

The definition of each filled tetrahedron follows the same line as for T_0, namely, in the preferred chart we fix a point which is a projection of two flags in two different faces. These two flags determine two angles. We then obtain a segment of flags defined by varying the angles between these two angles. The 3-simplex is the union of those segments.

We have to verify compatibilities in the definition, namely, that the side pairings map the edges between themselves and that the tetrahedra defined by the faces above do not intersect except in their common faces. We state the compatibility of the edges as a lemma whose proof is a straightforward computation.

Lemma 6.4. *The finite edges between the flags are compatible with the side pairings. That is, the face pairings map finite edges onto finite edges.*

Proof. The compatibility of the vertices is already verified by the definition of the tetrahedra in the computations in [Falbel et al. 2015]. We need to verify that in Figure 13, finite segments are mapped to finite segments by the side pairings. The side pairings are given by the four maps s_3, s_4, s_5, s_6; see the Appendix. Clearly, each transformation $g_a : (ijk)^m \to (i'j'k')^n$ is completely determined by the vertices of the two faces. The polyhedron side pairings s_i are determined by the g_a.

In our case, a tedious verification shows that all maps s_3, s_4, s_5, s_6 are such that they always map finite segments that are edges of a face of one tetrahedron of the

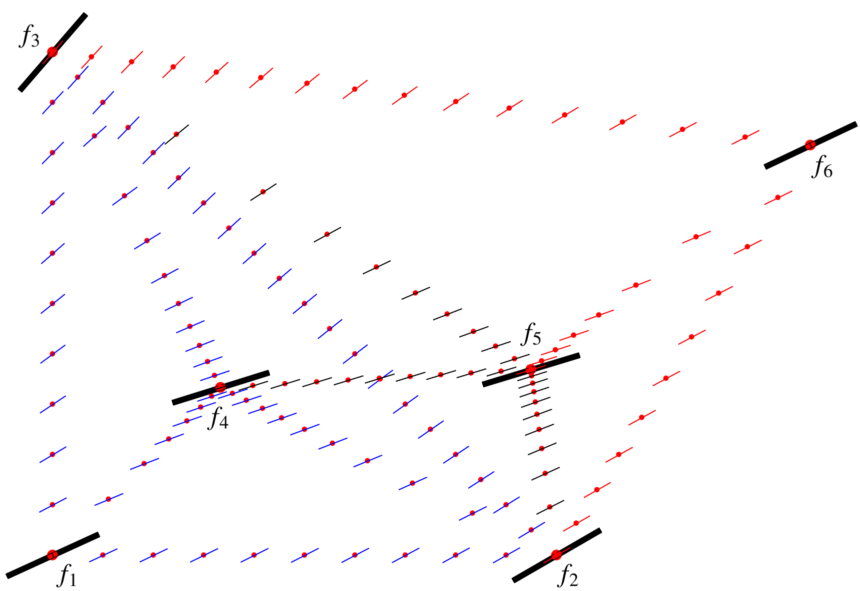


Figure 13. Gluing the 3 tetrahedra projected in the preferred chart.

polyhedron into finite segments that are edges of the corresponding face in another tetrahedron of the polyhedron. □

The next verification proves that T_0, $g_1(T_1)$, and $g_1g_2(T_2)$ are well defined and form a polyhedron in the flag space. That is, as for T_0, their faces intersect only at common edges. Finally, we prove that the three tetrahedra intersect only at common faces. The proof is a sequence of tedious arguments (Proposition A7), as in the proof that T_0 was well defined, but one can be convinced by carefully looking at Figure 13.

Proposition 6.5. *The gluing of the three tetrahedra T_0, $g_1(T_1)$, and $g_1g_2(T_2)$ forms a polyhedron in the flag space.*

6.6. ***The structure around the edges.*** There are three edges in the quotient manifold, represented by the edges E_{23}, E_{24}, and E_{34} in the first tetrahedron T_0. As far as the topological gluing is concerned, the number of tetrahedra around each edge are 8, 4, and 6 respectively (we show the schematic diagram of the gluing for each edge in Figure 14). To prove that we have a genuine flag structure on the quotient manifold, we should prove that the gluing of the tetrahedra around each of the three edges has no branching. That is, that the gluing around each edge gives a neighborhood of the edge.

We state the result in the following proposition. Its proof is a tedious verification and is given in one particular case; see Proposition A9. We use a slicing of the tetrahedra to describe the behavior of the structures around the edges. Heuristically, one can understand the neighborhood of an edge by following the vertices of the tetrahedra that one adjoins to the edge. Turning around the edge corresponds to turning the angle of the projected line of the flag in the vertex in such a way that increasing the angle makes the tetrahedron go up and decreasing the angle makes the tetrahedron go down.

In Figure 15, we show that the 4 tetrahedra around the edge E_{24} (the complete proof in this case is given in the Appendix). One can observe that the last point adjoined has the projected line of angle lower than the others. The tetrahedra adjoined will be below the original two. In Figure 16, we show 5 of the 6 tetrahedra around the edge E_{34}. Here we have to add three more points to the original 3 tetrahedra. Observe that the first two have lines of decreasing angle, but the last point increases the angle in order to complete the turn. In Figure 17, we show the vertices of the 8 tetrahedra around the edge E_{23}.

A detailed proof of the following proposition is given in Proposition A9 for the case E_{34}. For the others two gluings around E_{23} and E_{34}, the proof follows the same lines.

Proposition 6.7. *Along each of the three edges E_{23}, E_{24}, and E_{34} the gluing of the tetrahedra defines a neighborhood.*

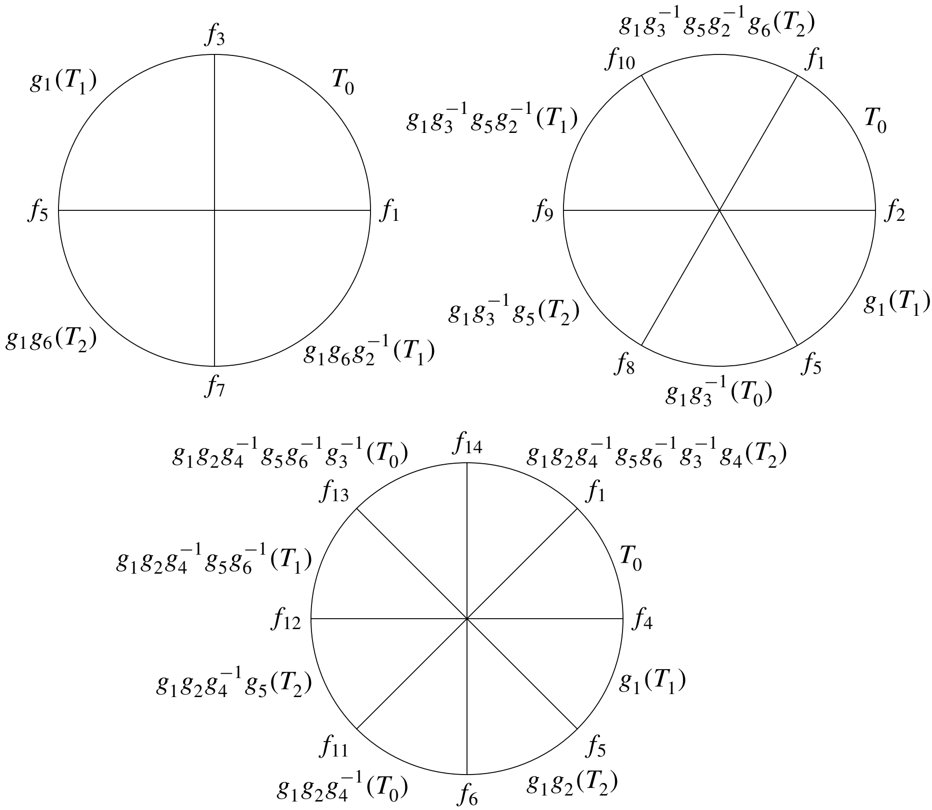


Figure 14. Top left: schematics of a neighborhood around the edge $E_{24} = [f_2, f_4]$, where the segments stand for the faces with the common edge E_{24} denoted by the origin and the arcs and the regions between two segments stand for the neighborhoods contained in one tetrahedron. The remaining diagrams represent a neighborhood around the edge $E_{23} = [f_2, f_3]$ (top right) and one around the edge $E_{34} = [f_3, f_4]$ (bottom).

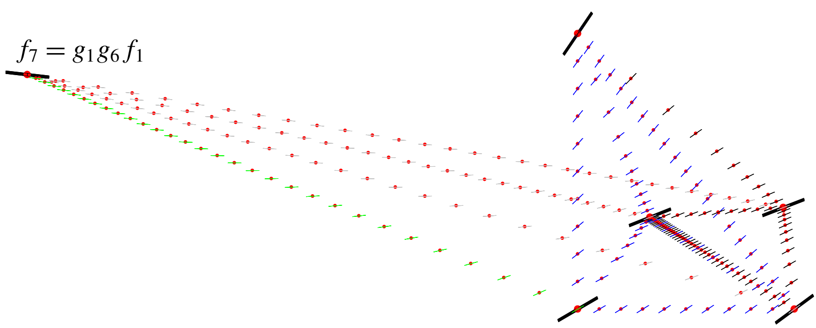


Figure 15. Tetrahedra around the edge E_{24}.

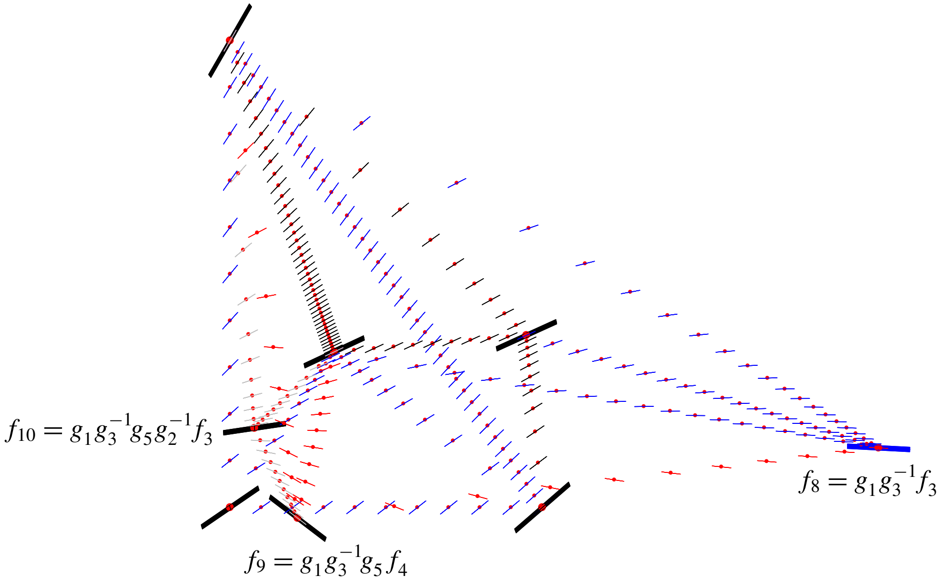


Figure 16. Tetrahedra around the edge E_{34}.

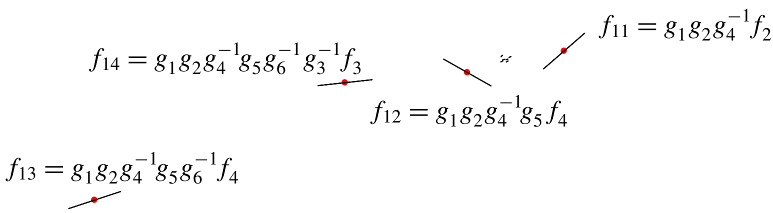


Figure 17. Vertices of tetrahedra around the edge E_{23}. The group of 6 points in the center can be zoomed to coincide with Figure 13.

As a consequence of the propositions, we obtain our conclusion:

Theorem 6.8. *The manifold m009 has a flag structure whose holonomy map is boundary unipotent.*

Appendix

A1. ***Generators and side pairings.*** To help the reader check computations we list explicitly the side pairings we use. Note that we simplify notation denoting matrices by the same letters as the maps. First we let

$$u_1 = 1 - \frac{1}{u_{12}} = \frac{2}{\gamma+3}, \qquad w_1 = 1 - \frac{1}{w_{12}} = \frac{2}{2-\gamma}, \qquad v_1 = 1 - \frac{1}{v_{12}} = -2-\gamma,$$

$$u_2 = 1 - u_{21} = 1-\gamma, \qquad w_2 = 1 - w_{21} = 2+\gamma, \qquad v_2 = 1 - v_{21} = \frac{\gamma-1}{\gamma-2},$$

$$u_3 = u_{34} = \frac{\gamma-2}{\gamma}, \qquad w_3 = w_{34} = \frac{\gamma+3}{\gamma+1}, \qquad v_3 = v_{34} = \frac{1}{\gamma+3},$$
$$u_4 = \frac{1}{1-u_{43}} = \frac{1}{2+\gamma}, \qquad w_4 = \frac{1}{1-w_{43}} = \frac{1}{1-\gamma}, \qquad v_4 = \frac{1}{1-v_{43}} = \frac{\gamma-2}{\gamma-1}.$$

The generators are given by

$$F^0_{234} = g_1(F^1_{243}) \quad g_1 = \begin{bmatrix} -\lambda_3 & 0 & \lambda_3 \\ -\lambda_1-\lambda_3 & \lambda_1 & \lambda_3 \\ -\lambda_3+\lambda_2 & 0 & \lambda_3 \end{bmatrix}, \quad \begin{aligned} \lambda_2 &= \lambda_1(v_3-1)(1-u_4), \\ \lambda_3 &= \frac{\lambda_1}{(v_4-1)(1-u_3)}, \end{aligned}$$

$$F^1_{142} = g_2(F^2_{241}), \quad g_2 = \begin{bmatrix} 0 & \delta_3 & \delta_2-\delta_3 \\ \delta_1 & 0 & \delta_2-\delta_1 \\ 0 & 0 & \delta_2 \end{bmatrix}, \quad \begin{aligned} \delta_2 &= \frac{\delta_1 v_1(w_2-1)}{w_2(v_1-1)}, \\ \delta_3 &= \frac{\delta_1(1-v_4)(1-w_4)}{v_4 w_4}, \end{aligned}$$

$$F^0_{142} = g_3(F^1_{314}), \quad g_3 = \begin{bmatrix} \alpha_2 & -\alpha_2-\alpha_1 & \alpha_1 \\ \alpha_2 & \alpha_3-\alpha_2 & 0 \\ \alpha_2 & -\alpha_2 & 0 \end{bmatrix}, \quad \begin{aligned} \alpha_2 &= \frac{\alpha_1 u_2 v_4}{1-u_2}, \\ \alpha_3 &= \frac{\alpha_1 u_4(v_1-1)}{u_4-1}, \end{aligned}$$

$$F^0_{134} = g_4(F^2_{143}), \quad g_4 = \begin{bmatrix} \beta_1 & -\beta_1-\beta_3 & \beta_3 \\ 0 & -\beta_3 & \beta_3 \\ 0 & \beta_2-\beta_3 & \beta_3 \end{bmatrix}, \quad \begin{aligned} \beta_2 &= \frac{\beta_1 u_4(1-w_3)}{w_3}, \\ \beta_3 &= \frac{\beta_1 u_3}{w_4(1-u_3)}, \end{aligned}$$

$$F^0_{123} = g_5(F^2_{213}), \quad g_5 = \begin{bmatrix} 0 & \epsilon_1 & 0 \\ \epsilon_2 & 0 & 0 \\ 0 & 0 & \epsilon_3 \end{bmatrix}, \quad \begin{aligned} \epsilon_2 &= \epsilon_1 u_3 w_3, \\ \epsilon_3 &= \frac{\epsilon_1 u_2}{w_1}, \end{aligned}$$

$$F^1_{123} = g_6(F^2_{342}), \quad g_6 = \begin{bmatrix} -\zeta_1 & 0 & \zeta_1 \\ \zeta_2 & 0 & 0 \\ -\zeta_3 & \zeta_3 & 0 \end{bmatrix}, \quad \begin{aligned} \zeta_2 &= \zeta_1 v_3(w_2-1), \\ \zeta_3 &= \zeta_1 v_2(1-w_4), \end{aligned}$$

Thinking of the generators as hyperbolic transformations, we can obtain a presentation of the fundamental group of m009. Indeed, the side pairings of the (hyperbolic) polyhedron formed by gluing the tetrahedra (as in Figure 13) according to $g_1 : (243)^1 \to (234)^0$ and $g_2 : (142)^2 \to (241)^1$ are

$$s_3 = g_3 g_1^{-1}, \quad s_4 = g_4 g_2^{-1} g_1^{-1}, \quad s_5 = g_5 g_2^{-1} g_1^{-1}, \quad s_6 = g_1 g_6 g_2^{-1} g_1^{-1}.$$

The three edge cycles give the relations

$$s_6 s_3^{-1}, \quad s_4^{-1} s_5 s_6^{-1} s_3^{-1} s_4 s_5^{-1}, \quad s_3^{-1} s_5 s_6 s_4^{-1},$$

and the presentation of the fundamental group $\Gamma = \pi_1(\mathrm{m009})$ of the manifold m009 can be simplified to be

$$\Gamma = \langle s_3, s_5 \mid [s_3^{-1}, s_5^{-1}] s_3^{-2} [s_3^{-1}, s_5] \rangle.$$

The manifold m009 is fibered over the circle. From the presentation, we know that its fundamental group Γ has abelianization

$$\Gamma/[\Gamma, \Gamma] = \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}.$$

Indeed, from the presentation we observe that $s_3^2 \in [\Gamma, \Gamma]$. We conclude that the image of s_5 in $\Gamma/[\Gamma, \Gamma]$ is nontrivial and generates an infinite cyclic group.

One can also check (using SnapPea for instance and comparing fundamental groups) that m009 is the same as the manifold b++RRL which is the punctured torus bundle defined by the pseudo-Anosov

$$\begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}.$$

However, a computation with the matrices of s_3 and s_5 (we warn the reader that we also write s_i for the image of s_i under the holonomy representation, by abuse of notation) shows that the holonomy group is contained in a triangle group of type $(3, 3, 5)$. Indeed, s_3 is of order 5, $s_3 s_5$ and $s_3^2 s_5$ are of order 3. On the other hand s_5 is unipotent.

A2. ***Slicing tetrahedra and proof of Proposition 6.7.*** In this section we describe a method to slice a tetrahedron according to directions of the flags contained on it. This is the main technical tool that we used to show the compatibility of the structure around an edge (Proposition 6.7) and can be implemented on a computer to check other examples.

In order to analyze gluings of tetrahedra properly, it will be convenient, in a fixed preferred chart, to deal with "constant angle" flags in a given face:

Definition A3. The *constant angle θ path $C_{123}(\theta)$* in a face F_{123} is the set of flags $f = (p, l) \in F_{123}$ such that, in the preferred chart, all lines l make an angle θ — see (2.2.1). That is,

$$C_{123}(\theta) = \{ f \in F_{123} \mid f = (p, l) = (p, [a, b, c]) \text{ with } \tan\theta = (b-a)/(a-c) \}.$$

The following lemma ensures that the set $C_{123}(\theta)$ is a path that, in the extreme cases, has only one flag.

Lemma A4. *Let F_{123} be a face with vertices $f_i = (p_i, l_i)$, $i = 1, 2, 3$ such that, in the preferred chart, each line l_i makes an angle θ_i with $\theta_i \neq \theta_j$, $i \neq j$. Then, to any angle θ such that*

$$\min\{\theta_1, \theta_2, \theta_3\} = \theta_m \leq \theta \leq \theta_M = \max\{\theta_1, \theta_2, \theta_3\},$$

the set $C_{123}(\theta)$ is a curve in the face and so we have the disjoint union

$$F_{123} = \bigcup_{\theta_m \le \theta \le \theta_M} C_{123}(\theta).$$

Proof. We use the preferred chart, and, without loss of generality, we consider the face F_{312} as in Figure 7. So, $p_1 = (0,0)$, $p_2 = (1,0)$, $p_3 = (0,1)$. Let $l_1 = (a_1 u,\, b_1 u)$, $l_2 = (a_2 v + 1,\, b_2 v)$, and $l_3 = (a_3 w,\, b_3 w + 1)$ for $u, v, w \in \mathbb{R}$. We first parametrize a flag $(p, l) \in F_{312}$ in terms of the coordinates in the preferred chart. The parametrization is in terms of the coordinates of the point p, which, by abuse of notation, we consider already in the preferred coordinate chart. For $p = (y, z) \in \pi_1(F_{312})$,

$$\begin{aligned}\overrightarrow{p_1 p_2} &= (r p_1 + (1-r) p_2) = ((1-r),\, 0)\\ \overrightarrow{p p_3} &= (s p + (1-s) p_3) = (s y,\, s z + (1-s)).\end{aligned}$$

Let $p_t = \overrightarrow{p_1 p_2} \cap \overrightarrow{p p_3}$, so $p_t = (y/(1-z),\, 0)$. Then,

$$\begin{aligned}p_{12} = l_1 \cap l_2 &= \left(\frac{a_1 b_2}{a_1 b_2 - a_2 b_1},\, \frac{b_1 b_2}{a_1 b_2 - a_2 b_1}\right),\\ \overrightarrow{p_t p_{12}} &= (r p_t + (1-r) p_{12})\\ &= \left(\frac{r y}{1-z} + \frac{(1-r) a_1 b_2}{a_1 b_2 - a_2 b_1},\, \frac{(1-r) b_1 b_2}{a_1 b_2 - a_2 b_1}\right).\end{aligned}$$

Let $q = \overrightarrow{p_t p_{12}} \cap l_3$. Then,

$$\begin{aligned}q = (q_y, q_z) = \Bigg(&\frac{a_3\big(y(b_1 b_2 + a_2 b_1 - a_1 b_2) + a_1 b_2 (1-z)\big)}{b_3 y (a_1 b_2 - a_2 b_1) + b_2 (1-z)(a_3 b_1 - a_1 b_3)},\\ &\frac{b_1 b_2 \big(y b_3 + a_3 (1-z)\big)}{b_3 y (a_1 b_2 - a_2 b_1) + b_2 (1-z)(a_3 b_1 - a_1 b_3)}\Bigg).\end{aligned}$$

Since $l = \overrightarrow{pq}$,

$$\begin{aligned}\tan\theta &= \frac{q_z - z}{q_y - y}\\ &= \frac{y b_3 \big(b_1 b_2 - z(a_1 b_2 - a_2 b_1)\big) + a_3 b_1 b_2 (1-z)^2 + z b_2 a_1 b_3 (1-z)}{y\big(a_3 b_1 b_2 + (a_3 + y b_3)(a_2 b_1 - a_1 b_2)\big) + b_2 (1-z)\big(a_3 a_1 + y(a_1 b_3 - a_3 b_1)\big)}.\end{aligned}$$

Fixing θ, we obtain the equation of a conic in the preferred chart which always contains $p_3 = (0,1)$. The part of this conic inside $\pi_1(F_{312})$ is the projection $\pi_1(C_{312}(\theta))$. As for all θ such that $\theta_m \le \theta \le \theta_M$, we have $C_{312}(\theta) \neq \varnothing$ and conclude that F_{312} is the disjoint union of paths, as claimed. Notice that F_{312} has only one flag with $\theta = \theta_m$ and another one with $\theta = \theta_M$. $\square$

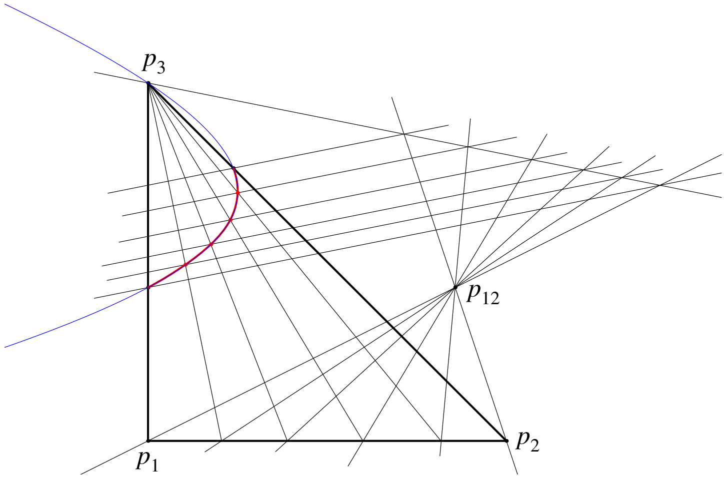


Figure 18. Path $C_{312}(\theta)$ with constant θ in the face F_{312}.

For example, in Figure 18, the red path is $\pi_1(C_{312}(\theta))$ for $a_1 = 2$, $b_1 = 1$, $a_2 = -1$, $b_2 = 3$, $a_3 = 5$, $b_3 = -1$, and $\tan\theta = 1/5$. The blue path is a branch of a hyperbola determined by the equation of Lemma A4. Notice that both lines coincide inside $\pi_1(F_{312})$.

Definition A5. A slice $S_{1234}(\theta)$ with constant θ in a tetrahedron T_{1234} is the set of flags $f = (p, l) \in T_{1234}$ such that, in the preferred chart, all lines l make an angle θ; that is,

$$S_{1234}(\theta) = \{f \in T_{1234} \mid f = (p, l) = (p, [a, b, c]) \text{ with } \tan\theta = (b - a)/(a - c)\}.$$

Proposition A6. *Suppose that T_{1234} is a finite tetrahedron with vertices $f_i = (p_i, l_i)$, $i = 1, 2, 3, 4$, such that, in the preferred chart, each line l_i makes an angle θ_i with $\theta_i \neq \theta_j$, $i \neq j$. Then, for any angle θ such that*

$$\min\{\theta_1, \theta_2, \theta_3, \theta_4\} = \theta_m \leq \theta \leq \theta_M = \max\{\theta_1, \theta_2, \theta_3, \theta_4\},$$

the set $S_{1234}(\theta)$ has three or four vertices (each one in a distinct edge). Furthermore, the tetrahedron is a disjoint union of slices, that is,

$$T_{1234} = \bigcup_{\theta_m \leq \theta \leq \theta_M} S_{1234}(\theta).$$

Proof. Without loss of generality, we can consider that the tetrahedron T_{1234} is T_0, as defined in 6.1, represented in the preferred chart in Figure 11 and analyzed in Proposition 6.2. Then, θ_i is such that $\theta_m = \theta_4 < \theta_1 < \theta_2 < \theta_3 = \theta_M$, and the tetrahedron's faces are the good faces F_{314}, F_{342}, F_{412}, and F_{312}. The slices of T_0 will be denoted $S_0(\theta)$.

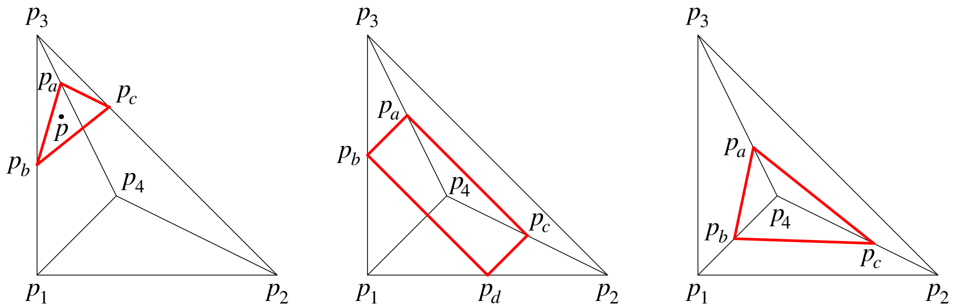


Figure 19. Slices in T_0 for $\theta_2 \le \theta \le \theta_3$, $\theta_1 \le \theta \le \theta_2$, and $\theta_4 \le \theta \le \theta_1$.

In the first case, let θ satisfy $\theta_2 \le \theta \le \theta_3 = \theta_M$. Then, by Lemma A4, there are three curves $C_{314}(\theta)$, $C_{342}(\theta)$, and $C_{312}(\theta)$ in the respective faces with the common vertex f_3. Considering the common edges between faces, let $f_a = C_{314} \cap C_{342} \in E_{34}$, $f_b = C_{314} \cap C_{312} \in E_{13}$, and $f_c = C_{342} \cap C_{312} \in E_{23}$. We obtain the slices as in Figure 19 (for simplicity the paths are depicted as straight lines).

By the monotonicity lemma (Lemma 3.1), each one of these curves separates the respective faces into two parts: one near the vertex f_3 where $\theta' > \theta$ and the other near the two other vertices of the face where $\theta'' < \theta$. Consider a point p in $\mathbb{P}(\mathbb{R}^3)$ as in Figure 19. There exist two flags in the faces of T_0 which project onto p, namely, $(p, d) \in F_{312}$ and $(p, d') \in F_{314}$ with $\theta_d > \theta$ and $\theta_{d'} < \theta$. Then, by the definition of T_0 as a 3-simplex (see Proposition 6.2), for all p inside the area delimited by $\pi_1(C_{312}(\theta))$, $\pi_1(C_{314}(\theta))$, and $\pi_1(C_{342}(\theta))$, there exists a flag $f' = (p, l') \in T_0$ such that $\theta' = \theta$, so $f' \in S_0(\theta)$. Clearly, $C_{312}(\theta) \subset S_0(\theta)$, $C_{314}(\theta) \subset S_0(\theta)$, and $C_{342}(\theta) \subset S_0(\theta)$. Furthermore $S_0(\theta)$ has three vertices: $f_a \in E_{34}$, $f_b \in E_{13}$, and $f_c \in E_{24}$, as claimed.

In the other cases, as in Figure 19, the argument is the same, and this concludes the first part of the proof.

The second part follows clearly, as for all θ such that $\theta_m \le \theta \le \theta_M$, we have $S_0(\theta) \ne \varnothing$. □

Figure 20 shows the exact slices in T_0 for $\tan\theta = 0.4$ and $\tan\theta = 0.5$.

Proposition A7. *The gluing of the three tetrahedra T_0, $g_1(T_1)$, and $g_1g_2(T_2)$ forms a polyhedron in the flag space.*

Proof. Observe that $\theta_5 \le \theta_i \le \theta_3$ for all $i = 1, \dots, 6$. We can construct the slices in each tetrahedron for θ from θ_5 to θ_3 and see that they only intersect in the common path contained in the common face between two tetrahedra. In Figure 21, we show the slices for $\tan\theta = 0.4$ ($\theta_4 < \theta < \theta_1$), $\tan\theta = 0.5$ ($\theta_6 < \theta < \theta_2$), and $\tan\theta = 0.8$ ($\theta_2 < \theta < \theta_3$). The other two cases $\theta_5 < \theta < \theta_4$ and $\theta_1 < \theta < \theta_6$ are similar. □

Now consider the tetrahedron 2457 with vertices f_2, f_4, $f_5 = g_1[p_1, l_1]$, and $f_7 = g_1g_6[p_1, l_1]$. It is one of the four tetrahedra around the edge E_{24} in Figure 15.

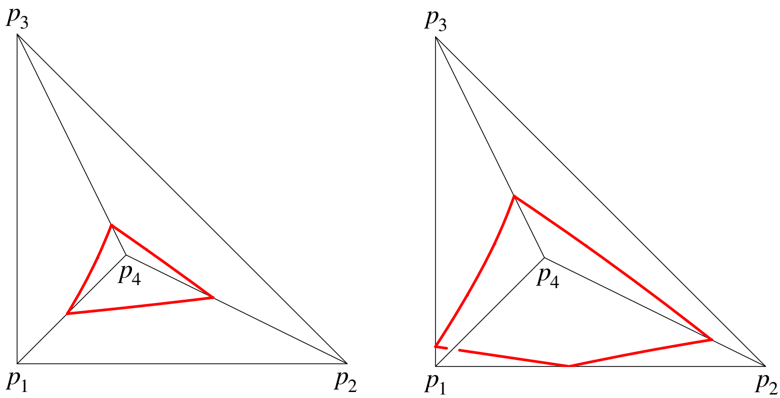


Figure 20. Left: slice in T_0 for $\tan\theta = 0.4$, exemplifying the case $\theta_4 < \theta < \theta_1 < \theta_2 < \theta_3$. Right: slice for $\tan\theta = 0.5$, corresponding to $\theta_4 < \theta_1 < \theta < \theta_2 < \theta_3$.

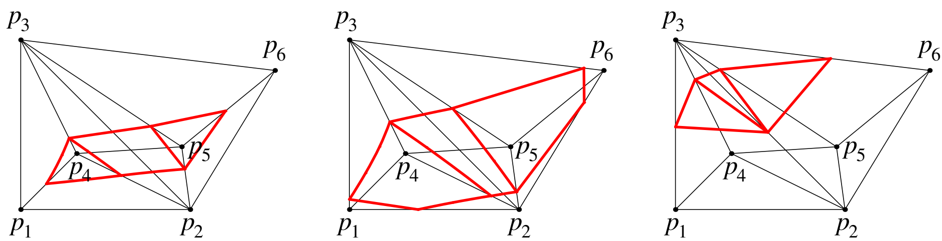


Figure 21. Slices in the gluing of T_0, $g_1(T_1)$ and $g_1g_2(T_2)$ for $\tan\theta = 0.4$, that is, $\theta_4 < \theta < \theta_1$, $\tan\theta = 0.5$, that is, $\theta_6 < \theta < \theta_2$, and $\tan\theta = 0.8$, that is, $\theta_2 < \theta < \theta_3$.

The following lemma describes the position of the projection of certain slices. It will be important in order to construct slices around an edge in the next proposition.

Lemma A8. *Consider the tetrahedron* 2457, *in the preferred chart*, *and the projection of the path* $C_{725}(\theta_p)$, *that is* $\pi_1(C_{725}(\theta_p))$. *Then*, *for all* $f_p = (p, l_p) \in E_{24}$, *we have* $q = \pi_1(E_{24}) \cap \pi_1(C_{725}(\theta_p)) \in [p_2, p]$.

In other words the projection of the path $C_{725}(\theta_p)$ intersects the projection of the edge E_{24}, that is $\pi_1(E_{24})$, between $p_2 = \pi_1(f_2)$ and $p = \pi_1(f_p)$, where $f_p = (p, l_p)$ is the flag in the edge E_{24} such that the line l_p makes an angle θ_p.

Proof. The projection of the tetrahedron is represented in Figure 22 with a detail in the lower diagram. Let $f_p = (p, l_p) \in E_{24}$, and let θ_p be the angle of l_p in the preferred chart. By definition, $p \in l_p$ and $p_{24} = l_2 \cap l_4 \in l_p$. As $\theta_7 < \theta_5 < \theta_4 < \theta_p < \theta_2$, the path $C_{725}(\theta_p)$ has one end in E_{25} and another in E_{27}, so clearly $\pi_1(C_{725}(\theta_p))$ intersects $\pi_1(E_{24})$ or $\pi_1(E_{45})$.

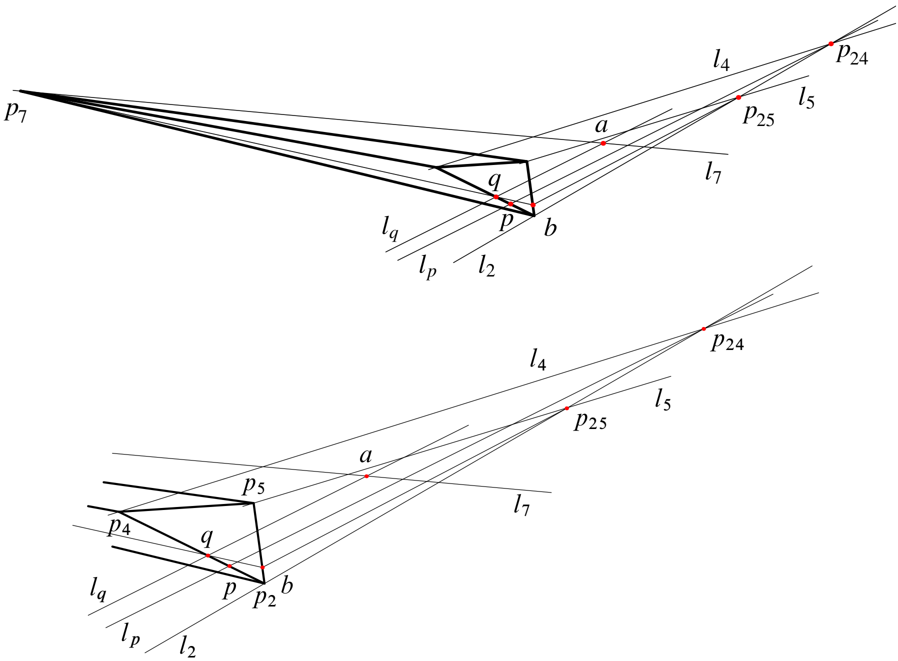


Figure 22. Top: tetrahedron 2457. Relative position of p and q in the edge E_{24}. Bottom: detail of the same tetrahedron, showing the relative position of p and q in the edge E_{24}.

Suppose first that it intersects E_{24}. Let $f_q = (q, l_q) \in C_{725}(\theta_p)$ such that $q = \pi_1(E_{24}) \cap \pi_1(C_{725}(\theta_p))$. Suppose, towards a contradiction, that $q \in [p, p_4]$. By definition, l_p and l_q are parallel. Let $a = l_q \cap l_7$ and let b be the intersection between $\pi_1(E_{25})$ and the straight line through p_7 and q. Notice that b and p_{25} are on opposite sides with respect to l_7 but are on the same side with respect to l_q. As $a = l_q \cap l_7$, the points a, b, and p_{25} don't lie in the same straight line, which contradicts the definition of the face F_{725}. Indeed, if $f_q \in F_{725}$, then a, b, and p_{25} should lie in the same straight line. We conclude that $q \notin [p, p_4]$.

By the same construction and arguments we obtain that $\pi_1(C_{725}(\theta_p))$ doesn't intersect $\pi_1(E_{45})$. Thus, $\pi_1(C_{725}(\theta_p))$ intersects $\pi_1(E_{24})$ between p and p_2. □

Proposition A9. *Along the edge E_{24}, the gluing of the tetrahedra defines a neighborhood.*

Proof. Consider the gluing of four tetrahedra along E_{24}, represented in Figure 15. Consider also the tetrahedron $g_1 g_6(T_2)$ with vertices f_2, f_4, f_5, and f_7 such that $\theta_7 < \theta_5 < \theta_4 < \theta_1 < \theta_2$. Observe first that, by Lemma A8, for all θ such that $\theta_4 < \theta < \theta_2$ (that is, $\tan\theta_4 \approx 0.313 < \tan\theta < \tan\theta_2 \approx 0.580$), the slice $S_{2457}(\theta)$ projects

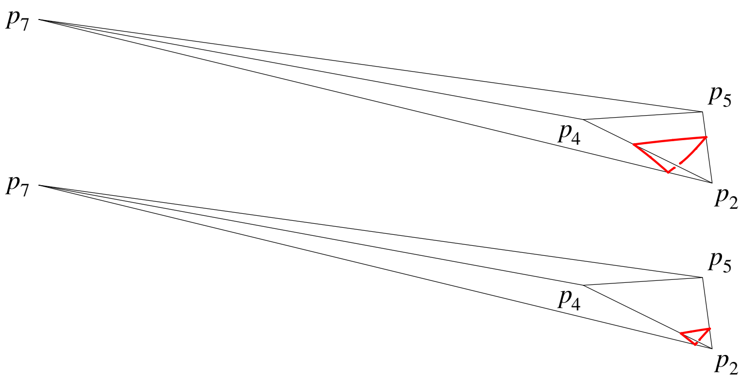


Figure 23. Slice $S_{2457}(\theta)$ in $g_1g_6(T_2)$ for $\tan\theta = 0.4$ (top) and $\tan\theta = 0.5$ (bottom).

to a preferred chart in a "triangle" with vertices in the edges $[p_7, p_2]$, $[p_2, p_4]$, and $[p_2, p_5]$, as shown in Figure 23 for $\tan\theta = 0.4 < \tan\theta_1 \approx 0.458$ (top diagram) and $\tan\theta = 0.5 > \tan\theta_1$ (bottom).

As a consequence, the unique form of the slices $S_{2457}(\theta)$ (they all project to the triangles described above), the form of the slices $S_{1234}(\theta)$, $S_{1247}(\theta)$, and $S_{2345}(\theta)$ are unique too. Indeed, the slices have common paths, for example, $C_{425}(\theta) = S_{2457}(\theta) \cap S_{2345}(\theta)$, and two paths inside a tetrahedron don't intersect. Therefore, we obtain slices in each one of the four tetrahedra as in Figure 24 for $\tan\theta = 0.4 < \tan\theta_1$ and in Figure 25 for $\tan\theta = 0.5 > \tan\theta_1$. Thus, all neighborhoods of a point $p \in \pi_1(E_{24})$ are as in Figure 26 for $\theta_4 < \theta < \theta_1 < \theta_2$ or as in Figure 27 for $\theta_4 < \theta_1 < \theta < \theta_2$. Figures 23 to 27 are obtained through exact computations. $\square$

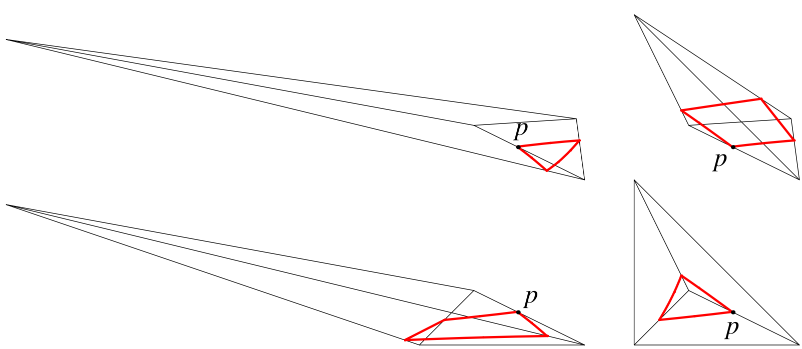


Figure 24. Slices in each tetrahedron with common edge E_{24} for $\tan\theta = 0.4$.

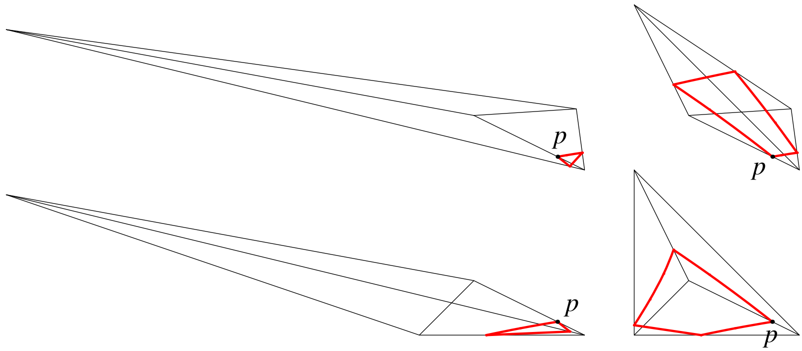


Figure 25. Slices in each tetrahedron with common edge E_{24} for $\tan\theta = 0.5$.

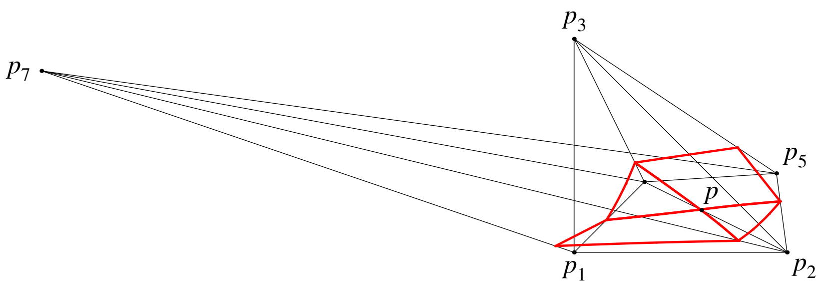


Figure 26. Neighborhood of p such that $f = (p, l) \in E_{24}$ for $\theta_7 < \theta_5 < \theta_4 < \theta < \theta_1 < \theta_2 < \theta_3$.

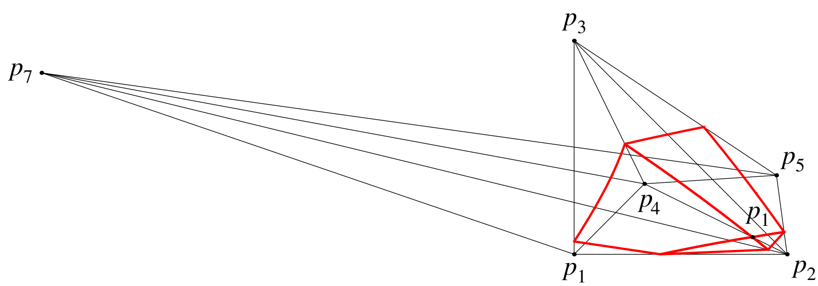


Figure 27. Neighborhood of p such that $f = (p, l) \in E_{24}$ for $\theta_7 < \theta_5 < \theta_4 < \theta_1 < \theta < \theta_2 < \theta_3$.

Acknowledgments

We thank N. Bergeron, M. Deraux, A. Guilloux, Neil Hoffman, and P.-V. Koseleff for stimulating discussions. We also thank an anonymous referee for many suggestions that improved the exposition.

This work was partially financed by an ANR project SGT (Structures géométriques triangulées). Santos Thebaldi is supported by CAPES Brazil (BEX 9599/11-8) and partially supported by UFOP. He also thanks FAPEMIG Brazil for financial support (CEX-APQ-01056-08).

References

[Barbot 2001] T. Barbot, "Flag structures on Seifert manifolds", *Geom. Topol.* **5** (2001), 227–266. MR 2002h:57035 Zbl 1032.57037

[Bergeron et al. 2014] N. Bergeron, E. Falbel, and A. Guilloux, "Tetrahedra of flags, volume and homology of SL(3)", *Geom. Topol.* **18**:4 (2014), 1911–1971. MR 3268771 Zbl 06356603

[Cooper et al. 2015] D. Cooper, D. D. Long, and S. Tillmann, "On convex projective manifolds and cusps", *Adv. Math.* **277** (2015), 181–251. MR 3336086 Zbl 06431971

[Deraux 2015] M. Deraux, "On Spherical CR Uniformization of 3-Manifolds", *Exp. Math.* **24**:3 (2015), 355–370. MR 3359222

[Falbel 2008] E. Falbel, "A spherical CR structure on the complement of the figure eight knot with discrete holonomy", *J. Differential Geom.* **79**:1 (2008), 69–110. MR 2009d:32037 Zbl 1148.57025

[Falbel et al. 2015] E. Falbel, P.-V. Koseleff, and F. Rouillier, "Representations of fundamental groups of 3-manifolds into PGL(3, $\mathbb{C}$): exact computations in low complexity", *Geom. Dedicata* **177** (2015), 229–255. MR 3370032 Zbl 06468790

[Falbel et al. 2015–] E. Falbel, S. Garoufalidis, A. Guilloux, M. Görner, P.-V. Koseleff, F. Rouillier, and C. Zickert, "CURVE", electronic reference, 2015–, Available at http://curve.unhyperbolic.org.

[Fock and Goncharov 2007] V. V. Fock and A. B. Goncharov, "Moduli spaces of convex projective structures on surfaces", *Adv. Math.* **208**:1 (2007), 249–273. MR 2008g:57015 Zbl 1111.32013

[Garoufalidis et al. 2011] S. Garoufalidis, D. Thurston, and C. Zickert, "The complex volume of SL(n, $\mathbb{C}$)-representations of 3-manifolds", preprint, 2011. arXiv 1111.2828v2

[Gelfand and MacPherson 1982] I. M. Gelfand and R. D. MacPherson, "Geometry in Grassmannians and a generalization of the dilogarithm", *Adv. in Math.* **44**:3 (1982), 279–312. MR 84b:57014

[Thurston 1981] W. Thurston, "The geometry and topology of 3-manifolds", Lecture notes, Princeton, 1981, Available at http://library.msri.org/books/gt3m.

Received April 26, 2014. Revised April 14, 2015.

ELISHA FALBEL
INSTITUT DE MATHÉMATIQUES
UNIVERSITÉ PIERRE ET MARIE CURIE
75252 PARIS CEDEX 5
FRANCE
elisha.falbel@imj-prg.fr

RAFAEL SANTOS THEBALDI
DEPARTAMENTO DE CIÊNCIAS EXATAS E APLICADAS
UNIVERSIDADE FEDERAL DE OURO PRETO
35931-008 JOÃO MONLEVADE, MG
BRAZIL
rafaelsantosthe@yahoo.com.br
rafael@decea.ufop.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.278.79

A NEW UPPER BOUND FOR THE DIRAC OPERATORS ON HYPERSURFACES

NICOLAS GINOUX, GEORGES HABIB AND SIMON RAULOT

Dedicated to Oussama Hijazi for his sixtieth birthday and to Sebastián Montiel.

We prove a new upper bound for the first eigenvalue of the Dirac operator of a compact hypersurface in any Riemannian spin manifold carrying a non-trivial twistor-spinor without zeros on the hypersurface. The upper bound is expressed as the first eigenvalue of a drifting Schrödinger operator on the hypersurface. Moreover, using a recent approach developed by O. Hijazi and S. Montiel, we completely characterize the equality case when the ambient manifold is the standard hyperbolic space.

1. Introduction

Let $M^n \overset{\iota}{\hookrightarrow} \widetilde{M}^{n+1}$ be an oriented, compact (without boundary), connected hypersurface of an $(n+1)$-dimensional Riemannian manifold $(\widetilde{M}^{n+1}, g)$ equipped with the induced Riemannian metric, also denoted by g.

It is by now a well-known approach to use the min-max characterization of eigenvalues to derive upper bounds for the spectrum of differential operators on M in terms of extrinsic geometric data. For example, if we consider the first positive eigenvalue $\lambda_1(\Delta)$ of the Laplace operator $\Delta := -\operatorname{tr}_g(\mathrm{Hess}_g)$, where Hess_g denotes the Hessian of M, a famous result of R.C. Reilly [1977] states that if $\widetilde{M}$ is the Euclidean space $\mathbb{R}^{n+1}$, then

$$\lambda_1(\Delta) \leq \frac{n}{\mathrm{Vol}(M)} \int_M H^2\, dv_g, \tag{1}$$

where H denotes the normalized mean curvature of M. The proof of this result uses, in an essential way, the Rayleigh characterization of $\lambda_1(\Delta)$ by choosing a modification of the coordinates functions as test functions. Moreover, it is a straightforward observation to see that equality occurs if and only if M is a totally umbilical round sphere. As observed in [El Soufi and Ilias 1992], this method

MSC2010: 53C27, 53C40.
Keywords: global analysis, spectral theory, Dirac operator, geometry of submanifolds.

directly applies for hypersurfaces in the unit sphere $\mathbb{S}^{n+1}$, leading to the counterpart of (1) in this situation:

$$\lambda_1(\Delta) \le \frac{n}{\mathrm{Vol}(M)} \int_M (H^2+1)\, dv_g. \tag{2}$$

If the ambient manifold $\widetilde{M}$ is the standard hyperbolic space, there is also an optimal upper bound proved by A. El Soufi and S. Ilias [1992, Theorem 1] which improves a previous result of E. Heintze [1988] and which states that

$$\lambda_1(\Delta) \le \frac{n}{\mathrm{Vol}(M)} \int_M (H^2-1)\, dv_g, \tag{3}$$

with equality if and only if M is a totally umbilical round sphere. All three estimates above follow actually from a much more general one, valid for submanifolds of any codimension [El Soufi and Ilias 1992], assuming solely that the ambient manifold is conformally equivalent to an open subset of the sphere of the same dimension: under that assumption,

$$\lambda_1(\Delta) \le \frac{n}{\mathrm{Vol}(M)} \int_M (H^2+R(\iota))\, dv_g, \tag{4}$$

[op. cit., Theorem 2], where $R(\iota)$ is the normalized trace of the ambient sectional curvature on the tangent planes; see the precise definition after (15).

Now if we assume the existence of a spin structure on $\widetilde{M}$ (which is the case for most classical ambient spaces), it induces a spin structure on the hypersurface M, and so we can define the spinor bundle ΣM over M as well as the associated Dirac operator D_M (see Section 2 and the references therein). When the ambient space $\widetilde{M}$ is the space form of constant sectional curvature $\kappa \in \{0, 1, -1\}$, C. Bär [1998] proved that

$$\lambda_1(D_M^2) \le \frac{n^2}{4\,\mathrm{Vol}(M)} \int_M (H^2+\kappa)\, dv_g \tag{5}$$

if $\kappa = 0, 1$ and

$$\lambda_1(D_M^2) \le \tfrac{1}{4} n^2 \sup_M (H^2+1) \tag{6}$$

for $\kappa = -1$. Here $\lambda_1(D_M^2)$ denotes the first nonnegative eigenvalue of the square of the Dirac operator D_M of (M, g). Those estimates are consequences of the min-max characterization of $\lambda_1(D_M^2)$ and the fact that the space forms $\mathbb{R}^{n+1}$, $\mathbb{S}^{n+1}$, and $\mathbb{H}^{n+1}$ carry, respectively, parallel, real Killing, and imaginary Killing spinors. In fact, taking the restriction of such a spinor field to the hypersurface as a test section in the Rayleigh quotient of $\lambda_1(D_M^2)$ gives the previous inequalities immediately. Note that these upper bounds hold for more general ambient manifolds since the proof only relies on the existence of one such particular field. For example, (5) with $\kappa = 0$

holds for compact oriented hypersurfaces in Calabi–Yau manifolds, hyper-Kähler and some other 7- and 8-dimensional special Riemannian manifolds. It also appears that both inequalities in (5) are sharp since round geodesic spheres in the Euclidean space $\mathbb{R}^{n+1}$ and in the round sphere $\mathbb{S}^{n+1}$ satisfy the equality case. O. Hijazi and S. Montiel [2013] proved that when $\kappa = 0$ those are the only hypersurfaces for which equality is achieved. The limiting case for hypersurfaces in the sphere seems to be out of reach at this time and could be considered as a spinorial analogue of the Yau conjecture about the first eigenvalue of the Laplace operator of minimal hypersurfaces in the unit sphere. However, there are nonminimal hypersurfaces in the sphere that satisfy the limiting case in (5); see, e.g., [Ginoux 2003a; Ginoux 2008].

Regarding the proof of (6), it is not difficult to observe that there are no hypersurfaces which satisfy the equality case. Modifying the computation of the Rayleigh quotient for $\lambda_1(D_M^2)$, this estimate can be improved [Ginoux 2003b, Theorem 1] into

$$\lambda_1(D_M^2) \leq \tfrac{1}{4}n^2 \sup_M (H^2 - 1), \tag{7}$$

where equality occurs for totally umbilical round spheres in $\mathbb{H}^{n+1}$. As we will see in Corollary 4.2, those are in fact the only hypersurfaces for which (7) is an equality.

In this paper, we prove a new upper bound for the first eigenvalue of the Dirac operator of M when the ambient manifold $\widetilde{M}$ carries a twistor-spinor; see Theorem 3.3. This bound coincides with the first eigenvalue of an elliptic differential operator of order two whose definition depends, among others, on the norm of the twistor-spinor along the hypersurface — see (15) — and which belongs to a particular class of operators: the drifting Schrödinger operators, that is, of the form drifting Laplacian plus potential; see Remarks 3.2. It is important to note that this estimate contains all the (up to date) known upper estimates *à la Reilly*; see Remarks 3.4. In a second part, we adapt the approach developed by Hijazi and Montiel [2013] to prove that, assuming the existence of imaginary Killing spinors for two opposite constants on $\widetilde{M}$, the only hypersurfaces satisfying the equality case in our previous estimate are the totally umbilical ones; see Theorem 4.1. In particular, only the geodesic hyperspheres satisfy that limiting case in the hyperbolic space; see Corollary 4.2. We also examine the setting of pseudohyperbolic spaces; see Corollary 4.7.

2. Preliminaries and notation

In this section, we briefly introduce the geometric setting and fix the notation of this paper. For more details on those preliminaries, see examples in [Lawson and Michelsohn 1989], [Friedrich 2000], or [Ginoux 2009, Chapter 1].

We consider $M^n \overset{\iota}{\hookrightarrow} \widetilde{M}^{n+1}$ an oriented n-dimensional Riemannian hypersurface with $n \geq 2$, isometrically immersed into an $(n+1)$-dimensional Riemannian spin manifold $(\widetilde{M}^{n+1}, g)$ with a fixed spin structure. We denote by ν the unit inner normal vector field induced by both orientations, that is, such that $(E_1, \dots, E_n, \nu_x)$ is an oriented basis of $T_x\widetilde{M}_{|M}$ if and only if $(E_1, \dots, E_n)$ is an oriented basis of T_xM for $x \in M$. We endow M with the spin structure induced by the one on $\widetilde{M}$ and let $\Sigma M \to M$ denotes the associated spinor bundle. Setting

$$\Sigma := \begin{cases} \Sigma M & \text{if } n \text{ is even,} \\ \Sigma M \oplus \Sigma M & \text{if } n \text{ is odd,} \end{cases}$$

the bundles Σ and the restriction $\Sigma\widetilde{M}_{|M}$ to M of the spinor bundle of $\widetilde{M}$ can be identified in such a way that:

- The natural Hermitian inner products — both of which we denote by $\langle\cdot\,,\cdot\rangle$ — coincide.
- Clifford multiplication "$\cdot$" on $\widetilde{M}$ and "$\underset{M}{\cdot}$" on M are related by

$$X \underset{\Sigma}{\cdot} := X \cdot \nu \cdot \simeq \begin{cases} X \underset{M}{\cdot} & \text{if } n \text{ is even,} \\ X \underset{M}{\cdot} \oplus -X \underset{M}{\cdot} & \text{if } n \text{ is odd,} \end{cases} \tag{8}$$

  for all $X \in TM$.
- The spin Levi–Civita connections $\widehat{\nabla}$ on $\Sigma\widetilde{M}$ and ∇ on Σ are related by the spin Gauss formula

$$\widehat{\nabla}_X\varphi = \nabla_X\varphi + \tfrac{1}{2}A(X)\cdot\nu\cdot\varphi, \tag{9}$$

  for all $X \in \Gamma(TM)$ and $\varphi \in \Gamma(\Sigma)$. Here $A := -\widehat{\nabla}\nu$ denotes the Weingarten map of the immersion.

The extrinsic Dirac operator of M is the first order elliptic differential operator of order one acting on sections of Σ locally given by

$$D := \sum_{j=1}^{n} e_j \cdot \nu \cdot \nabla_{e_j}.$$

It is a well-known fact that it defines an essentially self-adjoint operator with respect to the L^2-scalar product on Σ so that if M is compact, its spectrum is an unbounded sequence of real numbers. In this article, we adopt the convention that the spectrum $\mathrm{spec}(P)$ with multiplicities of a given elliptic self-adjoint operator P will be denoted by a sequence $(\lambda_k(P))_{k\geq 1}$, with the convention that $\lambda_1(P)$ is the smallest eigenvalue if $\mathrm{spec}(P)$ is bounded below and is the smallest *nonnegative* eigenvalue otherwise.

With respect to the previous identifications, the Dirac operator D is nothing but the Dirac operator D_M of (M, g) if n is even and $D_M \oplus -D_M$ if n is odd, so that studying the spectrum of the intrinsic Dirac operator D_M for the spin Riemannian structure induced on the hypersurface M is equivalent to study the spectrum of the extrinsic Dirac operator D on the hypersurface M. It is also relevant here to recall that the commutator of D and D^2 with functions are given by

$$D(f\varphi) = fD\varphi + \nabla f \cdot \nu \cdot \varphi \tag{10}$$

and

$$D^2(f\varphi) = fD^2\varphi - 2\nabla_{\nabla f}\varphi + (\Delta f)\varphi, \tag{11}$$

for all $f \in C^\infty(M)$ and $\varphi \in \Gamma(\Sigma)$. Here $H := \frac{1}{n}\operatorname{tr}(A)$ denotes the mean curvature function of M in $\widetilde{M}$.

Another operator of particular interest in this work is the Dirac–Witten operator $\hat{D}$ on M. It is also a first order elliptic operator acting on the restricted spinor bundle Σ and locally defined by $\hat{D} := \sum_{j=1}^n e_j \cdot \widehat{\nabla}_{e_j}$. It is related to the extrinsic Dirac operator by the following formula

$$D\varphi = -\nu \cdot \hat{D}\varphi + \tfrac{1}{2}nH\varphi \tag{12}$$

and to its square by

$$D^2\varphi = \hat{D}^2\varphi + \tfrac{1}{4}n^2H^2\varphi + \tfrac{1}{2}n\nabla H \cdot \nu \cdot \varphi, \tag{13}$$

for every $\varphi \in \Gamma(\Sigma)$.

3. Upper bounds in terms of a Laplace-type operator

In this section, we prove a new upper bound for the smallest eigenvalue of the squared Dirac operator D^2 when the ambient manifold $\widetilde{M}$ is endowed with a twistor-spinor. Recall that a *twistor-spinor* on a Riemannian spin manifold $(\widetilde{M}^{n+1}, g)$ is a section $\psi \in \Gamma(\Sigma\widetilde{M})$ satisfying

$$\widehat{\nabla}_X\psi = -\tfrac{1}{n+1}X \underset{\widetilde{M}}{\cdot} D_{\widetilde{M}}\psi \tag{14}$$

for all $X \in \Gamma(T\widetilde{M})$. Here $D_{\widetilde{M}}$ represents the Dirac operator of $\widetilde{M}$. Nonzero twistor-spinors have a discrete vanishing set and only exist for particular conformal classes; see, for example, the standard reference [Baum et al. 1991] or, for a short account, [Ginoux 2009, Appendix A]. It should also be pointed out that parallel spinors, and real and imaginary Killing spinors are twistor-spinors which are, in addition, eigensections for the Dirac operator $D_{\widetilde{M}}$ associated to the eigenvalue zero, or to real or purely imaginary eigenvalues, respectively. They exist on each simply connected complete space form of constant curvature. Assume now that such a spinor field ψ is given on $\widetilde{M}$ and also assume that it has no zero on the hypersurface M. We define the differential operator L_ψ acting on smooth functions on M by

$$L_\psi f := \Delta f - 2g(\nabla \ln|\psi|, \nabla f) + \tfrac{1}{4}n^2(H^2 + R(\iota)) f. \tag{15}$$

for $f \in C^\infty(M)$. Here

$$R(\iota) := \tfrac{1}{n(n-1)}(\tilde{S} - 2\,\widetilde{\mathrm{ric}}(\nu, \nu)),$$

$\tilde{S}$ and $\widetilde{\mathrm{ric}}$ are respectively the scalar curvature and the Ricci tensor (seen as a symmetric 2-tensor) of the manifold $\widetilde{M}$. Although this operator is not symmetric with respect to the L^2-scalar product on (M^n, g), we observe that it has the following interesting analytic properties:

Proposition 3.1. *The operator L_ψ is elliptic, and if M is closed, it is self-adjoint with respect to the L^2-scalar product on $(M^n, \bar{g})$, where $\bar{g} := |\psi|^{4/n} g$.*

Proof. Since L_ψ is of second order and its leading part is the scalar Laplacian, it is clearly elliptic. Because $\bar{g} = |\psi|^{4/n} g$, we have $dv_{\bar{g}} = |\psi|^2\, dv_g$ and for any $f, h \in C^\infty(M)$,

$$\int_M (L_\psi f) h\, dv_{\bar{g}} = \int_M \big(\Delta f - 2g(\nabla \ln|\psi|, \nabla f) + \tfrac{1}{4}n^2(H^2 + R(\iota)) f\big) h |\psi|^2\, dv_g.$$

Performing a partial integration, we have for the first term

$$\begin{aligned}\int_M (\Delta f) h |\psi|^2\, dv_g &= \int_M g(\nabla f, \nabla h)|\psi|^2 + g(\nabla f, \nabla(|\psi|^2)) h\, dv_g \\ &= \int_M g(\nabla f, \nabla h)|\psi|^2 + 2g(\nabla f, \nabla \ln|\psi|) h |\psi|^2\, dv_g.\end{aligned}$$

Therefore, the first-order term in $\nabla \ln|\psi|$ simplifies and we obtain

$$\int_M (L_\psi f) h\, dv_{\bar{g}} = \int_M \big(g(\nabla f, \nabla h) + \tfrac{1}{4}n^2(H^2 + R(\iota)) f h\big)|\psi|^2\, dv_g,$$

which is clearly symmetric in (f, h). This implies that L_ψ is formally self-adjoint with respect to the metric $\bar{g}$. Since M is closed, we conclude that L_ψ is essentially self-adjoint in $L^2(M)$. □

Remarks 3.2. (1) The operator L_ψ defined in (15) is of the form *drifting Laplacian* plus potential (the drifting Laplacian is also called *Laplacian with drift*, *Bakry–Émery Laplacian*, *weighted Laplacian*, or *Witten Laplacian* in the literature); this is the reason we refer to these operators as *drifting Schrödinger operators*. Indeed, a drifting Laplacian is an operator of the form

$$C^\infty(M) \xrightarrow{L_h} C^\infty(M), \quad f \mapsto \Delta f - g(\nabla h, \nabla f)$$

for some function $h \in C^\infty(M)$. It is elliptic and self-adjoint with respect to the measure $e^h d\mu_g$. Actually, a drifting Laplacian is always *unitarily equivalent* to a

Schrödinger operator: in the notation above, the operator L_h is unitarily equivalent to $\Delta - \frac{1}{2}\Delta h + \frac{1}{4}|\nabla h|_g^2$; see, for example, [Setti 1998, p. 28].

(2) Note that if $|\psi|$ is constant on M (which is the case if ψ is either a parallel or a real Killing spinor on $\widetilde{M}$), then the operator

$$L_\psi = \Delta + \tfrac{1}{4}n^2(H^2 + R(\iota))$$

does not depend on ψ.

Proposition 3.1 implies that the spectrum of L_ψ is purely discrete. We will denote by $\lambda_1(L_\psi)$ its first eigenvalue, which satisfies the min-max characterization

$$\lambda_1(L_\psi) = \inf_{f \in C^\infty(M)\smallsetminus\{0\}} \left(\frac{\int_M f(L_\psi f)\, dv_{\bar{g}}}{\int_M f^2\, dv_{\bar{g}}} \right). \tag{16}$$

We are now ready to give the precise statement of the first main result of this paper, namely:

Theorem 3.3. *Assume M is a closed oriented hypersurface isometrically immersed in a Riemannian spin manifold $(\widetilde{M}^{n+1}, g)$. If there exists a nontrivial twistor-spinor ψ on $\widetilde{M}$ with $\psi_x \neq 0$ for all $x \in M$, then*

$$\lambda_1(D_M^2) \le \lambda_1(L_\psi). \tag{17}$$

Proof. We apply the min-max characterization of $\lambda_1(D_M^2) = \lambda_1(D^2)$ using $f\psi$ as a test section, where $L_\psi f = \lambda_1(L_\psi) f$. The following computations rely on a large extent on those in the proof of [Ginoux 2009, Theorem 5.2.3].

First, if $f \in C^\infty(M)$ is an arbitrary smooth function on M, then using (11), (13), and (9) and the fact that ψ is a twistor-spinor on $\widetilde{M}$,

$$\begin{aligned} D^2(f\psi) &\overset{(11)}{=} f D^2\psi - 2\nabla_{\nabla f}\psi + (\Delta f)\psi \\ &\overset{(13)}{=} f\big(\hat{D}^2\psi + \tfrac{1}{4}n^2H^2\psi + \tfrac{1}{2}n\nabla H\cdot\nu\cdot\psi\big) \\ &\qquad - 2\nabla_{\nabla f}\psi + (\Delta f)\psi \\ &\overset{(9)}{=} f\big(\hat{D}^2\psi + \tfrac{1}{4}n^2H^2\psi + \tfrac{1}{2}n\nabla H\cdot\nu\cdot\psi\big) \\ &\qquad - 2\big(\widehat{\nabla}_{\nabla f}\psi - \tfrac{1}{2}A(\nabla f)\cdot\nu\cdot\psi\big) + (\Delta f)\psi \\ &= f\big(\hat{D}^2\psi + \tfrac{1}{4}n^2H^2\psi + \tfrac{1}{2}n\nabla H\cdot\nu\cdot\psi\big) + \tfrac{2}{n+1}\nabla f\cdot D_{\widetilde{M}}\psi \\ &\qquad + A(\nabla f)\cdot\nu\cdot\psi + (\Delta f)\psi. \end{aligned} \tag{18}$$

Next we compute $\hat{D}^2\psi$, again using the fact that ψ is a twistor-spinor, which implies, in particular, the following identity; see, e.g., [Ginoux 2009, Proposition A.2.1]:

$$\widehat{\nabla}_X(D_{\widetilde{M}}\psi) = \tfrac{n+1}{n-1}\big(-\tfrac{1}{2}\widetilde{\mathrm{Ric}}(X)\cdot\psi + \tfrac{1}{4n}\tilde{S}X\cdot\psi\big), \tag{19}$$

for every $X \in \Gamma(T\widetilde{M})$ and where $\widetilde{\mathrm{Ric}}$ denotes the Ricci tensor of $(\widetilde{M}^{n+1}, g)$ (seen as an endomorphism of the tangent bundle of $\widetilde{M}$). Thus we have

$$\begin{aligned}
\hat{D}^2\psi &= \hat{D}\Big(\sum_{j=1}^n e_j \cdot \widehat{\nabla}_{e_j}\psi\Big) \\
&\overset{(14)}{=} \tfrac{n}{n+1}\hat{D}(D_{\widetilde{M}}\psi) \\
&\overset{(19)}{=} \tfrac{n}{n-1}\sum_{j=1}^n\big(-\tfrac{1}{2}e_j\cdot\widetilde{\mathrm{Ric}}(e_j)\cdot\psi + \tfrac{1}{4n}\tilde{S}e_j\cdot e_j\cdot\psi\big) \\
&= \tfrac{n}{n-1}\big(\tfrac{1}{2}\tilde{S}\psi + \tfrac{1}{2}\nu\cdot\widetilde{\mathrm{Ric}}(\nu)\cdot\psi - \tfrac{1}{4}\tilde{S}\psi\big) \\
&= \tfrac{n}{n-1}\big(\tfrac{n(n-1)}{4}R(\iota)\psi + \tfrac{1}{2}\nu\cdot\widetilde{\mathrm{Ric}}(\nu)^{\top}\cdot\psi\big) \\
&= \tfrac{1}{4}n^2R(\iota)\psi + \tfrac{n}{2(n-1)}\nu\cdot\widetilde{\mathrm{Ric}}(\nu)^{\top}\cdot\psi,
\end{aligned} \tag{20}$$

where $\widetilde{\mathrm{Ric}}(\nu)^{\top} := \sum_{j=1}^n \widetilde{\mathrm{ric}}(\nu, e_j)e_j$ denotes the tangential projection of $\widetilde{\mathrm{Ric}}(\nu)$ on TM. Combining (18) with (20), we deduce that

$$\begin{aligned}
D^2(f\psi) = \tfrac{1}{4}n^2(H^2+R(\iota))f\psi + \tfrac{1}{2}nf\nabla H\cdot\nu\cdot\psi + \tfrac{n}{2(n-1)}f\nu\cdot\widetilde{\mathrm{Ric}}(\nu)^{\top}\cdot\psi \\
+ \tfrac{2}{n+1}\nabla f\cdot D_{\widetilde{M}}\psi + A(\nabla f)\cdot\nu\cdot\psi + (\Delta f)\psi.
\end{aligned} \tag{21}$$

Again, using the fact that ψ is a twistor-spinor on $(\widetilde{M}^{n+1}, g)$, for every $f \in C^\infty(M)$,

$$\begin{aligned}
\mathrm{Re}\langle D^2(f\psi), f\psi\rangle &\overset{(21)}{=} \tfrac{1}{4}n^2(H^2+R(\iota))f^2|\psi|^2 + \tfrac{2}{n+1}f\,\mathrm{Re}\langle\nabla f\cdot D_{\widetilde{M}}\psi, \psi\rangle \\
&\qquad + f(\Delta f)|\psi|^2 \\
&= \tfrac{1}{4}n^2(H^2+R(\iota))f^2|\psi|^2 - g\big(f\nabla f, \nabla(|\psi|^2)\big) \\
&\qquad + f(\Delta f)|\psi|^2 \\
&= f\big(\Delta f - 2g(\nabla f, \nabla\ln|\psi|) + \tfrac{1}{4}n^2(H^2+R(\iota))f\big)|\psi|^2 \\
&= f(L_\psi f)|\psi|^2.
\end{aligned}$$

The min-max principle for $\lambda_1(D^2)$ implies that, for any $f \in C^\infty(M)\setminus\{0\}$,

$$\lambda_1(D^2) \le \frac{\int_M \mathrm{Re}\langle D^2(f\psi), f\psi\rangle\, dv_g}{\int_M |f\psi|^2\, dv_g} = \frac{\int_M f(L_\psi f)\, dv_{\bar{g}}}{\int_M f^2\, dv_{\bar{g}}};$$

therefore,

$$\lambda_1(D^2) \le \inf_{f\in C^\infty(M,\mathbb{R})\setminus\{0\}}\left(\frac{\int_M f(L_\psi f)\, dv_{\bar{g}}}{\int_M f^2\, dv_{\bar{g}}}\right),$$

which from (16) gives the inequality (17). $\square$

Remarks 3.4. (1) The estimate (17) contains all known upper estimates *à la Reilly* for $\lambda_1(D_M^2)$. Indeed, we observe that by taking $f = 1$ in the Rayleigh quotient of L_ψ,

$$\lambda_1(L_\psi) \leq \frac{n^2}{4\,\mathrm{Vol}(M)} \int_M (H^2 + R(\iota))\, dv_g$$

if $|\psi|$ is constant, and

$$\lambda_1(L_\psi) \leq \tfrac{1}{4} n^2 \sup_M (H^2 + R(\iota))$$

otherwise. Those give exactly the inequalities (5) in [Bär 1998] and (7) in [Ginoux 2003b]. On the other hand, for $f = |\psi|^{-1}$ (with respect to the metric $\bar{g}$ defined above), we deduce that

$$\lambda_1(L_\psi) \leq \frac{n^2}{4\,\mathrm{Vol}(M)} \int_M (H^2 + R(\iota))\, dv_g + \frac{1}{\mathrm{Vol}(M)} \int_M \left|d \ln|\psi|\right|^2 dv_g,$$

which was proved in [Ginoux 2002, Theorem 1].

(2) It is interesting to compare (17) with (4). On the one hand, we do not obtain in the spinorial setting the exact analogue of (4) for $\widetilde{M}$ conformally equivalent to an open subset of the sphere $\mathbb{S}^{n+1}$. Of course, this must be expected since otherwise in dimension 2 this would mean that the Willmore functional bounds $\lambda_1(D_M^2) \cdot \mathrm{Area}(M^2, g)$ from above; but there is no conformal upper bound for the smallest positive Dirac eigenvalue on unit-area-metrics, as shown in [Ammann and Jammes 2012, Theorem 1.1]. Note that this does not prevent the analogue of (3) to possibly hold true for the Dirac operator, which is still an open question. On the other hand, our assumption on $\widetilde{M}$ in Theorem 3.3 is much more general since not only open subsets of spheres with conformal metrics allow twistor-spinors. We refer the reader to [Kühnel and Rademacher 1998] for the classification of Riemannian spin manifolds with twistor-spinors.

We now look at the equality case of the previous estimate in the situation where the twistor-spinor is also an eigenspinor for the Dirac operator of $\widetilde{M}$. More precisely, we prove:

Proposition 3.5. *Under the same assumptions as in Theorem 3.3, assume moreover that equality is achieved in (17).*

(1) *If ψ is a parallel spinor on $\widetilde{M}^{n+1}$, then*

$$A(\nabla \ln|f|) = -\tfrac{1}{2} n \nabla H$$

for any eigenfunction f of L_ψ associated with $\lambda_1(L_\psi)$.

(2) *If ψ is a real Killing spinor on $\widetilde{M} = \mathbb{S}^{n+1}$ or an imaginary Killing spinor on $\widetilde{M} = \mathbb{H}^{n+1}$, then the mean curvature H is constant and, in particular,*

$$\lambda_1(D_M^2) = \tfrac{1}{4}n^2(H^2+\kappa).$$

Proof. (1) If (17) is an equality and ψ is a parallel spinor, then the min-max principle yields $D^2(f\psi) = \lambda_1(D^2)f\psi$ for any eigenfunction f of L_ψ associated with $\lambda_1(L_\psi) = \lambda_1(D^2)$. But (21) together with $\widetilde{\mathrm{Ric}} = 0$ and $D_{\widetilde{M}}\psi = 0$ (both provided by $\widehat{\nabla}\psi = 0$) implies

$$\begin{aligned}\lambda_1(D^2)f\psi &= \tfrac{1}{4}n^2H^2f\psi + \tfrac{1}{2}nf\nabla H\cdot\nu\cdot\psi + A(\nabla f)\cdot\nu\cdot\psi + (\Delta f)\psi\\ &= (L_\psi f)\psi + \big(A(\nabla f) + \tfrac{1}{2}nf\nabla H\big)\cdot\nu\cdot\psi.\end{aligned}$$

With $\lambda_1(D^2) = \lambda_1(L_\psi)$, we deduce that

$$\big(A(\nabla f) + \tfrac{1}{2}nf\nabla H\big)\cdot\nu\cdot\psi = 0$$

which, since $\psi \neq 0$, gives $A(\nabla f) + \frac{1}{2}nf\nabla H = 0$. Since any eigenfunction for L_ψ associated with the eigenvalue $\lambda_1(L_\psi)$ is either positive or negative, we easily reach the conclusion.

(2) Assume first that $\widetilde{M}^{n+1}$ carries real Killing spinors and let ψ be a nonzero $(\varepsilon/2)$-Killing spinor for some $\varepsilon \in \{\pm 1\}$; that is, $\widehat{\nabla}_X\psi = \frac{\varepsilon}{2}X\cdot\psi$ for all $X \in \Gamma(T\widetilde{M})$. Again, one obtains $D^2(f\psi) = \lambda_1(D^2)f\psi$ for any eigenfunctions $f \in C^\infty(M)$ associated to $\lambda_1(L_\psi)$. Fixing such an f, the identity (21) yields

$$\lambda_1(D^2)f\psi = (L_\psi f)\psi + \big(A(\nabla f) + \tfrac{1}{2}nf\nabla H\big)\cdot\nu\cdot\psi - \varepsilon\nabla f\cdot\psi.$$

With $\lambda_1(D^2) = \lambda_1(L_\psi)$, we deduce that

$$\big(A(\nabla f) + \tfrac{1}{2}nf\nabla H\big)\cdot\nu\cdot\psi - \varepsilon\nabla f\cdot\psi = 0.$$

In particular, with the notation $Y_\varepsilon := -\varepsilon\nabla f$ and $X := A(\nabla f) + \frac{1}{2}nf\nabla H$, we have $(Y_\varepsilon + X\wedge\nu)\cdot\psi = 0$. At this point, we need the following claim:

Claim: Let $\alpha \in \bigwedge^*\mathbb{R}^{n+1}\otimes\mathbb{C}$. If n is odd, then $\delta_{n+1}(\alpha) = 0$ if and only if $\alpha = 0$. If n is even, then the same equivalence holds for $\alpha \in \bigwedge^*\mathbb{R}^n\otimes\mathbb{C}$.

Proof of Claim. Recall that the spinor representation $\delta_k : \mathrm{Cl}_k \longrightarrow \mathrm{End}_{\mathbb{C}}(\Sigma_k)$ of the complex Clifford algebra in dimension k is a complex-linear isomorphism for k even (but obviously not for k odd). So if n is odd, the claim follows directly from this fact. If n is even and $\alpha \in \bigwedge^*\mathbb{R}^n\otimes\mathbb{C}$, then $\Sigma_n \cong \Sigma_{n+1}$ and it is a simple trick to rewrite $\delta_{n+1}(\alpha)$ in the form $\delta_n(\check{\alpha})$ for a form $\check{\alpha} \in \bigwedge^*\mathbb{R}^n\otimes\mathbb{C}$ having the same coefficients as α in the canonical basis of $\bigwedge^*\mathbb{R}^n\otimes\mathbb{C}$, up to sign and some power of i. Namely, write

$$\alpha = \sum_{1\le j_1<\cdots<j_k\le n} \alpha_{j_1,\dots,j_k}\, e_{j_1}^*\wedge\cdots\wedge e_{j_k}^*,$$

where $(e_1, \dots, e_n, e_{n+1})$ is the canonical basis of $\mathbb{R}^{n+1}$. Let $\omega_n^{\mathbb{C}}$ denotes the complex volume form on $\mathbb{R}^n$ as defined in the proof of Proposition 3.5, which acts on Σ_n via $\delta_n(\omega_n^{\mathbb{C}}) = \mathrm{Id}_{\Sigma_n^+} \oplus -\mathrm{Id}_{\Sigma_n^-}$. Since, for all $v \in \mathbb{R}^n$, $\delta_{n+1}(ie_{n+1}) = \delta_n(\omega_n^{\mathbb{C}})$ and $\delta_n(v) = \delta_{n+1}(v) \circ \delta_{n+1}(e_{n+1})$, after some calculation,

$$\delta_{n+1}(\alpha) = \sum_{\substack{k \text{ even} \\ 1 \le j_1 < \dots < j_k \le n}} \alpha_{j_1,\dots,j_k} \delta_n(e_{j_1}) \circ \dots \circ \delta_n(e_{j_k}) + i \sum_{\substack{k \text{ odd} \\ 1 \le j_1 < \dots < j_k \le n}} \alpha_{j_1,\dots,j_k} \delta_n(e_{j_1}) \circ \dots \circ \delta_n(e_{j_k}) \circ \delta_n(\omega_n^{\mathbb{C}}).$$

Now it is an elementary computation to show that, for any $\beta \in \bigwedge^k \mathbb{R}^n$,

$$\delta_n(\beta) \circ \delta_n(e_1^* \wedge \dots \wedge e_n^*) = (-1)^{k(k+1)/2} \delta_n(*\beta),$$

where $* : \bigwedge^* \mathbb{R}^n \to \bigwedge^* \mathbb{R}^n$ is the Hodge-star operator. Therefore, we obtain

$$\delta_{n+1}(\alpha) = \sum_{\substack{k \text{ even} \\ 1 \le j_1 < \dots < j_k \le n}} \alpha_{j_1,\dots,j_k} \delta_n(e_{j_1}) \circ \dots \circ \delta_n(e_{j_k}) + c_{n,k} \sum_{\substack{k \text{ odd} \\ 1 \le j_1 < \dots < j_k \le n}} \alpha_{j_1,\dots,j_k} \delta_n(*(e_{j_1}^* \wedge \dots \wedge e_{j_k}^*)) = \delta_n(\check{\alpha}),$$

where we let $c_{n,k} := i^{n/2+1}(-1)^{k(k+1)/2}$ and

$$\check{\alpha} := \sum_{\substack{k \text{ even} \\ 1 \le j_1 < \dots < j_k \le n}} \alpha_{j_1,\dots,j_k} e_{j_1}^* \wedge \dots \wedge e_{j_k}^* + c_{n,k} \sum_{\substack{k \text{ odd} \\ 1 \le j_1 < \dots < j_k \le n}} \alpha_{j_1,\dots,j_k} *(e_{j_1}^* \wedge \dots \wedge e_{j_k}^*).$$

As a consequence, if $\delta_{n+1}(\alpha)\sigma = 0$ for all $\sigma \in \Sigma_{n+1} \cong \Sigma_n$, then $\delta_n(\check{\alpha}) = 0$, and the fact mentioned above implies $\check{\alpha} = 0$; since n is even, each form $*(e_{j_1}^* \wedge \dots \wedge e_{j_k}^*)$ is of odd degree when k is odd and therefore $\alpha_{j_1,\dots,j_k} = 0$ for all $1 \le j_1 < \dots < j_k \le n$; that is, $\alpha = 0$. This concludes the proof of the claim. □

If $\widetilde{M}^{n+1}$ is isometric to the standard round sphere $\mathbb{S}^{n+1}$, then it carries a *maximal* number (that is $2^{\lfloor \frac{n+1}{2} \rfloor}$) of linearly independent $(\varepsilon/2)$-Killing spinors. In that case, $(Y_\varepsilon + X \wedge \nu) \cdot \psi = 0$ holds pointwise for every $\psi \in \Sigma_x \widetilde{M}$. If n is odd, then the claim yields $Y_\varepsilon + X \wedge \nu = 0$, which implies $X = Y_\varepsilon = 0$; that is, f and H are constant. If n is even, then one may rewrite

$$Y_\varepsilon \cdot \psi + X \cdot \nu \cdot \psi = iY_\varepsilon \cdot i\nu \cdot \nu \cdot \psi + X \cdot \nu \cdot \psi = (X - iY_\varepsilon \lrcorner \omega_M^{\mathbb{C}}) \cdot \nu \cdot \psi,$$

where $\omega_M^{\mathbb{C}} := i^{\lfloor \frac{n+1}{2} \rfloor} e_1^* \wedge \dots \wedge e_n^* \in \Gamma(\bigwedge^n T^*M \otimes \mathbb{C})$ is the complex volume form on M. Again, the claim yields $X - iY_\varepsilon \lrcorner \omega_M^{\mathbb{C}} = 0$. If $n > 2$, then comparing the degrees yields $X = Y_\varepsilon = 0$; that is, f and H are constant. If $n = 2$, then an elementary computation gives $Z \lrcorner \omega_M^{\mathbb{C}} = iJ(Z)$ for every $Z \in \Gamma(TM)$, where J is the Kähler structure associated to the metric and the orientation on (M^2, g). In that case, one obtains $X + J(Y_\varepsilon) = 0$. However on the standard sphere $\mathbb{S}^3$, both spaces of $\pm\frac{1}{2}$-Killing spinors have maximal dimension 2, therefore $X + J(Y_\varepsilon) = 0$ for both $\varepsilon \in \{\pm 1\}$, which implies $X = Y_\varepsilon = 0$ and hence that f and H are constant.

The case of imaginary Killing spinors is much the same up to replacing ε by $i\varepsilon$. One obtains at the end $(iY_\varepsilon + X \wedge \nu) \cdot \psi = 0$ for all $(i\varepsilon/2)$-Killing spinors ψ on $\widetilde{M}^{n+1}$. The same arguments as above lead to $X = Y_\varepsilon = 0$. Notice that in the case $n = 2$, one does not need the existence of maximal spaces of $(i\varepsilon/2)$-Killing spinors for both $\varepsilon \in \{\pm 1\}$ since X and Y_ε are real vector fields on M. □

Remark 3.6. It is quite surprising that in the case where ψ is a parallel spinor we cannot conclude that the mean curvature of M must be constant. In fact, we are left to prove that if there exists a smooth positive function $f \in C^\infty(M)$ such that

$$\Delta f + \tfrac{1}{4}n^2 H^2 f = \lambda_1(D)^2 f \quad \text{and} \quad A(\nabla \ln f) = -\tfrac{1}{2}n\nabla H,$$

then f (or, equivalently, H) is constant on M.

4. The equality case in the presence of imaginary Killing spinors

In this section, we focus on the equality case of our estimate (17) when the ambient manifold $\widetilde{M}$ carries an imaginary Killing spinor. According to Proposition 3.5, it also corresponds to the equality case of the inequality (7). It is obvious to check that totally umbilical round spheres in the hyperbolic space $\mathbb{H}^{n+1}$ satisfy the equality in this estimate; however, it is still unknown if they are the only ones. In fact, if the hypersurface is embedded, then this result easily follows from the Alexandrov theorem in the hyperbolic space; see [Montiel 1999]. However, if the hypersurface is only assumed to be *immersed*, then the question is still open. In order to settle this problem, we adopt a method introduced in [Hijazi and Montiel 2013] which relies on the fact that such hypersurfaces are *critical points* for some eigenvalue functional associated to some Dirac-type operator on M. The main result of this section concerns the case when $\widetilde{M} = \mathbb{H}^{n+1}$ but actually we will prove the following more general statement:

Theorem 4.1. *Let M^n be an oriented, compact, connected hypersurface immersed in a Riemannian spin manifold $(\widetilde{M}^{n+1}, g)$. If $\widetilde{M}$ carries an $(i\varepsilon/2)$-Killing spinor for some $\varepsilon \in \{\pm 1\}$, then* (7) (*as well as* (17)) *holds and if equality holds then the mean curvature H is constant. Moreover, if $\widetilde{M}$ also carries a $(-i\varepsilon/2)$-Killing spinor, then equality holds if and only if M is totally umbilical with constant mean curvature.*

Since the standard hyperbolic space $\mathbb{H}^{n+1}$ has both $(i/2)$- and $(-i/2)$-Killing spinors (see, e.g., [Baum 1989a]), the previous result immediately implies the next:

Corollary 4.2. *The only oriented, compact, connected hypersurfaces immersed into the hyperbolic space* $\mathbb{H}^{n+1}$ *satisfying* $\lambda_1(D_M^2) = \frac{1}{4}n^2(H^2-1)$ *are the totally umbilical round spheres.*

In Section 4D, we will discuss the case of pseudohyperbolic spaces.

4A. ***The Hijazi–Montiel approach in the presence of imaginary Killing spinors.*** Assume that the ambient manifold $\widetilde{M}$ carries an $(i/2)$-Killing spinor $\Psi \in \Gamma(\Sigma\widetilde{M})$. After restriction to M, it is a straightforward computation to show that Ψ satisfies the modified Dirac equation

$$D_+\Psi = \tfrac{n}{2}H\Psi \tag{22}$$

where D_+ is a zero order modification of the extrinsic Dirac operator defined by

$$D_+\varphi := D\varphi - \tfrac{n}{2}i\nu\cdot\varphi \tag{23}$$

for $\varphi \in \Gamma(\Sigma)$. Note that we do not assume that the mean curvature H is constant for the moment. Suppose however that H is positive everywhere on M, and consider the metric conformally related to g on M, defined by $\bar{g} := H^2 g$. It is a well-known fact [Hitchin 1974; Hijazi 1986] that under a conformal change of the metric, there exists a bundle isometry $\varphi \mapsto \overline{\varphi}$, $\Sigma \to \overline{\Sigma}$, between the two extrinsic spinor bundles Σ and $\overline{\Sigma}$ over (M^n, g) and $(M^n, \bar{g})$. Under this identification, the extrinsic Dirac operators D and D^H associated to g and $\bar{g}$ and acting respectively on Σ and $\overline{\Sigma}$ are related by

$$D^H\overline{\varphi} = H^{-(n+1)/2}\,\overline{D(H^{(n-1)/2}\varphi)} \tag{24}$$

for all $\varphi \in \Gamma(\Sigma)$. Now consider on $\overline{\Sigma}$ the zero order modification of the extrinsic Dirac operator D^H given by

$$D_+^H\overline{\varphi} := D^H\overline{\varphi} - \tfrac{n}{2}H^{-1}\mathcal{I}_\nu\overline{\varphi}$$

where $\mathcal{I}_\nu$ is the Hermitian endomorphism of $\overline{\Sigma}$ defined by $\mathcal{I}_\nu\overline{\varphi} := \overline{i\nu\cdot\varphi}$ for all $\varphi \in \Gamma(\Sigma)$. Notice that D_+^H is an elliptic and self-adjoint differential operator of order one which, since M is assumed to be compact, has a discrete spectrum. In the following, we will denote by $\lambda_1(D_+^H)$ the first nonnegative eigenvalue of D_+^H. Now for every $\varphi \in \Gamma(\Sigma)$, consider the spinor field $\varphi_H := H^{-(n-1)/2}\varphi \in \Gamma(\Sigma)$ which is easily seen to satisfy

$$D_+^H\overline{\varphi}_H = H^{-(n+1)/2}\,\overline{D_+\varphi},$$

using the conformal covariance (24) of D. Using (22) on the $(i/2)$-Killing spinor $\Psi \in \Gamma(\Sigma\widetilde{M})$ in the previous identity gives $D_+^H\overline{\Psi}_H = \frac{n}{2}\overline{\Psi}_H$. This immediately implies that $\lambda_1(D_+^H) \le \frac{n}{2}$. Furthermore, if the mean curvature H is constant, it is an easy computation using $\{D, i\nu\cdot\}=0$ to show that

$$\mathrm{Spec}((D^H_+)^2)=\left\{\lambda_k((D^H_+)^2)=\frac{H^{-2}\big(\lambda_k(D)^2+\frac{1}{4}n^2\big)}{\lambda_k(D)}\in\mathrm{Spec}(D)\right\},$$

so that $\lambda_1(D^H_+)=\frac{n}{2}$ if and only if $\lambda_1(D^2)=\frac{n^2}{4}(H^2-1)$. Thus, we have proved this:

Proposition 4.3. *Let M be an orientable, compact, connected hypersurface immersed in a Riemannian spin manifold $(\widetilde{M}^{n+1}, g)$ admitting a $(i/2)$-Killing spinor, and suppose that the mean curvature of M, after a suitable choice of the unit normal, satisfies $H>0$. Then the first nonnegative eigenvalue of D^H_+ satisfies $\lambda_1(D^H_+)\le\frac{n}{2}$. Moreover, if H is constant, equality occurs if and only if equality occurs in (7).*

From this proposition, we deduce that any immersion for which (7) — or equivalently (17) — is an equality realizes a *maximum* for the map

$$\mathcal{F}^+_1:\mathrm{Imm}^+(M,\widetilde{M})\to\mathbb{R},\quad \iota\mapsto\lambda_1(D^{H_\iota}_+),$$

where $\mathrm{Imm}^+(M,\widetilde{M})$ denotes the space of isometric immersions of M in $\widetilde{M}$ with nonvanishing mean curvature H_ι. This characterization of hypersurfaces satisfying the equality case in (7) leads to the study of the critical points of the functional $\mathcal{F}^+_1$.

Remark 4.4. If the manifold $\widetilde{M}$ carries a $(-i/2)$-Killing spinor, then Proposition 4.3 is true with the operators D_+ and D^H_+ replaced respectively by

$$D_-:=D+\tfrac{n}{2}i\nu\cdot\,:\Gamma(\Sigma)\to\Gamma(\Sigma)$$

and

$$D^H_-:=D^H+\tfrac{n}{2}H^{-1}\mathcal{I}_\nu:\Gamma(\overline{\Sigma})\to\Gamma(\overline{\Sigma}). \tag{25}$$

In this situation, the corresponding functional is defined by

$$i\mathcal{F}^-_1:\iota\mapsto\lambda^-_1(D^{H_\iota}_-)$$

where $\lambda^-_1(D^{H_\iota}_-)$ is the first nonnegative eigenvalue of $D^{H_\iota}_-$.

4B. ***Derivatives of the functional $\mathcal{F}^{\pm}_1$.*** As explained in the previous section we are led to study the first derivatives of the functional $\mathcal{F}^{\pm}_1$ at least in a particular situation. As above, we start with an immersion $\iota=\iota_0:M\to\widetilde{M}$ with positive mean curvature (not necessarily constant) and such that $\lambda_1(D^H_+)=\frac{n}{2}$. Note that here *we do not assume the existence of imaginary Killing spinor fields on $\widetilde{M}$.*

Now we deform the immersion ι along normal geodesics; that is, we consider, for $\varepsilon>0$ sufficiently small, the map $F:\,]-\varepsilon,\varepsilon[\,\times M\to\widetilde{M}$, $(t,x)\mapsto\exp_{\iota(x)}(t\nu_x)$. Note that, choosing $\varepsilon>0$ sufficiently small, the map F is smooth and $F(t,\cdot):M\to\widetilde{M}$ is an immersion such that $F(0,\cdot)=\iota$. In fact, the map $t\mapsto F(t,x)$ is the geodesic starting from $\iota(x)$ with speed vector ν_x, and so it is analytic. For each $t \in]-\varepsilon, \varepsilon[$, we denote by $g_t := F(t,\cdot)^* g$ the induced metric on M, by ν_t the unit normal field inducing the orientation of M, by $H_t := -\frac{1}{n}\,\mathrm{tr}(\widehat{\nabla}\nu_t)$ the mean curvature of $F(t,\cdot)$ — which, up to making $\varepsilon > 0$ smaller, may be assumed to be positive on M for all $t \in]-\varepsilon, \varepsilon[$ — and we set $\bar{g}_t := H_t^2 g_t$.

We also denote by D^{H_t} the Dirac operator associated to the metric $\bar{g}_t$ and let $D_+^{H_t} := D^{H_t} - \frac{n}{2} H_t^{-1} \mathcal{I}_{\nu_t} : \Gamma(\overline{\Sigma}_t) \to \Gamma(\overline{\Sigma}_t)$, where $\mathcal{I}_{\nu_t}$ is the Hermitian endomorphism of $\overline{\Sigma}_t$ defined by $\mathcal{I}_{\nu_t}\overline{\varphi} := \overline{i\nu_t \cdot \varphi}$. Here $\overline{\Sigma}_t$ denotes the extrinsic spinor bundle over M endowed with the spin structure induced by $\widetilde{M}$ and the Riemannian metric $\bar{g}_t$. Since we perturb the immersion analytically, the family $(D_+^{H_t})$ with $t \in]-\varepsilon, \varepsilon[$ is an analytic family of unbounded closed self-adjoint operators with compact resolvent, therefore the spectrum of $D_+^{H_t}$ can be written as a sequence $(\mu_k^+(t))_{k\in\mathbb{N}}$, where each eigenvalue $\mu_k^+(t)$ depends analytically on t and where corresponding eigenvectors can be found to also depend analytically on t; see [Kato 1995]. We denote by $\lambda_1^+(t)$ any branch of that spectrum with $\lambda_1^+(0) = \lambda_1(D_+^H)$, the smallest nonnegative eigenvalue of $D_+^H = D_+^{H_0}$. Following [Bär et al. 2005], we denote by $\tau_0^t : \overline{\Sigma}_0 = \overline{\Sigma} \to \overline{\Sigma}_t$ the parallel transport along the curves $s \mapsto (s, x)$ in the so-called generalized cylinder $\big(]-\varepsilon, \varepsilon[\times M, dt^2 \oplus \bar{g}_t\big)$, for all $t \in]-\varepsilon, \varepsilon[$. Then for any analytic family $(\overline{\Phi}_t)_t$ of eigenvectors associated to $\lambda_1^+(t)$, differentiating the identity

$$\lambda_1^+(t) \int_M |\overline{\Phi}_t|^2 \, dv_{\bar{g}_t} = \int_M \mathrm{Re}\big\langle D_+^{H_t}\overline{\Phi}_t, \overline{\Phi}_t \big\rangle \, dv_{\bar{g}_t}$$

at $t = 0$ yields

$$\frac{d\lambda_1^+}{dt}(0) \int_M |\overline{\Phi}_0|^2 \, dv_{\bar{g}_0} = \int_M \mathrm{Re}\Big\langle \frac{d}{dt}\Big|_{t=0} (\tau_t^0 D_+^{H_t} \tau_0^t \overline{\Phi}_0), \overline{\Phi}_0 \Big\rangle \, dv_{\bar{g}_0}.$$

Now we have $\tau_t^0 D_+^{H_t} \tau_0^t = \tau_t^0 D^{H_t} \tau_0^t - \frac{n}{2} H_t^{-1} \tau_t^0 \mathcal{I}_{\nu_t} \tau_0^t$ and, since the variation of ι is a geodesic normal one, the vector field $\nu_t = \frac{\partial}{\partial t}$ is parallel along the curves $s \mapsto (s, x)$, so that $\tau_t^0 \mathcal{I}_{\nu_t} \tau_0^t = \mathcal{I}_{\nu_0} = \mathcal{I}_\nu$ for all $t \in]-\varepsilon, \varepsilon[$. With the formula for the first variation of the Dirac operator by J.-P. Bourguignon and P. Gauduchon [1992] (see also [Bär et al. 2005]), we deduce that

$$\begin{aligned} &\frac{d\lambda_1^+}{dt}(0) \int_M |\overline{\Phi}_0|^2 \, dv_{\bar{g}_0} \\ &\quad = -\frac{1}{2} \int_M \bar{g}_0\Big(T_{\overline{\Phi}_0}, \frac{\partial \bar{g}_t}{\partial t}(0)\Big) \, dv_{\bar{g}_0} + \frac{n}{2} \int_M H^{-2} \frac{\partial H_t}{\partial t}\Big|_{t=0} \mathrm{Re}\big\langle \mathcal{I}_\nu \overline{\Phi}_0, \overline{\Phi}_0 \big\rangle \, dv_{\bar{g}_0}, \end{aligned}$$

where

$$T_{\overline{\Phi}_0}(X, Y) := \tfrac{1}{2} \mathrm{Re}\big\langle X \cdot_{\overline{\Sigma}} \overline{\nabla}_Y \overline{\Phi}_0 + Y \cdot_{\overline{\Sigma}} \overline{\nabla}_X \overline{\Phi}_0, \overline{\Phi}_0 \big\rangle$$

is the so-called *energy–momentum* tensor associated to $\overline{\Phi}_0$. Here $\cdot_{\overline{\Sigma}}$ is the Clifford multiplication on $\overline{\Sigma}$ defined by (8) and $\overline{\nabla}$ is the spin Levi–Civita connection with respect to the metric $\bar{g}_0$. Note that we kept the same notation for the Hermitian scalar products on $\overline{\Sigma}$ and Σ. Now fix an eigenvector $\overline{\Phi}_0 \in \Gamma(\overline{\Sigma})$ for the Dirac-type operator D_+^H associated with $\lambda_1(D_+^H)$ and let $\overline{\Psi}_0 := H^{(n-1)/2}\overline{\Phi}_0$. We compute $d\lambda_1^+/dt(0)$ in terms of $\Psi_0 \in \Gamma(\Sigma)$ and of geometric quantities attached to ι. First, since $\partial F/\partial t(0,\cdot) = \nu$, we have on the one hand (see [Montiel 1999])

$$\frac{\partial \bar{g}_t}{\partial t}(0) = \frac{\partial}{\partial t}\Big|_{t=0}(H_t^2 g_t) = \tfrac{1}{n}2H\big(|A|^2 + \widetilde{\mathrm{ric}}(\nu,\nu)\big)g - 2H^2 g(A\cdot,\cdot).$$

On the other hand, using the isomorphism $\Sigma \to \overline{\Sigma}$, we may write (see, e.g., [Ginoux 2009, Section 1.3])

$$T_{\overline{\Phi}_0}(X,Y) = H^{-n+2}T_{\Psi_0}(X,Y),$$

for all $X, Y \in \Gamma(TM)$, where T_{Ψ_0} is the energy–momentum tensor associated to Ψ_0 defined by

$$T_{\Psi_0}(X,Y) := \tfrac{1}{2}\,\mathrm{Re}\big\langle X \underset{\Sigma}{\cdot} \nabla_Y\Psi_0 + Y \underset{\Sigma}{\cdot} \nabla_X\Psi_0, \Psi_0\big\rangle.$$

Therefore, assuming without loss of generality that $\int_M |\overline{\Phi}_0|^2\,dv_{\bar{g}_0} = 1$, we compute

$$\frac{d\lambda_1^+}{dt}(0) = \frac{1}{n}\int_M H^{-1}\big(|A|^2 + \widetilde{\mathrm{ric}}(\nu,\nu)\big)\big(\tfrac{n}{2}\,\mathrm{Re}\,\langle i\nu\cdot\Psi_0, \Psi_0\rangle - g(T_{\Psi_0}, g)\big)\,dv_g + \int_M g(T_{\Psi_0}, A)\,dv_g.$$

But since $g(T_{\Psi_0}, g) = \mathrm{tr}_g(T_{\Psi_0}) = \mathrm{Re}\,\langle D\Psi_0, \Psi_0\rangle$, we obtain

$$\frac{d\lambda_1^+}{dt}(0) = -\frac{1}{n}\int_M H^{-1}(|A|^2 + \widetilde{\mathrm{ric}}(\nu,\nu))\,\mathrm{Re}\,\langle D_+\Psi_0, \Psi_0\rangle\,dv_g + \int_M g(T_{\Psi_0}, A)\,dv_g.$$

However, since $\overline{\Phi}_0 \in \Gamma(\overline{\Sigma})$ is an eigenspinor for D_+^H associated with the eigenvalue $\lambda_1^+(0) = \frac{n}{2}$ and from the equivalence

$$D_+^H\overline{\Phi}_0 = \tfrac{n}{2}\overline{\Phi}_0 \iff D_+\Psi_0 = \tfrac{n}{2}H\Psi_0, \tag{26}$$

one concludes that

$$\frac{d\lambda_1^+}{dt}(0) = -\frac{1}{2}\int_M (|A|^2 + \widetilde{\mathrm{ric}}(\nu,\nu))\,|\Psi_0|^2\,dv_g + \int_M g(T_{\Psi_0}, A)\,dv_g. \tag{27}$$

To compute the remaining term $g(T_{\Psi_0}, A)$, we define a new covariant derivative by $\widehat{\nabla}_X^+ := \widehat{\nabla}_X - (i/2)X\cdot$ on Σ. Then a lengthy but direct calculation using the spin

Gauss formula (9) yields that for any $\varphi \in \Gamma(\Sigma)$,

$$\begin{aligned}|\widehat{\nabla}^+\varphi|^2 &:= \sum_{j=1}^n |\widehat{\nabla}^+_{e_j}\varphi|^2\\ &= \sum_{j=1}^n \big|\nabla_{e_j}\varphi + \tfrac{1}{2}A(e_j)\cdot\nu\cdot\varphi - \tfrac{i}{2}e_j\cdot\varphi\big|^2\\ &= |\nabla\varphi|^2 + \tfrac{1}{4}(|A|^2+n)|\varphi|^2 - g(T_\varphi, A) - \operatorname{Re}\big\langle i\nu\cdot(D\varphi - \tfrac{1}{2}nH\varphi), \varphi\big\rangle.\end{aligned}$$

For $\varphi = \Psi_0$, we deduce, using the right-hand side of (26), that

$$g(T_{\Psi_0}, A) = |\nabla\Psi_0|^2 - |\widehat{\nabla}^+\Psi_0|^2 + \tfrac{1}{4}(|A|^2 - n)|\Psi_0|^2.$$

Now integrating this identity over M with the help of the famous Schrödinger–Lichnerowicz formula

$$D^2 = \nabla^*\nabla + \tfrac{1}{4}S$$

gives

$$\begin{aligned}&\int_M g(T_{\Psi_0}, A)\,dv_g\\ &\qquad = \int_M \big(\operatorname{Re}\langle D^2\Psi_0, \Psi_0\rangle - \tfrac{1}{4}S|\Psi_0|^2 - |\widehat{\nabla}^+\Psi_0|^2 + \tfrac{1}{4}(|A|^2-n)|\Psi_0|^2\big)\,dv_g.\end{aligned}$$

Here S stands for the scalar curvature of (M^n, g). On the other hand, from (10), (26), and the anticommutativity rule $\{D, i\nu\cdot\} = 0^+$, we check that

$$D^2\Psi_0 = \tfrac{1}{4}n^2(H^2-1)\Psi_0 + \tfrac{1}{2}n\nabla H\cdot\nu\cdot\Psi_0,$$

so that $\operatorname{Re}\langle D^2\Psi_0, \Psi_0\rangle = \tfrac{1}{4}n^2(H^2-1)|\Psi_0|^2$, and hence

$$\int_M g(T_{\Psi_0}, A)\,dv_g = \int_M \big(\tfrac{1}{4}(n^2(H^2-1) - S + |A|^2 - n)|\Psi_0|^2 - |\widehat{\nabla}^+\Psi_0|^2\big)\,dv_g.$$

The Gauss formula for the scalar curvature provides

$$S = \tilde{S} - 2\widetilde{\operatorname{ric}}(\nu,\nu) + n^2H^2 - |A|^2,$$

from which

$$\begin{aligned}&\int_M g(T_{\Psi_0}, A)\,dv_g\\ &\quad = -\int_M \big(\tfrac{1}{4}(\tilde{S} + n(n+1)) - \tfrac{1}{2}(|A|^2 + \widetilde{\operatorname{ric}}(\nu,\nu))\big)|\Psi_0|^2\,dv_g - \int_M |\widehat{\nabla}^+\Psi_0|^2\,dv_g\end{aligned}$$

follows. Inserting this identity in (27), we finally deduce that

$$\frac{d\lambda_1^+}{dt}(0) = -\int_M \big(|\widehat{\nabla}^+\Psi_0|^2 + \tfrac{1}{4}(\tilde{S} + n(n+1))|\Psi_0|^2\big)\,dv_g.$$

It is worth noticing that this formula holds if we assume that it is the first nonnegative eigenvalue $\lambda_1(D_-^H)$ of D_-^H which satisfies $\lambda_1(D_-^H) = \frac{n}{2}$ instead of $\lambda_1(D_+^H)$; in this situation, $\widehat{\nabla}^+$ has to be replaced with the covariant derivative defined by $\widehat{\nabla}_X^- := \widetilde{\nabla}_X + \frac{i}{2} X \cdot$.

From this computation, it is now straightforward to give a necessary condition for an immersion ι to be a critical point of $\mathcal{F}_1^{\pm}$:

Theorem 4.5. *Let M be an oriented, compact, connected hypersurface isometrically immersed in a Riemannian spin manifold $(\widetilde{M}^{n+1}, g)$. Assume that the scalar curvature $\tilde{S}$ of $\widetilde{M}$ is greater or equal to $-n(n+1)$ and that the mean curvature H of M with respect to a suitable choice of the normal is positive. If $\lambda_1(D_\varepsilon^H) = \frac{n}{2}$ for some $\varepsilon \in \{\pm 1\}$ and it is critical for all the variations of the hypersurface M in $\widetilde{M}$, then $\tilde{S} = -n(n+1)$ and $\widetilde{\nabla}_X \Psi = \frac{i\varepsilon}{2} X \cdot \Psi$ for all $X \in \Gamma(TM)$ for all $\Psi \in \Gamma(\Sigma)$ satisfying*

$$D_\varepsilon \Psi = \frac{n}{2} H \Psi.$$

4C. *Proof of Theorem 4.1.* If $\widetilde{M}$ carries a $(i\varepsilon/2)$-Killing spinor for some $\varepsilon \in \{\pm 1\}$, then from Theorem 3.3 and Remarks 3.4, the inequalities (17) and (7) hold. Moreover, if equality holds in (17), then Proposition 3.5 implies that the mean curvature is constant and then $\lambda_1(D)^2 = \frac{1}{4} n^2 (H^2 - 1)$.

Assume now that $\widetilde{M}$ carries an $(i/2)$- as well as a $(-i/2)$-Killing spinor. From Proposition 4.3, we deduce that such an immersion is a maximum for the functional $\mathcal{F}_1^+$ and thus $d\lambda_1^+/dt(0) = 0$. Let Φ be a nonzero $(-i/2)$-Killing spinor on $\widetilde{M}$ so that $D_- \Phi = \frac{n}{2} H \Phi$. From this equation and since H is constant, a direct computation shows that the spinor $\tilde{\Phi} := H\Phi - i\nu \cdot \Phi$ satisfies $D_+ \tilde{\Phi} = \frac{n}{2} H \tilde{\Phi}$. On the other hand, since the existence of a $(\pm i/2)$-Killing spinor on $\widetilde{M}$ implies that $\widetilde{M}$ is an Einstein manifold with scalar curvature $\tilde{S} = -n(n+1)$ (see [Baum et al. 1991], for example), Theorem 4.5 applies and we get that $\widehat{\nabla}_X \tilde{\Phi} = \frac{i}{2} X \cdot \tilde{\Phi}$ for all $X \in \Gamma(TM)$; that is,

$$\begin{aligned} \tfrac{i}{2} X \cdot (H\Phi - i\nu \cdot \Phi) &= \widehat{\nabla}_X (H\Phi - i\nu \cdot \Phi) \\ &= H\big(-\tfrac{i}{2} X \cdot \Phi\big) + iA(X) \cdot \Phi - i\nu \cdot \big(-\tfrac{i}{2} X \cdot \Phi\big) \\ &= iA(X) \cdot \Phi - \tfrac{i}{2} H X \cdot \Phi - \tfrac{i}{2} X \cdot i\nu \cdot \Phi. \end{aligned}$$

This implies that $(A(X) - HX) \cdot \Phi = 0$ for all $X \in \Gamma(TM)$, and since Φ has no zero, M is totally umbilical. This concludes the proof of Theorem 4.1.

4D. *The case of pseudohyperbolic spaces.* In this section, we examine the case of other *complete* ambient manifolds $\widetilde{M}$ carrying imaginary Killing spinors. These manifolds have been classified by H. Baum [1989b; 1989a] and are known as *pseudohyperbolic spaces*. For the sake of completeness and since we need an additional argument for our purpose, we recall the result of those references and give a sketch of the proof:

Proposition 4.6. *Let $(\widetilde{M}^{n+1}, g)$ be a complete Riemannian spin manifold admitting a nonzero $(i\varepsilon/2)$-Killing spinor for some $\varepsilon \in \{\pm 1\}$. Then $(\widetilde{M}^{n+1}, g)$ is isometric to either the real hyperbolic space of constant sectional curvature -1 or to the warped product $(\mathbb{R} \times N, dt^2 \oplus e^{2t} g_N)$, where (N^n, g_N) is a complete nonflat Riemannian spin manifold carrying at least one nonzero parallel spinor. In the latter case, if n is odd, denote by $\mathcal{K}_0(N, g_N)$ the space of parallel spinors on (N^n, g_N) for the induced metric and spin structure, and if n is even, denote by $\mathcal{K}_0^\varepsilon(N, g_N)$ its projection onto the half-spinors bundle $\Sigma_\varepsilon N$. Then, the map*

$$\begin{cases} \mathcal{K}_0^\varepsilon(N, g_N) \to \left\{\frac{i\varepsilon}{2}\text{-Killing spinors on } \widetilde{M}\right\}, & \varphi \mapsto e^{\frac{t}{2}} \varphi \qquad \text{if } n \text{ is even}, \\ \mathcal{K}_0(N, g_N) \to \left\{\frac{i\varepsilon}{2}\text{-Killing spinors on } \widetilde{M}\right\}, & \varphi \mapsto e^{\frac{t}{2}}\left(\varphi \oplus \varepsilon i \frac{\partial}{\partial t} \cdot \varphi\right) \quad \text{if } n \text{ is odd}, \end{cases}$$

is a well-defined monomorphism. Moreover, if N is compact, then this is actually an isomorphism.

Proof. Let φ be a nonzero $(i\varepsilon/2)$-Killing spinor on the manifold $(\widetilde{M}^{n+1}, g)$. As Baum showed (see [1989b] and references therein), if $(\widetilde{M}, g)$ is *not* isometric to the hyperbolic space, then there must exist a unit smooth vector field ξ on $\widetilde{M}$ with $i\xi \cdot \varphi = \varepsilon\varphi$ on $\widetilde{M}$. From this relationship, the foliated structure of $\widetilde{M}$ can be deduced as follows. First note that $\xi = (\varepsilon V)/|V|$, where $g(V, X) := i\langle X \cdot \varphi, \varphi\rangle$ for all $X \in \Gamma(T\widetilde{M})$, and, in particular, $V = \varepsilon \nabla |\varphi|^2$ has no zeros on $\widetilde{M}$. Since $\widehat{\nabla}_X V = \varepsilon |\varphi|^2 X$ (that is V is a closed conformal vector field on $\widetilde{M}$), one deduces that $\widehat{\nabla}_X \xi = X - g(X, \xi)\xi$ for all $X \in \Gamma(T\widetilde{M})$, and as a consequence the flow of ξ — which is well-defined and complete since $(\widetilde{M}, g)$ is complete — preserves the level hypersurfaces of $|\varphi|^2 = |V|$. On the other hand, the second fundamental form of each such hypersurface with respect to ξ is $-\,\mathrm{Id}$, the Lie derivative of the metric in the direction of ξ is given by $\mathcal{L}_\xi g = 2g|_{\xi^\perp \times \xi^\perp}$, and hence, setting

$$N := \{x \in \widetilde{M} : |\varphi|^2(x) = 1\} \subset \widetilde{M},$$

the flow of ξ provides a diffeomorphism $\mathbb{R} \times N \to \widetilde{M}$ identifying ξ with $\frac{\partial}{\partial t}$ and pulling back the metric g onto $dt^2 \oplus e^{2t} g_N$, where g_N is the metric induced from g onto N. This done, the spin Gauss formula (9) implies that, for any $X \in \Gamma(TN)$,

$$\frac{i\varepsilon}{2} X \cdot \varphi = \widehat{\nabla}_X \varphi = \nabla_X^{\Sigma N} \varphi - \frac{1}{2} X \cdot \xi \cdot \varphi = \nabla_X^{\Sigma N} \varphi + \frac{i\varepsilon}{2} X \cdot \varphi,$$

from which $\nabla^{\Sigma N} \varphi|_N = 0$ follows: the restriction of φ onto any level hypersurface of $|\varphi|^2$ is a parallel spinor. Here $\nabla^{\Sigma N}$ stands for the spin Levi–Civita connection on $\Sigma := \Sigma\widetilde{M}|_N$. When n is even, the condition $i\xi \cdot \varphi = \varepsilon\varphi$ actually implies that $\varphi \in \Gamma(\Sigma_\varepsilon N)$ since $i\xi\,\cdot$ coincides with the Clifford action of the complex volume form of (N, g_N). When n is odd, the spinor $\varphi|_N$ can be rewritten in the form $\varphi|_N = \varphi_0 \oplus \varepsilon i \frac{\partial}{\partial t} \cdot \varphi_0$, where $\varphi_0 \in \Gamma(\Sigma N)$ is parallel. The dependence on t of φ is easily computed thanks to

$$\frac{\partial\varphi}{\partial t} = \widehat{\nabla}_{\frac{\partial}{\partial t}}\varphi = \tfrac{i\varepsilon}{2}\frac{\partial}{\partial t}\cdot\varphi = \tfrac{1}{2}\varphi,$$

from which $\varphi(t,\cdot) = e^{\frac{t}{2}}\varphi(0,\cdot)$ follows. This gives the formulas for the above map, which is obviously a right inverse to the "restriction" map

$$\begin{cases} \big\{\tfrac{i\varepsilon}{2}\text{-Killing spinors on } \widetilde{M}\big\} \to \mathcal{K}_0^{\varepsilon}(N, g_N), & \varphi \mapsto \varphi|_{\{0\}\times N} \quad \text{if } n \text{ is even,} \\ \big\{\tfrac{i\varepsilon}{2}\text{-Killing spinors on } \widetilde{M}\big\} \to \mathcal{K}_0(N, g_N), & \varphi \mapsto \varphi_+|_{\{0\}\times N} \quad \text{if } n \text{ is odd.} \end{cases}$$

In case N is compact, this restriction map is surjective, a remark which is missing in [Baum 1989a]. To establish this, let ψ be any further nonzero $(i\varepsilon/2)$-Killing spinor on $(\widetilde{M}^{n+1}, g)$. Again, ψ splits $(\widetilde{M}^{n+1}, g)$ as a warped product $(\mathbb{R}\times P^n, ds^2\oplus e^{2s}g_\Sigma)$, where (P^n, g_P) is complete, spin, and carries a nonzero parallel spinor. Now, using [Montiel 1999], the latter splitting must "coincide" (in a sense that is made precise below) with the former. Namely, for all $t\in\mathbb{R}$, the hypersurface $\{t\}\times N$ is a totally umbilical compact hypersurface of $\widetilde{M}$ with constant mean curvature. Therefore, by applying [op. cit., Lemma 4] to the foliation of $\widetilde{M}$ induced by ψ (whose leaves are not assumed to be compact), we easily conclude that for each $t\in\mathbb{R}$, there exists an $s\in\mathbb{R}$ such that $\{t\}\times N = \{s\}\times P$; in particular, P itself must be compact. The same argument shows that, for each $s\in\mathbb{R}$, there exists a $t\in\mathbb{R}$ with $\{s\}\times P = \{t\}\times N$. This yields that, if $\Phi : \mathbb{R}\times P \to \mathbb{R}\times N$, $(s,x)\mapsto(\phi_1(s,x), \phi_N(s,x))$, is the isometry induced by both splittings, then the component map ϕ_1 already only depends on s. By $\Phi^*(dt^2\oplus e^{2t}g_N) = ds^2\oplus e^{2s}g_P$ and the existence of an inverse map for Φ of a similar form, one deduces on the one hand that $\frac{\partial\phi_N}{\partial s}(s,x) = 0$ and hence $(\phi_1'(s))^2 = 1$ for all $s\in\mathbb{R}$, and on the other hand that $e^{2s}g_P = e^{2\phi_1(s)}(\phi_N)^*g_N$ holds for all $s\in\mathbb{R}$. This in turn implies the existence of an $s_0\in\mathbb{R}$ with $\phi_1(s) = s - s_0$ and $g_P = e^{-2s_0}(\phi_N)^*g_N$. Thus, up to homotheties on the metrics g_P and g_N, the Riemannian manifolds (P, g_P) and (N, g_N) are isometric and, up to translations in s, the splittings $\mathbb{R}\times P$ and $\mathbb{R}\times N$ coincide. By the first part of the proof, ψ must come from a parallel spinor on N and hence lie in the image of the map of Proposition 4.6. This concludes the proof. □

From the previous result, we deduce a characterization of hypersurfaces for which inequality (17) is an equality when $\widetilde{M}$ is a pseudohyperbolic space in several situations. In fact, as we will see, we are left with the case n is even, the manifold (N^n, g_N) has only positive (or only negative) nonzero parallel spinors, and M is only immersed in $\widetilde{M}$. Indeed:

Corollary 4.7. *Let $(\widetilde{M}^{n+1}, g) := (\mathbb{R}\times N, dt^2\oplus e^{2t}g_N)$, where (N^n, g_N) is a closed nonflat Riemannian spin manifold endowed with at least one nonzero parallel spinor and assume that $\widetilde{M}$ carries the induced spin structure (in particular, $(\widetilde{M}, g)$ admits an imaginary Killing spinor for at least one of the constants $\pm\frac{i}{2}$). Let $M^n \hookrightarrow \widetilde{M}$*

be any immersed closed orientable hypersurface carrying the induced metric and spin structure and suppose that one of the following supplementary conditions is fulfilled:

(a) *n is odd.*

(b) *n is even and (N^n, g_N) has nonzero positive and negative parallel spinors.*

(c) *n is even and M^n bounds a domain in $\widetilde{M}$.*

Then, M satisfies the equality case in (17) (and so in (7)) if and only if $M = \{t\} \times N$ for some $t \in \mathbb{R}$.

Proof. From Proposition 3.5, if $M^n \hookrightarrow \widetilde{M}^{n+1}$ satisfies the equality case in (17), then its mean curvature H must be constant. If either (a) or (b) is fulfilled, then by Proposition 4.6, the manifold $(\widetilde{M}, g)$ admits nonzero imaginary Killing spinors for *both* constants $\pm\frac{i}{2}$, therefore Proposition 3.5 implies that M is totally umbilical which, combined with [Montiel 1999, Lemma 4], yields $M = \{t\} \times N$ for some $t \in \mathbb{R}$. If (c) is fulfilled, this time [Montiel 1999, Theorem 10] applies and yields again $M = \{t\} \times N$ for some $t \in \mathbb{R}$. This shows the "only if" part of the corollary. The "if" part is easy to see since $\lambda_1(D_M) = 0$ because of parallel spinors on N, and on the other hand $|H| = 1$ by the explicit form of the metric. □

Acknowledgments

Part of this work was carried out during the Research in Pairs no. 998 at the Centre International de Rencontres Mathématiques (Luminy, France), which Ginoux and Habib would like to thank for their generous support and friendly welcome. Discussions at a latter stage of the paper were made possible thanks to a travel grant awarded to Ginoux by the German Academic Exchange Service (DAAD) and the hospitality of the Lebanese University at Beirut, whom Ginoux would like to thank.

References

[Ammann and Jammes 2012] B. Ammann and P. Jammes, "The supremum of first eigenvalues of conformally covariant operators in a conformal class", pp. 1–23 in *Variational problems in differential geometry*, edited by R. Bielawski et al., London Math. Soc. Lecture Note Ser. **394**, Cambridge Univ. Press, 2012. MR 2882767 Zbl 1243.53070

[Bär 1998] C. Bär, "Extrinsic bounds for eigenvalues of the Dirac operator", *Ann. Global Anal. Geom.* **16**:6 (1998), 573–596. MR 99k:58183 Zbl 0921.58065

[Bär et al. 2005] C. Bär, P. Gauduchon, and A. Moroianu, "Generalized cylinders in semi–Riemannian and Spin geometry", *Math. Z.* **249**:3 (2005), 545–580. MR 2006g:53065 Zbl 1068.53030

[Baum 1989a] H. Baum, "Complete Riemannian manifolds with imaginary Killing spinors", *Ann. Global Anal. Geom.* **7**:3 (1989), 205–226. MR 91k:58130 Zbl 0694.53043

[Baum 1989b] H. Baum, "Odd-dimensional Riemannian manifolds with imaginary Killing spinors", *Ann. Global Anal. Geom.* **7**:2 (1989), 141–153. MR 91b:53055 Zbl 0708.53039

[Baum et al. 1991] H. Baum, T. Friedrich, R. Grunewald, and I. Kath, *Twistors and Killing spinors on Riemannian manifolds*, Teubner-Texte zur Mathematik **124**, Teubner, Stuttgart, 1991. MR 94a:53077 Zbl 0734.53003

[Bourguignon and Gauduchon 1992] J.-P. Bourguignon and P. Gauduchon, "Spineurs, opérateurs de Dirac et variations de métriques", *Comm. Math. Phys.* **144**:3 (1992), 581–599. MR 93h:58164 Zbl 0755.53009

[El Soufi and Ilias 1992] A. El Soufi and S. Ilias, "Une inégalité du type "Reilly" pour les sous-variétés de l'espace hyperbolique", *Comment. Math. Helv.* **67**:2 (1992), 167–181. MR 93i:53059 Zbl 0758.53029

[Friedrich 2000] T. Friedrich, *Dirac operators in Riemannian geometry*, Graduate Studies in Mathematics **25**, Amer. Math. Soc., Providence, RI, 2000. MR 2001c:58017 Zbl 0949.58032

[Ginoux 2002] N. Ginoux, "Reilly-type spinorial inequalities", *Math. Z.* **241**:3 (2002), 513–525. MR 2004a:58037 Zbl 1033.58025

[Ginoux 2003a] N. Ginoux, "Remarques sur le spectre de l'opérateur de Dirac", *C. R. Math. Acad. Sci. Paris* **337**:1 (2003), 53–56. MR 2004h:58046 Zbl 1054.58021

[Ginoux 2003b] N. Ginoux, "Une nouvelle estimation extrinsèque du spectre de l'opérateur de Dirac", *C. R. Math. Acad. Sci. Paris* **336**:10 (2003), 829–832. MR 2004h:58045 Zbl 1054.58020

[Ginoux 2008] N. Ginoux, "The spectrum of the Dirac operator on $\mathrm{SU}_2/\mathrm{Q}_8$", *Manuscripta Math.* **125**:3 (2008), 383–409. MR 2008m:53116 Zbl 1143.53044

[Ginoux 2009] N. Ginoux, *The Dirac spectrum*, Lecture Notes in Mathematics **1976**, Springer, Berlin, 2009. MR 2010a:58039 Zbl 1186.58020

[Heintze 1988] E. Heintze, "Extrinsic upper bounds for λ_1", *Math. Ann.* **280**:3 (1988), 389–402. MR 89f:53091 Zbl 0628.53044

[Hijazi 1986] O. Hijazi, "A conformal lower bound for the smallest eigenvalue of the Dirac operator and Killing spinors", *Comm. Math. Phys.* **104**:1 (1986), 151–162. MR 87j:58096 Zbl 0593.58040

[Hijazi and Montiel 2013] O. Hijazi and S. Montiel, "A spinorial characterization of hyperspheres", *Calc. Var. Partial Differential Equations* **48**:3-4 (2013), 527–544. MR 3116021 Zbl 1286.53057

[Hitchin 1974] N. Hitchin, "Harmonic spinors", *Advances in Math.* **14** (1974), 1–55. MR 50 #11332 Zbl 0284.58016

[Kato 1995] T. Kato, *Perturbation theory for linear operators*, Classics in Mathematics, Springer, Berlin, 1995. MR 96a:47025 Zbl 0836.47009

[Kühnel and Rademacher 1998] W. Kühnel and H.-B. Rademacher, "Asymptotically Euclidean manifolds and twistor spinors", *Comm. Math. Phys.* **196**:1 (1998), 67–76. MR 99h:53060 Zbl 0929.53023

[Lawson and Michelsohn 1989] H. B. Lawson, Jr. and M.-L. Michelsohn, *Spin geometry*, Princeton Mathematical Series **38**, Princeton Univ. Press, 1989. MR 91g:53001 Zbl 0688.57001

[Montiel 1999] S. Montiel, "Unicity of constant mean curvature hypersurfaces in some Riemannian manifolds", *Indiana Univ. Math. J.* **48**:2 (1999), 711–748. MR 2001f:53131 Zbl 0973.53048

[Reilly 1977] R. C. Reilly, "On the first eigenvalue of the Laplacian for compact submanifolds of Euclidean space", *Comment. Math. Helv.* **52**:4 (1977), 525–533. MR 58 #2657 Zbl 0382.53038

[Setti 1998] A. G. Setti, "Eigenvalue estimates for the weighted Laplacian on a Riemannian manifold", *Rend. Sem. Mat. Univ. Padova* **100** (1998), 27–55. MR 2000a:58082 Zbl 0922.58084

Received March 19, 2014.

NICOLAS GINOUX
DÉPARTEMENT INFORMATIQUE
IUT DE METZ
UNIVERSITÉ DE LORRAINE
ILE DU SAULCY
CS10628
57045 METZ
FRANCE
nicolas.ginoux@univ-lorraine.fr

GEORGES HABIB
FACULTY OF SCIENCES II
DEPARTMENT OF MATHEMATICS
LEBANESE UNIVERSITY
P.O. BOX 90656
FANAR-MATN
LEBANON
ghabib@ul.edu.lb

SIMON RAULOT
LABORATOIRE DE MATHÉMATIQUES R. SALEM UMR 6085 CNRS
UNIVERSITÉ DE ROUEN AVENUE DE L'UNIVERSITÉ
BP.12 TECHNOPÔLE DU MADRILLET
76801 SAINT-ÉTIENNE-DU-ROUVRAY
FRANCE
simon.raulot@univ-rouen.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.278.103

GAMES AND ELEMENTARY EQUIVALENCE OF II_1-FACTORS

ISAAC GOLDBRING AND THOMAS SINCLAIR

We use Ehrenfeucht–Fraïssé games to give a local geometric criterion for elementary equivalence of II_1-factors. We obtain as a corollary that two II_1-factors are elementarily equivalent if and only their unitary groups are elementarily equivalent as $\mathbb{Z}_4$-metric spaces.

Introduction

While most mathematicians are concerned with determining when two objects in their field are isomorphic, logicians tend to be concerned with the coarser notion of *elementary equivalence*. Two (classical) structures M and N are said to be elementarily equivalent if and only if for any first-order sentence σ (in the language appropriate to the study of M and N), we have σ is true in M if and only if σ is true in N. For structures appearing in analysis, a continuous logic is used in which sentences can now take a continuum of "truth" values; the appropriate notion of elementary equivalence is that the truth values of all sentences are the same in both structures.

The model-theoretic study of tracial von Neumann algebras began in earnest in [Farah et al. 2013; 2014a; 2014b]. At the moment, there are only three distinct elementary equivalence classes of II_1-factors known. (This should not be so surprising as it took a while for many isomorphism classes of II_1-factors to be discovered and elementary equivalence is a much coarser notion.) Indeed, it was observed in [Farah et al. 2014b] that Property (Γ) and the property of being McDuff are both elementary properties (for separable II_1-factors). Thus, if we let M_{DL} be a separable II_1-factor that has Property (Γ) but is not McDuff (see [Dixmier and Lance 1969]), then M_{DL}, the hyperfinite II_1-factor $\mathcal{R}$ and the free group factor $L(\mathbb{F}_2)$ are mutually nonelementarily equivalent. Amongst those studying II_1-factors from a model-theoretic point of view, it is widely agreed that there should be more than three elementary equivalence classes of II_1-factors; in fact, there should probably be continuum many elementary equivalence classes. At the moment, we cannot even answer the question: is $\mathcal{R} \otimes L(\mathbb{F}_2)$ elementarily equivalent to $\mathcal{R}$? In

Goldbring's work was partially supported by NSF grant DMS-1007144. Sinclair was supported by an NSF RTG Assistant Adjunct Professorship.

MSC2010: primary 46L10; secondary 03C20.

Keywords: II_1-factors, model theory, Ehrenfeucht–Fraïssé games.

order to accomplish these goals, we need more tools for understanding elementary equivalence of II_1-factors.

Ehrenfeucht–Fraïssé games have long been a tool in model theory for establishing that structures are elementarily equivalent. In [Heinrich and Henson 1986], the authors exhibit an Ehrenfeucht–Fraïssé-type game used to establish elementary equivalence for Banach spaces. In this note, we adapt the game from [loc. cit.] and combine it with an argument of Kirchberg [1993] in order to characterize elementary equivalence for II_1-factors belonging to the class $\mathcal{K}_{op}$ (to be defined below). We should note that, currently, we do not know of a II_1-factor that does not belong to the class $\mathcal{K}_{op}$ and the existence of such a factor would already lead to two new theories of II_1-factors!

Recall Dye's theorem [1955], which states that any two factors not of type I_{2^n} (e.g., any two II_1-factors) are isomorphic if and only if their unitary groups are isomorphic (even as discrete groups). Combining Dye's theorem with the Keisler–Shelah theorem (which states that two structures are elementarily equivalent if and only if they have isomorphic ultrapowers) and the fact that the functors of taking ultrapowers and taking unitary groups commute, we see that two II_1-factors are elementarily equivalent if and only if their unitary groups are elementarily equivalent as metric groups (with respect to the ℓ_2 metric). Using the aforementioned Ehrenfeucht–Fraïssé games and some further arguments, our main result is that we can improve upon the previous sentence, essentially removing the group structure:

Theorem 0.1. *Suppose that M and N are II_1-factors belonging to the class $\mathcal{K}_{op}$. Then M and N are elementarily equivalent if and only if $U(M)$ and $U(N)$ are elementarily equivalent as $\mathbb{Z}_4$-metric* spaces.

Here, by a $\mathbb{Z}_4$-metric space, we mean a metric space X equipped with an action of $\mathbb{Z}_4$ on X by isometries. Unitary groups of von Neumann algebras will always be considered as $\mathbb{Z}_4$-metric spaces by having the generator of $\mathbb{Z}_4$ act by multiplication by i.

In this paper, we assume that the reader is familiar with some basic model theory and von Neumann algebra theory. Good references for continuous model theory are [Ben Yaacov et al. 2008] and [Farah et al. 2014a]; the latter is geared towards the model-theoretic study of operator algebras.

All normed spaces are assumed to be over the complex numbers, $\mathbb{C}$. For a normed space X, we denote the closed unit ball by $(X)_1 := \{x \in X : \|x\| \le 1\}$.

For the reader's convenience, we now recall the original notion of Ehrenfeucht–Fraïssé games in the context of continuous logic. This has not appeared in the literature but has appeared in some online lecture notes of Bradd Hart [2012]. Fix an arbitrary language $\mathcal{L}$ and atomic formulae $\varphi_1(\vec{x}), \dots, \varphi_k(\vec{x})$ in the variables $\vec{x} = (x_1, \dots, x_n)$ and $\varepsilon > 0$. The Ehrenfeucht–Fraïssé game $\mathfrak{G}(\varphi_1, \dots, \varphi_k, \varepsilon)$ is played with $\mathcal{L}$-structures M and N as follows: First Player I chooses $a_1 \in M$ or

$b_1 \in N$ respecting the sort of x_1. Player II chooses $b_2 \in N$ or $a_2 \in M$ respectively. The players alternate in this manner until they have produced sequences $a_1, \ldots, a_n \in M$ and $b_1, \ldots, b_n \in N$. Player II then wins the game if and only if for each $i = 1, \ldots, k$, we have $|\varphi_i(\vec{a})^M - \varphi_i(\vec{b})^N| \leq \varepsilon$. It is then a theorem that $M \equiv N$ if and only Player II has a winning strategy in each $\mathfrak{G}(\varphi_1, \ldots, \varphi_k, \varepsilon)$.

1. The class $\mathcal{K}_{op}$

Given a C* algebra A, recall that its opposite algebra A^{op} is the algebra obtained from A by multiplying elements in the opposite order; that is, for $a, b \in A$, we have $a \cdot_{op} b := b \cdot a$. It is immediate that A^{op} is once again a C* algebra. Furthermore, if A is a von Neumann algebra, then A^{op} is also a von Neumann algebra. Note also that if $(A_i : i \in I)$ is a family of C* algebras (resp. tracial von Neumann algebras) and $\mathcal{U}$ is an ultrafilter on I, then $\left(\prod_{\mathcal{U}} A_i\right)^{op} \cong \prod_{\mathcal{U}} A_i^{op}$ via the identity map, where the ultraproduct is understood to be the usual C* algebra ultraproduct (resp. tracial ultraproduct).

Many of the naturally occurring tracial von Neumann algebras are isomorphic to their opposites, e.g., $\mathcal{R}$ and $L(G)$ (G any group). There are examples of tracial von Neumann algebras that are not isomorphic to their opposites (see [Connes 1975]). During a seminar talk given by the first author at Vanderbilt University, Jesse Peterson asked whether or not the class of all tracial von Neumann algebras isomorphic to their opposites is an axiomatizable class. While we do not know the answer to this question (although we suspect the answer is negative), the answer is positive if one replaces the word "isomorphism" by "elementary equivalence" as we show in the following:

Proposition 1.1. *The class of all tracial von Neumann algebras that are elementarily equivalent to their opposites is an elementary class.*

Definition 1.2. We let $\mathcal{K}_{op}$ denote the class of all tracial von Neumann algebras elementarily equivalent to their opposites.

Proof of Proposition 1.1. We present a proof suggested to us by Todor Tsankov as well as independently by the anonymous referee. There is a collection of axioms for the class $\mathcal{K}_{op}$: for every term t, recursively define the term t^{op} by defining $(t_1 \cdot t_2)^{op} := t_2^{op} \cdot t_1^{op}$. Then one can recursively define, for any formula φ, the formula φ^{op}, the key clause being the atomic formulae, where one replaces every occurrence of a term t by the term t^{op}. Then the conditions $|\sigma - \sigma^{op}| = 0$, as σ ranges over all sentences, axiomatizes the class $\mathcal{K}_{op}$. □

We remark in passing that alternately by [Ben Yaacov et al. 2008, Proposition 5.14], it suffices to show that $\mathcal{K}_{op}$ is closed under isomorphisms, ultraproducts, and ultraroots. We leave it as an exercise to verify these properties for $\mathcal{K}_{op}$.

Since $\mathcal{R}$ and $L(\mathbb{F}_2)$ are isomorphic to their opposites, they belong to $\mathcal{K}_{op}$. Moreover, the example M_{DL} of a II_1-factor with Property (Γ) that is not McDuff given by Lance and Dixmier [1969] is also isomorphic to its opposite. Thus, we have this:

Corollary 1.3. *If there is a II_1-factor that does not belong to $\mathcal{K}_{op}$, then there are at least five theories of II_1-factors.*

Proof. If N is a II_1-factor that does not belong to $\mathcal{K}_{op}$, then the theories of N and N^{op} differ from each other and from the three known theories of II_1-factors. □

Question 1.4. Are there more "explicit" axioms for the class $\mathcal{K}_{op}$? Can one use typical model-theoretic preservation theorems to show that $\mathcal{K}_{op}$ is universally axiomatizable or $\forall\exists$-axiomatizable?

Question 1.5. Is there a single sentence σ such that adding the condition "$\sigma = 0$" to the axioms for II_1-factors gives an axiomatization of $\mathcal{K}_{op}$?

A negative answer to the last question implies that there must be infinitely many elementary equivalence classes of II_1-factors not belonging $\mathcal{K}_{op}$. Indeed, if there are only finitely many elementary equivalence classes of II_1-factors not belonging to $\mathcal{K}_{op}$, then the class of II_1-factors not belonging to $\mathcal{K}_{op}$ is readily verified to be elementary as well, whence a typical compactness argument is used to show that the last question has a positive answer.

2. Model theory of Banach pairs

In order to frame the main results of the paper in the next section on the model theory of II_1-factors, we introduce a class of linear (unbounded) metric structures ("Banach pairs") for which II_1-factors will be the primary set of examples. The important fact which we will see is that the theory of a II_1-factor regarded as a Banach pair will determine its theory as II_1-factor. For this reason we feel it is justified to introduce this treatment, despite several existing approaches in the literature for dealing with linear metric structures, e.g., [Ben Yaacov 2008; Ben Yaacov et al. 2008; Henson and Moore 1983], with at least one treatment [Farah et al. 2014a] being devoted to C*-algebras and tracial von Neumann algebras.

Definition 2.1. A *Banach pair* $(X, \mathcal{C})$ consist of a normed space X and a distinguished subset $\mathcal{C} \subset (X)_1$ which is

- complete;
- roundly convex, i.e., $\lambda x + \mu y \in \mathcal{C}$ for all $x, y \in \mathcal{C}$ and $\lambda, \mu \in \mathbb{C}$ with $|\lambda| + |\mu| \leq 1$;
- generating, i.e., $\bigcup_n n \cdot \mathcal{C} = X$.

The main examples of Banach pairs we will be interested in are where $X = M$, a tracial von Neumann algebra equipped with the 2-norm $\|x\|_2 := \mathrm{tr}(x^*x)^{1/2}$, and $\mathcal{C} = (M)_1$, the (norm) closed unit ball.

A Banach pair $(X, \mathcal{C})$ can be interpreted as a structure for the language $\mathcal{L}_{\mathrm{BP}}$ below:

- There is one sort each for $\mathbb{C}$ and X.
- There is a sequence of domains of quantification $\mathcal{C}_n$ for X.
- There are function symbols $\iota_{m,n} : \mathcal{C}_m \to \mathcal{C}_n$ for $m \le n$ to be interpreted as the usual inclusion maps.
- X is given the usual complex normed space axioms.
- There are axioms which show $0_X \in \mathcal{C}_1 \subset (X)_1$.
- There are axioms to show each $\mathcal{C}_n$ is roundly convex.

For a Banach pair $(X, \mathcal{C})$ and $x \in X$, we define $\|x\|_{\mathcal{C}} := \inf\{t > 0 : x \in t \cdot \mathcal{C}\}$, which can be checked to be a Banach norm on X. However, note that $\|\cdot\|_{\mathcal{C}}$ is a definable predicate if and only if it is uniformly continuous with respect to the usual norm. (In the case that X is a tracial von Neumann algebra, this will be the case if and only if X is finite-dimensional.)

As an $\mathcal{L}_{\mathrm{BP}}$-structure, the ultrapower $(X, \mathcal{C})^{\mathcal{U}}$ can be identified with the Banach pair $(X^{\mathcal{U}}, \mathcal{C}^{\mathcal{U}})$, where $X^{\mathcal{U}}$ is the quotient space of $\{(x_i) : \lim_{\mathcal{U}}\|x_i\|_{\mathcal{C}} < \infty\}$ modulo the subspace $\{(z_i) : \lim_{\mathcal{U}}\|z_i\|_{\mathcal{C}} < \infty,\ \lim_{\mathcal{U}}\|z_i\| = 0\}$ and $\mathcal{C}^{\mathcal{U}} \subset X^{\mathcal{U}}$ is defined in the obvious way.

We say that two Banach pairs $(X, \mathcal{C})$ and $(Y, \mathcal{D})$ are isomorphic (written $(X, \mathcal{C}) \cong (X, \mathcal{D})$) if they are isomorphic as $\mathcal{L}_{\mathrm{BP}}$-structures, that is, if there is an isometry $T : X \to Y$ so that $T(\mathcal{C}) = \mathcal{D}$. By definition, the aforementioned Banach pairs are elementarily equivalent (written $(X, \mathcal{C}) \equiv (Y, \mathcal{D})$) if $\mathrm{Th}(X, \mathcal{C}) = \mathrm{Th}(Y, \mathcal{D})$. As a consequence of the Keisler–Shelah theorem in continuous logic, we have that $(X, \mathcal{C}) \equiv (Y, \mathcal{D})$ if and only if there is an ultrafilter so that $(X, \mathcal{C})^{\mathcal{U}} \cong (Y, \mathcal{D})^{\mathcal{U}}$. See [Henson and Iovino 2002, §10] for a proof of this fact in the context of normed spaces or [Heinrich and Henson 1986, §3] for a more explicit construction for Banach spaces.

Our main observation in this section is that for Banach pairs $(X, \mathcal{C})$ and $(Y, \mathcal{D})$ elementary equivalence can be characterized in terms of the pairs "having the same local geometric structure" by the use of Ehrenfeucht–Fraïssé games. For the very similar case of Banach spaces, this was done by Heinrich and Henson [1986, Theorem 4] and the case of normed spaces is largely similar (see [Henson and Iovino 2002, Remark 10.10]).

We now describe precisely what we mean when we say that two Banach pairs $(X, \mathcal{C})$ and $(Y, \mathcal{D})$ have the same local geometric structure. For E a subspace of X and F a subspace of Y, we say that a linear bijection $T : E \to F$ is an *ε-almost isometry* if $\|T\|, \|T^{-1}\| \le 1+\varepsilon$ and $T(E \cap \mathcal{C}) \subset_{\varepsilon} F \cap \mathcal{D}$ and $T^{-1}(F \cap \mathcal{D}) \subset_{\varepsilon} E \cap \mathcal{C}$. (We write $A \subset_{\varepsilon} B$ if $\sup_{x\in A} \inf_{y\in B} \|x - y\| \le \varepsilon$.)

The following is adapted from [Heinrich and Henson 1986, §2]; see also [Henson and Moore 1983, §8]. We describe a game $\mathfrak{G}(n, \varepsilon)$ played by two players with Banach pairs $(X, \mathcal{C})$ and $(Y, \mathcal{D})$, where $\varepsilon > 0$ and n are fixed parameters.

Step 1. Player I chooses a one-dimensional subspace, either $E_1 \subset X$ or $F_1 \subset Y$. Player II then chooses a subspace, respectively $F_1 \subset Y$ or $E_1 \subset X$ and a linear bijection $T_1 : E_1 \to F_1$.

Step i. Player I chooses an at most one-dimensional extension, either $E_i \supset E_{i-1}$ or $F_i \supset F_{i-1}$. Player II then chooses a subspace, respectively $F_i \subset Y$ or $E_i \subset X$, and a linear bijection $T_i : E_i \to F_i$ which extends T_{i-1}.

Step n. The players make their choices, and the game terminates. Player II wins if $T_n : E_n \to F_n$ is an ε-almost isometry; otherwise, Player I wins.

During the course of proofs, we may speak of Player I playing $x_i \in X$, in which case we mean that Player I plays $\operatorname{span}(E_{i-1} \cup \{x_i\})$. We may then also say that Player II responds with $y_i \in Y$, in which case we mean that Player II plays the linear bijection T_i extending T_{i-1} that sends x_i to y_i.

Definition 2.2. We say that Banach pairs $(X, \mathcal{C})$ and $(Y, \mathcal{D})$ are *locally equivalent* (written $(X, \mathcal{C}) \cong_{\text{loc}} (Y, \mathcal{D})$) if for every $\varepsilon > 0$ and every n, Player II has a winning strategy for the game $\mathfrak{G}(n, \varepsilon)$.

Remark 2.3. Since ε is arbitrary, and we need only deal with at most one-dimensional extensions, we see that local isomorphism remains the same under an alternate version of ε-almost isometry, namely, the existence of linear bijections $T : E \to F$, $S : F \to E$ with strict containment $T(E \cap \mathcal{C}) \subsetneq F \cap \mathcal{D}$ and $S(F \cap \mathcal{D}) \subsetneq E \cap \mathcal{C}$ so that $\|ST - \mathrm{id}_E\|$, $\|TS - \mathrm{id}_F\| < \varepsilon$ and $\|T\|$, $\|S\| < 1 + \varepsilon$.

Proposition 2.4. *The following statements are equivalent*:

(1) $(X, \mathcal{C}) \equiv (Y, \mathcal{D})$.

(2) *There exists an ultrafilter so that* $(X, \mathcal{C})^{\mathcal{U}} \cong (Y, \mathcal{D})^{\mathcal{U}}$ *as Banach pairs.*

(3) $(X, \mathcal{C}) \cong_{\text{loc}} (Y, \mathcal{D})$.

As noted above, (1) $\Longleftrightarrow$ (2) is the Keisler–Shelah theorem applied to the language of Banach pairs. The proof of (2) $\Rightarrow$ (3) is straightforward using representing sequences. Therefore we only need to prove (3) $\Rightarrow$ (1). The proof is more or less identical to the Banach space version as in [Heinrich and Henson 1986]. However, since we are working in a different logic, we sketch a (nearly complete) proof here for the convenience of the reader.

Sketch of (3) $\Rightarrow$ (1). First, we work with the notion of ε-almost isometry as described in Remark 2.3. Let σ be a sentence of the form $\inf_{v_1} \sup_{v_2} \cdots Q_{v_n} \rho(v_1, \ldots, v_n)$, where Q is inf if n is odd and sup if n is even and where ρ is quantifier-free. (We suppress mention of the sorts $\mathcal{C}_i$ corresponding to each v_i.) Fix $\varepsilon > 0$. It suffices to show that $\sigma^{(Y,\mathcal{D})} \le \sigma^{(X,\mathcal{C})} + \varepsilon$ for all $\varepsilon > 0$. Indeed, by symmetry of the relation of local equivalence, this shows that all sentences of the above form have the same truth values in $(X, \mathcal{C})$ and $(Y, \mathcal{D})$. Since any sentence in prenex normal form is equivalent to one of the above form (by adding dummy variables) and since the set of sentences in prenex normal form is dense in the set of all sentences (see [Ben Yaacov et al. 2008, §6]), we obtain that $(X, \mathcal{C}) \equiv (Y, \mathcal{D})$.

Fix a sufficiently small $\delta > 0$. (We will see exactly how small δ needs to be in a moment.) Fix a winning strategy $\mathcal{S}$ for Player II in $\mathfrak{G}(n, \delta)$. Call a play of the game $\mathfrak{G}(n, \delta)$ *regular* if

- for odd i, Player I plays $x_i \in X$, while for even i, Player I plays $y_i \in Y$;
- for each i, Player I's move at Round i is always in the sort corresponding to the variable v_i;
- Player II always plays according to $\mathcal{S}$.

We say that sequences $x_1, \dots, x_k \in X$ and $y_1, \dots, y_k \in Y$ *correspond* if they are the results of the first k rounds of a regular play of $\mathfrak{G}(n, \delta)$.

For $0 \le l \le n$, let $\sigma_l(v_1, \dots, v_{n-l})$ denote the formula obtained from σ by removing the first $n - l$ quantifiers. One now proves, by induction on l ($0 \le l \le n$), that if $x_1, \dots, x_{n-l} \in X$ and $y_1, \dots, y_{n-l} \in Y$ correspond, then

$$\sigma_l(y_1, \dots, y_{n-1})^{(Y,\mathcal{D})} \le \sigma_l(x_1, \dots, x_{n-1})^{(X,\mathcal{C})} + \varepsilon.$$

The base case $l = 0$ follows from the fact that $T_n : \mathrm{span}(x_1, \dots, x_n) \to \mathrm{span}(y_1, \dots, y_n)$ is a δ-almost isometry if δ is chosen sufficiently small. We now prove the induction step. Suppose that the claim holds for l and that $x_1, \dots, x_{n-l-1} \in X$ and $y_1, \dots, y_{n-l-1} \in Y$ correspond. Let $r := \sigma_{l+1}(x_1, \dots, x_{n-l-1})^{(X,\mathcal{C})}$. First suppose that $n - l$ is odd, so that $\sigma_{l+1}(v_1, \dots, v_{n-l-1}) = \inf_{v_{n-l}} \sigma_l(v_1, \dots, v_{n-l})$. Fix $\eta > 0$ and let $x_{n_l} \in X$ be of the same sort as v_{n-l} so that $\sigma_l(x_1, \dots, x_{n-l})^{(X,\mathcal{C})} \le r + \eta$. Let $y_{n-l} \in Y$ be Player II's response to x_{n-l} according to the strategy $\mathcal{S}$. Then, by induction,

$$\sigma_l(y_1, \dots, y_{n-l})^{(Y,\mathcal{D})} \le \sigma_l(x_1, \dots, x_{n-l})^{(X,\mathcal{C})} + \varepsilon \le r + \varepsilon + \eta.$$

Letting η go to 0 yields the desired result. The case that $n - l$ is even is similar and is left to the reader. $\square$

3. Elementary equivalence of II_1-factors

We say that two tracial von Neumann algebras M and N are locally equivalent if the associated Banach pairs $(M, (M)_1)$ and $(N, (N)_1)$ are locally equivalent. Somewhat miraculously, it turns out that for II_1-factors belonging to $\mathcal{K}_{op}$, local equivalence is the same as elementary equivalence. This essentially follows from an argument of Kirchberg [1993]. First, we need to recall a fact about *Jordan morphisms* between von Neumann algebras.

Given a C* algebra A, the *special Jordan product* on A is the operation $\circ$ defined by $a \circ b := \frac{1}{2}(ab + ba)$ for all $a, b \in A$. If B is also a C* algebra, then a linear map $T : A \to B$ is a *Jordan morphism* if it preserves the special Jordan product and the involution. We need the following:

Fact 3.1 (See [Hanche-Olsen and Størmer 1984, Corollary 7.4.9]). If M and N are von Neumann algebras and $T : M \to N$ is a normal Jordan homomorphism, then T is the sum of a $*$-homomorphism and a $*$-antihomomorphism.

Recall that a map $A \to B$ between C* algebras is a $*$-antihomomorphism if and only if it is a $*$-homomorphism $A \to B^{op}$.

Suppose that M and N are von Neumann algebras and $T : M \to N$ is a unital, bijective, normal Jordan homomorphism. Write $T = T_1 + T_2$, where $T_1 : M \to N$ and $T_2 : M \to N^{op}$ are $*$-homomorphisms. Since $T_i(1)$ is a projection for $i = 1, 2$ and $T_1(1) + T_2(1) = 1$, we have that $T_1(1)$ and $T_2(1)$ are orthogonal projections. Since $T(M) = N$, it follows that each $T_i(1)$ is a central projection. Thus, if N is a factor, it follows that $\{T_1(1), T_2(1)\} = \{0, 1\}$, whence T is either an isomorphism or an anti-isomorphism.

The following is basically Proposition 4.6 in [Kirchberg 1993].

Proposition 3.2 (Kirchberg). *Suppose that M and N are II_1-factors. If there is an isometry $T : L^2(M, \mathrm{tr}_M) \to L^2(N, \mathrm{tr}_N)$ so that T maps M onto N contractively, then $M \cong N$ or $M \cong N^{op}$.*

Proof. We first show that T maps unitaries to unitaries. If $u \in M$ is a unitary, we have

$$1 = \|u\|_2^2 = \|T(u)\|_2^2 = \langle T(u), T(u)\rangle = \langle T(u)^* T(u), 1\rangle.$$

On the other hand,

$$\|T(u)^* T(u)\|_2 \leq \|T(u)\| \cdot \|T(u)\|_2 \leq 1.$$

It follows that $T(u)^* T(u) = 1$. We thus have that $T'(x) := T(1)^* T(x)$ is unital, contractive, trace-preserving, and takes unitaries to unitaries. By the same reasoning as in the proof of [Kirchberg 1993, Proposition 4.6], T' is a weakly continuous Jordan morphism and the result follows from the discussion preceding this proposition. □

Corollary 3.3. *Suppose that M and N are II_1-factors. Then M is locally equivalent to N if and only if M is elementarily equivalent to N or to N^{op}. In particular, if M and N are II_1-factors belonging to the class $\mathcal{K}_{op}$, then M is locally equivalent to N if and only if M is elementarily equivalent to N.*

Proof. By the downward Löwenheim–Skolem theorem (see [Farah et al. 2014a, Section 4.2]), we may suppose that M and N are separable. Suppose that M is locally equivalent to N. Then by Proposition 2.4, there is an isometry $L^2(M^{\mathcal{U}}) \to L^2(N^{\mathcal{U}})$ that maps $M^{\mathcal{U}}$ into $N^{\mathcal{U}}$ contractively. By Proposition 3.2, $M^{\mathcal{U}}$ is isomorphic to either $N^{\mathcal{U}}$ or $(N^{\mathcal{U}})^{op}$. It follows that M is elementarily equivalent to either N or N^{op}. The converse is trivial. □

We now introduce a more useful test for determining elementary equivalence which works in the more specific case of Banach pairs $(M, (M)_1)$, where M is a II_1-factor (or more generally a tracial von Neumann algebra) equipped with the 2-norm, and $(M)_1$ is the (operator norm) unit ball of M.

We define the game $\mathfrak{G}_{\mathrm{vN}}(n, \varepsilon)$ in parameters n and $\varepsilon > 0$ which is played by two players with II_1-factors M and N as follows.

Step i. Player I chooses a unitary, either $u_i \in U(M)$ or $v_i \in U(N)$. Player II then chooses a unitary, respectively $v_i \in U(N)$ or $u_i \in U(M)$, in the same manner.

Step n. The players make their choices, and the game terminates. Player II wins if $|\langle u_i, u_j\rangle - \langle v_i, v_j\rangle| < \varepsilon$ for all $1 \le i, j \le n$; otherwise, Player I wins.

Theorem 3.4. *The II_1-factors M and N are locally equivalent if and only if Player II has a winning strategy for the game $\mathfrak{G}_{\mathrm{vN}}(n, \varepsilon)$ for all parameters (n, ε).*

In order to prove this result we will first need one lemma.

Lemma 3.5. *Let M and N be II_1-factors, $E \subset M$ and $F \subset N$ be subspaces, and $T : (E, E \cap (M)_1) \to (F, F \cap (N)_1)$ be an ε-almost isometry. If $u \in E$ is a unitary, then there exists a unitary $v \in N$ so that $\|T(u) - v\|_2 \le 4\sqrt{\varepsilon}$.*

Proof. In a II_1-factor, a u is a unitary if and only if it is a contraction with $\|u\|_2 = 1$. By definition, we see that there exists a contraction $y \in N$ with $\|y - T(u)\|_2 \le \varepsilon$. In particular, $\|y\|_2 \ge 1 - 2\varepsilon$. By a standard estimate we have that

$$\big\|1-|y|\big\|_2^2 \le 1+\big\||y|\big\|_2^2 - 2\,\mathrm{tr}|y| = 1+\mathrm{tr}(|y|^2) - 2\,\mathrm{tr}|y| \le 1-\mathrm{tr}|y| \le 1-\|y\|_2 \le 2\varepsilon,$$

whence writing $y = v|y|$ for $v \in U(N)$ we have that $\|T(u) - v\|_2 \le 4\sqrt{\varepsilon}$. □

Proof of Theorem 3.4. First suppose that M and N are locally equivalent. Fix n and $\varepsilon > 0$; we describe a winning strategy for Player II in the game $\mathfrak{G}_{\mathrm{vN}}(n, \varepsilon)$. For simplicity, we suppose that $n = 2$ and describe a winning strategy for Player II; the general case is no more difficult, only the notation is more cumbersome. Fix δ sufficiently small (to be specified later) and fix a winning strategy $\mathcal{S}$ for Player II in the game $\mathfrak{G}(2, \delta)$. Suppose that Player I first plays $u_1 \in U(M)$. (The case that Player I first plays a unitary in N is similar.) Let $y_1 \in N$ be Player II's response to u_1 in the game $\mathfrak{G}(2, \delta)$ according to $\mathcal{S}$. Since $u_1 \mapsto y_1$ determines a δ-almost isometry, by Lemma 3.5, there is $v_1 \in U(N)$ such that $\|y_1 - v_1\|_2 \le 4\sqrt{\delta}$. Now suppose that Player II responds with $v_2 \in U(N)$. (The case that Player II responds with a unitary in M is similar.) Let $x_2 \in M$ be Player II's response to (u_1, y_1, v_2) in the game $\mathfrak{G}(2, \delta)$ according to $\mathcal{S}$. Since $u_1 \mapsto y_1, x_2 \mapsto v_2$ determines a δ-almost isometry, we once again have $u_2 \in U(M)$ such that $\|x_2 - u_2\|_2 \leq 4\sqrt{\delta}$.

We need to verify that $|\langle u_i, u_j\rangle - \langle v_i, v_j\rangle| < \varepsilon$ for $i, j = 1, 2$. If δ is chosen small enough so that a δ-almost isometry preserves inner products within an error of $\varepsilon/3$ (use, for example, the polarization identity) and such that perturbing entries of an inner product by a distance of no more than $4\sqrt{\delta}$ changes the inner product by an amount not exceeding $\varepsilon/3$, then the desired estimates hold. For example,

$$\langle u_1, u_2\rangle \sim_{\varepsilon/3} \langle u_1, x_2\rangle \sim_{\varepsilon/3} \langle y_1, v_2\rangle \sim_{\varepsilon/3} \langle v_1, v_2\rangle.$$

We now prove the converse. Suppose that Player II has a winning strategy in all of the games $\mathfrak{G}_{\mathrm{vN}}(n, \varepsilon)$; we show that M and N are elementarily equivalent as Banach pairs. By symmetry, it is enough to show that $\sigma^{(M,(M)_1)} \leq r$ implies that $\sigma^{(N,(N)_1)} \leq r$ for any positive real number r and any prenex normal form sentence σ. Since $\sigma \dot{-} r$ is equivalent to a prenex normal form sentence, it is enough to prove that $\sigma^{(M,(M)_1)} = 0$ implies $\sigma^{(N,(N)_1)} = 0$ for any prenex normal form sentence σ.

Towards this end, we introduce the "unitary transform" of a sentence in prenex normal form. Suppose that σ is a sentence in prenex normal form, say

$$\sigma = Q_1 x_1 \cdots Q_n x_n \varphi(\vec{x}),$$

where $\varphi(\vec{x})$ is quantifier-free. We form the new sentence σ^u as follows:

- If $Q_i = \inf$ and x_i is of sort n_i, replace each occurrence of the variable x_i by the term $t_i(u_i, v_i) := n_i \cdot ((u_i + v_i)/2)$, where u_i and v_i are variables of sort $\mathcal{C}_1$, and replace the quantifier $Q_i x_i$ by the quantifiers $Q_i u_i Q_i v_i$.
- The quantifier-free part of σ^u should now be

$$\max\bigl(\varphi, \max_i\bigl(\max(1 \dot{-} \|u_i\|_2, 1 \dot{-} \|v_i\|_2)\bigr)\bigr).$$

For example, if $\sigma = \sup_{x_1} \inf_{x_2} \varphi(x_1, x_2)$, where x_2 is of sort $\mathcal{C}_1$ (for simplicity), then $\sigma^u = \sup_{x_1} \inf_{u_2} \inf_{v_2} \varphi(x_1, (u_2 + v_2)/2)$.

Also, we let σ^{uu} be the "formula" defined in the exact same way as σ^u except that we only allow quantifiers over the unitary groups rather than the entire unit ball. (Formally, σ^{uu} is not a formula in the sense of continuous logic, but it will be useful in the remainder of the proof.)

Claim 1. We have $\sigma^{(M,(M)_1)} = 0$ if and only if $(\sigma^u)^{(M,(M)_1)} = 0$ (and the corresponding statement for $(N, (N)_1)$).

Claim 1 follows from the fact that, in a *finite* von Neumann algebra, any contraction is an average of two unitaries. Indeed if x is a contraction in a finite von Neumann algebra, then it has polar decomposition $x = u|x|$, where u is a unitary.

As $|x|$ is a self-adjoint contraction, by functional calculus it may be written as the average of two unitaries.

Claim 2. We have $(\sigma^u)^{(M,(M)_1)} = 0$ if and only if $(\sigma^{uu})^{(M,(M)_1)} = 0$ (and the corresponding statement for $(N, (N)_1)$).

The backwards direction of Claim 2 is trivial; the forwards direction follows from the fact that if x is a contraction in a finite factor and $\|x\|_2 \geq 1-\varepsilon$, then there is a unitary u so that $\|u - x\|_2 \leq 2\sqrt{\varepsilon}$.

Finally, suppose that σ is a sentence in prenex normal form and $\sigma^{(M,(M)_1)} = 0$. Then by Claims 1 and 2, we have $(\sigma^{uu})^{(M,(M)_1)} = 0$. Since atomic formulae are of the form $\|\lambda_1 x_1 + \cdots + \lambda_n x_n\|_2$ and arbitrary quantifier-free formulae are continuous combinations of atomic formulae, it follows from a winning strategy for Player II in $\mathfrak{G}_{\mathrm{vN}}(n, \varepsilon)$ (for suitably small ε) that $(\sigma^{uu})^{(N,(N)_1)} = 0$, whence $\sigma^{(N,(N)_1)} = 0$ by Claims 1 and 2 again. □

Suppose now that $\mathcal{L}_i = \{\Phi\}$, where Φ is a unary function symbol with modulus of uniform continuity $\Delta_\Phi(\varepsilon) = \varepsilon$. If M is a tracial von Neumann algebra, we view $U(M)$ as an $\mathcal{L}_i$-structure by interpreting Φ as multiplication by i. We then have this:

Corollary 3.6. *Let M and N be II_1-factors. Then M and N are locally equivalent if and only if $U(M)$ and $U(N)$ are elementarily equivalent as $\mathcal{L}_i$-structures.*

Proof. If M and N are locally equivalent, then M is elementarily equivalent to either N or N^{op}. It follows that there is an ultrafilter $\mathcal{U}$ such that $M^{\mathcal{U}}$ is isomorphic to $N^{\mathcal{U}}$ or $(N^{op})^{\mathcal{U}}$. In either case, $(U(M))^{\mathcal{U}} = U(M^{\mathcal{U}})$ is isomorphic to $(U(N))^{\mathcal{U}} = U(N^{\mathcal{U}})$ as $\mathcal{L}_i$-structures, whence $U(M)$ and $U(N)$ are elementarily equivalent as $\mathcal{L}_i$-structures.

Conversely, assume that $U(M)$ and $U(N)$ are elementarily equivalent as $\mathcal{L}_i$-structures. Then Player II has a winning strategy for the Ehrenfeucht–Fraïssé games for $U(M)$ and $U(N)$ as $\mathcal{L}_i$-structures. It then follows that Player II has a winning strategy in the games $\mathfrak{G}_{\mathrm{vN}}$ for M and N. Indeed, this follows from the fact that

$$\Re\langle u_i, u_j\rangle = 1 - \tfrac{1}{2} d(u_i, u_j)^2, \quad \Im\langle u_i, u_j\rangle = 1 - \tfrac{1}{2} d(u_i, i \cdot u_j)^2. \qquad \square$$

Remark 3.7. Notice that the proof of the previous corollary gives an alternative proof of the forward direction of Theorem 3.4.

Corollary 3.8. *Let M and N be II_1-factors in the class $\mathcal{K}_{op}$. Then M and N are elementarily equivalent if and only if $U(M)$ and $U(N)$ are elementarily equivalent as $\mathcal{L}_i$-structures.*

Corollary 3.9. *Let M and N be II_1-factors. Suppose that for every ε there is a $(1+\varepsilon)$-Lipschitz homeomorphism $f : U(M) \to U(N)$; that is, f is bijective with*

$$(1+\varepsilon)^{-1}\|u - v\|_2 \leq \|f(u) - f(v)\|_2 \leq (1+\varepsilon)\|u - v\|_2$$

that is further assumed to preserve the action by $\mathbb{Z}_4$. Then M and N are locally isomorphic.

We will say that M and N are *approximately Lipschitz isometric* if the condition of the previous corollary is satisfied. Although this relation ought to be in principle much stronger than elementary equivalence, to the best of our knowledge the results of [Farah et al. 2014b] heretofore furnish the only known examples of properties invariant under this relation namely, the McDuff property and property (Γ). It is, however, tempting to speculate that approximate Lipschitz isometry ought to be equivalent to isomorphism (up to opposites).

In lieu of this, it would be highly interesting to determine whether hyperfiniteness is an invariant of approximate Lipschitz isometry. If true, this would be in contrast with [Farah et al. 2014b, Theorem 4.3] which shows in particular that hyperfiniteness is not an invariant of elementary equivalence. Though one can show, essentially by Fact 3.1 and Proposition 3.2 (see also [Takesaki 2003, Chapter XIV.2]), that for every n, there exists $\varepsilon > 0$ so that for any ε-approximate Lipschitz embedding θ of M_n into a II_1-factor N, there is a $*$-homomorphism $\theta' : M_n \to N$ so that the image of the unit ball under θ is ε-contained in 2-norm in the image unit ball under θ' of M_n, this still does not seem sufficient, unless ε could be taken independent of n.

4. Further remarks and open problems

Of course, Corollary 3.8 raises the question: which $\mathbb{Z}_4$-metric spaces arise as unitary groups of II_1-factors? Even more importantly, what are the theories of such $\mathbb{Z}_4$-metric spaces? Ignoring the extra structure for a moment, an important example of a complete theory of (noncompact) metric spaces is the theory of the *Urysohn metric space*. (See, for example, [Ealy and Goldbring 2012].) Recall that the Urysohn metric space is the unique (up to isometry) complete, separable metric space that is universal (that is, every separable metric space isometrically embeds) and ultrahomogeneous (every isometry between finite — even compact — subspaces extends to an isometry of the entire space). However, the Urysohn space (or rather, its bounded counterpart, the Urysohn sphere) could never be isometric to the unitary group of a II_1-factor as the latter's metric is always negative definite.

Note that for M with separable predual, $U(M)$ isometrically embeds naturally in $\mathbb{S}^\infty$, the Hilbert sphere in ℓ^2. The space $\mathbb{S}^\infty$ is the "Hilbertian Urysohn sphere" in the sense described in [Nguyen Van Thé 2010, Section 1.4.2].

It is well worth pointing out the following proposition, which is an immediate consequence of Ozawa's fundamental result [2004] on the nonexistence of a universal, separable II_1-factor.

Proposition 4.1. *For any separable II_1-factor M, $U(M)$ is not universal among all $\mathbb{Z}_4$-metric spaces which embed* (*as $\mathbb{Z}_4$-metric spaces*) *in $\mathbb{S}^\infty$.*

Proof. Suppose, towards a contradiction, that there is a II_1-factor M for which $U(M)$ is universal among all $\mathbb{Z}_4$-metric spaces which embed in $\mathbb{S}^\infty$. In particular, for any II_1-factor N with separable predual, $U(N)$ isometrically embeds in $U(M)$ in a way which commutes with the action of i. Since this embedding respects the inner product, it is not hard to see it must extend to an isometric embedding $L^2(N) \to L^2(M)$ which takes N into M contractively. Thus, as above, there is a unital injective $*$-homomorphism $N \hookrightarrow pMp \oplus ((1-p)M(1-p))^{op}$, whence N embeds in either M or M^{op} since N is a factor. However, this would contradict the fact that there is no separable universal II_1-factor [Ozawa 2004] (pick $M \star M^{op}$). □

Question 4.2. Can $U(M)$ ever be universal among all metric spaces which embed in $\mathcal{S}^\infty$?

Proposition 4.1 is good evidence that the answer to the previous question is no. We remark that a positive answer to the previous question would be equivalent to demonstrating the existence of a separable II_1-factor for which there is an isometric embedding $\mathbb{S}^\infty \hookrightarrow U(M)$. We currently do not know whether $\mathbb{S}^\infty$ embeds isometrically in the unitary group of *any* II_1-factor. The existence of such an embedding ought to have striking consequences as the following proposition, which is similar in spirit, demonstrates.

Proposition 4.3. *Suppose M is a separable II_1-factor belonging to the class $\mathcal{K}_{op}$. Further suppose that, for each n, the n-dimensional complex spheres $\mathbb{S}^n$ isometrically embed in $U(M)$ with respect to the natural $\mathbb{Z}_4$-actions. Then M is a locally universal* II_1*-factor*; *that is*, *every separable II_1-factor embeds into an ultrapower of M. In particular*, *if*, *for each n*, *the n-dimensional complex spheres $\mathbb{S}^n$ isometrically embed in $U(\mathcal{R})$ with respect to the natural $\mathbb{Z}_4$-actions*, *then Connes' embedding problem has a positive answer.*

Proof. Suppose that M satisfies the assumption of the proposition and let N be a II_1-factor. Let F be any finite subset of $U(N)$. Then choosing an orthogonal projection P onto a suitably large finite-dimensional subspace so that $\|P(u)\| > 1-\varepsilon$ for all $u \in F \cup iF$, we can correct to an (effective in) ε-almost $\mathbb{Z}_4$-embedding of F into some $\mathbb{S}^n$, and therefore also in $U(M)$. But $\mathbb{Z}_4$-embeddings preserve inner products, whence pairs of inner products in F can be modeled arbitrarily well in $U(M)$. As above, Kirchberg's argument shows that N embeds in $M^{\mathcal{U}}$. □

We now remark how our main result recasts Kirchberg's characterization of $\mathcal{R}^\omega$-embeddability in a game-theoretical light. Let (A, tr) be an arbitrary tracial C*-algebra which we view as a normed space with respect to the 2-norm. To introduce a bit of terminology, we say that a subspace $E \subset A$ is *ε-almost representable* in $\mathcal{R}$ if there exists a subspace $F \subset \mathcal{R}$ and a linear bijection $T : E \to F$ so that $\|T\|, \|T^{-1}\| \leq 1+\varepsilon$ and $T(E \cap (A)_1) \subset_\varepsilon F \cap (\mathcal{R})_1$. Then by [Kirchberg

1993, Proposition 4.6], A is $\mathcal{R}^\omega$-embeddable if and only if for every $\varepsilon > 0$, every finite-dimensional subspace of A is ε-representable in $\mathcal{R}$.

Let us introduce the following "one-sided, one-round game" $\mathfrak{G}_{\mathcal{R}}(n, \varepsilon)$ for which the winning condition is that, for all $u_1, \ldots, u_n \in U(A)$ which are linearly independent, there exist n unitaries $v_1, \ldots, v_n \in U(\mathcal{R})$ so that the map

$$T : \operatorname{span}\{u_1, \ldots, u_n\} \to \operatorname{span}\{v_1, \ldots, v_n\}$$

defined by $T(u_i) = v_i$ satisfies $\|T\|, \|T^{-1}\| \le 1 + \varepsilon$.

Proposition 4.4. *There is a constant $N = N(n, \varepsilon)$ so that every n-dimensional subspace E of any tracial C^*-algebra (A, tr) is ε-almost representable in $\mathcal{R}$ if $\mathfrak{G}_{\mathcal{R}}(N, \varepsilon/4)$ is winnable.*

Proof. We first claim that there is a uniform constant $K(n, \varepsilon)$ so that for every n-dimensional subspace $E \subset A$ of any tracial C^*-algebra (A, tr) there exists a set of unitaries $\bar{u} = \{u_1, \ldots, u_l\} \subset U(A)$ with $l \le K$ so that every element of $E \cap (A)_1$ is ε-approximated in 2-norm by a convex combination of elements of $\bar{u}$.

Indeed, choose an $(\varepsilon/2)$-net $x_1, \ldots, x_m \in E \cap (A)_1$. The cardinality of such a net is bounded in particular by the $(\varepsilon/4)$-covering number of the unit ball in ℓ^2_n. We may perturb each x_i so that $\|x_i\| < 1 - \varepsilon/4$ and still have an ε-net for $E \cap (A)_1$. By the main result of [Popa 1981], there is a constant C depending only on ε so that each x_i is a convex combination of at most C unitaries in $U(A)$, whence the claim follows.

We next claim that if A is infinite-dimensional and if $E \subset A$ is a finite-dimensional subspace, then for every $\varepsilon > 0$ and $u \in U(A)$, there exists $u' \in U(A)$ with $\|u - u'\|_2 < \varepsilon$ and so that u' is linearly independent from E. To see this, let $P_E : L^2(A) \to E$ be the orthogonal projection onto E. By the Kaplansky density theorem, we have that $U(A)$ is 2-norm dense in $U(A'')$. Since $M := A'' \subset \mathcal{B}(L^2(A, \mathrm{tr}))$ is infinite-dimensional, it contains a diffuse abelian subalgebra. Therefore, there is a projection $p \in M$ with trace $\mathrm{tr}(p) = 1 - \varepsilon^2/2$ and a sequence of unitaries $v_n \in U(M)$ so that $v_n \to p$ weakly. Since P_E is a finite-rank operator, we thus have that $P_E(uv_n) \to P_E(up)$ strongly, whence

$$\|P_E(uv_n)\|_2 \to \|P_E(up)\|_2 \le \|p\|_2 = \sqrt{1 - \varepsilon^2/2}.$$

It is now easy to see that choosing n sufficiently large and $u' \in U(A)$ sufficiently close to uv_n works.

We now can proceed with the proof of the proposition. Let

$$E = \operatorname{span}\{u_1, \ldots, u_n\} \subset A.$$

(Every n-dimensional subspace of a C^*-algebra is a subspace of a space spanned by at most $4n$ unitaries, so we may assume this is the case without loss of generality.) By the previous claims, we can extend $u_1, \ldots, u_n$ to $u_1, \ldots, u_n, u_{n+1}, \ldots, u_s$

($s \leq n + K(n, \varepsilon)$) to a complete collection of linearly independent unitaries so that all elements in $E \cap (A)_1$ are 2ε-approximated in 2-norm by a convex combination of unitaries in the collection. If $\mathfrak{G}_R(s, \varepsilon/4)$ is winnable, then it is easy to check that for $S = T|_E$, we have that $S(E \cap (A)_1) \subset_\varepsilon S(E) \cap (\mathcal{R})_1$, and we are done. $\square$

Problem 4.5. Let $\mathcal{C} \subset \ell_n^2$ be a convex subset of the unit ball in n-dimensional Hilbert space. For every $\varepsilon > 0$, does there exist a II_1-factor M so that $(\ell_n^2, \mathcal{C})$ is ε-represented in M? Can one always choose a locally universal II_1-factor (in the sense of [Farah et al. 2014b]) or even $\mathcal{R}$?

References

[Ben Yaacov 2008] I. Ben Yaacov, "Continuous first order logic for unbounded metric structures", *J. Math. Log.* **8**:2 (2008), 197–223. MR 2011m:03065 Zbl 1191.03026

[Ben Yaacov et al. 2008] I. Ben Yaacov, A. Berenstein, C. W. Henson, and A. Usvyatsov, "Model theory for metric structures", pp. 315–427 in *Model theory with applications to algebra and analysis, II*, edited by Z. Chatzidakis et al., London Math. Soc. Lecture Note Ser. **350**, Cambridge Univ. Press, Cambridge, 2008. MR 2009j:03061 Zbl 1233.03045

[Connes 1975] A. Connes, "Sur la classification des facteurs de type II", *C. R. Acad. Sci. Paris Sér. A-B* **281**:1 (1975), Aii, A13–A15. MR 51 #13706 Zbl 0304.46042

[Dixmier and Lance 1969] J. Dixmier and E. C. Lance, "Deux nouveaux facteurs de type II_1", *Invent. Math.* **7** (1969), 226–234. MR 40 #1787 Zbl 0174.18701

[Dye 1955] H. A. Dye, "On the geometry of projections in certain operator algebras", *Ann. of Math.* (2) **61** (1955), 73–89. MR 16,598a Zbl 0064.11002

[Ealy and Goldbring 2012] C. Ealy and I. Goldbring, "Thorn-forking in continuous logic", *J. Symbolic Logic* **77**:1 (2012), 63–93. MR 2951630 Zbl 1257.03059

[Farah et al. 2013] I. Farah, B. Hart, and D. Sherman, "Model theory of operator algebras I: stability", *Bull. Lond. Math. Soc.* **45**:4 (2013), 825–838. MR 3081550 Zbl 1295.03019

[Farah et al. 2014a] I. Farah, B. Hart, and D. Sherman, "Model theory of operator algebras II: model theory", *Israel J. Math.* **201**:1 (2014), 477–505. MR 3265292 Zbl 1301.03037

[Farah et al. 2014b] I. Farah, B. Hart, and D. Sherman, "Model theory of operator algebras III: elementary equivalence and II_1 factors", *Bull. Lond. Math. Soc.* **46**:3 (2014), 609–628. MR 3210717 Zbl 1303.46049

[Hanche-Olsen and Størmer 1984] H. Hanche-Olsen and E. Størmer, *Jordan operator algebras*, Monographs and Studies in Mathematics **21**, Pitman (Advanced Publishing Program), Boston, 1984. MR 86a:46092 Zbl 0561.46031

[Hart 2012] B. Hart, "Continuous model theory and its applications", lecture notes, 2012, available at http://ms.mcmaster.ca/~bradd/courses/math712/index.html.

[Heinrich and Henson 1986] S. Heinrich and C. W. Henson, "Banach space model theory, II: Isomorphic equivalence", *Math. Nachr.* **125** (1986), 301–317. MR 87j:46139 Zbl 0633.46076

[Henson and Iovino 2002] C. W. Henson and J. Iovino, "Ultraproducts in analysis", pp. 1–110 in *Analysis and logic* (Mons, 1997), edited by C. Finet and C. Michaux, London Math. Soc. Lecture Note Ser. **262**, Cambridge Univ. Press, 2002. MR 2004d:03001 Zbl 1026.46007

[Henson and Moore 1983] C. W. Henson and L. C. Moore, Jr., "Nonstandard analysis and the theory of Banach spaces", pp. 27–112 in *Nonstandard analysis — recent developments* (Victoria, B.C., 1980), Lecture Notes in Math. **983**, Springer, Berlin, 1983. MR 85f:46033 Zbl 0511.46070

[Kirchberg 1993] E. Kirchberg, "On nonsemisplit extensions, tensor products and exactness of group C span $*$-algebras", *Invent. Math.* **112**:3 (1993), 449–489. MR 94d:46058 Zbl 0803.46071

[Nguyen Van Thé 2010] L. Nguyen Van Thé, "Structural Ramsey theory of metric spaces and topological dynamics of isometry groups", *Mem. Amer. Math. Soc.* **206**:968 (2010), x+140. MR 2011k:03096 Zbl 1203.05159

[Ozawa 2004] N. Ozawa, "There is no separable universal II_1-factor", *Proc. Amer. Math. Soc.* **132**:2 (2004), 487–490. MR 2004h:46074 Zbl 1041.46045

[Popa 1981] S. Popa, "On the Russo–Dye theorem", *Michigan Math. J.* **28**:3 (1981), 311–315. MR 82j:46074 Zbl 0499.46037

[Takesaki 2003] M. Takesaki, *Operator Algebras and Non-commutative Geometry, 8: Theory of operator algebras, III*, Encyclopaedia of Mathematical Sciences **127**, Springer, Berlin, 2003. MR 2004g:46080 Zbl 1059.46032

Received August 19, 2014. Revised March 2, 2015.

ISAAC GOLDBRING
DEPARTMENT OF MATHEMATICS, STATISTICS, AND COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS AT CHICAGO
SCIENCE AND ENGINEERING OFFICES M/C 249
851 SOUTH MORGAN STREET
CHICAGO, IL 60607-7045
UNITED STATES
isaac@math.uic.edu

THOMAS SINCLAIR
DEPARTMENT OF MATHEMATICS
PURDUE UNIVERSITY
150 N UNIVERSITY ST.
WEST LAFAYETTE, IN 47907-2067
UNITED STATES
tsincla@purdue.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.278.119

GROSSBERG–KARSHON TWISTED CUBES AND HESITANT WALK AVOIDANCE

MEGUMI HARADA AND EUNJEONG LEE

Let G be a complex semisimple simply connected linear algebraic group. Let λ be a dominant weight for G and $\mathcal{I} = (i_1, i_2, \ldots, i_n)$ a word decomposition for an element $w = s_{i_1} s_{i_2} \cdots s_{i_n}$ of the Weyl group of G, where the s_i are the simple reflections. In the 1990s, Grossberg and Karshon introduced a virtual lattice polytope associated to λ and $\mathcal{I}$, which they called a *twisted cube*, whose lattice points encode (counted with sign according to a density function) characters of representations of G. In recent work, Harada and Jihyeon Yang proved that the Grossberg–Karshon twisted cube is untwisted (so the support of the density function is a closed convex polytope) precisely when a certain torus-invariant divisor on a toric variety, constructed from the data of λ and $\mathcal{I}$, is basepoint-free. This corresponds to the situation in which the Grossberg–Karshon character formula is a true combinatorial formula, in the sense that there are no terms appearing with a minus sign. In this note, we translate this toric-geometric condition to the combinatorics of $\mathcal{I}$ and λ. More precisely, we introduce the notion of *hesitant λ-walks* and then prove that the associated Grossberg–Karshon twisted cube is untwisted precisely when $\mathcal{I}$ is hesitant-λ-walk-avoiding. Our combinatorial condition imposes strong geometric conditions on the Bott–Samelson variety associated to $\mathcal{I}$.

Introduction		120
1.	Background	121
2.	Diagram walks, hesitant walk avoidance, and statement of main theorem	124
3.	Proof of the main theorem: sufficiency	128
4.	Proof of the main theorem: necessity	130
5.	Open questions	134
Acknowledgements		135
References		135

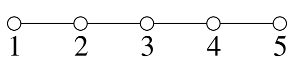
MSC2010: primary 20G05; secondary 52B20.
Keywords: Grossberg–Karshon twisted cubes, character formulae, pattern avoidance.

Introduction

Let G be a complex semisimple simply connected linear algebraic group. Building combinatorial models for G-representations is a fruitful technique in modern representation theory; a famous example is the theory of crystal bases and string polytopes. In a different direction, given a dominant weight λ and a choice of word expression $\mathcal{I} = (i_1, i_2, \ldots, i_n)$ of an element $w = s_{i_1} s_{i_2} \cdots s_{i_n}$ in the Weyl group, Grossberg and Karshon [1994] introduced a combinatorial object called a *twisted cube* $(C(\boldsymbol{c}, \ell), \rho)$, where $C(\boldsymbol{c}, \ell)$ is a subset of $\mathbb{R}^n$ and ρ is a support function with support precisely $C(\boldsymbol{c}, \ell)$. The lattice points of $C(\boldsymbol{c}, \ell)$ encode (counted with $\pm$ sign according to ρ) the character of the G-representation V_λ [Grossberg and Karshon 1994, Theorems 5 and 6]. Here the parameters $\boldsymbol{c}$ and ℓ are determined from λ and $\mathcal{I}$. These twisted cubes are combinatorially much simpler than general string polytopes but they are not true polytopes in the sense that their faces may have various angles and the intersection of faces may not be a face (cf. [Grossberg and Karshon 1994, §2.5 and Figure 1 therein]), and in general they may be neither closed nor convex (see Example 1.2). In particular, the Grossberg–Karshon character formula is not a purely combinatorial positive formula, since it may involve minus signs.

The main result of this note gives necessary and sufficient conditions on a dominant weight λ and a (not necessarily reduced) word expression $\mathcal{I} = (i_1, \ldots, i_n)$ of an element $w \in W$ such that the associated Grossberg–Karshon twisted cube is untwisted (cf. Definition 1.3), i.e., $C(\boldsymbol{c}, \ell)$ is a closed convex polytope and ρ is identically equal to 1 on $C(\boldsymbol{c}, \ell)$. This is precisely the situation in which the Grossberg–Karshon character formula is a true combinatorial formula, in the sense that it is a purely positive formula (with no terms appearing with a minus sign). In addition, an anonymous referee pointed out to us that the combinatorial condition on $\mathcal{I}$ and λ in our result also has interesting geometric consequences: it implies that (the image in a flag variety of) the corresponding Bott–Samelson variety is a *toric Schubert variety* in the sense of [Karuppuchamy 2013]; see Remark 2.10.

In order to state our result it is useful to introduce some terminology (see Section 2 for details). Roughly, we say that a word $\mathcal{I} = (i_1, \ldots, i_n)$ is a *diagram walk* (or simply *walk*) if successive roots are adjacent in the Dynkin diagram: for instance, in type A_5



the word $\mathcal{I} = (2, 4, 5)$ with corresponding simple roots (s_2, s_4, s_5) is not a walk since s_2 and s_4 are not adjacent, but $\mathcal{I} = (1, 2, 3, 2, 1)$ is a walk. Moreover, given a dominant weight $\lambda = \lambda_1 \varpi_1 + \cdots + \lambda_r \varpi_r$ written as a linear combination of the fundamental weights $\{\varpi_1, \ldots, \varpi_r\}$, we say $\mathcal{I} = (i_1, i_2, \ldots, i_n)$ is a λ-walk if it is a walk and if it ends at a root which appears in λ, i.e., $\lambda_{i_n} > 0$. A *hesitant λ-walk* is a word $\mathcal{I} = (i_0, i_1, \ldots, i_n)$ where $i_0 = i_1$, so there is a repetition at the first step, and the subword $(i_1, i_2, \dots, i_n)$ is a λ-walk. Finally, a word is *hesitant-λ-walk-avoiding* if there is no subword which is a hesitant λ-walk. With this terminology we can state the main result of this paper.

Theorem. *Let $\mathcal{I} = (i_1, i_2, \dots, i_n)$ be a word decomposition of an element $w = s_{i_1} s_{i_2} \cdots s_{i_n}$ of the Weyl group W and let $\lambda = \lambda_1 \varpi_1 + \lambda_2 \varpi_2 + \cdots + \lambda_r \varpi_r$ be a dominant weight. Then the corresponding Grossberg–Karshon twisted cube $(C(\boldsymbol{c}, \ell), \rho)$ is untwisted if and only if $\mathcal{I}$ is hesitant-λ-walk-avoiding.*

We note that pattern avoidance is an important notion in the study of Schubert varieties and Schubert calculus, first pioneered by Lakshmibai and Sandhya [1990] and further studied by many others (see, e.g., [Abe and Billey 2014] and references therein). It would be interesting to explore the relation between our notion of hesitant-λ-walk-avoidance with the other types of pattern avoidance in the theory of flag and Schubert varieties.

We additionally remark that Kiritchenko has recently defined *divided-difference operators D_i on polytopes* and, using these D_i inductively together with a fixed choice of reduced word decomposition for the longest element in the Weyl group of G, she constructs (possibly virtual) polytopes whose lattice points encode the character of irreducible G-representations [Kiritchenko 2013, Theorem 3.6]. Kiritchenko's virtual polytopes are generalizations of both Gel'fand–Cetlin polytopes and the Grossberg–Karshon twisted polytopes. It would be interesting to explore whether our methods can be further generalized to study Kiritchenko's virtual polytopes (see Section 5).

This paper is organized as follows. In Section 1 we recall the necessary definitions and background from previous papers. In particular, we recall the results of Harada and Yang [2015, Proposition 2.1 and Theorem 2.4] which characterize the untwistedness of the Grossberg–Karshon twisted cube in terms of the Cartier data associated to a certain toric divisor on a toric variety; this is a key tool for our proof. In Section 2 we introduce the notions of diagram walks and hesitant λ-walks and state our main theorem. We prove the sufficiency of hesitant-λ-walk-avoidance in Section 3. The proof of necessity, which occupies Section 4, is in part a case-by-case analysis according to Lie type. We briefly record some open questions in Section 5.

1. Background

We begin by recalling the definition of *twisted cube* given by Grossberg and Karshon [1994, §2.5]. We follow the exposition in [Harada and Yang 2015]. Fix a positive integer n. A twisted cube is a pair $(C(\boldsymbol{c}, \ell), \rho)$ where $C(\boldsymbol{c}, \ell)$ is a subset of $\mathbb{R}^n$ and $\rho : \mathbb{R}^n \to \mathbb{R}$ is a density function with support precisely equal to $C(\boldsymbol{c}, \ell)$. Here $\boldsymbol{c} = \{c_{jk}\}_{1 \le j < k \le n}$ and $\ell = \{\ell_1, \ell_2, \dots, \ell_n\}$ are fixed integers. (The general definition in [Grossberg and Karshon 1994] only requires the ℓ_i to be real numbers, but since we restrict our attention to the cases arising from representation theory, our ℓ_i will always be integers.) In order to simplify the notation in what follows, we define the following functions on $\mathbb{R}^n$:

$$A_n(x) = A_n(x_1, \ldots, x_n) = \ell_n, \tag{1-1}$$
$$A_j(x) = A_j(x_1, \ldots, x_n) = \ell_j - \sum_{k>j} c_{jk} x_k \quad \text{for all } 1 \le j \le n-1.$$

We also define a function $\operatorname{sgn} : \mathbb{R} \to \{\pm 1\}$ by $\operatorname{sgn}(x) = 1$ for $x < 0$ and $\operatorname{sgn}(x) = -1$ for $x \ge 0$.

We now give the precise definition.

Definition 1.1. Let n, $\boldsymbol{c}$, ℓ, and A_j be as above. Let $C(\boldsymbol{c}, \ell)$ denote the following subset of $\mathbb{R}^n$:

$$C(\boldsymbol{c}, \ell) := \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid \forall\, 1 \le j \le n,\ A_j(x) < x_j < 0 \text{ or } 0 \le x_j \le A_j(x)\}. \tag{1-2}$$

Moreover, we define a density function $\rho : \mathbb{R}^n \to \mathbb{R}$ by

$$\rho(x) = \begin{cases} (-1)^n \prod_{k=1}^n \operatorname{sgn}(x_k) & \text{if } x \in C(\boldsymbol{c}, \ell), \\ 0 & \text{else}. \end{cases} \tag{1-3}$$

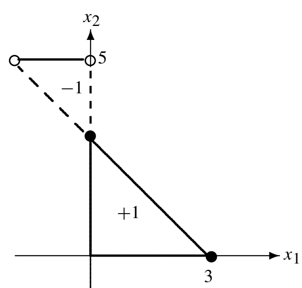
Evidently $\operatorname{supp}(\rho) = C(\boldsymbol{c}, \ell)$. We call the pair $(C(\boldsymbol{c}, \ell), \rho)$ the *twisted cube associated to $\boldsymbol{c}$ and ℓ*.

A twisted cube may not be a cube in the standard sense. In particular, the set C may be neither convex nor closed, as the following example shows. See also the discussion in [Grossberg and Karshon 1994, §2.5].

Example 1.2. Let $n = 2$ and let $\ell = (\ell_1 = 3, \ell_2 = 5)$ and $\boldsymbol{c} = \{c_{12} = 1\}$. Then

$$C = \{(x_1, x_2) \in \mathbb{R}^2 \mid 0 \le x_2 \le 5 \text{ and } (3 - x_2 < x_1 < 0 \text{ or } 0 \le x_1 \le 3 - x_2)\}.$$

See the figure. The value of the density function ρ is recorded within each region.



Note in particular that C does *not* contain the points $\{(0, x_2) \mid 3 < x_2 < 5\}$ and the points $\{(x_1, x_2) \mid 3 < x_2 < 5 \text{ and } x_1 = 3 - x_2\}$, so C is not closed, and it is also not convex.

As mentioned in the introduction, the main goal of this note is to give necessary and sufficient conditions for the *untwistedness* of the twisted cube, stated in terms of the combinatorics of the defining parameters. The following makes the notion precise.

Definition 1.3 (cf. [Harada and Yang 2015, Definition 2.2]). We say that the Grossberg–Karshon twisted cube $(C = C(\boldsymbol{c}, \ell), \rho)$ is *untwisted* if C is a closed convex polytope and if the support for ρ is constant and equal to 1 on C and 0 elsewhere. We say the twisted cube is *twisted* if it is not untwisted.

The main result of [Harada and Yang 2015] characterizes the untwistedness of the Grossberg–Karshon twisted cube in terms of the basepoint-freeness of a certain toric divisor on a toric variety constructed from the data of $\boldsymbol{c}$ and ℓ, which in turn can be stated in terms of the so-called Cartier data $\{m_\sigma\}$ associated to the divisor. In particular, in this paper we will not require the geometric perspective; instead we work with the integer vectors m_σ, which can be derived directly from the constants $\boldsymbol{c}$ and ℓ. Before quoting the relevant result from [Harada and Yang 2015] we need some terminology.

Let $\{e_1^+, \ldots, e_n^+\}$ be the standard basis of $\mathbb{R}^n$. For $\sigma = (\sigma_1, \ldots, \sigma_n) \in \{+, -\}^n$, define $m_\sigma = (m_{\sigma,1}, \ldots, m_{\sigma,n}) = \sum m_{\sigma,k} e_k^+ \in \mathbb{Z}^n$ as follows, using the functions $A_k(x)$ defined in (1-1):

$$m_{\sigma,k} = \begin{cases} 0 & \text{if } \sigma_k = +, \\ A_k(m_{\sigma,k+1}, \ldots, m_{\sigma,n}) & \text{if } \sigma_k = -. \end{cases} \tag{1-4}$$

We will also need a certain polytope P_D:

$$P_D = \{x \in \mathbb{R}^n \mid 0 \le x_j \le A_j(x) \text{ for all } 1 \le j \le n\} \subseteq \mathbb{R}^n. \tag{1-5}$$

Theorem 1.4 (cf. [Harada and Yang 2015, Proposition 2.1]). *Let n, $\boldsymbol{c}$, and ℓ be as above and let $(C(\boldsymbol{c}, \ell), \rho)$ denote the corresponding Grossberg–Karshon twisted polytope. Then $(C(\boldsymbol{c}, \ell), \rho)$ is untwisted if and only if $m_{\sigma,k} \ge 0$ for all $\sigma \in \{+, -\}^n$ and for all k with $1 \le k \le n$.*

Recall that the goal of this note is to analyze the case when the defining parameters for the Grossberg–Karshon twisted polytope arise from certain representation-theoretic data. We now briefly describe how to derive the $\boldsymbol{c}$ and ℓ in this case.

Following [Grossberg and Karshon 1994], let G be a complex semisimple simply connected linear algebraic group of rank r over an algebraically closed field $\boldsymbol{k}$. Choose a Cartan subgroup $H \subset G$ and a Borel subgroup. Let $\{\alpha_1, \ldots, \alpha_r\}$ denote the simple roots, $\{\alpha_1^\vee, \ldots, \alpha_r^\vee\}$ the coroots, and $\{\varpi_1, \ldots, \varpi_r\}$ the fundamental weights (characterized by the relation $\langle \varpi_i, \alpha_j^\vee \rangle = \delta_{ij}$). Let $s_{\alpha_i} \in W$ denote the simple reflection in the Weyl group corresponding to the root α_i.

Fix a choice $\lambda = \lambda_1\varpi_1 + \cdots + \lambda_r\varpi_r$ in the weight lattice, where $\lambda_i \in \mathbb{Z}$. Let $\mathcal{I} = (i_1, \ldots, i_n)$ be a sequence of elements in $[r] := \{1, 2, \ldots, r\}$; this corresponds to a (not necessarily reduced) decomposition of an element $w = s_{\alpha_{i_1}} s_{\alpha_{i_2}} \cdots s_{\alpha_{i_n}}$ in W. For simplicity, we introduce the notation $\beta_j := \alpha_{i_j}$, so β_j is the j-th simple root appearing in the word decomposition. For such λ and $\mathcal{I}$ we define constants $\boldsymbol{c}$ and ℓ by the formulas (cf. [Grossberg and Karshon 1994, §3.7])

$$c_{jk} = \langle \beta_k, \beta_j^\vee \rangle \tag{1-6}$$

for $1 \le j < k \le n$, and

$$\ell_1 = \langle \lambda, \beta_1^\vee \rangle, \ \ldots, \ \ell_n = \langle \lambda, \beta_n^\vee \rangle. \tag{1-7}$$

Note that if the j-th simple reflection in the given word decomposition $\mathcal{I}$ is equal to α_i, then $\ell_j = \lambda_i$, and that the constants c_{jk} are matrix entries in the Cartan matrix of G.

Example 1.5. Consider $G = \mathrm{SL}(3, \mathbb{C})$ with positive roots $\{\alpha_1, \alpha_2\}$, and let $\lambda = 2\varpi_1 + \varpi_2$ and $\mathcal{I} = (1, 2, 1)$. Then $(\beta_1, \beta_2, \beta_3) = (\alpha_1, \alpha_2, \alpha_1)$ and we have

$$\begin{aligned} c_{12} &= \langle \alpha_2, \alpha_1^\vee \rangle = -1, \\ c_{13} &= \langle \alpha_1, \alpha_1^\vee \rangle = 2, \\ c_{23} &= \langle \alpha_1, \alpha_2^\vee \rangle = -1, \\ \ell = (\ell_1, \ell_2, \ell_3) &= \big(\langle \lambda, \alpha_1^\vee \rangle = 2, \langle \lambda, \alpha_2^\vee \rangle = 1, \langle \lambda, \alpha_1^\vee \rangle = 2\big). \end{aligned} \tag{1-8}$$

As mentioned in the introduction, in the setting above Grossberg and Karshon derive a Demazure-type character formula for the irreducible G-representation corresponding to λ, expressed as a sum over the lattice points $\mathbb{Z}^n \cap C(\boldsymbol{c}, \ell)$ in the Grossberg–Karshon twisted cube $(C(\boldsymbol{c}, \ell), \rho)$ [Grossberg and Karshon 1994, Theorem 5 and Theorem 6]. The lattice points appear with a plus or minus sign according the density function ρ. Hence their formula is a *positive* formula if ρ is constant and equal to 1 on all of $C(\boldsymbol{c}, \ell)$. From the point of view of representation theory it is therefore of interest to determine conditions on the weight λ and the word decomposition $\mathcal{I} = (i_1, i_2, \ldots, i_n)$ for an element $w = s_{i_1} s_{i_2} \cdots s_{i_n}$ such that the associated Grossberg–Karshon twisted cube is in fact untwisted. This is the motivation for this note.

2. Diagram walks, hesitant walk avoidance, and statement of main theorem

In order to state our main theorem we introduce some terminology. In what follows, we fix an ordering on the simple roots as in Table 1; our conventions agree with those in the standard textbook of Humphreys [1972]. In particular, given an index i

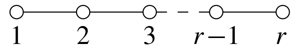
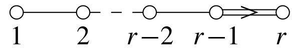
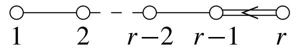
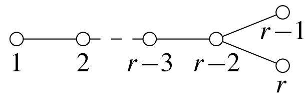
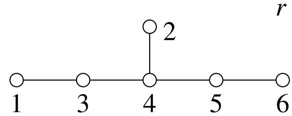
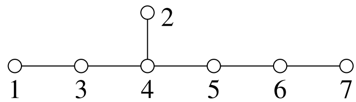
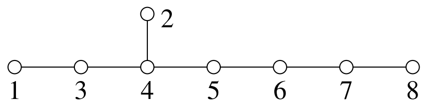
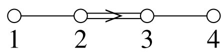
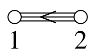
Φ	Dynkin diagram
A_r $(r \geq 1)$	
B_r $(r \geq 2)$	
C_r $(r \geq 3)$	
D_r $(r \geq 4)$	
E_6	
E_7	
E_8	
F_4	
G_2	

Table 1. Dynkin diagrams for all Lie types.

with $1 \leq i \leq r$, where r is the rank of G, we may refer to its corresponding simple reflection $s_i := s_{\alpha_i}$, where the index i refers to the ordering of the roots in Table 1.

Definition 2.1. Let $\mathcal{I} = (i_1, i_2, \ldots, i_n) \in [r]^n$ be a (not necessarily reduced) word decomposition of an element $w = s_{i_1} s_{i_2} \cdots s_{i_n}$ of the Weyl group W. We say that $\mathcal{I}$ is a *diagram walk* (or *walk*) if successive simple roots are adjacent in the corresponding Dynkin diagram, or more precisely, if for each $j \in [n-1] = \{1 \leq j \leq n-1\}$ the two successive roots α_{i_j} and $\alpha_{i_{j+1}}$ are distinct and there is an edge in the corresponding Dynkin diagram connecting α_{i_j} and $\alpha_{i_{j+1}}$. We call i_1 (or α_{i_1}) the *initial root* (*of the diagram walk* $\mathcal{I}$) and denote it by $\mathrm{IR}(\mathcal{I})$. We call i_n (or α_{i_n}) the *final root* (*of the diagram walk* $\mathcal{I}$) and denote it $\mathrm{FR}(\mathcal{I})$.

Example 2.2. (1) In type A, the words $s_2 s_3 s_4 s_5 s_4 s_3$ and $s_1 s_2 s_1 s_2 s_3$ are both diagram walks. Note that the second word is not reduced.

(2) In type B, $s_{r-2} s_{r-1} s_r$ is a diagram walk.

(3) In type E_8, $s_1 s_3 s_4 s_2 s_4 s_5$ is a diagram walk.

In what follows, we also find it useful to consider words which are almost diagram walks, except that the word begins with a repetition (thus disqualifying it from being a walk), i.e., the initial root appears twice.

Definition 2.3. Let $\mathcal{I} = (i_0, i_1, i_2, \ldots, i_n)$ be a (not necessarily reduced) word decomposition of an element $w = s_{i_0} s_{i_1} \cdots s_{i_n}$ of the Weyl group W. We say that $\mathcal{I}$ is a *hesitant (diagram) walk* if

- the length of the word is at least 2, i.e., $n \geq 1$,
- the first two roots are the same, i.e., $i_0 = i_1$, and
- the subword $(i_1, \ldots, i_n)$ is a diagram walk.

In other words, except for the hesitation at the first step, the remainder of the word is a diagram walk. We refer to the subword $(i_1, \ldots, i_n)$ as the *walking component* of the hesitant walk.

A few remarks are in order. First, we emphasize that a hesitant walk, despite the terminology, is not actually a diagram walk; it becomes a diagram walk only after deleting the first entry in the word. Furthermore, it is clear that a hesitant (diagram) walk is never a reduced word decomposition (because of the two repeated roots at the beginning). On the other hand, it is possible for a reduced word decomposition to *contain* a hesitant walk as a subword: for instance, for $G = \mathrm{SL}(4, \mathbb{C})$, the reduced word decomposition $s_1 s_2 s_3 s_1 s_2 s_1$ for the longest element in the Weyl group S_4 contains $s_1 s_1 s_2$ as a subword, which is a hesitant walk.

Definition 2.4. Let $\mathcal{I} = (i_1, i_2, \ldots, i_n)$ be a word decomposition of an element $w = s_{i_1} s_{i_2} \cdots s_{i_n}$ of the Weyl group W. We say that $\mathcal{I}$ is *hesitant-walk-avoiding* if there is no subword $\mathcal{J} = (i_{j_0}, i_{j_1}, \ldots, i_{j_s})$ of $\mathcal{I}$ which is a hesitant walk.

Example 2.5. Let $G = \mathrm{SL}(4, \mathbb{C})$ with Weyl group S_4. The reduced word decomposition $s_1 s_2 s_3$ is hesitant-walk-avoiding.

In what follows we will also be interested in dominant weights λ in the character lattice $X(H)$ associated to G. As in Section 1, we may express λ as a linear combination of the fundamental weights $\varpi_1, \ldots, \varpi_r$ corresponding to the simple roots $\alpha_1, \ldots, \alpha_r$. Thus we write

$$\lambda = \lambda_1 \varpi_1 + \cdots + \lambda_r \varpi_r$$

and since we assume λ is dominant, $\lambda_i \geq 0$ for all $i = 1, \ldots, r$.

Definition 2.6. Let λ be as above. We say that a simple root α_i *appears in* λ if the corresponding coefficient is strictly positive, i.e.,

$$\lambda_i = \langle \lambda, \alpha_i^\vee \rangle > 0. \tag{2-1}$$

We now introduce some terminology which relates diagram walks and hesitant walks to the dominant weight λ.

Definition 2.7. Let λ and $\mathcal{I}$ be as above. We will say that $\mathcal{I}$ is a λ-*walk* if

- $\mathcal{I}$ is a diagram walk, and
- the final root $\mathrm{FR}(\mathcal{I})$ of the walk $\mathcal{I}$ appears in λ.

Similarly, we say that $\mathcal{I}$ is a *hesitant* λ-*walk* if it is a hesitant walk and the final root of its walking component appears in λ. Finally, a word $\mathcal{I}$ is *hesitant-*λ*-walk-avoiding* if there is no subword $\mathcal{J}$ of $\mathcal{I}$ which is a hesitant λ-walk.

Example 2.8. Let $G = \mathrm{SL}(4, \mathbb{C})$ with Weyl group S_4. Consider the reduced word decomposition $\mathcal{I} = (1, 2, 3, 1, 2, 1)$ of the longest element $w_0 = s_1 s_2 s_3 s_1 s_2 s_1$ of S_4 and $\lambda = 3\varpi_3$. Then $\mathcal{I}$ is hesitant-λ-walk-avoiding.

Given the terminology introduced above we may now state our main theorem.

Theorem 2.9. *Let $\mathcal{I} = (i_1, i_2, \ldots, i_n)$ be a word decomposition of an element $w = s_{i_1} \cdots s_{i_n}$ of W and let $\lambda = \lambda_1\varpi_1 + \lambda_2\varpi_2 + \cdots + \lambda_r\varpi_r$ be a dominant weight. Let $\boldsymbol{c} = \{c_{jk}\}$ and $\ell = (\ell_1, \ldots, \ell_n)$ be determined from λ and $\mathcal{I}$ as in (1-6) and (1-7). Then the corresponding Grossberg–Karshon twisted cube $(C(\boldsymbol{c}, \ell), \rho)$ is untwisted if and only if $\mathcal{I}$ is hesitant-λ-walk-avoiding.*

The proof of the above theorem occupies Sections 3 and 4.

Remark 2.10. We thank the anonymous referee for pointing out that the combinatorial criterion of hesitant-λ-walk-avoidance has the following interesting geometric consequence. Since we have not introduced in this paper the objects in the following discussion, we keep our comments brief (the reader may consult, e.g., [Grossberg and Karshon 1994] for definitions). For a word $\mathcal{I} = (i_1, \ldots, i_n)$, let $Z(\mathcal{I})$ denote the associated Bott–Samelson variety and let $\pi_{\mathcal{I}} : Z(\mathcal{I}) \to G/B$ be the natural morphism. For a dominant weight λ, let $\varphi_\lambda : G/B \to \mathbb{P}(V_\lambda)$ denote the Plücker embedding. Let P_λ denote the parabolic subgroup of G corresponding to the set of all simple roots *not* appearing in λ in the sense of Definition 2.6; note that if λ is strictly dominant, then $P_\lambda = B$, and also that φ_λ factors through G/P_λ. Now let $\mathcal{I}'$ be the word obtained from $\mathcal{I}$ by deleting all the simple roots in $\mathcal{I}$ that do not appear in λ. If $\mathcal{I}$ is hesitant-λ-walk-avoiding, then in particular any simple root appearing in λ can occur at most once, so the simple roots occurring in $\mathcal{I}'$ are pairwise distinct. Note that by the definition of P_λ, the images of $Z(\mathcal{I})$ and $Z(\mathcal{I}')$ in G/P_λ are the same, and hence also in $\mathbb{P}(V_\lambda)$ via φ_λ. Furthermore, because the simple roots occurring in $\mathcal{I}'$ are pairwise distinct, from the classification of toric Schubert varieties in [Karuppuchamy 2013] it follows that the Schubert variety $X_{w(\mathcal{I}')}$ (as well as $Z(\mathcal{I}')$) is actually a toric variety. (Here $w(\mathcal{I}')$ denotes the product in the Weyl group W of the simple reflections in the word $\mathcal{I}'$ and $X_{w(\mathcal{I}')}$

denotes the corresponding Schubert variety.) Thus we see that the combinatorial criterion of Theorem 2.9 places quite strong conditions on the geometry of the associated Bott–Samelson variety and its images.

3. Proof of the main theorem: sufficiency

We begin the proof of Theorem 2.9 by first proving the "if" part of the statement, i.e., that hesitant-λ-walk-avoidance implies the untwistedness of the Grossberg–Karshon twisted cube.

We need some preliminary lemmas. Recall that the $m_\sigma = (m_{\sigma,1}, \ldots, m_{\sigma,n})$ are integer vectors defined by (1-4) associated to the defining constants $\boldsymbol{c}$ and ℓ of the twisted cube.

Lemma 3.1. *Let $\{c_{ij}\}_{1\le i<j\le n}$ and $\ell_1, \ldots, \ell_n$ be fixed integers. Assume that $\ell_i \ge 0$ for all i. If there exists an element σ of $\{+,-\}^n$ and $k \in [n]$ such that $m_{\sigma,k} > 0$ and $m_{\sigma,i} \ge 0$ for $i > k$, then there exists an increasing sequence $\mathcal{J}$ of indices $1 \le j_1 < j_2 < \cdots < j_s \le n$, with $s \ge 1$, such that*

(1) $j_1 = k$,

(2) $\ell_{j_s} > 0$, *and*

(3) $c_{j_t j_{t+1}} < 0$ *for* $t = 1, \ldots, s-1$.

Proof. Let σ and k be as above. We may explicitly construct the subsequence $\mathcal{J}$ as follows. First suppose $\ell_k > 0$. In this case, the subsequence $\mathcal{J} = (j_1 = k)$ satisfies the three required conditions (the third being vacuous), so we are done. If on the other hand $\ell_k = 0$, we set $j_1 = k$ and then define j_2 as follows. By assumption $m_{\sigma,k} > 0$, so we know $\sigma_k = -$, and by the definition of the m_σ we know

$$m_{\sigma,k} = \ell_k - \sum_{i>k} c_{ki} m_{\sigma,i} = -\sum_{i>k} c_{ki} m_{\sigma,i}. \tag{3-1}$$

Since $m_{\sigma,i} \ge 0$ for $i \ge k$ by assumption, in order for $m_{\sigma,k}$ to be strictly positive there must exist an index $J > k$ with $c_{kJ} < 0$ and $m_{\sigma,J} > 0$. Choose j_2 to be the minimal such index. If $\ell_{j_2} > 0$, then the sequence $\mathcal{J} = (j_1 = k, j_2)$ satisfies the three required conditions and we are done. Otherwise, we may repeat the above argument as many times as necessary (i.e., as long as $\ell_{j_t} = 0$). Since the indices j_t are bounded above by n, this process must stop, i.e., there must exist some $s \ge 1$ such that the sequence $\mathcal{J} = (j_1, \ldots, j_s)$ found in this manner satisfies the requirements. □

In the case when the constants $\boldsymbol{c}$ and ℓ are obtained from the data of a weight λ and a word $\mathcal{I}$ we can interpret Lemma 3.1 using the terminology introduced in Section 2.

Corollary 3.2. *Let $\mathcal{I} = (i_1, i_2, \ldots, i_n)$ be a word decomposition of an element $w = s_{i_1} \cdots s_{i_n}$ of W and let $\lambda = \lambda_1\varpi_1 + \lambda_2\varpi_2 + \cdots + \lambda_r\varpi_r$ be a dominant weight,* *i.e.*, $\lambda_i \geq 0$ *for all* i. *Let* $\boldsymbol{c}$, ℓ, *and* $\{m_\sigma\}_{\sigma\in\{+,-\}^n}$ *be determined from* $\mathcal{I}$ *and* λ *as in* (1-6), (1-7), *and* (1-4). *If there exist an element* σ *of* $\{+,-\}^n$ *and* $k \in [n]$ *such that* $m_{\sigma,k} > 0$ *and* $m_{\sigma,i} \geq 0$ *for* $i > k$, *then there exists a subword* $\mathcal{J} = (i_{j_1}, i_{j_2}, \dots, i_{j_s})$ *of* $\mathcal{I}$, *of length at least* 1 (*i.e.*, $s \geq 1$), *such that* $j_1 = k$ *and* $\mathcal{J}$ *is a* λ*-walk* (*i.e.*, *it is a diagram walk and the final root* $\mathrm{FR}(\mathcal{J})$ *appears in* λ).

Proof. First observe that by the definition of the ℓ_i (1-7) and by the assumption on λ, we have $\ell_i \geq 0$ for all i, and $\ell_i > 0$ exactly when β_i, the i-th simple root in $\mathcal{I}$, appears in λ. Let σ and k be as above. Then by Lemma 3.1 there exists a subword $\mathcal{J} = (i_{j_1} = i_k, i_{j_2}, \dots, i_{j_s})$ of length at least 1 such that $j_1 = k$ and $\mathrm{FR}(\mathcal{J})$ appears in λ. It remains to check that $\mathcal{J}$ is a diagram walk. Recall that by definition $c_{j\ell} = \langle \beta_\ell, \beta_j^\vee \rangle$. Hence $c_{j\ell} < 0$ if and only if there is an edge in the corresponding Dynkin diagram connecting the roots α_{i_j} and α_{i_ℓ}, so by the conditions on $\mathcal{J}$ in Lemma 3.1 we see that $\mathcal{J}$ is a diagram walk, as desired. □

The next result is the main technical fact we need.

Lemma 3.3. *Let* $\{c_{ij}\}_{1\leq i<j\leq n}$ *and* $\ell_1, \dots, \ell_n$ *be fixed integers and let* $(C(\boldsymbol{c}, \ell), \rho)$ *be the corresponding Grossberg–Karshon twisted cube. Assume that* $\ell_i \geq 0$ *for all* i. *If* $(C(\boldsymbol{c}, \ell), \rho)$ *is twisted, then there exists an increasing subsequence* $\mathcal{J} = (j_0 < j_1 < \cdots < j_s)$ *of indices of length at least* 2 (*i.e.*, $s \geq 1$) *such that*

(1) $\ell_{j_s} > 0$,

(2) $c_{j_0 j_1} > 0$, *and*

(3) $c_{j_t j_{t+1}} < 0$ *for all* $t = 1, \dots, s-1$.

Proof. By Theorem 1.4, there exist an element σ of $\{+,-\}^n$ and an index k such that $m_{\sigma,k} < 0$. For such a choice of σ we may assume without loss of generality that k is chosen to be the maximal such index, i.e., that $m_{\sigma,k} < 0$ and $m_{\sigma,s} \geq 0$ for $s > k$. Recall that by definition

$$m_{\sigma,k} = \ell_k - \sum_{s>k} c_{ks} m_{\sigma,s}.$$

By assumption $m_{\sigma,k} < 0$, so we have $\sum_{s>k} c_{ks} m_{\sigma,s} > \ell_k \geq 0$. Since $m_{\sigma,s} \geq 0$ for $s > k$, this implies that there exists some $p > k$ with $c_{kp} > 0$ and $m_{\sigma,p} > 0$. Applying Lemma 3.1 we obtain an increasing sequence $(j_1 = p, j_2, \dots, j_s)$ of indices with $s \geq 1$ such that $\ell_{j_s} > 0$ and $c_{j_t j_{t+1}} < 0$ for all $t = 1, \dots, s-1$. Then by choosing $j_0 = k < j_1 = p$ and since $c_{j_0 j_1} = c_{kp} > 0$ by construction of p, we obtain a sequence $\mathcal{J} = (j_0 = k, j_1 = p, \dots, j_s)$ satisfying the required conditions. □

The proof of the “if” part of Theorem 2.9 is a straightforward consequence of the above lemma.

Proof of the "if" part of Theorem 2.9. We will prove the contrapositive. Suppose the Grossberg–Karshon twisted cube $(C(\boldsymbol{c}, \ell), \rho)$ is twisted. By the dominance assumption on λ and by the definition of the ℓ_i, we know $\ell_i \geq 0$ for all i. Thus we may apply Lemma 3.3. Note also that $\ell_{j_s} > 0$ precisely when the root β_{j_s} appears in λ. Moreover, by definition, we know that $c_{j_0 j_1} := \langle \beta_{j_1}, \beta_{j_0}^\vee \rangle > 0$ if and only if $\beta_{j_0} = \beta_{j_1}$ (equivalently, $i_{j_0} = i_{j_1}$) and $c_{j_t j_{t+1}} < 0$ if and only if there is an edge in the corresponding Dynkin diagram connecting the roots β_{j_t} and $\beta_{j_{t+1}}$. Thus the subword $(i_{j_0}, i_{j_1}, \ldots, i_{j_s})$ of $\mathcal{I}$ corresponding to the subsequence $(j_0, j_1, \ldots, j_s)$ of indices obtained from Lemma 3.3 is a hesitant λ-walk, as desired. □

4. Proof of the main theorem: necessity

We now prove the "only if" part of Theorem 2.9, i.e., that untwistedness implies hesitant-λ-walk-avoidance. Part of the proof will be a case-by-case analysis of the possible Lie types of G.

For convenience, in Table 2 we recall the Cartan matrices for all Lie types (see, for example, [Humphreys 1972, pp. 58–59]).

In the discussion below it will be useful to restrict our attention to hesitant λ-walks which are minimal in an appropriate sense. We make this precise in the definition below.

Definition 4.1. Let λ be a dominant weight and let $\mathcal{I} = (i_0, \ldots, i_n)$ be a hesitant λ-walk. We say that $\mathcal{I}$ is *minimal* if

(1) $\{i_1, \ldots, i_n\}$ are all distinct, i.e., the walking component of $\mathcal{I}$ visits any given vertex of the Dynkin diagram at most once, and

(2) $\beta_0, \ldots, \beta_{n-1}$ do not appear in λ if $n \geq 2$.

Example 4.2. Let $G = \mathrm{SL}(6, \mathbb{C})$.

- Let $\lambda = \varpi_2$. The hesitant λ-walk $\mathcal{J} = (5, 5, 4, 3, 4, 3, 2)$ is not minimal since the walking component revisits some vertices multiple times, but the subword $\mathcal{J}' = (5, 5, 4, 3, 2)$ is minimal.
- Let $\lambda = \varpi_2 + \varpi_5$. In this case the hesitant λ-walk $(5, 5, 4, 3, 2)$ is not minimal since $\beta_0 = \beta_1 = \alpha_5$ already appears in λ. The subword $(5, 5)$ is minimal.

It is clear from the definition that for any dominant $\lambda \neq 0$ and a hesitant λ-walk $\mathcal{J}$, there exists a subword $\mathcal{J}'$ of $\mathcal{J}$ which is minimal in the sense of Definition 4.1.

Lemma 4.3. *Let $\lambda \neq 0$ be a dominant weight and $\mathcal{J} = (i_{j_0}, i_{j_1}, \ldots, i_{j_s})$ a hesitant λ-walk. Let $\boldsymbol{c}$ and ℓ be the constants associated to $\mathcal{J}$ and λ as defined in (1-6) and (1-7). If $\mathcal{J}$ is minimal, then*

(1) *$c_{j_p j_q} = 0$ if $|p-q| \geq 2$ and $1 \leq p, q \leq s$, and*

(2) *$\ell_{j_p} = 0$ for $0 \leq p \leq s-1$ if $s \geq 2$.*

$$A_r:\begin{bmatrix} 2 & -1 & 0 & & \cdots & & & 0\\ -1 & 2 & -1 & 0 & \cdots & & & 0\\ 0 & -1 & 2 & -1 & 0 & \cdots & \cdot & 0\\ \cdot & \cdot & \cdot & \cdot & \cdot\cdot\cdot\cdot & \cdot & \cdot & \cdot\\ 0 & 0 & 0 & & \cdots & -1 & 2 & -1\\ 0 & 0 & 0 & & \cdots & 0 & -1 & 2 \end{bmatrix} \qquad E_6:\begin{bmatrix} 2 & 0 & -1 & 0 & 0 & 0\\ 0 & 2 & 0 & -1 & 0 & 0\\ -1 & 0 & 2 & -1 & 0 & 0\\ 0 & -1 & -1 & 2 & -1 & 0\\ 0 & 0 & 0 & -1 & 2 & -1\\ 0 & 0 & 0 & 0 & -1 & 2 \end{bmatrix}$$

$$B_r:\begin{bmatrix} 2 & -1 & 0 & \cdots & & & 0\\ -1 & 2 & -1 & 0\cdots & & & 0\\ \cdot & \cdot & \cdot & \cdot\cdot\cdot\cdot & \cdot & \cdot & \cdot\\ 0 & 0 & 0 & \cdots & -1 & 2 & -2\\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix} \qquad E_7:\begin{bmatrix} 2 & 0 & -1 & 0 & 0 & 0 & 0\\ 0 & 2 & 0 & -1 & 0 & 0 & 0\\ -1 & 0 & 2 & -1 & 0 & 0 & 0\\ 0 & -1 & -1 & 2 & -1 & 0 & 0\\ 0 & 0 & 0 & -1 & 2 & -1 & 0\\ 0 & 0 & 0 & 0 & -1 & 2 & -1\\ 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{bmatrix}$$

$$C_r:\begin{bmatrix} 2 & -1 & 0 & \cdots & & & 0\\ -1 & 2 & -1 & 0\cdots & & & 0\\ \cdot & \cdot & \cdot & \cdot\cdot\cdot\cdot & \cdot & \cdot & \cdot\\ 0 & 0 & 0 & \cdots & -1 & 2 & -1\\ 0 & 0 & 0 & \cdots & 0 & -2 & 2 \end{bmatrix} \qquad E_8:\begin{bmatrix} 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0\\ 0 & 2 & 0 & -1 & 0 & 0 & 0 & 0\\ -1 & 0 & 2 & -1 & 0 & 0 & 0 & 0\\ 0 & -1 & -1 & 2 & -1 & 0 & 0 & 0\\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0\\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0\\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1\\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{bmatrix}$$

$$D_r:\begin{bmatrix} 2 & -1 & 0 & \cdot & \cdot & \cdot & & & 0\\ -1 & 2 & -1 & \cdot & \cdot & \cdot & & & 0\\ \cdot & \cdot & \cdot\cdot\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ 0 & 0 & & \cdot\cdot & -1 & 2 & -1 & 0 & 0\\ 0 & 0 & & \cdot\cdot & & -1 & 2 & -1 & -1\\ 0 & 0 & & \cdot\cdot & & 0 & -1 & 2 & 0\\ 0 & 0 & & \cdot\cdot & & 0 & -1 & 0 & 2 \end{bmatrix} \qquad F_4:\begin{bmatrix} 2 & -1 & 0 & 0\\ -1 & 2 & -2 & 0\\ 0 & -1 & 2 & -1\\ 0 & 0 & -1 & 2 \end{bmatrix} \qquad G_2:\begin{bmatrix} 2 & -1\\ -3 & 2 \end{bmatrix}$$

Table 2. Cartan matrices for all Lie types.

Proof. By the minimality assumption, and since Dynkin diagrams have no loops, we know that if $|p-q|\geq 2$ and $1\leq p,q\leq s$ (so j_p and j_q are in the walking component of $\mathcal{J}$) then the roots β_{j_p} are neither adjacent nor equal. This implies that the corresponding entry in the Cartan matrix is 0, as desired. The second statement is immediate from the minimality assumption since $\ell_{j_p}>0$ exactly when β_{j_p} appears in λ. $\square$

Lemma 4.4. *Let $\{c_{ij}\}_{1\leq i<j\leq n}$ and $\ell_1,\ldots,\ell_n$ be fixed integers and let $(C(\boldsymbol{c},\ell),\rho)$ be the corresponding Grossberg–Karshon twisted cube. Assume that $\ell_i\geq 0$ for all i. If there exist two distinct indices i and j, $1\leq i<j\leq n$, with $c_{ij}>1$ and $\ell_i=\ell_j>0$, then $(C(\boldsymbol{c},\ell),\rho)$ is twisted.*

Proof. By Theorem 1.4, it suffices to show that there exists an element σ of $\{+,-\}^n$ and some k with $1 \le k \le n$ such that $m_{\sigma,k} < 0$. Let $\sigma = (\sigma_1, \dots, \sigma_n) \in \{+,-\}^n$ be the element defined by

$$\sigma_k = \begin{cases} - & \text{if } k = i \text{ or } j, \\ + & \text{otherwise,} \end{cases}$$

and consider the associated $m_\sigma = (m_{\sigma,1}, \dots, m_{\sigma,n})$. Then by the definition of σ and m_σ we have

$$m_{\sigma,j} = \ell_j - \sum_{s>j} c_{js} m_{\sigma,s},$$

$$m_{\sigma,i} = \ell_i - \Biggl(c_{ij} m_{\sigma,j} - \sum_{\substack{s>i \\ s\neq j}} c_{is} m_{\sigma,s} \Biggr).$$

Since $\sigma_k = +$ for $k \neq i, j$, we have that $m_{\sigma,k} = 0$ for $k \neq i, j$. Hence the above equations can be simplified to

$$m_{\sigma,j} = \ell_j,$$

$$m_{\sigma,i} = \ell_i - c_{ij} m_{\sigma,j} = \ell_i - c_{ij} \ell_j.$$

By assumption $\ell_i = \ell_j$, so

$$m_{\sigma,i} = \ell_i (1 - c_{ij}).$$

Since $c_{ij} > 1$ and $\ell_i > 0$, we obtain $m_{\sigma,i} < 0$, as desired. □

As in the previous section, the above lemma can be interpreted in terms of hesitant λ-walks.

Corollary 4.5. *Let $\mathcal{I} = (i_1, i_2, \dots, i_n)$ be a word decomposition of an element $w = s_{i_1} \cdots s_{i_n}$ of W and let $\lambda = \lambda_1 \varpi_1 + \lambda_2 \varpi_2 + \cdots + \lambda_r \varpi_r$ be a dominant weight, i.e., $\lambda_i \ge 0$ for all i. Let $\boldsymbol{c} = \{c_{jk}\}$, $\ell = (\ell_1, \dots, \ell_n)$, and $\{m_\sigma\}_{\sigma \in \{+,-\}^n}$ be determined from $\mathcal{I}$ and λ as in (1-6), (1-7), and (1-4) and let $(C(\boldsymbol{c}, \ell), \rho)$ denote the corresponding Grossberg–Karshon twisted cube. If $\mathcal{I}$ contains a subword $\mathcal{J} = (j_0, j_1)$ of length 2 which is a hesitant λ-walk, then $(C(\boldsymbol{c}, \ell), \rho)$ is twisted.*

Proof. By the definition of hesitant λ-walk, if $\mathcal{J} = (j_0, j_1)$ is a hesitant λ-walk then $i_{j_0} = i_{j_1}$ (equivalently, $\beta_{j_0} = \beta_{j_1}$) and $\beta_{j_0} = \beta_{j_1}$ appears in λ. This implies $c_{j_0 j_1} = 2 > 1$ and $\ell_{j_0} = \ell_{j_1} > 0$. The result now follows from Lemma 4.4. □

Proof of the "only if" part of Theorem 2.9. Suppose $\mathcal{J} = \{i_{j_0}, i_{j_1}, \dots, i_{j_s}\}$ is a subword of $\mathcal{I}$ which is a hesitant λ-walk. We may without loss of generality assume that $\mathcal{J}$ is minimal in the sense of Definition 4.1. We then wish to show that $(C(\boldsymbol{c}, \ell), \rho)$ is twisted. If the length of $\mathcal{J}$ is 2, i.e., $s = 1$, then this follows from Corollary 4.5. Thus we may now assume that the length is at least 3, i.e., $s \ge 2$. To prove that $(C(\boldsymbol{c}, \ell), \rho)$ is twisted, by Theorem 1.4 it is enough to find an element σ

of $\{+,-\}^n$ and a $k \in [n]$ such that $m_{\sigma,k} < 0$. To achieve this, consider the element $\sigma = (\sigma_1, \ldots, \sigma_n) \in \{+,-\}^n$ defined by

$$\sigma_p = \begin{cases} - & \text{if } p \in \{j_0, j_1, \ldots, j_s\}, \\ + & \text{otherwise.} \end{cases}$$

By the definition of m_σ, we then have

$$\begin{aligned} m_{\sigma,j_s} &= \ell_{j_s} - \sum_{p>j_s} c_{j_s p} m_{\sigma,p}, \\ m_{\sigma,j_t} &= \ell_{j_t} - \Biggl(c_{j_t j_{t+1}} m_{\sigma,j_{t+1}} + \sum_{\substack{p>j_t \\ p\neq j_{t+1}}} c_{j_t p} m_{\sigma,p} \Biggr) \quad \text{for } 1 \le t \le s-1, \\ m_{\sigma,j_0} &= \ell_{j_0} - \Biggl(c_{j_0 j_1} m_{\sigma,j_1} + c_{j_0 j_2} m_{\sigma,j_2} + \sum_{\substack{p>j_0 \\ p\neq j_1,j_2}} c_{j_0 p} m_{\sigma,p} \Biggr). \end{aligned} \tag{4-1}$$

Since $\mathcal{J}$ is a hesitant λ-walk, we know $\ell_{j_s} > 0$. On the other hand, by the minimality assumption on $\mathcal{J}$ and Lemma 4.3, we know $\ell_{j_t} = 0$ for all t with $0 \le t \le s-1$. Moreover, again by minimality and Lemma 4.3, we know that $c_{j_t j_r} = 0$ for $j_r > j_t$ and $j_r \neq j_{t+1}$. Also, by construction of the σ, for $p \notin \mathcal{J} = \{j_0, j_1, \ldots, j_s\}$ we have $\sigma_p = +$ and hence $m_{\sigma,p} = 0$. Finally, since $\mathcal{J}$ is a hesitant λ-walk, we have $\beta_{j_0} = \beta_{j_1}$ and hence $c_{j_0 j_1} = \langle \beta_{j_0}, \beta_{j_1}^\vee \rangle = 2$. From these considerations we can simplify (4-1):

$$\begin{aligned} m_{\sigma,j_s} &= \ell_{j_s} > 0, \\ m_{\sigma,j_t} &= -c_{j_t j_{t+1}} m_{\sigma,j_{t+1}} \quad \text{for } 1 \le t \le s-1, \\ m_{\sigma,j_0} &= -(2m_{\sigma,j_1} + c_{j_0 j_2} m_{\sigma,j_2}). \end{aligned} \tag{4-2}$$

We now claim that $m_{\sigma,j_0} < 0$; as already noted, this suffices to prove the theorem. In order to prove this claim we need to know the values of the constants $c_{j_t j_{t+1}}$ and $c_{j_0 j_2}$ appearing in (4-2). By the assumption that $\mathcal{J}$ is a hesitant λ-walk, these constants are equal to the corresponding entry of the Cartan matrices for simple roots which are adjacent in the Dynkin diagram. For the case-by-case analysis below we refer to the list of Dynkin diagrams and Cartan matrices in Tables 1 and 2. Suppose first that the hesitant λ-walk only crosses edges of the form $\circ\!\!-\!\!\!-\!\!\!-\!\!\circ$ or that if it crosses a double edge $\circ\!\!=\!\!\!=\!\!\circ$ or triple edge $\circ\!\!\equiv\!\!\!\equiv\!\!\circ$ then it does so only by going in the direction *agreeing with* the arrow drawn on the edge in the Dynkin diagram (e.g., in type B, if $i_{j_t} = r-1$ and $i_{j_{t+1}} = r$, and in type G, if $i_{j_t} = 2$ and $i_{j_{t+1}} = 1$). In this situation, the corresponding constants $c_{j_t j_{t+1}}$ and $c_{j_0 j_2}$ are all equal to -1. So we consider this case first. In this setting we have

$$
\begin{aligned}
&m_{\sigma,j_s} = \ell_{j_s} > 0, \\
\text{(4-3)} \qquad &m_{\sigma,j_t} = m_{\sigma,j_{t+1}} \quad \text{for } 1 \le t \le s-1, \\
&m_{\sigma,j_0} = -(2m_{\sigma,j_1} - m_{\sigma,j_2}),
\end{aligned}
$$

so $m_{\sigma,j_1} = m_{\sigma,j_2} = \cdots = m_{\sigma,j_s} = \ell_{j_s}$ and $m_{\sigma,j_0} = -\ell_{j_s} < 0$, as desired.

Next we consider the possibility that the hesitant λ-walk crosses a double edge in a direction *against* the direction of the arrow on the edge. Since we assume the hesitant λ-walk is minimal, it can only cross such an edge once. In particular, in type B this implies that the hesitant λ-walk must be of the form $i_{j_0} = i_{j_1} = r$ and $i_{j_2} = r-1$, $i_{j_3} = r-2, \ldots, i_{j_s} = r-s+1$, while in type C it must be of the form $i_{j_0} = i_{j_1} = r-s+1$, $i_{j_2} = r-s+2, \ldots, i_{j_{s-1}} = r-1$ and $i_{j_s} = r$, for some $s \ge 2$. We consider these cases next.

In type B consider the hesitant λ-walk of the form $i_{j_0} = i_{j_1} = r$ and $i_{j_2} = r-1$, $i_{j_3} = r-2, \ldots, i_{j_s} = r-s+1$ for some $s \ge 2$. In this case the equations (4-2) become

$$
\begin{aligned}
m_{\sigma,j_s} &= \ell_{j_s} > 0, \\
m_{\sigma,j_{s-1}} &= \cdots = m_{\sigma,j_2} = \ell_{j_s}, \\
m_{\sigma,j_1} &= 2m_{\sigma,j_2} = 2\ell_{j_s}, \\
m_{\sigma,j_0} &= -(2m_{\sigma,j_1} + (-2)m_{\sigma,j_2}) = -2\ell_{j_s} < 0,
\end{aligned}
$$

so we obtain $m_{\sigma,j_0} < 0$, as desired. In type C, consider the hesitant λ-walk $i_{j_0} = i_{j_1} = r-s+1$, $i_{j_2} = r-s+2, \ldots, i_{j_{s-1}} = r-1$ and $i_{j_s} = r$ for $s \ge 2$. Note that the case $s = 2$ is already covered in the argument for type B above, so we may assume $s \ge 3$. It is straightforward to see that here we obtain from (4-2) that $m_{\sigma,j_s} = \ell_{j_s} > 0$, $m_{\sigma,j_{s-1}} = \cdots = m_{\sigma,j_1} = 2\ell_{j_s}$, and $m_{\sigma,j_0} = -2\ell_{j_s} < 0$. Thus $m_{\sigma,j_0} < 0$, as desired.

The only remaining cases are in the exceptional Lie types F and G, but many cases of hesitant λ-walks in type F are already handled by the considerations for types B and C above. Thus the only remaining cases are $(4, 4, 3, 2, 1)$ in type F and $(1, 1, 2)$ in type G. Both are straightforward and left to the reader. □

5. Open questions

The study of Grossberg–Karshon twisted cubes is related to representation theory and to the recent theory of Newton–Okounkov bodies and divided-difference operators on polytopes. In this paper we have introduced the notion of hesitant λ-walks as well as hesitant-λ-walk-avoidance. Below, we briefly mention some possible avenues for further exploration.

(1) The Grossberg–Karshon twisted cubes are a special case of the virtual polytopes produced by Kiritchenko's divided-difference operators [Kiritchenko

2013]. We may ask whether our methods generalize to Kiritchenko's setting to provide combinatorial conditions on a dominant weight λ and choice of word decomposition $\mathcal{I}$ which guarantee that the corresponding virtual polytope from Kiritchenko's construction is a true polytope. (See also Kiritchenko's discussion in [2013, §3.3].)

(2) In the cases when the Grossberg–Karshon twisted polytope is untwisted (i.e., it is a true polytope), it would be of interest to study the relationship between the Grossberg–Karshon polytope and other polytopes appearing in representation theory and Schubert calculus, such as Gel'fand–Cetlin polytopes, or (more generally) string polytopes, or (even more generally) Newton–Okounkov bodies of Bott–Samelson varieties (see [Kaveh 2011; Anderson 2013; Harada and Yang $\geq$ 2015]).

(3) Pattern avoidance is a recurring and important theme in the study of Schubert varieties. We may ask whether, and how, hesitant-λ-walk-avoidance relates to the known results in this direction [Abe and Billey 2014].

Acknowledgements

We thank Jihyeon Jessie Yang for useful conversations and Professor Dong Youp Suh for his support throughout the project. We are grateful to the anonymous referee for a careful reading of our initial manuscript and for improving the paper by providing the comments recorded in Remark 2.10. Harada was partially supported by an NSERC Discovery Grant (Individual), an Ontario Ministry of Research and Innovation Early Researcher Award, a Canada Research Chair (Tier 2) Award, and a Japan Society for the Promotion of Science Invitation Fellowship for Research in Japan (Fellowship ID L-13517). Lee was supported by the National Research Foundation of Korea Grant No. NRF-2013R1A1A2007780 funded by the Korean government (MOE). Harada additionally thanks the Osaka City University Advanced Mathematics Institute for its hospitality while part of this research was conducted.

References

[Abe and Billey 2014] H. Abe and S. Billey, "Consequences of the Lakshmibai–Sandhya theorem: the ubiquity of permutation patterns in Schubert calculus and related geometry", preprint, 2014. arXiv 1403.4345

[Anderson 2013] D. Anderson, "Okounkov bodies and toric degenerations", *Math. Ann.* **356**:3 (2013), 1183–1202. MR 3063911 Zbl 1273.14104

[Grossberg and Karshon 1994] M. Grossberg and Y. Karshon, "Bott towers, complete integrability, and the extended character of representations", *Duke Math. J.* **76**:1 (1994), 23–58. MR 96i:22030 Zbl 0826.22018

[Harada and Yang 2015] M. Harada and J. J. Yang, "Grossberg–Karshon twisted cubes and basepoint-free divisors", *J. Korean Math. Soc.* **52**:4 (2015), 853–868.

[Harada and Yang ≥ 2015] M. Harada and J. J. Yang, "Newton–Okounkov bodies of Bott-Samelson varieties", in preparation.

[Humphreys 1972] J. E. Humphreys, *Introduction to Lie algebras and representation theory*, Graduate Texts in Mathematics **9**, Springer, New York, 1972. MR 48 #2197 Zbl 0254.17004

[Karuppuchamy 2013] P. Karuppuchamy, "On Schubert varieties", *Comm. Algebra* **41**:4 (2013), 1365–1368. MR 3044412 Zbl 1277.14039

[Kaveh 2011] K. Kaveh, "Crystal bases and Newton–Okounkov bodies", preprint, 2011. To appear in *Duke Math. J.* arXiv 1101.1687

[Kiritchenko 2013] V. Kiritchenko, "Divided difference operators on polytopes", preprint, 2013. arXiv 1307.7234

[Lakshmibai and Sandhya 1990] V. Lakshmibai and B. Sandhya, "Criterion for smoothness of Schubert varieties in $\mathrm{Sl}(n)/B$", *Proc. Indian Acad. Sci. Math. Sci.* **100**:1 (1990), 45–52. MR 91c:14061 Zbl 0714.14033

Received October 3, 2014. Revised March 3, 2015.

MEGUMI HARADA
DEPARTMENT OF MATHEMATICS AND STATISTICS
MCMASTER UNIVERSITY
1280 MAIN STREET WEST
HAMILTON, ON L8S 4K1
CANADA
megumi.harada@math.mcmaster.ca

EUNJEONG LEE
DEPARTMENT OF MATHEMATICAL SCIENCES
KAIST
291 DAEHAK-RO YUSEONG-GU
DAEJEON 305-701
SOUTH KOREA
eunjeonglee@kaist.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.278.137

GAMMA FACTORS OF DISTINGUISHED REPRESENTATIONS OF $\mathrm{GL}_n(\mathbb{C})$

ALEXANDER KEMARSKY

Let (π, V) be a $\mathrm{GL}_n(\mathbb{R})$-distinguished, irreducible, admissible representation of $\mathrm{GL}_n(\mathbb{C})$, let π' be an irreducible, admissible, $\mathrm{GL}_m(\mathbb{R})$-distinguished representation of $\mathrm{GL}_m(\mathbb{C})$, and let ψ be a nontrivial character of $\mathbb{C}$ which is trivial on $\mathbb{R}$. We prove that the Rankin–Selberg gamma factor at $s = 1/2$ is $\gamma(1/2, \pi \times \pi'; \psi) = 1$. The result follows as a simple consequence from the characterization of $\mathrm{GL}_n(\mathbb{R})$-distinguished representations in terms of their Langlands data.

1.	Introduction	137
2.	Notation and preliminaries	140
3.	Some matrix spaces decompositions	142
4.	Proof of Theorem 1.2	145
5.	Calculation of Rankin–Selberg gamma factors	149
6.	Integral representation of Whittaker functions	154
7.	Archimedean Asai integrals	160
8.	Equality of two functionals	162
Appendix A.	Generic Langlands quotient	165
Appendix B.	Gamma factors: converse direction	166
Acknowledgements		170
References		170

1. Introduction

Let $G_n(\mathbb{C}) = \mathrm{GL}_n(\mathbb{C})$, $G_n(\mathbb{R}) = \mathrm{GL}_n(\mathbb{R})$. Let $B_n = B_n(\mathbb{C})$ be the Borel subgroup of upper triangular matrices in $G_n(\mathbb{C})$. Denote the complex conjugation by $x \to \bar{x}$. We identify $G_n(\mathbb{C})/G_n(\mathbb{R})$ with the space of matrices

$$X_n = \{x \in G_n(\mathbb{C}) : x \cdot \bar{x} = I_n\},$$

This research was supported by ISF grant No. 1394/12.
MSC2010: primary 22E30; secondary 20G05, 11F70.
Keywords: Rankin–Selberg gamma factor, distinguished representations, Langlands classification, Whittaker model.

via the isomorphism $gG_n(\mathbb{R}) \mapsto g \cdot \bar{g}^{-1}$. See [Serre 2002, Chapter 3, Section 1, Lemma 1] for the proof of the surjectivity of this map. Given a representation π of $G_n(\mathbb{C})$, the representation $\bar{\pi}$ is defined by the formula $\bar{\pi}(g) := \pi(\bar{g})$.

The group $G_n(\mathbb{C})$ acts on X_n by the twisted conjugation, where the action is induced by the natural action $l(g)g'G_n(\mathbb{R}) := gg'G_n(\mathbb{R})$. Namely, we have

$$g'G_n(\mathbb{R}) \leftrightarrow g' \cdot \bar{g'}^{-1} := x \quad \text{and} \quad l(g)(g'G_n(\mathbb{R})) := gg'G_n(\mathbb{R}) \leftrightarrow gg'\bar{g'}^{-1}\bar{g}^{-1}.$$

Hence, the action of $G_n(\mathbb{C})$ on X is given by $l(g)x := gx\bar{g}^{-1}$.

For a topological vector space V, we denote by V^* the topological dual of V, i.e., the space of all continuous maps from V to $\mathbb{C}$. In this paper we work with the category of the admissible smooth Fréchet representations of moderate growth (see [Wallach 1992, Section 11.5; Aizenbud et al. 2008, Section 2.1]).

A representation (π, V) of $G_n(\mathbb{C})$ is called $G_n(\mathbb{R})$-distinguished if there exists a nonzero continuous linear map $L : V \to \mathbb{C}$ such that

$$L(\pi(h)v) = L(v) \quad \text{for all } v \in V,\ h \in G_n(\mathbb{R}).$$

We denote the space of all such linear maps by $(V^*)^{G_n(\mathbb{R})}$. We denote the set of equivalence classes of irreducible representations of $G_n(\mathbb{C})$ by $\mathrm{Irr}(G_n(\mathbb{C}))$ and the set of equivalence classes of irreducible $G_n(\mathbb{R})$-distinguished representations of $G_n(\mathbb{C})$ by $\mathrm{Irr}_{G_n(\mathbb{R})}(G_n(\mathbb{C}))$.

Let $\psi : \mathbb{C} \to \mathbb{C}^\times$ be a nontrivial unitary character which is trivial on $\mathbb{R}$, for example

$$\psi(x) = e^{\pi(x-\bar{x})}.$$

We let $U_n(\mathbb{C})$ be the group of upper triangular matrices with unit diagonal and we denote by $\theta_{\psi,n}$ the character $\theta_{\psi,n} : U_n(\mathbb{C}) \to \mathbb{C}^\times$ defined by

$$\theta_{\psi,n}(u) = \psi\left(\sum_{i=1}^{n-1} u_{i,i+1}\right).$$

A ψ-form on V is a nonzero continuous linear form $\lambda : V \to \mathbb{C}$ such that

$$\lambda(\pi(u)v) = \theta_{\psi,n}(u)v,$$

for each $v \in V$ and each $u \in U_n(\mathbb{C})$. We say that π is a generic representation if there exists a ψ-form on V.

Theorem 1.1. *Let $\pi \in \mathrm{Irr}_{G_t(\mathbb{R})}(G_t(\mathbb{C}))$ and let $\pi' \in \mathrm{Irr}_{G_r(\mathbb{R})}(G_r(\mathbb{C}))$. If ψ is a nontrivial character of $\mathbb{C}$ with a trivial restriction to $\mathbb{R}$ then the value of the Rankin–Selberg gamma factor at $s = 1/2$ is*

$$\gamma\left(\frac{1}{2}, \pi \times \pi'; \psi\right) = 1.$$

A similar theorem is proved in [Offen 2011, Theorem 0.1] for the p-adic case (see also [Ok 1997]). See Section 5 for the definition of Rankin–Selberg integrals, Rankin–Selberg gamma factors and for the proof of Theorem 1.1.

We will deduce Theorem 1.1 from the characterization of irreducible $G_n(\mathbb{R})$-distinguished representations of $G_n(\mathbb{C})$. Let χ be a character of B_n. We denote by $I(\chi)$ the normalized parabolic induction representation

$$I(\chi) := \mathrm{Ind}_{B_n}^{G_n(\mathbb{C})}(\chi)$$

of the character $\chi = (\chi_1, \dots, \chi_n)$ from B_n to $G_n(\mathbb{C})$. We remind the reader that this space consists of smooth functions such that $f(bg) = (\chi\delta^{1/2})(b)f(g)$ for all $b \in B_n$ and all $g \in G_n(\mathbb{C})$. The group $G_n(\mathbb{C})$ acts on $I(\chi)$ by right translations, and the group of permutations on n elements, S_n, acts naturally on the characters of B_n. We will call a character $\chi = (\chi_1, \chi_2, \dots, \chi_n)$ of B_n dominant if

$$|\chi(t)| = |t_1|^{\lambda_1}|t_2|^{\lambda_2}\cdots|t_n|^{\lambda_n} \quad \text{with } \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

In Section 4 we will prove the following.

Theorem 1.2. *Let π be an element of $\mathrm{Irr}_{G_n(\mathbb{R})}(G_n(\mathbb{C}))$ and let $\chi = (\chi_1, \chi_2, \dots, \chi_n)$ be a dominant character of B_n. Suppose π is the Langlands quotient of $I(\chi)$, that is, the unique irreducible quotient of $I(\chi)$. Then there exists an involution $w \in S_n$ such that $w\chi = \overline{(\chi^{-1})}$. Moreover, we can choose this w such that for every fixed point i of w we have $\chi_i(-1) = 1$.*

Remark 1.3. Note that the conditions $w(i) = i$, $\chi_{w(i)} = \overline{\chi_i}^{-1}$ and $\chi_i(-1) = 1$ imply that χ_i is $\mathrm{GL}_1(\mathbb{R})$-distinguished. Indeed, $\chi_i = \overline{\chi_i}^{-1}$ implies that χ_i is $\mathbb{R}_+$-invariant. Together with the condition $\chi_i(-1) = 1$ this means that χ_i is $\mathbb{R}^\times$-invariant (i.e., $\mathrm{GL}_1(\mathbb{R})$-invariant).

As a consequence of Theorem 1.2 we obtain the following analogue of [Aizenbud and Lapid 2012, Theorem B.1].

Theorem 1.4. *Let $\pi \in \mathrm{Irr}_{G_n(\mathbb{R})}(G_n(\mathbb{C}))$ and suppose π is a generic representation of $G_n(\mathbb{C})$. Then $\bar{\pi} \simeq \tilde{\pi}$, where $\tilde{\pi}$ is the contragredient representation of π.*

Let (π, V) be an irreducible representation of $G_n(\mathbb{C})$. The existence of $I(\chi)$ with the properties stated in Theorem 1.2 is a well-known fact (see [Wallach 1988, Theorem 5.4.1]). It is also well-known that the Langlands quotient of $I(\chi)$ is generic if and only if $I(\chi)$ is irreducible (see Appendix A). Therefore, $\bar{\pi} = I(\bar{\chi})$ and $\tilde{\pi} = I(\chi^{-1})$. Since $\bar{\pi}$ is irreducible, for every $w_0 \in S_n$ we have $I(w_0(\bar{\chi})) \simeq I(\bar{\chi})$. In particular, for $w \in S_n$ such that $w(\bar{\chi}) = \chi^{-1}$, we have

$$\bar{\pi} \simeq I(\bar{\chi}) \simeq I(w(\bar{\chi})) \simeq I(\chi^{-1}) \simeq \tilde{\pi}.$$

A similar result was proved by Marie-Noelle Panichi in her Ph.D. thesis [2001, Theorem 3.3.6].

The structure of the paper is as follows. In Section 3 we recall basic facts about the structure of $G_n(\mathbb{C})$. In Section 4 we prove Theorem 1.2 by analyzing the geometry of the action of B_n on the variety $G_n(\mathbb{C})/G_n(\mathbb{R})$. In Section 5, as an application of our classification we deduce Theorem 1.1. In Section 8 we prove a new type of integral identity for Whittaker functions on generic $G_n(\mathbb{R})$-distinguished representations which in turn proves [Lapid and Mao 2014, Assumption 5.2]. A similar identity was proved in the p-adic case in [Offen 2011, Corollary 7.2]. Our proof is similar to the proof in the p-adic case, but in the archimedean case there are many analytical difficulties. We overcome them in Sections 5–7.

Finally, in Appendix B we prove a converse-type theorem. We prove that if $(\pi, V) = I(\chi)$ is an irreducible, generic, admissible unitary representation of $G_n(\mathbb{C})$ such that for every unitary character $\chi'(z) = (z/|z|)^{2m}$ with $m \in \mathbb{Z}$ we have

$$\gamma\left(\frac{1}{2}, \pi \times \chi', \psi\right) = 1,$$

then π is $G_n(\mathbb{R})$-distinguished. The proof is done by a combinatorial argument combined with the Tadic–Vogan classification of the unitary dual of $G_n(\mathbb{C})$.

2. Notation and preliminaries

Let $M(a \times b, F)$ be the space of matrices with a rows and b columns with entries in F, where F is either $\mathbb{R}$ or $\mathbb{C}$. Let $\eta_n = (0, 0, \dots, 1)$ be an element of $M(1 \times n, \mathbb{R})$, and let $P_n(\mathbb{R})$ be the subgroup of $G_n(\mathbb{R})$ consisting of all $n \times n$ matrices with the last row equal to η_n.

Let $U_n(F)$ be the group of all upper triangular matrices in $M(n \times n, F)$ with unit diagonal. Let

$$K_n = \{g \in G_n(\mathbb{C}) : g \cdot {}^t\bar{g} = I\}$$

be the standard maximal compact subgroup of $G_n(\mathbb{C})$.

For $g \in G_n(\mathbb{C})$, define

$$\|g\| := \sqrt{\sum_{i,j=1}^{n} |g_{ij}|^2} \quad \text{and} \quad \|g\|_H := \max\,(\|g\|, \|g\|^{-1}).$$

The value $\|g\|_H$ is called a norm on g (see [Wallach 1988, Section 2.A.2] for a general discussion on norms on a reductive group).

Let G be a group and H its subgroup. We say that a function $f : G \to \mathbb{C}$ is H-finite if the dimension of the space spanned by right H-translations of f is finite. In this work we will often consider K_n-finite functions on $G_n(\mathbb{C})$.

For V, a finite dimensional vector space over $\mathbb{R}$, we denote by $\mathcal{S}(V)$ the Schwartz space of all infinitely differentiable functions $f : V \to \mathbb{C}$ of rapid decay.

Let $\Phi \in \mathcal{S}(V)$, where $V = M(a \times b, \mathbb{C})$. We denote by $\hat{\Phi}$ the Fourier transform of Φ. It is a function on the same space, defined by

$$\hat{\Phi}(X) = \int \Phi(Y)\psi(-\mathrm{Tr}({}^tXY))\,dY.$$

For $\Phi \in \mathcal{S}(\mathbb{C}^n)$ and $g \in G_n(\mathbb{C})$ we denote by $(R(g)\Phi)(x) := \Phi(xg)$ the right translation of Φ by g.

For $z = x + iy \in \mathbb{C}$ we denote by $|z| = \sqrt{x^2 + y^2}$ the usual absolute value of z and by $|z|_{\mathbb{C}} = |z|^2 = x^2 + y^2$ the square of the usual absolute value. Note that $\mu(zA) = |z|_{\mathbb{C}}\mu(A)$, where $A \subset \mathbb{C}$ is an open set and μ is a Haar measure on $\mathbb{C}$.

Let W_n equal S_n and let $W_{n,2} = \{w \in W_n : w^2 = 1\}$ be the set of involutions in W_n. For $w \in W_{n,2}$ set

$$I_w = \{(i, j) : i > j,\ w(i) > w(j)\},$$

and define for any function $\kappa : I_w \to \mathbb{Z}_{\geq 0}$ a character α_κ of B_n by the formula

$$\alpha_\kappa(\mathrm{diag}(t_1, \dots, t_n)) = \prod_{(i,j)\in I_w} \left[\frac{t_i}{t_j}\right]^{\kappa(i,j)}.$$

We will identify α_κ with the one-dimensional representation of B_n on the vector space $\mathbb{C}$ with the action of α_κ. By abuse of notation we will denote both the function and the one-dimensional representation by the same symbol, α_κ.

For the convenience of the reader we write here notation and formulations of some of the theorems that appear in [Aizenbud and Lapid 2012], in versions that are suitable for this work.

Let G be an arbitrary group.

- For any G-set X and a point $x \in X$, we denote by $G(x)$ the G-orbit of x and by G^x the stabilizer of x.
- For any representation of G on a vector space V and a character χ of G, we denote by $V^{G,\chi}$ the subspace of (G, χ)-equivariant vectors in V.
- Given manifolds $L \subseteq M$, we denote by $N_L^M := (T_{M|L})/T_L$ the normal bundle to L in M and by $CN_L^M := (N_L^M)^*$ the conormal bundle. For any point $y \in L$, we denote by $N_{L,y}^M$ the normal space to L in M at the point y and by $CN_{L,y}^M$ the conormal space to L in M at the point y.
- The symmetric algebra of a vector space V is denoted by

$$\mathrm{Sym}(V) = \bigoplus_{k\geq 0} \mathrm{Sym}^k(V).$$

- The Fréchet space of Schwartz functions on a Nash manifold X is denoted by $\mathcal{S}(X)$ and the dual space of Schwartz distributions by $\mathcal{S}^*(X) := \mathcal{S}(X)^*$.
- For any Nash vector bundle E over X we denote by $\mathcal{S}(X, E)$ the space of Schwartz sections of E and by $\mathcal{S}^*(X, E)$ its dual space.

See [Aizenbud and Lapid 2012, p. 309] for more details.

Suppose X is a smooth manifold with G acting on X. Recall that $X = \bigcup_{i=1}^{l} X_i$ is called a G-invariant stratification if all sets X_i are G-invariant and there is some reordering $X_{i_1}, X_{i_2}, \ldots, X_{i_l}$ of $X_1, \ldots, X_l$ such that all the sets

$$X_{i_1}, X_{i_1} \cup X_{i_2}, \ldots, X_{i_1} \cup X_{i_2} \cup \cdots \cup X_{i_k}, \ldots, X = X_{i_1} \cup \cdots \cup X_{i_l}$$

are open in X.

Lemma 2.1. *Let a real algebraic Lie group G act on a real algebraic smooth manifold X. Let $X = \bigcup_{i=1}^{l} X_i$ be a G-invariant stratification, and let χ be a character of G. If*

$$\mathcal{S}^*(X)^{G,\chi} \neq 0,$$

then there exist an $1 \leq i \leq l$ and $k \geq 0$ such that

$$\mathcal{S}^*(X_i, \mathrm{Sym}^k(CN_{X_i}^X))^{G,\chi} \neq 0.$$

This lemma is a special case of [Aizenbud and Lapid 2012, Proposition B.3].

Theorem 2.2 [Aizenbud and Lapid 2012, Theorem B.6]. *Let G be a real algebraic group acting transitively on a real algebraic smooth manifold Z and let $\varphi : X \to Z$ be a G-equivariant smooth map. Fix $z \in Z$ and let X_z be the fiber of z. Let χ be a tempered character of G* [Aizenbud et al. 2008, Definition 5.1.1], *and let δ_G and δ_{G_z} be the modulus characters of the groups G and G_z respectively. Then $S^*(X)^{G,\chi}$ is canonically isomorphic to $S^*(X_z)^{G_z,\chi\delta_{G_z}^{-1}\delta_G}$.*

Moreover, for any G-equivariant bundle E on X, the space $S^*(X, E)^{G,\chi}$ is canonically isomorphic to $S^*(X_z, E|X_z)^{G_z,\chi\delta_{G_z}^{-1}\delta_G}$.

3. Some matrix spaces decompositions

In this section we obtain some matrix space decompositions that will be used in this work. In the following lemma we analyze the structure of orbits of the action of the Borel subgroup B_n on X_n. Let $W_n = S_n$ be the Weyl group of $G_n(\mathbb{C})$.

Lemma 3.1. *There is a bijection between $B_n \backslash X_n = B_n \backslash G_n(\mathbb{C})/G_n(\mathbb{R})$ and the space of involutions $W_{n,2} = \{w \in W_n : w^2 = 1\}$.*

Proof. Recall that $X_n = \{x \in G_n(\mathbb{C}) : x \cdot \bar{x} = I\}$ and let $x \in X_n$. Let

$$T = \{\mathrm{diag}(d_1, \ldots, d_n) : d_i \in \mathbb{C}^* \text{ for all } i\}$$

be a maximal torus in $G_n(\mathbb{C})$. From [Lapid and Rogawski 2003, Lemma 4.1.1] (see also [Springer 1984]), the B_n-orbit of x intersects the normalizer

$$N(T) := \{g \in G_n(\mathbb{C}) : gTg^{-1} = T\}.$$

It is a well-known fact that $N(T) = \{d \cdot w : d \in T,\ w \in W_n\}$. Thus, we may assume $x = dw$, where $w \in W_n$ and $d = \mathrm{diag}(d_1, d_2, \ldots, d_n)$. Note that w is uniquely determined by x. Since $x \cdot \bar{x} = I$, we have $dw = w^{-1}\bar{d}^{-1}$. We obtain $w = w^{-1}$ and therefore $w^2 = 1$, i.e., $w \in W_{n,2}$.

Therefore, we can assume that $w \in W_{n,2}$ in the decomposition $x = dw$. On the other hand, it is clear that different involutions $w, w' \in W_{n,2}$ belong to disjoint orbits of B_n. Indeed, $l(b)w := bw\bar{b}^{-1} \neq w'$ for all $b \in B_n$.

It remains to show that the B_n-orbit of $x = dw$ contains the point w, i.e., there is some $b \in B_n$ such that $l(b)x = w$. Since w is an involution it is enough to check the claim for 1×1 and 2×2 matrices. For the 1×1 matrix $x = (b)_{1\times 1}$, the assumption $x\bar{x} = I$ gives $b\bar{b} = 1$, and we want to prove that $b = \mu\bar{\mu}^{-1}$. Clearly, there is such a μ.

For a 2×2 matrix of the form $b = \left(\begin{smallmatrix} d_1 & 0 \\ 0 & d_2 \end{smallmatrix}\right)$, the assumption

$$x = b\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & d_1 \\ d_2 & 0 \end{pmatrix} \in X$$

gives the condition $d_1\bar{d}_2 = 1$ on the entries d_1, d_2. We seek an invertible matrix $\left(\begin{smallmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{smallmatrix}\right)$ such that

$$\begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}\begin{pmatrix} 0 & d_1 \\ d_2 & 0 \end{pmatrix}\begin{pmatrix} \overline{\mu_1}^{-1} & 0 \\ 0 & \overline{\mu_2}^{-1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{3-1}$$

Matrix multiplication gives the condition $d_1\mu_1\overline{\mu_2}^{-1} = 1$, and clearly there are such a μ_1 and μ_2. $\square$

In the next paragraph let us fix n and define $G = G_n(\mathbb{C}),\ H = G_n(\mathbb{R})$. Our goal is to obtain a generalized Cartan decomposition $G = KAH$, where K is a maximal compact subgroup of G consisting of all unitary matrices in G and A is a torus which we will now describe. Let $m = [n/2]$. Note that $H = G^\sigma$ and $K = G^\tau$, where $\sigma(g) = \bar{g}$ and $\tau(g) = g^* = {}^t\bar{g}$. Let $\mathfrak{g}$ be the Lie algebra of G over the field $\mathbb{C}$. Following [Kobayashi 2007, Fact 2.1, p. 7], we take $\mathfrak{a}$ to be a maximal abelian subspace in

$$\mathfrak{g}^{-\sigma,-\tau} = \{X \in \mathfrak{g} : \tau X = \sigma X = -X\}.$$

Following this recipe, let us define

$$\mathfrak{a} = \sum_{j=1}^{m} i\mathbb{R}(E_{2j+1,2j} - E_{2j,2j+1}).$$

Recall that

$$\exp\begin{pmatrix} 0 & it \\ -it & 0 \end{pmatrix} = \begin{pmatrix} \mathrm{ch}(t) & i\,\mathrm{sh}(t) \\ -i\,\mathrm{sh}(t) & \mathrm{ch}(t) \end{pmatrix}$$

and let A be the Lie group corresponding to $\mathfrak{a}$. Denote by $a(t_1, t_2, \ldots, t_m)$ the $n \times n$ matrix which consists of m 2×2 diagonal blocks of the form $\exp\left(\begin{smallmatrix} 0 & it_j \\ -it_j & 0 \end{smallmatrix}\right)$, where $j = 1, 2, \ldots, m$ if $n = 2m$ is even, and which consists of these blocks and $a_{nn} = 1$ in the last diagonal place if $n = 2m+1$ is odd. For example, if $n = 4$ then

$$a(t_1, t_2) = \begin{pmatrix} \cosh(t_1) & i\sinh(t_1) & 0 & 0 \\ -i\sinh(t_1) & \cosh(t_1) & 0 & 0 \\ 0 & 0 & \cosh(t_2) & i\sinh(t_2) \\ 0 & 0 & -i\sinh(t_2) & \cosh(t_2) \end{pmatrix}.$$

We have

$$A = \{a(t_1, t_2, \ldots, t_m) : t_1, t_2, \ldots, t_m \in \mathbb{R}\}.$$

Define

$$A^+ = \{a(t_1, t_2, \ldots, t_m) : t_1 \geq t_2 \geq \cdots \geq t_m \geq 0\}.$$

Theorem 3.2. *There is a decomposition $G = KA^+H$. That is, every element $g \in G$ can be written as*

$$g = kah, \quad \text{where } k \in K,\ a \in A^+,\ h \in H. \tag{3-2}$$

Moreover, the $a \in A^+$ in decomposition (3-2) is uniquely determined by g.

Remark 3.3. By taking the transpose of (3-2) we obtain a similar decomposition $G = HA^+K$. That is, every $g \in G$ can be written as

$$g = hak, \quad \text{where } h \in H,\ a \in A^+,\ k \in K, \tag{3-3}$$

and $a \in A^+$ in this decomposition is uniquely determined by g. Actually, after taking the transpose of (3-2), we obtain that $a^t \in A$ and in general $a^t \notin A^+$. But the permutation group S_n is naturally contained in both K and H and we can replace $a \in A$ with $a' = w_1 a w_2$ such that $a' \in A^+$.

Proof of Theorem 3.2. To prove the existence part we will show that $G = KAH$. Since permutation matrices are clearly in $H \cap K$, the equality $G = KA^+H$ will easily follow from the equality $G = KAH$. Let $g \in G$. Our goal is to achieve a decomposition $g = kah$ with $h \in H$, $a \in A$, and $k \in K$. Suppose that g is of such a form. Then, since $h^* = {}^t h$, $a^* = a$, and $k^* = k^{-1}$ we get

$$g^* g = {}^t h a^2 h, \quad h \in H,\ a \in A. \tag{3-4}$$

On the other hand, suppose that every matrix of the form $g^* g$ can be written as (3-4). Then write $g = ((g^*)^{-1}\,{}^t h a) a h$, and let us show that $k = (g^*)^{-1}\,{}^t h a$ is a unitary matrix. Indeed,

$$k^*k = a^*({}^th)^*((g^*)^{-1})^*(g^*)^{-1}\,{}^tha = ah(g^*g)^{-1}\,{}^tha = ah({}^tha^2h)^{-1}\,{}^tha = I.$$

Therefore, to prove the existence part of the theorem, it is enough to prove that every matrix of the form g^*g can be written in the form (3-4). For this purpose write $g^*g = x+iy,\ x, y \in H$. Then $x = {}^tx$ is symmetric, and $y = -{}^ty$ is antisymmetric. Also, ${}^tvg^*gv > 0$ for every $0 \neq v \in \mathbb{R}^n$. Hence, x is a positive definite matrix; that is, ${}^tvx({}^tx)v = {}^tvg^*gv > 0$ for every $0 \neq v \in \mathbb{R}^n$. Thus, there is a matrix $h \in H$ such that ${}^thxh = I$. Then ${}^thg^*gh = I + i({}^thyh)$. The matrix $z := {}^thyh$ is antisymmetric and it is a standard fact in linear algebra that it is diagonalizable by a real orthogonal matrix. Consequently, $h'zh'^{-1} = d$, with d consisting of $m = \lfloor n/2 \rfloor$ 2×2 blocks of the form

$$\begin{pmatrix} d_{2i-1,2i-1} & d_{2i-1,2i} \\ d_{2i,2i-1} & d_{2i,2i} \end{pmatrix} = \begin{pmatrix} 0 & \lambda_i \\ -\lambda_i & 0 \end{pmatrix}$$

in the case when n is even, and m such blocks and the last row zero in the case when n is odd. Note also that the numbers λ_i are uniquely determined up to a permutation by the matrix gg^* since they are eigenvalues of $hy\,{}^th$. Clearly, every 2×2 block of the form $\left(\begin{smallmatrix} 1 & i\lambda \\ -i\lambda & 1 \end{smallmatrix}\right)$ can be transformed by a diagonal matrix $\left(\begin{smallmatrix} d_1 & 0 \\ 0 & d_2 \end{smallmatrix}\right)$ to the form

$$\begin{pmatrix} \mathrm{ch}(\mu) & i\,\mathrm{sh}(\mu) \\ -i\,\mathrm{sh}(\mu) & \mathrm{ch}(\mu) \end{pmatrix} = \exp\left(\mu\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}\right).$$

Taking in every block a of the form $\exp\left(\frac{\mu}{2}\left(\begin{smallmatrix} 0 & i \\ -i & 0 \end{smallmatrix}\right)\right)$ proves the existence of the decomposition $g^*g = {}^tha^2h$ and thus establishes the existence of the decomposition $g = kah$.

We now prove the uniqueness part of the theorem. Note that H acts on the space of positive definite matrices of the form g^*g by $h \cdot x := h^txh$. Let us take $h, b, c \in H$ and suppose $h \cdot (I + ib) = I + ic$. Then h is an orthogonal matrix, ${}^thh = I$, and thus $c = h^{-1}bh$. In particular, the eigenvalues of b and c are equal. Now, to prove the uniqueness of $a \in A^+$ in the decomposition (3-2) let us write $a = \mathrm{Re}(a) + i\,\mathrm{Im}(a)$ and note that $H \cdot a = H \cdot (I + i\,\mathrm{Im}(a))$. Since the eigenvalues of $i\,\mathrm{Im}(a)$ are $\pm\sinh(\lambda_1), \ldots, \pm\sinh(\lambda_n)$ we see that if $a_1, a_2 \in A^+$ and $a_1 \neq a_2$, then $H \cdot (I + i\,\mathrm{Im}(a_1)) \neq H \cdot (I + i\,\mathrm{Im}(a_2))$, and therefore $H \cdot a_1 \neq H \cdot a_2$. It follows that the $a^2 \in A^+$ part in $g^*g = {}^tha^2h$ is uniquely determined by g. As a result, $a \in A^+$ is uniquely determined by g. $\square$

4. Proof of Theorem 1.2

In this paragraph n is fixed and $G = G_n(\mathbb{C}),\ H = G_n(\mathbb{R})$, and $B = B_n(\mathbb{C})$. We denote by M the standard maximal torus in G and by $W_2 = W_{n,2}$ the set of involutions in S_n. As a starting point of the proof, observe that

$$I(\chi)^* = \mathcal{S}^*(G)^{B,\chi\delta_0^{-\frac{1}{2}}},$$

where B acts on the space of tempered distributions $\mathcal{S}^*(G)$ from the left. We have

$$\mathrm{Hom}_H(I(\chi), \mathbb{C}) = \mathcal{S}^*(G/H)^{B,\chi\delta_0^{-\frac{1}{2}}}.$$

We will stratify $X := G/H$ by B-orbits. By Lemma 3.1, we have $B\backslash X = W_2$. Suppose $\mathrm{Hom}_H(I(\chi), \mathbb{C}) \neq 0$. By Lemma 2.1 there exists an involution $w \in W_2$ and a $k \geq 0$ such that

$$\mathcal{S}^*\big(B(w), \mathrm{Sym}^k(CN^X_{B(w)})\big)^{B,\chi\delta_0^{-\frac{1}{2}}} \neq 0.$$

Note that B acts on $B(w)$ transitively, the stabilizer of w under the action of B is B_w, and $\delta_0^{1/2}|_{B^w} = \delta_{B^w}$. Therefore, by Frobenius reciprocity (Theorem 2.2),

$$\begin{aligned}\mathcal{S}^*\big(B(w), \mathrm{Sym}^k(CN^X_{B(w)})\big)^{B,\chi\delta_0^{-\frac{1}{2}}} &= \mathcal{S}^*\big(\{w\}, \mathrm{Sym}^k(CN^X_{B(w),w})\big)^{B^w,\chi\delta_0^{-\frac{1}{2}}\delta_{B^w}^{-1}\delta_0} \\ &= \mathcal{S}^*\big(\{w\}, \mathrm{Sym}^k(CN^X_{B(w),w})\big)^{B^w,\chi} \\ &= \big(\mathrm{Sym}^k(N^X_{B(w),w}) \otimes_{\mathbb{R}} \mathbb{C}\big)^{B^w,\chi}.\end{aligned}$$

Observe that $M^w \subset B^w$. Hence $\big(\mathrm{Sym}^k(N^X_{B(w),w}) \otimes_{\mathbb{R}} \mathbb{C}\big)^{B^w,\chi} \neq 0$ implies

$$\big(\mathrm{Sym}^k(N^X_{B(w),w}) \otimes_{\mathbb{R}} \mathbb{C}\big)^{M^w,\chi} \neq 0.$$

Note that

$$\begin{aligned}M^w &= \{t \in M : t^{-1}w\bar{t} = w\} = \{t \in M : t = w\bar{t}w\} \\ &= \{t = \mathrm{diag}(t_1, t_2, \ldots, t_n) \in M : t_i = \overline{t_{w(i)}} \text{ for } 1 \leq i \leq n\}.\end{aligned}$$

It will be useful to obtain one more formula for M^w. It is easy to see, by examining the case of 1×1 and 2×2 matrices, that

$$M^w = \{t(w\bar{t}w)a : t \in M,\ a = \mathrm{diag}(a_1, a_2, \ldots, a_n), \quad a_i = 1 \text{ if } w(i) \neq i,\ a_i = \pm 1 \text{ if } w(i) = i\}. \tag{4-1}$$

In the next lemma we perform a calculation of the normal space $N^X_{B(w),w}$. Note that this is a finite-dimensional vector space over $\mathbb{R}$. Since the group M^w preserves the tangent space $T_w^{B(w)}$ and clearly preserves the tangent space T_w^X, there is an action of M^w on the normal space $N^X_{B(w),w}$. By taking the scalar extension $N^X_{B(w),w} \otimes \mathbb{C}$, we get a complex representation of M^w. Since M^w is abelian, this representation decomposes into a direct sum of irreducible, one-dimensional representations.

Lemma 4.1. *We have*

$$N^X_{B(w),w} \otimes_{\mathbb{R}} \mathbb{C} = \bigoplus_{\{(i,j)\in I_w\}} \alpha_{\delta(i,j)}$$

as a representation of M^w.

Before proving this lemma, we give the following corollary.

Corollary 4.2. *We have*

$$\mathrm{Sym}(N^X_{B(w),w} \otimes_{\mathbb{R}} \mathbb{C}) = \bigoplus_{\kappa : I_w \to \mathbb{Z}_{\geq 0}} \alpha_{\kappa}$$

as a representation of M^w.

Proof of Lemma 4.1. Let us denote by $e_{i,j}$ the elementary matrix with 1 at the (i, j)-th entry and zeros in all other entries. The tangent space of X at w is equal to

$$\begin{aligned} T^X_w &= \{A \in \mathrm{Mat}_n(\mathbb{C}) : Aw + w\bar{A} = 0\} = \{A \in \mathrm{Mat}_n(\mathbb{C}) : wAw = -\bar{A}\} \\ &= \mathrm{Span}_{\mathbb{R}}\{-e_{i,j} + e_{w(i),w(j)}, \sqrt{-1}(e_{i,j} + e_{w(i),w(j)}) : 1 \leq i, j \leq n\}. \end{aligned}$$

On the other hand,

$$T^{B(w)}_w = \{-Aw + w\bar{A} : A \in \mathrm{Mat}_n(\mathbb{C}),\ A \text{ is upper triangular}\}.$$

Since $e_{i,j}w = e_{i,w(j)}$ and $we_{i,j} = e_{w(i),j}$, we obtain that

$$\begin{aligned} T^{B(w)}_w &= \mathrm{Span}_{\mathbb{R}}\{-e_{i,w(j)} + e_{w(i),j}, \sqrt{-1}(e_{i,w(j)} + e_{w(i),j}) : 1 \leq i \leq j \leq n\} \\ &= \mathrm{Span}_{\mathbb{R}}\{-e_{i,j} + e_{w(i),w(j)}, \sqrt{-1}(e_{i,j} + e_{w(i),w(j)}) : i \leq w(j)\} \\ &= \mathrm{Span}_{\mathbb{C}}\{e_{i,j}, e_{w(i),w(j)} : i \leq w(j)\} \cap T^X_w \\ &= \mathrm{Span}_{\mathbb{C}}\{e_{i,j} : i \leq w(j) \text{ or } j \geq w(i)\} \cap T^X_w. \end{aligned}$$

Hence

$$\begin{aligned} N^X_{B(w),w} &\cong \mathrm{Span}_{\mathbb{C}}\{e_{i,j} : i > w(j),\ w(i) > j\} \cap T^X_w \\ &= \mathrm{Span}_{\mathbb{C}}\{e_{i,w(j)} : i > j,\ w(i) > w(j)\} \cap T^X_w \qquad (4\text{-}2) \\ &= \mathrm{Span}_{\mathbb{C}}\{e_{i,w(j)} : (i, j) \in I_w\} \cap T^X_w. \end{aligned}$$

Let us denote $V = \mathrm{Span}_{\mathbb{C}}\{e_{i,w(j)} : (i, j) \in I_w\}$. Note that if $e_{i,w(j)} \in V$ then also $e_{w(i),j} \in V$, since w is an involution and for an involution

$$(i, j) \in I_w \Longleftrightarrow (w(i), w(j)) \in I_w.$$

Let us use the lexicographic ordering on pairs (i, j): write $(i, j) < (i', j')$ if $i < i'$ or if $i = i'$ and $j < j'$. Then we may rewrite (4-2) as

$$N^X_{B(w),w} \cong \mathrm{Span}_{\mathbb{R}}\{\sqrt{-1}e_{i,w(j)} : (i, j) \in I_w,\ (i, j) = (w(i), w(j))\} \oplus \mathrm{Span}_{\mathbb{R}}\{e_{i,w(j)} - e_{w(i),j},\ \sqrt{-1}(e_{i,w(j)} + e_{w(i),j}) : (i, j) \in I_w,\ (i, j) < (w(i), w(j))\}. \tag{4-3}$$

For $t = \mathrm{diag}(t_1, \dots, t_n) \in M$ we have

$$te_{i,j}\bar{t}^{-1} = (t_i/\bar{t_j})e_{i,j},$$

and for $t \in M^w$ we also have

$$te_{i,w(j)}\bar{t}^{-1} = (t_i/\overline{t_{w(j)}})e_{i,w(j)} = (t_i/t_j)e_{i,w(j)}.$$

Therefore, the action of M^w on $e_{i,w(j)}$ is given by $\alpha_{\delta(i,j)}$. We obtain that, as a representation of M^w,

$$\begin{aligned} N^X_{B(w),w} \otimes_{\mathbb{R}} \mathbb{C} &\cong \bigoplus_{\{(i,j)\in I_w:(i,j)=(w(i),w(j))\}} \alpha_{\delta(i,j)} \oplus \bigoplus_{\{(i,j)\in I_w:(i,j)<(w(i),w(j))\}} (\alpha_{\delta(i,j)} \oplus \alpha_{\delta(w(i),w(j))}) \\ &= \bigoplus_{\{(i,j)\in I_w:(i,j)=(w(i),w(j))\}} \alpha_{\delta(i,j)} \oplus \bigoplus_{\{(i,j)\in I_w:(i,j)<(w(i),w(j))\}} \alpha_{\delta(i,j)} \oplus \bigoplus_{\{(i,j)\in I_w:(i,j)>(w(i),w(j))\}} \alpha_{\delta(i,j)} \\ &= \bigoplus_{\{(i,j)\in I_w\}} \alpha_{\delta(i,j)}. \end{aligned}$$

□

Lemma 4.3. *If* $(\mathrm{Sym}^k(N^X_{B(w),w}) \otimes_{\mathbb{R}} \mathbb{C})^{M^w,\chi} \neq 0$ *then* $k = 0$, $w(\chi) = \bar{\chi}^{-1}$, *and* $\chi_i(-1) = 1$ *for all* $1 \le i \le n$ *such that* $w(i) = i$.

Proof. Note that, for $t \in M$ and $w \in W_2$, the element wtw is also diagonal and its diagonal entries are the permutation of diagonal entries of t by w, i.e.,

$$(wtw)_{ii} = t_{w(i),w(i)}.$$

By (4-1), if $\alpha_\kappa|_{M^w} = \chi|_{M^w}$, then for every $t \in M$ we have

$$\alpha_\kappa(t(w\bar{t}w)) = \alpha_\kappa(t)\overline{w(\alpha_\kappa)(t)} = \chi(t)\overline{\chi(w(t))} = \chi(t(w\bar{t}w)).$$

That is,

$$(\alpha_\kappa|_{M^w} = \chi|_{M^w}) \Rightarrow (\overline{\alpha_\kappa} w(\alpha_\kappa) = \bar{\chi} w(\chi)). \tag{4-4}$$

To obtain (4-4) just put $a = 1$ in (4-1). The set of κs that satisfy

$$\overline{\alpha_\kappa} w(\alpha_\kappa) = \bar{\chi} w(\chi) \tag{4-5}$$

is $\{\kappa \equiv 0\}$ if $w(\chi) = \overline{\chi}^{-1}$ and is empty otherwise. Indeed, we take the absolute value on both sides of (4-5) to obtain

$$\prod_{(i,j)\in I_w} \left|\frac{t_i}{t_j}\right|^{\kappa(i,j)+\kappa(w(i),w(j))} = \prod_{i=1}^{n} |t_i|^{\lambda_i+\lambda_{w(i)}}. \tag{4-6}$$

First, we will deduce from the last equation that the right-hand side of this equation is 1. Note that from (4-6) it follows, by substituting $t_i = c$ for all i with a generic $c \in \mathbb{C}^*$, that

$$\lambda_1 + \cdots + \lambda_n = 0.$$

Since no pair $(1, i)$ belongs to I_w, it follows that $\lambda_1 + \lambda_{w(1)} \le 0$. Let i be the first index such that $\lambda_i + \lambda_{w(i)} > 0$. Then on the left-hand side of (4-6) the power of $|t_i|$ is positive, thus there is a j such that $(i, j) \in I_w$. Hence $i > j$, $w(i) > w(j)$ and from the assumption $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ we obtain $\lambda_i \le \lambda_j$ and $\lambda_{w(i)} \le \lambda_{w(j)}$. Thus

$$0 < \lambda_i + \lambda_{w(i)} \le \lambda_j + \lambda_{w(j)} \le 0,$$

a contradiction! Therefore, for every i, there is an inequality $\lambda_i + \lambda_{w(i)} \le 0$. Since the sum of all λs is equal to 0, we obtain $\lambda_i + \lambda_{w(i)} = 0$ for every i. Hence, $\lambda_1 \ge 0$ and $\lambda_n \le 0$.

Now, we can deduce $\kappa \equiv 0$. Let j be the minimal index such that there exists a pair $(i, j) \in I_w$ with the property

$$\kappa(i, j) \ne 0 \quad \text{or} \quad \kappa(w(i), w(j)) \ne 0.$$

The power of $|t_j|$ on the right-hand side of (4-5) must equal 0, thus there is a pair $(j, k) \in I_w$ such that $\kappa(j, k) \ne 0$ or $\kappa(w(j), w(k)) \ne 0$. In both cases we obtain a contradiction to the minimality of j. As a conclusion, we obtain that (4-5) implies $\kappa \equiv 0$.

Suppose now that $w(\chi) = \overline{\chi}^{-1}$ and thus $\kappa \equiv 0$. Then $\alpha_\kappa = 1$, the identity character. We want to prove that $\chi_i(-1) = 1$ for all i such that $w(i) = i$. This follows from $\chi(a) = \alpha_\kappa(a) = 1$ for $a = \mathrm{diag}(a_1, \dots, a_n)$, where $a_i = \pm 1$ whenever $w(i) = i$ and $a_i = 1$ otherwise. □

5. Calculation of Rankin–Selberg gamma factors

In this section we recall the notion of Rankin–Selberg integrals and apply the results of previous sections to calculate special values of Rankin–Selberg gamma factors. The exposition and notation follows [Jacquet 2009]. Let $\chi : B_n \to \mathbb{C}^\times$ be a multiplicative character and let $\lambda : I(\chi) \to \mathbb{C}$ be a ψ-form on $I(\chi)$. Recall that such a λ always exists and it is unique up to a scalar multiple. If $f \in V$, $g \in G_n$, we set

$$W_f(g) = \lambda(R(g)f).$$

Let $\mathcal{W}(I(\chi), \psi)$ be the space spanned by the functions of the form W_f.

For every n, we denote by w_n the $n \times n$ permutation matrix whose antidiagonal entries are 1. If $n > n'$, we define

$$w_{n,n'} = \begin{pmatrix} 1_{n'} & 0 \\ 0 & w_{n-n'} \end{pmatrix}.$$

If $f \in I(\chi)$, then the function $\tilde{f}$ is defined by

$$\tilde{f}(g) := f(w_n {}^t g^{-1}).$$

Let π be an irreducible representation of $G_n(\mathbb{C})$ and let π' be an irreducible representation of $G_m(\mathbb{C})$. Suppose π is the Langlands quotient of $I(\chi)$ and π' is the Langlands quotient of $I(\chi')$. We choose a ψ-form λ on $I(\chi)$ and a $\bar{\psi}$-form λ' on $I(\chi')$. Rankin–Selberg integrals are defined as follows. For $f \in I(\chi),\ f' \in I(\chi')$, set

$$W = W_f, \quad W' = W_{f'}.$$

For $W = W_f$, set

$$\widetilde{W}_f := W_f(w_n {}^t g^{-1}).$$

Note that $\widetilde{W}_f(g) = W_{\tilde{f}}(g)$ and $W_{\tilde{f}}(g) \in \mathcal{W}(I(\chi^{-1}), \bar{\psi})$.

If $n > n'$, we set

$$\Psi(s, W, W') = \int W\begin{pmatrix} g & 0 \\ 0 & 1_{n-n'} \end{pmatrix} W'(g) |\det g|_{\mathbb{C}}^{s-\frac{n-n'}{2}} \, dg. \tag{5-1}$$

In addition, for $0 \le j \le n - n' - 1$, we set

$$\Psi_j(s, W, W') = \int W\begin{pmatrix} g & 0 & 0 \\ X & 1_j & 0 \\ 0 & 0 & 1_{n-n'-j} \end{pmatrix} W'(g) |\det g|_{\mathbb{C}}^{s-\frac{n-n'}{2}} \, dX \, dg. \tag{5-2}$$

Here X is integrated over the space $M(m \times j, \mathbb{C})$ of matrices with m rows and j columns. In each integral, g is integrated over the quotient $U_n(\mathbb{C}) \backslash G_n(\mathbb{C})$.

If $n = n'$, we let Φ be a Schwartz function on $\mathbb{C}^n$ and we set

$$\Psi(s, W, W', \Phi) = \int W(g) W'(g) \Phi((0, 0, \dots, 0, 1)g) |\det g|_{\mathbb{C}}^{s} \, dg. \tag{5-3}$$

The Rankin–Selberg gamma factor $\gamma(s, \pi \times \pi', \psi)$ is a proportionality factor appearing in functional equations on certain Rankin–Selberg integrals. We quote here [Jacquet 2009, Theorem 2.1].

Theorem 5.1. (1) *The integrals* (5-1), (5-2), *and* (5-3) *converge for* $\mathrm{Re}(s) \gg 0$.

(2) *Each integral extends to a meromorphic function of s which is a holomorphic multiple of $L(s, \pi \times \pi')$ bounded at infinity in vertical strips. See* [Jacquet 2009] *for the definition of $L(s, \pi \times \pi')$.*

(3) *The following functional equations are satisfied. If $n > n'$,*

$$\Psi(1-s, \widetilde{W}, \widetilde{W'}) = \omega_{I(\chi)}(-1)^{n-1}\omega_{I(\chi')}(-1)\gamma(s, I(\chi) \times I(\chi'), \psi)\Psi(s, W, W').$$

If $n > n'+1$ and $\beta = n - n' - 1 - j$,

$$\Psi_j(1-s, \pi(w_{n,n'})\widetilde{W}, \widetilde{W'}) = \omega_{I(\chi)}(-1)^{n'}\omega_{I(\chi')}(-1)\gamma(s, \pi \times \pi', \psi)\Psi_\beta(s, W, W').$$

If $n = n'$,

$$\Psi(1-s, \widetilde{W}, \widetilde{W'}, \hat{\Phi}) = \omega_{I(\chi)}(-1)^{n-1}\gamma(s, \pi \times \pi', \psi)\Psi(s, W, W', \Phi).$$

We will calculate the special values of the Rankin–Selberg gamma factor of $G_n(\mathbb{R})$-distinguished representations. The main tool will be the classification of such representations obtained in Theorem 1.2 and basic properties of Rankin–Selberg gamma factors from [Jacquet 2009, Lemma 16.3].

Let us recall some facts about one-dimensional Tate gamma factors. Let χ be a one-dimensional character $\chi : \mathbb{C}^* \to \mathbb{C}^*$. We have the following functional equation for Tate gamma factors:

$$\gamma(s, \chi, \psi)\gamma(1-s, \overline{\chi^{-1}}, \overline{\psi^{-1}}) = 1. \tag{5-4}$$

Since we assume ψ is trivial on $\mathbb{R}$, we obtain $\overline{\psi^{-1}} = \psi$, and for $s = 1/2$ we get

$$\gamma\left(\frac{1}{2}, \chi, \psi\right)\gamma\left(\frac{1}{2}, \overline{\chi^{-1}}, \psi\right) = 1. \tag{5-5}$$

For a real character χ, that is, for χ satisfying $\chi^2 = 1$, we obtain $\gamma(1/2, \chi, \psi)^2 = 1$, and thus $\gamma(1/2, \chi, \psi) \in \{1, -1\}$. The value of $\gamma(1/2, \chi, \psi)$ depends on $\chi(-1)$. Whenever $\chi(-1) = 1$ we obtain

$$\gamma\left(\frac{1}{2}, \chi, \psi\right) = 1. \tag{5-6}$$

Proof of Theorem 1.1. Recall that if π is the Langlands quotient of $I(\chi)$ and π' is the Langlands quotient of $I(\chi')$, then

$$\gamma(s, \pi \times \pi', \psi) = \gamma(s, I(\chi) \times I(\chi'), \psi).$$

It is well-known that $\chi = (\chi_1, \ldots, \chi_t)$, where the χ_is are one-dimensional characters of $\mathbb{C}$. Similarly, $\chi' = (\chi'_1, \ldots, \chi'_r)$, where the χ'_is are one-dimensional characters of $\mathbb{C}$. Thus,

$$\gamma(s, I(\chi)\times I(\chi'), \psi) = \prod_{i=1}^{t} \gamma(s, \chi_i \times I(\chi'), \psi) = \prod_{i=1}^{t}\prod_{j=1}^{r} \gamma(s, \chi_i\chi'_j, \psi). \tag{5-7}$$

Using Theorem 1.2, there exist involutions $w \in S_t$ and $w' \in S_r$ such that $w(\chi) = \bar{\chi}^{-1}$ and $w'(\chi') = \bar{\chi}'^{-1}$ and for every fixed point i of w, and j of w', we have $\chi_i(-1) = 1$ and $\chi'_j(-1) = 1$. The formula in (5-7) may be rewritten as

$$\gamma(s, I(\chi)\times I(\chi'), \psi) = I_1 I_2,$$

where

$$I_1 = \prod_{\{(i,j):(w(i),w'(j))=(i,j)\}} \gamma\left(\frac{1}{2}, \chi_i\chi'_j, \psi\right),$$

$$I_2 = \prod_{\{(i,j):\, i<w(i)\ \text{ or }\ (i=w(i)\ \text{ and } w'(j)<j)\}} \gamma\left(\frac{1}{2}, \chi_i\chi'_j, \psi\right)\gamma\left(\frac{1}{2}, \chi_{w(i)}\chi'_{w'(j)}, \psi\right).$$

Let us prove that every term appearing in the product I_1 is 1. Indeed, by Theorem 1.2 the character $\chi_i\chi'_j$ appearing as the argument in the gamma factor in I_1 is a real character satisfying $\chi\chi'_j(-1) = 1$, and therefore, by (5-6), we get

$$\gamma\left(\frac{1}{2}, \chi_i\chi'_j, \psi\right) = 1.$$

Each term in the product I_2 also equals 1, since $\chi_{w(i)}\chi'_{w'(j)} = (\overline{\chi_i\chi'_j})^{-1}$ and by applying (5-5). Finally, $I_1 = I_2 = 1$ and we obtain

$$\gamma\left(\frac{1}{2}, \pi\times\pi', \psi\right) = 1. \qquad \square$$

We will need the following technical result about Rankin–Selberg integrals in Section 8.

Lemma 5.2. *Let (π, V), (π', V') be generic representations of $G_n(\mathbb{C})$ and let*

$$\mathcal{W}(\pi, \psi), \quad \mathcal{W}(\pi', \psi^{-1})$$

be their Whittaker models. Suppose (π, V) is unitarizable and (π', V') is tempered. Let $W \in \mathcal{W}(\pi, \psi))$, $W' \in \mathcal{W}(\pi', \psi^{-1})$, and $\Phi \in \mathcal{S}(\mathbb{C}^n)$. Then the Rankin–Selberg integral

$$\int_{U_n(\mathbb{C})\backslash G_n(\mathbb{C})} W(g)W'(g)\Phi((0,0,\dots,0,1)g)|\det g|_{\mathbb{C}}^{s}\, dg$$

converges absolutely at $s = 1/2$.

Proof. Define

$$T_n = \{\mathrm{diag}(t_1, \dots, t_n) : t_i \in \mathbb{R} \text{ and } t_1 \geq t_2 \geq \cdots \geq t_n > 0\}$$

and let K_n be a maximal compact subgroup of $G_n(\mathbb{C})$ consisting of all unitary matrices in $G_n(\mathbb{C})$. Let δ be the modular character of $B_n(\mathbb{C})$. By [Lapid and Mao 2014, Lemma 2.1], we know that there exist a $\lambda > -1/2$, a $d > 0$ and a continuous seminorm μ on $\mathcal{W}(\pi, \psi)$ such that

$$|W(tk)| \leq \delta^{\frac{1}{2}}(t) |\det t|_{\mathbb{C}}^{\lambda} |t_n|_{\mathbb{C}}^{-n\lambda} (1 + \|\log t\|)^d \mu(W)$$

for all $t \in T_n$, $k \in K_n$ and every $W \in \mathcal{W}(\pi, \psi)$. Similarly, there exist an $\epsilon < 0$ such that $\lambda + \epsilon > -1/2$, a $d' > 0$ and a continuous seminorm μ' on $\mathcal{W}(\pi', \psi^{-1})$ such that

$$|W'(tk)| \leq \delta^{\frac{1}{2}}(t) |\det t|_{\mathbb{C}}^{\epsilon} |t_n|_{\mathbb{C}}^{-n\epsilon} (1 + \|\log t\|)^{d'} \mu'(W')$$

for all $t \in T_n$, $k \in K_n$ and every $W' \in \mathcal{W}(\pi', \psi^{-1})$. Let us define $\lambda := \lambda + \epsilon$, $d := d + d'$. All that matters for the estimates is that $\lambda > 1/2$ and $d > 0$. There is a $\phi \in \mathcal{S}(\mathbb{R})$ such that

$$\begin{aligned} &|W(tk) W'(tk) \Phi((0, 0, \dots, 0, 1)g)| \\ &\qquad \leq \delta(t) |\det t|_{\mathbb{C}}^{\lambda} |t_n|_{\mathbb{C}}^{-n\lambda} (1 + \|\log t\|)^d \phi(t_n) \mu(W) \mu'(W') \end{aligned}$$

for all $t \in T_n$, $k \in K_n$, every $W \in \mathcal{W}(\pi, \psi)$ and every $W' \in \mathcal{W}(\pi', \psi^{-1})$. For fixed functions W, W' the numbers $\mu(W)$, $\mu'(W')$ are constant and we can move them to the function ϕ. Thus we can rewrite the last estimate as

$$|W(tk) W'(tk) \Phi((0, 0, \dots, 0, 1)g)| \leq \delta(t) |\det t|_{\mathbb{C}}^{\lambda} |t_n|_{\mathbb{C}}^{-n\lambda} (1 + \|\log t\|)^d \phi(t_n)$$

for all $t \in T_n$, $k \in K_n$. Let us rewrite the expression

$$\int\limits_{U_n(\mathbb{C}) \backslash G_n(\mathbb{C})} |W(g) W'(g) \Phi((0, 0, \dots, 0, 1)g)| \, |\det g|_{\mathbb{C}}^{\mathrm{Re}(s)} \, dg$$

using the Iwasawa decomposition to obtain

$$\int\limits_{K_n} \int\limits_{T_n} |W(tk) W'(tk) \Phi((0, 0, \dots, 0, 1)g)| \, |\det t|_{\mathbb{C}}^{\mathrm{Re}(s)} \delta^{-1}(t) \, dt \, dk. \tag{5-8}$$

For $f : G_n(\mathbb{C}) \to \mathbb{C}$ such that the following integrals are absolutely convergent, we have

$$\begin{aligned}\int_{G_n(\mathbb{C})} f(g)\,dg &= \int\limits_{U_n(\mathbb{C})}\int\limits_{T_n}\int\limits_{K_n} f(tuk)\,du\,dt\,dk\\ &= \int\limits_{U_n(\mathbb{C})}\int\limits_{T_n}\int\limits_{K_n} f((tut^{-1})tk)\,du\,dt\,dk\\ &= \int\limits_{U_n(\mathbb{C})}\int\limits_{T_n}\int\limits_{K_n} f(utk)\delta^{-1}(t)\,du\,dt\,dk.\end{aligned}$$

Let us define $\alpha_i(t) = t_i/t_{i+1}$ for $t \in T_n$. Note that

$$\det t = \prod_{i=1}^{n-1} \alpha_i(t)^i t_n^n.$$

The integrand in (5-8) is bounded by

$$\begin{aligned}&|W(tk)W'(tk)\Phi(\eta_n g)||\det t|_{\mathbb{C}}^{\mathrm{Re}(s)}\delta^{-1}(t)\\ &\qquad\le (1+\|\log t\|)^d\phi(t_n)\biggl(\prod_{i=1}^{n-1}\alpha_i(t)^{2i(\mathrm{Re}(s)+\lambda)}\biggr)t_n^{2n\ \mathrm{Re}(s)}.\end{aligned}$$

There exists an $e > 0$ such that

$$(1+\|\log t\|)^d \le \biggl(\prod_{j=1}^{n-1}(1+\log\alpha_j(t))^e\biggr)(1+\log|t_n|)^e$$

for all $t \in T_n$. We have the estimate

$$\begin{aligned}&\int_{T_n}(1+\|\log t\|)^e\phi(t_n)|\det t|_{\mathbb{C}}^{\mathrm{Re}(s)+\lambda}|t_n|_{\mathbb{C}}^{-n\lambda}\,dt\\ &\quad\le \prod_{j=1}^{n-1}\int_1^\infty(1+\log t_j)^e\, t_j^{2j(\mathrm{Re}(s)+\lambda)}\,d^\times t_j \times \int_0^\infty (1+|\log t_n|)^e\phi(t_n)\,t_n^{2nRe(s)}\,d^\times t_n.\end{aligned}$$

It follows that the integral absolutely converges for s satisfying $\mathrm{Re}(s) > -\lambda$ and $\mathrm{Re}(s) > 0$. As $\lambda > -1/2$, we obtain the absolute convergence of the Rankin–Selberg integral at $s = 1/2$. $\square$

6. Integral representation of Whittaker functions

Let $n \ge 2$ be fixed and let $K = U_n(\mathbb{C})$ be a maximal compact subgroup of $G_n(\mathbb{C})$. The next lemma gives a convenient formula for the $G_n(\mathbb{R})$-period of a unitary and generic representation $(\pi, V) \in \mathrm{Irr}_{G_n(\mathbb{R})}(G_n(\mathbb{C}))$. It is proved in [Lapid and Mao

2014, Lemma 1.2]. We state it and, for the convenience of the reader, provide a full proof here.

Lemma 6.1 [Lapid and Mao 2014, Lemma 1.2]. *Let $(\pi, V) \in \mathrm{Irr}_{G_n(\mathbb{R})}(G_n(\mathbb{C}))$ be unitarizable and generic and let $\mathcal{W}(\pi, \psi)$ be its Whittaker model. The functional*

$$\mu : W \mapsto \int_{U_{n-1}(\mathbb{R})\backslash G_{n-1}(\mathbb{R})} W(h)\, dh$$

defines a $P_n(\mathbb{R})$-invariant functional on $\mathcal{W}(\pi, \psi)$. Moreover, there exists an $N > 0$ and a seminorm ν on $\mathcal{W}(\pi, \psi)$ such that for all $g \in G_n(\mathbb{C})$ and $W \in \mathcal{W}(\pi, \psi)$ we have the inequality

$$|\mu(\pi(g)W)| \le \|g\|_H^N \nu(W).$$

Proof. By [Wallach 1992, Theorem 15.2.5], there exist a continuous seminorm ν' on $\mathcal{W}(\pi, \psi)$, a $\lambda > -1/2$ and a $d > 0$ such that

$$|W(tk)| \le \delta^{\frac{1}{2}}(t)|\det t|_{\mathbb{C}}^{\lambda} |t_n|_{\mathbb{C}}^{-n\lambda}(1 + \|\log t\|)^d \nu'(W)$$

for every $W \in \mathcal{W}(\pi, \psi)$, $t \in T_n$, and $k \in K_n$. For $g \in G_{n-1}(\mathbb{R})$ we have $t_n = 1$. Let us denote

$$T_{n-1} = \{t \in T_n : t_n = 1\}.$$

Let $\delta_0(t)$ be the modular character of $B_n(\mathbb{R})$. For $t \in B_n(\mathbb{R})$ we have that $\delta^{1/2}(t) = \delta_0(t)$. Multiplying ν' by a scalar we have the estimate

$$\begin{aligned} \int_{U_{n-1}(\mathbb{R})\backslash G_{n-1}(\mathbb{R})} |W(h)|\, dh &\le \int_{K_{n-1}(\mathbb{R})} \int_{T_{n-1}} |W(tk)| \delta_0^{-1}(t)\, dt\, dk \\ &\le \nu'(W) \int_{T_{n-1}} |\det t|_{\mathbb{C}}^{\lambda} (1 + \|\log t\|)^d\, dt \end{aligned} \tag{6-1}$$

for every $W \in \mathcal{W}(\pi, \psi)$. By the estimates of the previous lemma we obtain that the last integral converges absolutely for $\lambda > -1/2$. It follows that $|\mu(W)| \le \nu''(W)$ for a continuous seminorm ν'' on $\mathcal{W}(\pi, \psi)$. Since π is of moderate growth, there exist another continuous seminorm ν on $\mathcal{W}(\pi, \psi)$ and an $N > 0$ such that

$$|\mu(\pi(g)W)| \le \nu''(\pi(g)W) \le \|g\|_H^N \nu(W)$$

for every $W \in \mathcal{W}(\pi, \psi)$ and every $g \in G$. □

We will identify the functional μ on $\mathcal{W}(\pi, \psi)$ with the corresponding linear functional on V, which we will, by abuse of notation, also denote by μ. By the uniqueness of the Whittaker model, this identification defines $\mu \in V^*$ in a unique way, up to a scalar multiple. Since $\mu \in (V^*)^{P_n(\mathbb{R})}$ and $(V^*)^{P_n(\mathbb{R})} = (V^*)^{G_n(\mathbb{R})}$ (see [Kemarsky 2015, Theorem 1.1]), we obtain that $\mu \in (V^*)^{G_n(\mathbb{R})}$. Clearly, $\mu \neq 0$.

The functional μ defines an embedding of V into the space of functions on $G_n(\mathbb{R})\backslash G_n(\mathbb{C})$ via

$$V \ni v \mapsto (g \mapsto \mu(\pi(g)v)).$$

By abuse of notation we denote this embedding again by μ. Denote the image of the embedding μ by $\mathcal{C}_{G_n(\mathbb{R})}(\pi)$. In the other direction, we can define a map

$$\theta : \mathcal{C}_{G_n(\mathbb{R})}(\pi) \to \mathcal{W}(\pi, \psi)$$

by

$$\theta : f \mapsto \left(g \mapsto \int_{U_n(\mathbb{R})\backslash U_n(\mathbb{C})} f(ug)\psi^{-1}(u)\, du\right). \tag{6-2}$$

In this section we will prove that for every n there exists an irreducible representation (π, V) of $G_n(\mathbb{C})$ that is $G_n(\mathbb{R})$-distinguished and such that the integral (6-2) is absolutely convergent for every K-finite vector in (π, V).

Suppose that we have a generic representation $(\pi, V) \in \mathrm{Irr}_{G_n(\mathbb{R})}(G_n(\mathbb{C}))$ and that the integral (6-2) is absolutely convergent for a K-finite function $f \in \mathcal{C}_{G_n(\mathbb{R})}(\pi)$. Then, from [Lapid and Mao 2014], the composition of maps $\theta(\mu(f))$ is equal to cf for some constant $0 \neq c \in \mathbb{C}$.

Lemma 6.2. *Let $(\pi, V) \in \mathrm{Irr}_{G_n(\mathbb{R})}(G_n(\mathbb{C}))$ be a generic representation. Suppose the integral*

$$\int_{U_n(\mathbb{R})\backslash U_n(\mathbb{C})} W(u)\psi^{-1}(u)\, du$$

absolutely converges for every K_n-finite function $W \in \mathcal{W}(\pi, \psi)$. Then for every $W \in \mathcal{W}(\pi, \psi)$ there exists an $f \in \mathcal{C}_{G_n(\mathbb{R})}(\pi)$ such that

$$W(g) = \int_{U_n(\mathbb{R})\backslash U_n(\mathbb{C})} f(ug)\psi^{-1}(u)\, du.$$

Recall the decomposition (3-3):

$$G_n(\mathbb{C}) = G_n(\mathbb{R})A^+K.$$

The involution $g \to {}^tg^{-1}$ preserves this decomposition. Let $\check{W}(g) = W({}^tg^{-1})$. The Whittaker model $\mathcal{W}(\tilde{\pi}, \psi^{-1})$ of the contragredient representation of (π, V) is given by

$$\mathcal{W}(\tilde{\pi}, \psi^{-1}) = \{\check{W} : W \in \mathcal{W}(\pi, \psi)\}.$$

If the conditions of Lemma 6.2 are satisfied for $\mathcal{W}(\pi, \psi)$ then they are also satisfied for the contragredient representation $\mathcal{W}(\tilde{\pi}, \psi^{-1})$. Explicitly, if $W \in \mathcal{W}(\pi, \psi)$

is equal to

$$W(g) = \int_{U_n(\mathbb{R})\backslash U_n(\mathbb{C})} f(ug)\psi^{-1}(u)\,du,$$

then we have

$$\check{W}(g) = \int_{U_n(\mathbb{R})\backslash U_n(\mathbb{C})} \check{f}(ug)\psi(u)\,du,$$

where $\check{f}(g) := f({}^t g^{-1})$.

Lemma 6.3. *Let $N > 0$. Then there exists a $(\pi, V) \in \mathrm{Irr}_{G_n(\mathbb{R})}(G_n(\mathbb{C}))$ such that for every K-finite function $f \in V$ there is a constant $C > 0$, depending only on f, that satisfies, for every $k \in K$, $a \in A$, $h \in G_n(\mathbb{R})$, the inequality*

$$|f(hak)| \le C(f)\|a\|_H^{-N}. \tag{6-3}$$

Proof. By [Flensted-Jensen 1980], there exists a relatively discrete series $\mathcal{H} := L^2_{ds}(G_n(\mathbb{R})\backslash G_n(\mathbb{C}))$. Moreover, every irreducible representation in $\mathcal{H}$ is isomorphic to some $I(\chi)$, where

$$\chi(z) = ((z/|z|)^{i_1}, (z/|z|)^{i_2}, \dots, (z/|z|)^{i_n})$$

and $i_1, \dots, i_n \in \mathbb{Z}$. If $C > 0$ is big enough and if $|i_k - i_j| > C > 0$ for all $i \neq j$, then the $G_n(\mathbb{R})\backslash G_n(\mathbb{C})$ model of the space $I(\chi)$ lies in $\mathcal{H}$, and the $(\mathfrak{g}, K)$-module generated by a K-finite function $0 \neq f_\lambda \in I(\chi)$ satisfies the properties of the lemma. Indeed, by [Flensted-Jensen 1980, p. 254] (see also [Kassel and Kobayashi 2013, Proposition 5.1]), if $C > 0$ is big enough and if $|i_k - i_j| > C > 0$ for all $j \neq k$, then $f_\lambda(hak) \le C'\|a\|^{-N}$ for all $h \in H$, $a \in A^+$, and $k \in K$.

Clearly, f_λ and right translations of f_λ by K satisfy the properties of our lemma. We should prove that the derivatives of f_λ also satisfy similar growth properties. This is achieved by a classical idea, which is attributed to Harish-Chandra (see also an expository article by [Cowling et al. 1988]). The function f_λ is K-finite, hence there exists a smooth function e_α of compact support such that $f_\lambda * e_\alpha = f_\lambda$. Thus, for $X \in \mathfrak{g}$ we have $dX(f_\lambda) = f_\lambda * dX(e_\alpha)$. It follows that the derivatives of f_λ have the same decay properties that f_λ has.

Finally, the $(\mathfrak{g}, K)$-module generated by f_λ is of finite length. Consequently, it contains an irreducible admissible $(\mathfrak{g}, K)$-submodule satisfying the decay property (6-3). □

If every K-finite function in (π, V) satisfies (6-3) we say that the representation (π, V) decays faster than N. Note that if (π, V) decays faster than N, then its contragredient $(\tilde{\pi}, \tilde{V})$ also decays faster than N. Indeed, we can realize $(\tilde{\pi}, \tilde{V})$ as $\tilde{V} = \{\check{f} : f \in V\}$, where $\check{f}(g) := f({}^t g^{-1})$. If $g = hak$, then ${}^t g^{-1} = {}^t h^{-1} a\, {}^t k^{-1}$. Hence, the property of fast decay is true for $\check{f}$ if and only if it is true for f.

To obtain estimates of convergence of integrals over the unipotent matrices we need the next elementary result. Define Ω_n as the subset of all upper triangular unipotent matrices in $G_n(\mathbb{C})$ with u_{ij} purely imaginary for $j > i$. Note that Ω_n is a fundamental domain for $U_n(\mathbb{R})\backslash U_n(\mathbb{C})$.

Lemma 6.4. *There exist a $C > 0$ and $d > 0$, which depend only on n, such that for every $u \in \Omega_n$ we have*

$$\|u\| \le C\|u\bar{u}^{-1}\|^d.$$

Proof. The proof is by induction on n. For $n = 2$ it follows by direct computation: if $u = \left(\begin{smallmatrix} 1 & ix \\ 0 & 1 \end{smallmatrix}\right)$, then $u\bar{u}^{-1} = \left(\begin{smallmatrix} 1 & 2ix \\ 0 & 1 \end{smallmatrix}\right)$ and the claim is satisfied.

For a general n, let us define A_j to be the set of the entries in the j-th upper diagonal in g:

$$A_0 = \{g_{11}, g_{22}, \dots, g_{nn}\},\ A_1 = \{g_{12}, g_{23}, \dots, g_{(n-1)n}\}, \dots, A_{n-1} = \{g_{1n}\}.$$

Define $B_j := \bigcup_{0\le i<j} A_i$. The crucial observation is that entry (i, j) of $\bar{u}^{-1}$ with indices satisfying $j - i = k$ equals

$$\bar{u}^{-1}_{ij} = u_{ij} + P_{ij}(u),$$

where $P_{ij} \in \mathbb{C}[B_k]$ is a fixed polynomial which depends only on the entries u_{lm} with indices $l - m < k$. Similarly,

$$(u\bar{u}^{-1})_{ij} = 2u_{ij} + Q_{ij}(u),$$

where $Q_{ij} \in \mathbb{C}[B_k]$ is a fixed polynomial which depends only on the entries u_{lm} with indices $l - m < k$. For example, let $n = 3$. Then

$$u = \begin{pmatrix} 1 & ix & iy \\ 0 & 1 & iz \\ 0 & 0 & 1 \end{pmatrix},\ \bar{u}^{-1} = \begin{pmatrix} 1 & ix & iy - xz \\ 0 & 1 & iz \\ 0 & 0 & 1 \end{pmatrix},\ u\bar{u}^{-1} = \begin{pmatrix} 1 & ix & 2iy - 2xz \\ 0 & 1 & iz \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus $P_{12} = P_{23} = 0,\ P_{13} = -xz,\ Q_{12} = Q_{23} = 0,\ Q_{13} = -2xz$. Define $v = u\bar{u}^{-1}$ and define "partial seminorms" of u by

$$\|u\|_k = \sqrt{\sum_{(i,j):j-i\le k} |u_{ij}|^2}.$$

We will prove by induction on k, with base $k = 1$, that for every k, there exist $C_k, d_k > 0$ such that $\|u\|_k^2 \le C_k\|v\|_k^{d_k}$. As $\|u\|_n = \|u\|$, the result follows.

Let $k = 1$. For $C_1 = 1,\ d_1 = 1$ we obtain the desired inequality. Suppose the claim is true for $k - 1$; that is, suppose

$$\|u\|_{k-1}^2 \le C_{k-1}\|v\|_{k-1}^{d_{k-1}}.$$

We want to show a similar inequality for k. There exist $C, d > 0$ such that for every $1 \le i \le n-k$ we have $|v_{i,i+k}| \ge |u_{i,i+k}| - C\|u\|_{k-1}^d$. For example, one can choose

$$d = \max\{\deg(P_{ij}) : i - j = k\}$$

and a big enough constant C. Let u be a given upper triangular unipotent matrix with purely imaginary entries above the diagonal. There exist constants C', C'' such that if for all i we have $|u_{i,i+k}| \le 2C\|u\|_{k-1}^d$, then

$$\|u\|_k^2 = \|u\|_{k-1}^2 + \sum_i |u_{i,i+k}|^2 \le C'\|u\|_{k-1}^{2d} \le C''\|v\|_{k-1}^{2dd_{k-1}} \le C''\|v\|_k^{2dd_{k-1}}.$$

On the other hand, if for some i we have $|u_{i,i+k}| > 2C\|u\|_{k-1}^d$, then we have the inequality $|v_{i,i+k}| > |u_{i,i+k}|/2$ and there exists a constant C''' such that

$$\sum_{i=1}^{n-k} |u_{i,i+k}|^2 \le C''' \sum_{i=1}^{n-k} |v_{i,i+k}|^2 \le C'''\|v\|_k \le C'''\|v\|_k^{2dd_{k-1}}.$$

Therefore, in both cases there are constants C_k, d_k such that

$$\|u\|_k^2 \le C_k\|v\|_k^{d_k}. \qquad \square$$

Corollary 6.5. *There exist a $C > 0$ and $d > 0$, which depend only on n, such that for every $u \in \Omega_n$ we have*

$$\|u\|_H \le C\|u\bar{u}^{-1}\|_H^d.$$

Proof. From Lemma 6.4 we know that there exist $C_1, d_1 > 0$ such that for every $u \in \Omega_n$ we have $\|u\| < C_1\|u\bar{u}^{-1}\|^{d_1}$. Similarly, one proves that there exist $C_2, d_2 > 0$ such that for every $u \in \Omega_n$ we have $\|u\| < C_2\|\bar{u}u^{-1}\|^{d_2}$. Define $C = \max\{C_1, C_2\}$, $d = \max\{d_1, d_2\}$. Then $\|u\|_H \le C\|u\bar{u}^{-1}\|_H^d$ for every $u \in \Omega_n$. $\square$

Lemma 6.6. *Let $N > 0$ be sufficiently large. Then, for every irreducible, $G_n(\mathbb{R})$-distinguished representation (π, V) of $G_n(\mathbb{C})$ with decay faster than N, the integral*

$$\int_{U_n(\mathbb{R})\backslash U_n(\mathbb{C})} f(ug)\,du$$

absolutely converges for every $g \in G_n(\mathbb{C})$ and every K-finite function $f \in V$.

Proof. Let (π, V) be a $G_n(\mathbb{R})$-distinguished representation of $G_n(\mathbb{C})$ such that for every K-finite function $f \in V$ there exists a $C > 0$ depending only on f such that

$$|f(hak)| < C\|a\|_H^{-N}$$

for every $h \in G_n(\mathbb{R})$, $a \in A^+$, $k \in K$. Let $ug = hak$. Then $(\overline{ug})^{-1}ug = \bar{g}^{-1}(\bar{u}^{-1}u)g$. Since g is fixed, there exists a $C_1 > 0$ such that for every matrix $u \in G_n$ we have

$$C_1^{-1}\|\bar{u}^{-1}u\| < \|(\overline{ug})^{-1}ug\| < C_1\|\bar{u}^{-1}u\|.$$

By Lemma 6.4, for $u \in \Omega$ we have

$$\|u\| < C_2 \|\bar{u}^{-1} u\|^d.$$

On the other hand,

$$(\overline{ug})^{-1} ug = \bar{k}^{-1}(a\bar{a}^{-1})k = \bar{k}^{-1} a^2 k.$$

Note that $k \in K$ is a unitary matrix, and therefore

$$\|\bar{k}^{-1} a^2 k\| = \|a^2 k\| = \|a^2\|.$$

Combining these inequalities we get

$$\|a^2\| = \|(\overline{ug})^{-1} ug\| > C_3 \|u\bar{u}^{-1}\| > C_4 \|u\|^{1/d}.$$

Finally, we obtain that there exist constants C, d' such that for $ug = hak$, where u is in Ω_n and $g \in G_n$ is fixed, we have

$$\|a\| > C\|u\|^{1/d'}.$$

Therefore,

$$\int_{U_n(\mathbb{R})\backslash U_n(\mathbb{C})} |f(ug)|\, du \leq \int_{\Omega} C\|u\|_H^{-N/d'}\, du. \tag{6-4}$$

The integral in (6-4) converges for N big enough, thus the lemma is proved. □

Corollary 6.7. *Let $N > 0$ be sufficiently large. Then, for every irreducible $G_n(\mathbb{R})$-distinguished representation (π, V) of $G_n(\mathbb{C})$ with decay faster than N, the integral (6-2) is absolutely convergent.*

7. Archimedean Asai integrals

Nonarchimedean Asai integrals were introduced by Flicker [1988] who then used them to analyze the local and global Asai L and ϵ-factors [1993]. In this section we introduce an archimedean analog of Asai integrals and prove that they are of moderate growth. We also state a functional equation analogous to [Offen 2011, Lemma 4.2] that is satisfied by the integrals.

Let (π, V) be a generic irreducible unitarizable representation of $G_n(\mathbb{C})$ and let $\mathcal{W}(\pi, \psi)$ be its Whittaker model. For $W \in \mathcal{W}(\pi, \psi)$, we define an archimedean Asai integral to be

$$Z(s, W, \Phi) = \int_{U_n(\mathbb{R})\backslash G_n(\mathbb{R})} W(g)\Phi((0, 0, \ldots, 0, 1)g)|\det g|_{\mathbb{R}}^s\, dg. \tag{7-1}$$

Lemma 7.1. *Let $\Phi \in \mathcal{S}(\mathbb{C}^n)$ and $\mathrm{Re}(s) \geq 1$. Then $W \mapsto Z(s, W, \Phi)$ defines a continuous functional on $\mathcal{W}(\pi, \psi)$ for $\mathrm{Re}(s) \geq 1$. That is, there exist a continuous seminorm μ on $\mathcal{W}(\pi, \psi)$ and a continuous seminorm ν on $\mathcal{S}(\mathbb{C}^n)$ such that*

$$|Z(s, W, \Phi)| \leq \mu(W)\nu(\Phi)$$

for every $\Phi \in \mathcal{S}(\mathbb{C}^n)$ and every $W \in \mathcal{W}(\pi, \psi)$. As a consequence, there exist an $N > 0$, a continuous seminorm μ' on $\mathcal{W}(\pi, \psi)$ and a continuous seminorm ν on $\mathcal{S}(\mathbb{C}^n)$ such that

$$\int\limits_{U_n(\mathbb{R})\backslash G_n(\mathbb{R})} |W'(hg)\Phi((0, 0, \ldots, 0, 1)hg)|\,|\det h|_{\mathbb{R}}\, dh \leq \mu'(W)\nu'(\Phi)\, \|g\|_H^N$$

for every $g \in G$ and every $W' \in \mathcal{W}(\pi, \psi)$.

Proof. Let δ_0 be the modulus function of $B_n(\mathbb{R})$. Using the Iwasawa decomposition we obtain

(7-2)
$$|Z(s, W, \Phi)| \leq \int\limits_{K_n(\mathbb{R})} \int\limits_{T_n} |W(tk)\Phi((0, 0, \ldots, 0, 1)tk)|\,|\det(t)|_{\mathbb{R}}^{\mathrm{Re}(s)} \delta_0^{-1}(t)\, dt\, dk.$$

By [Lapid and Mao 2014, Corollary 2.2] there exist a $\lambda > -1/2$, a $d > 0$ and a continuous seminorm μ on $\mathcal{W}(\pi, \psi)$ such that

$$|W(tk)| \leq \delta^{\frac{1}{2}}(t)|\det t|_{\mathbb{R}}^{2\lambda} |t_n|_{\mathbb{R}}^{-2n\lambda}(1 + \|\log t\|)^d \mu(W)$$

for every $t \in T_n$, every $k \in K_n$ and every $W \in \mathcal{W}(\pi, \psi)$. Note that

$$\delta^{\frac{1}{2}}(t)\delta_0(t)^{-1} = 1$$

for all $t \in T_n$. The expression $\Phi((0, 0, \ldots, 0, 1)tk) = \Phi((0, 0, \ldots, 0, t_n)k)$ does not depend on $t_1, t_2, \ldots, t_{n-1}$. As in the proof of Lemma 5.2, we obtain for some $e > 0$

$$|Z(s, W, \Phi)| \leq \mu(W) \prod_{j=1}^{n-1} \int_1^\infty t_j^{j(2\lambda+\mathrm{Re}(s))}(1 + \log t)_j^e\, d^\times t_j$$
$$\times \int\limits_{K_n(\mathbb{R})} \int_0^\infty t_n^{nRe(s)}(1 + |\log t_n|)^e \Phi((0, 0, \ldots, 0, t_n)k)\, d^\times t_n\, dk.$$

Thus, $Z(s, W, \Phi)$ converges absolutely for $\mathrm{Re}(s) > \max(-2\lambda, 0)$. Since $\lambda > -1/2$ for π unitary and generic (see [Lapid and Mao 2014, p. 8]), the integral converges absolutely for $\mathrm{Re}(s) \geq 1$. For such s, there is a continuous seminorm ν on $\mathcal{S}(\mathbb{R}^n)$

such that

$$\left| \int_{K_n(\mathbb{R})} \int_0^\infty t_n^{nRe(s)} (1 + |\log t_n|)^e \Phi((0, 0, \ldots, 0, t_n)k) \, d^\times t_n \, dk \right| \le \nu(\Phi)$$

for every $\Phi \in \mathcal{S}(\mathbb{R}^n)$. As a consequence, there exist a continuous seminorm μ on $\mathcal{S}(\mathbb{R}^n)$ and a continuous seminorm ν on $\mathcal{W}(\pi, \psi)$ such that

$$|Z(s, W, \Phi)| \le \mu(\Phi)\nu(W)$$

for every $\Phi \in \mathcal{S}(\mathbb{R}^n)$ and every $W \in \mathcal{W}(\pi, \psi)$. Thus there exist $M_1, M_2 > 0$ and continuous seminorms μ' on $\mathcal{S}(\mathbb{R}^n)$ and ν' on $\mathcal{W}(\pi, \psi)$ such that

$$|Z(s, \pi(g)W, R(g)\Phi)| \le \mu(R(g)\Phi)\nu(\pi(g)W) \le \|g\|_H^{M_1} \|g\|_H^{M_2} \mu'(\Phi)\nu'(W)$$

for every $g \in G$ and every $W \in \mathcal{W}(\pi, \psi)$. □

The next lemma provides a functional equation for archimedean Asai integrals.

Lemma 7.2. *Let π be an irreducible, unitary, nondegenerate, $G_n(\mathbb{R})$-distinguished representation of $G_n(\mathbb{C})$. For every $\Phi \in \mathcal{S}(\mathbb{C}^n)$ and $W \in \mathcal{W}(\pi, \psi)$ we have*

$$Z(1, \tilde{W}, \hat{\Phi}|_{\mathbb{R}^n}) = c(\pi) Z(1, W, \Phi|_{\mathbb{R}^n}).$$

Proof. For the proof see [Offen 2011, Lemma 4.2]. □

We will use the following technical result in the next section.

Lemma 7.3. *Let (π', V') be nondegenerate unitary representation of $G_n(\mathbb{C})$ and let $\mathcal{W}(\pi', \psi^{-1})$ be its Whittaker model. Then there exists an $N > 0$ such that for every irreducible, $G_n(\mathbb{R})$-distinguished representation (π, V) of $G_n(\mathbb{C})$ with decay faster than N and every function $f \in \mathcal{C}_{G_n(\mathbb{R})}(\pi)$, the following integral is absolutely convergent*:

$$\int_{G_n(\mathbb{R})\backslash G_n(\mathbb{C})} |f(g)| |\det g|_{\mathbb{C}}^{\frac{1}{2}} \left(\int_{U_n(\mathbb{R})\backslash G_n(\mathbb{R})} |W'(hg)\Phi((0, 0, \ldots, 0, 1)hg)| |\det h|_{\mathbb{R}} \, dh \right) dg.$$

Proof. This is an immediate consequence of Lemma 7.1. □

8. Equality of two functionals

Let $(\pi, V) \in \mathrm{Irr}_{G_n(\mathbb{R})}(G_n(\mathbb{C}))$ be generic and unitarizable and let $\mathcal{W}(\pi, \psi)$ be its Whittaker model. Define linear functionals $\mu, \tilde{\mu} \in V^*$ on $\mathcal{W}(\pi, \psi))$ by

$$\mu : W \mapsto \int_{U_{n-1}(\mathbb{R})\backslash G_{n-1}(\mathbb{R})} W(g) \, dg \quad \text{and} \quad \tilde{\mu} : W \mapsto \int_{U_{n-1}(\mathbb{R})\backslash G_{n-1}(\mathbb{R})} W\left(\begin{pmatrix} 0 & 1 \\ I_{n-1} & 0 \end{pmatrix} \begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix} \right) dg.$$

Since $\mu, \tilde{\mu} \in (V^*)^{P_n(\mathbb{R})}$ and $(V^*)^{P_n(\mathbb{R})} = (V^*)^{G_n(\mathbb{R})}$ (see [Kemarsky 2015, Theorem 1.1]), we obtain that $\mu, \tilde{\mu} \in (V^*)^{G_n(\mathbb{R})}$. Clearly, the functionals $\mu, \tilde{\mu}$ are nonzero. The space of $G_n(\mathbb{R})$-invariant continuous functionals on V is one-dimensional (see [Aizenbud and Gourevitch 2009, Theorem 8.2.5]), thus there exists a proportionality constant $c(\pi) \neq 0$ such that $\tilde{\mu} = c(\pi)\mu$.

The goal of this section is to calculate the proportionality factor $c(\pi)$ by proving the following theorem.

Theorem 8.1. *Let $(\pi, V) \in \mathrm{Irr}_{G_n(\mathbb{R})}(G_n(\mathbb{C}))$. Then $c(\pi) = 1$.*

We now state an archimedean analogue of [Offen 2011, Lemma 6.1].

Lemma 8.2. *Let $\pi' \in \mathrm{Irr}_{G_n(\mathbb{R})}(G_n(\mathbb{C}))$ be generic and unitarizable. Then there exists a generic and unitarizable $\pi \in \mathrm{Irr}_{G_n(\mathbb{R})}(G_n(\mathbb{C}))$ such that*

$$\gamma\left(\frac{1}{2}, \pi \times \pi'; \psi\right) = c(\pi').$$

Note that for π, π' as in Lemma 8.2 we already know that $\gamma(1/2, \pi \times \pi', \psi) = 1$. As a result, the equality $c(\pi') = 1$ follows.

The proof of Lemma 8.2 is similar to the proof of [Offen 2011, Lemma 6.1]. However, in the archimedean case, there are convergence issues that we need to check.

Proof of Lemma 8.2. Let $W \in \mathcal{W}(\pi, \psi)$, $W' \in \mathcal{W}(\pi', \psi^{-1})$, and $\Phi \in \mathcal{S}(\mathbb{C}^n)$. The idea is to prove an equality of Rankin–Selberg integrals of the type

$$\Psi\left(\frac{1}{2}, \tilde{W}, \tilde{W}'; \hat{\Phi}\right) = c(\pi')\Psi\left(\frac{1}{2}, W, W'; \Phi\right). \tag{8-1}$$

Actually, it is enough to prove such an equality for a pair of functions $W \in \mathcal{W}(\pi, \psi)$ and $W' \in \mathcal{W}(\pi', \psi)$ such that at least one of the integrals $\Psi(1/2, \tilde{W}, \tilde{W}'; \hat{\Phi})$, $\Psi(1/2, W, W'; \Phi)$ is nonzero (and thus both integrals are nonzero).

We will obtain the necessary convergence estimates for every K_n-finite function $W \in \mathcal{W}(\pi, \psi)$ and every function $W' \in \mathcal{W}(\pi', \psi)$. By our classification of $G_n(\mathbb{R})$-distinguished representations of $G_n(\mathbb{C})$, the central character ω_π of the $G_n(\mathbb{R})$-distinguished representation satisfies $\omega_\pi(-1) = 1$. Thus, by Theorem 5.1, we have the equality

$$\Psi(1-s, \tilde{W}, \tilde{W}'; \hat{\Phi}) = \gamma(s, \pi \times \pi', \psi)\Psi(s, W, W'; \Phi).$$

Let $f \in \mathcal{C}_{G_n(\mathbb{R})}(\pi)$ be such that

$$W(g) = \int_{U_n(\mathbb{R})\backslash U_n(\mathbb{C})} f(ug)\psi^{-1}(u)\, du.$$

We will prove the absolute convergence of the following integrals at $s = 1/2$:

$$\int_{U_n(\mathbb{C})\backslash G_n(\mathbb{C})} |W(g)W'(g)\Phi((0,0,\dots,0,1)g)||\det g|_{\mathbb{C}}^{s}\,dg \tag{8-2}$$
$$\le \int_{U_n(\mathbb{C})\backslash G_n(\mathbb{C})} \biggl(\int_{U_n(\mathbb{R})\backslash U_n(\mathbb{C})} |f(ug)|du \biggr) \times |W'(g)\Phi((0,0,\dots,0,1)g)||\det g|_{\mathbb{C}}^{s}\,dg$$
$$= \int_{U_n(\mathbb{C})\backslash G_n(\mathbb{C})} |f(g)W'(g)\Phi((0,0,\dots,0,1)g)||\det g|_{\mathbb{C}}^{s}\,dg$$
$$= \int_{G_n(\mathbb{R})\backslash G_n(\mathbb{C})} |f(g)||\det g|_{\mathbb{C}}^{s} \times \biggl(\int_{U_n(\mathbb{R})\backslash G_n(\mathbb{R})} |W'(hg)\Phi((0,0,\dots,0,1)hg)||\det h|_{\mathbb{R}}^{2s}\,dh \biggr) dg.$$

The left-hand side of (8-2) is absolutely convergent by Lemma 5.2 and the integrals on the right-hand side of (8-2) are absolutely convergent by Lemmas 7.1 and 7.3. Using absolute convergence for $s = 1/2$ of the integrals appearing in (8-2) we obtain the equality (8-1) by the following argument:

$$\Psi\left(\frac{1}{2}, W, W'; \Phi\right)$$
$$= \int_{U_n(\mathbb{C})\backslash G_n(\mathbb{C})} W(g)W'(g)\Phi((0,0,\dots,0,1)g)|\det g|_{\mathbb{C}}^{\frac{1}{2}}\,dg$$
$$= \int_{U_n(\mathbb{C})\backslash G_n(\mathbb{C})} \biggl(\int_{U_n(\mathbb{R})\backslash U_n(\mathbb{C})} f(ug)\psi_n^{-1}(u)du \biggr) W'(g)\Phi((0,0,\dots,0,1)g)|\det g|_{\mathbb{C}}^{\frac{1}{2}}\,dg$$
$$= \int_{U_n(\mathbb{C})\backslash G_n(\mathbb{C})} f(g)W'(g)\Phi((0,0,\dots,0,1)g)|\det g|_{\mathbb{C}}^{\frac{1}{2}}\,dg$$
$$= \int_{G_n(\mathbb{R})\backslash G_n(\mathbb{C})} f(g)|\det g|_{\mathbb{C}}^{\frac{1}{2}} \biggl(\int_{U_n(\mathbb{R})\backslash G_n(\mathbb{R})} W'(hg)\Phi((0,0,\dots,0,1)hg)|\det h|_{\mathbb{R}}\,dh \biggr) dg$$
$$= \int_{G_n(\mathbb{R})\backslash G_n(\mathbb{C})} f(g)|\det g|_{\mathbb{C}}^{\frac{1}{2}} Z(1, \pi'(g)W', \Phi(\cdot g)|_{\mathbb{R}^n})\,dg.$$

Define $f^*(g) := f({}^t g^{-1})$. Then, clearly, $f^* \in \mathcal{C}_{G_n(\mathbb{R})}(\tilde{\pi})$. Applying the change of variables $u \to w_n {}^t u^{-1} w_n^{-1}$ and the fact that $f(w_n g) = f(g)$, it follows from the definitions that

$$\tilde{W}(g) = \int_{U_n(\mathbb{R})\backslash U_n(\mathbb{C})} f^*(ug)\psi(u)\,du.$$

The same computation applied to $\tilde{\pi}$ and $\tilde{\pi}'$ yields

$$\Psi\left(\frac{1}{2}, \tilde{W}, \tilde{W}'; \hat{\Phi}\right) = \int_{G_n(\mathbb{R})\backslash G_n(\mathbb{C})} f^*(g)|\det g|_{\mathbb{C}}^{\frac{1}{2}} Z(1, \tilde{\pi}'(g)\tilde{W}', \hat{\Phi}(\cdot g)|_{\mathbb{R}^n})\,dg.$$

By Lemma 7.2,

$$Z(1, \tilde{\pi}'({}^t g^{-1})\tilde{W}', \hat{\Phi}(\cdot\, {}^t g^{-1})|_{\mathbb{R}^n}) = c(\pi')|\det g|_{\mathbb{C}} Z(1, \pi'(g)W', \Phi(\cdot g)|_{\mathbb{R}^n}).$$

Finally, we obtain

$$\Psi\left(\frac{1}{2}, \tilde{W}, \tilde{W}'; \hat{\Phi}\right) = c(\pi')\Psi\left(\frac{1}{2}, W, W'; \Phi\right)$$

for every K_n-finite function $W \in \mathcal{W}(\pi, \psi)$, $W' \in \mathcal{W}(\pi', \psi^{-1})$ and every $\Phi \in \mathcal{S}(\mathbb{C}^n)$. It is well-known that there exist K_n-finite $W \in \mathcal{W}(\pi, \psi)$, $W' \in \mathcal{W}(\pi', \psi^{-1})$ such that $\Psi(1/2, W, W'; \Phi) \neq 0$. It follows that $c(\pi') = \gamma(1/2, \pi \times \pi'; \psi)$. ☐

Appendix A: Generic Langlands quotient

In this section we sketch a proof of the well-known fact that the Langlands quotient of $I(\chi)$ is generic if and only if $I(\chi)$ is irreducible. This fact follows from the papers of Kostant [1978] and Vogan [1978]. For the convenience of the reader we rewrite it here. Similar results for $\mathrm{GL}_n(\mathbb{R})$ were obtained by Casselman and Zuckerman.

Let $\mathfrak{g} = M_n(\mathbb{C})$ be the Lie algebra of $G_n(\mathbb{C})$ and let K be the standard maximal compact subgroup of $G_n(\mathbb{C})$.

Definition. An irreducible $(\mathfrak{g}, K)$-module X is called large if its annihilator in the universal enveloping algebra $U(\mathfrak{g})$ is a minimal primitive ideal. We will say that a smooth irreducible representation (π, V) of $G_n(\mathbb{C})$ is large if the corresponding $(\mathfrak{g}, K)$-module consisting of K-finite vectors in V is large.

Let $\chi = (\chi_1, \chi_2, \dots, \chi_n)$ be a character of $B_n(\mathbb{C})$ and suppose $|\chi_j(t)| = |t|^{\lambda_j}$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. By [Vogan 1978, Theorem 6.2], if (σ, W) is an irreducible subrepresentation of $I(\chi)$ then (σ, W) is large. Suppose (π, V) is the Langlands quotient of $I(\chi)$ and suppose (π, V) is generic. Then by Kostant's theorem (π, V) is large. On the other hand, [Vogan 1978, Corollary 6.7] states that there is a unique large composition factor in the composition series for $I(\chi)$. We obtain $(\pi, V) \simeq (\sigma, W)$ and thus $(\sigma, W) = I(\chi)$; that is, $I(\chi)$ is an irreducible representation.

Appendix B: Gamma factors: converse direction

Fix a smooth, irreducible, generic and admissible representation (π, V) of $G_n(\mathbb{C})$. Suppose we know that

$$\gamma\left(\frac{1}{2}, \pi \times \pi', \psi\right) = 1 \tag{B-1}$$

for every $m \leq k$ and every smooth irreducible $G_m(\mathbb{R})$-distinguished representation (π', V') of $G_m(\mathbb{C})$. What is the minimal k such that (B-1) implies that (π, V) is $G_n(\mathbb{R})$-distinguished? In this section we give an answer to this question in the case when (π, V) is a unitary representation.

In the following two theorems we prove that $k = 1$ is enough. Theorem B.1 is a particular case of Theorem B.2. Nevertheless we state and prove it since the proof of Theorem B.1 is simpler than and may aid in the understanding of the proof of Theorem B.2.

Theorem B.1. *Let $\chi = (\chi_1, \chi_2, \dots, \chi_n)$ be a unitary character of B_n and suppose that $\chi_j(z) = |z|_{\mathbb{C}}^{s_j}(z/|z|)^{k_j}$ with s_j purely imaginary and $k_j \in \mathbb{Z}$ for every $1 \leq j \leq n$. Suppose $(\pi, V) = I(\chi)$ is a smooth, generic and irreducible representation of $G_n(\mathbb{C})$. Finally, suppose*

$$\gamma\left(\frac{1}{2}, \pi \times \chi', \psi\right) = 1$$

for every $\mathbb{R}^\times$-distinguished unitary character $\chi' : \mathbb{C}^\times \to \mathbb{R}^\times$. Then there exists an involution $w \in S_n$ such that $w\chi = \overline{(\chi^{-1})}$. Moreover, one can find an involution $w \in S_n$ such that $w\chi = \overline{(\chi^{-1})}$ and such that, for every fixed point $w(i) = i$, the integer k_i is even.

Proof. Observe that every $\mathbb{R}^\times$-distinguished unitary character $\chi' : \mathbb{C} \to \mathbb{R}^\times$ is of the form $\chi(z) = (z/|z|)^{2m}$ for $m \in \mathbb{Z}$. By [Jacquet 2009, Lemma 16.3] we have

$$\gamma\left(\frac{1}{2}, \operatorname{Ind}(\chi) \times \chi', \psi\right) = \prod_{i=1}^{n} \gamma\left(\frac{1}{2}, \chi_i \chi', \psi\right),$$

where $\gamma(1/2, \chi_i \chi', \psi)$ is the one-dimensional Tate gamma factor. Following Tate [Cassels and Frölich 1967], denote $c_m(z) = (z/|z|)^m$ and recall that the Tate gamma factor is given by

$$\gamma(s, c_m, \psi) = \epsilon_m \frac{(2\pi)^{1-s}\Gamma\left(s + \frac{|m|}{2}\right)}{(2\pi)^s \Gamma\left((1-s) + \frac{|m|}{2}\right)},$$

where

$$\epsilon_m = \begin{cases} 1 & \text{if } m \text{ is even or } m > 0, \\ -1 & \text{if } m \text{ is odd and } m < 0. \end{cases}$$

Let us rewrite the equality $\gamma(1/2, \mathrm{Ind}(\chi)\times\chi', \psi) = 1$ as

$$\text{(B-2)} \qquad \prod_{i=1}^{n} \epsilon_{2m+k_i} \frac{(2\pi)^{\frac{1}{2}-s_i}}{(2\pi)^{\frac{1}{2}+s_i}} \frac{\Gamma\left(\frac{1}{2}+s_i+\frac{|k_i+2m|}{2}\right)}{\Gamma\left(\frac{1}{2}-s_i+\frac{|k_i+2m|}{2}\right)} = 1$$

for every $m \in \mathbb{Z}$. The product in (B-2) breaks into three products:

$$p_{m,1} = \prod_{i=1}^{n} \epsilon_{2m+k_i},$$

$$p_{m,2} = \prod_{i=1}^{n} \frac{(2\pi)^{\frac{1}{2}-s_i}}{(2\pi)^{\frac{1}{2}+s_i}} = (2\pi)^{-2s_1-2s_2-\cdots-2s_n},$$

$$p_{m,3} = \prod_{i=1}^{n} \frac{\Gamma\left(\frac{1}{2}+s_i+\frac{|k_i+2m|}{2}\right)}{\Gamma\left(\frac{1}{2}-s_i+\frac{|k_i+2m|}{2}\right)}.$$

Note that the term $p_{m,2}$ is constant (does not depend on m) and the term $p_{m,1}$ stabilizes (that is, $p_{m,1} = p_{m+1,1}$ for large enough and for small enough m). Also, we have $|k_i+m| = k_i+m$ for m large enough. Let us take m large enough and look at the expression

$$\frac{p_{m+1,1}p_{m+1,2}p_{m+1,3}}{p_{m,1}p_{m,2}p_{m,3}}.$$

By our assumption this fraction equals 1 for every m. For m large enough we have $p_{m+1,1}p_{m+1,2} = p_{m,1}p_{m,2}$, so $p_{m+1,3}/p_{m,3} = 1$. By the functional equation $\Gamma(z+1) = z\Gamma(z)$ we obtain

$$1 = \frac{p_{m+1,3}}{p_{m,3}} = \prod_{i=1}^{n} \frac{\left(\frac{1}{2}+s_i+\frac{k_i+2m}{2}\right)}{\left(\frac{1}{2}-s_i+\frac{k_i+2m}{2}\right)}.$$

Thus,

$$\prod_{i=1}^{n}\left(\frac{1}{2}+s_i+\frac{k_i+2m}{2}\right) = \prod_{i=1}^{n}\left(\frac{1}{2}-s_i+\frac{k_i+2m}{2}\right)$$

for large enough $m \in \mathbb{Z}$. Since both sides are polynomials in m, the polynomials are equal. As a consequence, the zeros of these two polynomials coincide; that is, for every $1 \le i \le n$ there exists a $1 \le j \le n$ such that

$$\frac{1}{2} - s_i + \frac{k_i}{2} = \frac{1}{2} + s_j + \frac{k_j}{2}.$$

By our assumption the s_is are purely imaginary and the k_is are integers. Thus, $-s_i = s_j$ and $k_i = k_j$. Note that $\bar{s_i} = -s_i$ and this means exactly that we can define $w(i) = j$, $w(j) = i$ and $\chi_j = \bar{\chi_i}^{-1} = \chi_{w(i)}$. Therefore, there exists an involution $w \in S_n$ such that $w(\chi) = \bar{\chi}^{-1}$.

From the proof of the existence of an involution w it follows that $\sum_{j=1}^{n} s_j = 0$ and that the products $p_{m,2} = 1$ and $p_{m,3} = 1$ for every $m \in \mathbb{Z}$. This establishes the existence of an involution $w \in S_n$ such that $w(\chi) = \bar{\chi}^{-1}$. It remains to establish the second property: existence of an involution such that, in addition, for every fixed point $w(j) = j$ of the involution the corresponding integer k_j is even. Note that if i is a fixed point of w then $s_i = 0$. Without loss of generality assume that if $w(i) = i$ and $w(j) = j$, then $k_i \neq k_j$. Otherwise we can define an involution w' by $w'(i) = j,\ w'(j) = i$ and $w'(l) = w(l)$ for $l \neq i, j$ and the new involution w' also satisfies $w'(\chi) = \bar{\chi}^{-1}$.

Assume on the contrary that $w(i) = i$ but that k_i is odd. Then take two consecutive products $p_{m,1}$ and $p_{m+1,1}$ for $m = (-k_i - 1)/2$. Observe that $\epsilon_{2m+k_i} = -\epsilon_{2(m+1)+k_i}$ and that the other terms appearing in the products $p_{m,1}$ and $p_{m+1,1}$ equal each other respectively. As a consequence, $p_{m+1,1} = -p_{m,1}$. But from the preceding paragraph we have $p_{m,2} = p_{m+1,2} = 1$ and also $p_{m,3} = p_{m+1,3} = 1$ and thus $p_{m+1,1} = p_{m,1} = 1$. Contradiction!

Therefore, if $w(i) = i$ then the integer k_i is even; that is, $\chi_i(-1) = 1$. □

A small modification of this proof gives a stronger theorem.

Theorem B.2. *Let $\chi = (\chi_1, \chi_2, \dots, \chi_n)$ be a character of B_n and suppose that $\chi_j(z) = |z|_{\mathbb{C}}^{s_j}(z/|z|)^{k_j}$ with $-1/2 < \mathrm{Re}(s_j) < 1/2$ and $k_j \in \mathbb{Z}$ for every $1 \leq j \leq n$. Suppose $(\pi, V) = I(\chi)$ is a smooth, generic, irreducible representation of $G_n(\mathbb{C})$. Finally, suppose*

$$\gamma\left(\frac{1}{2}, \pi \times \chi', \psi\right) = 1$$

for every $\mathbb{R}^\times$-distinguished unitary character $\chi' : \mathbb{C}^\times \to \mathbb{R}^\times$. Then there exists an involution $w \in S_n$ such that $w\chi = \overline{(\chi^{-1})}$. Moreover, one can find an involution $w \in S_n$ such that $w\chi = \overline{(\chi^{-1})}$ and such that, for every fixed point $w(i) = i$, the integer k_i is even.

Proof. By the same argument as in the previous theorem we obtain that for every $1 \leq i \leq n$ there exists a $1 \leq j \leq n$ such that

$$\frac{1}{2} - s_i + \frac{k_i}{2} = \frac{1}{2} + s_j + \frac{k_j}{2}.$$

By subtracting $1/2$ from both sides of this equality and taking real parts we can replace s_j by $\mathrm{Re}(s_j)$. Thus we can assume that for every $1 \leq i \leq n$ we have $-1/2 < s_i < 1/2$ and also for every $1 \leq i \leq n$ there exists a $1 \leq j \leq n$ such that $-s_i + k_i/2 = s_j + k_j/2$. Multiply both sides of this equation by 2 and replace s_i by $2s_i$. Then, we can assume that for every $1 \leq i \leq n$ we have $-1 < s_i < 1$ and also for every $1 \leq i \leq n$ there exists a $1 \leq j \leq n$ such that

$$-s_i + k_i = s_j + k_j.$$

Let us call this condition the "antisymmetry condition". The claim is that the "antisymmetry condition" implies that there exists an involution $w \in S_n$ such that $w(\chi) = \bar{\chi}^{-1}$; that is, if $w(i) = j$ then $s_i = -s_j$ and $k_i = k_j$. The proof of the existence of an involution w is by induction on n. Clearly, for $n = 1$ the condition $-s_1 + k_1 = s_1 + k_1$ gives us $s_1 = 0$ and thus the identity involution $w(1) = 1$ works. For a general n it is enough that the "antisymmetry condition" implies that there is a pair $i,\ j$ such that $s_i = -s_j$ and $k_i = k_j$. Note that if $i = j$ then $s_i = -s_i$ implies $s_i = 0$.

Suppose on the contrary that there are $\{s_i\}_{i=1}^n \subset (-1, 1)$ and $\{k_i\}_{i=1}^n \subset \mathbb{Z}$ that satisfy the "antisymmetry condition", but there is no pair of indices $1 \leq i,\ j \leq n$ that satisfy $s_i = -s_j$ and $k_i = k_j$. In particular, there is some $1 \leq i \leq n$ such that $-s_1 + k_1 = s_i + k_i$. By our assumption we have $i > 1$, so without loss of generality assume $i = 2$. Let us assume $s_1 > 0$; the proof in the case $s_1 < 0$ is similar and $s_1 = 0$ is not possible by our assumption. We obtain $k_1 - k_2 = s_1 + s_2$. The left-hand side is an integer and we have $-1 < s_1 + s_2 < 2$. Thus $s_1 + s_2 = 0$ or $s_1 + s_2 = 1$. The case $s_1 + s_2 = 0$ is not possible by our assumption, thus $s_1 + s_2 = 1$ and as a corollary $s_2 > 0$ and $k_2 = k_1 - 1$. Similarly, there is some $1 \leq i \leq n$ such that $-s_2 + k_2 = s_i + k_i$. By the same argument we obtain $s_i > 0$ and $k_i = k_2 - 1$. Thus $i \neq 1, 2$ and without loss of generality we can assume $i = 3$. Continuing in this manner we obtain an infinite sequence of integers k_j such that $k_j = k_1 + (j - 1)$. Contradiction!

Thus there is a pair of indices $1 \leq i,\ j \leq n$ such that $s_i = -s_j$ and $k_i = k_j$. Removing them from our sequence of length n we obtain a shorter sequence which satisfies the "antisymmetry condition".

Thus, we have proved that there is an involution $w \in S_n$ such that $w(\chi) = \bar{\chi}^{-1}$. The rest of the argument, that is, the proof of the existence of an involution w such that for every fixed point j of the involution the corresponding integer k_j is even, is the same as in the proof of the previous theorem. $\square$

As a corollary, using the Tadic–Vogan classification of the unitary dual of $G_n(\mathbb{C})$, we obtain the following theorem.

Theorem B.3. *Let $\chi = (\chi_1, \chi_2, \ldots, \chi_n)$ be a character of B_n and suppose that $(\pi, V) = \mathrm{Ind}(\chi)$ is a smooth, generic, irreducible, and unitary representation of $G_n(\mathbb{C})$. Suppose*

$$\gamma\left(\frac{1}{2}, \pi \times \chi', \psi\right) = 1$$

for every $\mathbb{R}^\times$-distinguished unitary character $\chi' : \mathbb{C}^\times \to \mathbb{R}^\times$. Then there exists an involution $w \in S_n$ such that $w\chi = \overline{(\chi^{-1})}$. Moreover, one can find an involution $w \in S_n$ such that $w\chi = \overline{(\chi^{-1})}$ and such that, for every fixed point $w(i) = i$, the integer k_i is even.

Proof. Let us define $\chi_j(z) = |z|_{\mathbb{C}}^{s_j}(z/|z|)^{k_j}$, where $s_j \in \mathbb{C}$ and $k_j \in \mathbb{Z}$. The theorem follows from Theorem B.2 and the fact that the unitaricity of $\mathrm{Ind}(\chi)$ implies $-1/2 < \mathrm{Re}(s_j) < 1/2$ for every $1 \le j \le n$ (see [Tadić 1985, Theorem A]). □

Finally, by [Panichi 2001, Theorem 3.3.6] we know that an irreducible tempered representation (π, V) of $G_n(\mathbb{C})$ is $G_n(\mathbb{R})$-distinguished if and only if there exists an involution $w \in S_n$ such that $w\chi = \overline{(\chi^{-1})}$ and such that, for every fixed point $w(i) = i$, the integer k_i is even. Therefore, an irreducible tempered representation (π, V) of $G_n(\mathbb{C})$ is $G_n(\mathbb{R})$-distinguished if and only if

$$\gamma\left(\frac{1}{2}, \pi \times \chi', \psi\right) = 1$$

for every $\mathbb{R}^\times$-distinguished unitary character $\chi' : \mathbb{C}^\times \to \mathbb{R}^\times$.

Acknowledgements

I would like to thank Omer Offen for posing to me this question and providing many explanations on the subject.

I am grateful to Dmitry Gourevitch and Erez Lapid for many fruitful discussions and their help.

During the conference in Jussieu, June 2014, I told the results of this paper to Herve Jacquet. I would like to thank him, as his comments were very helpful.

I also wish to thank Avraham Aizenbud, Uri Bader, Max Gurevich, Job Kuit, Nadir Matringe, Amos Nevo, Henrik Schlictkrull and Dror Speiser for useful conversations and remarks.

References

[Aizenbud and Gourevitch 2009] A. Aizenbud and D. Gourevitch, "Generalized Harish-Chandra descent, Gelfand pairs, and an Archimedean analog of Jacquet–Rallis's theorem", *Duke Math. J.* **149**:3 (2009), 509–567. MR 2011c:22026 Zbl 1221.22018

[Aizenbud and Lapid 2012] A. Aizenbud and E. Lapid, "Distinguished representations in the Archimedean case", 2012. Appendix D to B. Feigon, E. Lapid, and O. Offen, "On representations distinguished by unitary groups", *Publ. Math. Inst. Hautes Études Sci.* **115** (2012), 185–323. MR 2930996 Zbl 06112182

[Aizenbud et al. 2008] A. Aizenbud, D. Gourevitch, and E. Sayag, "$(\mathrm{GL}_{n+1}(F), \mathrm{GL}_n(F))$ is a Gelfand pair for any local field F", *Compos. Math.* **144**:6 (2008), 1504–1524. MR 2009k:22022 Zbl 1157.22004

[Cassels and Frölich 1967] J. W. S. Cassels and A. Frölich, *Algebraic number theory*, Academic Press, London, 1967. MR 35 #6500 Zbl 0153.07403

[Cowling et al. 1988] M. Cowling, U. Haagerup, and R. Howe, "Almost L^2 matrix coefficients", *J. Reine Angew. Math.* **387** (1988), 97–110. MR 89i:22008 Zbl 0638.22004

[Flensted-Jensen 1980] M. Flensted-Jensen, "Discrete series for semisimple symmetric spaces", *Ann. of Math.* (2) **111**:2 (1980), 253–311. MR 81h:22015 Zbl 0462.22006

[Flicker 1988] Y. Z. Flicker, "Twisted tensors and Euler products", *Bull. Soc. Math. France* **116**:3 (1988), 295–313. MR 89m:11049 Zbl 0674.10026

[Flicker 1993] Y. Z. Flicker, "On zeroes of the twisted tensor L-function", *Math. Ann.* **297**:2 (1993), 199–219. MR 95c:11065 Zbl 0786.11030

[Jacquet 2009] H. Jacquet, "Archimedean Rankin–Selberg integrals", pp. 57–172 in *Automorphic forms and L-functions II: Local aspects*, edited by D. Ginzburg et al., Contemp. Math. **489**, Amer. Math. Soc., Providence, RI, 2009. MR 2011a:11103 Zbl 1250.11051

[Kassel and Kobayashi 2013] F. Kassel and T. Kobayashi, "Poincaré series for non-Riemannian locally symmetric spaces", preprint, 2013. arXiv 1209.4075

[Kemarsky 2015] A. Kemarsky, "Distinguished representations of $GL_n(\mathbb{C})$", *Israel J. Math.* **207**:1 (2015), 435–448. MR 3358053 Zbl 06451119

[Kobayashi 2007] T. Kobayashi, "A generalized Cartan decomposition for the double coset space $(\mathrm{U}(n_1) \times \mathrm{U}(n_2) \times \mathrm{U}(n_3))\backslash \mathrm{U}(n)/(\mathrm{U}(p) \times \mathrm{U}(q))$", *J. Math. Soc. Japan* **59**:3 (2007), 669–691. MR 2008k:22017 Zbl 1124.22003

[Kostant 1978] B. Kostant, "On Whittaker vectors and representation theory", *Invent. Math.* **48**:2 (1978), 101–184. MR 80b:22020 Zbl 0405.22013

[Lapid and Mao 2014] E. Lapid and Z. Mao, "On a new functional equation for local integrals", pp. 261–294 in *Automorphic forms and related geometry: assessing the legacy of I. I. Piatetski-Shapiro*, edited by J. W. Cogdell et al., Contemp. Math. **614**, Amer. Math. Soc., Providence, RI, 2014. MR 3220931 Zbl 1306.22008

[Lapid and Rogawski 2003] E. M. Lapid and J. D. Rogawski, "Periods of Eisenstein series: the Galois case", *Duke Math. J.* **120**:1 (2003), 153–226. MR 2004m:11077 Zbl 1037.11033

[Offen 2011] O. Offen, "On local root numbers and distinction", *J. Reine Angew. Math.* **652** (2011), 165–205. MR 2012c:22025 Zbl 1278.22006

[Ok 1997] Y. Ok, *Distinction and gamma factors at* 1/2*: supercuspidal case*, Ph.D. thesis, Columbia University, 1997. MR 2716711

[Panichi 2001] M. N. Panichi, *Charactérisations du spectre tempéré de $GL_n(\mathbb{C})/GL_n(\mathbb{R})$*, Ph.D. thesis, Paris 7 University, 2001.

[Serre 2002] J.-P. Serre, *Galois cohomology*, 2nd ed., Springer, Berlin, 2002. MR 2002i:12004 Zbl 1004.12003

[Springer 1984] T. A. Springer, *Some results on algebraic groups with involutions*, Publications mathématiques de l'Université Pierre et Marie Curie **61**, Mathematisch Institut der Rÿksuniversiteit, 1984.

[Tadić 1985] M. Tadić, "Unitary representations of general linear group over real and complex field", preprint 22, Max Planck Institut für Mathematik, 1985, available at https://www.mpim-bonn.mpg.de/preblob/5395.

[Vogan 1978] D. A. Vogan, Jr., "Gelfand–Kirillov dimension for Harish-Chandra modules", *Invent. Math.* **48**:1 (1978), 75–98. MR 58 #22205 Zbl 0389.17002

[Wallach 1988] N. R. Wallach, *Real reductive groups, I*, Pure and Applied Mathematics **132**, Academic Press, Boston, 1988. MR 89i:22029 Zbl 0666.22002

[Wallach 1992] N. R. Wallach, *Real reductive groups, II*, Pure and Applied Mathematics **132**, Academic Press, Boston, 1992. MR 93m:22018 Zbl 0785.22001

Received October 12, 2014. Revised March 12, 2015.

ALEXANDER KEMARSKY
MATHEMATICS DEPARTMENT
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY
32000 HAIFA
ISRAEL
alexkem@tx.technion.ac.il

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.278.173

THE W-ENTROPY FORMULA FOR THE WITTEN LAPLACIAN ON MANIFOLDS WITH TIME DEPENDENT METRICS AND POTENTIALS

SONGZI LI AND XIANG-DONG LI

We develop a new approach to prove the W-entropy formula for the Witten Laplacian via warped product on Riemannian manifolds, giving a natural geometric interpretation of the quantities appearing in the W-entropy formula. We also prove the W-entropy formula for the Witten Laplacian on compact Riemannian manifolds with time dependent metrics and potentials, as well as for the backward heat equation associated with the Witten Laplacian on compact Riemannian manifolds equipped with Lott's modified Ricci flow. Our results extend to complete Riemannian manifolds with negative m-dimensional Bakry–Émery Ricci curvature, and to compact Riemannian manifolds with K-super m-dimensional Bakry–Émery Ricci flow. As an application, we prove that the optimal logarithmic Sobolev constant on compact manifolds equipped with the K-super m-dimensional Bakry–Émery Ricci flow is decreasing in time.

1. Introduction

Let M be a complete Riemannian manifold with a fixed Riemannian metric g and a fixed potential $\phi \in C^2(M)$. Let $d\mu = e^{-\phi}\, dv$, where dv is the Riemannian volume measure on (M, g). The Witten Laplacian (also called the weighted Laplacian),

$$L = \Delta - \nabla\phi \cdot \nabla,$$

is a self-adjoint and nonnegative operator on $L^2(M, \mu)$. By Itô's calculus, one can construct the symmetric diffusion process X_t associated to the Witten Laplacian by solving the SDE

$$dX_t = \sqrt{2}\, dW_t - \nabla\phi(X_t)\, dt,$$

S. Li was partially supported by the China Scholarship Council. X.-D. Li was supported by NSFC No. 11371351, Key Laboratory of Random Complex Structures and Data Science, AMSS, CAS, No. 2008DP173182, and a Hundred Talents Project of AMSS, CAS.

MSC2010: primary 53C44, 58J35, 58J65; secondary 60J60, 60H30.

Keywords: Bakry–Émery Ricci curvature, W-entropy, Witten Laplacian, modified Ricci flow.

where W_t is the Brownian motion on M. Moreover, it is well known that the transition probability density function of the diffusion process X_t is exactly the fundamental solution to the heat equation of L, i.e., the heat kernel of the Witten Laplacian L. In view of this, it is a fundamental problem to study the heat equation and related properties for the Witten Laplacian on manifolds.

In recent years, important progress has been made in the study of the heat equation associated with the Witten Laplacian by using new ideas and new methods from geometric analysis, PDEs and probability theory. In particular, F. Otto [2001] introduced an infinite dimensional Riemannian structure on the Wasserstein space of probability measures on $\mathbb{R}^n$ and proved that the heat equation

$$\partial_t u = Lu \tag{1}$$

can be realized as the reverse gradient flow of the Boltzmann–Shannon entropy[1]

$$H(u) = -\int_M u \log u \, d\mu.$$

See also [Otto and Villani 2000; Sturm 2005; von Renesse and Sturm 2005; Villani 2003; 2009] for the extension of Otto's work to Riemannian manifolds.

The Witten Laplacian is a natural extension of the standard Laplace–Beltrami operator and has a close connection to differential geometry, probability theory, quantum field theory and statistical mechanics. In view of this, it is natural to raise the question whether one can extend the results which hold for the standard Laplace–Beltrami operator to the Witten Laplacian on manifolds. The main tool which makes such an extension possible is the so-called Bakry–Émery Ricci curvature associated to L [Bakry and Émery 1985],

$$\mathrm{Ric}(L) = \mathrm{Ric} + \nabla^2 \phi,$$

which plays the same role as the Ricci curvature for the standard Laplace–Beltrami operator. We refer the reader to [Bakry and Qian 1999; Bakry and Ledoux 2006; Li 2005] for the Li–Yau Harnack estimates and the heat kernel estimates to the heat equation (1), and to [Li 2005] for the extension of S.-T. Yau's Strong Liouville theorem for the positive L-harmonic functions and the L^1-uniqueness of the heat equation on complete Riemannian manifolds. See also [Andrews and Ni 2012; Bakry and Qian 2000; Fang et al. 2009; Futaki et al. 2013; Otto and Villani 2000; von Renesse and Sturm 2005; Villani 2003; 2009; Wei and Wylie 2009] for other results on the Witten Laplacian and Bakry–Émery Ricci curvature on manifolds with weighted measures.

[1] Equivalently, the heat equation (1) is the gradient flow of $\mathrm{Ent}(u) = -H(u)$ on the Wasserstein space $\mathcal{P}_2(\mathbb{R}^n)$ equipped with Otto's infinite dimensional Riemannian metric.

The Bakry–Émery Ricci curvature has been essentially used in Perelman's work on the entropy formula for Ricci flow. Perelman [2002] first introduced the $\mathscr{F}$-functional on the space of Riemannian metrics and smooth functions, i.e., $\mathcal{M} = \{\text{Riemannian metric } g \text{ on } M\} \times C^\infty(M)$, as follows:

$$\mathscr{F}(g, f) = \int_M (R + |\nabla f|^2) e^{-f}\, dv,$$

where $f \in C^\infty(M)$, R denotes the scalar curvature on (M, g), and dv denotes the volume measure. Under the constraint condition which requires that

$$dm = e^{-f}\, dv$$

is a fixed weighted measure on (M, g), Perelman proved that the gradient flow of $\mathscr{F}$ with respect to the standard L^2-metric on $\mathcal{M}$ is given by the modified Ricci flow

$$\partial_t g = -2(\mathrm{Ric} + \nabla^2 f),$$

and f satisfies the so-called conjugate heat equation

$$\partial_t f = -\Delta f - R.$$

Moreover, Perelman [2002] introduced the W-entropy and proved its monotonicity for the Ricci flow on compact manifolds. This result plays an important role in the proof of the no local collapsing theorem and in the final resolution of the Poincaré and geometrization conjectures (see also [Cao and Zhu 2006; Morgan and Tian 2007; Kleiner and Lott 2008]). Since then, many people have derived the W-entropy formula for various geometric evolution equations and used it to study further analysis and geometric properties of manifolds. See, e.g., [Chow et al. 2006; Chang et al. 2011; Kleiner and Lott 2008; Ni 2004a; 2004b; Ecker 2007; Lu et al. 2009; Kotschwar and Ni 2009].

In [Li 2012] (see also [Li 2007; 2011; 2014]), inspired by Perelman's work on the W-entropy formula for Ricci flow, the second author proved the W-entropy formula for the fundamental solution of the Witten Laplacian on complete Riemannian manifolds with the bounded geometry condition. This extends a previous result due to Ni [2004b; 2004a], who proved an analogue of Perelman's W-entropy formula for the heat equation $\partial_t u = \Delta u$ on complete Riemannian manifolds with a fixed metric. More precisely, we have:

Theorem 1.1 [Li 2007; 2012; 2011; 2014]. *Let (M, g) be a compact Riemannian manifold, or a complete Riemannian manifold with the bounded geometry condition,*[2] *and $\phi \in C^4(M)$ with $\nabla\phi \in C_b^3(M)$. Let $m \geq n$, and $u = e^{-f}/(4\pi t)^{m/2}$ be a positive solution of the heat equation $\partial_t u = Lu$ when M is compact, or the*

[2] We say that (M, g) satisfies the bounded geometry condition if the Riemannian curvature tensor Riem and its covariant derivatives ∇^k Riem are uniformly bounded on M, $k = 1, 2, 3$.

fundamental solution associated with the Witten Laplacian, i.e., the heat kernel to the heat equation $\partial_t u = Lu$, *when* M *is complete. Let*

$$H_m(u,t) = -\int_M u \log u \, d\mu - \frac{m}{2}(1+\log(4\pi t)).$$

Define

$$W_m(u,t) = \frac{d}{dt}(tH_m(u)).$$

Then

$$W_m(u,t) = \int_M \big(t|\nabla f|^2 + f - m\big)\frac{e^{-f}}{(4\pi t)^{m/2}}\, d\mu \tag{2}$$

and

$$\begin{aligned}\frac{d}{dt}W_m(u,t) = {}& -2\int_M t\Big(\Big|\nabla^2 f - \frac{g}{2t}\Big|^2 + \mathrm{Ric}_{m,n}(L)(\nabla f,\nabla f)\Big)u\, d\mu \\ & - \frac{2}{m-n}\int_M t\Big(\nabla\phi\cdot\nabla f + \frac{m-n}{2t}\Big)^2 u\, d\mu,\end{aligned} \tag{3}$$

where

$$\mathrm{Ric}_{m,n}(L) = \mathrm{Ric} + \nabla^2\phi - \frac{\nabla\phi\otimes\nabla\phi}{m-n}$$

is the m*-dimensional Bakry–Émery Ricci curvature associated with the Witten Laplacian* L.

In particular, if (M, g, ϕ) *is compact or satisfies the bounded geometry condition and* $\mathrm{Ric}_{m,n}(L) \geq 0$, *then the W-entropy is decreasing in time* t, *i.e.*,

$$\frac{d}{dt}W_m(u,t) \leq 0, \quad \textit{for all } t \geq 0.$$

The purpose of this paper is to extend the W-entropy formula in Theorem 1.1 to the heat equation (1) associated with the time dependent Witten Laplacian on compact Riemannian manifolds equipped with time dependent metrics and potentials. In view of Perelman's work using the W-entropy formula for the Ricci flow to remove "the major stumbling block in Hamilton's approach to geometrization" [Perelman 2002], it might be possible that the W-entropy formula for the time dependent Witten Laplacian can bring some new insights to the study of geometric analysis on Riemannian manifolds with time dependent metrics and potentials. Our results can be regarded as the m-dimensional analogue of Perelman's results for the Ricci flow, where the Ricci curvature for the Ricci flow is replaced by the m-dimensional Bakry–Émery Ricci curvature, and the Laplacian is replaced by the Witten Laplacian.

We are now in a position to state the main results of this paper.

Theorem 1.2. *Let $(M, g(t), t \in [0, T])$ be a family of compact Riemannian manifolds with potential functions $\phi(t) \in C^\infty(M)$, $t \in [0, T]$. Suppose that $g(t)$ and $\phi(t)$ satisfy the conjugate equation*

$$\frac{\partial \phi}{\partial t} = \frac{1}{2} \operatorname{Tr}\Big(\frac{\partial g}{\partial t}\Big).$$

Let

$$L = \Delta_{g(t)} - \nabla_{g(t)}\phi(t) \cdot \nabla_{g(t)}$$

be the time dependent Witten Laplacian on $(M, g(t), \phi(t))$. Let u be a positive solution of the heat equation

$$\partial_t u = Lu$$

with initial data $u(0)$ satisfying $\int_M u(0)\, d\mu(0) = 1$. Let

$$H_m(u, t) = -\int_M u \log u \, d\mu - \frac{m}{2}(1 + \log(4\pi t)).$$

Define

$$W_m(u, t) = \frac{d}{dt}(t H_m(u)).$$

Then

$$W_m(u, t) = \int_M \big(t|\nabla \log u|^2 - \log u - \frac{m}{2}(2 + \log(4\pi t))\big) u \, d\mu,$$

and

$$\begin{aligned} \frac{d}{dt} W_m(u, t) = {} & -2t \int_M \Big|\nabla^2 \log u + \frac{g}{2t}\Big|^2 u \, d\mu \\ & - \frac{2t}{m-n} \int_M \Big(\nabla\phi \cdot \nabla \log u - \frac{m-n}{2t}\Big)^2 u \, d\mu \\ & - 2 \int_M t \Big(\frac{1}{2}\frac{\partial g}{\partial t} + \mathrm{Ric}_{m,n}(L)\Big)(\nabla \log u, \nabla \log u) u \, d\mu. \end{aligned} \tag{4}$$

In particular, if $\{g(t), \phi(t), t \in (0, T]\}$ satisfies the m-dimensional Perelman super Ricci flow and the conjugate equation

$$\frac{1}{2}\frac{\partial g}{\partial t} + \mathrm{Ric}_{m,n}(L) \geq 0, \tag{5}$$

$$\frac{\partial \phi}{\partial t} = \frac{1}{2} \operatorname{Tr}\Big(\frac{\partial g}{\partial t}\Big), \tag{6}$$

then $W_m(u, t)$ is decreasing in $t \in (0, T]$, i.e.,

$$\frac{d}{dt} W_m(u, t) \leq 0, \quad \textit{for all } t \in (0, T].$$

As an application of the W-entropy formula for the Witten Laplacian on manifolds with time dependent metrics and potentials, in the following theorem we prove that the optimal logarithmic Sobolev constant associated with the Witten Laplacian on compact manifolds equipped with the m-dimensional Perelman super Ricci flow is decreasing in time.

Theorem 1.3. *Let $(M, g(t), \phi(t), t \in [0, T])$ be as in Theorem 1.2. Then, for any fixed $t \in [0, T]$, there exists a positive and smooth function $u = e^{-v/2}$ such that v achieves the optimal logarithmic Sobolev constant $\mu(t)$ defined by*

$$\mu(t) := \inf\left\{W_m(u, t) : \int_M \frac{e^{-v}}{(4\pi t)^{m/2}}\, d\mu = 1\right\}.$$

Indeed, $u = e^{-v/2}$ is a solution to the nonlinear PDE

$$-4tLu - 2u \log u - mu = \mu(t)u.$$

Moreover, if $\{g(t), \phi(t), t \in [0, T]\}$ satisfies the m-dimensional Perelman super Ricci flow (5) and the conjugate equation (6), then $\mu(t)$ is decreasing in t on $[0, T]$.

Remark 1.4. We believe that, via the approach used in [Li 2012; 2011; 2014], it would be possible to further extend the W-entropy formula in Theorem 1.2 to the fundamental solution of the heat equation associated with the Witten Laplacian on complete Riemannian manifolds with time dependent metrics and potentials satisfying the bounded geometry condition. Technically, this would require some Hamilton-type gradient estimates for the logarithm of the heat kernel of the Witten Laplacian on complete Riemannian manifolds with time dependent metrics and potentials satisfying the uniformly bounded geometry condition.[3] We will study this problem in a forthcoming paper. If this can be verified, we can derive, for a family $\{g(t), \phi(t), t \in (0, T]\}$ of metrics and potentials satisfying (5) and (6) on a complete Riemannian manifold M with the uniformly bounded geometry condition, that

$$\frac{d}{dt} W_m(u, t) = 0, \quad \text{for some } t = \tau \in (0, T]$$

if and only if at time $t = \tau$, we have

$$\nabla^2 \log u = -\frac{g}{2\tau}, \quad \frac{1}{2}\frac{\partial g}{\partial t} + \mathrm{Ric}_{m,n}(L) = 0, \quad \text{and} \quad \nabla\phi \cdot \nabla \log u = \frac{m-n}{2\tau}.$$

By the same argument as used in [Li 2012; 2011; 2014], we can further prove the following rigidity result. Let $\{g(t), \phi(t), t \in (0, T]\}$ be a family of metrics and potentials satisfying (5) and (6) on a complete Riemannian manifold M with the

[3] We say that $(M, g(t), \phi(t), t \in [0, T])$ satisfies the uniformly bounded geometry condition if there exists some $N \in \mathbb{N}$ such that for all $\varepsilon \in (0, T)$, the k-th order covariant derivatives $\nabla^k \mathrm{Riem}(g(t))$ of the Riemannian curvature tensor $\mathrm{Riem}(g(t))$, as well as the k-th order covariant derivatives $\nabla^k \phi(t)$ of $\phi(t)$, are uniformly bounded on $[\varepsilon, T] \times M$ for $k = 0, \ldots, N$.

uniformly bounded geometry condition. Let u be the fundamental solution to the heat equation $\partial_t u = Lu$. Then

$$\frac{d}{dt} W_m(u, t) = 0, \quad \text{for some } t = \tau \in (0, T],$$

if and only if $(M, g(\tau))$ is isometric to $\mathbb{R}^n$, $\phi(\tau)$ is identically equal to a constant, $m = n$, $\partial g/\partial t = 0$, $\partial\phi/\partial t = 0$ at $t = \tau$, and

$$u(x, \tau) = \frac{e^{-\|x\|^2/4\tau}}{(4\pi\tau)^{n/2}}, \quad \text{for all } x \in M = \mathbb{R}^n.$$

The rest of this paper is organized as follows. In Section 2, we give a new proof of Theorem 1.1.[4] In Section 3, we prove the dissipation formula of the Boltzmann–Shannon entropy for the heat equation of the Witten Laplacian on compact manifolds with time dependent metrics and potentials. In Section 4, we prove Theorem 1.2 and Theorem 1.3. In Section 5, we use Perelman's W-entropy formula for Ricci flow to derive the W-entropy formula for the backward heat equation of the Witten Laplacian on compact Riemannian manifolds equipped with a modified Ricci flow introduced by Lott [2009]. In Section 6, we extend Theorem 1.1 and Theorem 1.2 to the case $\mathrm{Ric}_{m,n}(L) \geq -K$ and compact K-super m-dimensional Bakry–Émery Ricci flow.

2. A new proof of Theorem 1.1

To prove Theorem 1.1, we first recall some elementary geometric formulas on warped product metrics.

Let $m \in \mathbb{N}$, $m \geq n$. Let $\widetilde{M} = M \times N$, where (N, g_N) is a compact Riemannian manifold with dimension $q = m - n$. Let $\phi \in C^2(M)$. We consider the warped product metric

$$\tilde{g} = g_M \oplus e^{-2\phi/q} g_N. \tag{7}$$

on $\widetilde{M}$. Let ν_N be the volume measure on N. Then the volume measure on $(\widetilde{M}, \tilde{g})$ is given by

$$d\mathrm{vol}_{\widetilde{M}} = e^{-\phi}\, d\mathrm{vol}_M \otimes d\nu_N$$

Define

$$d\mu = e^{-\phi}\, d\mathrm{vol}_M.$$

Then

$$d\mathrm{vol}_{\widetilde{M}} = d\mu \otimes d\nu_N.$$

[4] One of the advantages of our new proof is that it gives a natural geometric interpretation of the third term appearing in the W-entropy formula (3). See Remark 2.2.

Without loss of generality, we may assume that

$$\nu_N(N) = 1.$$

Let $\widetilde{\Gamma}$ be the Christoffel symbol on $(\widetilde{M}, \tilde{g})$. By direct calculation, we verify that

$$\begin{aligned} &\widetilde{\Gamma}_{ij}^k = \Gamma_{ij}^k, \qquad &&\widetilde{\Gamma}_{\alpha\beta}^k = q^{-1} g^{kl} \partial_l \phi g_{\alpha\beta}, \qquad &&\widetilde{\Gamma}_{\alpha\beta}^\gamma = \Gamma_{\alpha\beta}^\gamma, \\ &\widetilde{\Gamma}_{ij}^\alpha = 0, &&\widetilde{\Gamma}_{i\alpha}^k = 0, &&\widetilde{\Gamma}_{i\alpha}^\beta = 0. \end{aligned}$$

Let $\widetilde{\nabla}$ be the Levi-Civita connection on $(\widetilde{M}, \tilde{g})$. For any $f \in C^2(M)$, using the formulas

$$\widetilde{\nabla}_{ij}^2 f = \partial_i \partial_j f - \widetilde{\Gamma}_{ij}^k \partial_k f, \quad \widetilde{\nabla}_{i\alpha}^2 f = \partial_i \partial_\alpha f - \widetilde{\Gamma}_{i\alpha}^k \partial_k f, \quad \widetilde{\nabla}_{\alpha\beta}^2 f = \partial_\alpha \partial_\beta f - \widetilde{\Gamma}_{\alpha\beta}^k \partial_k f,$$

we have

$$\widetilde{\nabla}_{ij}^2 f = \nabla_{ij}^2 f, \tag{8}$$

$$\widetilde{\nabla}_{\alpha\beta}^2 f = -q^{-1} g^{kl} \partial_l \phi \partial_k f g_{\alpha\beta}, \tag{9}$$

$$\nabla_{i\alpha}^2 f = 0. \tag{10}$$

Hence

$$\begin{aligned} \left|\widetilde{\nabla}^2 f - \frac{\tilde{g}}{2t}\right|^2 &= \left|\nabla^2 f - \frac{g}{2t}\right|^2 + \left|\widetilde{\nabla}_{\alpha\beta}^2 f - \frac{g_{\alpha\beta}}{2t}\right|^2 \\ &= \left|\nabla^2 f - \frac{g}{2t}\right|^2 + \left|\frac{g^{kl} \partial_l \phi \partial_k f g_{\alpha\beta}}{q} + \frac{g_{\alpha\beta}}{2t}\right|^2 \\ &= \left|\nabla^2 f - \frac{g}{2t}\right|^2 + \left|\left(\frac{\nabla\phi \cdot \nabla f}{m-n} + \frac{1}{2t}\right) g_{\alpha\beta}\right|^2 \\ &= \left|\nabla^2 f - \frac{g}{2t}\right|^2 + \frac{1}{m-n}\left(\nabla\phi \cdot \nabla f + \frac{m-n}{2t}\right)^2. \end{aligned} \tag{11}$$

The following result was obtained in a private discussion between Bing-Long Chen and the second author in January 2006.

Theorem 2.1. *The Laplace–Beltrami operator on* $(\widetilde{M}, \tilde{g})$ *is given by*

$$\Delta_{\widetilde{M}} = L + e^{-2\phi/(m-n)} \Delta_N.$$

Proof. The proof can be given by a direct calculation. □

Proof of Theorem 1.1. To avoid technical issues, we only consider the case of compact manifolds. Let $u = e^{-f}/(4\pi t)^{m/2} : M \to [0, \infty)$ be a positive solution to the heat equation $\partial_t u = Lu$. Then it satisfies the heat equation

$$\partial_t u = \Delta_{\widetilde{M}} u$$

on $(\widetilde{M}, \tilde{g})$. Since f depends only on the variable in the M-direction, we have $\widetilde{\nabla} f = \nabla f$. Therefore, the W-entropy functional $W_m(u, t)$ defined by (2) coincides with the W-entropy functional

$$\widetilde{W}_m(u,t) = \int_{\widetilde{M}} \big[t|\widetilde{\nabla} f|^2 + f - m\big] \frac{e^{-f}}{(4\pi t)^{m/2}}\, d\mathrm{vol}_{\widetilde{M}} \tag{12}$$

defined on $(\widetilde{M}, \tilde{g})$. Applying to $(\widetilde{M}, \tilde{g})$ the W-entropy formula for the heat equation $\partial_t u = \Delta u$ on compact Riemannian manifolds with fixed metrics due to Ni [2004b; 2004a], we have

$$\frac{d}{dt}\widetilde{W}_m(u,t) = -2\int_{\widetilde{M}} t\Big(\Big|\widetilde{\nabla}^2 f - \frac{\tilde{g}}{2t}\Big|^2 + \widetilde{\mathrm{Ric}}\big(\widetilde{\nabla}\log u, \widetilde{\nabla}\log u\big)\Big) u\, d\mu dv_N. \tag{13}$$

By (11), we have

$$\Big|\widetilde{\nabla}^2 f - \frac{\tilde{g}}{2t}\Big|^2 = \Big|\nabla^2 f - \frac{g}{2t}\Big|^2 + \frac{2}{m-n}\Big(\nabla\phi\cdot\nabla f + \frac{m-n}{2t}\Big)^2. \tag{14}$$

On the other hand, by [Besse 1987; Lott 2003; Li 2005], we have

$$\widetilde{\mathrm{Ric}}\big(\widetilde{\nabla}\log u, \widetilde{\nabla}\log u\big) = \mathrm{Ric}_{m,n}(L)(\nabla\log u, \nabla\log u). \tag{15}$$

From (13), (14) and (15), we obtain (3). This finishes the new proof of Theorem 1.1 in the case of compact manifolds. $\square$

Remark 2.2. One of the advantages of the above proof is that when $m \in \mathbb{N}$ and $m > n$, the quantity

$$\frac{1}{m-n}\Big(\nabla\phi\cdot\nabla f + \frac{m-n}{2t}\Big)^2$$

appearing in the W-entropy formula in Theorem 1.1 has a natural geometric interpretation. It corresponds to the vertical component of the quantity $\big|\widetilde{\nabla}^2 f - \tilde{g}/2t\big|^2$ on the warped product manifold $\widetilde{M} = M \times N$ equipped with the metric

$$\tilde{g} = g \oplus e^{-2\phi/(m-n)} g_N.$$

In the case where (M, g) is a complete Riemannian manifold with the bounded geometry condition, similarly to [Lott 2003; Charalambous and Lu 2015], by introducing a sequence of warped product metrics $\{\tilde{g}_\varepsilon\}$ on $\widetilde{M} = M \times N$ defined by

$$\tilde{g}_\varepsilon = g \oplus \varepsilon^2 e^{-2\phi/(m-n)} g_N,$$

and using the fact that the heat kernel of the Laplace–Beltrami $\Delta_{(\widetilde{M}, \tilde{g}_\varepsilon)}$ on $(\widetilde{M}, \tilde{g}_\varepsilon)$ (with renormalized volume measure) converges in the $C^{2,\alpha} \cap W^{2,p}$-topology to the heat kernel of the Witten Laplacian $L = \Delta_M - \nabla\phi\cdot\nabla$ on (M, g, μ), we can use the same approach as in the compact case to give a new proof of the W-entropy formula for the heat kernel of the Witten Laplacian on complete Riemannian manifolds satisfying the bounded geometry condition in Theorem 1.1. We will study this problem in detail in the future.

Remark 2.3. We would like to mention that, after the first version of this paper [Li and Li 2014b] was posted online in March 2013, N. Charalambous and Z. Lu posted a preprint [2015] in which they used the warped product approach to prove the Li–Yau differential Harnack inequality on complete Riemannian manifolds with weighted volume measure. Recently we also found a paper by H. Guo, R. Philipowski and A. Thalmaier [2015], in which they studied the Boltzmann entropy dissipation formula on manifolds with time dependent metrics. We would also like to point out that G. Huang and H. Li [2014] extended the W-entropy formula for the heat equation of the Witten Laplacian in Theorem 1.1 to the porous medium equation for the Witten Laplacian on compact Riemannian manifolds with fixed metric and potential. We can use the same method developed in Section 2 to give a new proof of their result. See also related works of Y.-Z. Wang et al. [2013; 2014].

3. Dissipation formula of the Boltzmann–Shannon entropy

Let $(M, g(t), \phi(t))$ be as in Theorem 1.2. Following [Bakry and Émery 1985; Lott 2003; Li 2005], we introduce the Bakry–Émery Ricci curvature associated with L as

$$\mathrm{Ric}(L) = \mathrm{Ric} + \nabla^2\phi.$$

The purpose of this section is to prove the following dissipation formula for the Boltzmann–Shannon entropy associated with the Witten Laplacian on manifolds with time dependent metrics and potentials.

Theorem 3.1. *Let u be a positive solution to the heat equation $\partial_t u = Lu$. Let*

$$H(u,t) = -\int_M u\log u\, d\mu$$

be the Boltzmann–Shannon entropy associated with the Witten Laplacian L. Then

$$\frac{\partial^2}{\partial t^2}H(u,t) = -2\int_M \Big(|\nabla^2\log u|^2 + \Big(\frac{1}{2}\frac{\partial g}{\partial t} + \mathrm{Ric}(L)\Big)(\nabla\log u, \nabla\log u)\Big)u\, d\mu. \tag{16}$$

Proof. By direct calculation, we have

$$\frac{\partial}{\partial t}H(u,t) = -\int_M \partial_t u(\log u + 1)\, d\mu = -\int_M Lu(\log u + 1)\, d\mu.$$

Integrating by parts yields

$$\frac{\partial}{\partial t}H(u,t) = \int_M \big|\nabla\log u\big|^2_{g(t)} u\, d\mu,$$

which further implies that, as $\partial_t(d\mu)=0$, we have

$$
\begin{aligned}
\frac{\partial^2}{\partial t^2}H(u,t) &= \int_M \frac{\partial}{\partial t}\big(|\nabla \log u|^2_{g(t)}u\big)\,d\mu \\
&= \int_M \Big[\frac{\partial g^{ij}}{\partial t}\nabla_i \log u \nabla_j \log u\Big]u\,d\mu + \int_M \frac{\partial}{\partial t}\Big[\frac{|\nabla u|^2}{u}\Big]_{g(t)\text{ fixed}} d\mu \\
&= \int_M \Big[-\frac{\partial g_{ij}}{\partial t}\nabla_i \log u \nabla_j \log u\Big]u\,d\mu + \int_M \frac{\partial}{\partial t}\Big[\frac{|\nabla u|^2}{u}\Big]_{g(t)\text{ fixed}} d\mu \\
&= \int_M \Big(-\frac{\partial g}{\partial t}(\nabla \log u, \nabla u) + \frac{\partial}{\partial t}\Big[\frac{|\nabla u|^2}{u}\Big]_{g(t)\text{ fixed}}\Big)\,d\mu,
\end{aligned}
\tag{17}
$$

where $[\cdot]_{g(t)\text{ fixed}}$ means that the quantity $|\nabla u|^2$ in $[\cdot]$ is defined under a fixed metric $g(t)$. We have also used the facts $|\nabla \log u|^2 = g^{ij}\nabla_i \log u \nabla_j \log u$ and $\partial_t g^{ij} = -\partial_t g_{ij}$.

By the entropy dissipation formula in [Bakry and Émery 1985; Li 2014], we have

$$
\begin{aligned}
\int_M \frac{\partial}{\partial t}\Big[\frac{|\nabla u|^2}{u}\Big]_{g(t)\text{ fixed}} d\mu \\
= -2\int_M \big[|\nabla^2 \log u|^2 + \mathrm{Ric}(L)(\nabla \log u, \nabla \log u)\big]u\,d\mu.
\end{aligned}
\tag{18}
$$

Combining (17) and (18) completes the proof of Theorem 3.1. □

As an easy consequence of Theorem 1.2, we have the following corollary.

Corollary 3.2. *Let $(M, g(t))$ be a closed manifold with a potential $\phi(t)$. Suppose that $(g(t), \phi(t))$ satisfies the Perelman super Ricci flow and the conjugate equation*:

$$
\frac{\partial g}{\partial t} \geq -2\,\mathrm{Ric}(L), \qquad \frac{\partial \phi}{\partial t} = \frac{1}{2}\,\mathrm{Tr}\Big(\frac{\partial g}{\partial t}\Big).
$$

Let u be a positive solution to the heat equation $\partial_t u = Lu$. Then the Boltzmann–Shannon entropy

$$
H(u,t) = -\int_M u \log u\,d\mu
$$

is concave in time t, i.e.,

$$
\frac{d^2}{dt^2}H(u,t) \leq 0.
$$

4. Proofs of Theorem 1.2 and Theorem 1.3

Following [Li 2014], we introduce

$$
W(u,t) = \frac{d}{dt}(tH(u,t)).
$$

By direct calculation, we can prove the following.

Proposition 4.1. *We have*

$$W(u,t)=\int_M \big[t|\nabla \log u|^2-\log u\big]u\,d\mu,$$

and

$$\frac{d}{dt}W(u,t) = -2\int_M t\Big(|\nabla^2 \log u|^2+\Big(\frac{1}{2}\frac{\partial g}{\partial t}+\mathrm{Ric}(L)\Big)(\nabla \log u,\nabla \log u)\Big)u\,d\mu + 2\int_M |\nabla \log u|^2 u\,d\mu. \tag{19}$$

Remark 4.2. From (19), we can derive that if

$$\frac{1}{2}\frac{\partial g}{\partial t}+\mathrm{Ric}(L)-\frac{1}{t}\geq 0,$$

then

$$\frac{d}{dt}W(u,t)\leq 0.$$

Let

$$H_m(u,t)=-\int_M u\log u\,d\mu-\frac{m}{2}\big(1+\log(4\pi t)\big).$$

Following [Perelman 2002; Ni 2004a; 2004b; Li 2012; 2014], we define $W_m(u,t)$ by the Boltzmann entropy formula

$$W_m(u,t)=\frac{d}{dt}(tH_m(u)). \tag{20}$$

We can verify that $W_m(u,t)$ coincides with the expression given in Theorem 1.2, namely

$$W_m(u,t)=\int_M \big(t|\nabla \log u|^2-\log u-\frac{m}{2}(2+\log(4\pi t))\big)u\,d\mu.$$

Proof of Theorem 1.2. By (20) and (16) in Theorem 3.1, we have

$$\frac{d}{dt}W_m(u,t) = -2\int_M t\Big(|\nabla^2 \log u|^2+\Big(\frac{1}{2}\frac{\partial g}{\partial t}+\mathrm{Ric}(L)\Big)(\nabla \log u,\nabla \log u)\Big)u\,d\mu + 2\int_M |\nabla \log u|^2 u\,d\mu-\frac{m}{2t}. \tag{21}$$

Note that

$$2t|\nabla^2 \log u|^2+\frac{m}{2t}=2t\Big|\nabla^2\log u+\frac{g}{2t}\Big|^2+\frac{m-n}{2t}-2\Delta \log u.$$

Hence

$$\begin{aligned}\frac{d}{dt}W_m(u,t) = -\frac{m-n}{2t} - 2t\int_M \left|\nabla^2 \log u + \frac{g}{2t}\right|^2 u\,d\mu \\ + 2\int_M |\nabla \log u|^2 u\,d\mu + 2\int_M (\Delta \log u)u\,d\mu \\ - 2\int_M t\Big(\frac{1}{2}\frac{\partial g}{\partial t} + \mathrm{Ric}(L)\Big)(\nabla \log u, \nabla \log u)u\,d\mu.\end{aligned}$$

Integrating by parts yields

$$\begin{aligned}\int_M (\Delta \log u)u\,d\mu &= \int_M (L\log u + \nabla\phi\cdot\nabla\log u)u\,d\mu \\ &= -\int_M |\nabla \log u|^2 u\,d\mu + \int_M (\nabla\phi\cdot\nabla\log u)u\,d\mu,\end{aligned}$$

whence

$$\begin{aligned}\frac{d}{dt}W_m(u,t) = -\frac{m-n}{2t} - 2t\int_M \left|\nabla^2 \log u + \frac{g}{2t}\right|^2 u\,d\mu + 2\int_M (\nabla\phi\cdot\nabla\log u)u\,d\mu \\ - 2\int_M t\Big(\frac{1}{2}\frac{\partial g}{\partial t} + \mathrm{Ric}(L)\Big)(\nabla \log u, \nabla \log u)u\,d\mu.\end{aligned}$$

Note that

$$\begin{aligned}\frac{m-n}{2t} + 2t\,\mathrm{Ric}(L)(\nabla\log u, \nabla\log u) - 2\nabla\phi\cdot\nabla\log u \\ = 2t\,\mathrm{Ric}_{m,n}(L)(\nabla\log u, \nabla\log u) + \frac{2t}{m-n}\Big(\nabla\phi\cdot\nabla\log u - \frac{m-n}{2t}\Big)^2.\end{aligned}$$

We conclude that

$$\begin{aligned}\frac{d}{dt}W_m(u,t) = -2t\int_M \left|\nabla^2 \log u + \frac{g}{2t}\right|^2 u\,d\mu - \frac{2t}{m-n}\int_M \Big(\nabla\phi\cdot\nabla\log u - \frac{m-n}{2t}\Big)^2 u\,d\mu \\ - 2\int_M t\Big(\frac{1}{2}\frac{\partial g}{\partial t} + \mathrm{Ric}_{m,n}(L)\Big)(\nabla \log u, \nabla \log u)u\,d\mu.\end{aligned}$$

This proves the W-entropy formula in Theorem 1.2, and the monotonicity result follows. □

Proof of Theorem 1.3. The proof is similar to the one used by Perelman [2002]. See also [Li 2012]. By definition, we have

$$\mu(t) = \inf_u \Big\{\int_M \big[4t|\nabla u|^2 - u^2\log u^2 - mu^2\big](4\pi t)^{-m/2}\,d\mu\Big\}, \tag{22}$$

where the infimum is taken over all u such that

$$\int_M (4\pi t)^{-m/2} u^2\,d\mu = 1.$$

Indeed, $\mu(t)$ is the optimal constant in the logarithmic Sobolev inequality stating that for all u satisfying the above condition,

$$\int_M u^2 \log u^2 (4\pi t)^{-m/2}\, d\mu \le \mu(t) + m + 4 \int_M t|\nabla u|^2 (4\pi t)^{-m/2}\, d\mu.$$

By a similar argument as used in [Perelman 2002; Cao and Zhu 2006; Kleiner and Lott 2008; Morgan and Tian 2007], we can prove that the minimization problem (22) has a nonnegative minimizer $u \in H^1(M, \mu)$, which satisfies the Euler–Lagrange equation

$$-4tLu - 2u \log u - mu = \mu(t)u.$$

By the regularity theory of elliptic PDEs, we have $u \in C^{1,\alpha}(M)$. By an argument due to Rothaus [1981], we can further prove that u is strictly positive and smooth. Hence $v = -2\log u$ is also smooth. Moreover, as a consequence of Theorem 1.2, we can derive that $\mu(t)$ is a decreasing function in t on $[0, T]$, provided that $\{g(t), \phi(t), t \in [0, T]\}$ satisfies the m-dimensional Perelman super Ricci flow (5) and the conjugate equation (6). This completes the proof of Theorem 1.3. □

Remark 4.3. Let $m \in \mathbb{N}$ and $m > n$. Let (N, g_N) be a compact Riemannian manifold of dimension $q = m - n$. Let $\mathcal{M} = M \times N$ be the product manifold equipped with the time dependent warped product metric

$$\tilde{g}(t) = g(t) \oplus e^{-2\phi(t)/(m-n)} g_N.$$

Similarly to Remark 2.2, the quantity

$$\frac{1}{m-n}\Big(\nabla\phi(t)\cdot\nabla\log u - \frac{m-n}{2t}\Big)^2$$

appearing in the W-entropy formula in Theorem 1.2 has a natural geometric interpretation. It corresponds to the vertical component of the quantity $\big|\widetilde{\nabla}^2 \log u + \tilde{g}(t)/2t\big|^2$ on $(\mathcal{M}, \tilde{g}(t))$.

Remark 4.4. Perelman [2002] gave an interpretation of the W-entropy using the Boltzmann entropy formula from statistical mechanics. In [Li 2012; 2011], the second author gave a probabilistic interpretation of the W-entropy for the Ricci flow, the heat equation of the Witten Laplacian and the Fokker–Planck heat equation. Note that, as in [Li 2012; 2011; 2014], we have

$$H_m(u, t) = H(u, t) - H(\gamma, t)$$

where $H(u, t)$ is the Boltzmann–Shannon entropy associated with the heat equation to the Witten Laplacian on $(M, g(t), \phi(t))$, and $H(\gamma, t)$ is the Boltzmann–Shannon entropy of the Gaussian heat kernel $\gamma(x, t)$ on $\mathbb{R}^m$ for $m \in \mathbb{N}$ with $m \ge n$,

$$\gamma(x, t) = \frac{1}{(4\pi t)^{m/2}} e^{-\|x\|^2/4t}, \quad x \in \mathbb{R}^m, t > 0.$$

Thus, in view of its definition (20), the W-entropy $W_m(u,t)$ can be regarded as the byproduct of the Boltzmann–Shannon entropy. This gives a probabilistic interpretation of the W-entropy $W_m(u,t)$.

On the other hand, similarly to [Perelman 2002], we can also give a heuristic interpretation of the W-entropy using the Boltzmann entropy formula from statistical mechanics. Suppose that there exists a canonical ensemble with a "density of state measure" $g(E)\,dE$ such that the partition function $Z_\beta = \int_{\mathbb{R}^+} e^{-\beta E} g(E)\,dE$ is given by

$$\log Z_\beta = H_m(u,t), \tag{23}$$

where $t = \beta^{-1}$. (Here, as in [Perelman 2002], we do not discuss the issue of whether such a "density of state measure" exists or not.) Then, formally applying the Boltzmann entropy formula from statistical mechanics, the thermodynamical entropy of this canonical ensemble is given by

$$S = \log Z_\beta - \beta \frac{\partial}{\partial \beta} \log Z_\beta .$$

Using the fact $\dfrac{\partial}{\partial \beta} = \dfrac{\partial}{\partial t}\dfrac{\partial t}{\partial \beta} = -\dfrac{1}{\beta^2}\dfrac{\partial}{\partial t} = -t^2 \dfrac{\partial}{\partial t}$, we can prove

$$S = W_m(u,t).$$

Moreover, formally using the formula

$$\frac{dS}{d\beta} = -\beta \frac{\partial^2}{\partial \beta^2} \log Z_\beta ,$$

from statistical mechanics, we can reprove the W-entropy formula in Theorem 1.2.

5. The W-entropy for the Ricci flow on warped product manifolds

Let $m \in \mathbb{N}$ and $m \geq n$. Let $\mathbb{T}^q$ be the q-dimensional torus with a fixed flat metric given in local coordinates by $\sum_{i=1}^q dx_i^2$, where $q = m - n$. Let $\widetilde{M} = M \times \mathbb{T}^q$ be equipped with a time dependent warped product metric

$$\tilde{g}(t) = \sum_{i,j=1}^n g_{ij}(t)\,dx^i dx^j + u(t)^{2/q} \sum_{\alpha=1}^q dx_\alpha^2 .$$

Lott [2009] studied the Ricci flow $\tilde{g}(t)$ on the warped product manifold $\widetilde{M} = M \times \mathbb{T}^q$, which consists of a modified Ricci flow for the Riemannian metric $g(t)$ and a forward heat equation for a potential function $\psi(t) = -\log u(t)$ on the manifold M. In this section, we use Perelman's W-entropy formula for the Ricci flow $\tilde{g}(t)$ on the warped product manifold $\widetilde{M}$ to derive the W-entropy formula for the backward heat equation associated with the Witten Laplacian $L = \Delta_{g(t)} - \nabla_{g(t)}\psi(t) \cdot \nabla_{g(t)}$

on the compact manifold M equipped with Lott's modified Ricci flow $g(t)$ and the time dependent potential $\psi(t)$.

We first recall Lott's Ricci flow on $\widetilde{M} = M \times \mathbb{T}^q$. Let $u = e^{-\psi}$. Let $\widetilde{\mathrm{Ric}}$ be the Ricci curvature on $(\widetilde{M}, \tilde{g})$, and Ric the Ricci curvature on (M, g). By calculation on warped product manifolds [Besse 1987; Lott 2003; 2009], we have

$$\widetilde{\mathrm{Ric}} = \mathrm{Ric}_\psi^q + \frac{1}{q}(\Delta\psi - |\nabla\psi|^2)u^{2/q}\sum_{i=1}^{q} dx_i^2, \tag{24}$$

where Ric_ψ^q is the m-dimensional Bakry–Émery Ricci curvature on (M, g) with respect to the potential function ψ, i.e.,

$$\mathrm{Ric}_\psi^q = \mathrm{Ric} + \mathrm{Hess}\,\psi - \frac{1}{q}\nabla\psi\otimes\nabla\psi.$$

See [Bakry and Émery 1985; Li 2012; 2011; 2014]. Below we will also use the notation Ric_q to denote $\widetilde{\mathrm{Ric}}$. By (24), the scalar curvature R_q on $(\widetilde{M}, \tilde{g})$ is given by

$$R_q = R + 2\Delta\psi - \Big(1 + \frac{1}{q}\Big)|\nabla\psi|^2.$$

The Ricci flow on $\widetilde{M}$ is defined by

$$\partial_t \tilde{g} = -2\widetilde{\mathrm{Ric}}. \tag{25}$$

According to [Lott 2009], the Ricci flow equation (25) is equivalent to the equations

$$\partial_t g = -2\,\mathrm{Ric}_\psi^q, \tag{26}$$

$$\partial_t \psi = \Delta\psi - |\nabla\psi|^2. \tag{27}$$

Note that the first equation (26) is indeed a modified Ricci flow equation for the metric $g(t)$ on M, and the second one (27) is a forward heat equation for the potential function $\psi(t)$ on $(M, g(t))$. The systems (26) and (27) are different from Perelman's (modified) Ricci flow and the conjugate heat equation introduced in [Perelman 2002], i.e.,

$$\frac{\partial g}{\partial t} = -2(\mathrm{Ric} + \nabla^2 f),$$
$$\frac{\partial f}{\partial t} = -\Delta f - R,$$

and are also different from the m-dimensional Perelman Ricci flow and the conjugate heat equation

$$\frac{\partial g}{\partial t} = -2\Big(\mathrm{Ric} + \nabla^2 f - \frac{\nabla f\otimes\nabla f}{m-n}\Big),$$
$$\frac{\partial f}{\partial t} = -\Delta f + \frac{|\nabla f|^2}{m-n} - R.$$

Let ϕ be a positive solution to the conjugate heat equation on $(\widetilde{M}, \tilde{g})$,

$$\partial_t \phi = -\Delta_{\widetilde{M}} \phi + R_q \phi. \tag{28}$$

Let $\tau \in [0, T]$ be such that

$$\partial_t \tau = -1,$$

and η such that

$$\phi = (4\pi\tau)^{-(n+q)/2} e^{-\eta}.$$

We then have

$$\partial_t \eta = -\Delta_{\widetilde{M}} \eta + |\nabla\eta|^2 - R_q + \frac{n+q}{2\tau}.$$

Following [Perelman 2002], the W-entropy for the Ricci flow $\tilde{g}(t)$ on the warped product manifold $\widetilde{M}$ is defined by

$$W(\tilde{g}, \eta, \tau) = \int_{\widetilde{M}} \big[\tau\big(|\widetilde{\nabla}\eta|^2_{\widetilde{M}} + R_q\big) + \eta - (n+q)\big]\phi \, d\mathrm{vol}_{\widetilde{M}},$$

where $d\mathrm{vol}_{\widetilde{M}}$ is the volume form $u \, d\mathrm{vol}_M d\mathrm{vol}_{\mathbb{T}^q}$ on $(\widetilde{M}, \tilde{g})$.

Applying Perelman's [2002] W-entropy formula for the Ricci flow to $(\widetilde{M}, \tilde{g})$, we have

$$\frac{d}{d\tau} W(\tilde{g}, \eta, \tau) = -2\tau \int_{\widetilde{M}} \left| \widetilde{\mathrm{Ric}} + \widetilde{\mathrm{Hess}}\, \eta - \frac{\tilde{g}}{2\tau} \right|^2_{\widetilde{M}} \phi \, d\mathrm{vol}_{\widetilde{M}}. \tag{29}$$

By Theorem 2.1, the Laplace–Beltrami on $(\widetilde{M}, \tilde{g})$ is given by

$$\Delta_{\widetilde{M}} = L + u^{-2/q} \Delta_{\mathbb{T}^q},$$

where

$$L = \Delta - \nabla\psi \cdot \nabla.$$

Here Δ and ∇ are the Laplace–Beltrami operator and the gradient operator on (M, g), respectively. In the case that ϕ is a function depending only on the variable of the horizontal direction, the conjugate heat equation (28) turns out to be the backward heat equation associated with the Witten Laplacian on $(M, g(t))$,

$$\partial_t \phi = -L\phi + R_q \phi. \tag{30}$$

In this case, η is a function depending only on the variable in M. Thus,

$$\begin{aligned} W(\tilde{g}, \eta, \tau) &= \int_{M \times \mathbb{T}^q} \big[\tau(|\nabla\eta|^2 + R_q) + \eta - (n+q)\big]\phi u \, d\mathrm{vol}_M d\mathrm{vol}_{\mathbb{T}^q} \\ &= \int_M \Big[\tau\Big(|\nabla\eta|^2 + R + 2\Delta\psi - \Big(1 + \frac{1}{q}\Big)|\nabla\psi|^2\Big) + \eta - (n+q)\Big]\phi \, d\mu. \end{aligned}$$

Here $d\mu = u\, d\mathrm{vol}_M$, and we assume $\mathrm{vol}(\mathbb{T}^q) = 1$. Note that for any vector field v on $\widetilde{M}$, by (8), (9) and (10) we have

$$\widetilde{\mathrm{Hess}}\, v = \mathrm{Hess}\, v - \frac{1}{q} u^{2/q} \langle \nabla\psi, \nabla v\rangle \sum_{\alpha=1}^{q} dx_\alpha^2. \tag{31}$$

Substituting (24) and (31) into (29), we have

$$\begin{aligned}
\frac{d}{d\tau} W(\tilde{g}, \eta, \tau) &= -2\tau \int_M \left| \widetilde{\mathrm{Ric}} + \widetilde{\mathrm{Hess}}\, \eta - \frac{\tilde{g}}{2\tau}\right|_{\widetilde{M}}^2 \phi\, d\mu \\
&= -2\tau \int_M \bigg| \mathrm{Ric}_\psi^q + \mathrm{Hess}\, \eta - \frac{g}{2\tau} \\
&\qquad\qquad + \frac{1}{q}\Big(\Delta\psi - |\nabla\psi|^2 - \langle \nabla\psi, \nabla\eta\rangle - \frac{q}{2\tau}\Big) u^{2/q} \sum_{\alpha=1}^{q} dx_\alpha^2 \bigg|_{\widetilde{M}}^2 \phi\, d\mu \\
&= -2\tau \int_M \bigg(\Big| \mathrm{Ric}_\psi^q + \mathrm{Hess}\, \eta - \frac{g}{2\tau}\Big|^2 \\
&\qquad\qquad\qquad + \frac{1}{q}\Big(\Delta\psi - |\nabla\psi|^2 - \langle \nabla\psi, \nabla\eta\rangle - \frac{q}{2\tau}\Big)^2 \bigg) \phi\, d\mu.
\end{aligned}$$

Thus we have proved the following W-entropy formula for the backward heat equation associated with the Witten Laplacian on compact manifolds equipped with Lott's modified Ricci flow and time dependent potentials.

Theorem 5.1. *Let $(M, g(t), \psi(t))$ be a compact manifold with a family of Riemannian metrics $g(t)$ and potentials $\psi(t)$ which satisfy*

$$\begin{aligned}
\partial_t g &= -2\Big(\mathrm{Ric} + \mathrm{Hess}\, \psi - \frac{1}{q} \nabla\psi \otimes \nabla\psi\Big), \\
\partial_t \psi &= \Delta\psi - |\nabla\psi|^2.
\end{aligned}$$

Let $d\mu = e^{-\psi} d\mathrm{vol}_M$, and $L = \Delta - \nabla\psi \cdot \nabla$. Let ϕ be a positive solution to the backward heat equation of the Witten Laplacian on M, i.e.,

$$\partial_t \phi = -L\phi + R_q \phi,$$

where $R_q = R + 2\Delta\psi - (1 + 1/q)|\nabla\psi|^2$. Define the W-entropy $W_q(g, \psi, \eta, \tau)$ by

$$W_q(g, \psi, \eta, \tau) = \int_M \big[\tau\big(|\nabla\eta|^2 + R_q\big) + \eta - (n+q)\big]\phi\, d\mu,$$

where $\phi = (4\pi\tau)^{-(n+q)/2} e^{-\eta}$, and (η, τ) satisfies

$$\partial_t \eta = -L\eta + |\nabla\eta|^2 - R_q + \frac{n+q}{2\tau}, \quad \partial_t \tau = -1.$$

Then

$$\frac{d}{d\tau}W_q(g,\psi,\eta,\tau)$$
$$= -2\tau\int_M\Big(\Big|\mathrm{Ric}_\psi^q+\mathrm{Hess}\,\eta-\frac{g}{2\tau}\Big|^2+\frac{1}{q}\Big(\Delta\psi-|\nabla\psi|^2-\langle\nabla\psi,\nabla\eta\rangle-\frac{q}{2\tau}\Big)^2\Big)\phi\,d\mu.$$

In particular, $W_q(g,\psi,\eta,\tau)$ *is decreasing in the backward time* τ, *and the monotonicity is strict unless*

$$\mathrm{Ric}_\psi^q+\mathrm{Hess}\,\eta=\frac{g}{2\tau},$$
$$\Delta\psi-|\nabla\psi|^2=\langle\nabla\psi,\nabla\eta\rangle-\frac{q}{2\tau}.$$

As an application of the W-entropy formula in Theorem 1.3, we have:

Theorem 5.2. *Let* $(M,g(t),\psi(t))$ *be a compact manifold with a family of Riemannian metrics* $g(t)$ *and potentials* $\psi(t)$ *which satisfy*

$$\partial_t g=-2\Big(\mathrm{Ric}+\mathrm{Hess}\,\psi-\frac{1}{q}\nabla\psi\otimes\nabla\psi\Big),$$
$$\partial_t\psi=\Delta\psi-|\nabla\psi|^2.$$

Then there exists a positive and smooth function $u=e^{-\eta/2}$ *such that* η *achieves the optimal logarithmic Sobolev constant* $\mu(\tau)$ *defined by*

$$\mu(\tau):=\inf\Big\{W_q(g,\psi,\eta,\tau):\int_M\frac{e^{-\eta}}{(4\pi\tau)^{(n+q)/2}}\,d\mu=1\Big\},$$

where

$$W_q(g,\psi,\eta,\tau)=\int_M\big(\tau(|\nabla\eta|^2+R_q)+\eta-(n+q)\big)\phi\,d\mu,$$

Indeed, $u=e^{-\eta/2}$ *is a solution to the nonlinear PDE*

$$-4\tau Lu+\tau R_q u-2u\log u-(n+q)u=\mu(\tau)u.$$

Moreover, $\mu(\tau)$ *is decreasing in* τ *on* $[0,T]$.

Proof. The proof is similar to Perelman's [2002] monotonicity theorem for the μ-invariant for Ricci flow. See also [Cao and Zhu 2006; Chow et al. 2006; Kleiner and Lott 2008; Morgan and Tian 2007] and the proof of Theorem 1.3. □

6. The W-entropy formula for the Witten Laplacian with negative Bakry–Émery Ricci curvature

The W-entropy formula (3) only implies the monotonicity of the W-entropy for the Witten Laplacian on complete Riemannian manifolds with nonnegative m-dimensional Bakry–Émery Ricci curvature, and the W-entropy formula (4) only implies the monotonicity of the W-entropy for the Witten Laplacian on compact Riemannian manifolds with time dependent metrics and potentials satisfying the super m-dimensional Bakry–Émery Ricci flow and the conjugate heat equation. On the other hand, J. Li and X. Xu [2011] introduced a W-entropy for the heat equation $\partial_t u = \Delta u$ on complete Riemannian manifolds with Ricci curvature bounded from below by a negative constant. In this section, we combine the ideas in [Li and Xu 2011; Li 2012; 2014] and Section 4 to extend Theorem 1.1 to the Witten Laplacian on complete Riemannian manifolds with $\mathrm{Ric}_{m,n}(L)$ bounded from below by a negative constant, and extend Theorem 1.2 to the Witten Laplacian on compact Riemannian manifolds with time dependent metrics and potentials satisfying the K-super m-dimensional Bakry–Émery Ricci flow and the conjugate heat equation.

Recall the following entropy dissipation formulas on complete Riemannian manifolds.

Theorem 6.1 [Li 2012; 2014]. *Let (M, g) be a complete Riemannian manifold with the bounded geometry condition, and $\phi \in C^4(M)$ with $\nabla\phi \in C_b^3(M)$. Let u be the fundamental solution to the heat equation $\partial_t u = Lu$. Let*

$$H(u, t) = -\int_M u \log u \, d\mu.$$

Then

$$\frac{d}{dt} H(u, t) = \int_M \frac{|\nabla u|^2}{u} \, d\mu,$$

and

$$\frac{d^2}{dt^2} H(u, t) = -2 \int_M \big(|\nabla^2 \log u|^2 + \mathrm{Ric}(L)(\nabla \log u, \nabla \log u)\big) u \, d\mu.$$

Proposition 6.2. *Let $m \geq n$ and $K \geq 0$ be constants. Under the same assumptions as in Theorem 6.1, define*

$$H_{m,K}(u, t) = -\int_M u \log u \, d\mu - \frac{m}{2}\big(1 + \log(4\pi t)\big) - \frac{m}{2} K t \Big(1 + \frac{Kt}{6}\Big),$$

Then

$$\frac{d}{dt} H_{m,K}(u, t) = \int_M \Big(\frac{|\nabla u|^2}{u^2} - \frac{m}{2t} - \frac{mK}{2}\Big(1 + \frac{Kt}{3}\Big)\Big) u \, d\mu.$$

In particular, if $\mathrm{Ric}_{m,n}(L) \geq -K$, *then*

$$\frac{d}{dt} H_{m,K}(u, t) \leq 0,$$

Proof. By Theorem 6.1, we have

$$\frac{d}{dt} H_{m,K}(u, t) = \int_M \Big(\frac{|\nabla u|^2}{u^2} - \frac{m}{2t} - \frac{mK}{2}\Big(1 + \frac{Kt}{3}\Big)\Big) u \, d\mu.$$

By the same argument as in [Li and Xu 2011], or using the warped product approach as in [Charalambous and Lu 2015] and the Li–Yau-type differential Harnack inequality obtained by J. Li and X. Xu [2011], we can prove that, if $\mathrm{Ric}_{m,n}(L) \geq -K$, then

$$\frac{|\nabla u|^2}{u^2} - \Big(1 + \frac{2Kt}{3}\Big)\frac{\partial_t u}{u} \leq \frac{m}{2t} + \frac{mK}{2}\Big(1 + \frac{Kt}{3}\Big).$$

From this we may use the fact $\int_M \partial_t u \, d\mu = \int_M Lu \, d\mu = 0$ to conclude

$$\frac{d}{dt} H_{m,K}(u,t) \leq 0. \qquad \square$$

We now prove the main result of this section.

Theorem 6.3. *Let $m \geq n$ and $K \geq 0$ be constants. Under the same assumptions as in Theorem 6.1, define the W-entropy by the Boltzmann formula*

$$W_{m,K}(u,t) = \frac{d}{dt}(t H_{m,K}(u)).$$

Set $u = \dfrac{e^{-f}}{(4\pi t)^{m/2}}$. Then

$$W_{m,K}(u,t) = \int_M \Big(t|\nabla f|^2 + f - m\Big(1 + \frac{Kt}{2}\Big)^2\Big) u \, d\mu, \tag{32}$$

and

$$\begin{aligned}\frac{d}{dt} W_{m,K}(u,t) = &\\ -2t \int_M \Big(\Big|\nabla^2 f - \Big(\frac{1}{2t} + \frac{K}{2}\Big)g\Big|^2 + (\mathrm{Ric}_{m,n}(L) + Kg)(\nabla f, \nabla f)\Big) u \, d\mu &\\ - \frac{2t}{m-n} \int_M \Big(\nabla\phi \cdot \nabla f + (m-n)\Big(\frac{1}{2t} + \frac{K}{2}\Big)\Big)^2 u \, d\mu.&\end{aligned} \tag{33}$$

In particular, if $\mathrm{Ric}_{m,n}(L) \geq -K$, then

$$\frac{d}{dt} W_{m,K}(u,t) \leq 0.$$

Proof. We can prove (32) by direct calculation. By Theorem 6.1, we have

$$\begin{aligned}\frac{d}{dt} W_{m,K}(u,t) &= \frac{d}{dt} W(u,t) - \frac{m}{2t} - mK\Big(1 + \frac{Kt}{2}\Big)\\ &= -2\int_M t\Big(|\nabla^2 \log u|^2 + \mathrm{Ric}(L)(\nabla \log u, \nabla \log u)\Big) u \, d\mu\\ &\qquad + 2\int_M |\nabla \log u|^2 u \, d\mu - \frac{m}{2t} - mK\Big(1 + \frac{Kt}{2}\Big).\end{aligned}$$

Defining

$$\kappa(t) = K\Big(1 + \frac{Kt}{2}\Big), \quad \lambda(u,t) = \Big|\nabla^2 \log u + \frac{g}{2t} + \frac{Kg}{2}\Big|^2,$$

note that

$$2t|\nabla^2 \log u|^2 + \frac{m}{2t} + m\kappa(t) = 2t\lambda(u, t) - 2(1 + Kt)\Delta \log u + (m - n)\Big(\frac{1}{2t} + \kappa(t)\Big).$$

Integrating by parts yields

$$\begin{aligned}\frac{d}{dt} W_{m,K}(u, t) = -2 \int_M & t\big(\lambda(u, t) + (\mathrm{Ric}(L) + Kg)(\nabla \log u, \nabla \log u)\big)u \, d\mu \\ & + 2(1 + Kt) \int_M (\nabla \log u \cdot \nabla\phi)u \, d\mu + (m - n)\Big(\frac{1}{2t} + \kappa(t)\Big) \\ = -2 \int_M & t\big(\lambda(u, t) + (\mathrm{Ric}_{m,n}(L) + Kg)(\nabla \log u, \nabla \log u)\big)u \, d\mu \\ & - \frac{2t}{m-n} \int_M \Big(\nabla \log u \cdot \nabla\phi - (m - n)\Big(\frac{1}{2t} + \frac{K}{2}\Big)\Big)^2 u \, d\mu.\end{aligned}$$

In particular, if $\mathrm{Ric}_{m,n}(L) \geq -Kg$, $W_{m,K}(u, t)$ is monotone decreasing. □

Remark 6.4. Suppose that $\mathrm{Ric}_{m,n}(L) \geq -K$. By Theorem 6.3, $\frac{d}{dt} W_{m,K}(u, t) = 0$ if and only if

$$\mathrm{Ric}_{m,n}(L) = -Kg, \quad \nabla^2 f = \Big(\frac{1}{2t} + \frac{K}{2}\Big)g, \quad \nabla\phi \cdot \nabla f = -(m - n)\Big(\frac{1}{2t} + \frac{K}{2}\Big).$$

In particular, if $m = n$ and $\phi = C$ is a constant, then (M, g) is an Einstein manifold with $\mathrm{Ric} = -K$, and the potential f satisfies the shrinking gradient Ricci soliton equation (see [Li and Xu 2011])

$$\frac{1}{2}\,\mathrm{Ric} + \nabla^2 f = \frac{g}{2t}.$$

In general, (M, g) is a quasi-Einstein manifold with the metric g such that $\mathrm{Ric}_{m,n}(L) = -Kg$, and the potential f satisfies the shrinking gradient quasi-Ricci soliton equation

$$\frac{1}{2}\,\mathrm{Ric}_{m,n}(L) + \nabla^2 f = \frac{g}{2t}.$$

Remark 6.5. Similarly to Section 2, in the case that $m \in \mathbb{N}$, $m \geq n$ and M is a compact Riemannian manifold, we can give a new proof of Theorem 6.3 via the warped product method. Let $\widetilde{M} = M \times N$, where (N, g_N) is a compact Riemannian manifold with dimension $q = m - n$. Consider the following warped product metric on $\widetilde{M}$:

$$\tilde{g} = g_M \oplus e^{-2\phi/q} g_N.$$

Applying the W-entropy formula due to J. Li and X. Xu [2011] for the heat equation $\partial_t u = \Delta_{\widetilde{M}} u$ on $(\widetilde{M}, \tilde{g})$, we have

$$\begin{aligned}(34)\quad &\frac{d}{dt}\widetilde{W}_{m,K}(u,t)\\ &= -2\int_{\widetilde{M}} t\Big(\Big|\widetilde{\nabla}^2 f - \frac{\tilde{g}}{2t} - \frac{K\tilde{g}}{2}\Big|^2 + \big(\widetilde{\mathrm{Ric}} + K\tilde{g}\big)\big(\widetilde{\nabla}\log u, \widetilde{\nabla}\log u\big)\Big) u\, d\mu dv_N.\end{aligned}$$

From (11), we get

$$\begin{aligned}(35)\quad &\Big|\widetilde{\nabla}^2 f - \frac{\tilde{g}}{2t} - \frac{K\tilde{g}}{2}\Big|^2\\ &\qquad = \Big|\nabla^2 f - \frac{g}{2t} - \frac{Kg}{2}\Big|^2 + \frac{2}{m-n}\Big(\nabla\phi\cdot\nabla f + (m-n)\Big(\frac{1}{2t} + \frac{K}{2}\Big)\Big)^2.\end{aligned}$$

On the other hand, by [Besse 1987; Lott 2003; Li 2005], we have

$$(36)\qquad \big(\widetilde{\mathrm{Ric}} + K\tilde{g}\big)\big(\widetilde{\nabla}\log u, \widetilde{\nabla}\log u\big) = \big(\mathrm{Ric}_{m,n}(L) + Kg\big)\big(\nabla\log u, \nabla\log u\big).$$

From (34), (35) and (36), we reprove (33). Note that (35) also gives a natural geometric interpretation of the third term in the W-entropy formula (33).

We now extend Theorem 6.3 to the Witten Laplacian on compact manifolds with time dependent metrics and potentials.

Theorem 6.6. *Let $m \geq n$ and $K \geq 0$ be constants. Under the same assumptions as in Theorem 1.2, define*

$$H_{m,K}(u,t) = -\int_M u\log u\, d\mu - \frac{m}{2}(1+\log(4\pi t)) - \frac{m}{2}Kt\Big(1+\frac{Kt}{6}\Big)$$

and

$$W_{m,K}(u,t) = \frac{d}{dt}\big(tH_{m,K}(u)\big).$$

Set $u = e^{-f}/(4\pi t)^{m/2}$. Then

$$W_{m,K}(u,t) = \int_M \Big(t|\nabla f|^2 + f - m\Big(1+\frac{Kt}{2}\Big)^2\Big)\frac{e^{-f}}{(4\pi t)^{m/2}}\, d\mu,$$

and

$$\begin{aligned}\frac{d}{dt}W_{m,K}(u,t) = &-2t\int_M \Big|\nabla^2 f - \frac{g}{2t} - \frac{Kg}{2}\Big|^2\frac{e^{-f}}{(4\pi t)^{m/2}}\, d\mu\\ &- 2t\int_M \Big(\frac{1}{2}\frac{\partial g}{\partial t} + \mathrm{Ric}_{m,n}(L) + Kg\Big)(\nabla f, \nabla f)\frac{e^{-f}}{(4\pi t)^{m/2}}\, d\mu\\ &- \frac{2t}{m-n}\int_M \Big(\nabla\phi\cdot\nabla f + (m-n)\Big(\frac{1}{2t}+\frac{K}{2}\Big)\Big)^2\frac{e^{-f}}{(4\pi t)^{m/2}}\, d\mu.\end{aligned}$$

In particular, if $\{g(t), \phi(t), t \in (0, T]\}$ *is the* K*-super* m*-dimensional Bakry–Émery Ricci flow and satisfies the conjugate equation*

$$\frac{1}{2}\frac{\partial g}{\partial t} + \mathrm{Ric}_{m,n}(L) \geq -Kg, \tag{37}$$

$$\frac{\partial \phi}{\partial t} = \frac{1}{2}\mathrm{Tr}\Big(\frac{\partial g}{\partial t}\Big), \tag{38}$$

then $W_{m,K}(u, t)$ *is decreasing in* $t \in (0, T]$, *i.e.*,

$$\frac{d}{dt}W_{m,K}(u, t) \leq 0, \quad \textit{for all } t \in (0, T].$$

Proof. By (19), and replacing $\mathrm{Ric}(L)$ by $\frac{1}{2}\frac{\partial g}{\partial t} + \mathrm{Ric}(L)$, the proof is similar to the proof of Theorem 6.3. □

Finally, as an application of Theorem 6.6, we have the following.

Theorem 6.7. *Let* $(M, g(t), \phi(t), t \in [0, T])$ *be as in Theorem 6.6. Then there exists a positive and smooth function* $u = e^{-v/2}$ *such that* v *achieves the optimal logarithmic Sobolev constant* $\mu_K(t)$ *defined by*

$$\mu_K(t) := \inf\Big\{W_{m,K}(u, t) : \int_M \frac{e^{-v}}{(4\pi t)^{m/2}}\, d\mu = 1\Big\}.$$

Indeed, $u = e^{-v/2}$ *is a solution to the nonlinear PDE*

$$-4tLu - 2u\log u - m\Big(1 + \frac{Kt}{2}\Big)^2 u = \mu_K(t)u.$$

Moreover, if $\{g(t), \phi(t), t \in [0, T]\}$ *satisfies the* K*-super* m*-dimensional Bakry–Émery Ricci flow* (37) *and the conjugate equation* (38), *then* $\mu_K(t)$ *is decreasing in* t *on* $[0, T]$.

Proof. The proof is similar to the proof of Theorem 1.3. □

Note added in proof

In a recent preprint, the authors introduced the W-entropy and proved the W-entropy formula for the heat equation of the Witten Laplacian on complete Riemannian manifolds with the $\mathrm{CD}(K, \infty)$ condition (i.e., $\mathrm{Ric}(L) \geq K$) and extended the corresponding result to the heat equation of the time dependent Witten Laplacian on compact Riemannian manifolds equipped with the K-super Perelman Ricci flow with respect to the Bakry–Émery Ricci curvature (i.e., $\frac{1}{2}\frac{\partial g}{\partial t} + \mathrm{Ric}(L) \geq Kg$). See [Li and Li 2014a].

Acknowledgements

This paper is an extended version of the preprint [Li and Li 2014b]. We would like to thank Prof. D. Bakry, Prof. B.-L. Chen and Dr. Yu-Zhao Wang for their interest and useful discussions, and Prof. Zhiqin Lu for the citation of the first version of this paper in [Charalambous and Lu 2015]. We are very grateful to the anonymous referees for their careful reading and helpful comments for the revision of the paper.

References

[Andrews and Ni 2012] B. Andrews and L. Ni, "Eigenvalue comparison on Bakry–Emery manifolds", *Comm. Partial Differential Equations* **37**:11 (2012), 2081–2092. MR 3005536 Zbl 1258.35153

[Bakry and Émery 1985] D. Bakry and M. Émery, "Diffusions hypercontractives", pp. 177–206 in *Séminaire de probabilités, XIX, 1983/84*, edited by J. Azéma and M. Yor, Lecture Notes in Mathematics **1123**, Springer, Berlin, 1985. MR 88j:60131 Zbl 0561.60080

[Bakry and Ledoux 2006] D. Bakry and M. Ledoux, "A logarithmic Sobolev form of the Li–Yau parabolic inequality", *Rev. Mat. Iberoam.* **22**:2 (2006), 683–702. MR 2008m:58051 Zbl 1116.58024

[Bakry and Qian 1999] D. Bakry and Z. M. Qian, "Harnack inequalities on a manifold with positive or negative Ricci curvature", *Rev. Mat. Iberoamericana* **15**:1 (1999), 143–179. MR 2000f:58052 Zbl 0924.58096

[Bakry and Qian 2000] D. Bakry and Z. Qian, "Some new results on eigenvectors via dimension, diameter, and Ricci curvature", *Adv. Math.* **155**:1 (2000), 98–153. MR 2002g:58048 Zbl 0980.58020

[Besse 1987] A. L. Besse, *Einstein manifolds*, Ergebnisse der Mathematik (3) **10**, Springer, Berlin, 1987. MR 88f:53087 Zbl 0613.53001

[Cao and Zhu 2006] H.-D. Cao and X.-P. Zhu, "A complete proof of the Poincaré and geometrization conjectures—application of the Hamilton–Perelman theory of the Ricci flow", *Asian J. Math.* **10**:2 (2006), 165–492. MR 2008d:53090 Zbl 1200.53057

[Chang et al. 2011] S.-C. Chang, T.-J. Kuo, and S.-H. Lai, "Li–Yau gradient estimate and entropy formulae for the CR heat equation in a closed pseudohermitian 3-manifold", *J. Differential Geom.* **89**:2 (2011), 185–216. MR 2863916 Zbl 1264.32030

[Charalambous and Lu 2015] N. Charalambous and Z. Lu, "Heat kernel estimates and the essential spectrum on weighted manifolds", *J. Geom. Anal.* **25**:1 (2015), 536–563. MR 3299292 Zbl 06406917

[Chow et al. 2006] B. Chow, P. Lu, and L. Ni, *Hamilton's Ricci flow*, vol. 77, Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2006. MR 2008a:53068 Zbl 1118:53001

[Ecker 2007] K. Ecker, "A formula relating entropy monotonicity to Harnack inequalities", *Comm. Anal. Geom.* **15**:5 (2007), 1025–1061. MR 2009c:53088 Zbl 1167.53055

[Fang et al. 2009] F. Fang, X.-D. Li, and Z. Zhang, "Two generalizations of Cheeger–Gromoll splitting theorem via Bakry–Émery Ricci curvature", *Ann. Inst. Fourier* (*Grenoble*) **59**:2 (2009), 563–573. MR 2010b:53057 Zbl 1166.53023

[Futaki et al. 2013] A. Futaki, H. Li, and X.-D. Li, "On the first eigenvalue of the Witten–Laplacian and the diameter of compact shrinking solitons", *Ann. Global Anal. Geom.* **44**:2 (2013), 105–114. MR 3073582 Zbl 1273.58018

[Guo et al. 2015] H. Guo, R. Philipowski, and A. Thalmaier, "An entropy formula for the heat equation on manifolds with time-dependent metric, application to ancient solutions", *Potential Anal.* **42**:2 (2015), 483–497. MR 3306693 Zbl 1309.53051

[Huang and Li 2014] G. Huang and H. Li, "Gradient estimates and entropy formulae of porous medium and fast diffusion equations for the Witten Laplacian", *Pacific J. Math.* **268**:1 (2014), 47–78. MR 3207600 Zbl 1297.35059

[Kleiner and Lott 2008] B. Kleiner and J. Lott, "Notes on Perelman's papers", *Geom. Topol.* **12**:5 (2008), 2587–2855. MR 2010h:53098 Zbl 1204.53033

[Kotschwar and Ni 2009] B. Kotschwar and L. Ni, "Local gradient estimates of p-harmonic functions, $1/H$-flow, and an entropy formula", *Ann. Sci. Éc. Norm. Supér.* (4) **42**:1 (2009), 1–36. MR 2010g:53121 Zbl 1182.53060

[Li 2005] X.-D. Li, "Liouville theorems for symmetric diffusion operators on complete Riemannian manifolds", *J. Math. Pures Appl.* (9) **84**:10 (2005), 1295–1361. MR 2006f:58046 Zbl 1082.58036

[Li 2007] X.-D. Li, *On Perelman's W-entropy functional on Riemannian manifolds with weighted volume measure*, in Thèse d'Habilitation à Diriger des Rechercher, Université Paul Sabatier, 2007.

[Li 2011] X.-D. Li, "Perelman's W-entropy for the Fokker–Planck equation over complete Riemannian manifolds", *Bull. Sci. Math.* **135**:6-7 (2011), 871–882. MR 2012m:53077 Zbl 1230.82039

[Li 2012] X.-D. Li, "Perelman's entropy formula for the Witten Laplacian on Riemannian manifolds via Bakry–Emery Ricci curvature", *Math. Ann.* **353**:2 (2012), 403–437. MR 2915542 Zbl 1280.58016

[Li 2014] X.-D. Li, "Hamilton's Harnack inequality and the W-entropy formula on complete Riemannian manifolds", 2014. To appear in *Stochastic Process. Appl.* arXiv 1303.1242

[Li and Li 2014a] S. Li and X.-D. Li, "Harnack inequalities and W-entropy formula for Witten Laplacian on Riemannian manifolds with K-super Perelman Ricci flow", preprint, 2014. arXiv 1412.7034

[Li and Li 2014b] S. Li and X.-D. Li, "W-entropy formula for the Witten Laplacian on manifolds with time dependent metrics and potentials", preprint, 2014. arXiv 1303.6019

[Li and Xu 2011] J. Li and X. Xu, "Differential Harnack inequalities on Riemannian manifolds I: linear heat equation", *Adv. Math.* **226**:5 (2011), 4456–4491. MR 2012h:53079 Zbl 1226.58009

[Lott 2003] J. Lott, "Some geometric properties of the Bakry–Émery–Ricci tensor", *Comment. Math. Helv.* **78**:4 (2003), 865–883. MR 2004i:53044 Zbl 1038.53041

[Lott 2009] J. Lott, "Optimal transport and Perelman's reduced volume", *Calc. Var. Partial Differential Equations* **36**:1 (2009), 49–84. MR 2011b:53160 Zbl 1171.53318

[Lu et al. 2009] P. Lu, L. Ni, J.-L. Vázquez, and C. Villani, "Local Aronson–Bénilan estimates and entropy formulae for porous medium and fast diffusion equations on manifolds", *J. Math. Pures Appl.* (9) **91**:1 (2009), 1–19. MR 2010j:35563 Zbl 1156.58015

[Morgan and Tian 2007] J. Morgan and G. Tian, *Ricci flow and the Poincaré conjecture*, Clay Mathematics Monographs **3**, American Mathematical Society, Providence, RI, 2007. MR 2008d:57020 Zbl 1179.57045

[Ni 2004a] L. Ni, "Addenda to: "The entropy formula for linear heat equation"", *J. Geom. Anal.* **14**:2 (2004), 369–374. MR 2004m:53118b Zbl 1062.58028

[Ni 2004b] L. Ni, "The entropy formula for linear heat equation", *J. Geom. Anal.* **14**:1 (2004), 87–100. MR 2004m:53118a Zbl 1044.58030

[Otto 2001] F. Otto, "The geometry of dissipative evolution equations: the porous medium equation", *Comm. Partial Differential Equations* **26**:1-2 (2001), 101–174. MR 2002j:35180 Zbl 0984.35089

[Otto and Villani 2000] F. Otto and C. Villani, "Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality", *J. Funct. Anal.* **173**:2 (2000), 361–400. MR 2001k:58076 Zbl 0985.58019

[Perelman 2002] G. Perelman, "The entropy formula for the Ricci flow and its geometric applications", preprint, 2002. arXiv math/0211159

[von Renesse and Sturm 2005] M.-K. von Renesse and K.-T. Sturm, "Transport inequalities, gradient estimates, entropy, and Ricci curvature", *Comm. Pure Appl. Math.* **58**:7 (2005), 923–940. MR 2006j:53048 Zbl 1078.53028

[Rothaus 1981] O. S. Rothaus, "Logarithmic Sobolev inequalities and the spectrum of Schrödinger operators", *J. Funct. Anal.* **42**:1 (1981), 110–120. MR 83f:58080b Zbl 0471.58025

[Sturm 2005] K.-T. Sturm, "Convex functionals of probability measures and nonlinear diffusions on manifolds", *J. Math. Pures Appl.* (9) **84**:2 (2005), 149–168. MR 2005k:49126 Zbl 1259.49074

[Villani 2003] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics **58**, American Mathematical Society, Providence, RI, 2003. MR 2004e:90003 Zbl 1106.90001

[Villani 2009] C. Villani, *Optimal transport: Old and new*, Grundlehren der Mathematischen Wissenschaften **338**, Springer, Berlin, 2009. MR 2010f:49001 Zbl 1156.53003

[Wang and Chen 2014] Y. Wang and W. Chen, "Gradient estimates and entropy monotonicity formula for doubly nonlinear diffusion equations on Riemannian manifolds", *Math. Methods Appl. Sci.* **37**:17 (2014), 2772–2781. MR 3271122 Zbl 06446828

[Wang et al. 2013] Y. Wang, J. Yang, and W. Chen, "Gradient estimates and entropy formulae for weighted p-heat equations on smooth metric measure spaces", *Acta Math. Sci. Ser. B Engl. Ed.* **33**:4 (2013), 963–974. MR 3072133 Zbl 1299.58018

[Wei and Wylie 2009] G. Wei and W. Wylie, "Comparison geometry for the Bakry–Emery Ricci tensor", *J. Differential Geom.* **83**:2 (2009), 377–405. MR 2011a:53064 Zbl 1189.53036

Received November 2, 2014. Revised March 24, 2015.

SONGZI LI
SCHOOL OF MATHEMATICAL SCIENCES
FUDAN UNIVERSITY
220 HANDAN ROAD
SHANGHAI 200433
CHINA

and

INSTITUT DE MATHÉMATIQUES DE TOULOUSE
UNIVERSITÉ PAUL SABATIER
118 ROUTE DE NARBONNE
31062 TOULOUSE CEDEX 9
FRANCE

Songzi.Li@math.univ-toulouse.fr

XIANG-DONG LI
ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCE
CHINESE ACADEMY OF SCIENCES
55 ZHONGGUANCUN EAST ROAD
BEIJING 100190
CHINA

xdli@amt.ac.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.278.201

A DIAGRAMMATIC CATEGORIFICATION OF THE AFFINE q-SCHUR ALGEBRA $\widehat{S}(n,n)$ FOR $n \geq 3$

MARCO MACKAAY AND ANNE-LAURE THIEL

This is a follow-up to our 2013 paper "Categorifications of the extended affine Hecke algebra and the affine quantum Schur algebra $\widehat{S}(n,r)$ for $3 \leq r < n$" in which we categorified the affine q-Schur algebra $\widehat{S}(n,r)$ for $2 < r < n$ using a quotient of the categorification of $U_q(\widehat{\mathfrak{sl}}_n)$ of Khovanov and Lauda (2009, 2010, 2011). In this paper we categorify $\widehat{S}(n,n)$ for $n \geq 3$ using an extension of the aforementioned quotient.

1. Introduction 201
2. Affine quantum algebras 202
3. A diagrammatic categorification of $\widehat{S}(n,n)$ 210
4. Two useful 2-functors 221
5. The Grothendieck group 229
Acknowledgements 231
References 232

1. Introduction

The affine q-Schur algebra $\widehat{S}(n,r)$ was first defined and studied by Ginzburg and Vasserot [1993] and later also studied by Green [1999] and Lusztig [1999]. Let us assume that $n, r \geq 3$. Then $\widehat{S}(n,r)$ is a quotient of $U_q(\widehat{\mathfrak{sl}}_n)$ and $U_q(\widehat{\mathfrak{gl}}_n)$ if $r < n$. In [Mackaay and Thiel 2013] we defined a quotient of Khovanov and Lauda's categorification $\mathcal{U}(\widehat{\mathfrak{sl}}_n)$, denoted $\widehat{\mathcal{S}}(n,r)$, and showed that the Grothendieck group of its Karoubi envelope (idempotent completion) was exactly isomorphic to $\widehat{S}(n,r)$ for $2 < r < n$. In order to establish the isomorphism, we used Doty and Green's [2007] idempotented presentation of $\widehat{S}(n,r)$ for $2 < r < n$.

In this paper we address the case $n = r$, which is slightly more complicated because $\widehat{S}(n,n)$ is not a quotient of $U_q(\widehat{\mathfrak{sl}}_n)$ or $U_q(\widehat{\mathfrak{gl}}_n)$ but of the strictly larger algebra $\widehat{U}_q(\widehat{\mathfrak{gl}}_n)$ called the *extended* affine general linear quantum algebra [Green

Mackaay and Thiel were supported by the FCT - Fundação para a Ciência e a Tecnologia, through project number PTDC/MAT/101503/2008, New Geometry and Topology.
MSC2010: 17B37, 16T99.
Keywords: categorification, affine Schur algebras, quantum algebra.

1999]. Therefore, we have to extend the Khovanov–Lauda calculus of the corresponding quotient of $\mathcal{U}(\widehat{\mathfrak{sl}}_n)$ by adding certain generating 1- and 2-morphisms and relations. We denote that extended 2-category by $\widehat{\mathcal{S}}(n,n)$ and show that the Grothendieck group of its Karoubi envelope is isomorphic to $\widehat{\boldsymbol{S}}(n,n)$ for $n \geq 3$. For that isomorphism we use Deng, Du and Fu's presentation of $\widehat{\boldsymbol{S}}(n,n)$ [Deng et al. 2012], which extends Doty and Green's.

A little warning should be made. The results in this paper are not sufficient to categorify $\widehat{\boldsymbol{U}}_q(\widehat{\mathfrak{gl}}_n)$ diagrammatically, because that would require a categorification of $\widehat{\boldsymbol{S}}(n,r)$ for $2 < n < r$ too. However, no presentation of $\widehat{\boldsymbol{S}}(n,r)$ of Drinfeld–Jimbo type is known in that case, so even on the decategorified level there is an open question that would need to be solved first. For more information on this problem, see Question 4.3.2 in [Green 1999] and Chapter 5 in [Deng et al. 2012].

There is another technical detail that we should explain beforehand. In [Mackaay and Thiel 2013], we introduced a new degree-2 variable y and a y-deformation of the relations in Khovanov and Lauda's $\mathcal{U}(\widehat{\mathfrak{sl}}_n)$, denoted $\mathcal{U}(\widehat{\mathfrak{sl}}_n)_{[y]}$. The corresponding Schur quotients were denoted $\widehat{\mathcal{S}}(n,r)_{[y]}$. This y-deformation was introduced in order to establish a precise relation between $\widehat{\mathcal{S}}(n,r)_{[y]}$ and an extension of the affine singular Soergel bimodules built from Soergel's reflection faithful representation of the affine Weyl group, which were defined and studied by Williamson [2011]. However, we also proved that the ideals generated by y are virtually nilpotent, so that the Grothendieck groups of $\mathcal{U}(\widehat{\mathfrak{sl}}_n)_{[y]}$ and $\widehat{\mathcal{S}}(n,r)_{[y]}$ are isomorphic to those of $\mathcal{U}(\widehat{\mathfrak{sl}}_n)$ and $\widehat{\mathcal{S}}(n,r)$. Furthermore, for $y = 0$, the 2-representations in [Mackaay and Thiel 2013] give 2-functors from $\mathcal{U}(\widehat{\mathfrak{sl}}_n)$ to certain extensions of the affine Soergel bimodules built from the geometric representation of the affine Weyl group, which is not reflection faithful but still has some nice properties (for more information on this topic, see Section 3.1 in [Elias and Williamson 2013] and the results in [Libedinsky 2008]). In order to keep the calculations simple in this paper, we put $y = 0$ here. It would not be hard to give the y-deformed relations in the definition of $\widehat{\mathcal{S}}(n,n)$, which would give a 2-category $\widehat{\mathcal{S}}(n,n)_{[y]}$, but some of the subsequent calculations would be much harder in the y-deformed setting, e.g., the ones in the proof of Proposition 3.5.

In general, it would be interesting to know more about the relation between $\widehat{\mathcal{S}}(n,r)$, for $n \geq r$, and its y-deformation and the 2-category of affine singular Soergel bimodules.

Knowing more about this relation might also help to establish a connection with the work by Lusztig [1999] and Ginzburg and Vasserot [1993] on perverse sheaves and affine quantum $\mathfrak{gl}_n$.

2. Affine quantum algebras

In this section, we first recall the definition of the extended affine quantum general linear algebra $\widehat{\boldsymbol{U}}_q(\widehat{\mathfrak{gl}}_n)$ and its subalgebras $\boldsymbol{U}_q(\widehat{\mathfrak{gl}}_n)$ and $\boldsymbol{U}_q(\widehat{\mathfrak{sl}}_n)$. After that, we recall the definition of the affine quantum Schur algebras $\widehat{\boldsymbol{S}}(n,r)$, due to Green [1999]. Furthermore, we recall an idempotented presentation of the affine quantum Schur algebras, due to Doty and Green [2007] for $n>r$ and to Deng, Du and Fu [Deng et al. 2012] for $n=r$.

The (extended) affine quantum general and special linear algebras. For the rest of this paper, let $n \geq 3$.

Since in this paper we are only interested in the affine quantum general and special linear algebras at level 0, i.e., the q-analogue of the loop algebras without central extension, we can work with the normal $\mathfrak{gl}_n$-weight lattice, which is isomorphic to $\mathbb{Z}^n$. Let $\varepsilon_i = (0,\dots,1,\dots,0) \in \mathbb{Z}^n$, with 1 being on the i-th coordinate, and $\alpha_i = \varepsilon_i - \varepsilon_{i+1} \in \mathbb{Z}^n$ for $i=1,\dots,n$, where the subscripts have to be understood modulo n; e.g., $\alpha_n = \varepsilon_n - \varepsilon_1 = (-1,0,\dots,0,1)$. We also define the Euclidean inner product on $\mathbb{Z}^n$ by $\langle \varepsilon_i, \varepsilon_j \rangle = \delta_{i,j}$.

Definition 2.1 [Green 1999]. The *extended quantum general linear algebra* $\widehat{\boldsymbol{U}}_q(\widehat{\mathfrak{gl}}_n)$ is the associative unital $\mathbb{Q}(q)$-algebra generated by $R^{\pm 1}$, $K_i^{\pm 1}$ and $E_{\pm i}$ for $i = 1,\dots,n$, subject to the relations

$$K_i K_j = K_j K_i, \quad K_i K_i^{-1} = K_i^{-1} K_i = 1, \tag{2-1}$$

$$E_i E_{-j} - E_{-j} E_i = \delta_{i,j} \frac{K_i K_{i+1}^{-1} - K_i^{-1} K_{i+1}}{q - q^{-1}}, \tag{2-2}$$

$$K_i E_{\pm j} = q^{\pm \langle \varepsilon_i, \alpha_j \rangle} E_{\pm j} K_i, \tag{2-3}$$

$$E_{\pm i}^2 E_{\pm j} - (q+q^{-1}) E_{\pm i} E_{\pm j} E_{\pm i} + E_{\pm j} E_{\pm i}^2 = 0 \quad \text{if } |i-j| = 1 \bmod n, \tag{2-4}$$

$$E_{\pm i} E_{\pm j} - E_{\pm j} E_{\pm i} = 0 \quad \text{else}, \tag{2-5}$$

$$R R^{-1} = R^{-1} R = 1, \tag{2-6}$$

$$R X_i R^{-1} = X_{i+1} \quad \text{for } X_i \in \{E_{\pm i}, K_i^{-1}\}. \tag{2-7}$$

In all equations, the subscripts have to be read modulo n.

Definition 2.2. The *affine quantum general linear algebra* $\boldsymbol{U}_q(\widehat{\mathfrak{gl}}_n) \subseteq \widehat{\mathbf{U}}_q(\widehat{\mathfrak{gl}}_n)$ is the unital $\mathbb{Q}(q)$-subalgebra generated by $E_{\pm i}$ and $K_i^{\pm 1}$ for $i = 1,\dots,n$.

The *affine quantum special linear algebra* $\boldsymbol{U}_q(\widehat{\mathfrak{sl}}_n) \subseteq \mathbf{U}_q(\widehat{\mathfrak{gl}}_n)$ is the unital $\mathbb{Q}(q)$-subalgebra generated by $E_{\pm i}$ and $K_i K_{i+1}^{-1}$ for $i = 1,\dots,n$.

Remark 2.3. A little warning about the notation is needed here. Our notation follows that of [Doty and Green 2007; Green 1999], which differs from that of [Deng et al. 2012]. What we call $\boldsymbol{U}_q(\widehat{\mathfrak{gl}}_n)$, Deng, Du and Fu call $U_\Delta(n)$. In [Deng et al. 2012, Remark 5.3.2] they define $\widehat{\boldsymbol{U}}$, which is equal to our $\widehat{\boldsymbol{U}}_q(\widehat{\mathfrak{gl}}_n)$. Finally, their $\boldsymbol{U}(\widehat{\mathfrak{gl}}_n)$ is the *quantum loop algebra* of $\mathfrak{gl}_n$ (see their Definition 2.3.1), which contains $U_\Delta(n)$, i.e., our $\boldsymbol{U}_q(\widehat{\mathfrak{gl}}_n)$, as a proper subalgebra. In their notation, $\widehat{\boldsymbol{U}}$ is not a subalgebra of $U(\widehat{\mathfrak{gl}}_n)$ because $R \in \widehat{U}$ would have to be equal to an infinite linear combination of generators of the latter.

We will also need the bialgebra structure on $\widehat{U}_q(\widehat{\mathfrak{gl}}_n)$.

Definition 2.4 [Green 1999]. $\widehat{U}_q(\widehat{\mathfrak{gl}}_n)$ is a bialgebra with counit $\varepsilon\colon \widehat{U}_q(\widehat{\mathfrak{gl}}_n) \to \mathbb{Q}(q)$ defined by

$$\varepsilon(E_{\pm i}) = 0, \qquad \varepsilon(R^{\pm 1}) = \varepsilon(K_i^{\pm 1}) = 1,$$

and coproduct $\Delta\colon \widehat{U}_q(\widehat{\mathfrak{gl}}_n) \to \widehat{U}_q(\widehat{\mathfrak{gl}}_n) \otimes \widehat{\mathbf{U}}_q(\widehat{\mathfrak{gl}}_n)$ defined by

$$\Delta(1) = 1 \otimes 1, \tag{2-8}$$

$$\Delta(E_i) = E_i \otimes K_i K_{i+1}^{-1} + 1 \otimes E_i, \tag{2-9}$$

$$\Delta(E_{-i}) = K_i^{-1} K_{i+1} \otimes E_{-i} + E_{-i} \otimes 1, \tag{2-10}$$

$$\Delta(K_i^{\pm 1}) = K_i^{\pm 1} \otimes K_i^{\pm 1}, \tag{2-11}$$

$$\Delta(R^{\pm 1}) = R^{\pm 1} \otimes R^{\pm 1}. \tag{2-12}$$

As a matter of fact, $\widehat{U}_q(\widehat{\mathfrak{gl}}_n)$ is even a Hopf algebra, but we do not need the antipode in this paper. Note that Δ and ε can be restricted to $U_q(\widehat{\mathfrak{gl}}_n)$ and $U_q(\widehat{\mathfrak{sl}}_n)$, which are bialgebras too.

At level 0, we can also work with the $U_q(\mathfrak{sl}_n)$-weight lattice, which is isomorphic to $\mathbb{Z}^{n-1}$. Suppose that V is a $U_q(\widehat{\mathfrak{gl}}_n)$-weight representation with weights $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{Z}^n$; i.e.,

$$V \cong \bigoplus_{\lambda} V_\lambda$$

and K_i acts as multiplication by q^{λ_i} on V_λ. Then V is also a $U_q(\widehat{\mathfrak{sl}}_n)$-weight representation with weights $\bar{\lambda} = (\bar{\lambda}_1, \ldots, \bar{\lambda}_{n-1}) \in \mathbb{Z}^{n-1}$ such that $\bar{\lambda}_j = \lambda_j - \lambda_{j+1}$ for $j = 1, \ldots, n-1$. Conversely, given a $U_q(\widehat{\mathfrak{sl}}_n)$-weight representation with weights $\mu = (\mu_1, \ldots, \mu_{n-1})$, there is not a unique choice of $U_q(\widehat{\mathfrak{gl}}_n)$-action on V. We can fix this by choosing the action of $K_1 \cdots K_n$. In terms of weights, this corresponds to the observation that, for any $r \in \mathbb{Z}$, the equations

$$\lambda_i - \lambda_{i+1} = \mu_i, \tag{2-13}$$

$$\sum_{i=1}^{n} \lambda_i = r \tag{2-14}$$

determine $\lambda = (\lambda_1, \ldots, \lambda_n)$ uniquely, if there exists a solution to (2-13) and (2-14) at all. To fix notation, we define the map $\varphi_{n,r}\colon \mathbb{Z}^{n-1} \to \mathbb{Z}^n \cup \{*\}$ by

$$\varphi_{n,r}(\mu) = \lambda \tag{2-15}$$

if (2-13) and (2-14) have a solution, and put $\varphi_{n,r}(\mu) = *$ otherwise. This map already appeared in [Mackaay and Thiel 2013] and [Mackaay et al. 2013].

As far as weight representations are concerned, we can restrict our attention to Beilinson, Lusztig, and MacPherson's idempotented version of these quantum groups [Beilinson et al. 1990], denoted $\dot{\widehat{U}}(\widehat{\mathfrak{gl}}_n)$, $\dot{U}(\widehat{\mathfrak{gl}}_n)$ and $\dot{U}(\widehat{\mathfrak{sl}}_n)$ respectively. To understand their definitions, recall that K_i acts as q^{λ_i} on the λ-weight space of any weight representation. For each $\lambda \in \mathbb{Z}^n$, adjoin an idempotent 1_λ to $\widehat{U}_q(\widehat{\mathfrak{gl}}_n)$ and add the relations

$$\begin{aligned} 1_\lambda 1_\mu &= \delta_{\lambda,\mu} 1_\lambda, \\ E_{\pm i} 1_\lambda &= 1_{\lambda \pm \alpha_i} E_{\pm i}, \\ K_i 1_\lambda &= q^{\lambda_i} 1_\lambda, \\ R 1_{(\lambda_1,\dots,\lambda_n)} &= 1_{(\lambda_n,\lambda_1,\dots,\lambda_{n-1})} R. \end{aligned}$$

Definition 2.5. The *idempotented extended affine quantum general linear algebra* is defined by

$$\dot{\widehat{U}}(\widehat{\mathfrak{gl}}_n) = \bigoplus_{\lambda,\mu \in \mathbb{Z}^n} 1_\lambda \widehat{U}_q(\widehat{\mathfrak{gl}}_n) 1_\mu.$$

Of course one defines $\dot{U}(\widehat{\mathfrak{gl}}_n) \subset \dot{\widehat{U}}(\widehat{\mathfrak{gl}}_n)$ as the idempotented subalgebra generated by 1_λ and $E_{\pm i} 1_\lambda$ for $i = 1, \dots, n$ and $\lambda \in \mathbb{Z}^n$. Similarly for $\widehat{U}_q(\mathfrak{sl}_n)$, adjoin an idempotent 1_λ for each $\lambda \in \mathbb{Z}^{n-1}$ and add the relations

$$\begin{aligned} 1_\lambda 1_\mu &= \delta_{\lambda,\mu} 1_\lambda, \\ E_{\pm i} 1_\lambda &= 1_{\lambda \pm \alpha_i} E_{\pm i}, \\ K_i K_{i+1}^{-1} 1_\lambda &= q^{\lambda_i} 1_\lambda. \end{aligned}$$

Definition 2.6. The *idempotented quantum special linear algebra* is defined by

$$\dot{U}(\widehat{\mathfrak{sl}}_n) = \bigoplus_{\lambda,\mu \in \mathbb{Z}^{n-1}} 1_\lambda U_q(\widehat{\mathfrak{sl}}_n) 1_\mu.$$

Just to fix notation for future use.

Notation 2.7. For $\underline{\boldsymbol{i}} = (\mu_1 i_1, \dots, \mu_m i_m)$, with $\mu_j = \pm$, define

$$E_{\underline{\boldsymbol{i}}} := E_{\mu_1 i_1} \cdots E_{\mu_m i_m},$$

and define $\underline{\boldsymbol{i}}_\Lambda \in \mathbb{Z}^n$ to be the n-tuple such that

$$E_{\underline{\boldsymbol{i}}} 1_\lambda = 1_{\lambda + \underline{\boldsymbol{i}}_\Lambda} E_{\underline{\boldsymbol{i}}}.$$

Following Khovanov and Lauda [2009; 2010; 2011], we call $\underline{\boldsymbol{i}}$ a *signed sequence* and denote the set of signed sequences by SSeq.

The affine q-Schur algebra. As we did in [Mackaay and Thiel 2013], we first copy some facts about the action of $\widehat{U}_q(\widehat{\mathfrak{gl}}_n)$ on tensor space from [Doty and Green 2007; Green 1999]. After that we define the quotient $\widehat{S}(n,r)$, for $n \geq r$, and give a presentation of that algebra. Note that the case $n = r$ was not considered in [Mackaay and Thiel 2013].

Tensor space. Let V be the $\mathbb{Q}(q)$-vector space freely generated by $\{e_t \mid t \in \mathbb{Z}\}$.

Definition 2.8 [Green 1999]. The following defines an action of $\widehat{\boldsymbol{U}}_q(\widehat{\mathfrak{gl}}_n)$ on V:

$$E_i e_{t+1} = e_t \qquad \text{if } i \equiv t \mod n, \tag{2-16}$$

$$E_i e_{t+1} = 0 \qquad \text{if } i \not\equiv t \mod n, \tag{2-17}$$

$$E_{-i} e_t = e_{t+1} \qquad \text{if } i \equiv t \mod n, \tag{2-18}$$

$$E_{-i} e_t = 0 \qquad \text{if } i \not\equiv t \mod n, \tag{2-19}$$

$$K_i^{\pm 1} e_t = q^{\pm 1} e_t \qquad \text{if } i \equiv t \mod n, \tag{2-20}$$

$$K_i^{\pm 1} e_t = e_t \qquad \text{if } i \not\equiv t \mod n, \tag{2-21}$$

$$R^{\pm 1} e_t = e_{t \pm 1} \qquad \text{for all } t \in \mathbb{Z}. \tag{2-22}$$

Note that V is clearly a weight-representation of $\widehat{\boldsymbol{U}}_q(\widehat{\mathfrak{gl}}_n)$, with e_t having weight equal to ε_i for $i \equiv t \mod n$. Therefore V is also a representation of $\dot{\widehat{\boldsymbol{U}}}(\widehat{\mathfrak{gl}}_n)$. Let $r \in \mathbb{N}_{>0}$ be arbitrary but fixed. As usual, one extends the above action to $V^{\otimes r}$ using the coproduct in $\widehat{\boldsymbol{U}}_q(\widehat{\mathfrak{gl}}_n)$. Again, this is a weight-representation, and therefore also a representation of $\dot{\widehat{\boldsymbol{U}}}(\widehat{\mathfrak{gl}}_n)$. There is also a right action of the extended affine Hecke algebra $\widehat{\mathcal{H}}_{\widehat{A}_{r-1}}$ on $V^{\otimes r}$, whose precise definition is not relevant here, which commutes with the left action of $\widehat{\mathbf{U}}_q(\widehat{\mathfrak{gl}}_n)$.

Definition 2.9 [Green 1999]. The *affine q-Schur algebra* $\widehat{\boldsymbol{S}}(n,r)$ is by definition the centralizing algebra

$$\mathrm{End}_{\widehat{\mathcal{H}}_{\widehat{A}_{r-1}}}(V^{\otimes r}).$$

It turns out that the image of the representation $\psi_{n,r} : \widehat{\boldsymbol{U}}_q(\widehat{\mathfrak{gl}}_n) \to \mathrm{End}(V^{\otimes r})$ is isomorphic to $\widehat{\boldsymbol{S}}(n,r)$. If $n > r$, then we can even restrict to $\boldsymbol{U}_q(\widehat{\mathfrak{sl}}_n) \subset \widehat{\boldsymbol{U}}_q(\widehat{\mathfrak{gl}}_n)$, i.e.,

$$\psi_{n,r}(\boldsymbol{U}_q(\widehat{\mathfrak{sl}}_n)) \cong \widehat{\boldsymbol{S}}(n,r).$$

If $n = r$, this is no longer true, as we will show below.

Presentation of $\widehat{\boldsymbol{S}}(n,r)$ for $n > r$. In this subsection, let $n > r$. As already mentioned, the map

$$\psi_{n,r} : \dot{\boldsymbol{U}}(\widehat{\mathfrak{gl}}_n) \to \mathrm{End}(V^{\otimes r}) \to \widehat{\boldsymbol{S}}(n,r)$$

is surjective. This observation gives rise to the following presentation of $\widehat{\boldsymbol{S}}(n,r)$. The proof can be found in [Doty and Green 2007, Theorem 2.6.1].

Theorem 2.10 [Doty and Green 2007]. *For $n > r$, the $\mathbb{Q}(q)$-algebra $\widehat{\boldsymbol{S}}(n,r)$ is isomorphic to the associative unital $\mathbb{Q}(q)$-algebra generated by 1_λ and $E_{\pm i}$ for $\lambda \in \Lambda(n,r)$ and $i = 1, \dots, n$, subject to the relations*

$$1_\lambda 1_\mu = \delta_{\lambda,\mu} 1_\lambda, \tag{2-23}$$

$$E_{\pm i} 1_\lambda = 1_{\lambda \pm \alpha_i} E_{\pm i}, \tag{2-24}$$

$$(E_i E_{-j} - E_{-j} E_i) 1_\lambda = \delta_{i,j} [\lambda_i - \lambda_{i+1}] 1_\lambda, \tag{2-25}$$

$$\big(E_{\pm i}^2 E_{\pm j} - (q+q^{-1}) E_{\pm i} E_{\pm j} E_{\pm i} + E_{\pm j} E_{\pm i}^2\big) 1_\lambda = 0 \quad \textit{if } |i-j| = 1 \bmod n, \tag{2-26}$$

$$(E_{\pm i} E_{\pm j} - E_{\pm j} E_{\pm i}) 1_\lambda = 0 \quad \textit{else}. \tag{2-27}$$

In all equations the subscripts i, j have to be read modulo n, and the equations hold for any $\lambda \in \Lambda(n, r)$. If $\lambda \pm \alpha_i \notin \Lambda(n, r)$, the corresponding idempotent is 0 *by convention.*

We can restrict $\psi_{n,r}$ even further and obtain a surjection $\psi_{n,r} : \dot{\boldsymbol{U}}(\widehat{\mathfrak{sl}}_n) \to \widehat{\boldsymbol{S}}(n, r)$, which can be given explicitly on the generators. For any $\lambda \in \mathbb{Z}^{n-1}$, we have

$$\psi_{n,r}(E_{\pm i} 1_\lambda) = E_{\pm i} 1_{\varphi_{n,r}(\lambda)},$$

where $\varphi_{n,r} : \mathbb{Z}^{n-1} \to \Lambda(n, r) \cup \{*\}$ is the map defined in (2-15). By convention, we put $1_* = 0$.

Presentation of $\widehat{\boldsymbol{S}}(n, n)$. A presentation of $\widehat{\boldsymbol{S}}(n, n)$ of Drinfeld–Jimbo type is harder to get, because

$$\psi_{n,n}(\boldsymbol{U}_q(\widehat{\mathfrak{sl}}_n)) = \psi_{n,n}(\boldsymbol{U}_q(\widehat{\mathfrak{gl}}_n))$$

is a proper subalgebra of $\widehat{\boldsymbol{S}}(n, n)$. Therefore Green [1999] introduced $\widehat{\boldsymbol{U}}_q(\widehat{\mathfrak{gl}}_n)$, which contains the new invertible element R, and proved that $\widehat{\boldsymbol{S}}(n, n)$ is a quotient of this extended algebra. As vector spaces, we get the $\mathbb{Q}(q)$-linear isomorphism

$$\widehat{\boldsymbol{S}}(n, n) \cong \psi_{n,n}(\boldsymbol{U}_q(\widehat{\mathfrak{sl}}_n)) \oplus \bigoplus_{t \neq 0} \mathbb{Q}[R^t, R^{-t}].$$

However, this is not an algebra isomorphism. In [Deng et al. 2012, Theorem 5.3.5], the authors show which relations need to be added in order to get a presentation of the algebra $\widehat{\boldsymbol{S}}(n, n)$. Let us first recall a slightly different presentation obtained by adding two new elements, $E_{\pm\delta}$, instead of $R^{\pm 1}$. This presentation, also due to Deng et al. [2012], turns out to be easier to categorify. As in [Mackaay and Thiel 2013], we write $1_n := 1_{(1^n)}$. Recall that the *divided powers* are defined by

$$E_{\pm i}^{(a)} := \frac{E_{\pm i}^a}{[a]!} \quad \text{for } i = 1, \dots, n.$$

Theorem 2.11 [Deng et al. 2012]. *The $\mathbb{Q}(q)$-algebra $\widehat{\boldsymbol{S}}(n, n)$ is generated by $E_{\pm\delta}$, $E_{\pm i}$ and 1_λ, for $i = 1, \dots, n$ and $\lambda \in \Lambda(n, n)$, subject to the relations* (2-23) *through* (2-27) *together with*

$$
\begin{aligned}
&\text{(i)} && E_{\pm\delta}1_{\lambda} = 1_{\lambda}E_{\pm\delta} = 0 \quad \textit{for all } \lambda \neq (1^n),\\
&\text{(ii)} && E_{\pm\delta}1_n = 1_n E_{\pm\delta},\\
&\text{(iii)} && E_{+\delta}E_{-\delta}1_n = E_{-\delta}E_{+\delta}1_n = 1_n,\\
&\text{(iv)} && E_i E_{+\delta}1_n = E_i^{(2)}E_{i-1}\cdots E_1 E_n \cdots E_{i+1}1_n,\\
&\text{(v)} && 1_n E_{+\delta}E_i = 1_n E_{i-1}\cdots E_1 E_n \cdots E_{i+1}E_i^{(2)},\\
&\text{(vi)} && E_{-i}E_{+\delta}1_n = E_{i-1}\cdots E_1 E_n \cdots E_{i+1}1_n,\\
&\text{(vii)} && 1_n E_{+\delta}E_{-i} = 1_n E_{i-1}\cdots E_1 E_n \cdots E_{i+1},\\
&\text{(viii)} && E_{-i}E_{-\delta}1_n = E_{-i}^{(2)}E_{-(i+1)}\cdots E_{-n}E_{-1}\cdots E_{-(i-1)}1_n,\\
&\text{(ix)} && 1_n E_{-\delta}E_{-i} = 1_n E_{-(i+1)}\cdots E_{-n}E_{-1}\cdots E_{-(i-1)}E_{-i}^{(2)},\\
&\text{(x)} && E_i E_{-\delta}1_n = E_{-(i+1)}\cdots E_{-n}E_{-1}\cdots E_{-(i-1)}1_n,\\
&\text{(xi)} && 1_n E_{-\delta}E_i = 1_n E_{-(i+1)}\cdots E_{-n}E_{-1}\cdots E_{-(i-1)}
\end{aligned}
$$

for any $i = 1, \ldots, n$.

To see that Theorem 2.11 really gives a presentation of $\widehat{S}(n,n)$, recall the following definition given in [Deng et al. 2012, (5.3.1.1) and (5.3.1.2)]. (They use the notation ρ where we use R):

Definition 2.12. Define

$$R^{-1} := E_{+\delta}1_n + \sum_{i=1}^{n} \sum_{\substack{(a_1,\ldots,a_n)\in\Lambda(n,n)\\ a_i=0}} E_{i-1}^{(a_{i-1})}\cdots E_1^{(a_1)}E_n^{(a_n)}\cdots E_{i+1}^{(a_{i+1})}1_{(a_n,a_1,\ldots,a_{n-1})}$$

and

$$R := E_{-\delta}1_n + \sum_{i=1}^{n} \sum_{\substack{(a_1,\ldots,a_n)\in\Lambda(n,n)\\ a_i=0}} E_{-(i-1)}^{(a_{i-1})}\cdots E_{-1}^{(a_1)}E_{-n}^{(a_n)}\cdots E_{-(i+1)}^{(a_{i+1})}1_{(a_1,\ldots,a_n)}.$$

Then note that

$$E_{i-1}^{(a_{i-1})}\cdots E_1^{(a_1)}E_n^{(a_n)}\cdots E_{i+1}^{(a_{i+1})}1_{(a_n,a_1,\ldots,a_{n-1})} = 1_{(a_1,\ldots,a_n)}E_{i-1}^{(a_{i-1})}\cdots E_1^{(a_1)}E_n^{(a_n)}\cdots E_{i+1}^{(a_{i+1})}$$

and

$$E_{i-1}^{(a_{i-1})}\cdots E_1^{(a_1)}E_n^{(a_n)}\cdots E_{i+1}^{(a_{i+1})}1_{\lambda} = 0$$

for all $\lambda \neq (a_n, a_1, \ldots, a_{n-1})$. Likewise, we have

$$E_{-(i-1)}^{(a_{i-1})}\cdots E_{-1}^{(a_1)}E_{-n}^{(a_n)}\cdots E_{-(i+1)}^{(a_{i+1})}1_{(a_1,\ldots,a_n)} = 1_{(a_n,a_1,\ldots,a_{n-1})}E_{-(i-1)}^{(a_{i-1})}\cdots E_{-1}^{(a_1)}E_{-n}^{(a_n)}\cdots E_{-(i+1)}^{(a_{i+1})}$$

and

$$E_{-(i-1)}^{(a_{i-1})}\cdots E_{-1}^{(a_1)}E_{-n}^{(a_n)}\cdots E_{-(i+1)}^{(a_{i+1})}1_{\lambda} = 0$$

for all $\lambda \neq (a_1, \dots, a_n)$. These remarks show that Proposition 5.3.3 and Corollary 5.3.4 in [Deng et al. 2012] imply that the presentation of $\widehat{S}(n,n)$ in Theorem 5.3.5 in that paper is equivalent to the one we have given in Theorem 2.11. In particular, the relations in Theorem 2.11 imply the following relations, which are exactly the ones in [Deng et al. 2012, Theorem 5.3.5]:

Corollary 2.13. *In $\widehat{S}(n,n)$, we have*

$$RR^{-1} = R^{-1}R = 1, \quad RE_{\pm i}R^{-1} = E_{\pm(i+1)}, \quad R1_\lambda R^{-1} = 1_{(\lambda_n,\lambda_1\dots,\lambda_{n-1})}.$$

As usual, we read the indices modulo n.

Therefore, the surjective algebra homomorphism

$$\psi_{n,n}\colon \dot{\widehat{U}}(\widehat{\mathfrak{gl}}_n) \to \widehat{S}(n,n)$$

can be defined as

$$\psi_{n,n}(1_\lambda) = \begin{cases} 1_\lambda & \text{if } \lambda \in \Lambda(n,n), \\ 0 & \text{else,} \end{cases}$$

and

$$\psi_{n,n}(E_{\pm i}1_\lambda) = E_{\pm i}\psi_{n,n}(1_\lambda), \quad \psi_{n,n}(R^{\pm 1}1_\lambda) = R^{\pm 1}\psi_{n,n}(1_\lambda).$$

In Lemma 3.2 and Corollary 5.6 in [Deng and Du 2013], the authors also show that there exists an embedding

$$\iota_n\colon \widehat{S}(n,n) \to \widehat{S}(n+1,n),$$

which gives an isomorphism of algebras

$$\widehat{S}(n,n) \cong \bigoplus_{\lambda,\mu\in\Lambda(n,n)} 1_{(\lambda,0)}\widehat{S}(n+1,n)1_{(\mu,0)}.$$

At that point of their paper they use a different presentation of the affine q-Schur algebras, but by [Deng and Du 2013, Proposition 7.1] it is not hard to work out the image under ι_n of the generators of $\widehat{S}(n,n)$ in Theorem 2.11. Note that we have multiplied their images of E_{+n} and E_{-n} by -1, which is more convenient for categorification and does not invalidate their results.

Proposition 2.14 [Deng and Du 2013]. *The $\mathbb{Q}(q)$-linear algebra homomorphism*

$$\iota_n\colon \widehat{S}(n,n) \to \widehat{S}(n+1,n)$$

defined by

$$\begin{aligned} 1_\lambda &\mapsto 1_{(\lambda,0)}, \\ E_{\pm i}1_\lambda &\mapsto E_{\pm i}1_{(\lambda,0)}, \\ E_n 1_\lambda &\mapsto E_n E_{n+1}1_{(\lambda,0)}, \end{aligned}$$

$$E_{-n}1_\lambda \mapsto E_{-(n+1)}E_{-n}1_{(\lambda,0)},$$
$$E_{+\delta}1_n \mapsto E_nE_{n-1}\cdots E_1E_{n+1}1_{(1^n,0)},$$
$$E_{-\delta}1_n \mapsto E_{-(n+1)}E_{-1}\cdots E_{-n}1_{(1^n,0)}$$

for any $1 \le i \le n-1$ *and* $\lambda \in \Lambda(n,n)$, *is an embedding and gives an isomorphism of algebras*

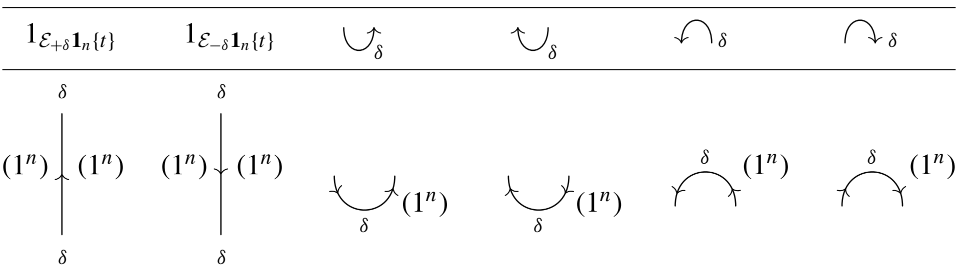
$$\widehat{\mathbf{S}}(n,n) \cong \bigoplus_{\lambda,\mu\in\Lambda(n,n)} 1_{(\lambda,0)}\widehat{\mathbf{S}}(n+1,n)1_{(\mu,0)}.$$

3. A diagrammatic categorification of $\widehat{S}(n, n)$

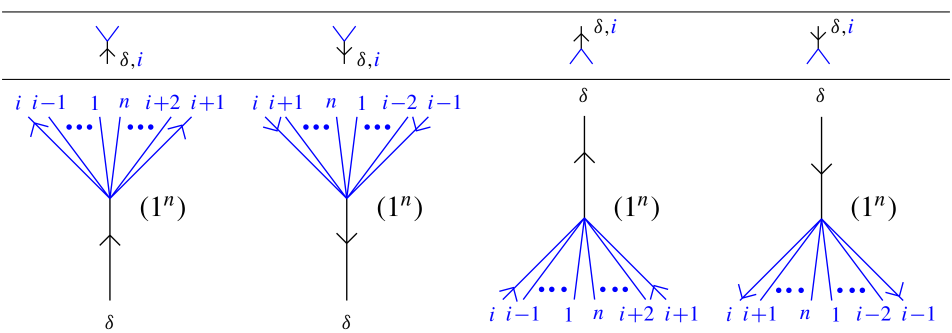
Definition 3.1. The 2-category $\widehat{\mathcal{S}}(n,n)$ is defined as the quotient of $\mathcal{U}(\widehat{\mathfrak{gl}}_n)$ by the ideal generated by all diagrams with regions whose labels are not contained in $\Lambda(n,n)$, just as in [Mackaay and Thiel 2013] (taking $y = 0$ in that paper), together with the generating 1-morphisms

$$\mathbf{1}_n\mathcal{E}_{+\delta}\mathbf{1}_n\{t\} \quad \text{and} \quad \mathbf{1}_n\mathcal{E}_{-\delta}\mathbf{1}_n\{t\},$$

for $t \in \mathbb{Z}$, the following generating 2-morphisms of degree 0 (with notation in the top row and the 2-morphisms below):

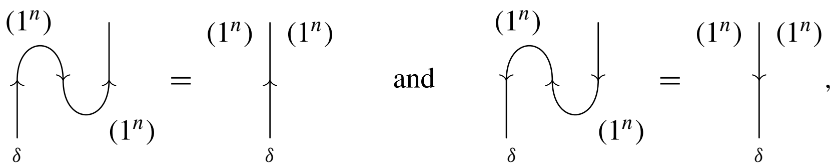


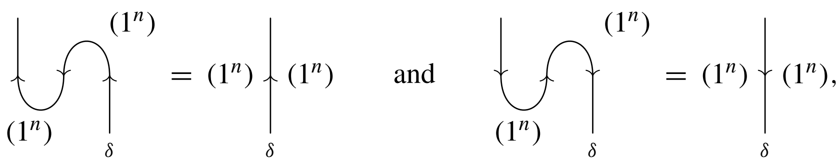
and the following generating 2-morphisms of degree 1 (again with notation in the top row and 2-morphisms below):

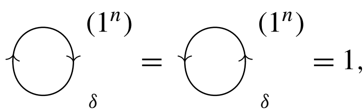


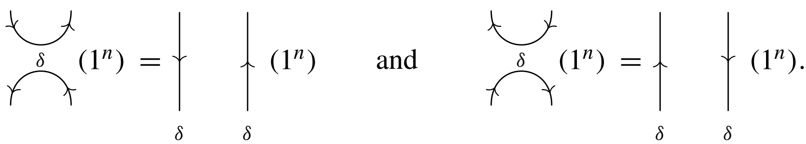
which are subject to the relations:

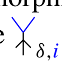
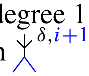
$\mathcal{E}_{+\delta}\mathbf{1}_n$ and $\mathcal{E}_{-\delta}\mathbf{1}_n$ are biadjoint inverses of each other,

(3-1) 

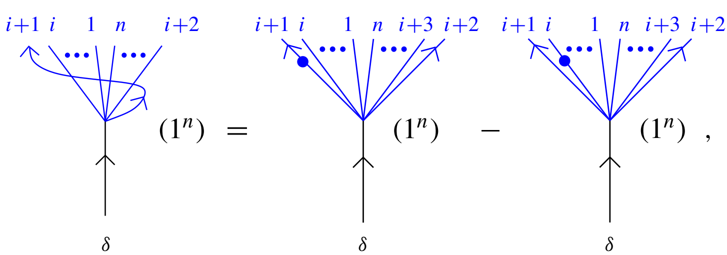
(3-2) 

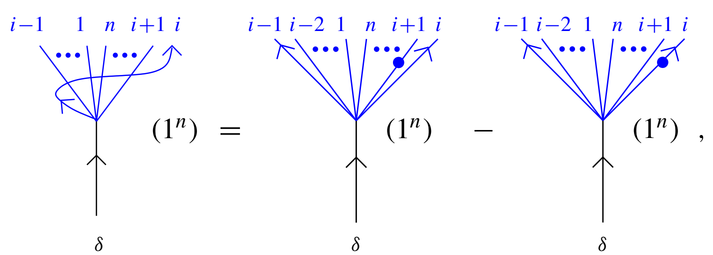
(3-3) 

(3-4) 

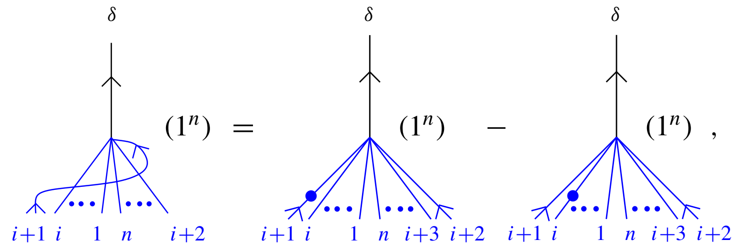
We impose full cyclicity with respect to our generating 2-morphisms of degree 1; for example, by using the adequate cups and caps we can rotate  to obtain .

Furthermore, we impose the relations

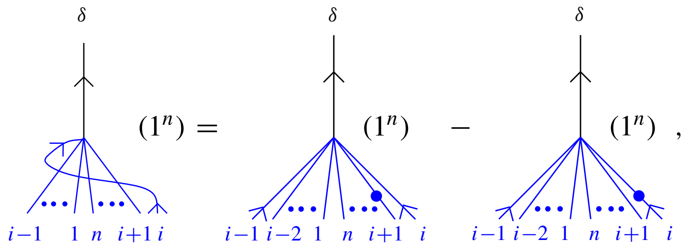
(3-5) 

(3-6) 

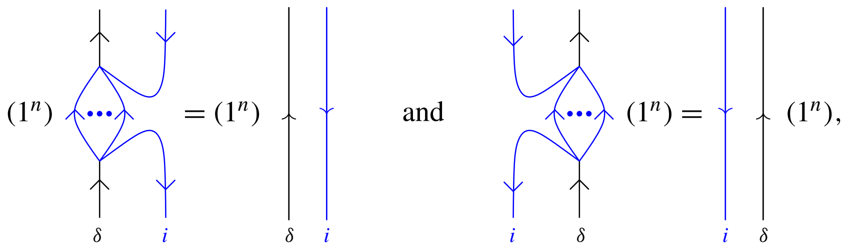
(3-7)



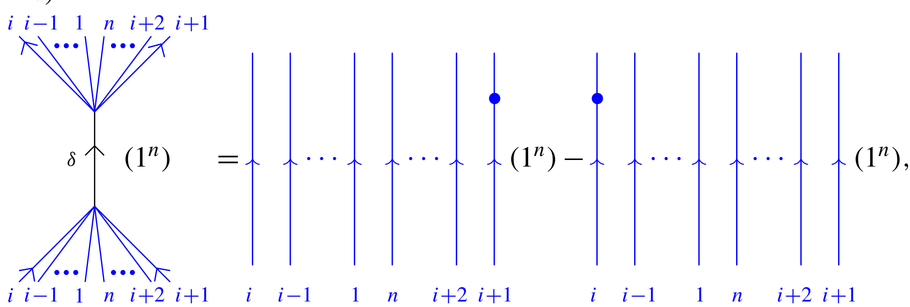
(3-8)



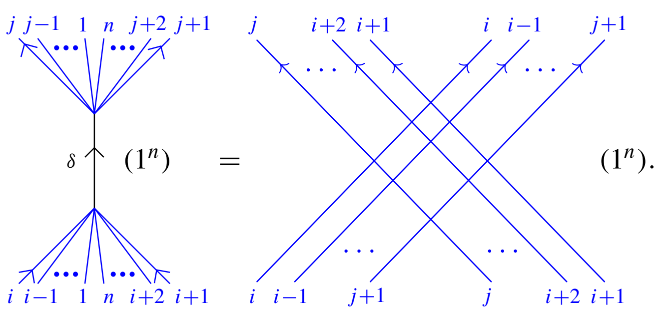
(3-9)



(3-10)



(3-11)

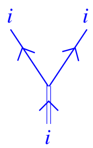
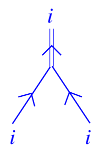


Note that cyclicity implies the analogous relations with all orientations reversed.

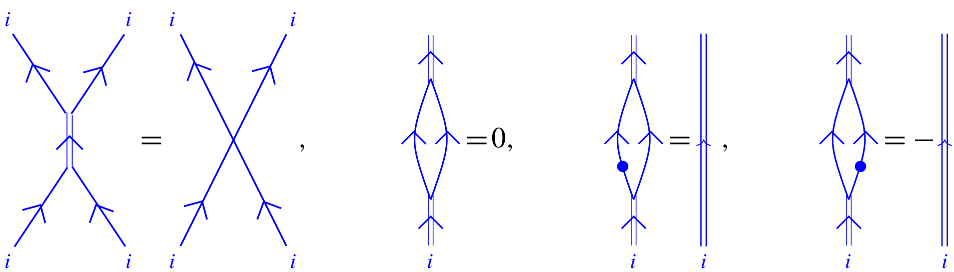
Before giving the following lemma, we recall that the Karoubi envelope (or idempotent completion) of Khovanov and Lauda's 2-categories, e.g., $\operatorname{Kar}\mathcal{U}(\mathfrak{sl}_n)$ and $\operatorname{Kar}\mathcal{U}(\mathfrak{gl}_n)$, contain the categorified divided powers $\mathcal{E}_{\pm i}^{(a)}$, which satisfy

$$\mathcal{E}_{\pm i}^{a} = (\mathcal{E}_{\pm i}^{(a)})^{\oplus [a]!}.$$

In [Khovanov et al. 2012] the 2-morphisms in $\operatorname{Kar}\mathcal{U}(\mathfrak{sl}_2)$ between the divided powers were worked out explicitly. Using the fact that $\operatorname{Kar}\mathcal{U}(\mathfrak{sl}_2)$ can be embedded into $\operatorname{Kar}\mathcal{U}(\widehat{\mathfrak{sl}}_n)$ for any choice of simple root, we can use the results in [loc. cit.]. We do not need much of that calculus in this paper, but we do have to recall the *splitters* (see the definitions below Lemma 2.2.3 and see (2.63) in [loc. cit.])


$: \mathcal{E}_{+i}^{(2)} \to \mathcal{E}_{+i}^{2}$ and

$: \mathcal{E}_{+i}^{2} \to \mathcal{E}_{+i}^{(2)}$

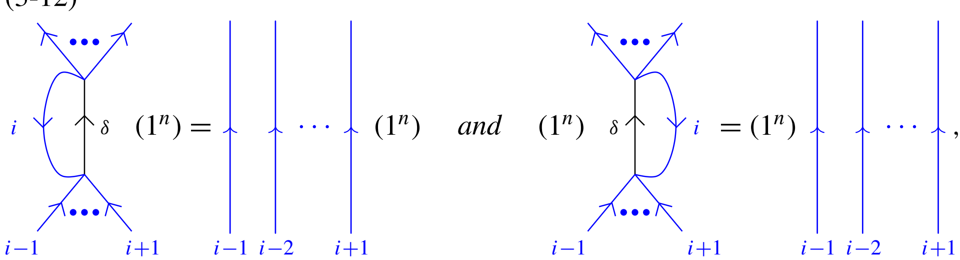
and the relations (see (2.36), (2.64) and (2.65) in [loc. cit.])



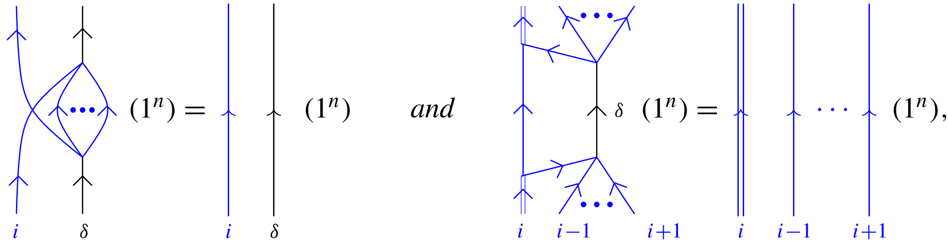
for any $i = 1, \dots, n$. By cyclicity, we get similar splitters and relations for $\mathcal{E}_{-i}^{(2)}$, $i = 1, \dots, n$.

Lemma 3.2.

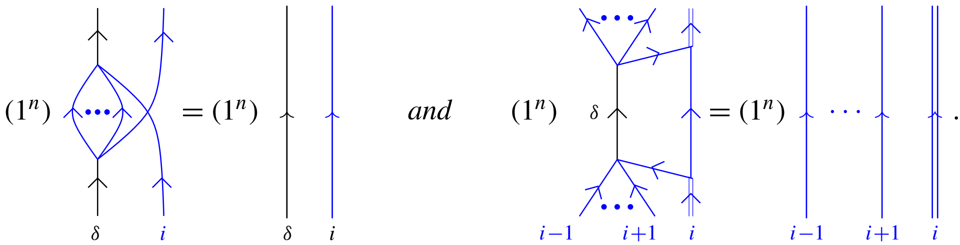
(3-12)



(3-13)



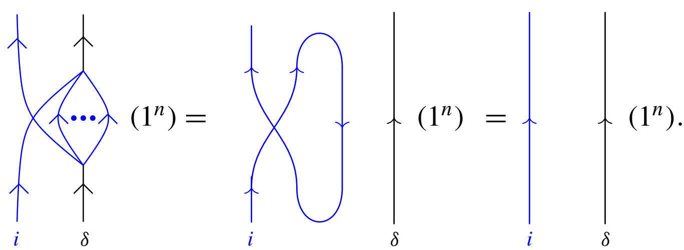
(3-14)



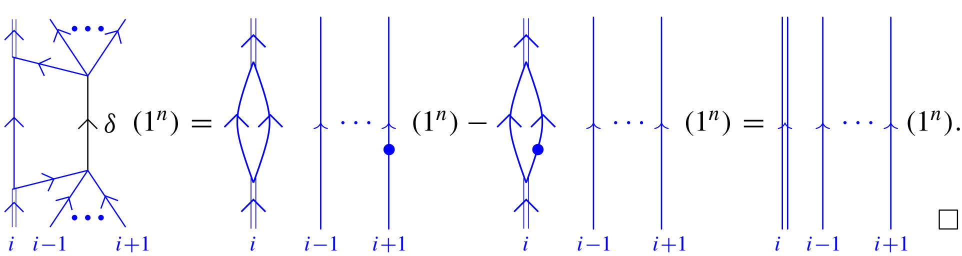
By cyclicity, we get the analogous relations with all orientations reversed.

Proof. The equations in (3-12) follow directly from (3-10) and the relations (3.39) and (3.40) in [Mackaay and Thiel 2013]. Note that one of the terms we get by applying (3-10) has a bubble of degree -2, which is equal to 0, and the other term has a bubble of degree 0 which is equal to -1 if it is counterclockwise and $+1$ if it is clockwise.

We only prove the equations in (3-13). The equations in (3-14) can be proved similarly. By the second relation in (3-9), curl removal and the evaluation of degree-0 bubbles, we get

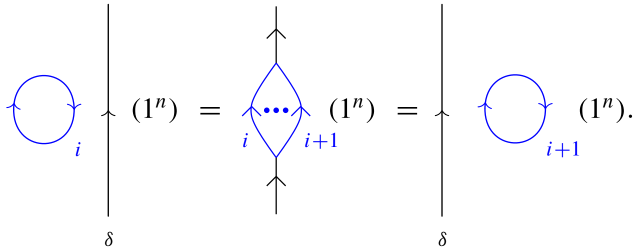


By (3-10) and the relations in (2.64) in [Khovanov et al. 2012], we get



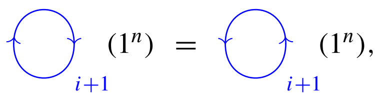
□

Lemma 3.3. *We have*



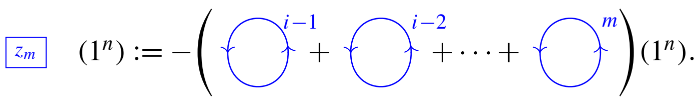
Proof. The first equality is a direct consequence of the first relation in (3-9).

The second is a consequence of the first relation in (3-9) and the fact that



which follows from the infinite Grassmannian relation for bubbles. □

In order to formulate the following results, define

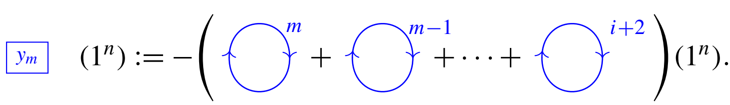


The sum of the bubbles is over the colors

$$\begin{cases} i-1, i-2, \dots, m & \text{if } 1 \le m \le i-1, \\ i-1, i-2, \dots, 1, n, n-1, \dots m & \text{if } m \ge i+1. \end{cases}$$

These are exactly the colors of all the strands in the diagram on the left-hand side of Lemma 3.4 between the strands $i-1$ and m. By definition we take $\boxed{z_i} = 0$ and use the convention that $0^0 = 1$.

Similarly, we define



The sum of the bubbles is over the colors

$$\begin{cases} m, m-1, \dots, i+2 & \text{if } i+2 \le m \le n, \\ m, m-1, \dots, 1, n, n-1, \dots i+2 & \text{if } m \le i+1. \end{cases}$$

These are exactly the colors of all the strands in the diagram on the left-hand side of Lemma 3.4 between the strands m and $i+2$. By definition we take $\boxed{y_{i+1}} = 0$ and use the convention that $0^0 = 1$.

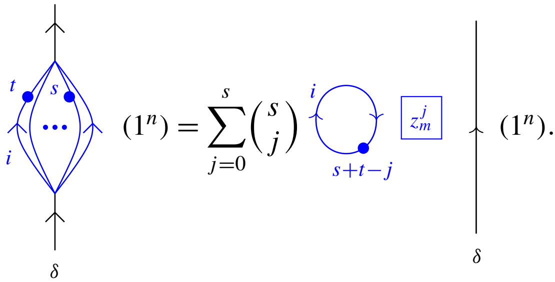
Note that

$$\boxed{y_{i-1}} = \boxed{z_{i+2}}$$

by the infinite Grassmannian relation.

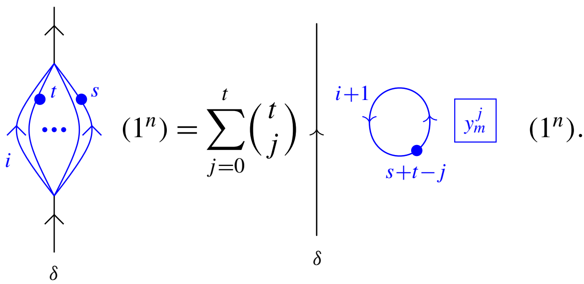
Lemma 3.4. *For any* $1 \le m \le n$ *and* $s, t \in \mathbb{N}$, *we have*

(3-15)



On the left-hand side of (3-15), *the* t *dots are on the* i*-th strand and the* s *dots are on the* m*-th strand. Similarly, we have*

(3-16)

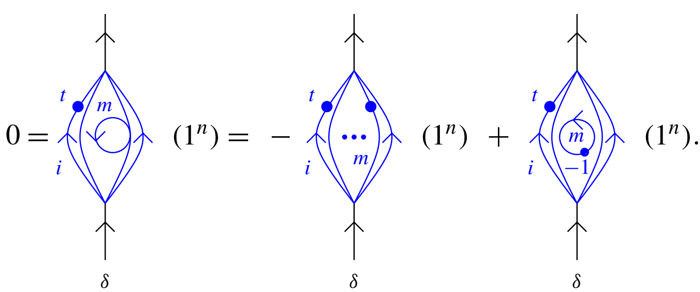


On the left-hand side of (3-15), *the* t *dots are on the* m*-th strand and the* s *dots are on the* $(i+1)$*-th strand.*

Proof. We only prove the first equation. The second can be proved in a similar way. The proof is by induction with respect to s. For $s = 0$ and any $1 \le m \le n$ and $t \in \mathbb{N}$, the result follows from (3-9).

Suppose $s > 0$, $t \in \mathbb{N}$ and $m \neq i+1$. The case $m = i$ follows from (3-9), so we can assume that $m \neq i$. First note the following:

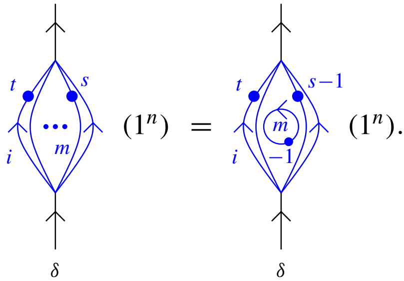
(3-17)



The first equality holds because the label of the region inside the curl does not belong to $\Lambda(n,n)$; its $(m+1)$-th entry equals -1. The second equality follows from resolving the curl. The minus sign is a consequence of our normalization of degree-0 bubbles in [Mackaay and Thiel 2013], because the label λ of the region just outside the bubble satisfies $\lambda_{m+1} = 0$. Note that the bubble in the second term has degree 2, since $\lambda_m - \lambda_{m+1} = 1$ for any $m \neq i, i+1$.

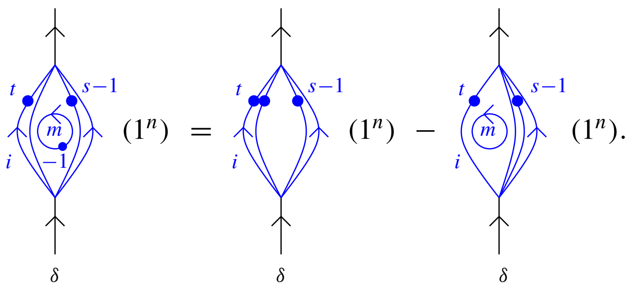
Equation (3-17) implies

(3-18)



Now slide the m-bubble to the left. Note that the strand directly to the left of the bubble has color $m+1$ (the colors are still taken modulo n). Thus, by the bubble-slide relations and the degree-0 bubble relations in [Mackaay and Thiel 2013], we get

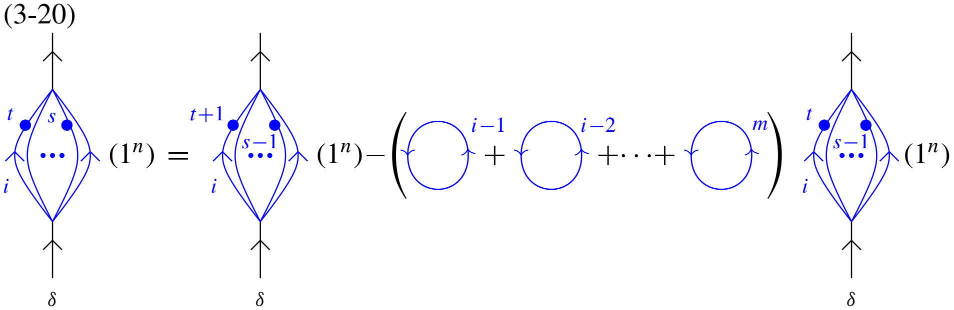
(3-19)



The new bubble, in the second diagram on the right-hand side of (3-19), still has color m of course. But now it is between the strands colored $m+2$ and $m+1$, reading from left to right. The label, λ, of the region between these two strands satisfies $\lambda_{m+1} = 1$. Thus, by the degree-0 bubble relations in [Mackaay and Thiel 2013], the counterclockwise degree-0 m-bubble in that region is equal to 1, which explains the positive sign of the first term on the right-hand side in (3-19). Note that the label of the region containing the m-bubble in the second term satisfies $\lambda_m - \lambda_{m+1} = 0$, so the dotless m-bubble has degree 2, as it should.

Note that the m-bubble in the second term in (3-19) can be slid completely to the left-hand side. After that, we can use (3-18) to eliminate the dot on the $(m+1)$-th strand and slide the $(m+1)$-bubble completely to the left-hand side. Repeating this for all strands between $i-1$ and m, we get the following result:

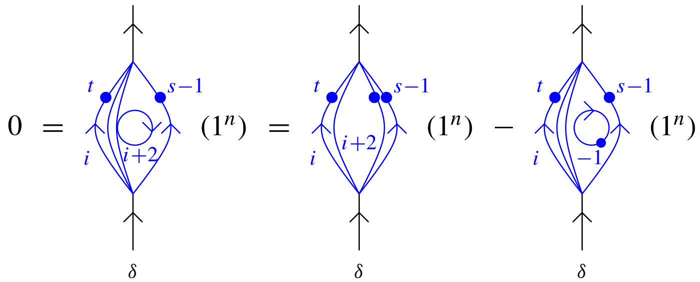
(3-20)



Induction then proves the result for $m \neq i+1$.

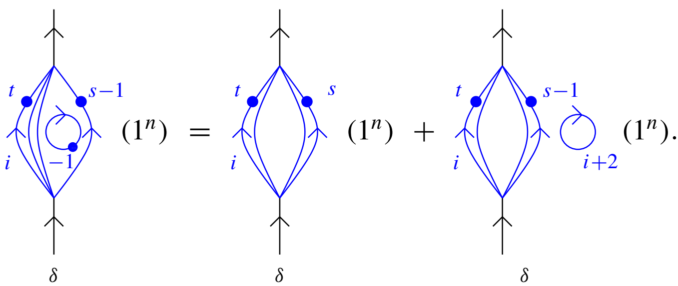
For $m = i+1$, we have to adapt our reasoning above, because the region between the $(i+2)$-th and the $(i+1)$-th strands has label $\lambda = (1^i, 2, 0, 1^{n-(i+2)})$. In particular $\lambda_{i+1} = 2$, so the left $(i+1)$-curl has degree 4 this time, which prevents us from using induction. Therefore, we use a slightly different argument involving a right curl.

We still assume that $s > 0$ holds. First note that, by the resolution of the curl and the degree-0 bubble relations in [Mackaay and Thiel 2013], we have

(3-21)


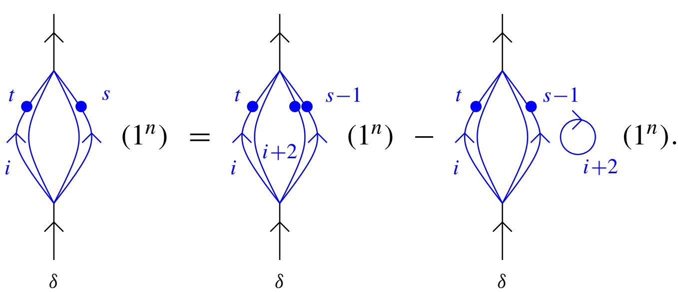
because the region between the $(i+2)$-th and the $(i+1)$-th strands is labeled $\lambda = (1^i, 2, 0, 1^{n-(i+2)})$. In particular, we have $\lambda_{i+2} - \lambda_{i+3} = -1$ and $\lambda_{i+3} = 1$, which explains the signs of the terms on the right-hand side of (3-21).

We now slide the $(i+2)$-bubble in the second term on the right-hand side of (3-21) to the right:

(3-22)


The sign of the first term on the right-hand side of (3-22) follows from the degree-0 bubble relations in [Mackaay and Thiel 2013].

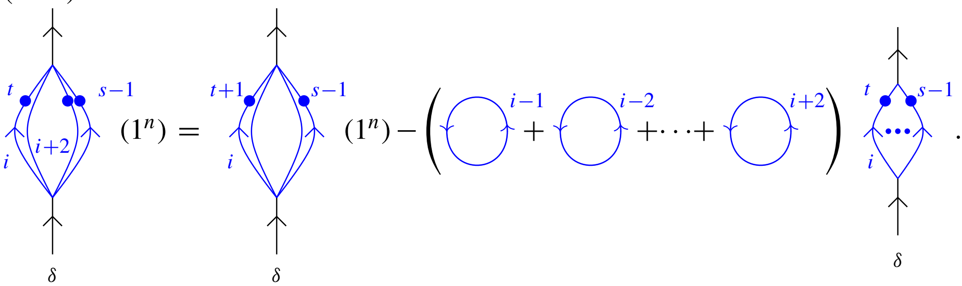
Putting (3-21) and (3-22) together, we get

(3-23)


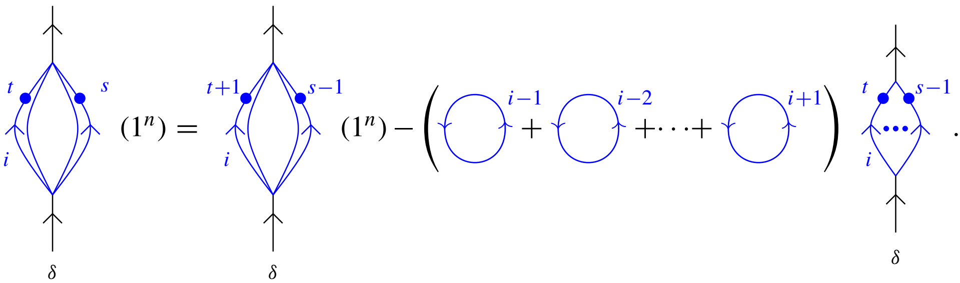
We can exchange the $(i+2)$-bubble on the right-hand side for an $(i+1)$-bubble on the left-hand side by Lemma 3.3, and invert its orientation by the infinite Grassmannian relation.

By the same reasoning as above, we get

(3-24)

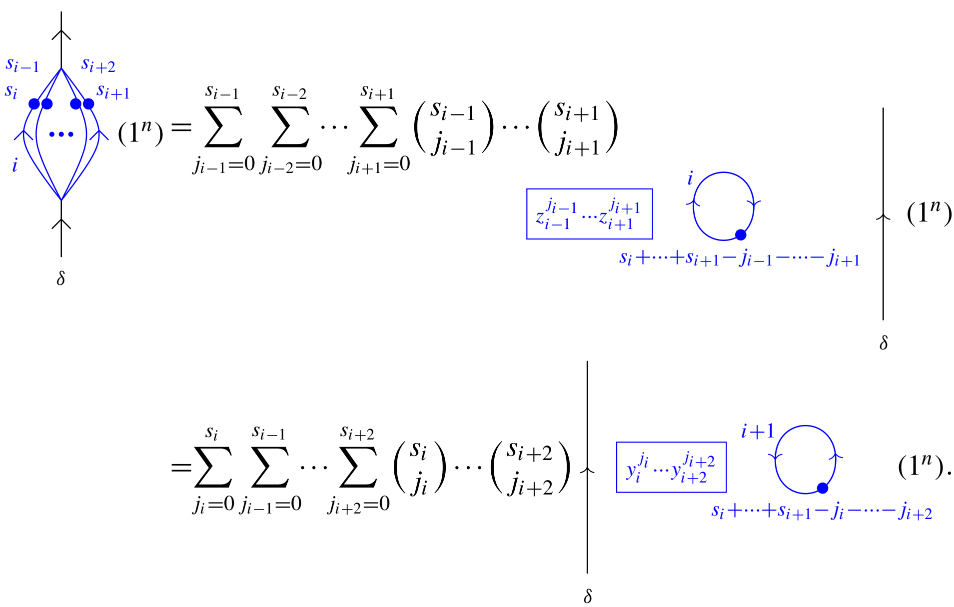


Putting (3-23) and (3-24) together, we obtain



As before, the result follows by induction. □

Proposition 3.5.



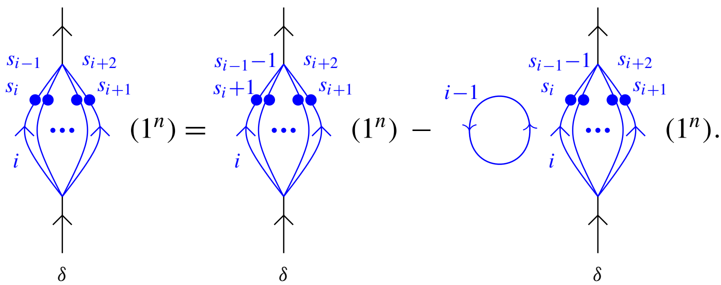
Proof. We only prove the first equation. The second can be proved by similar arguments.

We use induction with respect to the reverse lexicographical ordering of the dot sequences $(s_i, \dots, s_{i+1})$. The base of the induction, $s_i = \cdots = s_{i+1} = 0$, has been dealt with in Lemma 3.3.

The case $s_{i-1} = \cdots = s_{i+1} = 0$ has been dealt with in Lemma 3.4. Suppose there exists a $j \in \{i-1, \dots, i+1\}$ with $s_j > 0$. The argument below works for arbitrary j, but let us assume that $j = i-1$ for simplicity.

By the same arguments as used in the proof of Lemma 3.4, we get

(3-25)

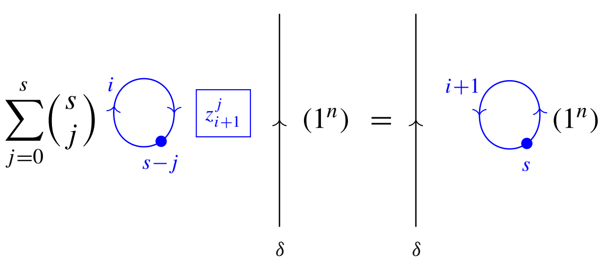


Induction on both terms on the right-hand side of (3-25) proves the proposition. $\square$

Proposition 3.5 also allows us to derive two bubble-slide formulas. The other two, for bubbles with the opposite orientation, can be obtained using the infinite Grassmannian relation and induction. Since we do not need them in this paper, we omit them.

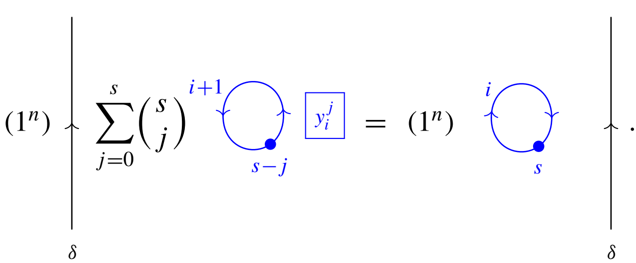
Corollary 3.6. *We have*

(3-26)



and

(3-27)



Proof. These two bubble-slide relations follow immediately from Lemma 3.4. For (3-26), apply (3-15) and (3-16) with $t=0$, $m=i+1$. For (3-27), apply (3-15) and (3-16) with $s=0$, $m=i$. $\square$

4. Two useful 2-functors

Definition 4.1. Let the 2-functor $\Psi_{n,n}\colon \mathcal{U}(\widehat{\mathfrak{sl}}_n)^* \to \widehat{\mathcal{S}}(n,n)^*$ be defined just as $\Psi_{n,r}$ in Section 3.5.3 in [Mackaay and Thiel 2013]; i.e., on objects and 1-morphisms it is determined by

$$\mu \mapsto \varphi_{n,n}(\mu) =: \lambda,$$
$$\mathcal{E}_{\underline{i}}\mathbf{1}_\mu \mapsto \mathcal{E}_{\underline{i}}\mathbf{1}_\lambda.$$

By convention, we put $1_* := 0$. On 2-morphisms it is determined by sending any diagram in $\mathcal{U}(\widehat{\mathfrak{sl}}_n)$ which is not a left cap or cup to the same diagram in $\widehat{\mathcal{S}}(n,n)$ and applying $\varphi_{n,n}$ to the labels of the regions in the diagram. The images of the left caps and cups also have to be multiplied by certain signs. To be more precise, define

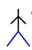
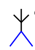
(4-1) $_{i,\mu} \mapsto (-1)^{\lambda_{i+1}+1}$$_{i,\lambda}$ and $_{i,\mu} \mapsto (-1)^{\lambda_{i+1}}$$_{i,\lambda}$.

We define any diagram in $\widehat{\mathcal{S}}(n,n)$ to be equal to 0 if it contains regions labeled $*$.

Note that, unlike $\Psi_{n,r}$ for $n>r$, $\Psi_{n,n}$ is not essentially surjective. However, it still has the following useful property.

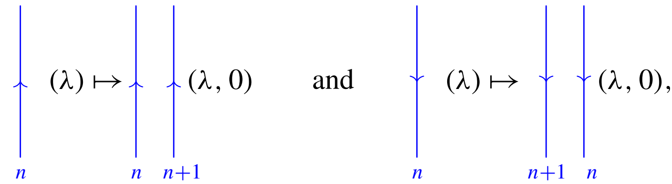
Lemma 4.2. *The 2-functor $\Psi_{n,n}$ is full.*

Proof. The proof follows from the following two observations, which show how to remove δ-strands from diagrams in $\mathrm{HOM}_{\widehat{\mathcal{S}}(n,n)}(\mathcal{E}_{\underline{i}}1_\lambda, \mathcal{E}_{\underline{j}}1_\lambda)$, for any signed sequences $\underline{\boldsymbol{i}}$ and $\underline{\boldsymbol{j}}$:

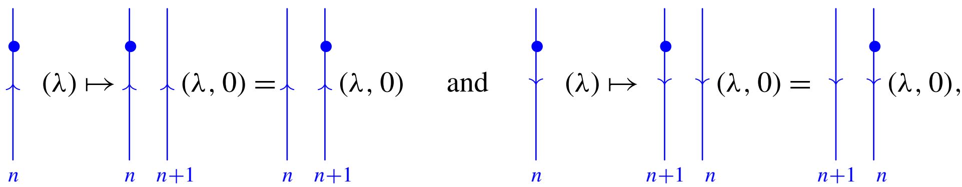
- Closed δ-diagrams always consist of disjoint δ-circles. By Corollary 3.6 we can move any closed i-diagram, which is always equivalent to a linear combination of disjoint i-circles, from the interior to the exterior of a δ-circle. By (3-3), we can then remove the δ-circles with empty interior.
- Any δ-strand which is not part of a δ-circle has to be part of a diagram obtained by gluing $_{\delta,j}$ on top of $^{\delta,i}$ or $_{\delta,i}$ on top of $^{\delta,j}$ for certain $1 \le i, j \le n$. In both cases we can remove the δ-strand by applying (3-10) or (3-11). $\square$

Definition 4.3. We define the 2-functor $\mathcal{I}_n\colon \widehat{\mathcal{S}}(n,n) \to \widehat{\mathcal{S}}(n+1,n)$ as follows:

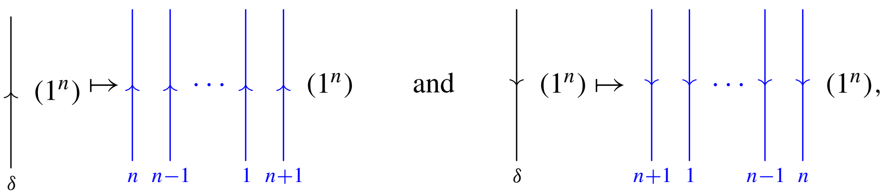
- On objects and 1-morphisms, use the map in Proposition 2.14.
- On 2-morphisms, take the identity on all i-strands, for $1 \le i \le n-1$, map all n-strands to two parallel strands labeled n and $n+1$, e.g.,

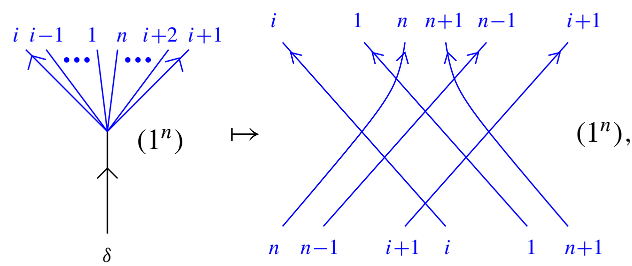


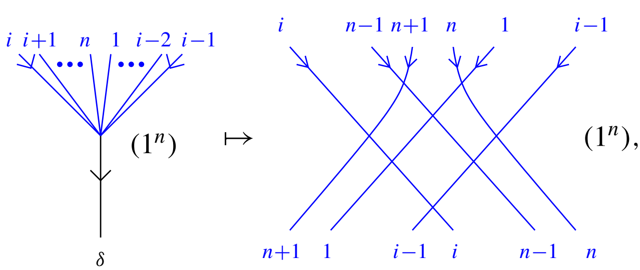
map dots on n-strands to dots on the corresponding pairs of parallel strands as follows:



and map the generators involving δ-strands as follows:







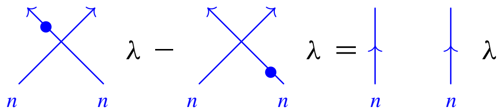
with the image of the other two δ-splitters being defined likewise using cyclicity.

Note that the two images of the dotted n-strands which are shown, are indeed equal in $\widehat{\mathcal{S}}(n+1, n)$. This follows from the relevant Reidemeister-2 relations, because the diagrams with the crossings in those relations are equal to 0 (the last entry of the labels of their middle regions is equal to -1).

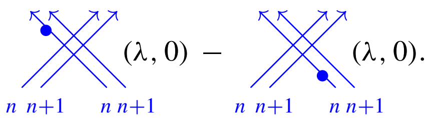
Lemma 4.4. *For any $n \geq 3$, $\mathcal{I}_n$ is well-defined.*

Proof. We only have to prove that $\mathcal{I}_n$ preserves the relations involving n and δ-strands, because $\mathcal{I}_n$ clearly preserves all other relations.

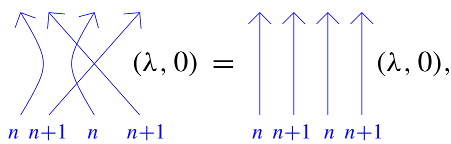
First consider the nil-Hecke relations which only involve n-strands. By cyclicity, we can assume that all strands are oriented upward. We give the proof of well-definedness with respect to one nil-Hecke relation in detail. The image of the left-hand side of

(4-2)


is given by



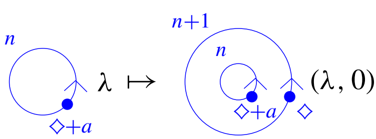
By the nil-Hecke relation for the n-strands, this is equal to



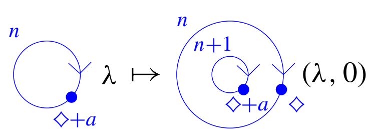
which is equal to the image of the right-hand side of (4-2). Note that in the last equality, we have omitted one term which is equal to 0 because it contains a region whose label has a negative entry.

Well-definedness with respect to the other two nil-Hecke relations for n-strands can be proved by similar arguments.

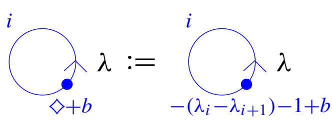
As for the other relations involving only n-strands, the first one we should look at is the infinite Grassmannian relation. The image of the n-bubbles is given by



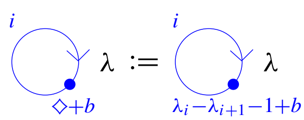
and



for any $a \in \mathbb{N}$ and $\lambda \in \Lambda(n, n)$. The notation $\diamond$ is defined by

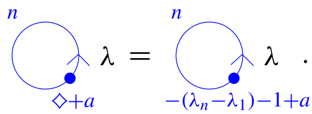


and

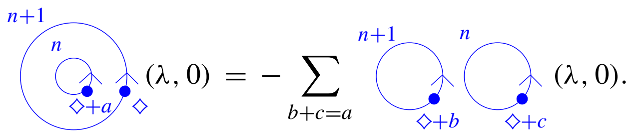


for any $b \in \mathbb{N}$.

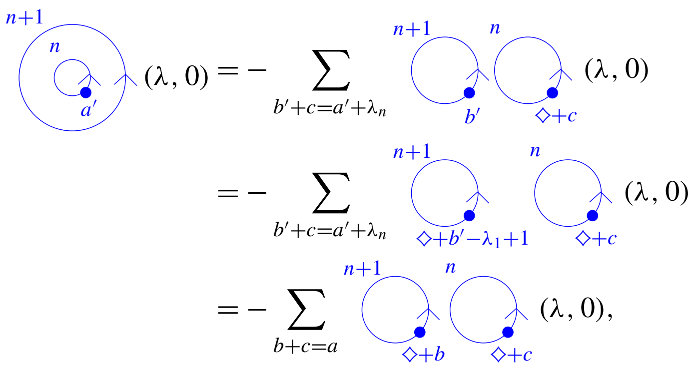
For $\diamondsuit + a < 0$, the image of the fake n-bubbles above is a definition. For $\diamondsuit + a \geq 0$, we have to prove that the image of the n-bubbles above is equal to the image assigned to them by $\mathcal{I}_n$. This is immediate if the two nested bubbles in the image are real (since the numbers of dots match), but one of them could be fake, in which case a proof is required. Let us give this proof for the counterclockwise n-bubbles. Note that

(4-3)


By the definition above, the image of the left-hand side of (4-3) is given by

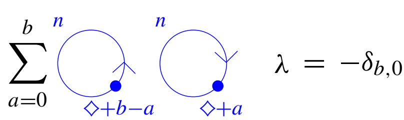


The equality is obtained by applying a bubble-slide relation. By the definition of $\mathcal{I}_n$, the image of the right-hand side of (4-3) is given by

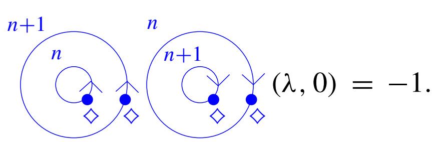


with $a' = -(\lambda_n - \lambda_1) - 1 + a$. The first equality is obtained by applying a bubble-slide relation, and the other equalities are obtained by reindexing. This finishes the proof that both definitions of the image of the counterclockwise nonfake n-bubbles are equal. The proof for the clockwise n-bubbles is similar and is left to the reader.

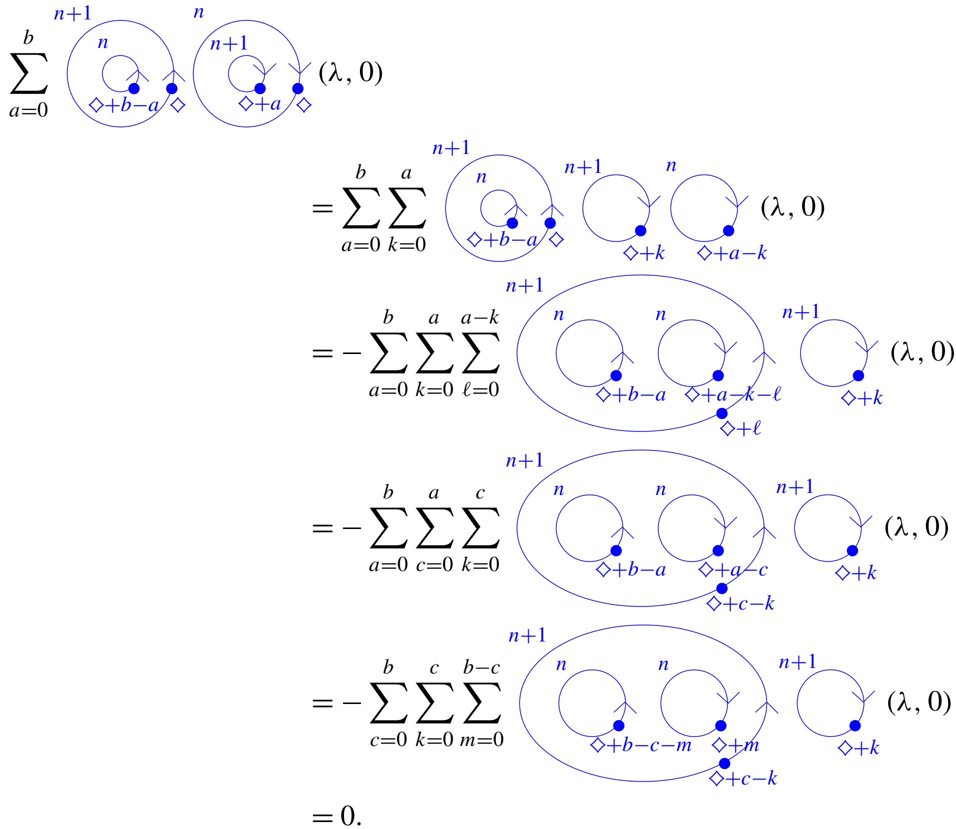
We now show that with the definitions above, the images of the bubbles satisfy the infinite Grassmannian relation. To be more precise, we have to show that the relation

(4-4)


is preserved, for any $b \in \mathbb{N}$. For $b = 0$, the image of (4-4) is given by



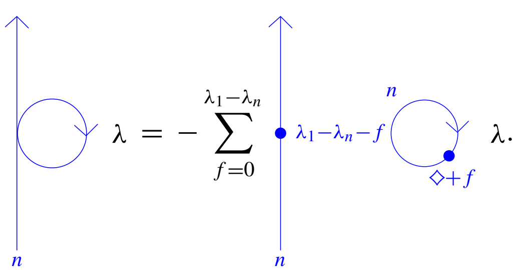
The equality follows immediately from the degree-0 bubble relations. For $b > 0$, the image of (4-4) is given by



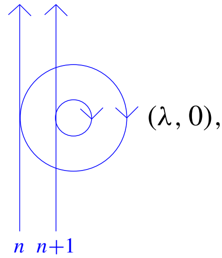
The first two equalities follow from bubble-slide relations. The next two equalities follow from reindexing, as indicated. The last equality follows from the infinite Grassmannian relation: for the n-bubbles, if $b > c$ (with c fixed), and for the $(n+1)$-bubbles if $b = c$.

Knowing the images of the fake bubbles allows us to prove the other relations involving only n-strands very easily. Let us do just one example; the other relations can be proved in a similar fashion. We show that $\mathcal{I}_n$ preserves the relation

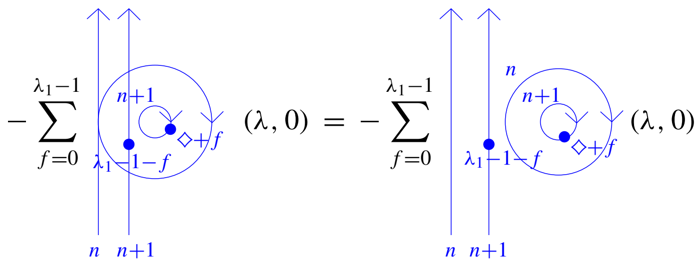
(4-5)

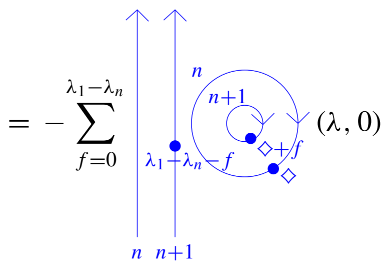


The image of the left-hand side of (4-5) is given by



which is equal to

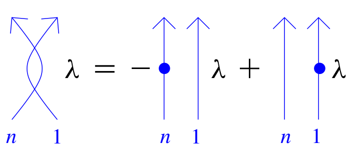




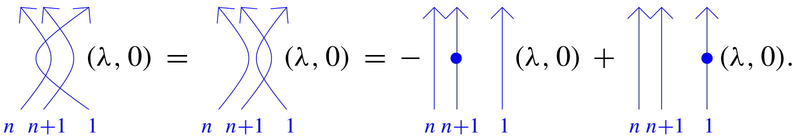
The first summation is obtained by resolving the $(n+1)$-curl. The second summation can then be obtained by applying a Reidemeister-3 relation to the strands colored n, $n+1$ and n. Note that only the terms which are shown survive; the other ones are 0 because they are given by diagrams which contain a region whose label has a negative entry. The last summation is obtained by first reindexing. Then an argument similar to the one we used below (4-3) ensures that the nested bubbles, before and after the equality, match and that the first $\lambda_n - 1$ terms of the reindexed summation vanish (indeed in those terms, bubbles of negative degree appear, and those are always 0). This last expression is equal to the image of the right-hand side of (4-5), which finishes our proof that $\mathcal{I}_n$ preserves (4-5).

Next let us have a look at the relations involving i-strands of more than one color. We just do one example in detail, the other relations can be proved in a similar fashion. Consider the relation

(4-6)

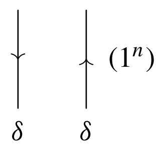


in $\widehat{\mathcal{S}}(n,n)$. The image of the term on the left-hand side is given by

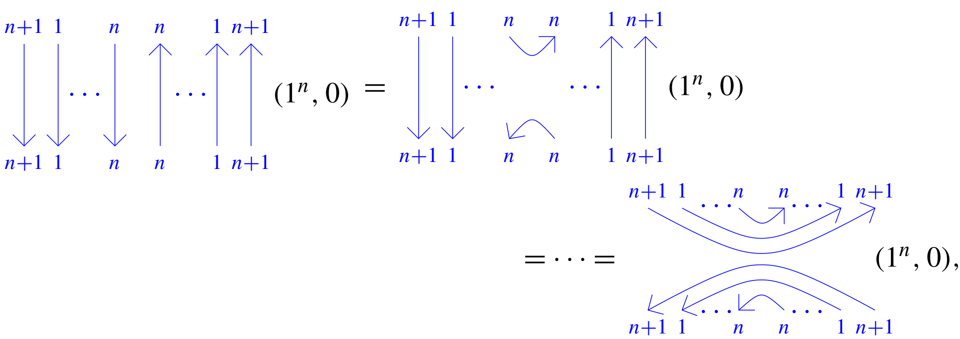


The first and the second equalities follow from the Reidemeister-2 relations in $\widehat{\mathcal{S}}(n+1,n)$. The linear combination at the end is exactly the image of the right-hand side in (4-6), which proves that (4-6) is preserved by $\mathcal{I}_n$.

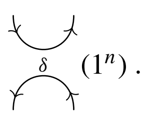
It remains to be proved that $\mathcal{I}_n$ preserves the relations involving δ-strands. For the relations (3-1) and (3-2), the proof follows immediately from the zigzag relations for i-strands with $i = 1, \ldots, n+1$. For the relations in (3-3), the proof follows immediately from the degree-0 i-bubble relations for $i = 1, \ldots, n+1$. Let us explain the first relation in (3-4) in more detail, the second being similar. The image of



is given by



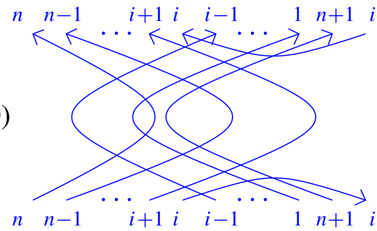
which is indeed equal to the image of



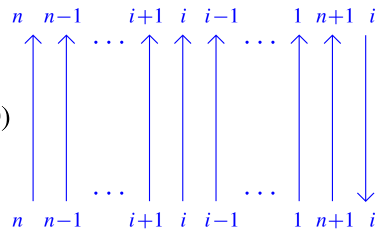
The equalities above are obtained by repeatedly applying Reidemeister-2 relations on the pairs of i-strands with $\lambda_i - \lambda_{i+1} = -1$ for all $i = 1, \ldots, n+1$. Note that the terms with two i-crossings are all equal to 0, because they contain a region whose label has one negative entry, and all bubbles in the other terms are of degree 0 and equal to -1.

The fact that relations (3-5), (3-6), (3-7) and (3-8) are preserved follows easily from applying Reidemeister-2 and -3 relations to the images of the terms on their left-hand side. The dots appear after applying the Reidemeister-2 relation involving the i and $(i+1)$-strands.

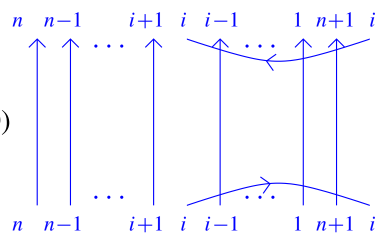
We prove the left relation in (3-9) for $1 \le i < n$. The proof for $i = n$ and the proof of the right relation in (3-9) are similar and are left to the reader. The image on the left-hand side of the first relation in (3-9) is given by

(4-7) $(1^n, 0)$ 

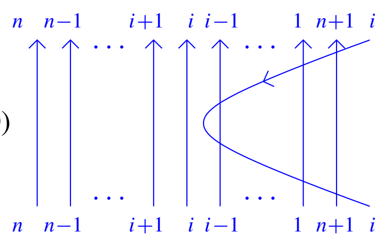
We claim that this is equal to

$(1^n, 0)$ 

which is indeed the image of the right-hand side of (3-9). This follows from first applying Reidemeister-2 relations to (4-7) in order to straighten all j-strands for $j \neq i$:

$(1^n, 0)$ 

then a Reidemeister-2 relation to the i-strands in the middle (note that the region at the top and the bottom between the i and the $(i-1)$-strand is labeled $(1, \ldots, 1, 0, 1, \ldots, 1)$ with 0 on the i-th position):

$(1^n, 0)$ 

and finally Reidemeister-2 relations in order to straighten the downward i-strand.

Finally, the fact that $\mathcal{I}_n$ preserves the relations (3-10) and (3-11) can be easily proved by applying Reidemeister-2 and -3 relations to the images of the diagrams on the left-hand sides of those two relations. $\square$

5. The Grothendieck group

In this section we prove that $\widehat{\mathcal{S}}(n,n)$ categorifies $\widehat{\mathbf{S}}(n,n)$ (Theorem 5.4). All the hard work has been done already, we just have to put everything together. In the following lemma, we show that all relations in $\widehat{\mathbf{S}}(n,n)$, which are listed in Theorem 2.11, hold up to isomorphism in $\widehat{\mathcal{S}}(n,n)$.

Lemma 5.1. *In $\widehat{\mathcal{S}}(n,n)$, we have*

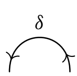
$$
\begin{array}{ll}
\text{(i)} & \mathcal{E}_{\pm\delta}\mathbf{1}_\lambda \cong \mathbf{1}_\lambda\mathcal{E}_{\pm\delta} \cong 0 \quad \textit{for all } \lambda \neq (1^n),\\
\text{(ii)} & \mathcal{E}_{\pm\delta}\mathbf{1}_n \cong \mathbf{1}_n\mathcal{E}_{\pm\delta},\\
\text{(iii)} & \mathcal{E}_{+\delta}\mathcal{E}_{-\delta}\mathbf{1}_n \cong \mathcal{E}_{-\delta}\mathcal{E}_{+\delta}\mathbf{1}_n \cong \mathbf{1}_n,\\
\text{(iv)} & \mathcal{E}_i\mathcal{E}_{+\delta}\mathbf{1}_n \cong \mathcal{E}_i^{(2)}\mathcal{E}_{i-1}\ldots\mathcal{E}_1\mathcal{E}_n\cdots\mathcal{E}_{i+1}\mathbf{1}_n,\\
\text{(v)} & \mathbf{1}_n\mathcal{E}_{+\delta}\mathcal{E}_i \cong \mathbf{1}_n\mathcal{E}_{i-1}\ldots\mathcal{E}_1\mathcal{E}_n\cdots\mathcal{E}_{i+1}\mathcal{E}_i^{(2)},\\
\text{(vi)} & \mathcal{E}_{-i}\mathcal{E}_{+\delta}\mathbf{1}_n \cong \mathcal{E}_{i-1}\cdots\mathcal{E}_1\mathcal{E}_n\cdots\mathcal{E}_{i+1}\mathbf{1}_n,\\
\text{(vii)} & \mathbf{1}_n\mathcal{E}_{+\delta}\mathcal{E}_{-i} \cong \mathbf{1}_n\mathcal{E}_{i-1}\cdots\mathcal{E}_1\mathcal{E}_n\cdots\mathcal{E}_{i+1},\\
\text{(viii)} & \mathcal{E}_{-i}\mathcal{E}_{-\delta}\mathbf{1}_n \cong \mathcal{E}_{-i}^{(2)}\mathcal{E}_{-(i+1)}\cdots\mathcal{E}_{-n}\mathcal{E}_{-1}\cdots\mathcal{E}_{-(i-1)}\mathbf{1}_n,\\
\text{(ix)} & \mathbf{1}_n\mathcal{E}_{-\delta}\mathcal{E}_{-i} \cong \mathbf{1}_n\mathcal{E}_{-(i+1)}\cdots\mathcal{E}_{-n}\mathcal{E}_{-1}\cdots\mathcal{E}_{-(i-1)}\mathcal{E}_{-i}^{(2)},\\
\text{(x)} & \mathcal{E}_i\mathcal{E}_{-\delta}\mathbf{1}_n \cong \mathcal{E}_{-(i+1)}\cdots\mathcal{E}_{-n}\mathcal{E}_{-1}\cdots\mathcal{E}_{-(i-1)}\mathbf{1}_n,\\
\text{(xi)} & \mathbf{1}_n\mathcal{E}_{-\delta}\mathcal{E}_i \cong \mathbf{1}_n\mathcal{E}_{-(i+1)}\cdots\mathcal{E}_{-n}\mathcal{E}_{-1}\cdots\mathcal{E}_{-(i-1)}
\end{array}
$$

for any $i = 1,\ldots,n$.

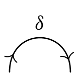
Proof. The isomorphisms in (i) and (ii) are immediate.

For (iii), consider the 2-morphisms

 $$(1^n)\colon \mathbf{1}_n \to \mathcal{E}_{-\delta}\mathcal{E}_{+\delta}\mathbf{1}_n,$$

 $$(1^n)\colon \mathcal{E}_{-\delta}\mathcal{E}_{+\delta}\mathbf{1}_n \to \mathbf{1}_n,$$

 $$(1^n)\colon \mathbf{1}_n \to \mathcal{E}_{+\delta}\mathcal{E}_{-\delta}\mathbf{1}_n,$$

 $$(1^n)\colon \mathcal{E}_{+\delta}\mathcal{E}_{-\delta}\mathbf{1}_n \to \mathbf{1}_n.$$

Relations (3-3) and (3-4) show that these 2-morphisms are 2-isomorphisms.

Similarly, the isomorphisms in (iv) and (v) follow from the relations in (3-13) and (3-14), and the isomorphisms in (vi) and (vii) follow from the relations in (3-9) and (3-12).

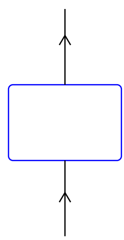
The isomorphisms in (viii)–(xi) follow from the ones above by biadjointness. $\square$

Recall that $\mathrm{END}(X)$ denotes the ring generated by all homogeneous 2-endomorphisms of a given 1-morphism X, whereas $\mathrm{End}(X) \subset \mathrm{END}(X)$ only contains the ones of degree 0.

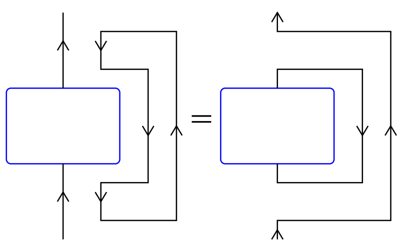
Lemma 5.2. *For any $t \in \mathbb{Z}$,*

$$\mathrm{END}(\mathcal{E}^t_{+\delta}\mathbf{1}_n) \cong 1_{\mathcal{E}^t_{+\delta}}\,\mathrm{END}(\mathbf{1}_n) \cong \mathrm{END}(\mathbf{1}_n)1_{\mathcal{E}^t_{+\delta}}.$$

Proof. Note that for $t = 0$ there is nothing to prove. Let us now explain the proof for $t = 1$. Given a diagram of the form



we can create a δ-bubble by (3-3) and apply (3-4) to obtain



This proves the lemma for $t = 1$. For $t > 1$, use the same trick repeatedly until you are left with a closed diagram and t upward δ-strands. For $t < 0$, a similar trick can be applied using the opposite orientation on the δ-strands. $\square$

Let $K_0(\mathrm{Kar}\,\widehat{\mathcal{S}}(n,n))$ be the split Grothendieck group of $\mathrm{Kar}\,\widehat{\mathcal{S}}(n,n)$. This is a $\mathbb{Z}[q,q^{-1}]$-module, where the action of q is defined by

$$q[X] := [X\{1\}].$$

Furthermore, let

$$K_0^{\mathbb{Q}(q)}(\mathrm{Kar}\,\widehat{\mathcal{S}}(n,n)) := K_0(\mathrm{Kar}\,\widehat{\mathcal{S}}(n,n)) \otimes_{\mathbb{Z}[q,q^{-1}]} \mathbb{Q}(q).$$

Definition 5.3. Define the $\mathbb{Q}(q)$-linear algebra homomorphism $\gamma_n \colon \widehat{\boldsymbol{S}}(n,n) \to K_0^{\mathbb{Q}(q)}(\operatorname{Kar}\widehat{\mathcal{S}}(n,n))$ by

$$\gamma_n(E_{\underline{i}}1_\lambda) := [\mathcal{E}_{\underline{i}}\mathbf{1}_\lambda] \otimes 1 \quad \text{and} \quad \gamma_n(E^t_{+\delta}1_n) := [\mathcal{E}^t_{+\delta}\mathbf{1}_n] \otimes 1$$

for any signed sequence $\underline{\boldsymbol{i}}$, $\lambda \in \Lambda(n,n)$ and $t \in \mathbb{Z}$.

Theorem 5.4. *The homomorphism γ_n is well-defined and bijective.*

Proof. Well-definedness follows from the corresponding statement for $\mathcal{U}(\widehat{\mathfrak{sl}}_n)$ by Khovanov and Lauda [2010] and from Theorem 2.11 and Lemma 5.1.

Let us now show surjectivity. By Lemma 5.1, any indecomposable object in $\operatorname{Kar}\widehat{\mathcal{S}}(n,n)$ is isomorphic to an object of the form (X, e), where X is either of the form $\mathcal{E}^t_{+\delta}$ for some $t \in \mathbb{Z}$ or of the form $\mathcal{E}_{\underline{i}}$ for some signed sequence $\underline{\boldsymbol{i}}$, and e is some idempotent in $\operatorname{End}(X)$. By Lemmas 4.2 and 5.2, we see that $\operatorname{End}(\mathcal{E}^t_{+\delta}) \cong \mathbb{Q}1_{\mathcal{E}^t_{+\delta}}$. Therefore $\mathcal{E}^t_{+\delta}$ is indecomposable in $\operatorname{Kar}\widehat{\mathcal{S}}(n,n)$. Note that its Grothendieck class lies indeed in the image of γ_n. By Lemma 4.2 we know that $\operatorname{End}_{\widehat{\mathcal{S}}(n,n)}(\mathcal{E}_{\underline{i}})$ is the surjective image of the analogous endomorphism ring in $\mathcal{U}(\widehat{\mathfrak{sl}}_n)$ for any signed sequence $\underline{\boldsymbol{i}}$. By [Khovanov and Lauda 2010, Theorem 1.1] and some general arguments which were explained in detail in [Mackaay et al. 2013], and also used in [Mackaay and Thiel 2013], this implies that the Grothendieck classes of all direct summands of $\mathcal{E}_{\underline{i}}$ in $\operatorname{Kar}\widehat{\mathcal{S}}(n,n)$ are contained in the image of γ_n. This concludes the proof that γ_n is surjective.

For injectivity, consider the following commutative diagram

$$\begin{tikzcd}
\widehat{\boldsymbol{S}}(n,n) \arrow[r, "\iota_n"] \arrow[d, "\gamma_n"] & \widehat{\boldsymbol{S}}(n+1,n) \arrow[d, "\gamma_{n+1}"] \\
K_0^{\mathbb{Q}(q)}(\operatorname{Kar}\widehat{\mathcal{S}}(n,n)) \arrow[r, "K_0(\mathcal{I}_n)\otimes 1"] & K_0^{\mathbb{Q}(q)}(\operatorname{Kar}\widehat{\mathcal{S}}(n+1,n))
\end{tikzcd}$$

where γ_{n+1} is the isomorphism from [Mackaay and Thiel 2013, Theorem 6.4] and $\mathcal{I}_n$ is defined in Definition 4.3. Since ι_n and γ_{n+1} are both injective, their composite is also injective. The commutativity of the diagram above then implies that γ_n is injective too. $\square$

Acknowledgements

We thank Jie Du and Qiang Fu for helpful exchanges of emails on the affine quantum Schur algebras. Additionally, we thank Ben Webster for sharing with us his ideas on a potentially interesting connection between our results in this paper and his work in [Webster 2012], the precise formulation of which remains to be worked out.

References

[Beilinson et al. 1990] A. A. Beilinson, G. Lusztig, and R. MacPherson, "A geometric setting for the quantum deformation of GL_n", *Duke Math. J.* **61**:2 (1990), 655–677. MR 91m:17012 Zbl 0713.17012

[Deng and Du 2013] B. Deng and J. Du, "Identification of simple representations for affine q-Schur algebras", *J. Algebra* **373** (2013), 249–275. MR 2995025 Zbl 06182885

[Deng et al. 2012] B. Deng, J. Du, and Q. Fu, *A double Hall algebra approach to affine quantum Schur–Weyl theory*, London Mathematical Society Lecture Note Series **401**, Cambridge Univ. Press, 2012. MR 3113018 Zbl 1269.20045

[Doty and Green 2007] S. R. Doty and R. M. Green, "Presenting affine q-Schur algebras", *Math. Z.* **256**:2 (2007), 311–345. MR 2008k:17014 Zbl 1148.17009

[Elias and Williamson 2013] B. Elias and G. Williamson, "Soergel calculus", preprint, 2013. arXiv 1309.0865

[Ginzburg and Vasserot 1993] V. Ginzburg and É. Vasserot, "Langlands reciprocity for affine quantum groups of type A_n", *Int. Math. Res. Not.* **1993**:3 (1993), 67–85. MR 94j:17011 Zbl 0785.17014

[Green 1999] R. M. Green, "The affine q-Schur algebra", *J. Algebra* **215**:2 (1999), 379–411. MR 2000b:20011 Zbl 0935.20003

[Khovanov and Lauda 2009] M. Khovanov and A. D. Lauda, "A diagrammatic approach to categorification of quantum groups, I", *Represent. Theory* **13** (2009), 309–347. MR 2010i:17023 Zbl 1188.81117

[Khovanov and Lauda 2010] M. Khovanov and A. D. Lauda, "A categorification of quantum $\mathrm{sl}(n)$", *Quantum Topol.* **1**:1 (2010), 1–92. MR 2011g:17028 Zbl 1206.17015

[Khovanov and Lauda 2011] M. Khovanov and A. D. Lauda, "A diagrammatic approach to categorification of quantum groups, II", *Trans. Amer. Math. Soc.* **363**:5 (2011), 2685–2700. MR 2012a:17021 Zbl 1214.81113

[Khovanov et al. 2012] M. Khovanov, A. D. Lauda, M. Mackaay, and M. Stošić, *Extended graphical calculus for categorified quantum* sl(2), vol. 219, Mem. Amer. Math. Soc. **1029**, 2012. MR 2963085 Zbl 1292.17013

[Libedinsky 2008] N. Libedinsky, "Équivalences entre conjectures de Soergel", *J. Algebra* **320**:7 (2008), 2695–2705. MR 2009h:20009 Zbl 1197.20004

[Lusztig 1999] G. Lusztig, "Aperiodicity in quantum affine $\mathfrak{gl}_n$", *Asian J. Math.* **3**:1 (1999), 147–177. MR 2000i:17027 Zbl 0983.17011

[Mackaay and Thiel 2013] M. Mackaay and A. L. Thiel, "Categorifications of the extended affine Hecke algebra and the affine quantum Schur algebra $\widehat{\mathbf{S}}(n,r)$ for $3 \leq r < n$", preprint, 2013. arXiv 1302.3102

[Mackaay et al. 2013] M. Mackaay, M. Stošić, and P. Vaz, "A diagrammatic categorification of the q-Schur algebra", *Quantum Topol.* **4**:1 (2013), 1–75. MR 2998837 Zbl 1272.81098

[Webster 2012] B. Webster, "Weighted Khovanov–Lauda–Rouquier algebras", preprint, 2012. arXiv 1209.2463

[Williamson 2011] G. Williamson, "Singular Soergel bimodules", *Int. Math. Res. Not.* **2011**:20 (2011), 4555–4632. MR 2844932 Zbl 1236.18009

Received July 22, 2014. Revised March 5, 2015.

MARCO MACKAAY
CENTER FOR MATHEMATICAL ANALYSIS, GEOMETRY, AND DYNAMICAL SYSTEMS
DEPARTAMENTO DE MATEMÁTICA
INSTITUTO SUPERIOR TÉCNICO
1049-001 LISBOA
PORTUGAL

and

DEPARTAMENTO DE MATEMÁTICA
FCT
UNIVERSIDADE DO ALGARVE
CAMPUS DE GAMBELAS
8005-139 FARO
PORTUGAL

mmackaay@ualg.pt

ANNE-LAURE THIEL
MATEMATISKA INSTITUTIONEN
UPPSALA UNIVERSITET
75106 UPPSALA
SWEDEN

anne-laure.thiel@math.uu.se

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.278.235

SHOWING DISTINCTNESS OF SURFACE LINKS BY TAKING 2-DIMENSIONAL BRAIDS

INASA NAKAMURA

For an oriented surface link S, we can take a satellite construction called a 2-dimensional braid over S, which is a surface link in the form of a covering over S. We demonstrate that 2-dimensional braids over surface links are useful for showing the distinctness of surface links. We investigate nontrivial examples of surface links with free abelian groups of rank two, concluding that their link types are infinitely many.

1. Introduction

A *surface link* is the image of a smooth embedding of a closed surface into Euclidean space $\mathbb{R}^4$. Two surface links are *equivalent* if there is an orientation-preserving self-diffeomorphism of $\mathbb{R}^4$ carrying one to the other. In this paper, we assume that surface links are oriented. In [Nakamura 2014a], we investigated a satellite construction called a 2-dimensional braid over an oriented surface link, and introduced its graphical presentation called an m-chart on a surface diagram. A 2-dimensional braid over a surface link S is a surface link in the form of a covering over S, and can be regarded as an analog to a double of a classical link. One of the expected applications of the notion of a 2-dimensional braid is that it will provide us with a method for showing the distinctness of surface links. The aim of this paper is to demonstrate such use for 2-dimensional braids.

Our main theorem is as follows. For a positive integer k, let $\sigma_1, \sigma_2, \dots, \sigma_k$ be the standard generators of the $(k+1)$-braid group. Take $X_1 = \sigma_1^2$ and $X_k = \sigma_1^2\sigma_2\sigma_3 \cdots \sigma_k$, and let Δ be a $(k+1)$-braid with a positive half twist. Consider $S_k = \mathcal{S}_{k+1}(X_k, \Delta^2)$, a torus-covering T^2-link determined from the $(k+1)$-braids X_k and Δ^2. We take the first component of S_k to be the one determined from the first strand of X_k, and likewise the second component from the second strand; see Section 2 for the construction. Here, a T^2*-link* is a surface link each of whose components is of genus one.

MSC2010: primary 57Q45; secondary 57Q35, 57M25.
Keywords: surface link, 2-dimensional braid, chart, Roseman move, triple linking.

Theorem 1.1. *Abelian T^2-links of rank two, S_k and S_l, are not equivalent for distinct positive integers k and l. Thus, the link types of abelian T^2-links of rank two are infinitely many.*

An *abelian surface link* of rank n is a surface link whose link group is a free abelian group of rank n [Ito and Nakamura 2014]; note that n is the number of the components. We remark that our abelian T^2-links of rank two cannot be distinguished by using link groups, and that by a homological argument we cannot show that their link types are infinitely many, but only that there are two link types; see Section 2B. Our abelian T^2-link S_k of rank two is a sublink of the surface link given in [Ito and Nakamura 2014], where we gave examples of abelian T^2-links of rank four, and we showed that their link types are infinitely many by calculations of triple linking numbers (see also Remark 2.3).

Triple linking numbers are integer-valued invariants of surface links with at least three components, so we cannot use them directly for our case. In order to overcome this situation, we take a 2-dimensional braid over S_k such that each component of S_k is split into two components. This has four components, so we can calculate triple linking numbers. A 2-dimensional braid over a surface link is obtained from the "standard" 2-dimensional braid by the addition of braiding information. Unfortunately, for the standard 2-dimensional braid, the triple linking number is trivial (Proposition 5.1). However, addition of braiding information makes a 2-dimensional braid with nontrivial triple linking, and enables us to show that S_k and S_l are not equivalent for distinct positive integers k and l. For a similar result, we refer to Suciu's paper [1985] where it is shown that there are infinitely many ribbon 2-knots in S^4 with knot group the trefoil knot group.

The paper is organized as follows. In Section 2, we review torus-covering links and explain our example S_k, and we review triple linking numbers of torus-covering links. In Section 3, we discuss the notion of a 2-dimensional braid over a surface link. In Section 4, we observe that a 2-dimensional braid of degree m over a surface link is presented by a finite graph called an m-chart on a surface diagram, and that 2-dimensional braids of degree m are equivalent if their surface diagrams with m-charts are related by local moves called Roseman moves. In Section 5, we prove Proposition 5.1. In Section 6, we calculate triple linking numbers of a certain 2-dimensional braid over S_k and prove Theorem 1.1.

2. Abelian T^2-links of rank two

The example S_k given in Theorem 1.1 is a surface link called a torus-covering link. In this section, we review torus-covering T^2-links; see [Nakamura 2011] for details. We briefly observe that S_k is an abelian surface link of rank two, and that we cannot show that the link types of our examples are infinitely many by using a homological argument. Further, we review a formula for the triple linking numbers of torus-covering links [Ito and Nakamura 2014].

2A. *Torus-covering links.* Let T be a standard torus in $\mathbb{R}^4$, the boundary of an unknotted (standardly embedded) solid torus in $\mathbb{R}^3 \times \{0\} \subset \mathbb{R}^4$.

Definition 2.1. A *torus-covering T^2-link S* is a surface link in the form of a 2-dimensional braid over the standard torus T, i.e., S is a T^2-link in $\mathbb{R}^4$ such that S is contained in a tubular neighborhood $N(T)$ and $\pi|_S : S \to T$ is an unbranched covering map, where $\pi : N(T) \to T$ is the natural projection.

Let S be a torus-covering T^2-link. Fix a base point $x_0 = (x_0', x_0'')$ of $T = S^1 \times S^1$. Take two simple closed curves on T, $\boldsymbol{m} = \partial B^2 \times \{x_0''\}$ and $\boldsymbol{l} = \{x_0'\} \times S^1$. Recall that T is embedded as $T = \partial(B^2 \times S^1) \subset \mathbb{R}^3 \times \{0\} \subset \mathbb{R}^4$. Let us consider the intersections $S \cap \pi^{-1}(\boldsymbol{m}) \subset B^2 \times \boldsymbol{m}$ and $S \cap \pi^{-1}(\boldsymbol{l}) \subset B^2 \times \boldsymbol{l}$. They are regarded as closed m-braids in the 3-dimensional solid tori, where m is the degree of the covering map $\pi|_S : S \to T$. Cutting open the solid tori along the 2-disk $\pi^{-1}(x_0) = B^2 \times \{x_0\}$, we obtain two m-braids a and b. The assumption that $\pi|_S$ is an unbranched covering implies that a and b commute. We call the commutative braids (a, b) the *basis braids* of S. Conversely, starting from a pair of commutative m-braids (a, b), we can uniquely construct a torus-covering T^2-link with basis braids (a, b) [Nakamura 2011, Lemma 2.8]. For commutative m-braids a and b, we denote by $\mathcal{S}_m(a, b)$ the torus-covering T^2-link with basis braids (a, b).

2B. *Our abelian T^2-links of rank two.* We can verify that $S_k = \mathcal{S}_{k+1}(X_k, \Delta^2)$ is an abelian surface link as follows. The link group of a torus-covering link $\mathcal{S}_m(a, b)$ is a quotient group of the classical link group of the closure of a such that the abelianization is a free abelian group [Nakamura 2011, Proposition 3.1]. Since the link group of the closure of X_k, a Hopf link, is a free abelian group of rank two, so is the link group of S_k.

We remark that a homological argument cannot show that our examples are infinitely many, but only that there are two link types. Let us consider the one-point compactification of $\mathbb{R}^4$, and regard S_k to be in the Euclidean 4-sphere S^4. Recall that we take the first and second components F_1 and F_2 of S_k to be determined from the first and second strands of X_k, respectively. By Alexander's duality, we see that $H_2(S^4 - F_1; \mathbb{Z}) \cong H_1(F_1; \mathbb{Z})$, whence $[F_2] = \mu + k\lambda \in H_2(S^4 - F_1; \mathbb{Z})$, where (μ, λ) is a preferred basis of $H_1(F_1; \mathbb{Z}) \cong H_2(S^4 - F_1; \mathbb{Z})$ represented by a meridian and a preferred longitude of F_1. Similarly, let us denote by F_1' and F_2' the first and second components of S_l. Then we can see that $[F_2'] = \mu' + l\lambda' \in H_2(S^4 - F_1'; \mathbb{Z})$, where (μ', λ') is a preferred basis of $H_1(F_1'; \mathbb{Z}) \cong H_2(S^4 - F_1'; \mathbb{Z})$ represented by a meridian and a preferred longitude of F_1'. Now, the standardly embedded tori F_1 and F_1' are related by an orientation-preserving self-diffeomorphism of S^4 if and only if

$$\begin{pmatrix} \mu' \\ \lambda' \end{pmatrix} = A \begin{pmatrix} \mu \\ \lambda \end{pmatrix}$$

for

$$A = \begin{pmatrix} \alpha & \beta \\ \delta & \gamma \end{pmatrix} \in GL_+(2;\mathbb{Z})$$

such that $\alpha+\beta+\gamma+\delta \equiv 0 \pmod 2$ [Montesinos 1983], which implies that $[F_2]=[F_2'] \in H_2(S^4 - F_1;\mathbb{Z})$ if and only if $k \equiv l \pmod 2$ (see [Iwase 1988]).

Remark 2.2. The abelian surface link S_1, i.e., $\mathcal{S}_2(\sigma_1^2, \sigma_1^2)$, is the twisted Hopf 2-link we will mention in the proof of Proposition 5.1; see also [Carter et al. 2001].

Remark 2.3. It is known [Kawauchi 1996, Theorem 6.3.1–Exercise 6.3.3] that for classical links, the rank of an abelian link is at most two, and for abelian links of rank two, there are exactly two link types: a positive and a negative Hopf link.

Remark 2.4. Set $T_m = \mathcal{S}_{k+1}(X_k, X_k^m)$ for an integer m. It is known ([Boyle 1993], see also [Iwase 1988; Nakamura 2011]) that T_m and T_n are equivalent for $m \equiv n \pmod 2$. Fix the first component of T_m in the form of the standard torus. By a homological argument as above, we see that T_m cannot be taken to T_n for $n \neq m$ by an orientation-preserving self-diffeomorphism of $\mathbb{R}^4$ relative to the first component.

2C. ***Triple linking numbers of torus-covering links.*** The triple linking number of a surface link S is defined as follows [Carter et al. 2003, Definition 9.1]. For the i-th, j-th, and k-th components F_i, F_j, F_k of S with $i \neq j$ and $j \neq k$, the *triple linking number* $\mathrm{tlk}_{i,j,k}(S)$ of the i-th, j-th, and k-th components of S is the total number of positive triple points minus the total number of negative triple points of a surface diagram of S such that the top, middle, and bottom sheet are from F_i, F_j, and F_k, respectively. The triple linking number is a link bordism invariant [Carter et al. 2004; 2001; Sanderson 1987; 1993]; for other properties, see [Carter et al. 2003; 2004]. Triple linking numbers are useful for showing the distinctness of surface links with at least three components [Ito and Nakamura 2014; Nakamura 2012; 2014b].

From [Ito and Nakamura 2014], we have a formula for the triple linking numbers of a torus-covering T^2-link $\mathcal{S}_m(a,b)$. Let A_i be the components of the closure of a which are from the i-th component of $\mathcal{S}_m(a,b)$. Take one of the connected components of A_i and denote it by A_i^1. We define by the classical linking number

$$\mathrm{lk}_{i,j}^a = \mathrm{lk}(A_i^1, A_j),$$

where we regard A_i^1 and A_j as oriented links in $\mathbb{R}^3$. The notation $\mathrm{lk}_{i,j}^b$ for the other basis braid is defined similarly. Note that $\mathrm{lk}_{i,j}^a$ does not depend on the choice of a connected component A_i^1 [Ito and Nakamura 2014, Remark 5.5], and note that $\mathrm{lk}^a_{i,j}$ is not always symmetric, i.e., $\mathrm{lk}^a_{i,j}$ is not always equal to $\mathrm{lk}^a_{j,i}$.

Now, for a torus-covering T^2-link, the triple linking number of the i-th, j-th, and k-th components is given by

$$\mathrm{tlk}_{i,j,k}(\mathcal{S}_m(a,b)) = -\mathrm{lk}^a_{j,i}\,\mathrm{lk}^b_{j,k} + \mathrm{lk}^a_{j,k}\,\mathrm{lk}^b_{j,i}, \tag{2-1}$$

where $i \neq k$ and $j \neq k$ [Ito and Nakamura 2014, Theorem 5.4 and Remark 5.7].

3. Two-dimensional braids over a surface link

A 2-dimensional braid, also called a simple braided surface, over a 2-disk is analogous to a classical braid [Kamada 1992; 2002; Rudolph 1983]. We can modify this notion to a 2-dimensional braid over a closed surface [Nakamura 2011], and further to a 2-dimensional braid over a surface link [Carter et al. 2004, Section 2.4.2; Nakamura 2014a]. In this section, we review this notion of a 2-dimensional braid over a surface link.

3A. ***Two-dimensional braids over a surface link.*** We use 2-dimensional braids without branch points over a closed surface, so our definition here is restricted to such surfaces; see [Nakamura 2011; 2014a] for the general definition.

Let Σ be a closed surface, let B^2 be a 2-disk, and let m be a positive integer.

Definition 3.1. A closed surface $\widetilde{\Sigma}$ embedded in $B^2 \times \Sigma$ is called a *2-dimensional braid over* Σ *of degree* m if the restriction $\pi|_{\widetilde{\Sigma}} : \widetilde{\Sigma} \to \Sigma$ is an unbranched covering map of degree m, where $\pi : B^2 \times \Sigma \to \Sigma$ is the natural projection.

Take a base point x_0 of Σ. Two 2-dimensional braids over Σ of degree m are *equivalent* if there is a fiber-preserving ambient isotopy of $B^2 \times \Sigma$ rel $\pi^{-1}(x_0)$ which carries one to the other.

A surface link is said to be *of type* Σ when it is the image of an embedding of Σ. Let S be a surface link of type Σ, and let $N(S)$ be a tubular neighborhood of S in $\mathbb{R}^4$.

Definition 3.2. A *2-dimensional braid* $\widetilde{S}$ *over* S is the image of a 2-dimensional braid over Σ in $B^2 \times \Sigma$ by an embedding $B^2 \times \Sigma \to \mathbb{R}^4$ which identifies $N(S)$ with $B^2 \times \Sigma$ as a B^2-bundle over a surface. We define the *degree* of $\widetilde{S}$ as that of S.

Two 2-dimensional braids $\widetilde{S}$ and $\widetilde{S}'$ over surface links S and S' are *equivalent* if there is an ambient isotopy of $\mathbb{R}^4$ carrying $\widetilde{S}$ to $\widetilde{S}'$ and $N(S) = B^2 \times S$ to $N(S') = B^2 \times S'$ as a B^2-bundle over a surface.

Equivalent 2-dimensional braids over surface links are also equivalent as surface links. A 2-dimensional braid $\widetilde{S}$ over S is a specific satellite with companion S; see [Carter et al. 2004, Section 2.4.2] as well as [Lickorish 1997, Chapter 1].

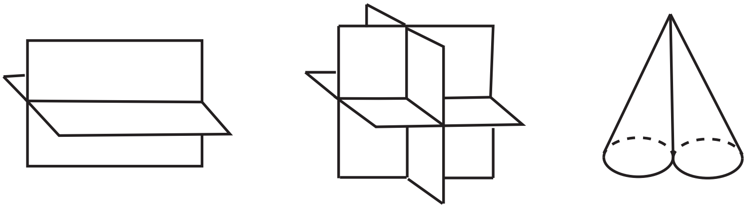


Figure 1. The singularity of a surface diagram.

3B. ***Standard 2-dimensional braids.*** In this section, we define the standard 2-dimensional braid over a surface link S. Using this notion, we will explain in the next section that a 2-dimensional braid is presented by a finite graph called an m-chart on a surface diagram D of S. The standard 2-dimensional braid over S is the 2-dimensional braid presented by an empty m-chart on D [Nakamura 2014a].

We first review surface diagrams of a surface link S; see [Carter et al. 2004]. For a projection $p : \mathbb{R}^4 \to \mathbb{R}^3$, the closure of the self-intersection set of $p(S)$ is called the singularity set. Let p be a generic projection, meaning that the singularity set of the image $p(S)$ consists of double points, isolated triple points, and isolated branch points; see Figure 1. The closure of the singularity set forms a union of immersed arcs and loops, called double point curves. Triple points form the intersection points of the double point curves, and branch points form the end points. A *surface diagram* of S is the image $p(S)$ equipped with over/under information along each double point curve with respect to the projection direction.

We define the $2m$-braid $\widetilde{\sigma}_1$ obtained from a 2-braid σ_1 as follows. More generally, we construct an mn-braid $\tilde{b}$ from an n-braid b, needed for the proof of Theorem 1.1. Let Q_m be m interior points of B^2. For a standard generator σ_i of an n-braid, let $\widetilde{\sigma}_i$ be the mn-braid obtained from σ_i in the form of a Q_m-bundle over σ_i by splitting each strand into a bundle of m parallel strands with a negative half twist at the initial points of each bundle; see Figure 2. The map taking σ_i to $\widetilde{\sigma}_i$ determines a homomorphism from the n-braid group to the mn-braid group. For an n-braid b, let $\tilde{b}$ denote the image of b by this homomorphism.

Definition 3.3. Let S be a surface link. A surface diagram D of S consists of the following local parts: around (1) a regular point, i.e., a nonsingular point,

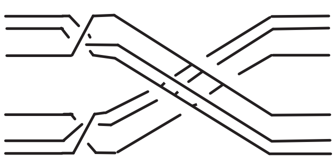


Figure 2. The $2m$-braid $\widetilde{\sigma}_1$.

(2) a double point curve, (3) a triple point, and (4) a branch point. The diagram around a regular point (1) consists of an embedded 2-disk B^2 with no singularities, and the diagram around a double-point curve (2) can be expressed as the product of a 2-braid σ_1 and an interval I.

We define the *standard 2-dimensional braid over S* locally for such local parts of D as follows: for (1), it is m parallel copies of B^2, and for (2), it is the product of the $2m$-braid $\widetilde{\sigma}_1$ and I. Then, for the other cases (3) and (4), the standard 2-dimensional braid is naturally defined [Nakamura 2014a, Definition 5.1 and Proposition 5.2].

4. Chart presentation of 2-dimensional braids and Roseman moves

In this section, we recall that a 2-dimensional braid of degree m over a surface link S is presented by a finite graph called an m-chart on a surface diagram D of S. For two 2-dimensional braids of degree m, they are equivalent if their surface diagrams with m-charts are related by a finite sequence of local moves called Roseman moves. See [Nakamura 2014a].

4A. ***Chart presentation of 2-dimensional braids over a surface link.*** The graphical method called an m-chart on a 2-disk was introduced to present a simple surface braid which is a 2-dimensional braid over a 2-disk with trivial boundary condition [Kamada 1992; 2002]. By regarding an m-chart on a 2-disk as drawn on a 2-sphere S^2, it presents a 2-dimensional braid over S^2 [Kamada 1992; 2002; Nakamura 2011]. This notion can be modified to an m-chart on a closed surface, and further to an m-chart on a surface diagram D of a surface link S [Nakamura 2011; 2014a]. A 2-dimensional braid over S is presented by an m-chart on D [Nakamura 2014a].

In this paper, we treat 2-charts with vertices of degree 2. We now review the graphical form of an m-chart of a 2-dimensional braid over a surface link. See [Nakamura 2014a] for details.

Let $\widetilde{S}$ be a 2-dimensional braid over a surface link S. Let D be a surface diagram of S by a projection $p: \mathbb{R}^4 \to \mathbb{R}^3$ which is generic with respect to both S and $\widetilde{S}$. We can assume that the singularity set of the surface diagram $p(\widetilde{S})$ is the union of the singularity set of the diagram of the standard 2-dimensional braid over S and some finite graph Γ [Nakamura 2014a, Theorem 5.5]. Project Γ to D by the projection $p(N(S)) = B^2 \times D \to D$. Thus we obtain a finite graph on the surface diagram D. An m-chart on a surface diagram D is such a finite graph equipped with certain additional information of orientations and labels assigned to the edges, where m is the degree of the 2-dimensional braid. Owing to the additional information, we can regain the original 2-dimensional braid from the m-chart on D [Nakamura 2014a] (see also [Kamada 2002]).

We can define an m-chart on D in graphical terms, where the labels of edges are from 1 to $m-1$; see [Nakamura 2014a, Definitions 5.3 and 5.4]. Around a double

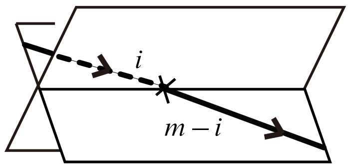


Figure 3. An m-chart around a double point curve. Here i is one of $1, \ldots, m-1$. For simplicity, we omit the over/under information of each sheet.

point curve, an m-chart is as in Figure 3, with a vertex of degree 2. A 2-dimensional braid over S is presented by an m-chart on D [Nakamura 2014a, Theorem 5.5].

4B. ***Roseman moves.*** Roseman moves are local moves of surface diagrams, as illustrated in Figure 4. It is known [Roseman 1998] that two surface links are equivalent if and only if their surface diagrams are related by a finite sequence of Roseman moves and ambient isotopies of the diagrams in $\mathbb{R}^3$. In [Nakamura 2014a], we introduced the notion of Roseman moves for surface diagrams with m-charts.

An m-chart is said to be *empty* if it is an empty graph.

Definition 4.1. We define *Roseman moves for surface diagrams with m-charts* by the local moves as shown in Figures 4 and 5, where we regard the diagrams in Figure 4 as equipped with empty m-charts.

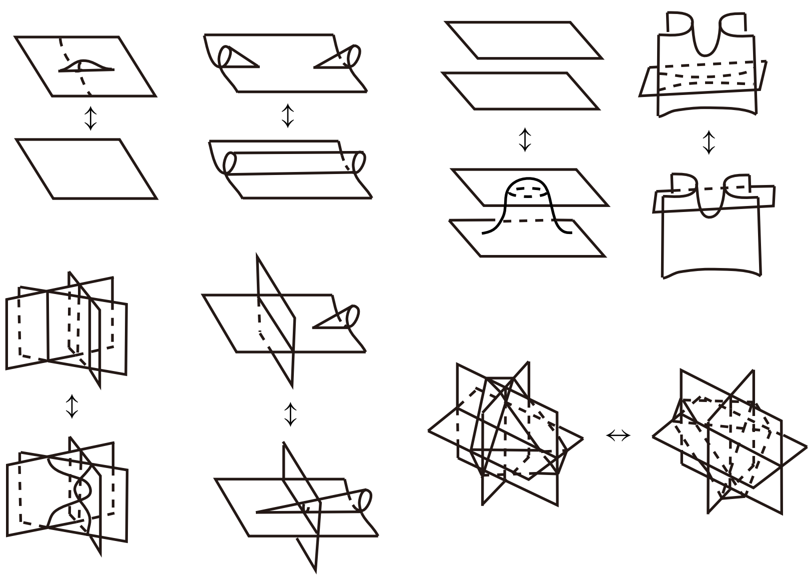


Figure 4. Roseman moves. We omit the over/under information of each sheet.

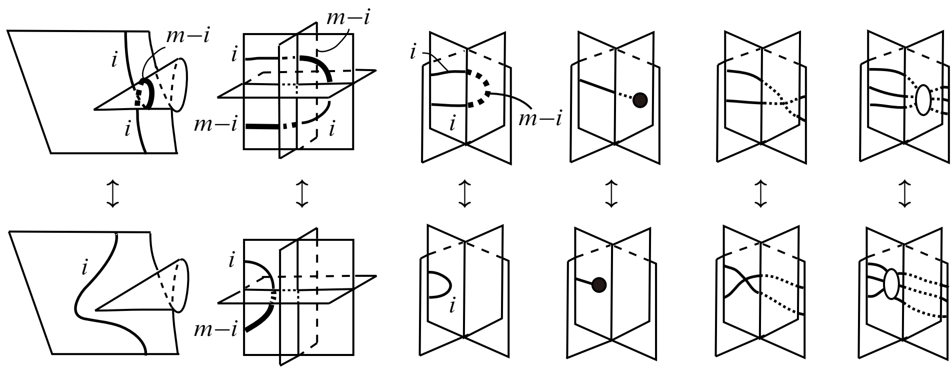


Figure 5. Roseman moves for surface diagrams with m-charts, where $i \in \{1, \dots, m-1\}$. We omit the over/under information of each sheet, and orientations and labels of edges of m-charts.

Roseman moves for surface diagrams with m-charts, as illustrated in Figures 4 and 5, are well-defined. That is, for each pair of Roseman moves, the m-charts on the indicated diagrams present equivalent 2-dimensional braids [Nakamura 2014a, Theorem 6.2].

5. Triple linking numbers of standard 2-dimensional braids

Recall the triple linking numbers from Section 2C. We will say that a surface link S has *trivial* triple linking if every triple linking number of S is zero or S consists of less than three components.

Proposition 5.1. *For the standard* 2*-dimensional braid* $\widetilde{S}$ *over a surface link* S, *if* S *has trivial triple linking*, *then so does* $\widetilde{S}$.

Proof. Assume that S has trivial triple linking. Recall from [Carter et al. 2001] that the link bordism class of a surface link is determined from triple linking numbers and double linking numbers (another kind of link bordism invariant), and that a surface link with trivial triple linking is link bordant to a split union of a finite number of trivial spheres and surface links called twisted Hopf 2-links, which has a surface diagram with no triple points (see also Remark 2.2). Hence, S is link bordant to a surface link S' whose surface diagram has no triple points. By the well-definedness of Roseman moves, $\widetilde{S}$ is link bordant to the standard 2-dimensional braid $\widetilde{S}'$ over S'. Since the surface diagram of a standard 2-dimensional braid has triple points only around triple points of the companion surface [Nakamura 2014a], the surface diagram of $\widetilde{S}'$ has no triple points. Thus $\widetilde{S}$ is link bordant to a surface link with no triple points, which implies that $\widetilde{S}$ has trivial triple linking. □

6. Proof of Theorem 1.1

In this section, we consider a 2-dimensional braid $\widetilde{S}$ over a surface link S presented by a 2-chart consisting of a finite number of loops on a surface diagram of S. Here, a *loop* is a union of edges connected by vertices of degree 2 as in Figure 3. For our 2-charts, the edges are labeled by 1 and the orientations are coherent around a vertex of degree 2, so we can ignore the label information, and we regard the 2-chart on a surface diagram of S as oriented loops. Further, we consider that the loops are on S itself. By the well-definedness of Roseman moves, a 2-dimensional braid presented by a 2-chart Γ on S is equivalent to the 2-dimensional braid presented by a 2-chart $f(\Gamma)$ on $f(S)$ for an orientation-preserving self-diffeomorphism f of $\mathbb{R}^4$.

For a component F of a torus-covering T^2-link, we take a preferred basis of $H_1(F;\mathbb{Z})$ represented by a pair of simple closed curves (μ,λ) such that μ is a connected component of $F \cap \pi^{-1}(\boldsymbol{m})$, and λ of $F\cap\pi^{-1}(\boldsymbol{l})$. Recall that $\pi : N(T) \to T$ is the natural projection for a standard torus T, and $\boldsymbol{m}$ and $\boldsymbol{l}$ are simple closed curves on T given in Section 2A. We will use the same notation (μ,λ) for the preferred basis, and we call simple closed curves in the homology classes μ and λ *meridians* and *preferred longitudes* of F, respectively. For a 2-chart Γ on F consisting of loops, we can assume that the intersections of the chart loops of Γ with a meridian μ and a preferred longitude λ of F are transverse. We assign each intersection point the sign $+1$ or -1 according to whether it presents a positive or negative crossing, and we denote by $I(\mu,\Gamma)$ and $I(\lambda,\Gamma)$ the sum of the signs of the intersection points of Γ with μ and λ. Note that we can assume that the chart loops are parallel by using local moves of charts called CI-moves of type (1) [Kamada 2002], and $I(\mu,\Gamma)$ and $I(\lambda,\Gamma)$ are well-defined for the homology classes μ and λ.

For the torus-covering T^2-link S, we take the first and second components of S as those determined from the first and second strands of each basis braid of S, respectively. Similarly, for the 2-dimensional braid $\widetilde{S}$, we take the i-th component of $\widetilde{S}$ as the one determined from the i-th strand of each basis braid of $\widetilde{S}$ for $i=1,2,3,4$.

We first calculate the triple linking numbers of a 2-dimensional braid of degree 2 over S_k as in Theorem 1.1.

Lemma 6.1. *For the torus-covering T^2-link S_k for a positive integer k, let us consider a 2-dimensional braid $\widetilde{S}_k$ of degree 2 over S_k, which is presented by a 2-chart Γ consisting of loops on S_k so that it consists of 4 components. Then* $\mathrm{tlk}_{i,j,3}(\widetilde{S}_k)=\mathrm{tlk}_{i,j,4}(\widetilde{S}_k)$ *for* $(i,j)=(1,2)$ *or* $(2,1)$, *and* $\mathrm{tlk}_{i,j,1}(\widetilde{S}_k)=\mathrm{tlk}_{i,j,2}(\widetilde{S}_k)$ *for* $(i,j)=(3,4)$ *or* $(4,3)$.

Proof. The 2-dimensional braid $\widetilde{S}_k$ is also a torus-covering T^2-link. We denote by (a,b) the basis braids presenting $\widetilde{S}_k$. Since $\mathrm{lk}^c_{j,3}=\mathrm{lk}^c_{j,4}$ for $j=2,1$, and $\mathrm{lk}^c_{j,1}=\mathrm{lk}^c_{j,2}$ for $j=4,3$ $(c=a,b)$, by (2-1) we have the result. □

Lemma 6.2. *For the torus-covering T^2-link S_k, denote by F_1 and F_2 the first and second components of S_k. Let (μ_i, λ_i) be a preferred basis of $H_1(F_i;\mathbb{Z})$ for $i = 1, 2$. Consider a 2-dimensional braid $\widetilde{S}_k$, as in Lemma 6.1, such that $I(\mu_i, \Gamma) = 2p_i$ and $I(\lambda_i, \Gamma) = 2q_i$ for integers p_i and q_i for $i = 1, 2$. Then we have* $\mathrm{tlk}_{1,2,3}(\widetilde{S}_k) = -kp_1 + q_1$ *and* $\mathrm{tlk}_{2,3,4}(\widetilde{S}_k) = -p_2 + q_2$.

Note that $\widetilde{S}_k$ consists of 4 components if and only if $I(\mu_i, \Gamma)$ and $I(\lambda_i, \Gamma)$ are even for $i = 1, 2$, since these conditions are equivalent to the condition that $\widetilde{S}_k \cap \pi_i^{-1}(\mu)$ and $\widetilde{S}_k \cap \pi_i^{-1}(\lambda)$ are closed pure braids, where $\pi_i : N(F_i) \to F_i$ is the natural projection.

Proof. The 2-dimensional braid $\widetilde{S}_k$ is also a torus-covering T^2-link. We denote by (a, b) the basis braids presenting $\widetilde{S}_k$. We use the notation given in Section 3B, taking $m = 2$ and $n = k + 1$. Then, $\mathrm{lk}^a_{2,1}$ is determined from the linking number coming from the linking consisting of $I(\mu_1, \Gamma)$ crossings and $\widetilde{X}_k$. That is,

$$\mathrm{lk}^a_{2,1} = p_1 + \mathrm{lk}^{\widetilde{X}_k}_{2,1},$$

and similarly,

$$\mathrm{lk}^b_{2,1} = q_1 + \mathrm{lk}^{\widetilde{\Delta}^2}_{2,1}.$$

By definition, for a braid c, the braid $\tilde{c}$ has a negative (respectively positive) half twist at the place which is a fiber of a point of each arc forming a positive (respectively negative) crossing of c. Hence,

$$\mathrm{lk}^{\widetilde{X}_k}_{2,1} = -\mathrm{lk}^{X_k}_{1,2}$$
$$\mathrm{lk}^{\widetilde{\Delta}^2}_{2,1} = -\mathrm{lk}^{\Delta^2}_{1,2},$$

from which it follows that

$$\mathrm{lk}^a_{2,1} = p_1 - \mathrm{lk}^{X_k}_{1,2}$$
$$\mathrm{lk}^b_{2,1} = q_1 - \mathrm{lk}^{\Delta^2}_{1,2}.$$

Further, $\mathrm{lk}^a_{2,3} = \mathrm{lk}^{X_k}_{1,2}$ and $\mathrm{lk}^b_{2,3} = \mathrm{lk}^{\Delta^2}_{1,2}$. Thus $\mathrm{tlk}_{1,2,3}(\widetilde{S}_k) = -p_1 \mathrm{lk}^{\Delta^2}_{1,2} + q_1 \mathrm{lk}^{X_k}_{1,2}$ by (2-1). Since $\mathrm{lk}^{X_k}_{1,2}$ is the linking number of the closure of X_k, $\mathrm{lk}^{X_k}_{1,2} = 1$. Since F_1 and F_2 are constructed by one strand and k strands of Δ^2, respectively, we have $\mathrm{lk}^{\Delta^2}_{1,2} = k$. Thus $\mathrm{tlk}_{1,2,3}(\widetilde{S}_k) = -kp_1 + q_1$.

By the same argument, we have $\mathrm{tlk}_{2,3,4}(\widetilde{S}_k) = -p_2 \,\mathrm{lk}^{\Delta^2}_{2,1} + q_2 \,\mathrm{lk}^{X_k}_{2,1}$ by (2-1), and $\mathrm{lk}^{X_k}_{2,1} = 1$. Since Δ^2 is a pure braid, we see that $\mathrm{lk}^{\Delta^2}_{2,1} = 1$. Thus $\mathrm{tlk}_{2,3,4}(\widetilde{S}_k) = -p_2 + q_2$. □

Proof of Theorem 1.1. Let k and l be positive integers. We denote by F_1 and F_2 the first and second components of S_k, and by F'_1 and F'_2 the first and second components of S_l.

First we show that for $k \neq l$, there does not exist an orientation-preserving self-diffeomorphism of $\mathbb{R}^4$ carrying F_1 to F'_1 and F_2 to F'_2. Assume that there is such a diffeomorphism f. Let us consider a 2-dimensional braid over S_k, denoted by $\widetilde{S}_k^1$, which is presented by a 2-chart Γ on S_k such that $\Gamma \cap F_1$ consists of loops with $I(\mu_1, \Gamma) = 2p$ and $I(\lambda_1, \Gamma) = 2q$, where (μ_1, λ_1) is a preferred basis of $H_1(F_1; \mathbb{Z})$, and $\Gamma \cap F_2 = \varnothing$. Note that $\widetilde{S}_k^1$ consists of 4 components.

Since f is an orientation-preserving diffeomorphism which carries F_1 to F'_1, $f|_{F_1}$ is an orientation-preserving diffeomorphism from a torus F_1 to a torus F'_1. Let

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{GL}_+(2, \mathbb{Z})$$

be a matrix determined by

$$\begin{pmatrix} \mu'_1 \\ \lambda'_1 \end{pmatrix} = A \begin{pmatrix} f_*(\mu_1) \\ f_*(\lambda_1) \end{pmatrix}, \tag{6-1}$$

where (μ'_1, λ'_1) is a preferred basis of $H_1(F'_1; \mathbb{Z})$.

Put $\Gamma' = f(\Gamma)$. By f, $\widetilde{S}_k^1$ is taken to a 2-dimensional braid $\widetilde{S}_l^1$ over S_l, presented by a 2-chart Γ' on S_l such that $\Gamma' \cap F'_1$ consists of loops and $\Gamma' \cap F'_2 = \varnothing$. We see that $I(f_*(\mu_1), \Gamma') = I(\mu_1, \Gamma) = 2p$, and $I(f_*(\lambda_1), \Gamma') = I(\lambda_1, \Gamma) = 2q$. Set $p' = I(\mu'_1, \Gamma')/2$ and $q' = I(\lambda'_1, \Gamma')/2$; note that p' and q' are integers, since $\widetilde{S}_l^1$ consists of 4 components. It follows from (6-1) that

$$\begin{pmatrix} p' \\ q' \end{pmatrix} = A \begin{pmatrix} p \\ q \end{pmatrix}. \tag{6-2}$$

Since the triple linking numbers $\mathrm{tlk}_{1,2,3}$ for $\widetilde{S}_k^1$ and $\widetilde{S}_l^1$ are the same, Lemma 6.2 implies that

$$-kp + q = -lp' + q'. \tag{6-3}$$

Hence, it follows from (6-2) that $kp - q = (\alpha l - \gamma)p + (\beta l - \delta)q$. Since this equation holds true for any integers p and q,

$$\begin{pmatrix} k \\ -1 \end{pmatrix} = A^T \begin{pmatrix} l \\ -1 \end{pmatrix}, \tag{6-4}$$

where A^T is the transposed matrix of A.

Next we consider another 2-dimensional braid over S_k, denoted by $\widetilde{S}_k^2$, presented by a 2-chart $\widetilde{\Gamma}$ on S_k such that $\widetilde{\Gamma} \cap F_1 = \varnothing$ and $\widetilde{\Gamma} \cap F_2$ consists of loops on F_2, and moreover that $\widetilde{\Gamma} \cap F_2$ is the preimage by the projection $N(T) \to T$ of a 2-chart Γ on the standard torus T consisting of loops with $I(\boldsymbol{m}, \Gamma) = 2p$ and $I(\boldsymbol{l}, \Gamma) = 2q$, where $(\boldsymbol{m}, \boldsymbol{l})$ is a preferred basis of T. Note that $I(\mu_2, \widetilde{\Gamma}) = 2kp$ and $I(\lambda_2, \widetilde{\Gamma}) = 2q$, where (μ_2, λ_2) is a preferred basis of $H_1(F_2; \mathbb{Z})$.

Let g be an orientation-preserving diffeomorphism of $\mathbb{R}^4$ which carries F_2 sufficiently close to F_1 and $(g|_{F_i})_* = \mathrm{id}_* : H_1(F_i; \mathbb{Z}) \to g_*(H_1(F_i); \mathbb{Z})$ for $i = 1, 2$. Further, we assume that T is sufficiently close to F_1. Then

$$\begin{pmatrix} \boldsymbol{m}' \\ \boldsymbol{l}' \end{pmatrix} = A \begin{pmatrix} (f \circ g)_*(\boldsymbol{m}) \\ (f \circ g)_*(\boldsymbol{l}) \end{pmatrix},$$

where $(\boldsymbol{m}', \boldsymbol{l}')$ is a preferred basis of $T' = (f \circ g)(T)$. Put $\Gamma' = (f \circ g)(\Gamma)$. Then we have

$$\begin{pmatrix} I(\boldsymbol{m}', \Gamma') \\ I(\boldsymbol{l}', \Gamma') \end{pmatrix} = A \begin{pmatrix} I(\boldsymbol{m}, \Gamma) \\ I(\boldsymbol{l}, \Gamma) \end{pmatrix}. \tag{6-5}$$

Put $S' = (f \circ g)(S_k)$. The surface link S' is in the form of a 2-dimensional braid over T' of degree $k+1$. For the natural projection $\pi' : N(T') = (f \circ g)(N(T)) \to T'$, a meridian $\boldsymbol{m}'$, and a preferred longitude $\boldsymbol{l}'$ of T', let us consider $S' \cap \pi'^{-1}(\boldsymbol{m}')$ and $S' \cap \pi'^{-1}(\boldsymbol{l}')$, which are closed $(k+1)$-braids in the 3-dimensional solid tori. In the same way as obtaining basis braids, we obtain $(k+1)$-braids from the closed braids by cutting open the solid tori along the 2-disk $\pi'^{-1}(x_0')$, where x_0' is the intersection point of $\boldsymbol{m}'$ and $\boldsymbol{l}'$. We denote the braids by a and b. Note that here T' is a standard torus, and hence (a, b) are basis braids, but we can apply the same argument if T' is not a standard torus. Since S' consists of two components, a and b satisfy one of the following three cases.

Case 1: The closure of a is a link consisting of two components, and b is a pure braid.

Case 2: Each of the closures of a and b is a link consisting of two components.

Case 3: The braid a is a pure braid, and the closure of b is a link consisting of two components.

Set $\widetilde{\Gamma}' = (f \circ g)(\widetilde{\Gamma})$. By $f \circ g$, $\widetilde{S}_k^2$ is taken to a 2-dimensional braid $\widetilde{S}'$ presented by a 2-chart $\widetilde{\Gamma}'$ on S'. We denote by F' the component $(f \circ g)(F_2)$ of S', and we denote by (μ', λ') a preferred basis of $H_1(F'; \mathbb{Z})$. Since $\widetilde{\Gamma} \cap F_2$ is in the form of the preimage by $N(T) \to T$ of the 2-chart Γ on T, $\widetilde{\Gamma}' \cap F'$ is in the form of the preimage by $N(T') \to T'$ of the 2-chart Γ' on T'. Hence, $I(\mu', \widetilde{\Gamma}') = i \cdot I(\boldsymbol{m}', \Gamma')$

and $I(\lambda', \widetilde{\Gamma}') = j \cdot I(\boldsymbol{l}', \Gamma')$ for $(i, j) = (k, 1)$ for Case 1, (k, k) for Case 2, and $(1, k)$ for Case 3. Thus

$$\begin{pmatrix} I(\mu', \widetilde{\Gamma}') \\ I(\lambda', \widetilde{\Gamma}') \end{pmatrix} = B \begin{pmatrix} I(\boldsymbol{m}', \Gamma') \\ I(\boldsymbol{l}', \Gamma') \end{pmatrix}, \tag{6-6}$$

where B is a diagonal matrix $\operatorname{diag}(i, j)$ such that $(i, j) = (k, 1)$ for Case 1, (k, k) for Case 2, and $(1, k)$ for Case 3.

Put $h = f \circ (f \circ g)^{-1}$. Then h is an orientation-preserving self-diffeomorphism of $\mathbb{R}^4$ which carries S' to S_l. In particular, h carries F' to the second component F_2' of S_l. Let

$$C = \begin{pmatrix} \alpha' & \beta' \\ \gamma' & \delta' \end{pmatrix} \in \mathrm{GL}_+(2, \mathbb{Z})$$

be a matrix determined by

$$\begin{pmatrix} \mu_2' \\ \lambda_2' \end{pmatrix} = C \begin{pmatrix} h_*(\mu') \\ h_*(\lambda') \end{pmatrix},$$

where (μ_2', λ_2') is a preferred basis of $H_1(F_2'; \mathbb{Z})$. Put $\Gamma'' = h(\widetilde{\Gamma}')$. Then

$$\begin{pmatrix} I(\mu_2', \Gamma'') \\ I(\lambda_2', \Gamma'') \end{pmatrix} = C \begin{pmatrix} I(\mu', \widetilde{\Gamma}') \\ I(\lambda', \widetilde{\Gamma}') \end{pmatrix}. \tag{6-7}$$

Set $p'' = I(\mu_2', \Gamma'')/2$ and $q'' = I(\lambda_2', \Gamma'')/2$, which are both integers. Since $I(\boldsymbol{m}, \Gamma) = 2p$ and $I(\boldsymbol{l}, \Gamma) = 2q$, together with (6-5)–(6-7), we have

$$\begin{pmatrix} p'' \\ q'' \end{pmatrix} = (CBA) \begin{pmatrix} p \\ q \end{pmatrix}. \tag{6-8}$$

By the composite diffeomorphism $h \circ f \circ g = f$, $\widetilde{S}_k^2$ is taken to a 2-dimensional braid over S_l, which will be denoted by $\widetilde{S}_l^2$. Since $\mathrm{tlk}_{2,3,4}$ is the same for $\widetilde{S}_k^2$ and $\widetilde{S}_l^2$, together with $I(\mu_2, \widetilde{\Gamma}) = 2kp$ and $I(\lambda_2, \widetilde{\Gamma}) = 2q$, Lemma 6.2 implies that

$$-kp + q = -p'' + q''. \tag{6-9}$$

Since this equation holds true for any integers p and q, it follows from (6-8) that

$$\begin{pmatrix} k \\ -1 \end{pmatrix} = (CBA)^T \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Thus, together with (6-4),

$$B^T C^T \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} l \\ -1 \end{pmatrix},$$

whence $i(\alpha'-\gamma')=l$ and $j(\beta'-\delta')=-1$. Let us assume $k>l>0$. For Cases 1 and 2, $k(\alpha'-\gamma')=l$ from the first equation. This contradicts the assumption that $k>l>0$. For Case 3, the second equation implies that $k(\delta'-\beta')=1$, which contradicts the assumption that $k>1$. Thus, for $k\neq l$, there does not exist an orientation-preserving self-diffeomorphism of R^4 which carries F_1 to F_1' and F_2 to F_2'.

Next we show that for $k\neq l$, there does not exist an orientation-preserving self-diffeomorphism of R^4 which carries F_1 to F_2' and F_2 to F_1'. We discuss a similar argument as in the former case of a diffeomorphism which carries F_1 to F_1' and F_2 to F_2', using the same notation except where noted.

Assume that there is such a diffeomorphism f, and consider Γ as before. Then, since $\mathrm{tlk}_{1,2,3}$ for $\widetilde{S}^1_k$ and $\mathrm{tlk}_{3,4,1}=\mathrm{tlk}_{4,3,2}$ (see Lemma 6.1) for $\widetilde{S}^1_l$ are the same, and since $\mathrm{tlk}_{4,3,2}=-\mathrm{tlk}_{2,3,4}$ [Carter et al. 2003], Lemma 6.2 implies that instead of (6-3) we have

$$-kp+q=p'-q', \tag{6-10}$$

where $p'=I(\mu_2',\Gamma')/2$ and $q'=I(\lambda_2',\Gamma')/2$. Hence instead of (6-4) we have

$$\begin{pmatrix}k\\-1\end{pmatrix}=A^T\begin{pmatrix}-1\\1\end{pmatrix}. \tag{6-11}$$

Next we will consider another 2-dimensional braid $\widetilde{S}^2_k$ over S_k, presented by the 2-chart $\widetilde{\Gamma}$ as in the former case. Then, by the same argument, we have (6-8), where $p''=I(\mu_1',\Gamma'')/2$ and $q''=I(\lambda_1',\Gamma'')/2$.

By the composite diffeomorphism $h\circ f\circ g$, $\widetilde{S}^2_k$ is carried to a 2-dimensional braid over S_l, which will be denoted by $\widetilde{S}^2_l$. Since $\mathrm{tlk}_{2,3,4}$ for $\widetilde{S}^2_k$ and $\mathrm{tlk}_{3,1,2}=\mathrm{tlk}_{3,2,1}$ (see Lemma 6.1) for $\widetilde{S}^2_l$ are the same, and since $\mathrm{tlk}_{3,2,1}=-\mathrm{tlk}_{1,2,3}$ [Carter et al. 2003], together with $I(\mu_2,\widetilde{\Gamma})=2kp$ and $I(\lambda_2,\widetilde{\Gamma})=2q$, Lemma 6.2 implies that

$$-kp+q=lp''-q''. \tag{6-12}$$

Since this equation holds true for any integers p and q, it follows from (6-8) that

$$\begin{pmatrix}k\\-1\end{pmatrix}=(CBA)^T\begin{pmatrix}-l\\1\end{pmatrix}.$$

Thus, together with (6-11),

$$B^TC^T\begin{pmatrix}-l\\1\end{pmatrix}=\begin{pmatrix}-1\\1\end{pmatrix},$$

whence $i(-l\alpha'+\gamma')=-1$ and $j(-l\beta'+\delta')=1$. Let us assume $k>l>0$. Since at least one of i and j is k for Cases 1, 2, and 3, these equations contradict the assumption that $k>1$. Thus, for $k\neq l$, there does not exist an orientation-preserving self-diffeomorphism of R^4 carrying F_1 to F_2' and F_2 to F_1'. Thus S_k and S_l are not equivalent for positive integers $k \neq l$. □

Acknowledgments

The author would like to thank Professors Seiichi Kamada, Shin Satoh, Hiroki Kodama, Takuya Sakasai, and the referees for their helpful comments. The author was supported by iBMath through the fund for Platform for Dynamic Approaches to Living Systems from MEXT.

References

[Boyle 1993] J. Boyle, "The turned torus knot in S^4", *J. Knot Theory Ramifications* **2**:3 (1993), 239–249. MR 94i:57037 Zbl 0995.57502

[Carter et al. 2001] J. S. Carter, S. Kamada, M. Saito, and S. Satoh, "A theorem of Sanderson on link bordisms in dimension 4", *Algebr. Geom. Topol.* **1** (2001), 299–310. MR 2002d:57018 Zbl 0973.57010

[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

[Carter et al. 2004] S. Carter, S. Kamada, and M. Saito, *Surfaces in 4-space*, Encyclopaedia of Mathematical Sciences **142**, Springer, Berlin, 2004. MR 2005e:57065 Zbl 1078.57001

[Ito and Nakamura 2014] T. Ito and I. Nakamura, "On surface links whose link groups are abelian", *Math. Proc. Cambridge Philos. Soc.* **157**:1 (2014), 63–77. MR 3211808 Zbl 1298.57018

[Iwase 1988] Z. Iwase, "Dehn-surgery along a torus T^2-knot", *Pacific J. Math.* **133**:2 (1988), 289–299. MR 89m:57020 Zbl 0629.57009

[Kamada 1992] S. Kamada, "Surfaces in $\mathbf{R}^4$ of braid index three are ribbon", *J. Knot Theory Ramifications* **1**:2 (1992), 137–160. MR 93h:57039 Zbl 0763.57013

[Kamada 2002] S. Kamada, *Braid and knot theory in dimension four*, Mathematical Surveys and Monographs **95**, American Mathematical Society, Providence, RI, 2002. MR 2003d:57050 Zbl 0993.57012

[Kawauchi 1996] A. Kawauchi, *A survey of knot theory*, Birkhäuser, Basel, 1996. MR 97k:57011 Zbl 0861.57001

[Lickorish 1997] W. B. R. Lickorish, *An introduction to knot theory*, Graduate Texts in Mathematics **175**, Springer, New York, 1997. MR 98f:57015 Zbl 0886.57001

[Montesinos 1983] J. M. Montesinos, "On twins in the four-sphere. I", *Quart. J. Math. Oxford Ser.* (2) **34**:134 (1983), 171–199. MR 86i:57025a Zbl 0522.57019

[Nakamura 2011] I. Nakamura, "Surface links which are coverings over the standard torus", *Algebr. Geom. Topol.* **11**:3 (2011), 1497–1540. MR 2012g:57048 Zbl 1230.57022

[Nakamura 2012] I. Nakamura, "Triple linking numbers and triple point numbers of certain T^2-links", *Topology Appl.* **159**:5 (2012), 1439–1447. MR 2879374 Zbl 1236.57030

[Nakamura 2014a] I. Nakamura, "Satellites of an oriented surface link and their local moves", *Topology Appl.* **164** (2014), 113–124. MR 3160466 Zbl 1295.57028

[Nakamura 2014b] I. Nakamura, "Surface links with free abelian groups", *J. Math. Soc. Japan* **66**:1 (2014), 247–256. MR 3161400 Zbl 1297.57056

[Roseman 1998] D. Roseman, “Reidemeister-type moves for surfaces in four-dimensional space”, pp. 347–380 in *Knot theory* (Warsaw, 1995), edited by V. F. R. Jones et al., Banach Center Publ. **42**, Polish Acad. Sci., Warsaw, 1998. MR 99f:57029 Zbl 0906.57010

[Rudolph 1983] L. Rudolph, “Braided surfaces and Seifert ribbons for closed braids”, *Comment. Math. Helv.* **58**:1 (1983), 1–37. MR 84j:57006 Zbl 0522.57017

[Sanderson 1987] B. J. Sanderson, “Bordism of links in codimension 2”, *J. London Math. Soc.* (2) **35**:2 (1987), 367–376. MR 88d:57023 Zbl 0622.57026

[Sanderson 1993] B. J. Sanderson, “Triple links in codimension 2”, pp. 457–471 in *Topology. Theory and applications, II* (Pécs, 1989), edited by Á. Császár, Colloq. Math. Soc. János Bolyai **55**, North-Holland, Amsterdam, 1993. MR 96a:57056 Zbl 0815.57016

[Suciu 1985] A. I. Suciu, “Infinitely many ribbon knots with the same fundamental group”, *Math. Proc. Cambridge Philos. Soc.* **98**:3 (1985), 481–492. MR 87a:57025 Zbl 0596.57013

Received October 7, 2014. Revised March 27, 2015.

INASA NAKAMURA
INSTITUTE FOR BIOLOGY AND MATHEMATICS OF DYNAMIC CELLULAR PROCESSES
INTERDISCIPLINARY CENTER FOR MATHEMATICAL SCIENCES
GRADUATE SCHOOL OF MATHEMATICAL SCIENCES
THE UNIVERSITY OF TOKYO
3-8-1 KOMABA, TOKYO 153-8914
JAPAN

inasa@ms.u-tokyo.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 278, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.278.253

CORRECTION TO MODULAR L-VALUES OF CUBIC LEVEL

ANDREW KNIGHTLY AND CHARLES LI

Volume **260**:2 (2012), 527–563

In the paper in question, equation (2-23) is incorrect, and hence so are equations (3-6) and (3-9). We correct the statements in this note. All other statements in the paper, including the main theorems, are unaffected.

In [Knightly and Li 2012], equation (2-23) was quoted from an early draft of [Knightly and Li 2015], which at the time contained an error. It should read

$$f_2^\sigma(zg) = \frac{(p+1)\zeta}{2\omega_p(zd)} \sum_{w\in(\mathbb{Z}/p\mathbb{Z})^*} \overline{\omega_p(w)}\theta_p\left(\frac{-\frac{c}{a}w - \frac{tb}{d}w^{-1}}{p}\right) \tag{2-23}$$

for $g = \left(\begin{smallmatrix} c & dp^{-2} \\ ap & b \end{smallmatrix}\right) \in \left(\begin{smallmatrix} \mathbb{Z}_p & (1/p^2)\mathbb{Z}_p^* \\ p\mathbb{Z}_p^* & \mathbb{Z}_p \end{smallmatrix}\right)$. It is a twisted Kloosterman sum.

Proposition 0.1 (Corrected Proposition 3.4). *Let* $\delta = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, *so that*

$$J_\delta(s, f^\sigma) = \int_{\mathbb{Q}_p^*}\int_{\mathbb{Q}_p} f^\sigma\left(\begin{pmatrix} 0 & -y \\ 1 & x \end{pmatrix}\right)\overline{\theta_p(rx)}\, dx\, \chi_p(y)\,|y|_p^{k/2-s}\, d^*y.$$

Then

$$J_\delta(s, f^\sigma) = J_\delta(s, f_2^\sigma) = \begin{cases} \dfrac{(p^3)^{k/2-s}\, p(p+1)\omega_p(-rp^2)}{2\zeta\,\chi_p(p^3)} & \text{if } p \nmid r, \\ 0 & \text{otherwise,} \end{cases} \tag{3-6}$$

and

$$I_\delta(s)_p = 0. \tag{3-7}$$

Proof. By (2-22), the matrix $\left(\begin{smallmatrix} 0 & -y \\ 1 & x \end{smallmatrix}\right)$ never belongs to $\mathrm{Supp}(f_1^\sigma)$, so $J_\delta(s, f^\sigma) = J_\delta(s, f_2^\sigma)$. Note that $\left(\begin{smallmatrix} 0 & -y \\ 1 & x \end{smallmatrix}\right) \in \mathrm{Supp}(f_2^\sigma)$ if and only if

$$\begin{pmatrix} 0 & -py \\ p & px \end{pmatrix} \in \begin{pmatrix} \mathbb{Z}_p & p^{-2}\mathbb{Z}_p^* \\ p\mathbb{Z}_p^* & \mathbb{Z}_p \end{pmatrix}.$$

MSC2010: primary 11F41; secondary 11F70.
Keywords: L-functions, relative trace formula, supercuspidal representations, Maass forms.

In this case, we may write $y = -p^{-3}u$ for $u \in \mathbb{Z}_p^*$, and $x' = px \in \mathbb{Z}_p$. Then $dx' = p^{-1}dx$, and dropping the $'$ from the notation, we have

$$\begin{aligned}
&J_\delta(s, f_2^\sigma)\\
&= p \int_{\mathbb{Z}_p^*}\int_{\mathbb{Z}_p} f_2^\sigma\left(\begin{pmatrix} p^{-1} & \\ & p^{-1}\end{pmatrix}\begin{pmatrix} 0 & p^{-2}u \\ p & x\end{pmatrix}\right)\theta_p\left(\frac{-rx}{p}\right) dx\, \chi_p(p^{-3})(p^3)^{k/2-s}\, d^*u\\
&= \frac{(p^3)^{k/2-s}p(p+1)\zeta\omega_p(p)}{2\chi_p(p^3)} \sum_{w\in(\mathbb{Z}/p\mathbb{Z})^*}\overline{\omega_p(w)}\int_{\mathbb{Z}_p^*}\overline{\omega_p(u)}\int_{\mathbb{Z}_p}\theta_p\left(\frac{-\frac{tx}{uw}}{p}\right)\theta_p\left(\frac{-rx}{p}\right) dx\, d^*u
\end{aligned}$$

by (2-23). Replacing u by $(-uw)^{-1}$, the above is equal to

$$\frac{(p^3)^{k/2-s}p(p+1)\zeta\omega_p(p)}{2\chi_p(p^3)} \sum_{w\in(\mathbb{Z}/p\mathbb{Z})^*}\overline{\omega_p(w)}\int_{\mathbb{Z}_p^*}\omega_p(-uw)\int_{\mathbb{Z}_p}\theta_p\left(\frac{(tu-r)x}{p}\right) dx\, d^*u.$$

Observe that w is eliminated, and the sum over w contributes $p-1$. Furthermore,

$$\int_{\mathbb{Z}_p}\theta_p\left(\frac{(tu-r)x}{p}\right) dx = \begin{cases} 1 & \text{if } u \in t^{-1}r + p\mathbb{Z}_p,\\ 0 & \text{otherwise.}\end{cases}$$

In particular, it vanishes if $p \mid r$. Assuming $p \nmid r$,

$$\begin{aligned}
J_\delta(s, f^\sigma) &= \frac{(p^3)^{k/2-s}p(p+1)\zeta\omega_p(p)}{2\chi_p(p^3)}(p-1)\int_{t^{-1}r+p\mathbb{Z}_p}\omega_p(-u)\, d^*u\\
&= \frac{(p^3)^{k/2-s}p(p+1)\zeta\omega_p(p)}{2\chi_p(p^3)}\omega_p(-t^{-1}r)
\end{aligned}$$

since the coset has multiplicative measure $1/(p-1)$. Equality (3-6) now follows, using the fact that

$$\frac{\zeta\omega_p(p)}{\omega_p(t)} = \frac{\zeta\omega_p(p)^2}{\omega_p(pt)} = \frac{\omega_p(p)^2}{\zeta}.$$

For fixed t, summing (3-6) over $\pm\zeta$ gives 0. Thus $I_\delta(s)_p = \sum_\sigma J_\delta(s, f^\sigma) = 0$. □

Proposition 0.2 (Corrected Proposition 3.5). *For $a \in \mathbb{Q}^*$, let $\delta_a = \left(\begin{smallmatrix} a & -1 \\ 1 & 0\end{smallmatrix}\right)$, so that*

$$J_{\delta_a}(s, f^\sigma) = \int_{\mathbb{Q}_p^*}\int_{\mathbb{Q}_p} f^\sigma\left(\begin{pmatrix} ya & y(xa-1) \\ 1 & x\end{pmatrix}\right)\overline{\theta_p(rx)}\, dx\, \chi_p(y)\,|y|_p^{k/2-s}\, d^*y.$$

Then $J_{\delta_a}(s, f_1^\sigma)$ vanishes unless $a \in p^2\mathbb{Z}_p$ and $p \nmid r$. In this case, writing $a = p^{a_p}a_0$ for $a_0 \in \mathbb{Z}_p^* \cap \mathbb{Q}^*$, we have

$$J_{\delta_a}(s, f_1^\sigma) = \begin{cases} \dfrac{|a|_p^{2s-k} p(p+1)\,\omega_p(p^{a_p})}{2\chi_p(a^2)} \theta_p\left(\dfrac{ta}{rp^3} - \dfrac{r}{a}\right) & \textit{if } a_0 \equiv 1 \bmod p, \\ 0 & \textit{otherwise.} \end{cases} \tag{3-8}$$

The integral $J_{\delta_a}(s, f_2^\sigma)$ *vanishes unless* $a \in p^2\mathbb{Z}_p$. *For such* a,

$$J_{\delta_a}(s, f_2^\sigma) = \begin{cases} \dfrac{(p^3)^{k/2-s} p(p+1)\,\omega_p(-p^2 r)}{2\chi_p(p^3)\zeta} \theta_p\left(-\dfrac{ta}{rp^3}\right) & \textit{if } p \nmid r, \\ 0 & \textit{otherwise.} \end{cases} \tag{3-9}$$

Finally, $I_{\delta_a}(s)_p$ *vanishes unless* $p \nmid r$ *and* $a = p^{a_p} a_0$ *for* $a_p \geq 2$ *and* $a_0 \equiv 1 \bmod p\mathbb{Z}_p$. *If these conditions are satisfied, then*

(3-10)

$$I_{\delta_a}(s)_p = \frac{|a|_p^{2s-k} p(p+1)\,\omega_p(p^{a_p})\theta_p\left(-\frac{r}{a}\right)}{\chi_p(a^2)} \Delta_p(a) \quad \textit{for } \Delta_p(a) = \begin{cases} p-1 & \textit{if } a_p > 2, \\ -1 & \textit{if } a_p = 2. \end{cases}$$

Remark. Only (3-9) differs from the original statement.

Proof of (3-9). Consider

$$J_{\delta_a}(s, f_2^\sigma) = \int_{\mathbb{Q}_p^*} \int_{\mathbb{Q}_p} f_2^\sigma\left(\begin{pmatrix} ya & y(xa-1) \\ 1 & x \end{pmatrix}\right) \overline{\theta_p(rx)}\, dx\, \chi_p(y)\, |y|_p^{k/2-s}\, d^*y.$$

By (2-23), the integrand is nonzero precisely when

$$\begin{pmatrix} p & \\ & p \end{pmatrix} \begin{pmatrix} ya & y(xa-1) \\ 1 & x \end{pmatrix} = \begin{pmatrix} pya & py(xa-1) \\ p & px \end{pmatrix} \in \begin{pmatrix} \mathbb{Z}_p & p^{-2}\mathbb{Z}_p^* \\ p\mathbb{Z}_p^* & \mathbb{Z}_p \end{pmatrix}.$$

Taking the determinant, this says, in particular, that $p^2 y \in p^{-1}\mathbb{Z}_p^*$, so we may write $y = u/p^3$ for $u \in \mathbb{Z}_p^*$. Replacing px by x, we have

$$J_{\delta_a}(s, f_2^\sigma) = \frac{(p^3)^{k/2-s} p\,\omega_p(p)}{\chi_p(p^3)} \int_{\mathbb{Z}_p^*} \int_{\mathbb{Z}_p} f_2^\sigma\left(\begin{pmatrix} \frac{ua}{p^2} & \frac{u}{p^2}\left(\frac{xa}{p} - 1\right) \\ p & x \end{pmatrix}\right) \theta_p\left(\frac{-rx}{p}\right) dx\, d^*u.$$

From the upper left entry, the integrand is nonzero only if $a_p \geq 2$. Assuming the latter, we also have $xa/p - 1 \in \mathbb{Z}_p^*$, so the upper right entry belongs to $p^{-2}\mathbb{Z}_p^*$ as required. Hence by (2-23), the above is equal to

$$\frac{(p^3)^{k/2-s} p(p+1)\zeta\,\omega_p(p)}{2\chi_p(p^3)} \sum_{w \in (\mathbb{Z}/p\mathbb{Z})^*} \overline{\omega_p(w)} \int_{\mathbb{Z}_p^*} \overline{\omega_p(u)}\theta_p\left(-\frac{uaw}{p^3}\right)$$

$$\int_{\mathbb{Z}_p} \overline{\omega_p\left(\frac{xa}{p} - 1\right)} \theta_p\left(\frac{-txu^{-1}\left(\frac{xa}{p} - 1\right)^{-1} w^{-1}}{p}\right) \theta_p\left(\frac{-rx}{p}\right) dx\, d^*u.$$

Note that $\omega_p(xa/p-1)=\omega_p(-1)$ since $p^2\mid a$. For the same reason,

$$\theta_p\left(\frac{-txu^{-1}\left(\frac{xa}{p}-1\right)^{-1}w^{-1}}{p}\right)=\theta_p\left(\frac{tu^{-1}w^{-1}x}{p}\right).$$

Therefore the above integral over $\mathbb{Z}_p$ equals

$$\omega_p(-1)\int_{\mathbb{Z}_p}\theta_p\left(\frac{(-r+tu^{-1}w^{-1})x}{p}\right)dx=\begin{cases}\omega_p(-1) & \text{if } u\in tr^{-1}w^{-1}+p\mathbb{Z}_p,\\ 0 & \text{otherwise.}\end{cases}$$

In particular, $J_{\delta_a}(s,f_2^\sigma)=0$ if $p\mid r$. Assuming $p\nmid r$, $J_{\delta_a}(s,f_2^\sigma)$ equals

$$\begin{aligned}&\frac{(p^3)^{k/2-s}p(p+1)\zeta\omega_p(-p)}{2\chi_p(p^3)}\sum_{w\in(\mathbb{Z}/p\mathbb{Z})^*}\overline{\omega_p(w)}\int_{tr^{-1}w^{-1}+p\mathbb{Z}_p}\overline{\omega_p(u)}\theta_p\left(-\frac{uaw}{p^3}\right)d^*u\\
&=\frac{(p^3)^{k/2-s}p(p+1)\zeta\omega_p(-p)}{2\chi_p(p^3)}\sum_{w\in(\mathbb{Z}/p\mathbb{Z})^*}\overline{\omega_p(w)}\,\overline{\omega_p(tr^{-1}w^{-1})}\theta_p\left(-\frac{tr^{-1}a}{p^3}\right)\frac{1}{p-1}\\
&=\frac{(p^3)^{k/2-s}p(p+1)\zeta\omega_p(-p)}{2\chi_p(p^3)}\overline{\omega_p(tr^{-1})}\theta_p\left(-\frac{tr^{-1}a}{p^3}\right).\end{aligned}$$

Equation (3-9) follows upon using $\zeta\omega_p(p)/\omega_p(t)=\omega_p(p)^2/\zeta$. □

References

[Knightly and Li 2012] A. Knightly and C. Li, "Modular L-values of cubic level", *Pacific J. Math.* **260**:2 (2012), 527–563. MR 3001804 Zbl 1286.11069

[Knightly and Li 2015] A. Knightly and C. Li, "Simple supercuspidal representations of GL(n)", *Taiwanese J. Math.* **19**:4 (2015), 995–1029.

Received March 1, 2015.

ANDREW KNIGHTLY
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MAINE
NEVILLE HALL
ORONO, ME 04469-5752
UNITED STATES
knightly@math.umaine.edu

CHARLES LI
DEPARTMENT OF MATHEMATICS
THE CHINESE UNIVERSITY OF HONG KONG
SHATIN
HONG KONG
charles@charlesli.org

Guidelines for Authors

Authors may submit articles at msp.org/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 278 No. 1 November 2015

Growth tight actions 1

GOULNARA N. ARZHANTSEVA, CHRISTOPHER H. CASHEN and JING TAO

A flag structure on a cusped hyperbolic 3-manifold 51

ELISHA FALBEL and RAFAEL SANTOS THEBALDI

A new upper bound for the Dirac operators on hypersurfaces 79

NICOLAS GINOUX, GEORGES HABIB and SIMON RAULOT

Games and elementary equivalence of II_1-factors 103

ISAAC GOLDBRING and THOMAS SINCLAIR

Grossberg–Karshon twisted cubes and hesitant walk avoidance 119

MEGUMI HARADA and EUNJEONG LEE

Gamma factors of distinguished representations of $\mathrm{GL}_n(\mathbb{C})$ 137

ALEXANDER KEMARSKY

The W-entropy formula for the Witten Laplacian on manifolds with time dependent metrics and potentials 173

SONGZI LI and XIANG-DONG LI

A diagrammatic categorification of the affine q-Schur algebra $\hat{\mathbf{S}}(n, n)$ for $n \geq 3$ 201

MARCO MACKAAY and ANNE-LAURE THIEL

Showing distinctness of surface links by taking 2-dimensional braids 235

INASA NAKAMURA

Correction to Modular L-values of cubic level 253

ANDREW KNIGHTLY and CHARLES LI