Pacific Journal of Mathematics

Volume 292 No. 1 January 2018

PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm

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

The subscription price for 2018 is US $475/year for the electronic version, and $640/year for print and electronic.
Subscriptions, requests for back issues and changes of subscriber address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

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

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 292, No. 1, 2018
dx.doi.org/10.2140/pjm.2018.292.1

NEW CHARACTERIZATIONS OF LINEAR WEINGARTEN SPACELIKE HYPERSURFACES IN THE DE SITTER SPACE

LUIS J. ALÍAS, HENRIQUE F. DE LIMA AND FÁBIO R. DOS SANTOS

We deal with complete linear Weingarten spacelike hypersurfaces immersed in the de Sitter space, that is, spacelike hypersurfaces of de Sitter space whose mean and scalar curvatures are linearly related. In this setting, we apply a suitable extension of the generalized maximum principle of Omori–Yau to show that either such a spacelike hypersurface must be totally umbilical or there holds a sharp estimate for the norm of its total umbilicity tensor, with equality characterizing hyperbolic cylinders of de Sitter space. We also study the parabolicity of these spacelike hypersurfaces with respect to a Cheng–Yau modified operator.

1. Introduction

The last few decades have seen a steadily growing interest in the study of the geometry of spacelike hypersurfaces immersed into a Lorentzian space. Recall that a hypersurface M^n immersed into a Lorentzian space is said to be *spacelike* if the metric induced on M^n from that of the ambient space is positive definite. Apart from physical motivations, from the mathematical point of view this interest is mostly due to the fact that such hypersurfaces exhibit nice Bernstein-type properties, and one can truly say that the first remarkable results in this branch were the rigidity theorems of E. Calabi [1970] and S. Y. Cheng and S. T. Yau [1976], who showed (the former for $n \leq 4$, and the latter for general n) that the only maximal complete, noncompact, spacelike hypersurfaces of the Lorentz–Minkowski space $\mathbb{L}^{n+1}$ are the spacelike hyperplanes. However, in the case that the mean curvature is a positive constant, A. E. Treibergs [1982] astonishingly showed that there are many entire solutions of the corresponding constant mean curvature equation in $\mathbb{L}^{n+1}$, which he was able to classify by their projective boundary values at infinity.

When the ambient is the de Sitter space $\mathbb{S}_1^{n+1}$, A. J. Goddard [1977] conjectured that every complete spacelike hypersurface with constant mean curvature H in $\mathbb{S}_1^{n+1}$ should be totally umbilical. Although the conjecture turned out to be false in its

MSC2010: primary 53C42; secondary 53A10, 53C20, 53C50.

Keywords: de Sitter space, linear Weingarten hypersurfaces, spacelike hypersurfaces, totally umbilical hypersurfaces, hyperbolic cylinders, parabolicity.

original statement, it motivated a great deal of work of several authors trying to find a positive answer to the conjecture under appropriate additional hypotheses. For instance, J. Ramanathan [1987] proved Goddard's conjecture for $\mathbb{S}_1^3$ and $0 \leq H \leq 1$. Moreover, if $H > 1$ he showed that the conjecture is false as can be seen from an example due to M. Dajczer and K. Nomizu [1981]. K. Akutagawa [1987] proved that Goddard's conjecture is true when $n = 2$ and $H^2 \leq 1$ or when $n \geq 3$ and $H^2 < 4(n-1)/n^2$. He also constructed complete spacelike rotation surfaces in $\mathbb{S}_1^3$ with constant H satisfying $H > 1$ which are not totally umbilical.

S. Montiel [1988] proved that Goddard's conjecture is true provided that M^n is compact. Furthermore, he exhibited examples of complete spacelike hypersurfaces in $\mathbb{S}_1^{n+1}$ with constant H satisfying $H^2 \geq 4(n-1)/n^2$ and being nontotally umbilical, the so-called *hyperbolic cylinders*, which are isometric to the Riemannian product $\mathbb{H}^1(r) \times \mathbb{S}^{n-1}(\sqrt{1+r^2})$ of a hyperbolic line of radius $r > 0$ and an $(n-1)$-dimensional Euclidean sphere of radius $\sqrt{1+r^2}$. S. Montiel [1996] characterized such hyperbolic cylinders as the only complete noncompact spacelike hypersurfaces in $\mathbb{S}_1^{n+1}$ with constant mean curvature $H = 2\sqrt{n-1}/n$ and having at least two ends. A. Brasil Jr., A. G. Colares and O. Palmas [Brasil et al. 2003] obtained a sort of extension of Montiel's result, showing that the hyperbolic cylinders are the only complete spacelike hypersurfaces in $\mathbb{S}_1^{n+1}$ with constant mean curvature, nonnegative Ricci curvature and having at least two ends. They also characterized all complete spacelike hypersurfaces of constant mean curvature with two distinct principal curvatures as rotation hypersurfaces or generalized hyperbolic cylinders $\mathbb{H}^k(r) \times \mathbb{S}^{n-k}(\sqrt{1+r^2})$. Proceeding with the ideas related to Goddard's conjecture, it is interesting to obtain characterizations of the so-called *linear Weingarten* spacelike hypersurfaces (that is, spacelike hypersurfaces whose mean and scalar curvatures are linearly related) immersed in a certain Lorentzian space. Many authors have approached problems in this branch. For instance, when the ambient space is $\mathbb{S}_1^{n+1}$, we refer to the readers the works [Chao 2013; Cheng 1990; de Lima and Velásquez 2013; Hou and Yang 2010; Li 1997].

Here, our purpose is to obtain new characterizations concerning complete linear Weingarten spacelike hypersurfaces immersed in $\mathbb{S}_1^{n+1}$. Under appropriated constrains on the scalar curvature function, we apply a suitable extension of the generalized maximum principle of Omori–Yau (see Proposition 7) in order to give a sharp estimate of the total umbilicity tensor of the hypersurface, which allows us to characterize hyperbolic cylinders

$$\mathbb{H}^1(r) \times \mathbb{S}^{n-1}(\sqrt{1+r^2})$$

of $\mathbb{S}_1^{n+1}$ when $n \geq 3$ (see Theorem 8 and Corollary 9) and totally umbilic 2-spheres in $\mathbb{S}_1^3$ when $n = 2$ (see Theorem 10 and Corollary 11). We also study the parabolicity of these spacelike hypersurfaces with respect to a Cheng–Yau modified operator (see Theorem 12 and Proposition 13).

2. Preliminaries

Let $\mathbb{R}_1^{n+2}$ be an $(n+2)$-dimensional real vector space endowed with an inner product of index 1 given by

$$\langle x, y\rangle = \sum_{j=1}^{n+1} x_j y_j - x_{n+2} y_{n+2},$$

where $x = (x_1, x_2, \dots, x_{n+2})$ is the natural coordinate of $\mathbb{R}_1^{n+2}$.

$\mathbb{R}_1^{n+2} = \mathbb{L}^{n+2}$ is called the $(n+2)$-dimensional *Lorentz–Minkowski space*. We define the $(n+1)$-dimensional *de Sitter space* $\mathbb{S}_1^{n+1}$ as the following hyperquadric of $\mathbb{L}^{n+2}$:

$$\mathbb{S}_1^{n+1} = \{(x_1, x_2, \dots x_{n+2}) \in \mathbb{R}_1^{n+2} : \langle x, x\rangle = 1\}.$$

The induced metric $\langle\,\cdot\,,\cdot\,\rangle$ makes $\mathbb{S}_1^{n+1}$ a Lorentzian manifold with constant sectional curvature 1.

An n-dimensional hypersurface M^n in $\mathbb{S}_1^{n+1}$ is said to be *spacelike* if the metric on M^n induced from the metric of $\mathbb{S}_1^{n+1}$ is positive definite.

From now on, we will consider complete spacelike hypersurfaces M^n of $\mathbb{S}_1^{n+1}$. We choose a local field of semi-Riemannian orthonormal frame $\{e_A\}_{1\le A\le n+1}$ in $\mathbb{S}_1^{n+1}$, with dual coframe $\{\omega_A\}_{1\le A\le n+1}$, such that, at each point of M^n, $e_1, \dots, e_n$ are tangent to M^n and e_{n+1} is normal to M^n. We will use the following convention for the indices

$$1 \le A, B, C, \dots \le n+1, \quad 1 \le i, j, k, \dots \le n.$$

In this setting, denoting by $\{\omega_{AB}\}$ the connection forms of $\mathbb{S}_1^{n+1}$, the structure equations of $\mathbb{S}_1^{n+1}$ are given by

$$d\omega_A = \sum_i \omega_{Ai}\wedge\omega_i - \omega_{An+1}\wedge\omega_{n+1}, \quad \omega_{AB}+\omega_{BA}=0,$$

$$d\omega_{AB} = \sum_C \varepsilon_C\omega_{AC}\wedge\omega_{CB} - \frac{1}{2}\sum_{C,D}\varepsilon_C\varepsilon_D K_{ABCD}\omega_C\wedge\omega_D,$$

$$K_{ABCD} = \varepsilon_A\varepsilon_B(\delta_{AC}\delta_{BD} - \delta_{AD}\delta_{BC}),$$

where $\varepsilon_i = 1$ and $\varepsilon_{n+1} = -1$.

Next, we restrict all the tensors to M^n. First of all, $\omega_{n+1} = 0$ on M^n, so $\sum_i \omega_{n+1i}\wedge\omega_i = d\omega_{n+1} = 0$ and by Cartan's lemma we can write

$$\omega_{n+1i} = \sum_j h_{ij}\omega_j, \quad h_{ij} = h_{ji}. \tag{2-1}$$

This gives the second fundamental form of M^n, $A = \sum_{ij} h_{ij}\omega_i \otimes \omega_j e_{n+1}$. Furthermore, the mean curvature H of M^n is defined by $H = 1/n \sum_i h_{ii}$.

The structure equations of M^n are given by

$$d\omega_i = \sum_j \omega_{ij} \wedge \omega_j, \quad \omega_{ij} + \omega_{ji} = 0,$$
$$d\omega_{ij} = \sum_k \omega_{ik} \wedge \omega_{kj} - \frac{1}{2} \sum_{k,l} R_{ijkl}\omega_k \wedge \omega_l.$$

Using the structure equations we obtain the Gauss equation

$$R_{ijkl} = \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk} - h_{ik}h_{jl} + h_{il}h_{jk}, \tag{2-2}$$

where R_{ijkl} are the components of the curvature tensor of M^n.

The Ricci curvature and the normalized scalar curvature of M^n are given, respectively, by

$$R_{ij} = (n-1)\delta_{ij} - nHh_{ij} + \sum_k h_{ik}h_{kj} \tag{2-3}$$

and

$$R = \frac{1}{n(n-1)} \sum_i R_{ii}. \tag{2-4}$$

From (2-3) and (2-4) we obtain

$$S = n^2H^2 + n(n-1)(R-1) = nH^2 + n(n-1)(H^2 - H_2), \tag{2-5}$$

where $S = \sum_{i,j} h_{ij}^2$ is the square of the length of the second fundamental form A of M^n, and $H_2 = 2S_2/(n(n-1))$ denotes the mean value of the second elementary symmetric function S_2 on the eigenvalues of A. In particular, it follows from (2-5) that M^n is totally umbilical if and only if $S = nH^2$.

The components h_{ijk} of the covariant derivative ∇A satisfy

$$\sum_k h_{ijk}\omega_k = dh_{ij} + \sum_k h_{ik}\omega_{kj} + \sum_k h_{jk}\omega_{ki}. \tag{2-6}$$

Observe that,

$$|\nabla A|^2 = \sum_{i,j,k} h_{ijk}^2.$$

Then, by exterior differentiation of (2-1), we obtain the *Codazzi equation*

$$h_{ijk} = h_{jik} = h_{ikj}. \tag{2-7}$$

Similarly, the components h_{ijkl} of the second covariant derivative $\nabla^2 B$ are given by

$$\sum_l h_{ijkl}\omega_l = dh_{ijk} + \sum_l h_{ljk}\omega_{li} + \sum_l h_{ilk}\omega_{lj} + \sum_l h_{ijl}\omega_{lk}.$$

By exterior differentiation of (2-6), we can get the following *Ricci formula*

$$h_{ijkl} - h_{ijlk} = \sum_m h_{im}R_{mjkl} + \sum_m h_{jm}R_{mikl}. \tag{2-8}$$

The Laplacian Δh_{ij} of h_{ij} is defined by $\Delta h_{ij} = \sum_k h_{ijkk}$. From (2-7) and (2-8), we have

$$\Delta h_{ij} = \sum_k h_{kkij} + \sum_{k,l} h_{kl}R_{lijk} + \sum_{k,l} h_{li}R_{lkjk}. \tag{2-9}$$

Since $\Delta S = 2\big(\sum_{i,j} h_{ij}\Delta h_{ij} + \sum_{i,j,k} h_{ijk}^2\big)$, from (2-9) we get

$$\tfrac{1}{2}\Delta S = |\nabla A|^2 + \sum_{i,i,k} h_{ij}h_{kkij} + \sum_{i,j,k,l} h_{ij}h_{lk}R_{lijk} + \sum_{i,j,k,l} h_{ij}h_{il}R_{lkjk}. \tag{2-10}$$

Consequently, taking a (local) orthonormal frame $\{e_1, \ldots, e_n\}$ on M^n such that $h_{ij} = \lambda_i\delta_{ij}$, from (2-10) we obtain the following Simons-type formula:

$$\tfrac{1}{2}\Delta S = |\nabla A|^2 + \sum_i \lambda_i(nH)_{,ii} + \frac{1}{2}\sum_{i,j} R_{ijij}(\lambda_i - \lambda_j)^2. \tag{2-11}$$

Now, let $\phi = \sum_{i,j}\phi_{ij}\omega_i\omega_j$ be a symmetric tensor on M^n defined by

$$\phi_{ij} = nH\delta_{ij} - h_{ij}.$$

Following [Cheng and Yau 1977], we introduce a operator $\square$ associated to ϕ acting on any smooth function f by

$$\square f = \sum_{i,j}\phi_{ij}f_{ij} = \sum_{i,j}(nH\delta_{ij} - h_{ij})f_{ij}. \tag{2-12}$$

Setting $f = nH$ in (2-12) and taking into account equations (2-5) and (2-11), with a straightforward computation we obtain

$$\square(nH) = |\nabla A|^2 - n^2|\nabla H|^2 - \tfrac{1}{2}n(n-1)\Delta R + \frac{1}{2}\sum_{i,j} R_{ijij}(\lambda_i - \lambda_j)^2. \tag{2-13}$$

3. Linear Weingarten hypersurfaces in $\mathbb{S}_1^{n+1}$

In what follows, we will consider M^n as being a *linear Weingarten* spacelike hypersurface immersed in $\mathbb{S}_1^{n+1}$, that is, a spacelike hypersurface of $\mathbb{S}_1^{n+1}$ whose mean curvature H and normalized scalar curvature R satisfy

$$R = aH + b,$$

for some $a, b \in \mathbb{R}$. We first state some auxiliary results.

Lemma 1 [de Lima and Velásquez 2013]. *Let M^n be a linear Weingarten spacelike hypersurface in $\mathbb{S}_1^{n+1}$, such that $R = aH + b$ for some $a, b \in \mathbb{R}$. Suppose that*

$$(n-1)a^2 + 4n(1-b) \geq 0. \tag{3-1}$$

Then

$$|\nabla A|^2 \geq n^2 |\nabla H|^2. \tag{3-2}$$

Moreover, if the inequality (3-1) is strict and the equality holds in (3-2) on M^n, then H is constant on M^n.

Now, we will consider the following Cheng–Yau's modified operator:

$$L = \square + \frac{n-1}{2} a \Delta. \tag{3-3}$$

In other words, for any $u \in \mathcal{C}^2(M)$,

$$L(u) = \operatorname{tr}(P \circ \nabla^2 u), \tag{3-4}$$

with

$$P = \left(nH + \frac{n-1}{2} a \right) I - A, \tag{3-5}$$

where I is the identity in the algebra of smooth vector fields on M^n and $\nabla^2 u$ stands for the self-adjoint linear operator metrically equivalent to the Hessian of u.

Remark 2. From Equation (2-5), since $R = aH + b$, we have that

$$n^2 H^2 = S - n(n-1)(aH + b - 1). \tag{3-6}$$

When $b < 1$, it follows from (3-6) that $H(p) \neq 0$ for every $p \in M^n$. In this case, we can choose the orientation of M^n such that $H > 0$. On the other hand, when $b = 1$, we will assume that H does not change sign on M^n and we will choose the orientation of M^n such that $H \geq 0$.

The next lemma gives a sufficient criterion for the ellipticity of the operator L.

Lemma 3. *Let M^n be a linear Weingarten spacelike hypersurface in $\mathbb{S}_1^{n+1}$ such that $R = aH + b$. Let μ_- and μ_+ be, respectively, the minimum and the maximum of the eigenvalues of the operator P at every point $p \in M^n$. If $b < 1$, then the operator L is elliptic, with*

$$\mu_- > 0 \quad \textit{and} \quad \mu_+ < 2nH + (n-1)a.$$

In the case where $b = 1$, assume further that the mean curvature function H does not change sign and $R \geq 1$. Then the operator L is semielliptic, with

$$\mu_- \geq 0 \quad \textit{and} \quad \mu_+ \leq 2nH + (n-1)a,$$

unless M^n is totally geodesic.

Proof. Let us consider $b < 1$ and choose the orientation on M^n for which $H > 0$ (see Remark 2). From (3-6), we have that

$$n^2H^2 = S + n(n-1)(1 - aH - b) \geq \lambda_i^2 - n(n-1)aH,$$

for each principal curvature λ_i of M^n, $i = 1, \ldots, n$.

On the other hand, with a straightforward computation we verify that

$$\begin{aligned} \lambda_i^2 &\leq n^2H^2 + n(n-1)aH \\ &= \left(nH + \frac{n-1}{2}a\right)^2 - \frac{(n-1)^2}{4}a^2 \\ &\leq \left(nH + \frac{n-1}{2}a\right)^2. \end{aligned} \tag{3-7}$$

From (3-6) we also have that

$$nH(nH + (n-1)a) = S + n(n-1)(1-b) > 0. \tag{3-8}$$

We claim that $nH + \frac{1}{2}(n-1)a > 0$. When $a \geq 0$, our assertion is immediate since

$$nH + \left(\frac{n-1}{2}\right)a \geq nH > 0.$$

When $a < 0$, from (3-8) we get $nH + (n-1)a > 0$. So, $nH + \frac{1}{2}(n-1)a > nH + (n-1)a > 0$.

So, from (3-7) we obtain

$$-nH - \left(\frac{n-1}{2}\right)acn - 12a < \lambda_i < nH + \left(\frac{n-1}{2}\right)acn - 12a, \quad i = 1, \ldots, n.$$

Therefore, for every i, we get

$$0 < nH + \left(\frac{n-1}{2}\right)a - \lambda_i < 2nH + (n-1)a.$$

However, $\mu_i = nH + \frac{1}{2}(n-1)a - \lambda_i$ are precisely the eigenvalues of P. In particular, we conclude that $\mu_- > 0$ and $\mu_+ < 2nH + (n-1)a$.

If $b = 1$, then choose the orientation of M^n for which $H \geq 0$. From (3-6), we have that

$$n^2H^2 = S - n(n-1)aH \geq \lambda_i^2 - n(n-1)aH,$$

for each principal curvature λ_i of M^n, $i = 1, \ldots, n$ and

$$\lambda_i^2 \leq \left(nH + \frac{n-1}{2}a\right)^2.$$

From (3-6) we also have that

$$nH(nH + (n-1)a) = S \geq 0.$$

Since $R = aH + 1 \geq 1$, we have $aH \geq 0$. If $a \geq 0$ then $nH + \frac{1}{2}(n-1)a \geq 0$ and, similarly as in the case $b < 1$, we conclude that $\mu_- \geq 0$ and $\mu_+ \leq 2nH + (n-1)a$.

On the other hand, if $a < 0$ we have $H \equiv 0$ and then $R \equiv 1$ and $S \equiv 0$, which means that M^n must be totally geodesic. □

Remark 4. Also related to the ellipticity of operator L, when M^n is totally geodesic, we observe that the operator L reduces to $L = \frac{1}{2}(n-1)a\Delta$, which is elliptic if and only if $a > 0$. For this reason, in order to keep the validity of Lemma 3 when $b = 1$, even in the totally geodesic case, we will choose a to be a positive constant.

We close this section recalling a classic algebraic lemma due to M. Okumura [1974], which was completed with the equality case by H. Alencar and M. P. do Carmo [1994].

Lemma 5. *Let $\kappa_1, \ldots, \kappa_n$ be real numbers such that $\sum_i \kappa_i = 0$ and $\sum_i \kappa_i^2 = \beta^2$, with $\beta \geq 0$. Then,*

$$-\frac{(n-2)}{\sqrt{n(n-1)}}\beta^3 \leq \sum_i \kappa_i^3 \leq \frac{(n-2)}{\sqrt{n(n-1)}}\beta^3,$$

and equality holds if and only if at least $(n-1)$ of the numbers κ_i are equal.

4. Characterizations of linear Weingarten spacelike hypersurfaces

From now on, we will also consider the following symmetric tensor

$$\Phi = \sum_{i,j} \Phi_{ij}\omega_i \otimes \omega_j$$

associated to the second fundamental form of a hypersurface M^n in $\mathbb{S}_1^{n+1}$, whose components are given by $\Phi_{ij} = h_{ij} - H\delta_{ij}$. Let $|\Phi|^2 = \sum_{i,j} \Phi_{ij}^2$ be the square of the length of Φ. It is not difficult to check that Φ is traceless and, from (2-4), we get

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

In particular, $|\Phi|^2 = 0$ at the umbilic points of M^n. For that reason Φ is usually called the total umbilicity tensor of M^n.

In order to establish our characterization results, we will need the following lower bound for the operator L acting on the square length of the traceless operator of a linear Weingarten hypersurface.

Proposition 6. *Let M^n be a linear Weingarten spacelike hypersurface immersed in $\mathbb{S}_1^{n+1}$, $n \geq 2$, such that $R = aH + b$ with $b \leq 1$. In the case where $b = 1$, assume that the mean curvature function H does not change sign and $R \geq 1$. Then,*

$$L(|\Phi|^2) \geq 2(n-1)|\Phi|^2 \varphi_{a,b}(|\Phi|)\sqrt{\frac{|\Phi|^2}{n(n-1)} + \frac{a^2}{4} - b + 1},$$

where

$$\varphi_{a,b}(x) = \frac{n-2}{n-1}x^2 + \left(na - \frac{n(n-2)}{\sqrt{n(n-1)}}x\right)\sqrt{\frac{x^2}{n(n-1)} + \frac{a^2}{4} - b + 1} + \frac{n(n-2)}{\sqrt{n(n-1)}}\frac{a}{2}x - n\left(\frac{a^2}{2} - b\right). \tag{4-2}$$

Proof. Let us choose a local orthonormal frame $\{e_1, \ldots, e_n\}$ on M^n such that $h_{ij} = \lambda_i \delta_{ij}$. Taking into account equations (2-10) and (2-13), we get from (3-3) that

$$L(nH) = \sum_{i,j,k} h_{ijk}^2 - n^2|\nabla H|^2 + \frac{1}{2}\sum_{i,j} R_{ijij}(\lambda_i - \lambda_j)^2 \tag{4-3}$$

On the one hand, by a straightforward computation we can check

$$\frac{1}{2}\sum_{i,j} R_{ijij}(\lambda_i - \lambda_j)^2 = \frac{1}{2}\sum_{i,j}(1-\lambda_i\lambda_j)(\lambda_i - \lambda_j)^2 = S^2 - nH\sum_i \lambda_i^3 + n|\Phi|^2. \tag{4-4}$$

But, on the other hand, since we are assuming that $b \leq 1$, we have that the relation (3-1) holds, and hence we can apply Lemma 1 to guarantee that

$$\sum_{i,j,k} h_{ijk}^2 - n^2|\nabla H|^2 \geq 0. \tag{4-5}$$

Thus, from (4-3), (4-4) and (4-5) we have

$$L(nH) \geq S^2 - nH\sum_i \lambda_i^3 + n|\Phi|^2. \tag{4-6}$$

Taking a (local) orthonormal frame $\{e_1, \ldots, e_n\}$ at $p \in M^n$ such that

$$h_{ij} = \lambda_i \delta_{ij} \quad \text{and} \quad \phi_{ij} = \kappa_i \delta_{ij},$$

it is not difficult to verify the algebraic relations

$$\sum_i \kappa_i = 0, \quad \sum_i \kappa_i^2 = |\Phi|^2 \quad \text{and} \quad \sum_i \kappa_i^3 = \sum_i \lambda_i^3 - 3H|\Phi|^2 - nH^3. \tag{4-7}$$

When $n \geq 3$, using Lemma 5 and equations (4-1) and (4-7) we have

$$\begin{aligned} S^2 - nH\sum_{i=1}^{n}\lambda_i^3 &= (|\Phi|^2 + nH^2)^2 - nH\sum_i \kappa_i^3 - 3nH^2|\Phi|^2 - n^2H^4 \\ &= |\Phi|^4 - nH^2|\Phi|^2 - nH\sum_i \kappa_i^3 \\ &\geq |\Phi|^2\left(|\Phi|^2 - \frac{n(n-2)}{\sqrt{n(n-1)}}H|\Phi| - nH^2\right). \end{aligned} \tag{4-8}$$

In the case that $n = 2$, since $\kappa_1 + \kappa_2 = 0$ we also have $\kappa_1^3 + \kappa_2^3 = 0$, and from (4-1) and (4-7) we obtain

$$\begin{aligned} S^2 - 2H\sum_{i=1}^{2}\lambda_i^3 &= (\lambda_1^2 + \lambda_2^2)^2 - (\lambda_1 + \lambda_2)(\lambda_1^3 + \lambda_2^3) \\ &= |\Phi|^2(|\Phi|^2 - 2H^2). \end{aligned} \tag{4-9}$$

Hence, inserting either (4-8), when $n \geq 3$, or (4-9), when $n = 2$, into (4-6) we get

$$L(nH) \geq |\Phi|^2\left(|\Phi|^2 - \frac{n(n-2)}{\sqrt{n(n-1)}}H|\Phi| - n(H^2 - 1)\right). \tag{4-10}$$

On the other hand, from (4-1) and $R = aH + b$, we have

$$\frac{1}{n-1}|\Phi|^2 = nH^2 + naH + n(b-1). \tag{4-11}$$

If M^n is totally geodesic then the operator L reduces to $L = \frac{1}{2}(n-1)a\Delta$ where $a > 0$ is any positive constant (see Remark 4). In this case $|\Phi|^2 \equiv 0$ and the inequality in Proposition 6 holds trivially. On the other hand, if M^n is not totally geodesic then Lemma 3 guarantees that the operator P is positive definite if $b < 1$, and P is positive semidefinite if $b = 1$. In any case, from (4-11) we have

$$\begin{aligned} \frac{1}{n-1}L(|\Phi|^2) &= 2HL(nH) + 2n\langle P(\nabla H), \nabla H\rangle + aL(nH) \\ &\geq 2\left(H + \frac{a}{2}\right)L(nH), \end{aligned} \tag{4-12}$$

once (3-4) guarantees that $L(u^2) = 2uL(u) + 2\langle P(\nabla u), \nabla u\rangle$ for every $u \in \mathcal{C}^2(M)$.

Therefore, taking into account that $H + a/2 \geq 0$, from (4-10) and (4-12) we get

$$\frac{1}{2(n-1)}L(|\Phi|^2) \geq \left(H + \frac{a}{2}\right)|\Phi|^2\left(|\Phi|^2 - \frac{n(n-2)}{\sqrt{n(n-1)}}H|\Phi| - n(H^2-1)\right). \tag{4-13}$$

Besides, from (4-11) we have

$$H^2 = \frac{1}{n(n-1)}|\Phi|^2 - aH - b + 1,$$

and consequently, we can write

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

From (4-14) and (4-11), after a straightforward computation, we get

$$|\Phi|^2 - \frac{n(n-2)}{\sqrt{n(n-1)}} H|\Phi| - n(H^2 - 1) = \varphi_{a,b}(|\Phi|), \tag{4-15}$$

where $\varphi_{a,b}(x)$ is defined as in (4-2). Therefore, replacing (4-14) and (4-15) in (4-13), we obtain the desired inequality. □

Let us consider on a Riemannian manifold M^n a semielliptic operator of the form $\mathcal{L} = \mathrm{tr}(\mathcal{P} \circ \mathrm{Hess})$, where $\mathcal{P} : TM \to TM$ is a positive semidefined symmetric tensor. We say that *the Omori–Yau maximum principle* holds on M^n for the operator $\mathcal{L}$ if, for any function $u \in \mathcal{C}^2(M)$ with $u^* = \sup_M u < \infty$, there exists a sequence $\{p_k\}_{k\in\mathbb{N}} \subset M^n$ with the properties

$$u(p_k) > u^* - \frac{1}{k}, \quad |\nabla u(p_k)| < \frac{1}{k} \quad \text{and} \quad \mathcal{L}u(p_k) < \frac{1}{k}$$

for every $k \in \mathbb{N}$.

As an application of Theorem 6.13 of [Alías et al. 2016] (see also Lemma 4.2 of [Alías et al. 2012]), we establish the following Omori–Yau maximum principle which will be our analytical key tool for the proofs of our main results.

Proposition 7. *Let M^n be complete noncompact linear Weingarten spacelike hypersurface immersed in $\mathbb{S}_1^{n+1}$ such that $R = aH + b$ with $b \le 1$. In the case where $b = 1$, assume that the mean curvature function H does not change sign and $R \ge 1$. If $\sup_M |\Phi|^2 < +\infty$, then the Omori–Yau maximum principle holds on M^n for the operator L.*

Proof. Taking into account that $R = aH + b$, from (4-1) we get

$$|\Phi|^2 = n(n-1)(H^2 + aH) + n(n-1)(b-1). \tag{4-16}$$

Since we are assuming $\sup_M |\Phi|^2 < +\infty$, from (4-16) it follows that $\sup_M H < +\infty$. Thus, from (3-5) we have

$$\mathrm{tr}(P) = n(n-1)H + \frac{n(n-1)}{2}a$$

and hence,

$$\sup_M \mathrm{tr}(P) < +\infty. \tag{4-17}$$

On the other hand, recall from the proof of Lemma 3 that $nH+\frac{1}{2}(n-1)a>0$ and

$$-nH-\frac{n-1}{2}a<\lambda_i<nH+\frac{n-1}{2}a,\quad i=1,\dots,n.$$

Therefore, from (2-2) we see that the sectional curvatures of M^n satisfy

$$R_{ijij}=1-\lambda_i\lambda_j\geq 1-\left(nH+\frac{n-1}{2}a\right)^2>-\infty. \tag{4-18}$$

Furthermore, Lemma 3 guarantees us that the operator L is semielliptic. Therefore, taking into account (3-4), (4-17) and (4-18), we can apply Theorem 6.13 of [Alías et al. 2016] in the particular case where the sectional curvatures of M^n are bounded from below by a constant to conclude the desired result. □

Now, we are in position to state and prove our main characterization result concerning linear Weingarten hypersurfaces immersed in $\mathbb{S}_1^{n+1}$.

Theorem 8. *Let M^n be a complete linear Weingarten spacelike hypersurface isometrically immersed in the de Sitter space $\mathbb{S}_1^{n+1}$, $n\geq 3$, such that $R=aH+b$ with $0<b\leq R<(n-2)/n$. Then*

(i) *either $\sup_M|\Phi|^2=0$ and M^n is a totally umbilical hypersurface,*

(ii) *or*

$$\sup_M|\Phi|^2\geq\alpha(n,a,b)>0,$$

where $\alpha(n,a,b)$ is a positive constant depending only on n, a and b.

In (ii), a necessary and sufficient condition for equality to hold and the supremum to be attained at some point of M^n is that M^n be isometric to a hyperbolic cylinder $\mathbb{H}^1(r)\times\mathbb{S}^{n-1}(\sqrt{1+r^2})$ of radius $r>0$.

Proof. If $\sup_M|\Phi|^2=0$, then M^n is totally umbilical and, hence, item (i) holds. If $\sup_M|\Phi|^2=+\infty$, then (ii) is trivially satisfied. So, let us suppose that $0<\sup_M|\Phi|^2<+\infty$ and let us take $u=|\Phi|^2$. Then, from Proposition 6 we get

$$L(u)\geq f(u), \tag{4-19}$$

where

$$f(u)=2(n-1)u\varphi_{a,b}(\sqrt{u})\sqrt{\frac{u}{n(n-1)}+1-b+\frac{a^2}{4}}$$

and $\varphi_{a,b}(x)$ is given by (4-2).

If M^n is compact, there exists a point $p_0\in M^n$ such that $u(p_0)=u^*=\sup_M u$. Consequently, $\nabla u(p_0)=0$ and $Lu(p_0)\leq 0$. Therefore, from (4-19) we get $f(u^*)\leq 0$. Now, assume that M^n is complete and noncompact. Since $u^*<+\infty$, Proposition 7 guarantees that there exists a sequence of points $\{p_k\}_{k\in\mathbb{N}} \subset M^n$ satisfying

$$u(p_k) > u^* - \frac{1}{k} \quad \text{and} \quad Lu(p_k) < \frac{1}{k}, \tag{4-20}$$

for every $k \in \mathbb{N}$. Therefore from (4-19) and (4-20), we get

$$\frac{1}{k} > Lu(p_k) \geq f(u(p_k)). \tag{4-21}$$

Taking the limit $k \to +\infty$ in (4-21), by continuity, we have

$$f(u^*) = 2(n-1)u^*\varphi_{a,b}(\sqrt{u^*})\sqrt{\frac{u^*}{n(n-1)} + 1 - b + \frac{a^2}{4}} \leq 0.$$

Since $u^* > 0$ and $b < 1$, we obtain that

$$\varphi_{a,b}(\sqrt{u^*}) \leq 0. \tag{4-22}$$

Recall from Remark 2 that $H > 0$ on M^n. Thus, since we are assuming that $n \geq 3$ and $0 < b \leq R < (n-2)/n$, it is not difficult to verify that $\varphi_{a,b}$ has an unique positive root $x_0 = \sqrt{\alpha(n,a,b)} > 0$. Moreover, we have that $\varphi_{a,b}(x) > 0$, for $0 \leq x < x_0$, and $\varphi_{a,b}(x) < 0$, for $x > x_0$.

Therefore, (4-22) implies

$$u^* \geq x_0^2 = \alpha(n,a,b),$$

that is,

$$\sup_M |\Phi|^2 \geq \alpha(n,a,b).$$

This proves the inequality of item (ii).

Moreover, the equality $\sup_M |\Phi|^2 = \alpha(n,a,b)$ holds if and only if $\sqrt{u^*} = x_0$. Thus $\varphi_{a,b}(\sqrt{u}) \geq 0$ on M^n, which jointly with (4-19) implies that

$$L(u) \geq 0 \quad \text{on } M^n.$$

On the other hand, since $b < 1$ we know from Lemma 3 that the operator L is elliptic. Therefore, if there exists a point $p_0 \in M^n$ such that $|\Phi(p_0)| = \sup_M |\Phi|$, from the maximum principle the function $u = |\Phi|^2$ must be constant and, consequently, $|\Phi| \equiv x_0$. Thus,

$$0 = L(|\Phi|^2) \geq 2(n-1)|\Phi|^2\varphi_{a,b}(|\Phi|)\sqrt{\frac{|\Phi|^2}{n(n-1)} + 1 - b + \frac{a^2}{4}}.$$

Hence, all the inequalities in the proof of Proposition 6 must be equalities. In particular, since L is elliptic if and only if P is positive definite, returning to (4-12) we obtain that $\nabla H = 0$ and H is constant. Moreover, equality occurs in (4-5) as well, or equivalently,

$$|\nabla A|^2 = \sum_{i,j,k} h_{ijk}^2 = n^2 |\nabla H|^2 = 0.$$

So, it follows that λ_i is constant for every $i = 1, \dots, n$, that is, M^n is an isoparametric hypersurface. Finally, (4-8) must also be an equality, which guarantees that the equality in Lemma 5 occurs. This implies that the hypersurface has exactly two distinct principal curvatures one of which is simple. Therefore, we can apply a classical congruence theorem due to Abe et al. [1987, Theorem 5.1] to conclude that M^n must be one of the two following standard product embeddings into $\mathbb{S}_1^{n+1}$:

(a) $\mathbb{H}^1(r) \times \mathbb{S}^{n-1}(\sqrt{1+r^2})$,
(b) $\mathbb{H}^{n-1}(r) \times \mathbb{S}^1(\sqrt{1+r^2})$,

in either case with a positive radius $r > 0$. In case (a), for a given radius $r > 0$ the standard product embedding $\mathbb{H}^1(r) \times \mathbb{S}^{n-1}(\sqrt{1+r^2}) \hookrightarrow \mathbb{S}_1^{n+1}$ has constant principal curvatures given by

$$\lambda_1 = \frac{\sqrt{1+r^2}}{r}, \quad \lambda_2 = \cdots = \lambda_n = \frac{r}{\sqrt{1+r^2}}.$$

Therefore,

$$nH = \frac{1+nr^2}{r\sqrt{1+r^2}}, \quad S = \frac{1+2r^2+nr^4}{r^2(1+r^2)}, \quad \text{and} \quad |\Phi|^2 = \frac{n-1}{nr^2(1+r^2)},$$

and its constant scalar curvature is given

$$R = \frac{n-2}{n(1+r^2)},$$

which satisfies our hypothesis, since

$$0 < R < \frac{n-2}{n} < 1$$

for every $r > 0$. On the other hand, in case (b) and for a given radius $r > 0$ the standard product embedding $\mathbb{H}^{n-1}(r) \times \mathbb{S}^1(\sqrt{1+r^2}) \hookrightarrow \mathbb{S}_1^{n+1}$ has constant principal curvatures given by

$$\lambda_1 = \cdots = \lambda_{n-1} = \frac{\sqrt{1+r^2}}{r}, \quad \lambda_n = \frac{r}{\sqrt{1+r^2}}.$$

Therefore,

$$nH = \frac{(n-1)+nr^2}{r\sqrt{1+r^2}} \quad S = \frac{n-1+2(n-1)r^2+nr^4}{r^2(1+r^2)}, \quad \text{and} \quad |\Phi|^2 = \frac{n-1}{nr^2(1+r^2)},$$

and its constant scalar curvature is given by

$$R = -\frac{n-2}{nr^2} < 0,$$

which does not satisfy our hypothesis. □

When the spacelike hypersurface has constant scalar curvature (which corresponds to the case $a = 0$), we also have the following consequence of Theorem 8.

Corollary 9. *Let M^n be a complete spacelike hypersurface isometrically immersed in de Sitter space $\mathbb{S}_1^{n+1}$, $n \geq 3$, with constant scalar curvature R satisfying $0 < R < (n-2)/n$. Then*

(i) *either $\sup_M |\Phi|^2 = 0$ and M^n is a totally umbilical hypersurface,*

(ii) *or*

$$\sup_M |\Phi|^2 \geq \beta(n, R) > 0,$$

where

$$\beta(n, R) = \alpha(n, 0, R) = \frac{n(n-1)R^2}{(n-2)(n-2-nR)}.$$

In (ii), a necessary and sufficient condition for equality to hold and the supremum to be attained at some point of M^n is that M^n be isometric to a hyperbolic cylinder $\mathbb{H}^1(r) \times \mathbb{S}^{n-1}(\sqrt{1+r^2})$ of radius $r > 0$.

For the proof of Corollary 9 simply observe that when $a = 0$ (and hence $R = b$) the positive root x_0 of $\varphi_{0,R}(x) = 0$ is given explicitly by

$$x_0^2 = \frac{n(n-1)R^2}{(n-2)(n-2-nR)}.$$

On the other hand, when $n = 2$ it is easy to see that, supposing $0 < b < 1$ and $R \geq b$, the function $\varphi_{a,b}(x)$ is increasing for $x \geq 0$, with $\varphi_{a,b}(x) \geq \varphi_{a,b}(0) > 0$. Therefore in this case, and taking into account that $R = K$ is the Gaussian curvature of M^2, Theorem 8 can be written as follows.

Theorem 10. *Let M^2 be a complete linear Weingarten spacelike surface isometrically immersed in the de Sitter space $\mathbb{S}_1^3$ such that $K = aH + b$ with $0 < b < 1$ and $K \geq b$. If $\sup_M |\Phi|^2 < +\infty$ then M^2 is a totally umbilical surface.*

In other words, taking into account that the only totally umbilical surfaces in $\mathbb{S}_1^3$ having $K > 0$ are the totally umbilical 2-spheres $\mathbb{S}^2(r) \subset \mathbb{S}_1^3$, with radius $r > 1$, Theorem 10 says:

> *The only complete linear Weingarten spacelike surfaces in de Sitter space $\mathbb{S}_1^3$ satisfying $K = aH + b$ with $0 < b < 1$ and $K \geq b$ for which $|\Phi|^2$ is bounded are the totally umbilical 2-spheres.*

The proof of Theorem 10 follows from that of Theorem 8 after observing that when $n=2$ it cannot happen that $0<\sup_M|\Phi|^2<+\infty$ because that would imply $0<\varphi_{a,b}(\sqrt{u^*})\le 0$. Thus if $\sup_M|\Phi|^2<+\infty$ we must have $|\Phi|^2\equiv 0$ and M^2 is a totally umbilical surface.

Finally, when $a=0$ and $n=2$, from Theorem 10 we also obtain the following:

Corollary 11. *The only complete spacelike surfaces in the de Sitter space $\mathbb{S}^3_1$ with constant Gaussian curvature $0<K<1$ for which $|\Phi|^2$ is bounded (or, equivalently, H is bounded) are the totally umbilical 2-spheres $\mathbb{S}^2(r)\subset\mathbb{S}^3_1$, with radius $r>1$.*

5. L-parabolicity of linear Weingarten hypersurfaces

Recall that a Riemannian manifold M^n is said to be parabolic if the constant functions are the only subharmonic functions on M^n which are bounded from above; that is, for a function $u\in\mathcal{C}^2(M)$

$$\Delta u\ge 0\quad\text{and}\quad u\le u^*<+\infty\quad\text{imply } u=\text{constant}.$$

So, considering the Cheng–Yau modified operator L given in (3-3), we say that M^n is L-parabolic if the only solutions of the inequality $L(u)\ge 0$ which are bounded from above are the constant functions. In this setting, and motivated by Theorem 3 in [Alías et al. 2012], we have the following result.

Theorem 12. *Let M^n be a complete linear Weingarten spacelike hypersurface immersed in de Sitter space $\mathbb{S}^{n+1}_1$, $n\ge 3$, such that $R=aH+b$ with $0<b\le R<(n-2)/n$. Suppose that M^n is not totally umbilical. If M^n is L-parabolic, then*

$$\sup_M|\Phi|^2\ge\alpha(n,a,b)>0, \tag{5-1}$$

with equality if and only if M^n is isometric to a hyperbolic cylinder $\mathbb{H}^1(r)\times\mathbb{S}^{n-1}(\sqrt{1+r^2})$ of radius $r>0$.

Proof. If $\sup_M|\Phi|^2=+\infty$ then there is nothing to prove. Since M^n is not totally umbilical, we consider the case that $0<\sup_M|\Phi|^2<+\infty$. In this case, reasoning as in the first part of the proof of Theorem 8, we guarantee that $\sup_M|\Phi|^2\ge\alpha(n,a,b)$. Moreover, if equality holds in (5-1), then we have $\varphi_{a,b}(|\Phi|)\ge 0$ and, consequently, $L(|\Phi|^2)\ge 0$ on M^n. Hence, from the L-parabolicity of M^n we conclude that the function $u=|\Phi|^2$ must be constant and equal to $\alpha(n,a,b)$. Therefore, at this point, we can reason as in the proof of Theorem 8. □

We close our paper establishing the following L-parabolicity criterion.

Proposition 13. *Let M^n be a complete linear Weingarten spacelike hypersurface immersed in $\mathbb{S}^{n+1}_1$ such that $R=aH+b$ with $b\le 1$. In the case $b=1$, assume further that mean curvature function H does not change sign and $b\le R$. If*

$\sup_M |\Phi|^2 < +\infty$ *and, for some reference point* $o \in M^n$,

$$\int_0^{+\infty} \frac{dr}{\mathrm{vol}(\partial B_r)} = +\infty, \tag{5-2}$$

then M^n *is* L*-parabolic. Here* B_r *denotes the geodesic ball of radius* r *in* M^n *centered at the origin* o*.*

Proof. By a straightforward computation we can check from (3-4) that

$$L(u) = \mathrm{div}(P(\nabla u)), \tag{5-3}$$

for any $u \in \mathcal{C}^2(M)$, where P is defined in (3-5).

Now, we consider on M^n the symmetric (0, 2) tensor field h given by $h(X, Y) = \langle PX, Y\rangle$, or, equivalently, $h(\nabla u, \cdot)^\sharp = P(\nabla u)$, where $^\sharp : T^*M \to TM$ denotes the musical isomorphism. Thus, from (5-3) we get

$$L(u) = \mathrm{div}(h(\nabla u, \cdot)^\sharp).$$

On the other hand, as $\sup_M |\Phi|^2 < +\infty$, from (4-16), we have that $\sup_M H < +\infty$. So, we can define a positive continuous function h_+ on $[0, +\infty)$, by

$$h_+(r) = 2n \sup_{\partial B_r} H + (n-1)a. \tag{5-4}$$

Thus, from (5-4) we have

$$h_+(r) = 2n \sup_{\partial B_r} H + (n-1)a \le 2n \sup_M H + (n-1)a < +\infty. \tag{5-5}$$

Hence, from (5-2) and (5-5) we get

$$\int_0^{+\infty} \frac{dr}{h_+(r)\mathrm{vol}(\partial B_r)} = +\infty.$$

Therefore, we can apply Theorem 2.6 of [Pigola et al. 2005] to conclude the proof. □

Acknowledgements

The authors would like to thank the anonymous referee for valuable suggestions and corrections which greatly improved the final version of this paper.

This work is a result of the activity developed within the framework of the Programme in Support of Excellence Groups of the Región de Murcia, Spain, by Fundación Séneca, Science and Technology Agency of the Región de Murcia. The first author was partially supported by MINECO/FEDER project reference MTM2015-65430-P and Fundación Séneca project reference 19901/GERM/15, Spain, and Ciência sem Fronteiras, Programa PVE, project A012/2013, CAPES, Brazil. The second author was partially supported by CNPq, Brazil, grant 303977/2015-9.

References

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

[Akutagawa 1987] K. Akutagawa, "On spacelike hypersurfaces with constant mean curvature in the de Sitter space", *Math. Z.* **196**:1 (1987), 13–19. MR Zbl

[Alencar and do Carmo 1994] H. Alencar and M. do Carmo, "Hypersurfaces with constant mean curvature in spheres", *Proc. Amer. Math. Soc.* **120**:4 (1994), 1223–1229. MR Zbl

[Alías et al. 2012] L. J. Alías, D. Impera, and M. Rigoli, "Spacelike hypersurfaces of constant higher order mean curvature in generalized Robertson–Walker spacetimes", *Math. Proc. Cambridge Philos. Soc.* **152**:2 (2012), 365–383. MR Zbl

[Alías et al. 2016] L. J. Alías, P. Mastrolia, and M. Rigoli, *Maximum principles and geometric applications*, Springer, Cham, 2016. MR Zbl

[Brasil et al. 2003] A. Brasil, Jr., A. G. Colares, and O. Palmas, "Complete spacelike hypersurfaces with constant mean curvature in the de Sitter space: a gap theorem", *Illinois J. Math.* **47**:3 (2003), 847–866. MR Zbl

[Calabi 1970] E. Calabi, "Examples of Bernstein problems for some nonlinear equations", pp. 223–230 in *Global analysis* (Berkeley, CA, 1968), edited by S.-s. Chern and S. Smale, Proc. Sympos. Pure Math. **15**, American Mathematical Society, Providence, RI, 1970. MR Zbl

[Chao 2013] X. Chao, "Complete spacelike hypersurfaces in the de Sitter space", *Osaka J. Math.* **50**:3 (2013), 715–723. MR Zbl

[Cheng 1990] Q. M. Cheng, "Complete space-like hypersurfaces of a de Sitter space with $r = kH$", *Mem. Fac. Sci. Kyushu Univ. Ser. A* **44**:2 (1990), 67–77. MR Zbl

[Cheng and Yau 1976] S. Y. Cheng and S. T. Yau, "Maximal space-like hypersurfaces in the Lorentz–Minkowski spaces", *Ann. of Math.* (2) **104**:3 (1976), 407–419. MR Zbl

[Cheng and Yau 1977] S. Y. Cheng and S. T. Yau, "Hypersurfaces with constant scalar curvature", *Math. Ann.* **225**:3 (1977), 195–204. MR Zbl

[Dajczer and Nomizu 1981] M. Dajczer and K. Nomizu, "On flat surfaces in S_1^3 and H_1^3", pp. 71–108 in *Manifolds and Lie groups* (Notre Dame, IN, 1980), edited by S. Murakami et al., Progr. Math. **14**, Birkhäuser, Boston, 1981. MR Zbl

[Goddard 1977] A. J. Goddard, "Some remarks on the existence of spacelike hypersurfaces of constant mean curvature", *Math. Proc. Cambridge Philos. Soc.* **82**:3 (1977), 489–495. MR Zbl

[Hou and Yang 2010] Z. H. Hou and D. Yang, "Linear Weingarten spacelike hypersurfaces in de Sitter space", *Bull. Belg. Math. Soc. Simon Stevin* **17**:5 (2010), 769–780. MR Zbl

[Li 1997] H. Li, "Global rigidity theorems of hypersurfaces", *Ark. Mat.* **35**:2 (1997), 327–351. MR Zbl

[de Lima and Velásquez 2013] H. F. de Lima and M. A. L. Velásquez, "On the geometry of linear Weingarten spacelike hypersurfaces in the de Sitter space", *Bull. Braz. Math. Soc.* (*N.S.*) **44**:1 (2013), 49–65. MR Zbl

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

[Montiel 1996] S. Montiel, "A characterization of hyperbolic cylinders in the de Sitter space", *Tohoku Math. J.* (2) **48**:1 (1996), 23–31. MR Zbl

[Okumura 1974] M. Okumura, "Hypersurfaces and a pinching problem on the second fundamental tensor", *Amer. J. Math.* **96** (1974), 207–213. MR Zbl

[Pigola et al. 2005] S. Pigola, M. Rigoli, and A. G. Setti, "A Liouville-type result for quasi-linear elliptic equations on complete Riemannian manifolds", *J. Funct. Anal.* **219**:2 (2005), 400–432. MR Zbl

[Ramanathan 1987] J. Ramanathan, "Complete spacelike hypersurfaces of constant mean curvature in de Sitter space", *Indiana Univ. Math. J.* **36**:2 (1987), 349–359. MR Zbl

[Treibergs 1982] A. E. Treibergs, "Entire spacelike hypersurfaces of constant mean curvature in Minkowski space", *Invent. Math.* **66**:1 (1982), 39–56. MR Zbl

Received December 6, 2016.

LUIS J. ALÍAS
DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD DE MURCIA
CAMPUS DE ESPINARDO
MURCIA
SPAIN
ljalias@um.es

HENRIQUE F. DE LIMA
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CAMPINA GRANDE, PB
BRAZIL
henrique@mat.ufcg.edu.br

FÁBIO R. DOS SANTOS
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CAMPINA GRANDE, PB
BRAZIL
fabio@mat.ufcg.edu.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 292, No. 1, 2018
dx.doi.org/10.2140/pjm.2018.292.21

CELLULAR STRUCTURES USING U_q-TILTING MODULES

HENNING HAAHR ANDERSEN,
CATHARINA STROPPEL AND DANIEL TUBBENHAUER

We use the theory of U_q-tilting modules to construct cellular bases for centralizer algebras. Our methods are quite general and work for any quantum group U_q attached to a Cartan matrix and include the nonsemisimple cases for q being a root of unity and ground fields of positive characteristic. Our approach also generalizes to certain categories containing infinite-dimensional modules. As applications, we give a new semisimplicity criterion for centralizer algebras, and we recover the cellularity of several known algebras (with partially new cellular bases) which all fit into our general setup.

Andersen and Tubbenhauer were partially supported by the center of excellence grant "Centre for Quantum Geometry of Moduli Spaces (QGM)" from the "Danish National Research Foundation (DNRF)". Stroppel was supported by the Max-Planck-Gesellschaft. Tubbenhauer was partially supported by a research funding of the "Deutsche Forschungsgemeinschaft (DFG)" during the last part of this work.

MSC2010: primary 17B10, 17B37, 20G05; secondary 16S50, 20C08, 20G42.

Keywords: cellular algebras, cellular bases, tilting modules, quantum enveloping algebras.

1. Introduction

Fix any field $\mathbb{K}$ and set $\mathbb{K}^* = \mathbb{K} - \{0, -1\}$ if $\mathrm{char}(\mathbb{K}) > 2$ and $\mathbb{K}^* = \mathbb{K} - \{0\}$ otherwise. Let $\boldsymbol{U}_q(\mathfrak{g})$ be the quantum group over $\mathbb{K}$ for a fixed, arbitrary parameter $q \in \mathbb{K}^*$ associated to a simple Lie algebra $\mathfrak{g}$. The main result in this paper is the following:

Theorem (Cellularity of endomorphism algebras). *Let T be a $\boldsymbol{U}_q(\mathfrak{g})$-tilting module. Then* $\mathrm{End}_{\boldsymbol{U}_q(\mathfrak{g})}(T)$ *is a cellular algebra* (in the sense of [Graham and Lehrer 1996]).

It is important to note that cellular bases are not unique. In particular, a single algebra can have many cellular bases. As a concrete application, see Section 5B, we construct (several) new cellular bases for the Temperley–Lieb algebra depending on the ground field and the choice of deformation parameter. These bases differ therefore for instance from the construction in [Graham and Lehrer 1996, Section 6] of cellular bases for the Temperley–Lieb algebras. Moreover, we also show that some of our bases for the Temperley–Lieb algebra can be equipped with a $\mathbb{Z}$-grading which is in contrast to Graham and Lehrer's bases. Our bases also depend heavily on the characteristic of $\mathbb{K}$ (and on $q \in \mathbb{K}^*$). Hence, they see more of the characteristic (and parameter) depended representation theory, but are also more difficult to construct explicitly.

We stress that the cellularity itself can be deduced from general theory. Namely, any $\boldsymbol{U}_q(\mathfrak{g})$-tilting module T is a summand of a full $\boldsymbol{U}_q(\mathfrak{g})$-tilting module $\tilde{T}$. By [Ringel 1991, Theorem 6], $\mathrm{End}_{\boldsymbol{U}_q(\mathfrak{g})}(\tilde{T})$ is quasihereditary and comes equipped with an involution as we explain in Section 3C. Thus, it is cellular; see [König and Xi 1998]. By their Theorem 4.3, this induces the cellularity of the idempotent truncation $\mathrm{End}_{\boldsymbol{U}_q(\mathfrak{g})}(T)$. In contrast, our approach provides the existence and a method of construction of many cellular bases. It generalizes to the infinite-dimensional Lie theory situation and has other nice consequences that we will explore in this paper. In particular, our results give a novel semisimplicity criterion for $\mathrm{End}_{\boldsymbol{U}_q(\mathfrak{g})}(T)$; see Theorem 4.13. This together with the Jantzen sum formula gives rise to a new way to obtain semisimplicity criteria for these algebras (we explain and explore this in [Andersen et al. 2017] where we recover semisimplicity criteria for several algebras using the results of this paper). Here a crucial fact is that the tensor product of $\boldsymbol{U}_q$-tilting modules is again a $\boldsymbol{U}_q$-tilting module; see [Paradowski 1994]. This implies that our results also vastly generalize [Westbury 2009] to the nonsemisimple cases (where our main theorem is nontrivial).

The framework. Given any simple, complex Lie algebra $\mathfrak{g}$, we can assign to it a *quantum deformation* $\boldsymbol{U}_v = \boldsymbol{U}_v(\mathfrak{g})$ of its universal enveloping algebra by deforming its Serre presentation. (Here v is a generic parameter and $\boldsymbol{U}_v$ is an $\mathbb{Q}(v)$-algebra.) The representation theory of $\boldsymbol{U}_v$ shares many similarities with that of $\mathfrak{g}$. In particular, the category[1] U_v-**Mod** is semisimple.

But one can spice up the story drastically: the quantum group $U_q = U_q(\mathfrak{g})$ is obtained by specializing v to an arbitrary $q \in \mathbb{K}^*$. In particular, we can take q to be a root of unity[2]. In this case U_q-**Mod** is not semisimple anymore, which makes the representation theory much more interesting. It has many connections and applications in different directions, e.g., the category has a neat combinatorics, is related to the corresponding almost-simple, simply connected algebraic group G over $\mathbb{K}$ with char($\mathbb{K}$) prime; see for example [Andersen et al. 1994] or [Lusztig 1989], to the representation theory of affine Kac–Moody algebras, see [Kazhdan and Lusztig 1994] or [Tanisaki 2004], and to (2+1)-TQFTs and the Witten–Reshetikhin–Turaev invariants of 3-manifolds; see for example [Turaev 2010].

Semisimplicity in light of our main result means the following. If we take $\mathbb{K} = \mathbb{C}$ and $q = \pm 1$, then our result says that the algebra $\mathrm{End}_{U_q}(T)$ is cellular for any U_q-module $T \in U_q$-**Mod** because in this case all U_q-modules are U_q-tilting modules. This is no surprise: when T is a direct sum of simple U_q-modules, then $\mathrm{End}_{U_q}(T)$ is a direct sum of matrix algebras $M_n(\mathbb{K})$. Likewise, for any $\mathbb{K}$, if $q \in \mathbb{K}^* - \{1\}$ is not a root of unity, then U_q-**Mod** is still semisimple and our result is (almost) standard. But even in the semisimple case we can say more: we get an Artin–Wedderburn basis as a cellular basis for $\mathrm{End}_{U_q}(T)$, i.e., a basis realizing the decomposition of $\mathrm{End}_{U_q}(T)$ into its matrix components; see Section 5A.

On the other hand, if $q = 1$ and char($\mathbb{K}$) > 0 or if $q \in \mathbb{K}^*$ is a root of unity, then U_q-**Mod** is far from being semisimple and our result gives many interesting cellular algebras.

For example, if $G = \mathrm{GL}(V)$ for some n-dimensional $\mathbb{K}$-vector space V, then $T = V^{\otimes d}$ is a G-tilting module for any $d \in \mathbb{Z}_{\geq 0}$. By Schur–Weyl duality we have

(1) $$\Phi_{\mathrm{SW}} : \mathbb{K}[S_d] \twoheadrightarrow \mathrm{End}_G(T) \quad \text{and} \quad \Phi_{\mathrm{SW}} : \mathbb{K}[S_d] \xrightarrow{\cong} \mathrm{End}_G(T), \text{ if } n \geq d,$$

where $\mathbb{K}[S_d]$ is the group algebra of the symmetric group S_d in d letters. We can realize this as a special case in our framework by taking $q = 1$, $n \geq d$ and $\mathfrak{g} = \mathfrak{gl}_n$ (although $\mathfrak{gl}_n$ is not a simple, complex Lie algebra, our approach works fine for it as well). On the other hand, by taking q arbitrary in $\mathbb{K}^* - \{1\}$ and $n \geq d$, the group algebra $\mathbb{K}[S_d]$ is replaced by the type A_{d-1} *Iwahori–Hecke algebra* $\mathcal{H}_d(q)$ over $\mathbb{K}$ and our theorem gives cellular bases for this algebra as well. Note that one underlying fact why (1) stays true in the nonsemisimple case is that $\dim(\mathrm{End}_G(T))$ is independent of the characteristic of $\mathbb{K}$ (and of the parameter q in the quantum case), since T is a G-tilting module.

[1]For any algebra A we denote by A-**Mod** the category of finite-dimensional, left A-modules. If not stated otherwise, all modules are assumed to be finite-dimensional, left modules.

[2]In our terminology: The two cases $q = \pm 1$ are special and do not count as roots of unity. Moreover, for technical reasons, we always exclude $q = -1$ in case char($\mathbb{K}$) > 2.

Of course, both $\mathbb{K}[S_d]$ and $\mathcal{H}_d(q)$ are known to be cellular (these cases were one of the main motivations of Graham and Lehrer to introduce the notion of cellular algebras), but the point we want to make is that they fit into our more general framework.

The following known cellularity properties can also be recovered directly from our approach. And moreover in most of the examples we either have no or only some mild restrictions on $\mathbb{K}$ and $q \in \mathbb{K}^*$.

- As sketched above, the algebras $\mathbb{K}[S_d]$ and $\mathcal{H}_d(q)$ and their quotients under Φ_{SW}.
- The *Temperley–Lieb algebras* $\mathcal{TL}_d(\delta)$ introduced in [Temperley and Lieb 1971].
- Other less well-known endomorphism algebras for $\mathfrak{sl}_2$-related tilting modules appearing in more recent work, e.g., [Andersen et al. 2015a], [Andersen and Tubbenhauer 2017] or [Rose and Tubbenhauer 2016].
- *Spider algebras* in the sense of [Kuperberg 1996].
- Quotients of the group algebras of $\mathbb{Z}/r\mathbb{Z} \wr S_d$ and its quantum version $\mathcal{H}_{d,r}(q)$, the *Ariki–Koike algebras* introduced in [Ariki and Koike 1994]. This includes the Ariki–Koike algebras themselves and thus, the Hecke algebras of type B. This also includes Martin and Saleur's [1994] *blob algebras* $\mathcal{BL}_d(q, m)$ and (*quantized*) *rook monoid algebras* (also called *Solomon algebras*) $\mathcal{R}_d(q)$ in the spirit of [Solomon 1990].
- *Brauer algebras* $\mathcal{B}_d(\delta)$, introduced in [Brauer 1937] in the context of classical invariant theory, and related algebras, e.g., the *walled Brauer algebras* $\mathcal{B}_{r,s}(\delta)$ as in [Koike 1989] and [Turaev 1989], and the *Birman–Murakami–Wenzl* algebras $\mathcal{BMW}_d(\delta)$, in the sense of [Birman and Wenzl 1989] and [Murakami 1987].

Our methods also apply for some categories containing infinite-dimensional modules. For example, with a little bit more care, one could allow T to be a not necessarily finite-dimensional $\boldsymbol{U}_q$-tilting module. Moreover, our methods also include the *BGG category* $\mathcal{O}$, its *parabolic subcategories* $\mathcal{O}^{\mathfrak{p}}$ and its quantum cousin $\mathcal{O}_q$ from [Andersen and Mazorchuk 2015]. For example, using the "big projective tilting" in the principal block, we get a cellular basis for the coinvariant algebra of the Weyl group associated to $\mathfrak{g}$. In fact, we get a vast generalization of this, e.g., we can fit *generalized Khovanov arc algebras*, see, e.g., [Brundan and Stroppel 2011a], $\mathfrak{sl}_n$*-web algebras*, see, e.g., [Mackaay et al. 2014], *cyclotomic Khovanov–Lauda and Rouquier algebras* of type A, see [Khovanov and Lauda 2009; 2011] or [Rouquier 2008], for which we obtain cellularity via the connection to cyclotomic quotients of the degenerate affine Hecke algebra, see [Brundan and Kleshchev 2009], cyclotomic $\mathbb{W}_d$*-algebras*, see, e.g., [Ehrig and Stroppel 2013] and *cyclotomic quotients of affine Hecke algebras* $\mathbf{H}^s_{\mathbb{K},d}$, see, e.g., [Rouquier et al. 2016], into our framework as well; see Section 5A. However, we will for simplicity focus on the finite-dimensional world. Here we provide all necessary tools and arguments in great detail, sometimes, for brevity, only in an extra file [Andersen et al. 2015b]. See also Remark 1.

Following Graham and Lehrer's approach, our cellular bases for $\mathrm{End}_{U_q}(T)$ provide also $\mathrm{End}_{U_q}(T)$-cell modules, the classification of simple $\mathrm{End}_{U_q}(T)$-modules, etc. We give an interpretation of this in our setting as well; see Section 4. For instance, we deduce a new criterion for semisimplicity of $\mathrm{End}_{U_q}(T)$; see Theorem 4.13.

Remark 1. Instead of working with the infinite-dimensional algebra U_q, we could also work with a finite-dimensional, quasihereditary algebra (with a suitable anti-involution). By using results summarized in [Donkin 1998, Appendix], our constructions will go through very much in the same spirit as for U_q. However, using U_q has some advantages. For example, we can construct an abundance of cellular bases (for the explicit construction of our basis we need "weight spaces" such that, e.g., (2) or Lemma 3.4 work). Having several cellular bases is certainly an advantage, although calculating these is in general a nontrivial task. (For example, getting an explicit understanding of the endomorphisms giving rise to the cellular basis is a tough challenge, but see [Riche and Williamson 2015] for some crucial steps in this direction.) As a direct consequence of the existence of many cellular bases most of the algebras appearing in our list of examples above can be additionally equipped with a $\mathbb{Z}$-grading. The basis elements from Theorem 3.9 can be chosen such that our approach leads to a $\mathbb{Z}$-graded cellular basis in the sense of [Hu and Mathas 2010]. We make this more precise in the case of the Temperley–Lieb algebras, but one could for instance also recover the $\mathbb{Z}$-graded cellular bases of the Brauer algebras from [Ehrig and Stroppel 2016a] from our approach. We stress that in both cases the cellular bases in [Graham and Lehrer 1996, § 4 and 6] are not $\mathbb{Z}$-graded. To keep the paper within reasonable boundaries, we do not treat the graded setup in detail.

2. Quantum groups, their representations and tilting modules

We briefly recall some facts we need in this paper. Details can be found, e.g., in [Andersen et al. 1991] and [Jantzen 1996], or [Donkin 1998] and [Jantzen 2003]. For notations and arguments adopted to our situation see [Andersen et al. 2015b]. See also [Ringel 1991] and [Donkin 1993] for the classical treatment of tilting modules (in the modular case). As before, we fix a field $\mathbb{K}$ over which we work throughout.

2A. ***The quantum group U_q.*** Let Φ be a finite *root system* in an Euclidean space E. We fix a choice of *positive roots* $\Phi^+ \subset \Phi$ and *simple roots* $\Pi \subset \Phi^+$. We assume that we have n simple roots that we denote by $\alpha_1, \ldots, \alpha_n$. For each $\alpha \in \Phi$, we denote by $\alpha^\vee \in \Phi^\vee$ the corresponding *coroot*. Then $A = (\langle \alpha_i, \alpha_j^\vee \rangle)_{i,j=1}^n$ is called the *Cartan matrix*.

By the set of (*integral*) *weights* we mean $X = \{\lambda \in E \mid \langle \lambda, \alpha_i^\vee \rangle \in \mathbb{Z} \text{ for all } \alpha_i \in \Pi\}$. The *dominant* (*integral*) *weights* X^+ are those $\lambda \in X$ such that $\langle \lambda, \alpha_i^\vee \rangle \geq 0$ for all $\alpha_i \in \Pi$.

Recall that there is a *partial ordering* on X given by $\mu \leq \lambda$ if and only if $\lambda - \mu$ is an $\mathbb{Z}_{\geq 0}$-valued linear combination of the simple roots, that is, $\lambda - \mu = \sum_{i=1}^n a_i \alpha_i$ with $a_i \in \mathbb{Z}_{\geq 0}$.

We denote by $\boldsymbol{U}_q = \boldsymbol{U}_q(\boldsymbol{A})$ the *quantum enveloping algebra* attached to a Cartan matrix $\boldsymbol{A}$ and specialized at $q \in \mathbb{K}^*$, where we follow [Andersen et al. 1991] with our conventions. Note $\boldsymbol{U}_q$ always means the quantum group over $\mathbb{K}$ defined via Lusztig's divided power construction. (Thus, we have generators K_i, E_i and F_i for all $i = 1, \dots, n$ as well as divided power generators.) We have a decomposition $\boldsymbol{U}_q = \boldsymbol{U}_q^- \boldsymbol{U}_q^0 \boldsymbol{U}_q^+$, with subalgebras generated by the F, K and E respectively (and some divided power generators; see, e.g., their Section 1). Note we can recover the generic case $\boldsymbol{U}_v = \boldsymbol{U}_v(\boldsymbol{A})$ by choosing $\mathbb{K} = \mathbb{Q}(v)$ and $q = v$.

It is worth noting that $\boldsymbol{U}_q$ is a Hopf algebra, so its module category is a monoidal category with duals. We denote by $\boldsymbol{U}_q$-**Mod** the category of finite-dimensional $\boldsymbol{U}_q$-modules (of type 1, see [Andersen et al. 1991, Section 1.4]). We consider only such $\boldsymbol{U}_q$-modules in what follows.

Recall that there is a contravariant, character-preserving *duality functor* $\mathcal{D}$ that is defined on the $\mathbb{K}$-vector space level via $\mathcal{D}(M) = M^*$ (the $\mathbb{K}$-linear dual of M) and an action of $\boldsymbol{U}_q$ on $\mathcal{D}(M)$ is defined as follows: Let $\omega : \boldsymbol{U}_q \to \boldsymbol{U}_q$ be the automorphism of $\boldsymbol{U}_q$ which interchanges E_i and F_i and interchanges K_i and K_i^{-1} (see, e.g., [Jantzen 1996, Lemma 4.6], which extends to our setup without difficulties). Then define $uf = m \mapsto f(\omega(S(u))m)$ for $u \in \boldsymbol{U}_q$, $f \in \mathcal{D}(M)$, $m \in M$. Given any $\boldsymbol{U}_q$-homomorphism f between $\boldsymbol{U}_q$-modules, we also write $\mathrm{i}(f) = \mathcal{D}(f)$. This duality gives rise to the involution in our cellular datum from Section 3C.

Assumption 2.1. If q is a root of unity, then, to avoid technicalities, we assume that q is a primitive root of unity of odd order l. A treatment of the even case, that can be used to repeat everything in this paper in the case where l is even, can be found in [Andersen 2003]. Moreover, in case of type G_2 we additionally assume that l is prime to 3.

For each $\lambda \in X^+$ there is a *Weyl* $\boldsymbol{U}_q$*-module* $\Delta_q(\lambda)$ and a *dual Weyl* $\boldsymbol{U}_q$*-module* $\nabla_q(\lambda)$ satisfying $\mathcal{D}(\Delta_q(\lambda)) \cong \nabla_q(\lambda)$. The $\boldsymbol{U}_q$-module $\Delta_q(\lambda)$ has a unique simple head $L_q(\lambda)$ which is the unique simple socle of $\nabla_q(\lambda)$. Thus, there is a (up to scalars) unique $\boldsymbol{U}_q$-homomorphism

$$c^\lambda : \Delta_q(\lambda) \to \nabla_q(\lambda) \quad \text{(mapping head to socle).} \tag{2}$$

This relies on the fact that $\Delta_q(\lambda)$ and $\nabla_q(\lambda)$ both have one-dimensional λ-weight spaces. The same fact implies that $\mathrm{End}_{\boldsymbol{U}_q}(L_q(\lambda)) \cong \mathbb{K}$ for all $\lambda \in X^+$; see [Andersen et al. 1991, Corollary 7.4]. This last property fails for quasihereditary algebras in general when $\mathbb{K}$ is not algebraically closed.

Theorem 2.2 (Ext-vanishing). *We have for all $\lambda, \mu \in X^+$ that*

$$\mathrm{Ext}^i_{\boldsymbol{U}_q}(\Delta_q(\lambda), \nabla_q(\mu)) \cong \begin{cases} \mathbb{K}c^\lambda & \text{if } i = 0 \text{ and } \lambda = \mu, \\ 0 & \text{else}. \end{cases}$$

We have to enlarge the category $\boldsymbol{U}_q$-**Mod** by not necessarily finite-dimensional $\boldsymbol{U}_q$-modules to have enough injectives such that the $\mathrm{Ext}^i_{\boldsymbol{U}_q}$-functors make sense by using q-analogous arguments as in [Jantzen 2003, Part I, Chapter 3]. However, $\boldsymbol{U}_q$-**Mod** has enough injectives in characteristic zero; see [Andersen 1992, Proposition 5.8] for a treatment of the nonsemisimple cases.

Proof. Similar to the modular analog treated in [Jantzen 2003, Proposition II.4.13] (a proof in our notation can be found in [Andersen et al. 2015b]). □

2B. ***Tilting modules and* Ext*-vanishing.*** We say that a $\boldsymbol{U}_q$-module M has a Δ_q-*filtration* if there exists some $k \in \mathbb{Z}_{\geq 0}$ and a finite descending sequence of $\boldsymbol{U}_q$-submodules

$$M = M_0 \supset M_1 \supset \cdots \supset M_{k'} \supset \cdots \supset M_{k-1} \supset M_k = 0,$$

such that $M_{k'}/M_{k'+1} \cong \Delta_q(\lambda_{k'})$ for all $k' = 0, \ldots, k-1$ and some $\lambda_{k'} \in X^+$. A ∇_q-*filtration* is defined similarly, but using a finite ascending sequence of $\boldsymbol{U}_q$-submodules and the $\nabla_q(\lambda)$ instead of the $\Delta_q(\lambda)$. We denote by $(M : \Delta_q(\lambda))$ and $(N : \nabla_q(\lambda))$ the corresponding multiplicities, which are well-defined by Corollary 2.3. Note that a $\boldsymbol{U}_q$-module M has a Δ_q-filtration if and only if its dual $\mathcal{D}(M)$ has a ∇_q-filtration.

A corollary of the Ext-vanishing theorem is the following, whose proof is left to the reader or can be found in [Andersen et al. 2015b]. (Note that the proof of Corollary 2.3 therein gives, in principle, a method to find and construct bases of $\mathrm{Hom}_{\boldsymbol{U}_q}(M, \nabla_q(\lambda))$ and $\mathrm{Hom}_{\boldsymbol{U}_q}(\Delta_q(\lambda), N)$ respectively.)

Corollary 2.3. *Let $M, N \in \boldsymbol{U}_q$-**Mod** and $\lambda \in X^+$. Assume that M has a Δ_q-filtration and N has a ∇_q-filtration. Then*

$$\dim(\mathrm{Hom}_{\boldsymbol{U}_q}(M, \nabla_q(\lambda))) = (M : \Delta_q(\lambda)) \quad \text{and} \quad \dim(\mathrm{Hom}_{\boldsymbol{U}_q}(\Delta_q(\lambda), N)) = (N : \nabla_q(\lambda)).$$

In particular, $(M : \Delta_q(\lambda))$ and $(N : \nabla_q(\lambda))$ are independent of the choice of filtrations.

Proposition 2.4 (Donkin's Ext-criteria). *The following are equivalent*:

(a) *An $M \in \boldsymbol{U}_q$-**Mod** has a Δ_q-filtration (resp. $N \in \boldsymbol{U}_q$-**Mod** has a ∇_q-filtration).*

(b) *We have $\mathrm{Ext}^i_{\boldsymbol{U}_q}(M, \nabla_q(\lambda)) = 0$ (resp. $\mathrm{Ext}^i_{\boldsymbol{U}_q}(\Delta_q(\lambda), N) = 0$) for all $\lambda \in X^+$ and all $i > 0$.*

(c) *We have $\mathrm{Ext}^1_{\boldsymbol{U}_q}(M, \nabla_q(\lambda)) = 0$ (resp. $\mathrm{Ext}^1_{\boldsymbol{U}_q}(\Delta_q(\lambda), N) = 0$) for all $\lambda \in X^+$.*

Proof. As in [Jantzen 2003, Proposition II.4.16]. A proof in our notation can be found in [Andersen et al. 2015b]. $\square$

A $\boldsymbol{U}_q$-module T which has both, a Δ_q- and a ∇_q-filtration, is called a *$\boldsymbol{U}_q$-tilting module*. Following [Donkin 1993], we are now ready to define the *category of $\boldsymbol{U}_q$-tilting modules* that we denote by $\mathcal{T}$. This category is our main object of study.

Definition 2.5 (Category of $\boldsymbol{U}_q$-tilting modules). The category $\mathcal{T}$ is the full subcategory of $\boldsymbol{U}_q$-**Mod** whose objects are given by all $\boldsymbol{U}_q$-tilting modules.

From Proposition 2.4 we obtain directly an important statement.

Corollary 2.6. *Let $T \in \boldsymbol{U}_q$-**Mod**. Then*

$$T \in \mathcal{T} \quad \textit{if and only if} \quad \mathrm{Ext}^1_{\boldsymbol{U}_q}(T, \nabla_q(\lambda)) = 0 = \mathrm{Ext}^1_{\boldsymbol{U}_q}(\Delta_q(\lambda), T) \quad \textit{for all } \lambda \in X^+.$$

When $T \in \mathcal{T}$, the corresponding higher Ext*-groups vanish as well.*

The indecomposable $\boldsymbol{U}_q$-modules in $\mathcal{T}$, that we denote by $T_q(\lambda)$, are indexed by $\lambda \in X^+$. The $\boldsymbol{U}_q$-tilting module $T_q(\lambda)$ is determined by the property that it is indecomposable with λ as its unique maximal weight. In fact, $(T_q(\lambda) : \Delta_q(\lambda)) = 1$, and $(T_q(\lambda) : \Delta_q(\mu)) \neq 0$ only if $\mu \leq \lambda$. (Dually for ∇_q-filtrations.)

Note that the duality functor $\mathcal{D}$ from above restricts to $\mathcal{T}$. Moreover, as a consequence of the classification of indecomposable $\boldsymbol{U}_q$-modules in $\mathcal{T}$, we have $\mathcal{D}(T) \cong T$ for $T \in \mathcal{T}$. In particular, we have for all $\lambda \in X^+$ that

$$(T : \Delta_q(\lambda)) = \dim(\mathrm{Hom}_{\boldsymbol{U}_q}(T, \nabla_q(\lambda))) = \dim(\mathrm{Hom}_{\boldsymbol{U}_q}(\Delta_q(\lambda), T)) = (T : \nabla_q(\lambda)).$$

It is known that $\mathcal{T}$ is a Krull–Schmidt category, closed under finite direct sums, taking summands and finite tensor products (the latter is a nontrivial fact, see [Paradowski 1994, Theorem 3.3]).

For a fixed $\lambda \in X^+$ we have $\boldsymbol{U}_q$-homomorphisms

$$\Delta_q(\lambda) \xhookrightarrow{\iota^\lambda} T_q(\lambda) \xrightarrow{\pi^\lambda} \nabla_q(\lambda),$$

where ι^λ is the inclusion of the first $\boldsymbol{U}_q$-submodule in a Δ_q-filtration of $T_q(\lambda)$ and π^λ is the surjection onto the last quotient in a ∇_q-filtration of $T_q(\lambda)$. Note that these are only defined up to scalars and we fix scalars in the following such that $\pi^\lambda \circ \iota^\lambda = c^\lambda$ (where c^λ is again the $\boldsymbol{U}_q$-homomorphism from (2)).

Remark 2. Let $T \in \mathcal{T}$. An easy argument (based on Theorem 2.2) shows the following crucial fact:

$$\begin{aligned} &\mathrm{Ext}^1_{\boldsymbol{U}_q}(\Delta_q(\lambda), T) = 0 = \mathrm{Ext}^1_{\boldsymbol{U}_q}(T, \nabla_q(\lambda)) \\ \Rightarrow\ &\mathrm{Ext}^1_{\boldsymbol{U}_q}(\mathrm{coker}(\iota^\lambda), T) = 0 = \mathrm{Ext}^1_{\boldsymbol{U}_q}(T, \ker(\pi^\lambda)) \end{aligned} \tag{3}$$

for all $\lambda \in X^+$. Consequently, we see that any $\boldsymbol{U}_q$-homomorphism $g : \Delta_q(\lambda) \to T$ extends to a $\boldsymbol{U}_q$-homomorphism $\bar{g} : T_q(\lambda) \to T$ whereas any $\boldsymbol{U}_q$-homomorphism $f : T \to \nabla_q(\lambda)$ factors through $T_q(\lambda)$ via some $\bar{f} : T \to T_q(\lambda)$.

Remark 3. In [Andersen et al. 2015b] it is described in detail how to compute $(T_q(\lambda) : \Delta_q(\mu))$ for $\lambda, \mu \in X^+$. This can be done algorithmically in case q is a complex, primitive l-th root of unity, i.e., one can use Soergel's version of the *affine parabolic Kazhdan–Lusztig polynomials*. For brevity, we do not recall the definition of these polynomials here, but refer to [Soergel 1997, Section 3] where the relevant polynomials are denoted $n_{y,x}$ (and where all the other relevant notions are defined). The main point for us is the following theorem due to Soergel [1998, Theorem 5.12] (see also [Soergel 1997, Conjecture 7.1]): Suppose $\mathbb{K} = \mathbb{C}$ and q is a complex, primitive l-th root of unity. For each pair $\lambda, \mu \in X^+$ with λ being an l-regular $\boldsymbol{U}_q$-weight (that is, $T_q(\lambda)$ belongs to a regular block of $\mathcal{T}$) we have (with $n_{\mu\lambda}$ equal to the relevant $n_{y,x}$)

$$(T_q(\lambda) : \Delta_q(\mu)) = n_{\mu\lambda}(1) = (T_q(\lambda) : \nabla_q(\mu)).$$

From this one obtains a method to find the indecomposable summands of $\boldsymbol{U}_q$-tilting modules with known characters (e.g., tensor products of minuscule representations).

3. Cellular structures on endomorphism algebras

In this section we give our construction of cellular bases for endomorphism rings $\mathrm{End}_{\boldsymbol{U}_q}(T)$ of $\boldsymbol{U}_q$-tilting modules T and prove our main result, that is, Theorem 3.9.

The main tool is Theorem 3.1. The proof of the latter needs several ingredients which we establish in the form of separate lemmas collected in Section 3B.

3A. ***The basis theorem.*** As before, we consider the category $\boldsymbol{U}_q$-**Mod**. Moreover, we fix two $\boldsymbol{U}_q$-modules M, N, where we assume that M has a Δ_q-filtration and N has a ∇_q-filtration. Then, by Corollary 2.3, we have

$$\dim(\mathrm{Hom}_{\boldsymbol{U}_q}(M, N)) = \sum_{\lambda \in X^+} (M : \Delta_q(\lambda))(N : \nabla_q(\lambda)). \tag{4}$$

We point out that the sum in (4) is actually finite since $(M : \Delta_q(\lambda)) \neq 0$ for only a finite number of $\lambda \in X^+$. (Dually, $(N : \nabla_q(\lambda)) \neq 0$ for only finitely many $\lambda \in X^+$.)

Given $\lambda \in X^+$, we define for $(N : \nabla_q(\lambda)) > 0$, respectively for $(M : \Delta_q(\lambda)) > 0$, the two sets

$$\mathcal{I}^\lambda = \{1, \ldots, (N : \nabla_q(\lambda))\} \quad \text{and} \quad \mathcal{J}^\lambda = \{1, \ldots, (M : \Delta_q(\lambda))\}.$$

By convention, $\mathcal{I}^\lambda = \varnothing$ and $\mathcal{J}^\lambda = \varnothing$ if $(N : \nabla_q(\lambda)) = 0$, respectively if $(M : \Delta_q(\lambda)) = 0$.

We can fix a basis of $\mathrm{Hom}_{\boldsymbol{U}_q}(M, \nabla_q(\lambda))$ indexed by $\mathcal{J}^\lambda$. We denote this fixed basis by $F^\lambda = \{f_j^\lambda : M \to \nabla_q(\lambda) \mid j \in \mathcal{J}^\lambda\}$. By Proposition 2.4 and (3), we see that all elements of F^λ factor through the $\boldsymbol{U}_q$-tilting module $T_q(\lambda)$, i.e., we have commuting diagrams

$$\begin{tikzcd} M \arrow[r, dotted, "\exists\, \overline{f}_j^{\lambda}"] \arrow[dr, "f_j^{\lambda}"'] & T_q(\lambda) \arrow[d, two heads, "\pi^{\lambda}"] \\ & \nabla_q(\lambda). \end{tikzcd}$$

We call $\overline{f}_j^{\lambda}$ a *lift* of f_j^λ. (Note that a lift $\overline{f}_j^{\lambda}$ is not unique.) Dually, we can choose a basis of $\mathrm{Hom}_{\boldsymbol{U}_q}(\Delta_q(\lambda), N)$ as $G^\lambda = \{g_i^\lambda : \Delta_q(\lambda) \to N \mid i \in \mathcal{I}^\lambda\}$, which extends to give (a nonunique) lift $\bar{g}_i^\lambda : T_q(\lambda) \to N$ such that $\bar{g}_i^\lambda \circ \iota^\lambda = g_i^\lambda$ for all $i \in \mathcal{I}^\lambda$.

We can use this setup to define a basis for $\mathrm{Hom}_{\boldsymbol{U}_q}(M, N)$ which, when $M = N$, turns out to be a cellular basis; see Theorem 3.9. For each $\lambda \in X^+$ and all $i \in \mathcal{I}^\lambda$, $j \in \mathcal{J}^\lambda$ set

$$c_{ij}^\lambda = \bar{g}_i^\lambda \circ \overline{f}_j^{\lambda} \in \mathrm{Hom}_{\boldsymbol{U}_q}(M, N).$$

Our main result here is now the following.

Theorem 3.1 (Basis theorem). *For any choice of F^λ and G^λ as above and any choice of lifts of the f_j^λ and the g_i^λ (for all $\lambda \in X^+$), the set*

$$GF = \{c_{ij}^\lambda \mid \lambda \in X^+,\ i \in \mathcal{I}^\lambda,\ j \in \mathcal{J}^\lambda\}$$

is a basis of $\mathrm{Hom}_{\boldsymbol{U}_q}(M, N)$.

Proof. This follows from Proposition 3.3 combined with Lemmas 3.6 and 3.7 from below. □

The basis GF for $\mathrm{Hom}_{\boldsymbol{U}_q}(M, N)$ can be illustrated in a commuting diagram as

$$\begin{tikzcd} & \Delta_q(\lambda) \arrow[d, hook, "\iota^{\lambda}"'] \arrow[dr, "g_i^{\lambda}"] & \\ M \arrow[r, dotted, "\overline{f}_j^{\lambda}"] \arrow[dr, "f_j^{\lambda}"'] & T_q(\lambda) \arrow[r, dotted, "\bar{g}_i^{\lambda}"'] \arrow[d, two heads, "\pi^{\lambda}"] & N \\ & \nabla_q(\lambda) & \end{tikzcd}$$

Since $\boldsymbol{U}_q$-tilting modules have both a Δ_q- and a ∇_q-filtration, we get as an immediate consequence a key result for our purposes.

Corollary 3.2. *Let $T \in \mathcal{T}$. Then GF is, for any choices involved, a basis of* $\mathrm{End}_{\boldsymbol{U}_q}(T)$.

3B. ***Proof of the basis theorem.*** We first show that, given lifts $\overline{f}_j{}^{\lambda}$, there is a consistent choice of lifts $\bar{g}_i^{\lambda}$ such that GF is a basis of $\mathrm{Hom}_{U_q}(M, N)$.

Proposition 3.3 (Basis theorem — dependent version). *For any choice of F^{λ} and any choice of lifts of the f_j^{λ} (for all $\lambda \in X^+$) there exist a choice of a basis G^{λ} and a choice of lifts of the g_i^{λ} such that $GF = \{c_{ij}^{\lambda} \mid \lambda \in X^+,\ i \in \mathcal{I}^{\lambda},\ j \in \mathcal{J}^{\lambda}\}$ is a basis of* $\mathrm{Hom}_{U_q}(M, N)$.

The corresponding statement with the roles of the f and the g swapped clearly holds as well.

Proof. We will construct GF inductively. For this purpose, let

$$0 = N_0 \subset N_1 \subset \cdots \subset N_{k-1} \subset N_k = N$$

be a ∇_q-filtration of N, i.e., $N_{k'+1}/N_{k'} \cong \nabla_q(\lambda_{k'})$ for some $\lambda_{k'} \in X^+$ and all $k' = 0, \ldots, k-1$.

Let $k = 1$ and $\lambda_1 = \lambda$. Then $N_1 = \nabla_q(\lambda)$ and $\{c^{\lambda} : \Delta_q(\lambda) \to \nabla_q(\lambda)\}$ gives a basis of $\mathrm{Hom}_{U_q}(\Delta_q(\lambda), \nabla_q(\lambda))$, where c^{λ} is again the U_q-homomorphism chosen in (2). Set $g_1^{\lambda} = c^{\lambda}$ and observe that $\bar{g}_1^{\lambda} = \pi^{\lambda}$ satisfies $\bar{g}_1^{\lambda} \circ \iota^{\lambda} = g_1^{\lambda}$. Thus, we have a basis and a corresponding lift. This clearly gives a basis of $\mathrm{Hom}_{U_q}(M, N_1)$, since, by assumption, we have that F^{λ} gives a basis of $\mathrm{Hom}_{U_q}(M, \nabla_q(\lambda))$ and $\pi^{\lambda} \circ \overline{f}_j{}^{\lambda} = f_j^{\lambda}$.

Hence, it remains to consider the case $k > 1$. Set $\lambda_k = \lambda$ and observe that we have a short exact sequence of the form

$$0 \longrightarrow N_{k-1} \xhookrightarrow{\text{inc}} N_k \xrightarrow{\text{pro}} \nabla_q(\lambda) \longrightarrow 0. \tag{5}$$

By Theorem 2.2 (and the usual implication as in (3)) this leads to a short exact sequence

$$0 \to \mathrm{Hom}_{U_q}(M, N_{k-1}) \xhookrightarrow{\text{inc}_*} \mathrm{Hom}_{U_q}(M, N_k) \xrightarrow{\text{pro}_*} \mathrm{Hom}_{U_q}(M, \nabla_q(\lambda)) \to 0. \tag{6}$$

By induction, we get from (6) for all $\mu \in X^+$ a basis of $\mathrm{Hom}_{U_q}(\Delta_q(\mu), N_{k-1})$ consisting of the g_i^{μ} with lifts $\bar{g}_i^{\mu}$ such that

$$\{c_{ij}^{\mu} = \bar{g}_i^{\mu} \circ \overline{f}_j{}^{\mu} \mid \mu \in X^+,\ i \in \mathcal{I}_{k-1}^{\mu},\ j \in \mathcal{J}^{\mu}\} \tag{7}$$

is a basis of $\mathrm{Hom}_{U_q}(M, N_{k-1})$ (here we use $\mathcal{I}_{k-1}^{\mu} = \{1, \ldots, (N_{k-1} : \nabla_q(\mu))\}$). We define $g_i^{\mu}(N_k) = \mathrm{inc} \circ g_i^{\mu}$ and $\bar{g}_i^{\mu}(N_k) = \mathrm{inc} \circ \bar{g}_i^{\mu}$ for each $\mu \in X^+$ and each $i \in \mathcal{I}_{k-1}^{\mu}$.

We now have to consider two cases, namely $\lambda \neq \mu$ and $\lambda = \mu$. In the first case we see that $\mathrm{Hom}_{U_q}(\Delta_q(\mu), \nabla_q(\lambda)) = 0$, so that, by using (5) and the usual implication from (3),

$$\mathrm{Hom}_{U_q}(\Delta_q(\mu), N_{k-1}) \cong \mathrm{Hom}_{U_q}(\Delta_q(\mu), N_k).$$

Thus, our basis from (7) gives a basis of $\mathrm{Hom}_{\boldsymbol{U}_q}(\Delta_q(\mu), N_k)$ and also gives the corresponding lifts. On the other hand, if $\lambda = \mu$, then

$$(N_k : \nabla_q(\lambda)) = (N_{k-1} : \nabla_q(\lambda)) + 1.$$

By Theorem 2.2 (and the corresponding implication as in (3)), we can choose $g^\lambda : \Delta_q(\lambda) \to N_k$ such that $\mathrm{pro} \circ g^\lambda = c^\lambda$. Then any choice of a lift $\bar{g}^\lambda$ of g^λ will satisfy $\mathrm{pro} \circ \bar{g}^\lambda = \pi^\lambda$.

Adjoining g^λ to the basis from (7) gives a basis of $\mathrm{Hom}_{\boldsymbol{U}_q}(\Delta_q(\lambda), N_k)$ which satisfies the lifting property. Note that we know from the case $k = 1$ that

$$\{\mathrm{pro} \circ \bar{g}^\lambda \circ \bar{f}_j{}^\lambda = \pi^\lambda \circ \bar{f}_j{}^\lambda \mid j \in \mathcal{J}^\lambda\}$$

is a basis of $\mathrm{Hom}_{\boldsymbol{U}_q}(M, \nabla_q(\lambda))$. Combining everything: we have that

$$\{c_{ij}^\lambda = \bar{g}_i^\lambda(N_k) \circ \bar{f}_j{}^\lambda \mid \lambda \in X^+,\ i \in \mathcal{I}^\lambda,\ j \in \mathcal{J}^\lambda\}$$

is a basis of $\mathrm{Hom}_{\boldsymbol{U}_q}(M, N_k)$ (by enumerating $\bar{g}^\lambda_{(N:\nabla_q(\lambda))}(N_k) = \bar{g}^\lambda$ in the $\lambda = \mu$ case). □

We assume in the following that we have fixed some choices as in Proposition 3.3.

Let $\lambda \in X^+$. Given $\varphi \in \mathrm{Hom}_{\boldsymbol{U}_q}(M, N)$, we denote by $\varphi_\lambda \in \mathrm{Hom}_{\boldsymbol{U}_q^0}(M_\lambda, N_\lambda)$ the induced $\boldsymbol{U}_q^0$-homomorphism (that is, $\mathbb{K}$-linear maps) between the λ-weight spaces M_λ and N_λ. In addition, we denote by $\mathrm{Hom}_{\mathbb{K}}(M_\lambda, N_\lambda)$ the $\mathbb{K}$-linear maps between these λ-weight spaces.

Lemma 3.4. *For any $\lambda \in X^+$ the induced set $\{(c_{ij}^\lambda)_\lambda \mid c_{ij}^\lambda \in GF\}$ is a linearly independent subset of* $\mathrm{Hom}_{\mathbb{K}}(M_\lambda, N_\lambda)$.

Proof. We proceed as in the proof of Proposition 3.3.

If $N = \nabla_q(\lambda)$ (this was $k = 1$ above), then $c_{1j}^\lambda = \pi^\lambda \circ \bar{f}_j{}^\lambda = f_j^\lambda$ and the c_{1j}^λ form a basis of $\mathrm{Hom}_{\boldsymbol{U}_q}(M, \nabla_q(\lambda))$. By the q-Frobenius reciprocity from [Andersen et al. 1991, Proposition 1.17] we have

$$\mathrm{Hom}_{\boldsymbol{U}_q}(M, \nabla_q(\lambda)) \cong \mathrm{Hom}_{\boldsymbol{U}_q^- \boldsymbol{U}_q^0}(M, \mathbb{K}_\lambda) \subset \mathrm{Hom}_{\boldsymbol{U}_q^0}(M, \mathbb{K}_\lambda) = \mathrm{Hom}_{\mathbb{K}}(M_\lambda, \mathbb{K}).$$

Hence, because $N_\lambda = \mathbb{K}$ in this case, we have the base of the induction.

Assume now $k > 1$. The construction of the set $\{c_{ij}^\mu(N_k)\}_{\mu,i,j}$ in the proof of Proposition 3.3 shows that it consists of two separate parts: one being the basis from (7) coming from a basis for $\mathrm{Hom}_{\boldsymbol{U}_q}(M, N_{k-1})$ and the second part (which only occurs when $\lambda = \mu$) coming from a basis from $\mathrm{Hom}_{\boldsymbol{U}_q}(\Delta_q(\lambda), N_k)$.

By (6) there is a short exact sequence

$$0 \to \mathrm{Hom}_{\mathbb{K}}(M_\lambda, (N_{k-1})_\lambda) \xhookrightarrow{\mathrm{inc}_*} \mathrm{Hom}_{\mathbb{K}}(M_\lambda, (N_k)_\lambda) \xrightarrow{\mathrm{pro}_*} \mathrm{Hom}_{\mathbb{K}}(M_\lambda, \mathbb{K}) \to 0.$$

Thus, we can proceed as in the proof of Proposition 3.3. □

We need another piece of notation: we define for each $\lambda \in X^+$

$$\mathrm{Hom}_{\boldsymbol{U}_q}(M,N)^{\leq\lambda} = \{\varphi \in \mathrm{Hom}_{\boldsymbol{U}_q}(M,N) \mid \varphi_\mu = 0 \text{ unless } \mu \leq \lambda\}.$$

In words: a $\boldsymbol{U}_q$-homomorphism $\varphi \in \mathrm{Hom}_{\boldsymbol{U}_q}(M,N)$ belongs to $\mathrm{Hom}_{\boldsymbol{U}_q}(M,N)^{\leq\lambda}$ if and only if φ vanishes on all $\boldsymbol{U}_q$-weight spaces M_μ with $\mu \not\leq \lambda$. In addition to the notation above, we use the evident notation $\mathrm{Hom}_{\boldsymbol{U}_q}(M,N)^{<\lambda}$. We arrive at the following:

Lemma 3.5. *For any fixed $\lambda \in X^+$ the sets*

$$\{c_{ij}^\mu \mid c_{ij}^\mu \in GF,\ \mu \leq \lambda\} \quad \textit{and} \quad \{c_{ij}^\mu \mid c_{ij}^\mu \in GF,\ \mu < \lambda\}$$

are bases of $\mathrm{Hom}_{\boldsymbol{U}_q}(M,N)^{\leq\lambda}$ *and* $\mathrm{Hom}_{\boldsymbol{U}_q}(M,N)^{<\lambda}$ *respectively.*

Proof. As c_{ij}^μ factors through $T_q(\mu)$ and $T_q(\mu)_\nu = 0$ unless $\nu \leq \mu$ (which follows using the classification of indecomposable $\boldsymbol{U}_q$-tilting modules), we see that $(c_{ij}^\mu)_\nu = 0$ unless $\nu \leq \mu$. Moreover, by Lemma 3.4, each $(c_{ij}^\mu)_\mu$ is nonzero. Thus, $c_{ij}^\mu \in \mathrm{Hom}_{\boldsymbol{U}_q}(M,N)^{\leq\lambda}$ if and only if $\mu \leq \lambda$. Now choose any $\varphi \in \mathrm{Hom}_{\boldsymbol{U}_q}(M,N)^{\leq\lambda}$. By Proposition 3.3 we may write

$$\varphi = \sum_{\mu,i,j} a_{ij}^\mu c_{ij}^\mu, \quad a_{ij}^\mu \in \mathbb{K}. \tag{8}$$

Choose $\mu \in X^+$ maximal with the property that there exist $i \in \mathcal{I}^\lambda$, $j \in \mathcal{J}^\lambda$ such that $a_{ij}^\mu \neq 0$.

We claim that $a_{ij}^\nu (c_{ij}^\nu)_\mu = 0$ whenever $\nu \neq \mu$. This is true because, as observed above, $(c_{ij}^\nu)_\mu = 0$ unless $\mu \leq \nu$, and for $\mu < \nu$ we have $a_{ij}^\nu = 0$ by the maximality of μ. We conclude $\varphi_\mu = \sum_{i,j} a_{ij}^\mu (c_{ij}^\mu)_\mu$ and thus, $\varphi_\mu \neq 0$ by Lemma 3.4. Hence, $\mu \leq \lambda$, which gives by (8) that $\varphi \in \mathrm{span}_{\mathbb{K}}\{c_{ij}^\mu \mid c_{ij}^\mu \in GF,\ \mu \leq \lambda\}$ as desired. This shows that $\{c_{ij}^\mu \mid c_{ij}^\mu \in GF,\ \mu \leq \lambda\}$ spans $\mathrm{Hom}_{\boldsymbol{U}_q}(M,N)^{\leq\lambda}$. Since it is clearly a linearly independent set, it is a basis.

The second statement follows analogously, so the details are omitted. □

We need the following two lemmas to prove that all choices in Proposition 3.3 lead to bases of $\mathrm{Hom}_{\boldsymbol{U}_q}(M,N)$. As before we assume that we have, as in Proposition 3.3, constructed $\{g_i^\lambda, i \in \mathcal{I}^\lambda\}$ and the corresponding lifts $\bar{g}_i^\lambda$ for all $\lambda \in X^+$.

Lemma 3.6. *Suppose that we have other $\boldsymbol{U}_q$-homomorphisms $\tilde{g}_i^\lambda : T_q(\lambda) \to N$ such that $\tilde{g}_i^\lambda \circ \iota^\lambda = g_i^\lambda$. Then the following set is also a basis of* $\mathrm{Hom}_{\boldsymbol{U}_q}(M,N)$:

$$\{\tilde{c}_{ij}^\lambda = \tilde{g}_i^\lambda \circ \bar{f}_j^\lambda \mid \lambda \in X^+,\ i \in \mathcal{I}^\lambda,\ j \in \mathcal{J}^\lambda\}.$$

Proof. As $(\bar{g}_i^\lambda - \tilde{g}_i^\lambda) \circ \iota^\lambda = 0$, we see that $\bar{g}_i^\lambda - \tilde{g}_i^\lambda \in \mathrm{Hom}_{\boldsymbol{U}_q}(T_q(\lambda),N)^{<\lambda}$. Hence, we have $c_{ij}^\lambda - \tilde{c}_{ij}^\lambda \in \mathrm{Hom}_{\boldsymbol{U}_q}(M,N)^{<\lambda}$. Thus, by Lemma 3.5, there is a unitriangular change-of-basis matrix between $\{c_{ij}^\lambda\}_{\lambda,i,j}$ and $\{\tilde{c}_{ij}^\lambda\}_{\lambda,i,j}$. □

Now assume that we have chosen another basis $\{h_i^\lambda \mid i \in \mathcal{I}^\lambda\}$ of the spaces $\mathrm{Hom}_{U_q}(\Delta_q(\lambda), N)$ for each $\lambda \in X^+$ and the corresponding lifts $\bar{h}_i^\lambda$ as well.

Lemma 3.7. *The following set is also a basis of* $\mathrm{Hom}_{U_q}(M, N)$:

$$\{d_{ij}^\lambda = \bar{h}_i^\lambda \circ \bar{f}_j^\lambda \mid \lambda \in X^+,\ i \in \mathcal{I}^\lambda,\ j \in \mathcal{J}^\lambda\}.$$

Proof. Write $g_i^\lambda = \sum_{k=1}^{(N:\nabla_q(\lambda))} b_{ik}^\lambda h_k^\lambda$ with $b_{ik}^\lambda \in \mathbb{K}$ and set $\tilde{g}_i^\lambda = \sum_{k=1}^{(N:\nabla_q(\lambda))} b_{ik}^\lambda \bar{h}_k^\lambda$. Then the $\tilde{g}_i^\lambda$ are lifts of the g_i^λ. Hence, by Lemma 3.6, the elements $\tilde{g}_i^\lambda \circ \bar{f}_j^\lambda$ form a basis of $\mathrm{Hom}_{U_q}(M, N)$. Thus, this proves the lemma, since, by construction, $\{d_{ij}^\lambda\}_{\lambda,i,j}$ is related to this basis by the invertible change-of-basis matrix $(b_{ik}^\lambda)_{i,k=1;\lambda\in X^+}^{(N:\nabla_q(\lambda))}$. □

In total, we established Proposition 3.3.

3C. *Cellular structures on endomorphism algebras of U_q-tilting modules.* This section finally contains the statement and proof of our main theorem. We keep on working over a field $\mathbb{K}$ instead of a ring as for example Graham and Lehrer [1996] do. (This avoids technicalities, e.g., the theory of indecomposable U_q-tilting modules over rings is much more subtle than over fields. See, e.g., [Donkin 1993, Remark 1.7].)

Definition 3.8 (Cellular algebras). Suppose A is a finite-dimensional $\mathbb{K}$-algebra. A *cell datum* is an ordered quadruple $(\mathcal{P}, \mathcal{I}, \mathcal{C}, \mathrm{i})$, where $(\mathcal{P}, \leq)$ is a finite poset, $\mathcal{I}^\lambda$ is a finite set for all $\lambda \in \mathcal{P}$, i is a $\mathbb{K}$-linear anti-involution of A and $\mathcal{C}$ is an injection

$$\mathcal{C} : \coprod_{\lambda\in\mathcal{P}} \mathcal{I}^\lambda \times \mathcal{I}^\lambda \to A,\ (i, j) \mapsto c_{ij}^\lambda.$$

The whole data should be such that the c_{ij}^λ form a basis of A with $\mathrm{i}(c_{ij}^\lambda) = c_{ji}^\lambda$ for all $\lambda \in \mathcal{P}$ and all $i, j \in \mathcal{I}^\lambda$. Moreover, for all $a \in A$ and all $\lambda \in \mathcal{P}$ we have

$$ac_{ij}^\lambda = \sum_{k\in\mathcal{I}^\lambda} r_{ik}(a) c_{kj}^\lambda \pmod{A^{<\lambda}} \quad \text{for all } i, j \in \mathcal{I}^\lambda. \tag{9}$$

Here $A^{<\lambda}$ is the subspace of A spanned by the set $\{c_{ij}^\mu \mid \mu < \lambda \text{ and } i, j \in \mathcal{I}(\mu)\}$ and the scalars $r_{ik}(a) \in \mathbb{K}$ are supposed to be independent of j.

An algebra A with such a quadruple is called a *cellular algebra* and the c_{ij}^λ are called a *cellular basis* of A (with respect to the $\mathbb{K}$-linear anti-involution i).

Let us fix $T \in \mathcal{T}$ in the following. We will now construct cellular bases of $\mathrm{End}_{U_q}(T)$ in the semisimple as well as in the nonsemisimple case.

To this end, we need to specify the cell datum. Set

$$(\mathcal{P}, \leq) = (\{\lambda \in X^+ \mid (T : \nabla_q(\lambda)) = (T : \Delta_q(\lambda)) \neq 0\}, \leq),$$

where $\leq$ is the usual partial ordering on X^+; see at the beginning of Section 2A. Note that $\mathcal{P}$ is finite since T is finite-dimensional. Moreover, motivated by Theorem 3.1, for each $\lambda \in \mathcal{P}$ define $\mathcal{I}^\lambda = \{1, \dots, (T : \nabla_q(\lambda))\} = \{1, \dots, (T : \Delta_q(\lambda))\} = \mathcal{J}^\lambda$.

Recalling that we write $\mathrm{i}(\cdot) = \mathcal{D}(\cdot)$ (for $\mathcal{D}$ being the duality functor from Section 2A that exchanges Weyl and dual Weyl U_q-modules and fixes all U_q-tilting modules), the assignment $\mathrm{i} : \mathrm{End}_{U_q}(T) \to \mathrm{End}_{U_q}(T)$, $\phi \mapsto \mathcal{D}(\phi)$ is clearly a $\mathbb{K}$-linear anti-involution. Choose any basis G^λ of $\mathrm{Hom}_{U_q}(\Delta_q(\lambda), T)$ as above and any lifts $\bar{g}_i^\lambda$. Then $\mathrm{i}(G^\lambda)$ is a basis of $\mathrm{Hom}_{U_q}(T, \nabla_q(\lambda))$ and $\mathrm{i}(\bar{g}_i^\lambda)$ is a lift of $\mathrm{i}(g_i^\lambda)$. By Corollary 3.2 we see that

$$\{c_{ij}^\lambda = \bar{g}_i^\lambda \circ \mathrm{i}(\bar{g}_j^\lambda) = \bar{g}_i^\lambda \circ \bar{f}_j^\lambda \mid \lambda \in \mathcal{P},\ i, j \in \mathcal{I}^\lambda\}$$

is a basis of $\mathrm{End}_{U_q}(T)$. Finally let $\mathcal{C} : \mathcal{I}^\lambda \times \mathcal{I}^\lambda \to \mathrm{End}_{U_q}(T)$ be given by $(i, j) \mapsto c_{ij}^\lambda$.

Now we are ready to state and prove our main theorem.

Theorem 3.9 (A cellular basis for $\mathrm{End}_{U_q}(T)$). *The quadruple* $(\mathcal{P}, \mathcal{I}, \mathcal{C}, \mathrm{i})$ *defined above is a cell datum for* $\mathrm{End}_{U_q}(T)$.

Proof. As mentioned above, the sets $\mathcal{P}$ and $\mathcal{I}^\lambda$ are finite for all $\lambda \in \mathcal{P}$. Moreover, i is a $\mathbb{K}$-linear anti-involution of $\mathrm{End}_{U_q}(T)$ and the c_{ij}^λ form a basis of $\mathrm{End}_{U_q}(T)$ by Corollary 3.2. Because the functor $\mathcal{D}(\cdot)$ is contravariant, we see that

$$\mathrm{i}(c_{ij}^\lambda) = \mathrm{i}(\bar{g}_i^\lambda \circ \mathrm{i}(\bar{g}_j^\lambda)) = \bar{g}_j^\lambda \circ \mathrm{i}(\bar{g}_i^\lambda) = c_{ji}^\lambda.$$

Thus, only the condition (9) remains to be proven. For this purpose, let $\varphi \in \mathrm{End}_{U_q}(T)$. Since $\varphi \circ \bar{g}_i^\lambda \circ \iota^\lambda = \varphi \circ g_i^\lambda \in \mathrm{Hom}_{U_q}(\Delta_q(\lambda), T)$, we have coefficients $r_{ik}^\lambda(\varphi) \in \mathbb{K}$ such that

$$\varphi \circ g_i^\lambda = \sum_{k \in \mathcal{I}^\lambda} r_{ik}^\lambda(\varphi) g_k^\lambda, \tag{10}$$

because we know that the g_i^λ form a basis of $\mathrm{Hom}_{U_q}(\Delta_q(\lambda), T)$. But this implies then that $\varphi \circ \bar{g}_i^\lambda - \sum_{k \in \mathcal{I}^\lambda} r_{ik}^\lambda(\varphi) \bar{g}_k^\lambda \in \mathrm{Hom}_{U_q}(T_q(\lambda), T)^{<\lambda}$, so that

$$\varphi \circ \bar{g}_i^\lambda \circ \bar{f}_j^\lambda - \sum_{k \in \mathcal{I}^\lambda} r_{ik}^\lambda(\varphi) \bar{g}_k^\lambda \circ \bar{f}_j^\lambda \in \mathrm{Hom}_{U_q}(T, T)^{<\lambda} = \mathrm{End}_{U_q}(T)^{<\lambda},$$

which proves (9). The theorem follows. □

4. The cellular structure and $\mathrm{End}_{U_q}(T)$-Mod

The goal of this section is to present the representation theory of cellular algebras for $\mathrm{End}_{U_q}(T)$ from the viewpoint of U_q-tilting theory. In fact, most of the results in this section are not new and have been proved for general cellular algebras, see, e.g., [Graham and Lehrer 1996, Section 3]. However, they take a nice and easy form in our setup. The last theorem, the semisimplicity criterion from Theorem 4.13, is new and has potentially many applications; see for example [Andersen et al. 2017].

4A. ***Cell modules for*** $\mathbf{End}_{U_q}(T)$**.** We study now the representation theory for $\mathrm{End}_{U_q}(T)$ via the cellular structure we have found for it. We denote its module category by $\mathrm{End}_{U_q}(T)$-**Mod**.

Definition 4.1 (Cell modules). Let $\lambda \in \mathcal{P}$. The *cell module* associated to λ is the left $\mathrm{End}_{U_q}(T)$-module given by $C(\lambda) = \mathrm{Hom}_{U_q}(\Delta_q(\lambda), T)$. The right $\mathrm{End}_{U_q}(T)$-module given by $C(\lambda)^* = \mathrm{Hom}_{U_q}(T, \nabla_q(\lambda))$ is called the *dual cell module* associated to λ.

The link to the definition of cell modules from [Graham and Lehrer 1996, Definition 2.1] is given via our choice of basis $\{g_i^\lambda\}_{i\in\mathcal{I}^\lambda}$. In this basis the action of $\mathrm{End}_{U_q}(T)$ on $C(\lambda)$ is given by

$$\varphi \circ g_i^\lambda = \sum_{k\in\mathcal{I}^\lambda} r_{ik}^\lambda(\varphi) g_k^\lambda, \quad \varphi \in \mathrm{End}_{U_q}(T); \tag{11}$$

see (10). Here the coefficients are the same as those appearing when we consider the left action of $\mathrm{End}_{U_q}(T)$ on itself in terms of the cellular basis $\{c_{ij}^\lambda\}_{i,j\in\mathcal{I}^\lambda}^{\lambda\in\mathcal{P}}$, that is,

$$\varphi \circ c_{ij}^\lambda = \sum_{k\in\mathcal{I}^\lambda} r_{ik}^\lambda(\varphi) c_{kj}^\lambda \pmod{\mathrm{End}_{U_q}(T)^{<\lambda}}, \quad \varphi \in \mathrm{End}_{U_q}(T). \tag{12}$$

In a completely similar fashion the dual cell module $C(\lambda)^*$ has a basis consisting of $\{f_j^\lambda\}_{j\in\mathcal{I}^\lambda}$ with $f_j^\lambda = \mathrm{i}(g_j^\lambda)$. In this basis the right action of $\mathrm{End}_{U_q}(T)$ is given via

$$f_j^\lambda \circ \varphi = \sum_{k\in\mathcal{I}^\lambda} r_{kj}^\lambda(\mathrm{i}(\varphi)) f_k^\lambda, \quad \varphi \in \mathrm{End}_{U_q}(T). \tag{13}$$

We can use the unique U_q-homomorphism from (2) and the duality functor $\mathcal{D}(\,\cdot\,)$ to define the following *cellular pairing* in the spirit of [Graham and Lehrer 1996, Definition 2.3].

Definition 4.2 (Cellular pairing). Let $\lambda \in \mathcal{P}$. Then we denote by ϑ^λ the $\mathbb{K}$-bilinear form $\vartheta^\lambda : C(\lambda) \otimes C(\lambda) \to \mathbb{K}$ determined by the property

$$\mathrm{i}(h) \circ g = \vartheta^\lambda(g, h) c^\lambda, \quad g, h \in C(\lambda) = \mathrm{Hom}_{U_q}(\Delta_q(\lambda), T).$$

We call ϑ^λ the *cellular pairing* associated to $\lambda \in \mathcal{P}$.

Lemma 4.3. *The cellular pairing ϑ^λ is well-defined*, *symmetric and contravariant.*

Proof. That ϑ^λ is well-defined follows directly from the uniqueness of c^λ.

Applying i to the defining equation of ϑ^λ gives

$$\vartheta^\lambda(g, h)\mathrm{i}(c^\lambda) = \mathrm{i}(\vartheta^\lambda(g, h)c^\lambda) = \mathrm{i}(\mathrm{i}(h) \circ g) = \mathrm{i}(g) \circ h = \vartheta^\lambda(h, g)c^\lambda,$$

and thus, $\vartheta^\lambda(g, h) = \vartheta^\lambda(h, g)$, because $c^\lambda = \mathrm{i}(c^\lambda)$. (Recall that $c^\lambda : \Delta_q(\lambda) \to \nabla_q(\lambda)$ is unique up to scalars. Hence, we can fix scalars accordingly such that $c^\lambda = \mathrm{i}(c^\lambda)$.)

Similarly, contravariance of $\mathcal{D}(\,\cdot\,)$ gives

$$\vartheta^{\lambda}(\varphi \circ g, h) = \vartheta^{\lambda}(g, \mathrm{i}(\varphi) \circ h), \quad \varphi \in \mathrm{End}_{\boldsymbol{U}_q}(T),\ g, h \in C(\lambda),$$

which shows contravariance of the cellular pairing. □

Proposition 4.4. *Let $\lambda \in \mathcal{P}$. Then $T_q(\lambda)$ is a summand of T if and only if $\vartheta^{\lambda} \neq 0$.*

Proof. (See also [Andersen 1997, Proposition 1.5].) Assume $T \cong T_q(\lambda) \oplus \text{rest}$. We denote by $\bar{g} : T_q(\lambda) \to T$ and by $\bar{f} : T \to T_q(\lambda)$ the corresponding inclusion and projection respectively. As usual, set $g = \bar{g} \circ \iota^{\lambda}$ and $f = \pi^{\lambda} \circ \bar{f}$. Then we have $f \circ g : \Delta_q(\lambda) \hookrightarrow T_q(\lambda) \hookrightarrow T \twoheadrightarrow T_q(\lambda) \twoheadrightarrow \nabla_q(\lambda) = c^{\lambda}$ (mapping head to socle), giving $\vartheta^{\lambda}(g, \mathrm{i}(f)) = 1$. This shows that $\vartheta^{\lambda} \neq 0$.

Conversely, assume that there exist $g, h \in C(\lambda)$ with $\vartheta^{\lambda}(g, h) \neq 0$. Then the commuting "bow tie diagram", i.e.,

$$\begin{tikzcd}
\Delta_q(\lambda) \arrow[d, hook, "\iota^{\lambda}"'] \arrow[dr, "g"] & & \\
T_q(\lambda) \arrow[r, dotted, "\bar{g}"'] & T \arrow[r, dotted, "\overline{\mathrm{i}(h)}"] \arrow[dr, "\mathrm{i}(h)"'] & T_q(\lambda), \arrow[d, two heads, "\pi^{\lambda}"] \\
& & \nabla_q(\lambda)
\end{tikzcd}$$

shows that $\overline{\mathrm{i}(h)} \circ \bar{g}$ is nonzero on the λ-weight space of $T_q(\lambda)$, because $\mathrm{i}(h) \circ g = \vartheta^{\lambda}(g, h)c^{\lambda}$. Thus, $\overline{\mathrm{i}(h)} \circ \bar{g}$ must be an isomorphism (because $T_q(\lambda)$ is indecomposable and has therefore only invertible or nilpotent elements in $\mathrm{End}_{\boldsymbol{U}_q}(T_q(\lambda))$) showing that $T \cong T_q(\lambda) \oplus \text{rest}$. □

In view of Proposition 4.4, it makes sense to study the set

$$\mathcal{P}_0 = \{\lambda \in \mathcal{P} \mid \vartheta^{\lambda} \neq 0\} \subset \mathcal{P}. \tag{14}$$

Hence, if $\lambda \in \mathcal{P}_0$, then we have $T \cong T_q(\lambda) \oplus \text{rest}$ for some $\boldsymbol{U}_q$-tilting module called rest. Note also that $\mathrm{End}_{\boldsymbol{U}_q}(T)$ is quasihereditary if and only if $\mathcal{P} = \mathcal{P}_0$, see, e.g., [Graham and Lehrer 1996, Remark 3.10].

4B. *The structure of* $\mathbf{End}_{\boldsymbol{U}_q}(\boldsymbol{T})$ *and its cell modules.* Recall that, for any $\lambda \in \mathcal{P}$, we have that $\mathrm{End}_{\boldsymbol{U}_q}(T)^{\leq\lambda}$ and $\mathrm{End}_{\boldsymbol{U}_q}(T)^{<\lambda}$ are two-sided ideals in $\mathrm{End}_{\boldsymbol{U}_q}(T)$ (this follows from (9) and its right-handed version obtained by applying i), as in any cellular algebra. In our case we can also see this as follows. If $\varphi \in \mathrm{End}_{\boldsymbol{U}_q}(T)^{\leq\lambda}$, then $\varphi_{\mu} = 0$ unless $\mu \leq \lambda$. Hence, for any $\varphi, \psi \in \mathrm{End}_{\boldsymbol{U}_q}(T)$ we have $(\varphi \circ \psi)_{\mu} = \varphi_{\mu} \circ \psi_{\mu} = 0 = \psi_{\mu} \circ \varphi_{\mu} = (\psi \circ \varphi)_{\mu}$ unless $\mu \leq \lambda$. As a consequence, $\mathrm{End}_{\boldsymbol{U}_q}(T)^{\lambda} = \mathrm{End}_{\boldsymbol{U}_q}(T)^{\leq\lambda} / \mathrm{End}_{\boldsymbol{U}_q}(T)^{<\lambda}$ is an $\mathrm{End}_{\boldsymbol{U}_q}(T)$-bimodule.

Recall that, for any $g \in C(\lambda)$ and any $f \in C(\lambda)^*$, we denote by $\bar{g} : T_q(\lambda) \to T$ and $\overline{f} : T \to T_q(\lambda)$ a choice of lifts which satisfy $\bar{g} \circ \iota^\lambda = g$ and $\pi^\lambda \circ \overline{f} = f$, respectively.

Lemma 4.5. *Let $\lambda \in \mathcal{P}$. Then the pairing map*

$$\langle \cdot, \cdot \rangle^\lambda : C(\lambda) \otimes C(\lambda)^* \to \mathrm{End}_{\boldsymbol{U}_q}(T)^\lambda, \quad \langle g, f \rangle^\lambda = \bar{g} \circ \overline{f} + \mathrm{End}_{\boldsymbol{U}_q}(T)^{<\lambda},$$

with $g \in C(\lambda)$, $f \in C(\lambda)^*$ is an isomorphism of $\mathrm{End}_{\boldsymbol{U}_q}(T)$*-bimodules.*

Proof. First we note that $\bar{g} \circ \overline{f} + \mathrm{End}_{\boldsymbol{U}_q}(T)^{<\lambda}$ does not depend on the choices for the lifts $\overline{f}, \bar{g}$, because the change-of-basis matrix from Lemma 3.6 is unitriangular (and works for swapped roles of the f and the g as well). This makes the pairing well-defined.

Note that the pairing $\langle \cdot, \cdot \rangle^\lambda$ takes, by definition, the basis $(g_i^\lambda \otimes f_j^\lambda)_{i,j \in \mathcal{I}^\lambda}$ of $C(\lambda) \otimes C(\lambda)^*$ to the basis $\{c_{ij}^\lambda + \mathrm{End}_{\boldsymbol{U}_q}(T)^{<\lambda}\}_{i,j \in \mathcal{I}^\lambda}$ of $\mathrm{End}_{\boldsymbol{U}_q}(T)^\lambda$ (where the latter is a basis by Lemma 3.5).

So we only need to check that $\langle \varphi \circ g_i^\lambda, f_j^\lambda \circ \psi \rangle^\lambda = \varphi \circ c_{ij}^\lambda \circ \psi \pmod{\mathrm{End}_{\boldsymbol{U}_q}(T)^{<\lambda}}$ for any $\varphi, \psi \in \mathrm{End}_{\boldsymbol{U}_q}(T)$. But this is a direct consequence of (11), (12) and (13). $\square$

The next lemma is straightforward by Lemma 4.5. Details are left to the reader.

Lemma 4.6. *We have the following*:

(a) *There is an isomorphism of $\mathbb{K}$-vector spaces* $\mathrm{End}_{\boldsymbol{U}_q}(T) \cong \bigoplus_{\lambda \in \mathcal{P}} \mathrm{End}_{\boldsymbol{U}_q}(T)^\lambda$.

(b) *If $\varphi \in \mathrm{End}_{\boldsymbol{U}_q}(T)^{\leq \lambda}$, then we have $r_{ik}^\mu(\varphi) = 0$ for all $\mu \not\leq \lambda$, $i, k \in \mathcal{I}(\mu)$. Equivalently*, $\mathrm{End}_{\boldsymbol{U}_q}(T)^{\leq \lambda} C(\mu) = 0$ *unless $\mu \leq \lambda$.*

In the following we assume that $\lambda \in \mathcal{P}_0$ as in (14). Define m_λ via

$$T \cong T_q(\lambda)^{\oplus m_\lambda} \oplus T', \tag{15}$$

where T' is a $\boldsymbol{U}_q$-tilting module containing no summands isomorphic to $T_q(\lambda)$.

Choose now a basis of $C(\lambda) = \mathrm{Hom}_{\boldsymbol{U}_q}(\Delta_q(\lambda), T)$ as follows. For $i = 1, \dots, m_\lambda$, let $\bar{g}_i^\lambda$ be the inclusion of $T_q(\lambda)$ into the i-th summand of $T_q(\lambda)^{\oplus m_\lambda}$, and set $g_i^\lambda = \bar{g}_i^\lambda \circ \iota^\lambda$. Then extend $\{g_1^\lambda, \dots, g_{m_\lambda}^\lambda\}$ to a basis of the cell module $C(\lambda)$ by adding an arbitrary basis of $\mathrm{Hom}_{\boldsymbol{U}_q}(\Delta_q(\lambda), T')$. Thus, in our usual notation, we have $c_{ij}^\lambda = \bar{g}_i^\lambda \circ \overline{f}_j^\lambda$ with $\overline{f}_j^\lambda = \mathrm{i}(\bar{g}_j^\lambda)$.

In particular, $\overline{f}_j^\lambda$ projects onto the j-th summand in $T_q(\lambda)^{\oplus m_\lambda}$ for $j = 1, \dots, m_\lambda$. Thus, the c_{ii}^λ for $i \leq m_\lambda$ are idempotents in $\mathrm{End}_{\boldsymbol{U}_q}(T)$ corresponding to the i-th summand in $T_q(\lambda)^{\oplus m_\lambda}$. Since $\lambda \in \mathcal{P}_0$ (which implies $1 \leq m_\lambda$), c_{11}^λ is always such an idempotent. This is crucial for the following lemma, which will play an important role in the proof of Proposition 4.8.

Lemma 4.7. *In the above notation*,

(a) $c_{i1}^\lambda \circ g_1^\lambda = g_i^\lambda$ *for all $i \in \mathcal{I}^\lambda$*,

(b) $c_{ij}^\lambda \circ g_1^\lambda = 0$ *for all $i, j \in \mathcal{I}^\lambda$ with $j \neq 1$.*

Proof. We have $\bar{f}_1^\lambda \circ g_1^\lambda = \bar{f}_1^\lambda \circ \bar{g}_1^\lambda \circ \iota^\lambda = \iota^\lambda$. This implies $c_{i1}^\lambda \circ g_1^\lambda = \bar{g}_i^\lambda \circ \iota^\lambda = g_i^\lambda$. Next, if $j \neq 1$, then $\bar{f}_j^\lambda \circ g_1^\lambda = 0$, since $\bar{f}_j^\lambda$ is zero on $T_q(\lambda)$. Thus, $c_{ij}^\lambda \circ g_1^\lambda = 0$ for all $i, j \in \mathcal{I}^\lambda$ with $j \neq 1$. □

Proposition 4.8 (Homomorphism criterion). *Let $\lambda \in \mathcal{P}_0$ and fix $M \in \mathrm{End}_{\boldsymbol{U}_q}(T)$-**Mod**. Then there is an isomorphism of $\mathbb{K}$-vector spaces*

$$\mathrm{Hom}_{\mathrm{End}_{\boldsymbol{U}_q}(T)}(C(\lambda), M) \cong \{m \in M \mid \mathrm{End}_{\boldsymbol{U}_q}(T)^{<\lambda} m = 0 \text{ and } c_{11}^\lambda m = m\}. \tag{16}$$

Proof. Let $\psi \in \mathrm{Hom}_{\mathrm{End}_{\boldsymbol{U}_q}(T)}(C(\lambda), M)$. Then $\psi(g_1^\lambda)$ belongs to the right-hand side, because, by (b) of Lemma 4.6, we get $\mathrm{End}_{\boldsymbol{U}_q}(T)^{<\lambda} C(\lambda) = 0$, and we have $c_{11}^\lambda \circ g_1^\lambda = g_1^\lambda$ by (a) of Lemma 4.7. Conversely, if $m \in M$ belongs to the right-hand side in (16), then we may define $\psi \in \mathrm{Hom}_{\mathrm{End}_{\boldsymbol{U}_q}(T)}(C(\lambda), M)$ by $\psi(g_i^\lambda) = c_{i1}^\lambda m$, $i \in \mathcal{I}^\lambda$. Moreover, the fact that this definition gives an $\mathrm{End}_{\boldsymbol{U}_q}(T)$-homomorphism follows from (10), (11) and (12) via a direct computation, since $\mathrm{End}_{\boldsymbol{U}_q}(T)^{<\lambda} m = 0$. Clearly these two operations are mutually inverse. □

Corollary 4.9. *Let $\lambda \in \mathcal{P}_0$. Then $C(\lambda)$ has a unique simple head, denoted by $L(\lambda)$.*

Proof. Set $\mathrm{Rad}(\lambda) = \{g \in C(\lambda) \mid \vartheta^\lambda(g, C(\lambda)) = 0\}$. As the cellular pairing ϑ^λ from Definition 4.2 is contravariant by Lemma 4.3, we see that $\mathrm{Rad}(\lambda)$ is an $\mathrm{End}_{\boldsymbol{U}_q}(T)$-submodule of $C(\lambda)$. Since $\vartheta^\lambda \neq 0$ for $\lambda \in \mathcal{P}_0$, we have $\mathrm{Rad}(\lambda) \subsetneq C(\lambda)$. We claim that $\mathrm{Rad}(\lambda)$ is the unique maximal proper $\mathrm{End}_{\boldsymbol{U}_q}(T)$-submodule of $C(\lambda)$.

Let $g \in C(\lambda) - \mathrm{Rad}(\lambda)$. Moreover, choose $h \in C(\lambda)$ with $\vartheta^\lambda(g, h) = 1$. Then $\mathrm{i}(h) \circ g = c^\lambda$ so that $\overline{\mathrm{i}(h)} \circ g = \iota^\lambda \pmod{\mathrm{End}_{\boldsymbol{U}_q}(T)^{<\lambda}}$. Therefore,

$$g' = \bar{g}' \circ \overline{\mathrm{i}(h)} \circ g \pmod{\mathrm{End}_{\boldsymbol{U}_q}(T)^{<\lambda}} \quad \text{for all } g' \in C(\lambda).$$

This implies $C(\lambda) = \mathrm{End}_{\boldsymbol{U}_q}(T)^{\leq\lambda} g$. Thus, any proper $\mathrm{End}_{\boldsymbol{U}_q}(T)$-submodule of $C(\lambda)$ is contained in $\mathrm{Rad}(\lambda)$ which implies the desired statement. □

Corollary 4.10. *Let $\lambda \in \mathcal{P}_0$, $\mu \in \mathcal{P}$ and assume that $\mathrm{Hom}_{\mathrm{End}_{\boldsymbol{U}_q}(T)}(C(\lambda), M) \neq 0$ for some $\mathrm{End}_{\boldsymbol{U}_q}(T)$-module M isomorphic to a subquotient of $C(\mu)$. Then we have $\mu \leq \lambda$. In particular, all composition factors $L(\lambda)$ of $C(\mu)$ satisfy $\mu \leq \lambda$.*

Proof. By Proposition 4.8 the assumption in the corollary implies the existence of an element $m \in M$ with $c_{11}^\lambda m = m$. But if $\mu \not\leq \lambda$, then c_{11}^λ vanishes on the $\boldsymbol{U}_q$-weight space T_μ and hence, $c_{11}^\lambda g$ kills the highest weight vector in $\Delta_q(\mu)$ for all $g \in C(\mu)$. This makes the existence of such an $m \in M$ impossible unless $\mu \leq \lambda$. □

4C. *Simple* $\mathrm{End}_{\boldsymbol{U}_q}(\boldsymbol{T})$*-modules and semisimplicity of* $\mathrm{End}_{\boldsymbol{U}_q}(\boldsymbol{T})$. Let $\lambda \in \mathcal{P}_0$. Note that Corollary 4.9 shows that $C(\lambda)$ has a unique simple head $L(\lambda)$. We then arrive at the following classification of all simple modules in $\mathrm{End}_{\boldsymbol{U}_q}(T)$-**Mod**.

Theorem 4.11 (Classification of simple $\mathrm{End}_{\boldsymbol{U}_q}(T)$-modules). *The set $\{L(\lambda) \mid \lambda \in \mathcal{P}_0\}$ forms a complete set of pairwise nonisomorphic, simple $\mathrm{End}_{\boldsymbol{U}_q}(T)$-modules.*

Proof. We have to show three statements, namely that the $L(\lambda)$ are simple, that they are pairwise nonisomorphic and that every simple $\mathrm{End}_{U_q}(T)$-module is one of the $L(\lambda)$.

Because the first statement follows directly from the definition of $L(\lambda)$ (see Corollary 4.9), we start by showing the second. Thus, assume that $L(\lambda) \cong L(\mu)$ for some $\lambda, \mu \in \mathcal{P}_0$. Then

$$\mathrm{Hom}_{\mathrm{End}_{U_q}(T)}(C(\lambda), C(\mu)/\operatorname{Rad}(\mu)) \neq 0 \neq \mathrm{Hom}_{\mathrm{End}_{U_q}(T)}(C(\mu), C(\lambda)/\operatorname{Rad}(\lambda)).$$

By Corollary 4.10, we get $\mu \leq \lambda$ and $\lambda \leq \mu$ from the left- and right-hand side. Thus, $\lambda = \mu$.

Suppose that $L \in \mathrm{End}_{U_q}(T)$-**Mod** is simple. Then we can choose $\lambda \in \mathcal{P}$ minimal such that (recall that $\mathrm{End}_{U_q}(T)^{\leq\lambda}$ is a two-sided ideal)

$$\mathrm{End}_{U_q}(T)^{<\lambda} L = 0 \quad \text{and} \quad \mathrm{End}_{U_q}(T)^{\leq\lambda} L = L. \tag{17}$$

We claim that $\lambda \in \mathcal{P}_0$. Indeed, if not, then, by Proposition 4.4, we see that $T_q(\lambda)$ is not a summand of T. Hence, in our usual notation, all $\bar{f}_j^{\lambda} \circ \bar{g}_{i'}^{\lambda}$ vanish on the λ-weight space. It follows that $c_{ij}^{\lambda} c_{i'j'}^{\lambda}$ also vanishes on the λ-weight space for all $i, j, i', j' \in \mathcal{I}^{\lambda}$. This means that we have $\mathrm{End}_{U_q}(T)^{\leq\lambda}\mathrm{End}_{U_q}(T)^{\leq\lambda} \subset \mathrm{End}_{U_q}(T)^{<\lambda}$ making (17) impossible.

For $\lambda \in \mathcal{P}_0$ we see by Lemma 4.7 that

$$c_{i1}^{\lambda} c_{1j}^{\lambda} = c_{ij}^{\lambda} \pmod{\mathrm{End}_{U_q}(T)^{<\lambda}}. \tag{18}$$

Hence, by (17), there exist $i, j \in \mathcal{I}^{\lambda}$ such that $c_{ij}^{\lambda} L \neq 0$. By (18) we also have $c_{i1}^{\lambda} L \neq 0 \neq c_{1j}^{\lambda} L$. This in turn (again by (18)) ensures that $c_{11}^{\lambda} L \neq 0$. Take then $m \in c_{11}^{\lambda} L - \{0\}$ and observe that $c_{11}^{\lambda} m = m$. Hence, by Proposition 4.8, there is a nonzero $\mathrm{End}_{U_q}(T)$-homomorphism $C(\lambda) \to L$. The conclusion follows now from Corollary 4.9. □

Recall from Section 4B the notation m_{λ} (the multiplicity of $T_q(\lambda)$ in T) and the choice of basis for $C(\lambda)$ (in the paragraphs before Lemma 4.7). Then we get the following connection between the decomposition of T as in (15) and the simple $\mathrm{End}_{U_q}(T)$-modules $L(\lambda)$.

Theorem 4.12 (Dimension formula). *If* $\lambda \in \mathcal{P}_0$, *then* $\dim(L(\lambda)) = m_{\lambda}$.

Note that this result is implicit in [Graham and Lehrer 1996] and has also been observed in, e.g., [Erdmann 1995] and [Soergel 1999].

Proof. We use the notation from Section 4B. Since T' has no summands isomorphic to $T_q(\lambda)$, we see that $\mathrm{Hom}_{U_q}(\Delta_q(\lambda), T') \subset \operatorname{Rad}(\lambda)$ (see the proof of Corollary 4.9). On the other hand, $g_i^{\lambda} \notin \operatorname{Rad}(\lambda)$ for $1 \leq i \leq m_{\lambda}$ because for these i we have $f_i^{\lambda} \circ g_i^{\lambda} = c^{\lambda}$ by construction. Thus, the statement follows. □

Theorem 4.13 (Semisimplicity criterion). *The cellular algebra* $\mathrm{End}_{U_q}(T)$ *is semisimple if and only if* T *is a semisimple* U_q*-module.*

Proof. Note that the $T_q(\lambda)$ are simple if and only if $T_q(\lambda) \cong \Delta_q(\lambda) \cong L_q(\lambda) \cong \nabla_q(\lambda)$. Hence, T is semisimple as a U_q-module if and only if $T = \bigoplus_{\lambda \in \mathcal{P}_0} \Delta_q(\lambda)^{\oplus m_\lambda}$ with m_λ as in Section 4B above.

Thus, we see that, if T decomposes into simple U_q-modules, then $\mathrm{End}_{U_q}(T)$ is semisimple by the Artin–Wedderburn theorem (since $\mathrm{End}_{U_q}(T)$ will decompose into a direct sum of matrix algebras in this case).

On the other hand, if $\mathrm{End}_{U_q}(T)$ is semisimple, then we know by Corollary 4.9 that the cell modules $C(\lambda)$ are simple, i.e., $C(\lambda) = L(\lambda)$ for all $\lambda \in \mathcal{P}_0$. Then

$$T \cong \bigoplus_{\lambda \in \mathcal{P}_0} T_q(\lambda)^{\oplus m_\lambda}, \tag{19}$$

$$\text{with } m_\lambda = \dim(L(\lambda)) = \dim(C(\lambda)) = \dim(\mathrm{Hom}_{U_q}(\Delta_q(\lambda), T))$$

by Theorem 4.12. Assume now that there exists a summand $T_q(\lambda')$ of T as in (19) with $T_q(\lambda') \ncong \Delta_q(\lambda')$ and choose $\lambda' \in \mathcal{P}_0$ minimal with this property.

Then there exists a $\mu < \lambda'$ such that $\mathrm{Hom}_{U_q}(\Delta_q(\mu), T_q(\lambda')) \neq 0$. Choose also μ minimal among those. By our usual construction this then gives in turn a nonzero U_q-homomorphism $\bar{g} \circ \bar{f} : T_q(\lambda') \to T_q(\mu) \to T_q(\lambda')$. By (19), we can extend $\bar{g} \circ \bar{f}$ to an element of $\mathrm{End}_{U_q}(T)$ by defining it to be zero on all other summands.

Clearly, by construction, $(\bar{g} \circ \bar{f})C(\mu') = 0$ for $\mu' \in \mathcal{P}_0$ with $\mu' \neq \lambda'$ and $\mu' \not\leq \mu$. If $\mu' \leq \mu$, then consider $\varphi \in C(\mu')$. Then $(\bar{g} \circ \bar{f}) \circ \varphi = 0$ unless φ has some nonzero component $\varphi' : \Delta_q(\mu') \to T_q(\lambda')$. This forces $\mu' = \mu$ by minimality of μ. But since $\Delta_q(\mu') \cong T_q(\mu')$, by minimality of λ', we conclude that $\bar{f} \circ \varphi = 0$ (otherwise $T_q(\mu')$ would be a summand of $T_q(\lambda')$).

Hence, the nonzero element $\bar{g} \circ \bar{f} \in \mathrm{End}_{U_q}(T)$ kills all $C(\mu')$ for $\mu' \in \mathcal{P}_0$. This contradicts the semisimplicity of $\mathrm{End}_{U_q}(T)$: as noted above, $C(\lambda) = L(\lambda)$ for all $\lambda \in \mathcal{P}_0$, which implies $\mathrm{End}_{U_q}(T) \cong \bigoplus_{\lambda \in \mathcal{P}_0} C(\lambda)^{\oplus k_\lambda}$ for some $k_\lambda \in \mathbb{Z}_{\geq 0}$. □

5. Cellular structures: examples and applications

In this section we provide many examples of cellular algebras arising from our main theorem. This includes several renowned examples where cellularity is known (but usually proved case by case spread over the literature and with cellular bases which differ in general from ours), and also new ones. In the first section we give a full treatment of the semisimple case and describe how to obtain all the examples from the introduction using our methods. In the second section we focus on the Temperley–Lieb algebras $\mathcal{TL}_d(\delta)$ and give a detailed account how to apply our results to these.

5A. ***Cellular structures using U_q-tilting modules: several examples.*** In the following let ω_i for $i = 1, \dots, n$ denote the fundamental weights (of the corresponding type).

5A.1. *The semisimple case.* Suppose the category $\boldsymbol{U}_q$-**Mod** is semisimple, that is, q is not a root of unity in $\mathbb{K}^* - \{1\}$ or $q = \pm 1 \in \mathbb{K}$ with $\mathrm{char}(\mathbb{K}) = 0$.

In this case, $\boldsymbol{\mathcal{T}} = \boldsymbol{U}_q$-**Mod** and any $T \in \boldsymbol{\mathcal{T}}$ has a decomposition

$$T \cong \bigoplus_{\lambda \in X^+} \Delta_q(\lambda)^{\oplus m_\lambda} \quad \text{with the multiplicities } m_\lambda = (T : \Delta_q(\lambda)).$$

This induces an Artin–Wedderburn decomposition

$$\mathrm{End}_{\boldsymbol{U}_q}(T) \cong \bigoplus_{\lambda \in X^+} M_{m_\lambda}(\mathbb{K}) \tag{20}$$

into matrix algebras. A natural choice of basis for $\mathrm{Hom}_{\boldsymbol{U}_q}(\Delta_q(\lambda), T)$ is

$$G^\lambda = \{g_1^\lambda, \dots, g_{m_\lambda}^\lambda \mid g_i^\lambda : \Delta_q(\lambda) \hookrightarrow T \text{ is the inclusion into the } i\text{-th summand}\}.$$

Then our cellular basis consisting of the c_{ij}^λ as in Section 3C (no lifting is needed in this case) is an Artin–Wedderburn basis, that is, a basis of $\mathrm{End}_{\boldsymbol{U}_q}(T)$ that realizes the decomposition as in (20) in the following sense. The basis element c_{ij}^λ is the matrix $\boldsymbol{E}_{ij}^\lambda$ (in the λ-summand on the right-hand side in (20)) which has all entries zero except one entry equals 1 in the i-th row and j-th column. Note that, as expected in this case, $\mathrm{End}_{\boldsymbol{U}_q}(T)$ has, by the Theorems 4.11 and 4.12, one simple $\mathrm{End}_{\boldsymbol{U}_q}(T)$-module $L(\lambda)$ of dimension m_λ for all summands $\Delta_q(\lambda)$ of T.

5A.2. *The symmetric group and the Iwahori–Hecke algebra.* Let us fix $d \in \mathbb{Z}_{\geq 0}$ and let us denote by S_d the symmetric group in d letters and by $\mathcal{H}_d(q)$ its associated Iwahori–Hecke algebra. We note that $\mathbb{K}[S_d] \cong \mathcal{H}_d(1)$. Moreover, let $\boldsymbol{U}_q = \boldsymbol{U}_q(\mathfrak{gl}_n)$. The vector representation of $\boldsymbol{U}_q$, which we denote by $V = \mathbb{K}^n = \Delta_q(\omega_1)$, is a $\boldsymbol{U}_q$-tilting module (since ω_1 is minimal in X^+). Set $T = V^{\otimes d}$, which is again a $\boldsymbol{U}_q$-tilting module. Quantum Schur–Weyl duality (see [Du et al. 1998, Theorem 6.3] for surjectivity and use Ext-vanishing for the fact that $\dim(\mathrm{End}_{\boldsymbol{U}_q}(T))$ is obtained via base change from $\mathbb{Z}[v, v^{-1}]$ to $\mathbb{K}$ for all $\mathbb{K}$ and $q \in \mathbb{K}^*$) states that

$$\Phi_{q\mathrm{SW}} : \mathcal{H}_d(q) \twoheadrightarrow \mathrm{End}_{\boldsymbol{U}_q}(T) \quad \text{and} \quad \Phi_{q\mathrm{SW}} : \mathcal{H}_d(q) \xrightarrow{\cong} \mathrm{End}_{\boldsymbol{U}_q}(T), \text{ if } n \geq d. \tag{21}$$

Thus, our main result implies that $\mathcal{H}_d(q)$, and in particular $\mathbb{K}[S_d]$, are cellular for any $q \in \mathbb{K}^*$ and any field $\mathbb{K}$ (by taking $n \geq d$).

In this case the cell modules for $\mathrm{End}_{\boldsymbol{U}_q}(T)$ are usually called Specht modules $S_{\mathbb{K}}^\lambda$ and our Theorem 4.12 gives the following:

- If $q = 1$ and $\mathrm{char}(\mathbb{K}) = 0$, then the dimension $\dim(S^{\lambda}_{\mathbb{K}})$ is equal to the multiplicity of the simple U_1-module $\Delta_1(\lambda) \cong L_1(\lambda)$ in $V^{\otimes d}$ for all $\lambda \in \mathcal{P}^0$. These numbers are given by known formulas (e.g., the hook length formula).

- If $q = 1$ and $\mathrm{char}(\mathbb{K}) > 0$, then the dimension of the simple head of $S^{\lambda}_{\mathbb{K}}$, usually denoted $D^{\lambda}_{\mathbb{K}}$, is the multiplicity with which $T_1(\lambda)$ occurs as a summand in $V^{\otimes d}$ for all $\lambda \in \mathcal{P}_0$, see also [Erdmann 1995]. It is a wide open problem to determine these numbers. (See however [Riche and Williamson 2015].)

- If q is a complex, primitive root of unity, then we can compute the dimension of the simple $\mathcal{H}_d(q)$-modules by using the algorithm as in [Andersen et al. 2015b]. In particular, this connects with the LLT algorithm from [Lascoux et al. 1996].

- If q is a root of unity and $\mathbb{K}$ is arbitrary, then not much is known. Still, our methods apply and we get a way to calculate the dimensions of the simple $\mathcal{H}_d(q)$-modules, if we can decompose T into its indecomposable summands.

5A.3. *The Temperley–Lieb algebra and other $\mathfrak{sl}_2$-related algebras.* Let U_q be $U_q(\mathfrak{sl}_2)$ and let T be as in Section 5A.2 with $n = 2$. For any $d \in \mathbb{Z}_{\geq 0}$ we have $\mathcal{TL}_d(\delta) \cong \mathrm{End}_{U_q}(T)$ by Schur–Weyl duality, where $\mathcal{TL}_d(\delta)$ is the Temperley–Lieb algebra in d-strands with parameter $\delta = q + q^{-1}$. This works for all $\mathbb{K}$ and all $q \in \mathbb{K}^*$ (this can be deduced from, for example, [Du et al. 1998, Theorem 6.3]). Hence, $\mathcal{TL}_d(\delta)$ is always cellular. We discuss this case in more detail in Section 5B.

Furthermore, if we are in the semisimple case, then $\Delta_q(i)$ is a U_q-tilting module for all $i \in \mathbb{Z}_{\geq 0}$ and so is $T = \Delta_q(i_1) \otimes \cdots \otimes \Delta_q(i_d)$. Thus, we obtain that $\mathrm{End}_{U_q}(T)$ is cellular.

The algebra $\mathrm{End}_{U_q}(T)$ is known to give a diagrammatic presentation of the (tensor) category of U_q-modules, see [Rose and Tubbenhauer 2016], and can be used to define the colored Jones polynomial.

If $q \in \mathbb{K}$ is a root of unity and l is the order of q^2, then, for any $0 \leq i < l$, $\Delta_q(i)$ is a U_q-tilting module (since it is simple) and so is $T = \Delta_q(i)^{\otimes d}$. The endomorphism algebra $\mathrm{End}_{U_q}(T)$ is cellular. This reproves parts of [Andersen et al. 2015a, Theorem 1.1] using our general approach.

In characteristic 0, another family of U_q-tilting modules was studied in [Andersen and Tubbenhauer 2017]. For any $d \in \mathbb{Z}_{\geq 0}$, fix any $\lambda_0 \in \{0, \ldots, l-2\}$ and consider $T = T_q(\lambda_0) \oplus \cdots \oplus T_q(\lambda_d)$, where λ_k is the unique integer $\lambda_k \in \{kl, \ldots, (k+1)l-2\}$ linked to λ_0. We again obtain that $\mathrm{End}_{U_q}(T)$ is cellular. Note that $\mathrm{End}_{U_q}(T)$ can be identified with a so-called (type A) zig-zag algebra A_d, see [Andersen and Tubbenhauer 2017, Proposition 3.9], introduced in [Huerfano and Khovanov 2001]. These algebras are naturally graded making $\mathrm{End}_{U_q}(T)$ into a graded cellular algebra in the sense of [Hu and Mathas 2010] and are special examples arising from the family of generalized Khovanov arc algebras whose cellularity is studied in [Brundan and Stroppel 2011a].

5A.4. *Spider algebras.* Let $\boldsymbol{U}_q = \boldsymbol{U}_q(\mathfrak{sl}_n)$ (or, alternatively, $\boldsymbol{U}_q(\mathfrak{gl}_n)$). One has for any $q \in \mathbb{K}^*$ that all $\boldsymbol{U}_q$-representations $\Delta_q(\omega_i)$ are $\boldsymbol{U}_q$-tilting modules (because the ω_i are minimal in X^+). Hence, for any $k_i \in \{1, \dots, n-1\}$, $T = \Delta_q(\omega_{k_1}) \otimes \cdots \otimes \Delta_q(\omega_{k_d})$ is a $\boldsymbol{U}_q$-tilting module. Thus, $\mathrm{End}_{\boldsymbol{U}_q}(T)$ is cellular. These algebras are related to type A_{n-1} spider algebras as in [Kuperberg 1996], are connected to the Reshetikhin–Turaev $\mathfrak{sl}_n$-link polynomials and give a diagrammatic description of the representation theory of $\mathfrak{sl}_n$, see [Cautis et al. 2014], providing a link from our work to low-dimensional topology and diagrammatic algebra. Note that cellular bases (which, in this case, coincide with our cellular bases) of these were found in [Elias 2015, Theorem 2.57].

More general: In any type we have that $\Delta_q(\lambda)$ are $\boldsymbol{U}_q(\mathfrak{g})$-tilting modules for minuscule $\lambda \in X^+$, see [Jantzen 2003, Part II, Chapter 2, Section 15]. Moreover, if q is a root of unity "of order l big enough" (ensuring that the ω_i are in the closure of the fundamental alcove), then the $\Delta_q(\omega_i)$ are $\boldsymbol{U}_q(\mathfrak{g})$-tilting modules by the linkage principle; see [Andersen 2003, Corollaries 4.4 and 4.6]. So in these cases we can generalize the above results to other types.

Still more generally, we may take (for any type and any $q \in \mathbb{K}^*$) arbitrary $\lambda_j \in X^+$ (for $j = 1, \dots, d$) and obtain a cellular structure on $\mathrm{End}_{\boldsymbol{U}_q}(T)$ for $T = T_q(\lambda_1) \otimes \cdots \otimes T_q(\lambda_d)$.

5A.5. *The Ariki–Koike algebra and related algebras.* Take $\mathfrak{g} = \mathfrak{gl}_{m_1} \oplus \cdots \oplus \mathfrak{gl}_{m_r}$ (which can be easily fit into our context) with $m_1 + \cdots + m_r = m$ and let V be the vector representation of $\boldsymbol{U}_1(\mathfrak{gl}_m)$ restricted to $\boldsymbol{U}_1 = \boldsymbol{U}_1(\mathfrak{g})$. This is again a $\boldsymbol{U}_1$-tilting module and so is $T = V^{\otimes d}$. Then we have a cyclotomic analog of (21), namely

$$
\begin{aligned}
&\Phi_{\mathrm{cl}} : \mathbb{C}[\mathbb{Z}/r\mathbb{Z} \wr S_d] \twoheadrightarrow \mathrm{End}_{\boldsymbol{U}_1}(T) \qquad \text{and} \\
&\Phi_{\mathrm{cl}} : \mathbb{C}[\mathbb{Z}/r\mathbb{Z} \wr S_d] \xrightarrow{\cong} \mathrm{End}_{\boldsymbol{U}_1}(T), \quad \text{if } m \geq d,
\end{aligned}
\tag{22}
$$

where $\mathbb{C}[\mathbb{Z}/r\mathbb{Z} \wr S_d]$ is the group algebra of the complex reflection group $\mathbb{Z}/r\mathbb{Z} \wr S_d \cong (\mathbb{Z}/r\mathbb{Z})^d \rtimes S_d$; see [Mazorchuk and Stroppel 2016, Theorem 9]. Thus, we can apply our main theorem and obtain a cellular basis for these quotients of $\mathbb{C}[\mathbb{Z}/r\mathbb{Z} \wr S_d]$. If $m \geq d$, then (22) is an isomorphism (see Lemma 11 of [loc. cit.]) and we obtain that $\mathbb{C}[\mathbb{Z}/r\mathbb{Z} \wr S_d]$ itself is a cellular algebra for all r, d. In the extremal case $m_1 = m - 1$ and $m_2 = 1$, the resulting quotient of (22) is known as Solomon's algebra introduced in [Solomon 1990] (also called the algebra of the inverse semigroup or the rook monoid algebra) and we obtain that Solomon's algebra is cellular. In the extremal case $m_1 = m_2 = 1$, the resulting quotient is a specialization of the blob algebra $\mathcal{BL}_d(1, 2)$ (in the notation used in [Ryom-Hansen 2010]). To see this, note that both algebras are quotients of $\mathbb{C}[\mathbb{Z}/r\mathbb{Z} \wr S_d]$. The kernel of the quotient to $\mathcal{BL}_d(1, 2)$

is described explicitly by Ryom-Hansen [2010, (1)] and is by [Mazorchuk and Stroppel 2016, Lemma 11] contained in the kernel of Φ_{cl} from (22). Since both algebras have the same dimensions, they are isomorphic.

Let $U_q = U_q(\mathfrak{g})$. We get in the quantized case (for $q \in \mathbb{C} - \{0\}$ not a root of unity)

$$\text{(23)}\quad \Phi_{q\mathrm{cl}} : \mathcal{H}_{d,r}(q) \twoheadrightarrow \mathrm{End}_{U_q}(T) \quad \text{and} \quad \Phi_{q\mathrm{cl}} : \mathcal{H}_{d,r}(q) \xrightarrow{\cong} \mathrm{End}_{U_q}(T), \text{ if } m \geq d,$$

where $\mathcal{H}_{d,r}(q)$ is the Ariki–Koike algebra introduced in [Ariki and Koike 1994]. A proof of (23) can for example be found in [Sakamoto and Shoji 1999, Theorem 4.1]. Thus, as before, our main theorem applies and we obtain: the Ariki–Koike algebra $\mathcal{H}_{d,r}(q)$ is cellular (by taking $m \geq d$), the quantized rook monoid algebra $\mathcal{R}_d(q)$ from [Halverson and Ram 2001] is cellular and the blob algebra $\mathcal{BL}_d(q, m)$ is cellular (which follows as above). Note that the cellularity of $\mathcal{H}_{d,r}(q)$ was obtained in [Dipper et al. 1998] and the cellularity of the quantum rook monoid algebras and of the blob algebra can be found in [Paget 2006] and in [Ryom-Hansen 2012] respectively.

In fact, (23) is still true in the nonsemisimple cases, see [Hu and Stoll 2004, Theorem 1.10 and Lemma 2.12], as long as $\mathbb{K}$ satisfies a certain separation condition (which implies that the algebra in question has the right dimension, see [Ariki 1999]). Again, our main theorem applies.

5A.6. *The Brauer algebras and related algebras.* Consider $U_q = U_q(\mathfrak{g})$ where $\mathfrak{g}$ is either an orthogonal $g = \mathfrak{o}_{2n}$ and $g = \mathfrak{o}_{2n+1}$ or the symplectic $g = \mathfrak{sp}_{2n}$ Lie algebra. Let $V = \Delta_q(\omega_1)$ be the quantized version of the corresponding vector representation. In both cases, V is a U_q-tilting module (for type B and $q = 1$ this requires $\mathrm{char}(\mathbb{K}) \neq 2$, see [Jantzen 1973, Page 20]) and hence, so is $T = V^{\otimes d}$. We first take $q = 1$ and set $\delta = 2n$ in case $\mathfrak{g} = \mathfrak{o}_{2n}$, and $\delta = 2n+1$ in case $\mathfrak{g} = \mathfrak{o}_{2n+1}$ and $\delta = -2n$ in case $\mathfrak{g} = \mathfrak{sp}_{2n}$ respectively. Then (see [Dipper et al. 2008, Theorem 1.4] and [Doty and Hu 2009, Theorem 1.2] for infinite $\mathbb{K}$, or [Ehrig and Stroppel 2016b, Theorem 5.5] for $\mathbb{K} = \mathbb{C}$)

$$\text{(24)}\qquad \Phi_{\mathrm{Br}} : \mathcal{B}_d(\delta) \twoheadrightarrow \mathrm{End}_{U_1}(T) \quad \text{and} \quad \Phi_{\mathrm{Br}} : \mathcal{B}_d(\delta) \xrightarrow{\cong} \mathrm{End}_{U_1}(T), \text{ if } n > d,$$

where $\mathcal{B}_d(\delta)$ is the Brauer algebra in d strands (for $\mathfrak{g} \neq \mathfrak{o}_{2n}$ the isomorphism in (24) already holds for $n = d$). Thus, we get cellularity of $\mathcal{B}_d(\delta)$ by observing that in characteristic p we can always assume that n is large because $\mathcal{B}_d(\delta) = \mathcal{B}_d(\delta + p)$.

Similarly, let $U_q = U_q(\mathfrak{gl}_n)$, $q \in \mathbb{K}^*$ arbitrary, and $T = \Delta_q(\omega_1)^{\otimes r} \otimes \Delta_q(\omega_{n-1})^{\otimes s}$. By [Dipper et al. 2014, Theorem 7.1 and Corollary 7.2] we have

$$\text{(25)}\qquad \begin{aligned} &\Phi_{\mathrm{wBr}} : \mathcal{B}^n_{r,s}([n]) \twoheadrightarrow \mathrm{End}_{U_q}(T) \quad \text{and} \\ &\Phi_{\mathrm{wBr}} : \mathcal{B}^n_{r,s}([n]) \xrightarrow{\cong} \mathrm{End}_{U_q}(T), \quad \text{if } n \geq r+s. \end{aligned}$$

Here $\mathcal{B}^n_{r,s}([n])$ is the quantized walled Brauer algebra for $[n]=q^{1-n}+\cdots+q^{n-1}$. Since T is a $\boldsymbol{U}_q$-tilting module, we get from (25) cellularity of $\mathcal{B}^n_{r,s}([n])$ and of its quotients under Φ_{wBr}.

The walled Brauer algebra $\mathcal{B}^n_{r,s}(\delta)$ over $\mathbb{K}=\mathbb{C}$ for arbitrary parameter $\delta\in\mathbb{Z}$ appears as the centralizer of $\mathrm{End}_{\mathfrak{gl}(m|n)}(T)$ for $T=V^{\otimes r}\otimes(V^*)^{\otimes s}$ where V is the vector representation of the superalgebra $\mathfrak{gl}(m|n)$ with $\delta=m-n$. That is, we have

$$(26)\qquad \begin{aligned} &\Phi_{\mathrm{s}}:\mathcal{B}^n_{r,s}(\delta)\twoheadrightarrow \mathrm{End}_{\mathfrak{gl}(m|n)}(T) && \text{and}\\ &\Phi_{\mathrm{s}}:\mathcal{B}^n_{r,s}(\delta)\xrightarrow{\cong} \mathrm{End}_{\mathfrak{gl}(m|n)}(T), && \text{if } (m+1)(n+1)\geq r+s,\end{aligned}$$

see [Brundan and Stroppel 2012a, Theorem 7.8]. It can be shown that T is a $\mathfrak{gl}(m|n)$-tilting module and thus, our main theorem applies and hence, by (26), $\mathcal{B}^n_{r,s}(\delta)$ is cellular. Similarly for the quantized version.

Quantizing the Brauer case, taking $q\in\mathbb{K}^*$, $\mathfrak{g}$, $V=\Delta_q(\omega_1)$ and T as before (without the restriction $\mathrm{char}(\mathbb{K})\neq 2$ for type B) gives us a cellular structure on $\mathrm{End}_{\boldsymbol{U}_q}(T)$. The algebra $\mathrm{End}_{\boldsymbol{U}_q}(T)$ is a quotient of the Birman–Murakami–Wenzl algebra $\mathcal{BMW}_d(\delta)$ (for appropriate parameters), see [Lehrer and Zhang 2012, (9.6)] for the orthogonal case (which works for any $q\in\mathbb{C}-\{0,\pm1\}$) and [Hu 2011, Theorem 1.5] for the symplectic case (which works for any $q\in\mathbb{K}^*-\{1\}$ and infinite $\mathbb{K}$). Again, taking $n\geq d$ (or $n>d$), we recover the cellularity of $\mathcal{BMW}_d(\delta)$.

5A.7. *Infinite-dimensional modules — highest weight categories.* Observe that our main theorem does not use the specific properties of $\boldsymbol{U}_q$-**Mod**, but works for any $\mathrm{End}_{A\text{-}\mathbf{Mod}}(T)$ where T is an A-tilting module for some finite-dimensional, quasi-hereditary algebra A over $\mathbb{K}$ or $T\in\mathcal{C}$ for some highest weight category $\mathcal{C}$ in the sense of [Cline et al. 1988]. For the explicit construction of our basis we however need a notion like "weight spaces" such that Lemma 3.4 makes sense.

The most famous example of such a category is the BGG category $\mathcal{O}=\mathcal{O}(\mathfrak{g})$ attached to a complex semisimple or reductive Lie algebra $\mathfrak{g}$ with a corresponding Cartan $\mathfrak{h}$ and fixed Borel subalgebra $\mathfrak{b}$. We denote by $\Delta(\lambda)\in\mathcal{O}$ the Verma module attached to $\lambda\in\mathfrak{h}^*$. In the same vein, pick a parabolic $\mathfrak{p}\supset\mathfrak{b}$ and denote for any $\mathfrak{p}$-dominant weight λ the corresponding parabolic Verma module by $\Delta^{\mathfrak{p}}(\lambda)$. It is the unique quotient of the Verma module $\Delta(\lambda)$ which is locally $\mathfrak{p}$-finite, i.e., contained in the parabolic category $\mathcal{O}^{\mathfrak{p}}=\mathcal{O}^{\mathfrak{p}}(\mathfrak{g})\subset\mathcal{O}$ (see, e.g., [Humphreys 2008]).

There is a contravariant, character preserving duality functor ${}^{\vee}:\mathcal{O}^{\mathfrak{p}}\to\mathcal{O}^{\mathfrak{p}}$ which allows us to set $\nabla^{\mathfrak{p}}(\lambda)=\Delta^{\mathfrak{p}}(\lambda)^{\vee}$. Hence, we can play the same game again since the $\mathcal{O}$-tilting theory works in a very similar fashion as for $\boldsymbol{U}_q$-**Mod** (see [Humphreys 2008, Chapter 11] and the references therein). In particular, we have indecomposable $\mathcal{O}$-tilting modules $T(\lambda)$ for any $\lambda\in\mathfrak{h}^*$. Similarly for $\mathcal{O}^{\mathfrak{p}}$ giving an indecomposable $\mathcal{O}^{\mathfrak{p}}$-tilting module $T(\lambda)$ for any $\mathfrak{p}$-dominant $\lambda\in\mathfrak{h}^*$.

We give a few examples where our approach leads to cellular structures on interesting algebras. For this purpose, let $\mathfrak{p} = \mathfrak{b}$ and $\lambda = 0$. Then $T(0)$ has Verma factors of the form $\Delta(w.0)$ (for $w \in W$, where W is the Weyl group associated to $\mathfrak{g}$). Each of these appears with multiplicity 1. Hence, $\dim(\mathrm{End}_{\mathcal{O}}(T(0))) = |W|$ by the analog of (4). Then we have $\mathrm{End}_{\mathcal{O}}(T(0)) \cong S(\mathfrak{h}^*)/S_+^W$. The algebra $S(\mathfrak{h}^*)/S_+^W$ is called the coinvariant algebra. (For the notation, the conventions and the result see [Soergel 1990] — this is Soergel's famous Endomorphismensatz.) Hence, our main theorem implies that $S(\mathfrak{h}^*)/S_+^W$ is cellular, which is no big surprise since all finite-dimensional, commutative algebras are cellular, see [König and Xi 1998, Proposition 3.5].

There is also a quantum version of this result: replace $\mathcal{O}$ by its quantum cousin $\mathcal{O}_q$ from [Andersen and Mazorchuk 2015] (which is the analog of $\mathcal{O}$ for $U_q(\mathfrak{g})$). This works over any field $\mathbb{K}$ with $\mathrm{char}(\mathbb{K}) = 0$ and any $q \in \mathbb{K}^* - \{1\}$ (which can be deduced from Section 6 therein). There is furthermore a characteristic p version of this result: Consider the G-tilting module $T(p\rho)$ in the category of finite-dimensional G-modules (here G is an almost simple, simply connected algebraic group over $\mathbb{K}$ with $\mathrm{char}(\mathbb{K}) = p$). Its endomorphism algebra is isomorphic to the corresponding coinvariant algebra over $\mathbb{K}$, see [Andersen et al. 1994, Proposition 19.8].

Returning to $\mathbb{K} = \mathbb{C}$, we can generalize the example of the coinvariant algebra. To this end, note that, if T is an $\mathcal{O}^{\mathfrak{p}}$-tilting module, then so is $T \otimes M$ for any finite-dimensional $\mathfrak{g}$-module M; see [Humphreys 2008, Proposition 11.1 and Section 11.8] (and the references therein). Thus, $\mathrm{End}_{\mathcal{O}^{\mathfrak{p}}}(T \otimes M)$ is cellular by our main theorem.

A special case is $\mathfrak{g}$ is of classical type, $T = \Delta^{\mathfrak{p}}(\lambda)$ is simple (hence, $\mathcal{O}^{\mathfrak{p}}$-tilting), V is the vector representation of $\mathfrak{g}$ and $M = V^{\otimes d}$. Let first $\mathfrak{g} = \mathfrak{gl}_n$ with standard Borel $\mathfrak{b}$ and parabolic $\mathfrak{p}$ of block size $(n_1, \ldots, n_\ell)$. Then one can find a certain $\mathfrak{p}$-dominant weight λ_{I}, called Irving-weight, such that $T = \Delta^{\mathfrak{p}}(\lambda_{\mathrm{I}})$ is $\mathcal{O}^{\mathfrak{p}}$-tilting. Moreover, $\mathrm{End}_{\mathcal{O}^{\mathfrak{p}}}(T \otimes V^{\otimes d})$ is isomorphic to a sum of blocks of cyclotomic quotients of the degenerate affine Hecke algebra $\mathcal{H}_d/\Pi_{i=1}^{\ell}(x_i - n_i)$; see [Brundan and Kleshchev 2008, Theorem 5.13]. In the special case of level $\ell = 2$, these algebras can be explicitly described in terms of generalizations of Khovanov's arc algebra (which Khovanov [2002] introduced to give an algebraic structure underlying Khovanov homology and which categorifies the Temperley–Lieb algebra $\mathcal{TL}_d(\delta)$) and have an interesting representation theory; see [Brundan and Stroppel 2010; 2011a; 2011b; 2012b]. A consequence of this is that, using the results from [Sartori 2014, Theorem 6.9] and [Sartori and Stroppel 2015, Theorem 1.1], one can realize the walled Brauer algebra from Section 5A.6 for arbitrary parameter $\delta \in \mathbb{Z}$ as endomorphism algebras of some $\mathcal{O}^{\mathfrak{p}}$-tilting module and hence, using our main theorem, deduce cellularity again.

If $\mathfrak{g}$ is of another classical type, then the role of the (cyclotomic quotients of the) degenerate affine Hecke algebra is played by (cyclotomic quotients of) degenerate

BMW algebras or so-called (cyclotomic quotients of) $\mathbb{W}_d$-algebras (also called Nazarov–Wenzl algebras). These are still poorly understood and technically quite involved; see [Ariki et al. 2006]. In [Ehrig and Stroppel 2013] special examples of level $\ell = 2$ quotients were studied and realized as endomorphism algebras of some $\mathcal{O}^{\mathfrak{p}}(\mathfrak{so}_{2n})$-tilting module $\Delta^{\mathfrak{p}}(\underline{\delta}) \otimes V \in \mathcal{O}^{\mathfrak{p}}(\mathfrak{so}_{2n})$ where V is the vector representation of $\mathfrak{so}_{2n}$, $\underline{\delta} = \frac{1}{2}\delta \sum_{i=1}^{n} \epsilon_i$ and $\mathfrak{p}$ is a maximal parabolic subalgebra of type A (see Theorem B therein). Hence, our theorem implies cellularity of these algebras. Soergel's theorem is therefore just a shadow of a rich world of endomorphism algebras whose cellularity can be obtained from our approach.

Our methods also apply to (parabolic) category $\mathcal{O}^{\mathfrak{p}}(\hat{\mathfrak{g}})$ attached to an affine Kac–Moody algebra $\hat{\mathfrak{g}}$ over $\mathbb{K}$ and related categories. In particular, one can consider a (level-dependent) quotient $\hat{\mathfrak{g}}_\kappa$ of $\boldsymbol{U}(\hat{\mathfrak{g}})$ and a category, denoted by $\mathbf{O}^{\nu,\kappa}_{\mathbb{K},\tau}$, attached to it (we refer the reader to [Rouquier et al. 2016, Sections 5.2 and 5.3] for the details). Then there is a subcategory $\mathbf{A}^{\nu,\kappa}_{\mathbb{K},\tau} \subset \mathbf{O}^{\nu,\kappa}_{\mathbb{K},\tau}$ and a $\mathbf{A}^{\nu,\kappa}_{\mathbb{K},\tau}$-tilting module $\mathbf{T}_{\mathbb{K},d}$ defined in Section 5.5 of [loc. cit.] such that

$$\Phi_{\mathrm{aff}} : \mathbf{H}^{s}_{\mathbb{K},d} \longrightarrow \mathrm{End}_{\mathbf{A}^{\nu,\kappa}_{\mathbb{K},\tau}}(\mathbf{T}_{\mathbb{K},d}) \quad \text{and}$$
$$\Phi_{\mathrm{aff}} : \mathbf{H}^{s}_{\mathbb{K},d} \xrightarrow{\cong} \mathrm{End}_{\mathbf{A}^{\nu,\kappa}_{\mathbb{K},\tau}}(\mathbf{T}_{\mathbb{K},d}), \quad \text{if } \nu_p \geq d,\ p = 1, \dots N;$$

see Theorem 5.37 and Proposition 8.1 of [loc. cit.]. Here $\mathbf{H}^{s}_{\mathbb{K},d}$ denotes an appropriate cyclotomic quotient of the affine Hecke algebra. Again, our main theorem applies for $\mathbf{H}^{s}_{\mathbb{K},d}$ in case $\nu_p \geq d$.

5A.8. *Graded cellular structures.* A striking property which arises in the context of (parabolic) category $\mathcal{O}$ (or $\mathcal{O}^{\mathfrak{p}}$) is that all the endomorphism algebras from Section 5A.7 can be equipped with a $\mathbb{Z}$-grading as in [Stroppel 2003] arising from the Koszul grading of category $\mathcal{O}$ (or of $\mathcal{O}^{\mathfrak{p}}$). We might choose our cellular basis compatible with this grading and obtain a grading on the endomorphism algebras turning them into graded cellular algebras in the sense of [Hu and Mathas 2010, Definition 2.1].

For the cyclotomic quotients this grading is nontrivial and in fact is the type A KL-R grading in the spirit of Khovanov and Lauda and independently Rouquier (see [Khovanov and Lauda 2009; 2011] or [Rouquier 2008]), which can be seen as a grading on cyclotomic quotients of degenerate affine Hecke algebras; see [Brundan and Kleshchev 2009]. See [Brundan and Stroppel 2011b] for level $\ell = 2$ and [Hu and Mathas 2015] for all levels where the authors construct explicit graded cellular bases. For gradings on (cyclotomic quotients of) $\mathbb{W}_d$-algebras see Section 5 in [Ehrig and Stroppel 2013] and for gradings on Brauer algebras see [Ehrig and Stroppel 2016a] or [Li 2014].

In the same spirit, it should be possible to obtain the higher level analogs of the generalizations of Khovanov's arc algebra, known as $\mathfrak{sl}_n$-web (or, alternatively,

$\mathfrak{gl}_n$-web) algebras (see [Mackaay et al. 2014] and [Mackaay 2014]), from our setup as well using the connections from cyclotomic KL-R algebras to these algebras in [Tubbenhauer 2014a; 2014b]. Although details still need to be worked out, this can be seen as the categorification of the connections to the spider algebras from Section 5A.4: the spiders provide the setup to study the corresponding Reshetikhin–Turaev $\mathfrak{sl}_n$-link polynomials; the $\mathfrak{sl}_n$-web algebras provide the algebraic setup to study the Khovanov–Rozansky $\mathfrak{sl}_n$-link homologies. This would emphasize the connection between our work and low-dimensional topology.

5B. ***(Graded) cellular structures and the Temperley–Lieb algebras: a comparison.*** Finally we want to present one explicit example, the Temperley–Lieb algebras, which is of particular interest in low-dimensional topology and categorification. Our main goal is to construct new (graded) cellular bases, and use our approach to establish semisimplicity conditions, and construct and compute the dimensions of its simple modules in new ways.

We start by briefly recalling the necessary definitions. The reader unfamiliar with these algebras might consider for example [Graham and Lehrer 1996, Section 6] (or [Andersen et al. 2015b], where we recall the basics in detail using the usual Temperley–Lieb diagrams and our notation).

Fix $\delta = q + q^{-1}$ for $q \in \mathbb{K}^*$.[3] Recall that the *Temperley–Lieb algebra* $\mathcal{TL}_d(\delta)$ in d strands with parameter δ is the free diagram algebra over $\mathbb{K}$ with basis consisting of all possible nonintersecting tangle diagrams with d bottom and top boundary points modulo boundary preserving isotopy and the local relation for evaluating circles given by the parameter[4] δ.

Recall from Section 5A.3 (whose notation we use now) that, by quantum Schur–Weyl duality, we can use Theorem 3.9 to obtain cellular bases of $\mathcal{TL}_d(\delta) \cong \mathrm{End}_{U_q}(T)$ (we fix the isomorphism coming from quantum Schur–Weyl duality from now on). The aim now is to compare our cellular bases to the one given by Graham and Lehrer [1996, Theorem 6.7], where we point out that we do not obtain their cellular basis: our cellular basis depends for instance on whether $\mathcal{TL}_d(\delta)$ is semisimple or not. In the nonsemisimple case, at least for $\mathbb{K} = \mathbb{C}$, we obtain a nontrivially $\mathbb{Z}$-graded cellular basis in the sense of [Hu and Mathas 2010, Definition 2.1]; see Proposition 5.8.

Before stating our cellular basis, we provide a criterion which tells precisely whether $\mathcal{TL}_d(\delta)$ is semisimple or not. Recall that there is a known criteria for which Weyl modules $\Delta_q(i)$ are simple; see, e.g., [Andersen et al. 1991, Corollary 4.6].

[3] The $\mathfrak{sl}_2$ case works with any $q \in \mathbb{K}^*$, including even roots of unity, see, e.g., [Andersen and Tubbenhauer 2017, Definition 2.3].

[4] We point out that there are two different conventions about circle evaluations in the literature: evaluating to δ or to $-\delta$. We use the first convention because we want to stay close to the cited literature.

Proposition 5.1 (Semisimplicity criterion for $\mathcal{TL}_d(\delta)$). *We have the following*:

(a) *Let $\delta \neq 0$. Then $\mathcal{TL}_d(\delta)$ is semisimple if and only if $[i] = q^{1-i} + \cdots + q^{i-1} \neq 0$ for all $i = 1, \ldots, d$ if and only if q is not a root of unity with $d < l = \mathrm{ord}(q^2)$, or $q = 1$ and $\mathrm{char}(\mathbb{K}) > d$.*

(b) *Let $\mathrm{char}(\mathbb{K}) = 0$. Then $\mathcal{TL}_d(0)$ is semisimple if and only if d is odd (or $d = 0$).*

(c) *Let $\mathrm{char}(\mathbb{K}) = p > 0$. Then $\mathcal{TL}_d(0)$ is semisimple if and only if $d = 0$ or $d \in \{1, 3, 5, \ldots, 2p-1\}$.*

Proof. (a): We want to show that $T = V^{\otimes d}$ decomposes into simple $\boldsymbol{U}_q$-modules if and only if $d < l$, or $q = 1$ and $\mathrm{char}(\mathbb{K}) > d$, which is clearly equivalent to the nonvanishing of $[i]$ for $i = 1, \ldots, d$.

Assume that $d < l$. Since the maximal $\boldsymbol{U}_q$-weight of $V^{\otimes d}$ is d and since all Weyl $\boldsymbol{U}_q$-modules $\Delta_q(i)$ for $i < l$ are simple, we see that all indecomposable summands of $V^{\otimes d}$ are simple.

Otherwise, if $l \leq d$, then $T_q(d)$ (or $T_q(d-2)$ in the case $d \equiv -1 \mod l$) is a nonsimple, indecomposable summand of $V^{\otimes d}$ (note that this arguments fails if $l = 2$, i.e., $\delta = 0$).

The case $q = 1$ works similarly, and we can now use Theorem 4.13 to finish the proof of (a).

(b): Since $\delta = 0$ if and only if $q = \pm\sqrt[2]{-1}$, we can use the linkage from, e.g., [Andersen and Tubbenhauer 2017, Theorem 2.23] in the case $l = 2$ to see that $T = V^{\otimes d}$ decomposes into a direct sum of simple $\boldsymbol{U}_q$-modules if and only if d is odd (or $d = 0$). This implies that $\mathcal{TL}_d(0)$ is semisimple if and only if d is odd (or $d = 0$) by Theorem 4.13.

(c): If $\mathrm{char}(\mathbb{K}) = p > 0$ and $\delta = 0$ (for $p = 2$ this is equivalent to $q = 1$), then we have $\Delta_q(i) \cong L_q(i)$ if and only if $i = 0$ or $i \in \{2ap^n - 1 \mid n \in \mathbb{Z}_{\geq 0}, 1 \leq a < p\}$. In particular, this means that for $d \geq 2$ we have that either $T_q(d)$ or $T_q(d-2)$ is a simple $\boldsymbol{U}_q$-module if and only if $d \in \{3, 5, \ldots, 2p-1\}$. Hence, using the same reasoning as above, we see that $T = V^{\otimes d}$ is semisimple if and only if $d \in \{0, 1, 3, 5, \ldots, 2p-1\}$. By Theorem 4.13 we see that $\mathcal{TL}_d(0)$ is semisimple if and only if $d \in \{0, 1, 3, 5, \ldots, 2p-1\}$. □

Example 5.2. We have that $[k] \neq 0$ for all $k = 1, 2, 3$ is satisfied if and only if q is not a fourth or a sixth root of unity. By Proposition 5.1 we see that $\mathcal{TL}_3(\delta)$ is semisimple as long as q is not one of these values from above. The other way around is only true for q being a sixth root of unity (the conclusion from semisimplicity to nonvanishing of the quantum numbers above does not work in the case $q = \pm\sqrt[2]{-1}$).

Remark 4. The semisimplicity criterion for $\mathcal{TL}_d(\delta)$ was already found, using quite different methods, in [Westbury 1995, Section 5] in the case $\delta \neq 0$, and in the case $\delta = 0$ in [Martin 1991, Chapter 7] or [Ridout and Saint-Aubin 2014, above Proposition 4.9]. For us it is an easy application of Theorem 4.13.

A direct consequence of Proposition 5.1 is that the Temperley–Lieb algebra $\mathcal{TL}_d(\delta)$ for $q \in \mathbb{K}^* - \{1\}$ not a root of unity is semisimple (or $q = \pm 1$ and $\mathrm{char}(\mathbb{K}) = 0$), regardless of d.

5B.1. *Temperley–Lieb algebra: the semisimple case.* Assume $q \in \mathbb{K}^* - \{1\}$ is not a root of unity (or $q = \pm 1$ and $\mathrm{char}(\mathbb{K}) = 0$). Thus, we are in the semisimple case.

Let us compare our cell datum $(\mathcal{P}, \mathcal{I}, \mathcal{C}, \mathrm{i})$ to the one of Graham and Lehrer (indicated by a subscript GL) from [Graham and Lehrer 1996, Section 6]. They have the poset $\mathcal{P}_{\mathrm{GL}}$ consisting of all length-two partitions of d, and we have the poset $\mathcal{P}$ consisting of all $\lambda \in X^+$ such that $\Delta_q(\lambda)$ is a factor of T. The two sets are clearly the same: an element $\lambda = (\lambda_1, \lambda_2) \in \mathcal{P}_{\mathrm{GL}}$ corresponds to $\lambda_1 - \lambda_2 \in \mathcal{P}$. Similarly, an inductive reasoning shows that $\mathcal{I}_{\mathrm{GL}}$ (standard fillings of the Young diagram associated to λ) is also the same as our $\mathcal{I}$ (to see this one can use the facts listed in [Andersen and Tubbenhauer 2017, Section 2]). One directly checks that the $\mathbb{K}$-linear anti-involution i_{GL} (turning diagrams upside-down) is also our involution i. Thus, except for $\mathcal{C}$ and $\mathcal{C}_{\mathrm{GL}}$, the cell data agree.

In order to state how our cellular bases for $\mathcal{TL}_d(\delta)$ look like, recall that the so-called *generalized Jones–Wenzl projectors* $JW_{\vec{\epsilon}}$ are indexed by d-tuples (with $d > 0$) of the form $\vec{\epsilon} = (\epsilon_1, \ldots, \epsilon_d) \in \{\pm 1\}^d$ such that $\sum_{j=1}^{k} \epsilon_j \geq 0$ for all $k = 1, \ldots, d$, see, e.g., [Cooper and Hogancamp 2015, Section 2]. In case $\vec{\epsilon} = (1, \ldots, 1)$, one recovers the usual Jones–Wenzl projectors introduced by Jones [1983] and then further studied by Wenzl [1987].

Now, in [Cooper and Hogancamp 2015, Proposition 2.19 and Theorem 2.20] it is shown that there exist nonzero scalars $a_{\vec{\epsilon}} \in \mathbb{K}$ such that $JW'_{\vec{\epsilon}} = a_{\vec{\epsilon}} JW_{\vec{\epsilon}}$ are well-defined idempotents forming a complete set of mutually orthogonal, primitive idempotents in $\mathcal{TL}_d(\delta)$. (Cooper and Hogancamp [2015] work over $\mathbb{C}$, but as long as $q \in \mathbb{K}^* - \{1\}$ is not a root of unity their arguments work in our setup as well.) These project to the summands of $T = V^{\otimes d}$ of the form $\Delta_q(i)$ for $i = \sum_{j=1}^{k} \epsilon_j$. In particular, the usual Jones–Wenzl projectors project to the highest weight summand $\Delta_q(d)$ of $T = V^{\otimes d}$.

Proposition 5.3 ((New) cellular bases). *The datum given by the quadruple* $(\mathcal{P}, \mathcal{I}, \mathcal{C}, \mathrm{i})$ *for* $\mathcal{TL}_d(\delta) \cong \mathrm{End}_{U_q}(T)$ *is a cell datum for* $\mathcal{TL}_d(\delta)$. *Moreover*, $\mathcal{C} \neq \mathcal{C}_{\mathrm{GL}}$ *for all* $d > 1$ *and all choices involved in the definition of* $\mathrm{im}(\mathcal{C})$. *In particular, there is a choice such that all generalized Jones–Wenzl projectors* $JW'_{\vec{\epsilon}}$ *are part of* $\mathrm{im}(\mathcal{C})$.

Proof. That we get a cell datum as stated follows from Theorem 3.9 and the discussion above.

That our cellular basis $\mathcal{C}$ will never be $\mathcal{C}_{\mathrm{GL}}$ for $d > 1$ is due to the fact that Graham and Lehrer's cellular basis always contains the identity (which corresponds to the unique standard filling of the Young diagram associated to $\lambda = (d, 0)$).

In contrast, let $\lambda_k = (d-k, k)$ for $0 \leq k \leq \lfloor \frac{1}{2} d \rfloor$. Then

$$T = V^{\otimes d} \cong \Delta_q(d) \oplus \bigoplus_{0<k\leq\lfloor d/2\rfloor} \Delta_q(d-2k)^{\oplus m_{\lambda_k}} \tag{27}$$

for some multiplicities $m_{\lambda_k} \in \mathbb{Z}_{>0}$, we see that for $d > 1$ the identity is never part of any of our bases: all the $\Delta_q(i)$ are simple $\boldsymbol{U}_q$-modules and each c^k_{ij} factors only through $\Delta_q(k)$. In particular, the basis element c^λ_{11} for $\lambda = \lambda_d$ has to be (a scalar multiple) of $JW_{(1,\dots,1)}$.

As in Section 5A.1 we can choose for $\mathcal{C}$ an Artin–Wedderburn basis of $\mathcal{TL}_d(\delta) \cong \mathrm{End}_{\boldsymbol{U}_q}(T)$. Hence, by the above, the corresponding basis consists of the projectors $JW_{\vec{\epsilon}}$. □

Note the following classification result (see for example [Ridout and Saint-Aubin 2014, Corollary 5.2] for $\mathbb{K} = \mathbb{C}$).

Corollary 5.4. *We have a complete set of pairwise nonisomorphic, simple $\mathcal{TL}_d(\delta)$-modules $L(\lambda)$, where $\lambda = (\lambda_1, \lambda_2)$ is a length-two partition of d. Moreover,* $\dim(L(\lambda)) = |\mathrm{Std}(\lambda)|$, *where* $\mathrm{Std}(\lambda)$ *is the set of all standard tableaux of shape λ.*

Proof. This follows directly from Proposition 5.3 and Theorems 4.11 and 4.12 because $m_\lambda = |\mathrm{Std}(\lambda)|$. □

5B.2. *Temperley–Lieb algebra: the nonsemisimple case.* Let us assume that we have fixed $q \in \mathbb{K}^* - \{1, \pm\sqrt[2]{-1}\}$ to be a critical value such that $[k] = 0$ for some $k = 1, \dots, d$. Then, by Proposition 5.1, the algebra $\mathcal{TL}_d(\delta)$ is no longer semisimple. In particular, to the best of our knowledge, there is no diagrammatic analog of the Jones–Wenzl projectors in general.

Proposition 5.5 ((New) cellular basis — the second). *The datum* $(\mathcal{P}, \mathcal{I}, \mathcal{C}, \mathrm{i})$ *with $\mathcal{C}$ as in Theorem 3.9 for $\mathcal{TL}_d(\delta) \cong \mathrm{End}_{\boldsymbol{U}_q}(T)$ is a cell datum for $\mathcal{TL}_d(\delta)$. Moreover, $\mathcal{C} \neq \mathcal{C}_{\mathrm{GL}}$ for all $d > 1$ and all choices involved in the definition of our basis. Thus, there is a choice such that all generalized, nonsemisimple Jones–Wenzl projectors are part of* $\mathrm{im}(\mathcal{C})$.

Proof. As in the proof of Proposition 5.3 and left to the reader. □

Hence, directly from Proposition 5.5 and Theorems 4.11 and 4.12, we obtain:

Corollary 5.6. *We have a complete set of pairwise nonisomorphic, simple $\mathcal{TL}_d(\delta)$-modules $L(\lambda)$, where $\lambda = (\lambda_1, \lambda_2)$ is a length-two partition of d. Moreover,* $\dim(L(\lambda)) = m_\lambda$, *where m_λ is the multiplicity of $T_q(\lambda_1 - \lambda_2)$ as a summand of $T = V^{\otimes d}$.*

Note that we can do better: one gets a decompositions

$$\mathcal{T} \cong \mathcal{T}_{-1} \oplus \mathcal{T}_0 \oplus \mathcal{T}_1 \oplus \cdots \oplus \mathcal{T}_{l-3} \oplus \mathcal{T}_{l-2} \oplus \mathcal{T}_{l-1}, \tag{28}$$

where the blocks $\mathcal{T}_{-1}$ and $\mathcal{T}_{l-1}$ are semisimple if $\mathbb{K} = \mathbb{C}$. (This follows from the linkage principle. For notation and the statement see [Andersen and Tubbenhauer 2017, Section 2].)

Fix $\mathbb{K} = \mathbb{C}$. As explained in [loc. cit., Section 3.5] each block in the decomposition (28) can be equipped with a nontrivial $\mathbb{Z}$-grading coming from the zig-zag algebra [Huerfano and Khovanov 2001]. Hence, we have the following.

Lemma 5.7. *The $\mathbb{C}$-algebra $\mathrm{End}_{U_q}(T)$ can be equipped with a nontrivial $\mathbb{Z}$-grading. Thus, $\mathcal{TL}_d(\delta)$ over $\mathbb{C}$ can be equipped with a nontrivial $\mathbb{Z}$-grading.*

Proof. The second statement follows directly from the first using quantum Schur–Weyl duality. Hence, we only need to show the first.

Note that $T = V^{\otimes d}$ decomposes as in (27), but with the $T_q(k)$ instead of the $\Delta_q(k)$, and we can order this decomposition by blocks. Each block carries a $\mathbb{Z}$-grading coming from the zig-zag algebra (as explained in [Andersen and Tubbenhauer 2017, Section 3]). In particular, we can choose the basis elements c_{ij}^{λ} in such a way that we get the $\mathbb{Z}$-graded basis obtained in Corollary 4.23 therein. Since there is no interaction between different blocks, the statement follows. $\square$

Recall from [Hu and Mathas 2010, Definition 2.1] that a $\mathbb{Z}$-graded cell datum of a $\mathbb{Z}$-graded algebra is a cell datum for the algebra together with an additional *degree function* $\deg : \coprod_{\lambda\in\mathcal{P}} \mathcal{I}^{\lambda} \to \mathbb{Z}$, such that $\deg(c_{ij}^{\lambda}) = \deg(i) + \deg(j)$. For us the choice of $\deg(\cdot)$ is as follows.

If $\lambda \in \mathcal{P}$ is in one of the semisimple blocks, then we simply set $\deg(i) = 0$ for all $i \in \mathcal{I}^{\lambda}$.

Assume that $\lambda \in \mathcal{P}$ is not in the semisimple blocks. It is known that every $T_q(\lambda)$ has precisely two Weyl factors. The g_i^{λ} that map $\Delta_q(\lambda)$ into a higher $T_q(\mu)$ should be indexed by a 1-colored i whereas the g_i^{λ} mapping $\Delta_q(\lambda)$ into $T_q(\lambda)$ should have 0-colored i. Similarly for the f_j^{λ}. Then the degree of the elements $i \in \mathcal{I}^{\lambda}$ should be the corresponding color. We get the following. (Here $\mathcal{C}$ is as in Theorem 3.9.)

Proposition 5.8 (Graded cellular basis). *The datum $(\mathcal{P}, \mathcal{I}, \mathcal{C}, \mathrm{i})$ supplemented with the function $\deg(\cdot)$ from above is a $\mathbb{Z}$-graded cell datum for the $\mathbb{C}$-algebra $\mathcal{TL}_d(\delta) \cong \mathrm{End}_{U_q}(T)$.*

Proof. The hardest part is cellularity which directly follows from Theorem 3.9. That the quintuple $(\mathcal{P}, \mathcal{I}, \mathcal{C}, \mathrm{i}, \deg)$ gives a $\mathbb{Z}$-graded cell datum follows from the construction. $\square$

Remark 5. Our grading and the one found by Plaza and Ryom-Hansen [2014] agree (up to a shift of the indecomposable summands). To see this, note that our algebra is isomorphic to the algebra $K_{1,n}$ studied in [Brundan and Stroppel 2011a] which is by (4.8) there and [Brundan and Stroppel 2011b, Theorem 6.3] a quotient of some particular cyclotomic KL-R algebra (the compatibility of the grading follows for example from [Hu and Mathas 2015, Corollary B.6]). The same holds, by construction, for the grading in [Plaza and Ryom-Hansen 2014].

Acknowledgements

We would like to thank Ben Cooper, Michael Ehrig, Matt Hogancamp, Johannes Kübel, Gus Lehrer, Paul Martin, Andrew Mathas, Volodymyr Mazorchuk, Steen Ryom-Hansen and Paul Wedrich for helpful comments and discussions, and the referees for further useful comments. H.H.A. would like to thank the Institut Mittag-Leffler for the hospitality he enjoyed there during the final stages of this work. C.S. is very grateful to the Max Planck Institute in Bonn for the extraordinary support and excellent working conditions. A large part of her research was worked out during her stay there. D.T. would like to thank the dark Danish winter for very successfully limiting his nonwork options and the Australian long blacks for pushing him forward.

References

[Andersen 1992] H. H. Andersen, "Tensor products of quantized tilting modules", *Comm. Math. Phys.* **149**:1 (1992), 149–159. MR Zbl

[Andersen 1997] H. H. Andersen, "Filtrations and tilting modules", *Ann. Sci. École Norm. Sup.* (4) **30**:3 (1997), 353–366. MR Zbl

[Andersen 2003] H. H. Andersen, "The strong linkage principle for quantum groups at roots of 1", *J. Algebra* **260**:1 (2003), 2–15. MR Zbl

[Andersen and Mazorchuk 2015] H. H. Andersen and V. Mazorchuk, "Category $\mathcal{O}$ for quantum groups", *J. Eur. Math. Soc.* **17**:2 (2015), 405–431. MR Zbl

[Andersen and Tubbenhauer 2017] H. H. Andersen and D. Tubbenhauer, "Diagram categories for $\mathbb{U}_q$-tilting modules at roots of unity", *Transform. Groups* **22**:1 (2017), 29–89.

[Andersen et al. 1991] H. H. Andersen, P. Polo, and K. X. Wen, "Representations of quantum algebras", *Invent. Math.* **104**:1 (1991), 1–59. MR Zbl

[Andersen et al. 1994] H. H. Andersen, J. C. Jantzen, and W. Soergel, *Representations of quantum groups at a pth root of unity and of semisimple groups in characteristic p: independence of p*, Astérisque **220**, Société Mathématique de France, Paris, 1994. MR Zbl

[Andersen et al. 2015a] H. H. Andersen, G. I. Lehrer, and R. Zhang, "Cellularity of certain quantum endomorphism algebras", *Pacific J. Math.* **279**:1-2 (2015), 11–35. MR Zbl

[Andersen et al. 2015b] H. H. Andersen, C. Stroppel, and D. Tubbenhauer, "Additional notes for the paper 'Cellular structures using $\mathbb{U}_q$-tilting modules'", 2015, Available at http://tinyurl.com/tubben. Appendix to H. H. Andersen, C. Stroppel, and D. Tubbenhauer, "Cellular structures using $\mathbb{U}_q$-tilting modules", preprint, 2015.

[Andersen et al. 2017] H. H. Andersen, C. Stroppel, and D. Tubbenhauer, "Semisimplicity of Hecke and (walled) Brauer algebras", *J. Aust. Math. Soc.* **103**:1 (2017), 1–44.

[Ariki 1999] S. Ariki, "Cyclotomic q-Schur algebras as quotients of quantum algebras", *J. Reine Angew. Math.* **513** (1999), 53–69. MR Zbl

[Ariki and Koike 1994] S. Ariki and K. Koike, "A Hecke algebra of $(\mathbb{Z}/r\mathbb{Z}) \wr \mathfrak{S}_n$ and construction of its irreducible representations", *Adv. Math.* **106**:2 (1994), 216–243. MR Zbl

[Ariki et al. 2006] S. Ariki, A. Mathas, and H. Rui, "Cyclotomic Nazarov–Wenzl algebras", *Nagoya Math. J.* **182** (2006), 47–134. MR Zbl

[Birman and Wenzl 1989] J. S. Birman and H. Wenzl, "Braids, link polynomials and a new algebra", *Trans. Amer. Math. Soc.* **313**:1 (1989), 249–273. MR Zbl

[Brauer 1937] R. Brauer, "On algebras which are connected with the semisimple continuous groups", *Ann. of Math.* (2) **38**:4 (1937), 857–872. MR Zbl

[Brundan and Kleshchev 2008] J. Brundan and A. Kleshchev, "Schur–Weyl duality for higher levels", *Selecta Math.* (*N.S.*) **14**:1 (2008), 1–57. MR Zbl

[Brundan and Kleshchev 2009] J. Brundan and A. Kleshchev, "Blocks of cyclotomic Hecke algebras and Khovanov–Lauda algebras", *Invent. Math.* **178**:3 (2009), 451–484. MR Zbl

[Brundan and Stroppel 2010] J. Brundan and C. Stroppel, "Highest weight categories arising from Khovanov's diagram algebra, II: Koszulity", *Transform. Groups* **15**:1 (2010), 1–45. MR Zbl

[Brundan and Stroppel 2011a] J. Brundan and C. Stroppel, "Highest weight categories arising from Khovanov's diagram algebra, I: Cellularity", *Mosc. Math. J.* **11**:4 (2011), 685–722. MR Zbl

[Brundan and Stroppel 2011b] J. Brundan and C. Stroppel, "Highest weight categories arising from Khovanov's diagram algebra, III: Category $\mathcal{O}$", *Represent. Theory* **15** (2011), 170–243. MR Zbl

[Brundan and Stroppel 2012a] J. Brundan and C. Stroppel, "Gradings on walled Brauer algebras and Khovanov's arc algebra", *Adv. Math.* **231**:2 (2012), 709–773. MR Zbl

[Brundan and Stroppel 2012b] J. Brundan and C. Stroppel, "Highest weight categories arising from Khovanov's diagram algebra, IV: The general linear supergroup", *J. Eur. Math. Soc.* **14**:2 (2012), 373–419. MR Zbl

[Cautis et al. 2014] S. Cautis, J. Kamnitzer, and S. Morrison, "Webs and quantum skew Howe duality", *Math. Ann.* **360**:1-2 (2014), 351–390. MR Zbl

[Cline et al. 1988] E. Cline, B. Parshall, and L. Scott, "Finite-dimensional algebras and highest weight categories", *J. Reine Angew. Math.* **391** (1988), 85–99. MR Zbl

[Cooper and Hogancamp 2015] B. Cooper and M. Hogancamp, "An exceptional collection for Khovanov homology", *Algebr. Geom. Topol.* **15**:5 (2015), 2659–2707. MR Zbl

[Dipper et al. 1998] R. Dipper, G. James, and A. Mathas, "Cyclotomic q-Schur algebras", *Math. Z.* **229**:3 (1998), 385–416. MR Zbl

[Dipper et al. 2008] R. Dipper, S. Doty, and J. Hu, "Brauer algebras, symplectic Schur algebras and Schur–Weyl duality", *Trans. Amer. Math. Soc.* **360**:1 (2008), 189–213. MR Zbl

[Dipper et al. 2014] R. Dipper, S. Doty, and F. Stoll, "The quantized walled Brauer algebra and mixed tensor space", *Algebr. Represent. Theory* **17**:2 (2014), 675–701. MR Zbl

[Donkin 1993] S. Donkin, "On tilting modules for algebraic groups", *Math. Z.* **212**:1 (1993), 39–60. MR Zbl

[Donkin 1998] S. Donkin, *The q-Schur algebra*, London Mathematical Society Lecture Note Series **253**, Cambridge Univ. Press, 1998. MR Zbl

[Doty and Hu 2009] S. Doty and J. Hu, "Schur–Weyl duality for orthogonal groups", *Proc. Lond. Math. Soc.* (3) **98**:3 (2009), 679–713. MR Zbl

[Du et al. 1998] J. Du, B. Parshall, and L. Scott, "Quantum Weyl reciprocity and tilting modules", *Comm. Math. Phys.* **195**:2 (1998), 321–352. MR Zbl

[Ehrig and Stroppel 2013] M. Ehrig and C. Stroppel, "Nazarov–Wenzl algebras, coideal subalgebras and categorified skew Howe duality", preprint, 2013. arXiv

[Ehrig and Stroppel 2016a] M. Ehrig and C. Stroppel, "Koszul gradings on Brauer algebras", *Int. Math. Res. Not.* **2016**:13 (2016), 3970–4011. MR

[Ehrig and Stroppel 2016b] M. Ehrig and C. Stroppel, "Schur–Weyl duality for the Brauer algebra and the ortho-symplectic Lie superalgebra", *Math. Z.* **284**:1-2 (2016), 595–613. MR Zbl

[Elias 2015] B. Elias, "Light ladders and clasp conjectures", preprint, 2015. arXiv

[Erdmann 1995] K. Erdmann, "Tensor products and dimensions of simple modules for symmetric groups", *Manuscripta Math.* **88**:3 (1995), 357–386. MR Zbl

[Graham and Lehrer 1996] J. J. Graham and G. I. Lehrer, "Cellular algebras", *Invent. Math.* **123**:1 (1996), 1–34. MR Zbl

[Halverson and Ram 2001] T. Halverson and A. Ram, "q-rook monoid algebras, Hecke algebras, and Schur–Weyl duality", *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*) **283**:6 (2001), 224–250. MR Zbl

[Hu 2011] J. Hu, "BMW algebra, quantized coordinate algebra and type C Schur–Weyl duality", *Represent. Theory* **15** (2011), 1–62. MR Zbl

[Hu and Mathas 2010] J. Hu and A. Mathas, "Graded cellular bases for the cyclotomic Khovanov–Lauda–Rouquier algebras of type A", *Adv. Math.* **225**:2 (2010), 598–642. MR Zbl

[Hu and Mathas 2015] J. Hu and A. Mathas, "Quiver Schur algebras for linear quivers", *Proc. Lond. Math. Soc.* (3) **110**:6 (2015), 1315–1386. MR Zbl

[Hu and Stoll 2004] J. Hu and F. Stoll, "On double centralizer properties between quantum groups and Ariki–Koike algebras", *J. Algebra* **275**:1 (2004), 397–418. MR Zbl

[Huerfano and Khovanov 2001] R. S. Huerfano and M. Khovanov, "A Category for the Adjoint Representation", *J. Algebra* **246**:2 (2001), 514–542.

[Humphreys 2008] J. E. Humphreys, *Representations of semisimple Lie algebras in the BGG category $\mathcal{O}$*, Graduate Studies in Mathematics **94**, American Mathematical Society, Providence, RI, 2008. MR Zbl

[Jantzen 1973] J. C. Jantzen, *Darstellungen halbeinfacher algebraischer Gruppen und zugeordnete kontravariante Formen*, Bonn. Math. Schr. **67**, Mathematischen Institut der Universität, Bonn, 1973. MR Zbl

[Jantzen 1996] J. C. Jantzen, *Lectures on quantum groups*, Graduate Studies in Mathematics **6**, American Mathematical Society, Providence, RI, 1996. MR Zbl

[Jantzen 2003] J. C. Jantzen, *Representations of algebraic groups*, 2nd ed., Mathematical Surveys and Monographs **107**, American Mathematical Society, Providence, RI, 2003. MR Zbl

[Jones 1983] V. F. R. Jones, "Index for subfactors", *Invent. Math.* **72**:1 (1983), 1–25. MR Zbl

[Kazhdan and Lusztig 1994] D. Kazhdan and G. Lusztig, "Tensor structures arising from affine Lie algebras, IV", *J. Amer. Math. Soc.* **7**:2 (1994), 383–453. MR Zbl

[Khovanov 2002] M. Khovanov, "A functor-valued invariant of tangles", *Algebr. Geom. Topol.* **2** (2002), 665–741. MR Zbl

[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 Zbl

[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 Zbl

[Koike 1989] K. Koike, "On the decomposition of tensor products of the representations of the classical groups: by means of the universal characters", *Adv. Math.* **74**:1 (1989), 57–86. MR Zbl

[König and Xi 1998] S. König and C. Xi, "On the structure of cellular algebras", pp. 365–386 in *Algebras and modules, II* (Geiranger, Norway, 1996), edited by I. Reiten et al., CMS Conf. Proc. **24**, American Mathematical Society, Providence, RI, 1998. MR Zbl

[Kuperberg 1996] G. Kuperberg, "Spiders for rank 2 Lie algebras", *Comm. Math. Phys.* **180**:1 (1996), 109–151. MR Zbl

[Lascoux et al. 1996] A. Lascoux, B. Leclerc, and J.-Y. Thibon, "Hecke algebras at roots of unity and crystal bases of quantum affine algebras", *Comm. Math. Phys.* **181**:1 (1996), 205–263. MR Zbl

[Lehrer and Zhang 2012] G. Lehrer and R. Zhang, "The second fundamental theorem of invariant theory for the orthogonal group", *Ann. of Math.* (2) **176**:3 (2012), 2031–2054. MR Zbl

[Li 2014] G. Li, "A KLR grading of the Brauer algebras", preprint, 2014. arXiv

[Lusztig 1989] G. Lusztig, "Modular representations and quantum groups", pp. 59–77 in *Classical groups and related topics* (Beijing, 1987), edited by A. J. Hahn et al., Contemp. Math. **82**, American Mathematical Society, Providence, RI, 1989. MR Zbl

[Mackaay 2014] M. Mackaay, "The $\mathfrak{sl}_N$-web algebras and dual canonical bases", *J. Algebra* **409** (2014), 54–100. MR Zbl

[Mackaay et al. 2014] M. Mackaay, W. Pan, and D. Tubbenhauer, "The $\mathfrak{sl}_3$-web algebra", *Math. Z.* **277**:1-2 (2014), 401–479. MR Zbl

[Martin 1991] P. Martin, *Potts models and related problems in statistical mechanics*, Series on Advances in Statistical Mechanics **5**, World Scientific, Teaneck, NJ, 1991. MR Zbl

[Martin and Saleur 1994] P. Martin and H. Saleur, "The blob algebra and the periodic Temperley–Lieb algebra", *Lett. Math. Phys.* **30**:3 (1994), 189–206. MR Zbl

[Mazorchuk and Stroppel 2016] V. Mazorchuk and C. Stroppel, "$G(\ell, k, d)$-modules via groupoids", *J. Algebraic Combin.* **43**:1 (2016), 11–32. MR Zbl

[Murakami 1987] J. Murakami, "The Kauffman polynomial of links and representation theory", *Osaka J. Math.* **24**:4 (1987), 745–758. MR Zbl

[Paget 2006] R. Paget, "Representation theory of q-rook monoid algebras", *J. Algebraic Combin.* **24**:3 (2006), 239–252. MR Zbl

[Paradowski 1994] J. Paradowski, "Filtrations of modules over the quantum algebra", pp. 93–108 in *Algebraic groups and their generalizations: quantum and infinite-dimensional methods* (University Park, PA, 1991), edited by W. J. Haboush and B. J. Parshall, Proc. Sympos. Pure Math. **56, part 2**, American Mathematical Society, Providence, RI, 1994. MR Zbl

[Plaza and Ryom-Hansen 2014] D. Plaza and S. Ryom-Hansen, "Graded cellular bases for Temperley–Lieb algebras of type A and B", *J. Algebraic Combin.* **40**:1 (2014), 137–177. MR Zbl

[Riche and Williamson 2015] S. Riche and G. Williamson, "Tilting modules and the p-canonical basis", preprint, 2015, Available at https://hal.archives-ouvertes.fr/hal-01249796v3/document. arXiv

[Ridout and Saint-Aubin 2014] D. Ridout and Y. Saint-Aubin, "Standard modules, induction and the structure of the Temperley–Lieb algebra", *Adv. Theor. Math. Phys.* **18**:5 (2014), 957–1041. MR Zbl

[Ringel 1991] C. M. Ringel, "The category of modules with good filtrations over a quasi-hereditary algebra has almost split sequences", *Math. Z.* **208**:2 (1991), 209–223. MR Zbl

[Rose and Tubbenhauer 2016] D. E. V. Rose and D. Tubbenhauer, "Symmetric webs, Jones–Wenzl recursions, and q-Howe duality", *Int. Math. Res. Not.* **2016**:17 (2016), 5249–5290. MR

[Rouquier 2008] R. Rouquier, "2-Kac–Moody algebras", preprint, 2008. arXiv

[Rouquier et al. 2016] R. Rouquier, P. Shan, M. Varagnolo, and E. Vasserot, "Categorifications and cyclotomic rational double affine Hecke algebras", *Invent. Math.* **204**:3 (2016), 671–786. MR Zbl

[Ryom-Hansen 2010] S. Ryom-Hansen, "The Ariki–Terasoma–Yamada tensor space and the blob algebra", *J. Algebra* **324**:10 (2010), 2658–2675. MR Zbl

[Ryom-Hansen 2012] S. Ryom-Hansen, "Cell structures on the blob algebra", *Represent. Theory* **16** (2012), 540–567. MR Zbl

[Sakamoto and Shoji 1999] M. Sakamoto and T. Shoji, "Schur–Weyl reciprocity for Ariki–Koike algebras", *J. Algebra* **221**:1 (1999), 293–314. MR Zbl

[Sartori 2014] A. Sartori, "The degenerate affine walled Brauer algebra", *J. Algebra* **417** (2014), 198–233. MR Zbl

[Sartori and Stroppel 2015] A. Sartori and C. Stroppel, "Walled Brauer algebras as idempotent truncations of level 2 cyclotomic quotients", *J. Algebra* **440** (2015), 602–638. MR Zbl

[Soergel 1990] W. Soergel, "Kategorie $\mathcal{O}$, perverse Garben und Moduln über den Koinvarianten zur Weylgruppe", *J. Amer. Math. Soc.* **3**:2 (1990), 421–445. MR Zbl

[Soergel 1997] W. Soergel, "Kazhdan–Lusztig polynomials and a combinatoric[s] for tilting modules", *Represent. Theory* **1** (1997), 83–114. MR

[Soergel 1998] W. Soergel, "Character formulas for tilting modules over Kac–Moody algebras", *Represent. Theory* **2** (1998), 432–448. MR Zbl

[Soergel 1999] W. Soergel, "Character formulas for tilting modules over quantum groups at roots of one", pp. 161–172 in *Current developments in mathematics* (Cambridge, 1997), edited by R. Bott et al., International Press, Boston, 1999. MR Zbl

[Solomon 1990] L. Solomon, "The Bruhat decomposition, Tits system and Iwahori ring for the monoid of matrices over a finite field", *Geom. Dedicata* **36**:1 (1990), 15–49. MR Zbl

[Stroppel 2003] C. Stroppel, "Category $\mathcal{O}$: gradings and translation functors", *J. Algebra* **268**:1 (2003), 301–326. MR Zbl

[Tanisaki 2004] T. Tanisaki, "Character formulas of Kazhdan–Lusztig type", pp. 261–276 in *Representations of finite dimensional algebras and related topics in Lie theory and geometry*, edited by V. Dlab and C. M. Ringel, Fields Inst. Commun. **40**, American Mathematical Society, Providence, RI, 2004. MR Zbl

[Temperley and Lieb 1971] H. N. V. Temperley and E. H. Lieb, "Relations between the 'percolation' and 'colouring' problem and other graph-theoretical problems associated with regular planar lattices: some exact results for the 'percolation' problem", *Proc. Roy. Soc. London Ser. A* **322**:1549 (1971), 251–280. MR Zbl

[Tubbenhauer 2014a] D. Tubbenhauer, "$\mathfrak{sl}_3$-web bases, intermediate crystal bases and categorification", *J. Algebraic Combin.* **40**:4 (2014), 1001–1076. MR Zbl

[Tubbenhauer 2014b] D. Tubbenhauer, "$\mathfrak{sl}_n$-webs, categorification and Khovanov–Rozansky homologies", preprint, 2014. arXiv

[Turaev 1989] V. G. Turaev, "Operator invariants of tangles, and R-matrices", *Izv. Akad. Nauk SSSR Ser. Mat.* **53**:5 (1989), 1073–1107. In Russian; translated in *Math. USSR-Izv.* **35**:2 (1990), 411–444. MR Zbl

[Turaev 2010] V. G. Turaev, *Quantum invariants of knots and 3-manifolds*, revised 2nd ed., de Gruyter Studies in Mathematics **18**, de Gruyter, Berlin, 2010. MR Zbl

[Wenzl 1987] H. Wenzl, "On sequences of projections", *C. R. Math. Rep. Acad. Sci. Canada* **9**:1 (1987), 5–9. MR Zbl

[Westbury 1995] B. W. Westbury, "The representation theory of the Temperley–Lieb algebras", *Math. Z.* **219**:4 (1995), 539–565. MR Zbl

[Westbury 2009] B. W. Westbury, "Invariant tensors and cellular categories", *J. Algebra* **321**:11 (2009), 3563–3567. MR Zbl

Received October 1, 2016. Revised February 22, 2017.

HENNING HAAHR ANDERSEN
QGM, DET NATURVIDENSKABELIGE FAKULTET
AARHUS UNIVERSITET
AARHUS
DENMARK
h.h.andersen@gmail.com

CATHARINA STROPPEL
MATHEMATISCHES INSTITUT
UNIVERSITÄT BONN
BONN
GERMANY
stroppel@math.uni-bonn.de

DANIEL TUBBENHAUER
MATHEMATISCHES INSTITUT
UNIVERSITÄT BONN
BONN
GERMANY
dtubben@math.uni-bonn.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 292, No. 1, 2018
dx.doi.org/10.2140/pjm.2018.292.61

MERIDIONAL RANK AND BRIDGE NUMBER FOR A CLASS OF LINKS

MICHEL BOILEAU, YEONHEE JANG AND RICHARD WEIDMANN

We prove that links with meridional rank 3 whose 2-fold branched covers are graph manifolds are 3-bridge links. This gives a partial answer to a question by S. Cappell and J. Shaneson on the relation between the bridge numbers and meridional ranks of links. To prove this result, we also show that the meridional rank of any satellite knot is at least 4.

1. Introduction

An *n-bridge sphere* of a link L in the 3-sphere S^3 is a 2-sphere which meets L in $2n$ points and cuts (S^3, L) into n-string trivial tangles. Here, an *n-string trivial tangle* is a pair (B^3, t) of the 3-ball B^3 and n arcs properly embedded in B^3 parallel to the boundary of B^3. It is known that every link admits an n-bridge sphere for some positive integer n. We call a link L an *n-bridge link* if L admits an n-bridge sphere and does not admit an $(n-1)$-bridge sphere. We call n the *bridge number* of the link L and denote it by $b(L)$.

If a link admits an n-bridge sphere, then it is easy to see that $\pi_1(S^3 \setminus L)$ can be generated by n meridians, where a meridian is an element of the fundamental group that is represented by a curve that is freely homotopic to a meridian of L. This implies that the minimal number of meridians needed to generate the group $\pi_1(S^3 \setminus L)$ is less than or equal to $b(L)$. We denote by $w(L)$ the minimal number of meridians needed to generate $\pi_1(S^3 \setminus L)$ and call it the *meridional rank* of L. Thus for any link L we have $b(L) \geq w(L)$.

S. Cappell and J. Shaneson [Kirby 1978, Problem 1.11], as well as K. Murasugi, have asked whether the converse holds:

Question 1.1. Does the equality $b(L) = w(L)$ hold for any link L in S^3?

This is known to be true for (generalized) Montesinos links by [Boileau and Zieschang 1985], torus links by [Rost and Zieschang 1987] and for another class

Boileau was partially supported by ANR projects 12-BS01-0003-01 and 12-BS01-0004-01 .
MSC2010: 57M25.
Keywords: knot, link, bridge number, meridian, meridional rank, 2-fold branched cover, graph manifold.

of knots (also referred to as generalized Montesinos knots) by [Lustig and Moriah 1993]. More recently the equality has been established for a large class of iterated torus knots using knot contact homology [Cornwell and Hemminger 2016]; see also [Cornwell 2014]. It is a consequence of Dehn's Lemma that $b(L) = 1$ if and only if $w(L) = 1$. Moreover in [Boileau and Zimmermann 1989] it is proved that $b(L) = 2$ if and only if $w(L) = 2$.

The main purpose of this paper is to prove the following theorem.

Theorem 1.2. *Let L be a link in the 3-sphere S^3, and suppose that the 2-fold branched cover of S^3 branched along L is a graph manifold. If $w(L) = 3$, then $b(L) = 3$, i.e., L is a 3-bridge link.*

Here a graph manifold is a compact orientable prime 3-manifold whose geometric decomposition contains only Seifert fibered pieces.

The above theorem, together with the result in [Boileau and Zimmermann 1989], implies that $b(L) = 3$ if and only if $w(L) = 3$ for links whose 2-fold branched covers are graph manifolds. In particular we obtain the following:

Corollary 1.3. *Let $L \subset S^3$ be a link whose 2-fold branched cover is a graph manifold. If $b(L) = 4$, then $w(L) = 4$.*

We obtain also the following corollary which answers [Boileau and Weidmann 2008, Question 2] positively for graph manifolds.

Corollary 1.4. *For a closed orientable graph manifold M, any inversion of $\pi_1(M)$ is hyperelliptic.*

We remark that Question 1.1, posed by Cappell and Shaneson, is related, by taking the 2-fold branched covering, to the question of whether or not the Heegaard genus of a 3-manifold equals the rank of its fundamental group. For the latter question many counterexamples are known; see [Boileau and Weidmann 2005; Boileau and Zieschang 1983; 1984; Li 2013; Schultens and Weidmann 2007; Weidmann 2003]. Thus there exist manifolds such that the ranks of their fundamental groups are smaller than their Heegaard genera. To the question of Cappell and Shaneson, however, no counterexample is known to date.

We also remark that if we replace $w(L)$ with the rank of the link group $\pi_1(S^3 \setminus L)$ then we can easily find examples where the differences between the two numbers are arbitrarily large. For example, the rank of the group $\pi_1(S^3 \setminus K(p,q))$ of a torus link $K(p,q)$ is 2 while $b(K(p,q)) = \min(p,q)$ by [Schubert 1954].

To prove Theorem 1.2 we distinguish two cases, namely the case when the link L is arborescent in the sense of Bonahon and Siebenmann [2016] and the case when L is not arborescent. We will make use of the following theorem, which is interesting in its own right.

Theorem 1.5. *Let K be a prime knot such that $S^3 \setminus K$ has a nontrivial JSJ-decomposition and let m_1, m_2, m_3 be meridians. Then one of the following holds*:

(1) *$\langle m_1, m_2, m_3 \rangle$ is free.*

(2) *$\langle m_1, m_2, m_3 \rangle$ is conjugate into the subgroup of $\pi_1(S^3 \setminus K)$ corresponding to the peripheral piece of $S^3 \setminus K$.*

Corollary 1.6. *Let K be a prime knot such that $S^3 \setminus K$ has a nontrivial JSJ-decomposition. Then $w(K) \geq 4$.*

Corollary 1.7. *Let $K \subset S^3$ be a knot. If $w(K) \leq 3$, then K is either a hyperbolic knot or a torus knot or a connected sum of two 2-bridge knots.*

Theorem 1.5 suggests this strengthening of Question 1.1 for a hyperbolic knot:

Question 1.8. Let $K \subset S^3$ be a hyperbolic knot. Is a subgroup of $\pi_1(S^3 \setminus K)$ generated by at most $b(K) - 1$ meridians free?

In the case of torus knots the conclusion of Question 1.8 has been established by M. Rost and H. Zieschang [1987]. The case of hyperbolic 3-bridge knots follows from a general result for subgroups generated by two meridians in a knot group; see Proposition 4.2. It should be noted that the conclusion of Question 1.8 does obviously not hold for connected sums of knots, and it is moreover not difficult to come up with examples of prime knots with nontrivial JSJ-decomposition for which the conclusion does not hold either.

There is a natural partial order on the set of links in S^3 given by degree-one maps: We say that a link $L \subset S^3$ *dominates* a link $L' \subset S^3$ and write $L \geq L'$ if there is a proper degree-one map $f : E(L) \to E(L')$ between the exteriors of L and L' whose restriction to the boundary is a homeomorphism which extends to the regular neighborhoods of L and L'. It defines a partial order on the set of links in S^3, and it is an open problem to characterize minimal elements. In particular the behavior of the bridge number with respect to this order is far from being understood:

Question 1.9. Let L and L' be links in S^3. Does $L \geq L'$ imply $b(L) \geq b(L')$?

It follows from the definition that the epimorphism $f_\star : \pi_1(S^3 \setminus L) \to \pi_1(S^3 \setminus L')$ induced by the degree-one map $f : E(L) \to E(L')$ preserves the meridians and so that $w(L) \geq w(L')$ whenever $L \geq L'$. Therefore an affirmative answer to Question 1.1 would imply an affirmative answer to Question 1.9.

The answer to Question 1.9 is certainly positive when $b(L') = 2$ as in this case any knot L with $L \geq L'$ cannot be trivial. Our results moreover imply the following:

Proposition 1.10. *Let $L \geq L'$ be two links in S^3.*

(a) *If $b(L') = 3$, then $b(L) \geq 3$.*

(b) *If $b(L') = 4$ and the 2-fold cover of S^3 branched along L is a graph manifold, then $b(L) \geq 4$.*

In Section 2, we recall the definition and some properties of arborescent links and show that an arborescent link L with $w(L) = 3$ is hyperbolic. Section 3 is devoted to the proof of Theorem 1.2 for arborescent links. Section 4 contains the proof of Theorem 1.5. In Section 5 we complete the proof of Theorem 1.2 for the case of non-arborescent links. Then Section 6 contains the proof of Proposition 1.10.

2. Arborescent links

A (3, 1)-*manifold pair* is a pair (M, L) of a compact oriented 3-manifold M and a proper 1-submanifold L of M. By a *surface* F *in* (M, L), we mean a surface F in M intersecting L transversely. Two surfaces F and F' in (M, L) are said to be *pairwise isotopic* (*isotopic*, in brief) if there is a homeomorphism $f : (M, L) \to (M, L)$ such that $f(F) = F'$ and f is pairwise isotopic to the identity. We call a (3, 1)-manifold pair a *tangle* if M is homeomorphic to B^3.

A *trivial tangle* is a (3, 1)-manifold pair (B^3, L), where L is the union of two properly embedded arcs in the 3-ball B^3 which together with arcs on the boundary of B^3 bound disjoint disks. A *rational tangle* is a trivial tangle (B^3, L) endowed with a homeomorphism from $\partial(B^3, L)$ to the “standard” pair of the 2-sphere and the union of four points on the sphere. It is well known that rational tangles (up to isotopy fixing the boundaries) correspond to elements of $\mathbb{Q} \cup \{\infty\}$, called the *slopes* of the rational tangles. For example, the rational tangle of slope β/α can be illustrated as in Figure 1, where α, β are defined by the continued fraction

$$(*) \qquad \begin{aligned} \frac{\beta}{\alpha} &= -a_0 + [a_1, -a_2, \dots, \pm a_m] \\ &:= -a_0 + \cfrac{1}{a_1 + \cfrac{1}{-a_2 + \cfrac{1}{\dots + \cfrac{1}{\pm a_m}}}} \end{aligned}$$

together with the condition that α and β are relatively prime and $\alpha \geq 0$. Here, the numbers a_i denote the numbers of right-hand half twists.

A *Montesinos pair* is a (3, 1)-manifold pair which is built from the pair in Figure 2 (left) or Figure 2 (right) by plugging some of the holes with rational tangles of finite slopes. We say that a Montesinos pair is *trivial* if it is homeomorphic to a rational tangle or $(S, P) \times I$, where S is a 2-sphere, P is the union of four distinct points on S and I is a closed interval. A *Montesinos link* is a link obtained by plugging the remaining holes of a Montesinos pair in Figure 2 (left) with rational tangles of finite slopes, as shown in Figure 3. Unless otherwise stated, we assume that the slope β_i/α_i of each rational tangle is not an integer, that is, $\alpha_i > 1$. The above Montesinos link is denoted by $L(-b\,; \beta_1/\alpha_1, \dots, \beta_r/\alpha_r)$. (We note that this is denoted by $m(0 \mid b\,; (\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_r, \beta_r))$ in [Boileau and Zieschang 1985].)

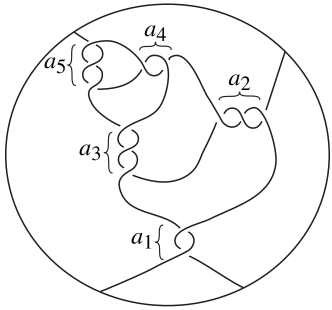


Figure 1. Rational tangle of slope $\beta/\alpha = 31/50$, which has the expression (∗) with $m = 5$, $a_0 = 0$, $a_1 = 2$, $a_2 = 3$, $a_3 = 3$, $a_4 = 2$, $a_5 = 3$.

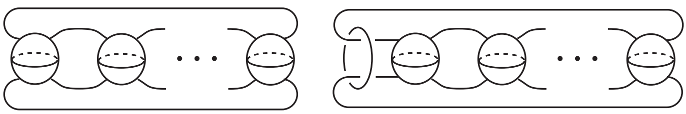


Figure 2. Starting points for a Montesinos pair.

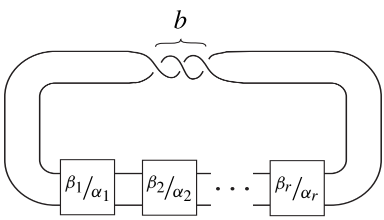


Figure 3. A Montesinos link with $b = 3$.

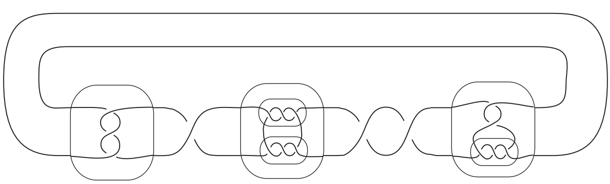


Figure 4. An arborescent link.

An *arborescent link* is a link in S^3 obtained by gluing some Montesinos pairs in their boundaries as in Figure 4; see [Bonahon and Siebenmann 2016].

The main result of this section is the following proposition which is used to prove Theorem 1.2 in Section 3 when the link L is an arborescent link.

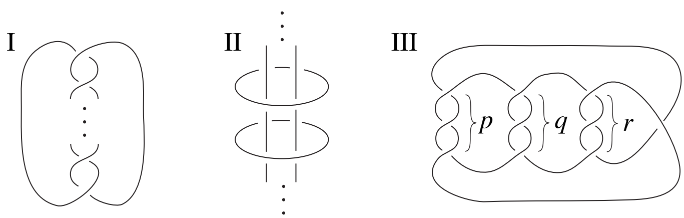
Proposition 2.1. *Let L be an arborescent link which is not a generalized Montesinos link, and suppose that $w(L) = 3$. Then L is hyperbolic.*

Proof. Let L be an arborescent link which is not a generalized Montesinos link, and suppose that $w(L) = 3$. Assume on the contrary that L is not hyperbolic. By [Bonahon and Siebenmann 2016] (see also [Futer and Guéritaud 2009] or [Jang 2011, Proposition 3]), we are in one of three cases, illustrated below:

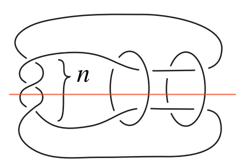
(I) L is a torus knot or link of type $(2, n)$ for some integer n.

(II) L has two parallel components, each of which bounds a twice-punctured disk properly embedded in $S^3 \setminus L$.

(III) L or its reflection is the pretzel link $P(p, q, r, -1) := L\big(-1; \frac{1}{p}, \frac{1}{q}, \frac{1}{r}\big)$, where $p, q, r \geq 2$ and $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} \geq 1$.



By the assumptions that L is not a generalized Montesinos link and that $w(L) = 3$, L must be equivalent to a link in case II, namely, L has two parallel components, each of which bounds a twice-punctured disk properly embedded in $S^3 \setminus L$. Moreover, since $w(L) = 3$, L must have 3 components. Recall that the 2-fold branched cover of S^3 branched along L is a graph manifold. By [Boileau and Weidmann 2008, Proposition 20(2)], the union of any two components of L is a 2-bridge link. Then, by arguments in the proof of [Jang 2011, Proposition 4(1)], we see that L is equivalent to this link:



However, this link is a generalized Montesinos link, which contradicts the assumption. Hence, L is hyperbolic. □

3. Proof of Theorem 1.2 for arborescent links

Let L be an arborescent link and suppose that $w(L) = 3$. If L is a generalized Montesinos link, then we have $b(L) = 3$ by [Boileau and Zieschang 1985]. Thus we assume that L is not a generalized Montesinos link in the remainder of this proof. Then, by Proposition 2.1, L is hyperbolic. Let $P = P_1 \cup \cdots \cup P_k$ be the union of Conway spheres which gives the characteristic decomposition of L. (See [Bonahon and Siebenmann 2016] for a definition of the characteristic decomposition of a link; by [Boileau et al. 2003], this decomposition corresponds to the geometric decomposition of the 3-orbifold with underlying space S^3 and singular locus L with branching index 2.) Let $M := M_2(L)$ be the 2-fold cover of S^3 branched along L, and let T_i be the preimage of P_i in M ($i = 1, \ldots, k$). Then each T_i is a separating torus in M and $T = T_1 \cup \cdots \cup T_k$ gives the JSJ-decomposition of M, by [Jang 2011, Proposition 4]. Let τ_L be the covering involution of the 2-fold branched cover. By construction, the following hold.

(T1) Each T_i is τ_L-invariant and $\tau|_{T_i}$ is hyperelliptic.

(T2) τ_L preserves each JSJ piece and each exceptional fiber of Seifert pieces.

Recall that we have an exact sequence

$$1 \to \pi_1(M) \to \pi_1(S^3 \setminus L)/N \to \mathbb{Z}/2\mathbb{Z} \to 1,$$

where N is the subgroup of $\pi_1(S^3 \setminus L)$ normally generated by the squares of the meridians. Let m_1, m_2 and m_3 be meridians of $\pi_1(S^3 \setminus L)$ generating the group. For $1 \leq i \leq 3$ we denote the image of m_i in $\pi_1(S^3 \setminus L)/N$ again by m_i. Since $\pi_1(M)$ can be regarded as an index-2 subgroup of $\pi_1(S^3 \setminus L)/N$ by the above exact sequence, any element of $\pi_1(M)$ can be represented as a product of even numbers of m_1, m_2 and m_3. Set $g_1 := m_1 m_2$ and $g_2 := m_1 m_3$. Then g_1 and g_2 generate $\pi_1(M)$. Let α be the automorphism of $\pi_1(S^3 \setminus L)/N$ induced by the conjugation by m_1. Then τ_L is a realization of α. We see $\alpha(g_i) = m_1 g_i m_1^{-1} = g_i^{-1}$ for each $i = 1, 2$, and hence, $\alpha|_{\pi_1(M)}$ is an automorphism of $\pi_1(M)$ which sends each generator g_i to g_i^{-1}. Namely, α is an *inversion* of $\pi_1(M)$ (see [Boileau and Weidmann 2008]). Since M is a graph manifold which admits an inversion, the Heegaard genus of M is 2 by [Boileau and Weidmann 2008, Theorem 3]. Recall that $T = T_1 \cup \cdots \cup T_k$ gives the nontrivial JSJ-decomposition of M, where each T_i is a separating torus in M. By [Jang 2011, Proposition 4], M satisfies one of the following conditions (M1), (M2), (M3) and (M4) which originally come from [Kobayashi 1984].

(M1) M is obtained from a Seifert fibered space M_1 over a disk with two exceptional fibers and the exterior M_2 of a nonhyperbolic 1-bridge knot K in a lens space by gluing their boundaries so that the meridian of K is identified with the regular fiber of M_1.

(M2) M is obtained from a Seifert fibered space M_1 over a disk with two or three exceptional fibers and the exterior M_2 of a nonhyperbolic 2-bridge knot K in S^3 by gluing their boundaries so that the meridian of K is identified with the regular fiber of M_1.

(M3) M is obtained from a Seifert fibered space M_1 over a Möbius band with one or two exceptional fibers and the exterior M_2 of a nonhyperbolic 2-bridge knot K in S^3 by gluing their boundaries so that the meridian of K is identified with the regular fiber of M_1.

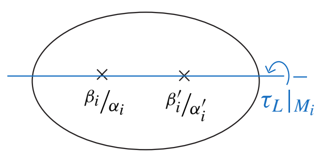
(M4) M is obtained from two Seifert fibered spaces M_1 and M_2 over a disk with two exceptional fibers and the exterior M_3 of a nonhyperbolic 2-bridge link $L = K_1 \cup K_2$ in S^3 by gluing $\partial(M_1 \cup M_2)$ and ∂M_3 so that the meridian of K_i is identified with the regular fiber of M_i $(i = 1, 2)$.

Assume that M satisfies the condition (M1). That is, M is obtained from a Seifert fibered space M_1 over a disk with two exceptional fibers and the exterior M_2 of a nonhyperbolic 1-bridge knot K in a lens space by gluing their boundaries so that the meridian of K is identified with the regular fiber of M_1. By [Kobayashi 1984], M_2 satisfies one of the following.

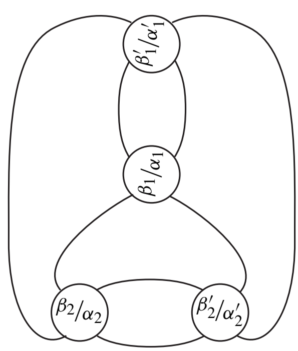
(M1-a) M_2 is a Seifert fibered space over a disk with two exceptional fibers, or

(M1-b) M_2 is a Seifert fibered space over a Möbius band with one exceptional fiber.

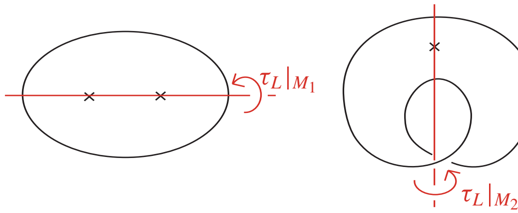
First we assume that M_2 satisfies (M1-a). Recall that the covering involution τ_L satisfies the conditions (T1) and (T2). Since the center of $\pi_1(M)$ is trivial, the strong equivalence class of τ_L is determined by its image in the mapping class group by [Tollefson 1981, Theorem 7.1]. By [Jang 2011, Lemma 4(1)] (or [Jang 2011, Proposition 6(1)]), we may assume that the restriction $\tau_L|_{M_i}$ $(i = 1, 2)$ is a fiber-preserving involution of M_i which induces the involution on the base orbifold:



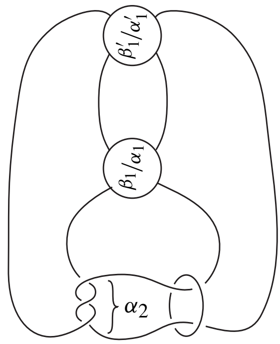
Each quotient orbifold $(M_i, \mathrm{Fix}\tau_L|_{M_i})/\tau_L|_{M_i}$ $(i = 1, 2)$ is a Montesinos pair with two rational tangles. By gluing them so that the image of the meridian of K is identified with the image of the regular fiber of M_1, we see that L must be a 3-bridge link like this (see also [Jang 2011, Section 7, Case 1.1]):



Assume that M_2 satisfies (M1-b). By [Jang 2011, Lemma 4(1) and (2)] together with [Tollefson 1981, Theorem 7.1], we may assume that the restriction $\tau_L|_{M_i}$ ($i=1,2$) is a fiber-preserving involution of M_i that induces the involution on the base orbifold as illustrated here:



By considering the quotient orbifold $(M, \mathrm{Fix}\tau_L)/\tau_L$, we see that L is equivalent to a 3-bridge link of this form (see also [Jang 2011, Section 7, Case 1.2]):



The remaining cases can be treated similarly except for the case where M satisfies the condition (M3). Thus, in the rest of this section, we assume that M satisfies the condition (M3). That is, M is obtained from a Seifert fibered space M_1 over a Möbius band with one or two exceptional fibers and the exterior M_2 of a nonhyperbolic 2-bridge knot K in S^3 by gluing their boundaries so that the meridian of K is identified with the regular fiber of M_1. By an argument similar to those for the previous cases, we can see that L is equivalent to the link in Figure 5 on the next page. For that link, we may assume that the rational number β_1/α_1 is not an integer, and that the rational number β_2/α_2 is an integer or not an integer according to whether the number of the exceptional fibers of M_1 is one or two. We can see that the bridge number of the link $K_1 \cup K_2$ in the figure is at least 4, since K_1 is a 3-bridge link by [Boileau and Zieschang 1985] and [Jang 2011]. However, by [Boileau and Zieschang 1985, Lemma 1.7] and [Boileau and Zimmermann 1989, Corollary 3.3], we have $w(K_1 \cup K_2) \geq w(K_1) + w(K_2) = 3 + 1 = 4$, which contradicts the assumption that $w(L) = 3$.

This completes the proof of Theorem 1.2 for arborescent knots.

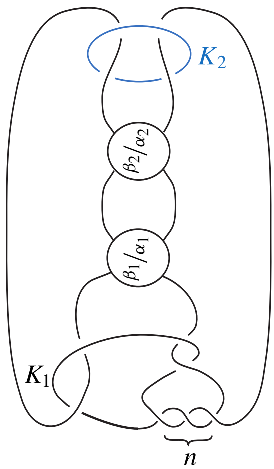


Figure 5. Link equivalent to L when M satisfies condition (M3); see previous page.

4. Subgroups generated by meridians

In this section we study subgroups of knot and link groups that are generated by two or three meridians and we give a proof of Theorem 1.5.

For L a link in S^3 and the link space $E(L)$, choose annuli and tori as follows:

(1) Let $\{A_1, \dots, A_n\}$ be a maximal collection of disjoint nonparallel and properly embedded essential annuli in $E(L)$ whose boundaries are meridians. Thus the closures of the components of $E(L) \setminus \bigcup_{1 \le i \le n} A_i$ are the link spaces $E(L_1), \dots, E(L_k)$ of the prime factors L_i of L.

(2) Let $\{T_1, \dots, T_m\}$ be the union of the characteristic families of tori of the manifolds $E(L_i)$ for $1 \le i \le n$.

Thus the closures of the components of

$$E(L) \setminus \Biggl(\biggl(\bigcup_{1\le i\le n} A_i\biggr) \cup \biggl(\bigcup_{1\le i\le m} T_i\biggr)\Biggr)$$

are the pieces of the JSJ-decompositions of the link spaces $E(L_i)$ with $1 \le i \le n$. We call such a piece peripheral if it meets a boundary component of $E(L)$.

Now, let $G = \pi_1(E(L))$. Let $\mathbb{A}_L$ be the graph of group decomposition of G corresponding to the splitting of $E(L)$ along the A_i and T_i. Thus the vertex groups are the fundamental groups of pieces of the JSJ-decompositions of the $E(L_i)$ and the edge groups are infinite cyclic or isomorphic to $\mathbb{Z}^2$.

Lemma 4.1. *Let L be as above, $G := \pi_1(E(L))$ and $m_1, \dots, m_k \in G$ be meridians (not necessarily corresponding to the same component of L).*

Then either $\langle m_1, \dots, m_k \rangle$ *is free or there exist meridians* $m'_1, \dots, m'_k \in G$ *such that the following hold*:

(1) $(m_1, \dots, m_k)$ *is Nielsen-equivalent to* $(m'_1, \dots, m'_k)$ *and* m_i *is conjugate to* m'_i *for* $1 \le i \le k$.

(2) *There exist* $i \neq j \in \{1, \dots, k\}$ *such that* $\langle m'_i, m'_j \rangle$ *is conjugate to a vertex group of* $\mathbb{A}_L$ *that corresponds to a peripheral piece of some* $E(L_i)$. *Moreover* m'_i *and* m'_j *are conjugate to meridians in this vertex group.*

Proof. We consider the action of G on the Bass–Serre tree T corresponding to $\mathbb{A}_L$. Any m_i acts elliptically and the fixed point set of m_i coincides with the fixed point set of m_i^n for any $n \neq 0$. This is true as m_i is a peripheral element and therefore not a proper root of the regular fiber of any Seifert piece.

Moreover for all $i \in \{1, \dots, k\}$ the element m_i (and therefore also m_i^n with $n \neq 0$) fixes no edge corresponding to a canonical torus of the JSJ-decomposition of some $E(L_i)$ as no power of the meridian is freely homotopic to a curve in one of these tori.

It now follows from [Weidmann 2002, Theorem 7] applied to $(\{m_1\}, \dots, \{m_k\}, \varnothing)$ that either $\langle m_1, \dots, m_k \rangle$ is free or that there exist elements $m'_1, \dots, m'_k$ such that the following hold:

(1) $(m_1, \dots, m_k)$ is Nielsen-equivalent to $(m'_1, \dots, m'_k)$.

(2) m_i is conjugate to m'_i for $1 \le i \le k$.

(3) There exist $i \neq j \in \{1, \dots, k\}$ such that nontrivial powers of m'_i and m'_j fix a common vertex of T.

This implies in particular that m'_i is a meridian for $1 \le i \le k$. The above remark further implies that not only powers of m'_i and m'_j but m'_i and m'_j themselves fix a common vertex v of T that is therefore also fixed by $\langle m'_i, m'_j \rangle$. As both m'_i and m'_j only fix vertices of T that correspond to peripheral pieces, it follows that v corresponds to a peripheral piece. As no meridian is conjugate in a peripheral piece to an element corresponding to one of the characteristic tori it follows moreover that m'_i and m'_j are conjugate to meridians in the stabilizer of v. □

Proposition 4.2. *Let* K *be a knot in* S^3 *and* $G := \pi_1(E(K))$. *If* $m_1, m_2 \in G$ *are meridians that generate a nonfree subgroup of* G *then* K *has a prime factor* K_1 *that is a* 2*-bridge knot and* $\langle m_1, m_2 \rangle$ *is conjugate to the subgroup of* G *corresponding to* K_1.

Proof. It follows from Lemma 4.1 that $\langle m_1, m_2 \rangle$ lies in the subgroup corresponding to a peripheral piece of $E(K)$. Thus $\langle m_1, m_2 \rangle$ is contained in the subgroup corresponding to the peripheral piece M of the JSJ-decomposition of a prime factor K_1

of K. Moreover m_1 and m_2 are in this subgroup conjugate to the meridian. We distinguish two cases:

Suppose first that M is Seifert fibered. Thus M is a torus knot space or a cable space. In the first case it follows from [Rost and Zieschang 1987] that either $\langle m_1, m_2\rangle$ is free or that $\langle m_1, m_2\rangle = \pi_1(M)$ and that M is the exterior of a 2-bridge knot which proves the claim. In the second case M is the mapping torus of a disk with finitely many punctures with respect to an automorphism of finite order. Moreover (like all elements conjugate to a meridian) both m_1 and m_2 lie in the free fundamental group of the fiber which implies that $\langle m_1, m_2\rangle$ is free.

Suppose now that M is hyperbolic. We may assume that $\langle m_1, m_2\rangle$ is not abelian as two conjugates of the meridian that generate an abelian group must lie in the same conjugate of the same peripheral subgroup and therefore generate a cyclic subgroup.

It follows from Proposition 2 of [Boileau and Weidmann 2005] that either $\langle m_1, m_2\rangle = \pi_1(M)$ and that M is the exterior of a 2-bridge knot or that $|\pi_1(M) : \langle m_1, m_2\rangle| = 2$ and the 2-sheeted cover $\tilde{M}$ of M corresponding to $\langle m_1, m_2\rangle$ is the exterior of a 2-bridge link with 2 components.

In the first case the conclusion is immediate. Suppose now that the second case occurs. As m_1 and m_2 is conjugate in $\pi_1(M)$ it follows that both boundary components of $\tilde{M}$ cover the same boundary component of M, in particular M is a knot exterior. Now $\langle m_1, m_2\rangle$ contains a conjugate of the peripheral subgroup of $\pi_1(M)$ and is normal in $\pi_1(M)$. It follows that $\langle m_1, m_2\rangle$ contains all parabolic elements of $\pi_1(M)$. As $\pi_1(M)$ is a knot group, it is generated by parabolic elements. It follows that $\pi_1(M) = \langle m_1, m_2\rangle$ which yields a contradiction. □

The rest of this section is devoted to the proofs of Theorem 1.5 and Corollary 1.7.

Proof of Theorem 1.5. It follows from Lemma 4.1 that we may assume that $\langle m_1, m_2\rangle$ fixes a vertex v of the Bass–Serre tree that corresponds to the peripheral piece M of $S^3 \setminus K$ and m_1 and m_2 are conjugate to meridians in $\pi_1(M)$. By Proposition 4.2 the group $\langle m_1, m_2\rangle$ is free.

Choose a torus T of the characteristic family of tori for $S^3 \setminus K$ such that T cuts $S^3 \setminus K$ into two pieces, a geometric knot space N and its complement $\widehat{M}$. Clearly M is contained in $\widehat{M}$. Let $W = S^3 \setminus \mathrm{int(N)}$ be the solid torus containing M. Since m_1 and m_2 are conjugate to meridians in $\pi_1(M)$, they are null-homologous in W and so is any element of $\langle m_1, m_2\rangle$. The meridian of N and its powers are the only elements of $\pi_1(T) = \partial W$ which are null-homologous in W, therefore the subgroup $\langle m_1, m_2\rangle$ intersects any conjugate of the free abelian subgroup $\pi_1(T) \subset G = \pi_1(S^3 \setminus K)$ at most in a subgroup of the cyclic group generated by the meridian of N. Consider the action of G on the Bass–Serre tree corresponding to the amalgamated product $\pi_1(N) *_{\pi_1(T)} \pi_1(\widehat{M})$. Let v be the vertex fixed by $\langle m_1, m_2\rangle$, note that v corresponds to $\pi_1(\widehat{M})$. As the meridian of N does not agree with the fiber of N if N is Seifert fibered, it follows that no element of $\langle m_1, m_2\rangle$ fixes a vertex at distance more than 1 from v. Moreover m_3 fixes a single vertex that corresponds to $\pi_1(\widehat{M})$. By applying Theorem 7 of [Weidmann 2002] to $(\{m_1, m_2\}, \{m_3\})$ it follows that either m_3 also fixes v or that $\langle m_1, m_2, m_3\rangle \cong \langle m_1, m_2\rangle * \langle m_3\rangle \cong F_3$. This proves the claim. □

Corollary 1.6 is a direct consequence of Theorem 1.5. We prove now Corollary 1.7.

Proof of Corollary 1.7. Let $K \subset S^3$ be a knot such that $w(K) = 3$. If K is prime, then Theorem 1.5 implies that K is a hyperbolic knot or a torus knot. If $K = K_1 \sharp K_2$ is a nontrivial connected sum, then the 2-fold cover $M_2(K)$ of S^3 branched along K is the nontrivial connected sum $M_2(K_1) \sharp M_2(K_2)$ of the 2-fold branched covers of K_1 and K_2. Since $w(K) = 3$, it follows that $\pi_1(M_2(K))$ is generated by two elements. Since

$$\pi_1(M_2(K)) = \pi_1(M_2(K_1)) * \pi_1(M_2(K_2))$$

is a free product of nontrivial groups, by the orbifold theorem (see [Boileau and Porti 2001]), it follows that each group $\pi_1(M_2(K_1))$ and $\pi_1(M_2(K_2))$ is cyclic. Again the orbifold theorem allows us to conclude that K_1 and K_2 are 2-bridge knots. □

5. Proof of Theorem 1.2

Let L be a link in S^3, and suppose that the 2-fold branched cover $M := M_2(L)$ of S^3 branched along L is a graph manifold. Since we have already treated the case when L is an arborescent link in Section 3, we assume here that L is not an arborescent link and that $w(L) = 3$.

We first assume that M is a Seifert fibered space. Then L is either a (generalized) Montesinos link or a Seifert link, i.e., $S^3 \setminus L$ admits a Seifert fibration. If L is a (generalized) Montesinos link or a torus link, then we have $b(L) = 3$ by [Boileau and Zieschang 1985; Rost and Zieschang 1987]. So we assume that L is a Seifert link which is not a torus link. By [Burde and Murasugi 1970], we see that L is the union of a torus knot of type $(2, b)$ and its *core* of index 2, in which case it is easy to see that $b(L) = 3$.

Next we assume that M is not a Seifert fibered space. Let $T = T_1 \cup \cdots \cup T_k$ be tori which give the JSJ-decomposition of M. As in Section 3, we can see that M is a genus-2 manifold and the covering involution τ_L is a realization of an inversion of $\pi_1(M)$. Let $\alpha := (\tau_L)_*$ be the automorphism of $\pi_1(M)$ and let g and h be a pair of generators for $\pi_1(M)$. By [Boileau and Weidmann 2008, Proposition 20], τ_L respects the JSJ-decomposition of M and the Seifert fibered structures on the JSJ pieces. Let Q be the oriented circle bundle over the Möbius band. We follow the argument in [Boileau and Weidmann 2005, Section 3], under the assumption that M is a genus-2 closed manifold. We first deal with the following case.

5.1. ***The JSJ-decomposition has a separating torus and no piece homeomorphic to Q.*** Let T_1 be the separating torus by changing order if necessary, and let M_A and M_B be the two submanifold of M divided by T_1. By the argument in [Boileau and Weidmann 2005], we see that M_A is a Seifert fibered space, g is a root of a fiber of M_A and $g^n \in \pi_1(T_1)$. Moreover, one of the following holds.

(i) M_A is a Seifert fibered space over a disk with two exceptional fibers and M_B is the exterior of a 1-bridge knot in a lens space.

(ii) M_A is a Seifert fibered space over a disk with two exceptional fibers and M_B is the exterior of a nonhyperbolic 2-bridge knot in S^3.

(iii) M_A is a Seifert fibered space over a disk with two exceptional fibers and M_B is decomposed by T_2 into two pieces $M_B^{(1)}$ and $M_B^{(2)}$, where $M_B^{(1)}$ is the exterior of a 2-component nonhyperbolic 2-bridge link in S^3 and where $M_B^{(2)}$ is a Seifert fibered space over a disk with two exceptional fibers.

(iv) M_A is a Seifert fibered space over a Möbius band with one or two exceptional fibers and M_B is the exterior of a nonhyperbolic 2-bridge knot in S^3.

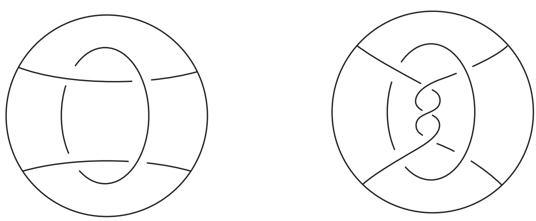
(v) M_A is a Seifert fibered space over a disk with three exceptional fibers and M_B is the exterior of a nonhyperbolic 2-bridge knot in S^3.

Here, the boundaries of M_A and M_B are glued so that the fiber of M_A is identified with the meridian of M_B.

First assume that (i) is satisfied. Since

$$\alpha(g^n) = g^{-n},$$

we see that $\tau_L|_{T_1}$ is hyperelliptic. Note that $\tau_L|_{T_1}$ extends to M_B in a unique way and the quotient of M_B by $\tau_L|_{M_B}$ gives a tangle as in the figure below, right (see

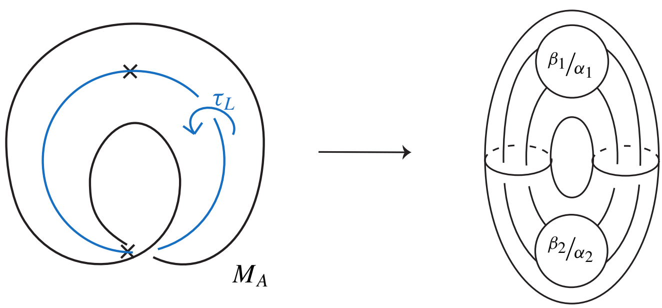


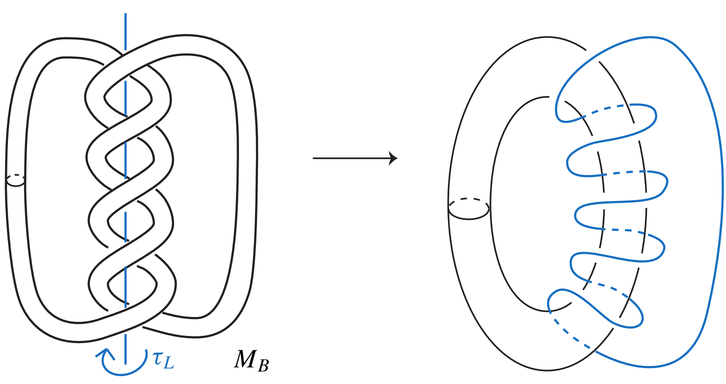
[Jang 2011, Lemma 9]). Since we assume that L is not an arborescent link, we see that τ_L exchanges the two exceptional fibers of M_A. This implies that the two exceptional fibers of M_A have the same index. Then the quotient of M_A by $\tau_L|_{M_A}$ is obtained from the tangle in the figure above, left, by applying Dehn surgery along the loop component in the tangle, where the surgery slope is the reciprocal of the index of the exceptional fibers of M_A. Hence the quotient of M by τ_L is a nontrivial lens space, a contradiction.

Assume that (ii) is satisfied. Note that M_B is a Seifert fibered space over a disk with two exceptional fibers of indices $1/2$ and $-n/(2n+1)$. Thus the involution on M_B which is hyperelliptic on the boundary is unique (see [Jang 2011, Lemma 4(1)] for example). By an argument similar to that for the previous case, we can lead to a contradiction.

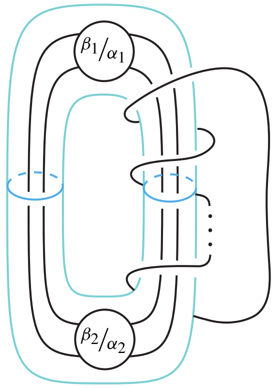
Assume that (iii) is satisfied. Then we see that either $\tau_L(T_i) = T_i$ and $\tau_L|_{T_i}$ is hyperelliptic ($i = 1, 2$) or $\tau_L(T_1) = T_2$. In the former case, we can use arguments similar to those in the previous cases to lead to a contradiction. In the latter case, M_A and $M_B^{(2)}$ are homeomorphic and τ_L interchanges the two pieces. Denote by N the quotient of $M_B^{(1)}$ by $\tau_L|_{M_B^{(1)}}$, which is a solid torus, and denote by F the image of the fixed point set. Then the exterior of F in N is homeomorphic to the exterior of a torus link of type $(2, 2m)$. The quotient of M by τ_L, which is supposed to be S^3, is obtained by gluing M_A and a solid torus, which implies that M_A is homeomorphic to the exterior of a torus knot (see [Burde and Murasugi 1970]). Thus L is a nontrivial cable knot of a torus knot. By Corollary 1.6, we have $w(L) \geq 4$, a contradiction.

Assume that (iv) is satisfied. By arguments similar to those for the previous cases, we can see that $\tau_L|_{M_A}$ and $\tau_L|_{M_B}$ are equivalent to the involutions illustrated here:





Hence, the quotient of M_A gives a 2-bridge link in a solid torus and the quotient of M_B gives a component of a torus link of type $(2, 2m)$ with the regular neighborhood of the other component removed. Then we obtain this 3-bridge link (see [Jang 2012]):



Assume that (v) is satisfied. We can lead to a contradiction by arguments similar to those for the previous cases.

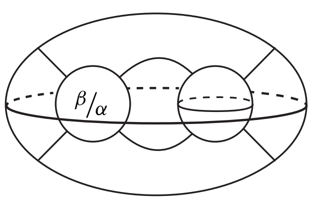
5.2. ***The JSJ-decomposition has a nonseparating torus.*** Since the genus of M is 2, M consists of one or two Seifert pieces.

We first deal with the case when M consists of one Seifert piece. By an argument of [Boileau and Weidmann 2005], we have the following two cases.

(i) The torus T cuts M into the exterior of a 2-component nonhyperbolic 2-bridge link, and g and hgh^{-1} are the meridians.

(ii) The torus T cuts M into a Seifert fibered space over an annulus with two exceptional fibers, whose boundary components are glued so that the fibers are identified.

When (ii) holds, M is a Seifert fibered space, a contradiction. Hence assume that (i) holds. Note that the closure of $M \setminus T$ is a Seifert fibered space, say M', over an annulus with one exceptional fiber. Since we assume that M is not a Seifert fibered space, the fibers on the two boundary components of M' do not match. Since g is a meridian of the 2-bridge link, we can see that $\tau_L|_T$ is hyperelliptic. Then the quotient of M' by $\tau_L|_{M'}$ gives a (3, 1)-manifold pair in the following diagram:



The quotient of M by τ_L is obtained from $S^3 \setminus (B_1 \cup B_2)$, where B_1 and B_2 are open 3-balls, by gluing the two 2-spheres ∂B_1 and ∂B_2, and hence the quotient of M cannot be homeomorphic to S^3, a contradiction.

Next we deal with the case when M consists of two Seifert pieces M_A and M_B. By [Kobayashi 1984], M_A is a Seifert fibered space over an annulus with one or two exceptional fibers and M_B is the exterior of a 2-component nonhyperbolic 2-bridge link. By arguments similar to those for previous cases (compare [Jang 2012]), we can see that L is equivalent to a link having the form shown at the top of the previous page.

5.3. ***There exists a piece homeomorphic to Q.*** By [Kobayashi 1984], we have the following cases.

(i) M consists of two JSJ pieces homeomorphic to Q.

(ii) M consists of two JSJ pieces, one of which is homeomorphic to Q, and the other is either a Seifert fibered space over a disk with two exceptional fibers or a Seifert fibered space over a Möbius band with one exceptional fiber.

(iii) M consists of three JSJ pieces, one of which is homeomorphic to Q, the second piece is the exterior of a 2-component nonhyperbolic 2-bridge link and the third piece is a Seifert fibered space over a disk with two exceptional fibers.

Assume that (i) is satisfied. By [Boileau and Weidmann 2005, Lemma 17], the regular fibers of the two pieces, considered as a Seifert fibered space over a disk with two exceptional fibers, intersect in one point, and g^2 is a fiber of one piece. Then we see that $\tau_L|_T$ is hyperelliptic, and we can lead to a contradiction by using arguments similar to those in the previous cases.

Assume that (ii) is satisfied. By an argument in [Boileau and Weidmann 2005, Proof of Lemma 18], we can see that $\tau_L|_T$ is hyperelliptic, and we can lead to a contradiction by using arguments similar to those in the previous cases.

Assume that (iii) is satisfied. Similarly, we can see that either $\tau_L(T_i) = T_i$ and $\tau_L|_{T_i}$ is hyperelliptic ($i = 1, 2$) or $\tau_L(T_1) = T_2$. In the former case, we can lead to a contradiction by using arguments similar to those in the previous cases. In the latter case, we can see that the quotient of M by τ_L is the union of Q and a solid torus, which cannot be homeomorphic to S^3, a contradiction.

This completes the proof of Theorem 1.2.

Proof of Corollary 1.4. Let M be a closed orientable graph manifold which admits an inversion, i.e., $\pi_1(M)$ is generated by two elements g and h and there exists an automorphism α of $\pi_1(M)$ which sends g and h to g^{-1} and h^{-1}, respectively. If M is a Seifert fibered space, then α is hyperelliptic by [Boileau and Weidmann 2008, Theorem 5]. If M is not a Seifert fibered space, then α is hyperelliptic by Theorem 1.2 and [Boileau and Weidmann 2008, Proposition 20(3)]. □

6. Degree-one maps

Proof of Proposition 1.10. (a) Let $L' \subset S^3$ such that $b(L') = 3$, then $w(L') = 3$ by [Boileau and Zimmermann 1989]. Thus if $L \geq L'$, then $b(L) \geq w(L) \geq w(L') = 3$.

(b) Let $L' \subset S^3$ such that $b(L') = 4$. Assume that $L \geq L'$ and that the 2-fold branched cover M of L is a graph manifold. The degree-one map $f : E(L) \to E(L')$ between the exteriors of L and L' which preserves the meridians lifts to a degree-one map $\tilde{f} : \tilde{E}(L) \to \tilde{E}(L')$ between their 2-fold covers, which extends to a degree-one map $\tilde{f} : M \to M'$ between their 2-fold branched covers $M := M_2(L)$ and $M' = M_2(L')$. Since M is a graph manifold, its simplicial volume $\|M\| = 0$. The existence of the degree-one map $\tilde{f} : M \to M'$ implies that $\|M'\| \leq \|M\|$ and thus $\|M'\| = 0$. By the orbifold theorem [Boileau and Porti 2001] M' admits a geometric decomposition and thus is a connected sum of graph manifolds. Therefore L' is a connected sum of links whose 2-fold branched covers are graph manifolds.

If L' is prime, it follows from Corollary 1.3 that $w(L') = 4$ and therefore $b(L) \geq w(L) \geq w(L') = 4$.

If L' is not prime, then $L' = L'_1 \sharp L'_2$ with $b(L'_1) = 2 = w(L'_1)$ and $b(L'_2) = 3 = w(L'_2)$ by [Boileau and Zimmermann 1989]. The exterior $E(L')$ can be split along a properly embedded essential annulus A into two pieces homeomorphic to $E(L'_1)$ and $E(L'_2)$ so that $\pi_1(E(L') = \pi_1(E(L'_1)) *_{\pi_1(A)} \pi_1(E(L'_2)$, where $\pi_1(A) \cong \mathbb{Z}$ is generated by a meridian of L'_1 and L'_2. By killing the meridians of L'_2 which are not conjugate to the generator of $\pi_1(A)$, one can define an epimorphism $\phi_1 : \pi_1(E(L')) \to \pi_1(E(L'_1))$ such that the restriction of ϕ_1 to $\pi_1(E(L'_1))$ is the identity and $\phi_1(\pi_1(E(L'_2))) = \pi_1(A)$. In the same way one can define an epimorphism $\phi_2 : \pi_1(E(L')) \to \pi_1(E(L'_2))$ such that the restriction of ϕ_2 to $\pi_1(E(L'_2))$ is the identity and $\phi_2(\pi_1(E(L'_1))) = \pi_1(A)$. These epimorphisms imply that $w(L') = w(L'_1) + w(L'_2) - 1 = 4$, and thus $b(L) \geq w(L) \geq w(L') = 4$. □

References

[Boileau and Porti 2001] M. Boileau and J. Porti, *Geometrization of 3-orbifolds of cyclic type*, Astérisque **272**, Société Mathématique de France, Paris, 2001. MR Zbl

[Boileau and Weidmann 2005] M. Boileau and R. Weidmann, "The structure of 3-manifolds with two-generated fundamental group", *Topology* **44**:2 (2005), 283–320. MR Zbl

[Boileau and Weidmann 2008] M. Boileau and R. Weidmann, "On invertible generating pairs of fundamental groups of graph manifolds", pp. 105–128 in *The Zieschang Gedenkschrift*, edited by M. Boileau et al., Geom. Topol. Monogr. **14**, 2008. MR Zbl

[Boileau and Zieschang 1983] M. Boileau and H. Zieschang, "Genre de Heegaard d'une variété de dimension 3 et générateurs de son groupe fondamental", *C. R. Acad. Sci. Paris Sér. I Math.* **296**:22 (1983), 925–928. MR Zbl

[Boileau and Zieschang 1984] M. Boileau and H. Zieschang, "Heegaard genus of closed orientable Seifert 3-manifolds", *Invent. Math.* **76**:3 (1984), 455–468. MR Zbl

[Boileau and Zieschang 1985] M. Boileau and H. Zieschang, "Nombre de ponts et générateurs méridiens des entrelacs de Montesinos", *Comment. Math. Helv.* **60**:2 (1985), 270–279. MR Zbl

[Boileau and Zimmermann 1989] M. Boileau and B. Zimmermann, "The π-orbifold group of a link", *Math. Z.* **200**:2 (1989), 187–208. MR Zbl

[Boileau et al. 2003] M. Boileau, S. Maillot, and J. Porti, *Three-dimensional orbifolds and their geometric structures*, Panoramas et Synthèses **15**, Société Mathématique de France, Paris, 2003. MR Zbl

[Bonahon and Siebenmann 2016] F. Bonahon and L. C. Siebenmann, "New geometric splittings of classical knots, and the classification and symmetries of arborescent knots", preprint, 2016, Available at http://tinyurl.com/bonsieb.

[Burde and Murasugi 1970] G. Burde and K. Murasugi, "Links and Seifert fiber spaces", *Duke Math. J.* **37** (1970), 89–93. MR Zbl

[Cornwell 2014] C. R. Cornwell, "Knot contact homology and representations of knot groups", *J. Topol.* **7**:4 (2014), 1221–1242. MR Zbl

[Cornwell and Hemminger 2016] C. R. Cornwell and D. R. Hemminger, "Augmentation rank of satellites with braid pattern", *Comm. Anal. Geom.* **24**:5 (2016), 939–967. MR Zbl

[Futer and Guéritaud 2009] D. Futer and F. Guéritaud, "Angled decompositions of arborescent link complements", *Proc. Lond. Math. Soc.* (3) **98**:2 (2009), 325–364. MR Zbl

[Jang 2011] Y. Jang, "Classification of 3-bridge arborescent links", *Hiroshima Math. J.* **41**:1 (2011), 89–136. MR Zbl

[Jang 2012] Y. Jang, "Characterization of 3-bridge links with infinitely many 3-bridge spheres", *Topology Appl.* **159**:4 (2012), 1132–1145. MR Zbl

[Kirby 1978] R. Kirby, "Problems in low dimensional manifold theory", pp. 273–312 in *Algebraic and geometric topology, II* (Stanford, 1976), edited by R. J. Milgram, Proc. Sympos. Pure Math. **32**, American Mathematical Society, Providence, RI, 1978. MR Zbl

[Kobayashi 1984] T. Kobayashi, "Structures of the Haken manifolds with Heegaard splittings of genus two", *Osaka J. Math.* **21**:2 (1984), 437–455. MR Zbl

[Li 2013] T. Li, "Rank and genus of 3-manifolds", *J. Amer. Math. Soc.* **26**:3 (2013), 777–829. MR Zbl

[Lustig and Moriah 1993] M. Lustig and Y. Moriah, "Generalized Montesinos knots, tunnels and $\mathcal{N}$-torsion", *Math. Ann.* **295**:1 (1993), 167–189. MR Zbl

[Rost and Zieschang 1987] M. Rost and H. Zieschang, "Meridional generators and plat presentations of torus links", *J. London Math. Soc.* (2) **35**:3 (1987), 551–562. MR Zbl

[Schubert 1954] H. Schubert, "Über eine numerische Knoteninvariante", *Math. Z.* **61** (1954), 245–288. MR Zbl

[Schultens and Weidmann 2007] J. Schultens and R. Weidmann, "On the geometric and the algebraic rank of graph manifolds", *Pacific J. Math.* **231**:2 (2007), 481–510. MR Zbl

[Tollefson 1981] J. L. Tollefson, "Involutions of sufficiently large 3-manifolds", *Topology* **20**:4 (1981), 323–352. MR Zbl

[Weidmann 2002] R. Weidmann, "The Nielsen method for groups acting on trees", *Proc. London Math. Soc.* (3) **85**:1 (2002), 93–118. MR Zbl

[Weidmann 2003] R. Weidmann, "Some 3-manifolds with 2-generated fundamental group", *Arch. Math.* (*Basel*) **81**:5 (2003), 589–595. MR Zbl

Received November 7, 2015. Revised May 3, 2017.

MICHEL BOILEAU
AIX-MARSEILLE UNIV.
CNRS, CENTRALE MARSEILLE
MARSEILLE
FRANCE
michel.boileau@univ-amu.fr

YEONHEE JANG
DEPARTMENT OF MATHEMATICS
NARA WOMEN'S UNIVERSITY
KITAUOYANISHI-MACHI
NARA
JAPAN
yeonheejang@cc.nara-wu.ac.jp

RICHARD WEIDMANN
MATHEMATISCHES SEMINAR
CHRISTIAN-ALBRECHTS-UNIVERSITÄT ZU KIEL
KIEL
GERMANY
weidmann@math.uni-kiel.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 292, No. 1, 2018
dx.doi.org/10.2140/pjm.2018.292.81

POINTWISE CONVERGENCE OF ALMOST PERIODIC FOURIER SERIES AND ASSOCIATED SERIES OF DILATES

CHRISTOPHE CUNY AND MICHEL WEBER

Let $\mathcal{S}^2$ be the Stepanov space with norm $\|f\|_{\mathcal{S}^2} = \sup_{x\in\mathbb{R}}\big(\int_x^{x+1}|f(t)|^2\,dt\big)^{1/2}$, $\lambda_n \uparrow \infty$, and let $(a_n)_{n\ge1}$ satisfy Wiener's condition $\sum_{n\ge1}\big(\sum_{k:n\le\lambda_k\le n+1}|a_k|\big)^2 < \infty$. We establish the following maximal inequality:

$$\Big\|\sup_{N\ge1}\Big|\sum_{n=1}^{N} a_n \mathrm{e}^{i\lambda_n t}\Big|\Big\|_{\mathcal{S}^2} \le C\bigg(\sum_{n\ge1}\bigg(\sum_{k:n\le\lambda_k\le n+1}|a_k|\bigg)^2\bigg)^{1/2},$$

where $C>0$ is a universal constant. Moreover, the series $\sum_{n\ge1} a_n\mathrm{e}^{it\lambda_n}$ converges for λ-a.e. $t\in\mathbb{R}$. We give a simple and direct proof. This contains as a special case Hedenmalm and Saksman's result for Dirichlet series. We also obtain maximal inequalities for corresponding series of dilates. Let $(\lambda_n)_{n\ge1}$, $(\mu_n)_{n\ge1}$, be nondecreasing sequences of real numbers greater than 1. We prove the following interpolation theorem. Let $1\le p,q\le 2$ be such that $1/p+1/q=\frac{3}{2}$. There exists $C>0$ such that for any sequences $(\alpha_n)_{n\ge1}$ and $(\beta_n)_{n\ge1}$ of complex numbers such that $\sum_{n\ge1}\big(\sum_{k:n\le\lambda_k<n+1}|\alpha_k|\big)^p<\infty$ and $\sum_{n\ge1}\big(\sum_{k:n\le\mu_k<n+1}|\beta_k|\big)^q<\infty$, we have

$$\Big\|\sup_{N\ge1}\Big|\sum_{n=1}^{N}\alpha_n D(\lambda_n t)\Big|\Big\|_{\mathcal{S}^2} \le C\bigg(\sum_{n\ge1}\bigg(\sum_{k:n\le\lambda_k<n+1}|\alpha_k|\bigg)^p\bigg)^{1/p}\bigg(\sum_{n\ge1}\bigg(\sum_{k:n\le\mu_k<n+1}|\beta_k|\bigg)^q\bigg)^{1/q},$$

where $D(t)=\sum_{n\ge1}\beta_n\mathrm{e}^{i\mu_n t}$ is defined in $\mathcal{S}^2$. Moreover, $\sum_{n\ge1}\alpha_n D(\lambda_n t)$ converges in $\mathcal{S}^2$ and for λ-a.e. $t\in\mathbb{R}$. We further show that if $\{\lambda_k,\ k\ge1\}$ satisfies the condition

$$\sum_{\substack{k\ne\ell,\ k'\ne\ell'\\(k,\ell)\ne(k',\ell')}}(1-|(\lambda_k-\lambda_\ell)-(\lambda_{k'}-\lambda_{\ell'})|)_+^2<\infty,$$

then the series $\sum_k a_k\mathrm{e}^{i\lambda_k t}$ converges on a set of positive Lebesgue measure only if the series $\sum_{k=1}^{\infty}|a_k|^2$ converges. The above condition is in particular fulfilled when $\{\lambda_k,\ k\ge1\}$ is a Sidon sequence.

MSC2010: primary 42A75; secondary 42A24, 42B25.

Keywords: almost periodic function, Stepanov space, Carleson theorem, Dirichlet series, dilated function, series, almost everywhere convergence.

1. Introduction

We study almost everywhere convergence properties of almost periodic Fourier series in the Stepanov space $\mathcal{S}^2$ and of corresponding series of dilates. This space is defined as the subspace of functions f of $L^2_{\text{loc}}(\mathbb{R})$ verifying the following analogue of the Bohr almost periodicity property: *For all $\varepsilon > 0$, there exists $K_\varepsilon > 0$ such that for any $x_0 \in \mathbb{R}$, there exists $\tau \in [x_0, x_0 + K_\varepsilon]$ such that $\|f(\cdot + \tau) - f(\cdot)\|_{\mathcal{S}^2} \le \varepsilon$.* The Stepanov norm in $\mathcal{S}^2$ is defined by

$$\|f\|_{\mathcal{S}^2} = \sup_{x \in \mathbb{R}} \left(\int_x^{x+1} |f(t)|^2 \, dt \right)^{1/2}.$$

Recall some basic facts. By the fundamental theorem on almost periodic functions, see [Besicovitch 1932, p. 88], the Stepanov space $\mathcal{S}^2$ coincides with the closure of the set of generalized trigonometric polynomials $\left\{ \sum_{k=1}^n a_k \mathrm{e}^{i\lambda_k t} : \alpha_k \in \mathbb{C},\ \lambda_k \in \mathbb{R} \right\}$ with respect to this norm. It is clear by considering for instance $f = \chi_{[0,1]}$ that the space $\{f \in L^2_{\text{loc}}(\mathbb{R}) : \|f\|_{\mathcal{S}^2} < \infty\}$ is strictly larger than $\mathcal{S}^2$. Introduce also the Besicovitch seminorm of order 2 of $f \in L^2_{\text{loc}}(\mathbb{R})$

$$\|f\|_{\mathcal{B}^2} = \limsup_{T \to \infty} \left(\frac{1}{2T} \int_{-T}^{T} |f(t)|^2 \, dt \right)^{1/2}. \tag{1-1}$$

For every $\lambda \in \mathbb{R}$ and every $f \in L^1_{\text{loc}}(\mathbb{R})$ define the Fourier coefficient $\hat{f}(\lambda)$ of the exponent λ of f by

$$\hat{f}(\lambda) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(x) \mathrm{e}^{-i\lambda x} \, dx, \tag{1-2}$$

whenever the limit exists. It is easily seen, by approximating by generalized trigonometric polynomials in the Stepanov norm, that the above limit exists for every $f \in \mathcal{S}^2$ and every $\lambda \in \mathbb{R}$. Moreover, for any finite family $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, we have by the Parseval equation in $\mathcal{B}^2$, see [Bellman 1944, p. 109],

$$\sum_{k=1}^n |\hat{f}(\lambda_k)|^2 \le \|f\|^2_{\mathcal{B}^2} \le \|f\|^2_{\mathcal{S}^2}.$$

In particular, for $f \in \mathcal{S}^2$, $\Lambda := \{\lambda \in \mathbb{R} : \hat{f}(\lambda) \neq 0\}$ is countable. We call Λ the (set of) Fourier exponents of f. Let $f \in \mathcal{S}^2$ have of Fourier exponents Λ. Then

$$\sum_{\lambda \in \Lambda} |\hat{f}(\lambda)|^2 \le \|f\|^2_{\mathcal{B}^2} \le \|f\|^2_{\mathcal{S}^2}. \tag{1-3}$$

We then define formally the Fourier series of $f \in \mathcal{S}^2$ as

$$\sum_{\lambda \in \Lambda} \hat{f}(\lambda) e^{i\lambda \cdot}.$$

Notice that the set $\Lambda \cap [-A, A]$ may be infinite for a given $A > 0$.

In this paper we are interested in the convergence of the Fourier series of f (to f) either in the Stepanov sense or in the almost everywhere sense, and the same sort of consideration will motivate us in the study of associated series of dilates. This second question is actually our main objective. See Section 3.

Concerning convergence of the Fourier series, it is necessary to recall Bredihina's extension to $\mathcal{S}^2$ of Kolmogorov's theorem asserting that if $s_n(x)$ are the partial sums of the Fourier series of a function $f \in L^2(\mathbb{T})$, then $s_{m_n}(x)$ converges almost everywhere to f provided that $m_{n+1}/m_n \geq q > 1$. Bredihina [1968] showed that the Fourier series of a function in $\mathcal{S}^2$ with α-separated frequencies ($\alpha > 0$), namely $|\lambda_k - \lambda_\ell| \geq \alpha > 0$ for all $k, \ell, k \neq \ell$, converges almost everywhere along any exponentially increasing subsequence. That is, for every $\rho > 1$, the sequence $\left\{\sum_{1\leq k\leq \rho^n} \hat{f}(\lambda_k)\mathrm{e}^{i\lambda_k t}, n \geq 1\right\}$ converges for λ-almost every $t \in \mathbb{R}$. The corresponding maximal inequality has been recently obtained by Bailey [2014] who also considered Stepanov spaces of higher order.

Remark 1.1. For a short proof of Kolmogorov's Theorem, see Marcinkiewicz [1933], who showed that this follows from Fejér's Theorem, see [Zygmund 1968, Theorem 3.4-(III)], and the classical fact that if a series $\sum u_n$ with partial sums s_n has infinitely many lacunary gaps and is summable $(C, 1)$ to sum s, then $s_n \to s$. See Theorem 1.27 in Chapter III of [Zygmund 1968].

In view of Carleson's theorem, a natural question is whether the "full" series converges for any $f \in \mathcal{S}^2$.

That question has been addressed in the very specific situation of Dirichlet series by Hedenmalm and Saksman [2003]. A simplified proof may be found in [Konyagin and Queffélec 2001/02] (see also below). They proved the following. Let λ denote here and throughout the Lebesgue measure on the real line.

Theorem 1.2. *Let $(a_n)_{n\geq 1}$ be complex numbers such that $\sum_{n\geq 1} n|a_n|^2 < \infty$. Then the series $\sum_{n\geq 1} a_n n^{it}$ converges λ-almost everywhere.*

Their condition is optimal when $(a_n)_{n\geq 1}$ is nonincreasing. However, if $(a_n)_{n\geq 1}$ is supported say on $\{2^n : n \geq N\}$ the corresponding series is a standard (periodic) trigonometric series and in that case, the optimality is lost, since the condition is much stronger than Carleson's condition.

On the other hand, it follows from [Wiener 1926] that the series $\sum_{n\geq 1} a_n n^{it}$ converges in $\mathcal{S}^2$ provided that

$$\sum_{n\geq 0}\left(\sum_{k=2^n}^{2^{n+1}-1} |a_k|\right)^2 < \infty. \tag{1-4}$$

More precisely, the sequence of partial sums converges in $\mathcal{S}^2$ to a limit $f \in \mathcal{S}^2$. If $a_n > 0$ for every n, the converse is also true; see [Tornehave 1954].

Our first goal (see the next section) is to prove that (1-4) is sufficient for λ-a.e. convergence and to provide the corresponding maximal inequality. Moreover, it will turn out that the problem of the λ-almost everywhere convergence of the series $\sum_{n\geq 1} a_n \mathrm{e}^{i\lambda_n t}$ can be reduced to the study of Dirichlet series.

In doing so, we obtain a Carleson-type theorem for almost periodic series and make the link with the study of almost everywhere convergence of the Fourier series associated with Stepanov's almost periodic functions.

Then, in Section 3, we consider associated series of dilates and obtain a sufficient condition for almost everywhere convergence. We further prove an interpolation theorem. Finally, in Section 4, we obtain a general necessary condition for the convergence almost everywhere of series of functions. The condition involves correlations of order 4. As an application, we show for instance that if $\{\lambda_k, k \geq 1\}$ is a Sidon sequence, and the series $\sum_k a_k \mathrm{e}^{i\lambda_k t}$ converges on a set of positive λ-measure, then the series $\sum_{k=1}^{\infty} |a_k|^2$ converges.

2. Almost everywhere convergence of almost periodic Fourier series

We start with the proof by Konyagin and Queffélec of Hedenmalm and Saksman's result, to which we add a maximal inequality.

Proposition 2.1. *There exists $C > 0$ such that for any sequence $(a_n)_{n\geq 1}$ of complex numbers such that $\sum_{n\geq 1} n|a_n|^2 < \infty$,*

$$\Big\| \sup_{n\geq 1} \Big| \sum_{k=1}^{n} a_k k^{i\cdot} \Big| \Big\|_{\mathcal{S}^2} \leq C \Big(\sum_{n\geq 1} n|a_n|^2 \Big)^{1/2}. \tag{2-1}$$

Before giving the proof, it is necessary to recall some classical but important facts. Let $g \in L^p(\mathbb{T}),\ 1 < p < \infty$. Consider the maximal operator

$$T^* g(x) = \sup_{L=0}^{\infty} \Big| \sum_{|k|\leq L} \hat{g}(k) \mathrm{e}^{2i\pi kx} \Big|.$$

For $f \in L^p(\mathbb{R})$ consider analogously the maximal operator

$$C^* f(x) = \sup_{T>0} \Big| \int_{-T}^{T} \hat{f}(t) \mathrm{e}^{ixt}\, dt \Big|.$$

An operator U on L^p is called strong (p, p) if $\|Uf\|_p \leq C_p \|f\| p$ for all $f \in L^p$. The fact that strong (p, p), $1 < p < \infty$, for T^* is equivalent to strong (p, p) for C^* follows from known elementary arguments, see [Auscher and Carro 1992, p. 166]. We refer to [Hunt 1968, Theorem 1] concerning the deep fact that T^* is strong (p, p), $1 < p < \infty$ and we shall call it "the Carleson–Hunt theorem" when $p = 2$. We will freely use the fact the C^* is consequently strong (p, p), $1 < p < \infty$.

Proof. We first notice that it is enough to prove that

$$\Big\| \sup_{n\geq 1} \Big| \sum_{k=1}^{n} a_k k^{i\cdot} \Big| \Big\|_{L^2[0,1]} \leq C \Big(\sum_{n\geq 1} n|a_n|^2 \Big)^{1/2}. \tag{2-2}$$

Indeed, then the desired result follows from the fact that

$$\sum_{k=1}^{n} a_k k^{i(t+x)} = \sum_{k=1}^{n} (a_k k^{ix}) k^{it},$$

since we may apply the above estimate to the sequence $(a_n n^{ix})_{n\geq 1}$ whose moduli are the same as the ones of the sequence $(a_n)_{n\geq 1}$.

Let us prove (2-2). Define $h \in L^2(\mathbb{R})$ by setting $h \equiv 0$ on $(-\infty, 1)$ and for every $n \in \mathbb{N},\ h(x) = a_n$ whenever $x \in [n, n+1)$.

Let $N \geq 1$. We have

$$\begin{aligned}\sum_{n=1}^{N} a_n n^{it} &= \sum_{n=1}^{N} a_n \int_n^{n+1} (\mathrm{e}^{it\log n} - \mathrm{e}^{it\log x})\, dx + \int_1^{N+1} h(x) \mathrm{e}^{it\log x}\, dx \\ &= \sum_{n=1}^{N} a_n \int_n^{n+1} (\mathrm{e}^{it\log n} - \mathrm{e}^{it\log x})\, dx + \int_0^{\log(N+1)} \mathrm{e}^x h(\mathrm{e}^x) \mathrm{e}^{itx}\, dx.\end{aligned}$$

Now, for every $x \in [n, n+1)$,

$$|\mathrm{e}^{it\log n} - \mathrm{e}^{it\log x}| \leq \frac{t}{n}.$$

Hence,

$$\sum_{n\geq 1} \Big| a_n \int_n^{n+1} (\mathrm{e}^{it\log n} - \mathrm{e}^{it\log x})\, dx \Big| \leq t \Big(\sum_{n\geq 1} n|a_n|^2 \Big)^{1/2} \Big(\sum_{n\geq 1} \frac{1}{n^3} \Big)^{1/2}.$$

On the other hand, $\int_0^{+\infty} \mathrm{e}^{2x} |h|^2(\mathrm{e}^x)\, dx = \int_1^{+\infty} u|h|^2(u)\, du \leq \sum_{n\geq 1} (n+1)|a_n|^2 < \infty$. Hence, since C^* is strong $(2-2)$,

$$\Big\| \sup_{N\geq 1} \Big| \int_0^{\log(N+1)} \mathrm{e}^x h(\mathrm{e}^x) \mathrm{e}^{itx}\, dx \Big| \Big\|_{2,dt}^2 \leq C \int_0^{+\infty} \mathrm{e}^{2x} |h|^2(\mathrm{e}^x)\, dx.$$

Hence (2-1) follows. □

We now derive an improved version of Proposition 2.1.

Theorem 2.2. *There exists $C > 0$ such that for every sequence $(a_n)_{n\geq 1}$ of complex numbers satisfying (1-4),*

$$\Big\| \sup_{n\geq 1} \Big| \sum_{k=1}^{n} a_k k^{i\cdot} \Big| \Big\|_{\mathcal{S}^2} \leq C \Big(\sum_{n\geq 0} \Big(\sum_{k=2^n}^{2^{n+1}-1} |a_k| \Big)^2 \Big)^{1/2}. \tag{2-3}$$

Moreover, $\sum_{n\geq 1} a_n n^{it}$ *converges for* λ*-a.e.* $t \in \mathbb{R}$.

Remarks 2.3. The proof of Theorem 2.2 makes use of the Carleson–Hunt theorem (T^* is strong $(2-2)$) and of Proposition 2.1. The latter was proved using that C^* is strong $(2-2)$, which is equivalent to the Carleson–Hunt theorem. On the other hand, given any sequence $(b_n)_{n\geq 1} \in \ell^2$, applying Theorem 2.2 with $(a_n)_{n\geq 1}$ such that $a_{2^k} = b_k$ and $a_n = 0$ otherwise, we see that Theorem 2.2 implies the Carleson–Hunt theorem, hence is equivalent to it. We shall see below that Theorem 2.2 allows one to treat almost everywhere convergence of series $\sum_{n\geq 1} b_n \mathrm{e}^{it\lambda_n}$ for non-decreasing sequences $(\lambda_n)_{n\geq 1}$. Notice that Theorem 2.2 corresponds to the case where $\lambda_n = \log n$. For more on the Carleson–Hunt theorem we refer to [Lacey 2004]. See also [Jørsboe and Mejlbro 1982].

Proof. As in the previous proof, it is enough to prove a maximal inequality in $L^2([0,1])$. We shall first work along the subsequence $(2^n - 1)_{n\geq 1}$.

Let $n \geq 1$ and define $S_{k,n} := \sum_{\ell=2^n}^{k} a_k$ for every $2^n \leq k \leq 2^{n+1}-1$ and $S_{2^n-1,n} = 0$. In particular, for every $2^n \leq k \leq 2^{n+1}-1$,

$$|S_{k,n}| \leq \sum_{j=2^n}^{2^{n+1}-1} |a_j|,$$

a fact that will be used freely in the sequel.

By Abel summation by parts, we have

$$\sum_{k=2^n}^{2^{n+1}-1} a_k k^{it} = \sum_{k=2^n}^{2^{n+1}-1} (S_{k,n} - S_{k-1,n}) k^{it} = \sum_{k=2^n}^{2^{n+1}-1} S_{k,n}(k^{it} - (k+1)^{it}) + 2^{(n+1)it} S_{2^{n+1}-1,n}.$$

Since $2^{(n+1)it} = \mathrm{e}^{i(n+1)t\log 2}$ and by our assumption $\sum_{n\geq 1} |S_{2^{n+1}-1,n}|^2 < \infty$, it follows from Carleson's theorem that

$$\Big\| \sup_{N\geq 1} \sum_{n=1}^{N} S_{2^{n+1}-1,n} 2^{(n+1)it} \Big\|_{L^2([0,1],dt)} \leq C \Big(\sum_{n\geq 1} |S_{2^{n+1}-1,n}|^2 \Big)^{1/2}.$$

Hence, we are back to controlling the L^2-norm of

$$\sup_{N\geq 1} \Big| \sum_{n=1}^{N} \sum_{k=2^n}^{2^{n+1}-1} S_{k,n}(k^{it} - (k+1)^{it}) \Big|.$$

But we have

$$\begin{aligned} k^{it} - (k+1)^{it} &= \mathrm{e}^{it\log k} - \mathrm{e}^{it\log(k+1)} \\ &= \mathrm{e}^{it\log k}\Big(1 - \mathrm{e}^{it\log(1+1/k)} + \frac{it}{k}\Big) - \frac{it}{k}\mathrm{e}^{it\log k} = u_k(t) - \frac{it}{k}\mathrm{e}^{it\log k}. \end{aligned}$$

Now there exists $C > 0$ such that $|u_k(t)| \le C(t+t^2)/k^2$. Hence,

$$\sum_{n\ge 1}\sum_{k=2^n}^{2^{n+1}-1} |S_{k,n}||u_k(t)| \le C(t+t^2)\sum_{n\ge 1}\frac{\sum_{k=2^n}^{2^{n+1}-1}|a_k|}{2^n}$$
$$\le C(t+t^2)\Biggl(\sum_{n\ge 1}\Biggl(\sum_{k=2^n}^{2^{n+1}-1}|a_k|\Biggr)^2\Biggr)^{1/2}.$$

It remains to control

$$\sup_{N\ge 1}\Biggl|\sum_{n=1}^{N}\sum_{k=2^n}^{2^{n+1}-1}\frac{S_{k,n}}{k}\mathrm{e}^{it\log k}\Biggr|.$$

But we are exactly in the situation of Proposition 2.1. Hence

$$\Biggl\|\sup_{N\ge 1}\Biggl|\sum_{n=1}^{N}\sum_{k=2^n}^{2^{n+1}-1}\frac{S_{k,n}}{k}\mathrm{e}^{it\log k}\Biggr|\Biggr\|_{L^2([0,1],dt)} \le C\Biggl(\sum_{n\ge 1}\sum_{k=2^n}^{2^{n+1}-1} k\frac{|S_{k,n}|^2}{k^2}\Biggr)^{1/2}$$
$$\le \Biggl(\sum_{n\ge 1}\Biggl(\sum_{k=2^n}^{2^{n+1}-1}|a_k|\Biggr)^2\Biggr)^{1/2} < \infty.$$

Let $n \ge 1$ and $2^n \le \ell \le 2^{n+1}-1$. We have

$$\Biggl|\sum_{k=1}^{\ell} a_n k^{it} - \sum_{k=1}^{2^n-1} a_n k^{it}\Biggr| \le \sum_{k=2^n}^{2^{n+1}-1}|a_k|.$$

Hence,

$$\sup_{N\ge 1}\Biggl|\sum_{n=1}^{N} a_n \mathrm{e}^{it\log n}\Biggr| \le \sup_{N\ge 1}\Biggl|\sum_{n=1}^{2^N-1} a_n \mathrm{e}^{it\log n}\Biggr| + \Biggl(\sum_{n\ge 1}\Biggl(\sum_{k=2^n}^{2^{n+1}-1}|a_k|\Biggr)^2\Biggr)^{1/2}.$$

So, (2-3) is proved. The λ-almost everywhere convergence may be proved by a standard procedure thanks to the maximal inequality. Alternatively, following all the steps of the proof of the maximal inequality lets us give a more direct proof. □

As a corollary we deduce:

Theorem 2.4. *Let $(\lambda_n)_{n\ge 1}$ be an increasing sequence of positive real numbers tending to ∞. Let $(a_n)_{n\ge 1}$ be such that*

(2-4) $$\sum_{n\ge 1}\Biggl(\sum_{k:n\le\lambda_k\le n+1}|a_k|\Biggr)^2 < \infty.$$

There exists a universal constant $C>0$ *such that*

$$\Big\|\sup_{N\ge 1}\Big|\sum_{n=1}^N a_n\mathrm{e}^{i\lambda_n t}\Big|\Big\|_{\mathcal{S}^2}\le C\Big(\sum_{n\ge 1}\Big(\sum_{k:n\le\lambda_k\le n+1}|a_k|\Big)^2\Big)^{1/2}. \tag{2-5}$$

Moreover, the series $\sum_{n\ge 1}a_n\mathrm{e}^{it\lambda_n}$ *converges for* λ*-a.e.* $t\in\mathbb{R}$.

Proof. Write $u_n:=[2^{\lambda_n}]$. Hence $(u_n)_{n\ge 1}$ is a nondecreasing sequence of integers. That sequence may overlap from time to time. So let $(v_k)_{k\ge 1}$ be a strictly increasing sequence of integers with same range as $(u_n)_{n\ge 1}$.

Define a sequence $(b_n)_{n\ge 1}$ as follows. Let $n\ge 1$ be such that there exists $k\ge 1$ such that $n=v_k$. Then set $b_n:=\sum_{\ell:u_\ell=v_k}a_\ell$. If there is no $k\ge 1$ such that $n=v_k$, set $b_n:=0$.

We first control

$$\sup_{N\ge 1}\Big|\sum_{n=1}^N b_n\mathrm{e}^{it\log_2 n}\Big|,$$

where $\log_2$ stands for the logarithm in base 2.

By Theorem 2.2, we have

$$\Big\|\sup_{N\ge 1}\Big|\sum_{n=1}^N b_n\mathrm{e}^{i\log_2 n\cdot}\Big|\Big\|_{\mathcal{S}^2}^2\le C\sum_{n\ge 0}\Big(\sum_{k=2^n}^{2^{n+1}-1}|b_k|\Big)^2=\sum_{n\ge 0}\Big(\sum_{\ell:2^n\le u_\ell\le 2^{n+1}-1}|b_\ell|\Big)^2.$$

Now, if $2^n\le u_\ell\le 2^{n+1}-1$, then $n\le\lambda_\ell\le n+1$ and our first step is proved.

Let $q\ge p$ be integers. There exist integers $q'\ge p'$ such that $v_{p'}=u_p$ and $v_{q'}=u_q$. We have

$$\Big|\sum_{k=p}^q a_k\mathrm{e}^{it\lambda_k}-\sum_{k=v_{p'}}^{v_{q'}}b_k\mathrm{e}^{it\log_2 u_k}\Big|\le\sum_{k:u_k=u_p}|a_k|+\sum_{k:u_k=u_q}|a_k|+\sum_{\ell=p'}^{q'}\sum_{k:u_k=v_\ell}|a_k||\mathrm{e}^{it\lambda_k}-\mathrm{e}^{it\log_2 u_k}|.$$

Clearly, it suffices to control

$$\sum_{n\ge 0}\sum_{\ell:2^n\le v_\ell\le 2^{n+1}-1}\sum_{k:u_k=v_\ell}|a_k||\mathrm{e}^{it\lambda_k}-\mathrm{e}^{it\log_2 u_k}|.$$

Now, for $2^n\le v_\ell\le 2^{n+1}-1$ and $u_k=v_\ell$, using that $u_k\le 2^{\lambda_k}\le u_k+1$, we see that $|\log_2(2^{\lambda_k})-\log_2 u_k|\le C/u_k$ and that

$$|\mathrm{e}^{it\lambda_k}-\mathrm{e}^{it\log_2 u_k}|=|\mathrm{e}^{it\log_2(2^{\lambda_k})}-\mathrm{e}^{it\log_2 u_k}|\le\frac{C|t|}{u_k}\le\frac{C|t|}{2^n}.$$

Hence, using Cauchy–Schwarz,

$$\sum_{n\geq 0}\sum_{\ell:2^n\leq v_\ell\leq 2^{n+1}-1}\sum_{k:u_k=v_\ell}|a_k||\mathrm{e}^{it\lambda_k}-\mathrm{e}^{it\log u_k}|\leq Ct\sum_{n\geq 0}2^{-n}\sum_{k:2^n\leq u_k\leq 2^{n+1}-1}|a_k|$$
$$\leq Ct\biggl(\sum_{n\geq 0}\biggl(\sum_{k:2^n\leq u_k\leq 2^{n+1}-1}|a_k|\biggr)^2\biggr)^{1/2},$$

which converges by our assumption. □

We shall now derive an almost everywhere convergence result concerning the Fourier series of an almost periodic function in $\mathcal{S}^2$. We shall first recall known results about norm convergence.

Let $(\lambda_n)_{n\geq 1}$ be a (not necessarily increasing) sequence of positive real numbers. As already mentioned (in the case of Dirichlet series), by [Wiener 1926], see also [Tornehave 1954], if

$$\sum_{n\geq 0}\biggl(\sum_{k\geq 1:n\leq\lambda_k<n+1}|a_k|\biggr)^2<\infty, \tag{2-6}$$

then $\sum_{n\geq 1}a_n\mathrm{e}^{i\lambda_n t}$ is the Fourier series of an element of $f\in\mathcal{S}^2$.

On the other hand, if $f\in\mathcal{S}^2$ admits a sequence of positive real numbers $(\lambda_n)_{n\geq 1}$ as frequencies and such that $\hat{f}(\lambda_n)\geq 0$ for every $n\geq 1$, then, see [Tornehave 1954],

$$\sum_{n\geq 0}\biggl(\sum_{k\geq 1:n\leq\lambda_k<n+1}|\hat{f}(\lambda_k)|\biggr)^2\leq C\|f\|^2_{\mathcal{S}^2}.$$

Hence, (2-6) holds.

Condition (2-6) is thus optimal for deciding whether $\sum_{n\geq 1}a_n\mathrm{e}^{i\lambda_n t}$ is the Fourier series of an element of $\mathcal{S}^2$ or not. One can not however expect that it is always necessary, so we provide a counterexample in Proposition 2.7 below.

Let $f\in\mathcal{S}^2$ be such that $\Lambda\subset[0,+\infty)$ (that restriction may be obviously removed). Assume that Λ is α-separated for some $\alpha>0$ and write $\Lambda:=\{\lambda_1<\lambda_2\cdots\}$. Then,

$$\frac{\alpha}{C}\sum_{n\geq 0}\biggl(\sum_{k\geq 1:n\leq\lambda_k<n+1}|\hat{f}(\lambda_k)|\biggr)^2\leq\sum_{n\geq 1}|\hat{f}(\lambda_n)|^2\leq\|f\|^2_{\mathcal{S}^2}$$
$$\leq C\sum_{n\geq 0}\biggl(\sum_{k\geq 1:n\leq\lambda_k<n+1}|\hat{f}(\lambda_k)|\biggr)^2.$$

In particular, we have the following direct consequence of Theorem 2.2:

Corollary 2.5. *Let $f \in \mathcal{S}^2$ be such that $\Lambda \subset [0,+\infty)$. Assume that Λ is α-separated for some $\alpha > 0$. There exists $C > 0$, independent of f and α such that*

$$\Big\|\sup_{N\ge 1}\Big|\sum_{n=1}^{N}\hat f(\lambda_n)\mathrm{e}^{i\lambda_n\cdot}\Big|\Big\|_{\mathcal{S}^2} \le C\frac{\|f\|_{\mathcal{S}^2}}{\alpha}.$$

Moreover, the series $\sum_{n\ge 1}\hat f(\lambda_n)\mathrm{e}^{i\lambda_n\cdot}$ converges for λ-almost every $t \in \mathbb{R}$.

We now give an example of Fourier series converging in $\mathcal{S}^2$ while (2-6) does not hold. Let us first recall the following result of Halász; see [Queffélec 1984].

Lemma 2.6. *There exists $C > 0$ such that for every sequence of iid Rademacher variables $(\varepsilon_n)_{n\ge1}$*

$$\mathbb{E}\Big(\sup_{t\in\mathbb{R}}\Big|\sum_{k=1}^{n}\varepsilon_k k^{it}\Big|\Big) \le C\frac{n}{\log(n+1)}. \tag{2-7}$$

Proposition 2.7. *Let $(\varepsilon_n)_{n\ge1}$ be iid Rademacher variables on $(\Omega,\mathcal{F},\mathbb{P})$. For $\mathbb{P}$-almost all $\omega\in\Omega$, $\sum_{n\ge1}\varepsilon_n(\omega)n^{it}/n\sqrt{\log(n+1)}$ converges in $\mathcal{S}^2$, while (2-4) is not satisfied (with $a_n = \varepsilon_n(\omega)/n\sqrt{\log(n+1)}$).*

Proof. For every $n \ge 1$, every $2^n \le k \le 2^{n+1}$ and every $\omega\in\Omega$, we have

$$\Big\|\sum_{\ell=2^n}^{k}\frac{\varepsilon_\ell(\omega)\ell^{it}}{\ell\sqrt{\log(\ell+1)}}\Big\|_{\mathcal{S}^2} \le \sum_{\ell=2^n}^{k}\frac{1}{\ell\sqrt{\log(\ell+1)}} \le \frac{2}{\sqrt{n}} \underset{n\to+\infty}{\longrightarrow} 0.$$

Hence, it suffices to prove that for $\mathbb{P}$-almost every $\omega\in\Omega$,

$$\Big(\sum_{n=1}^{2^N}\frac{\varepsilon_n(\omega)n^{it}}{n\sqrt{\log(n+1)}}\Big)_{N\ge1} \quad \text{converges in } \mathcal{S}^2.$$

Let $S_n(t) := \sum_{k=1}^{n}\varepsilon_k k^{it}$ ($S_0(t)=0$) and $u_n := \big(n\sqrt{\log(n+1)}\big)^{-1}$. We have

$$\sum_{n=1}^{2^N}\frac{\varepsilon_n(\omega)n^{it}}{n\sqrt{\log(n+1)}} = \sum_{n=1}^{2^N}(S_n(t)-S_{n-1}(t))u_n = \sum_{n=1}^{2^N}S_n(t)(u_n-u_{n+1}) + S_{2^N}(t)u_{2^N+1}.$$

It follows from (2-7) that

$$\mathbb{E}\Big(\sum_{n\ge1}\sup_{t\in\mathbb{R}}|S_n(t)(u_n-u_{n+1})|\Big) < \infty, \quad \mathbb{E}\Big(\sum_{n\ge1}\sup_{t\in\mathbb{R}}|S_{2^N}(t)u_{2^N+1}|\Big) < \infty,$$

and the result follows. ∎

3. Convergence almost everywhere of associated series of dilates

Theorem 3.1. *Let $(\lambda_n)_{n\geq 1}$ and $(\mu_n)_{n\geq 1}$ be nondecreasing sequences of real numbers greater than* 1. *Let $(\alpha_n)_{n\geq 1}$ be a sequence of complex numbers such that*

$$\sum_{n\geq 1}\Big(\sum_{k:n\leq\lambda_k<n+1}|\alpha_k|\Big)^2<\infty.$$

Let $(\beta_n)_{n\geq 1}\in\ell^1$. Then $D(t):=\sum_{n\geq 1}\beta_n\mathrm{e}^{i\mu_n t}$ defines a continuous function on $\mathbb{R}$ (and in $\mathcal{S}^2$) and there exists a universal constant $C>0$ such that

$$\Big\|\sup_{N\geq 1}\Big|\sum_{n=1}^{N}\alpha_n D(\lambda_n t)\Big|\Big\|_{\mathcal{S}^2}\leq C\Big(\sum_{n\geq 1}|\beta_n|\Big)\Big(\sum_{n\geq 1}\Big(\sum_{k:n\leq\lambda_k<n+1}|\alpha_k|\Big)^2\Big)^{1/2}. \tag{3-1}$$

Moreover, the series $\sum_{n\geq 1}\alpha_n D(\lambda_n t)$ converges in $\mathcal{S}^2$ and for λ-a.e. $t\in\mathbb{R}$.

Proof. Let $x\in\mathbb{R}$. The fact that D is a continuous function in $\mathcal{S}^2$ follows easily from the fact that $(\beta_n)_{n\geq 1}\in\ell^1$. We also have, for every $N\geq 1$,

$$\Big|\sum_{n=1}^{N}\alpha_n D(\lambda_n t)\Big|\leq\sum_{k\geq 1}|\beta_k|\Big|\sum_{n=1}^{N}\alpha_n\mathrm{e}^{it\lambda_n\mu_k}\Big|.$$

By Theorem 2.4, we have

$$\begin{aligned}\int_x^{x+1}\sup_{N\geq 1}\Big|\sum_{n=1}^{N}\alpha_n\mathrm{e}^{it\lambda_n\mu_k}\Big|^2dt&=\frac{1}{\mu_k}\int_{\mu_k x}^{\mu_k(x+1)}\sup_{N\geq 1}\Big|\sum_{n=1}^{N}\alpha_n\mathrm{e}^{it\lambda_n}\Big|^2dt\\&\leq\frac{[\mu_k]+1}{\mu_k}\Big\|\sup_{N\geq 1}\Big|\sum_{n=1}^{N}\alpha_n\mathrm{e}^{it\lambda_n}\Big|\Big\|_{\mathcal{S}^2}^2,\end{aligned}$$

and (3-1) follows.

The convergence almost everywhere and in $\mathcal{S}^2$ follows by standard arguments. □

We also have the following obvious corollary of Theorem 2.4, whose proof is left to the reader:

Proposition 3.2. *Let $(\lambda_n)_{n\geq 1}$ and $(\mu_n)_{n\geq 1}$ be nondecreasing sequences of real numbers greater than* 1. *Let $(\beta_n)_{n\geq 1}$ be a sequence of complex numbers such that*

$$\sum_{n\geq 1}\Big(\sum_{k:n\leq\mu_k<n+1}|\beta_k|\Big)^2<\infty.$$

Let $(\alpha_n)_{n\geq 1}\in \ell^1$. Then, $D(t):=\sum_{n\geq 1}\beta_n \mathrm{e}^{i\mu_n t}$ converges in $\mathcal{S}^2$ and there exists a universal constant $C>0$ such that

$$\Big\|\sup_{N\geq 1}\Big|\sum_{n=1}^{N}\alpha_n D(\lambda_n t)\Big|\Big\|_{\mathcal{S}^2}\leq C\Big(\sum_{n\geq 1}|\alpha_n|\Big)\Big(\sum_{n\geq 1}\Big(\sum_{k:n\leq \mu_k<n+1}|\beta_k|\Big)^2\Big)^{1/2}. \tag{3-2}$$

Moreover, the series $\sum_{n\geq 1}\alpha_n D(\lambda_n t)$ converges in $\mathcal{S}^2$ and for λ-a.e. $t\in\mathbb{R}$.

Theorem 3.3. *Let $(\lambda_n)_{n\geq 1}$ and $(\mu_n)_{n\geq 1}$ be nondecreasing sequences of real numbers greater than 1. Let $1\leq p,q\leq 2$ satisfy $1/p+1/q=\frac{3}{2}$. There exists $C>0$ such that for any sequences $(\alpha_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ of complex numbers such that*

$$\sum_{n\geq 1}\Big(\sum_{k:n\leq \lambda_k<n+1}|\alpha_k|\Big)^p<\infty \quad \textit{and} \quad \sum_{n\geq 1}\Big(\sum_{k:n\leq \mu_k<n+1}|\beta_k|\Big)^q<\infty, \tag{3-3}$$

we have

$$\Big\|\sup_{N\geq 1}\Big|\sum_{n=1}^{N}\alpha_n D(\lambda_n t)\Big|\Big\|_{\mathcal{S}^2} \leq C\Big(\sum_{n\geq 1}\Big(\sum_{k:n\leq \lambda_k<n+1}|\alpha_k|\Big)^p\Big)^{1/p}\Big(\sum_{n\geq 1}\Big(\sum_{k:n\leq \mu_k<n+1}|\beta_k|\Big)^q\Big)^{1/q}, \tag{3-4}$$

where $D(t):=\sum_{n\geq 1}\beta_n \mathrm{e}^{i\mu_n t}$ is defined in $\mathcal{S}^2$. Moreover, the series $\sum_{n\geq 1}\alpha_n D(\lambda_n t)$ converges in $\mathcal{S}^2$ and for λ-a.e. $t\in\mathbb{R}$.

Before doing the proof let us mention the following immediate corollaries. We first apply Theorem 3.3 with the choice $\mu_n=\log n,\ n\geq 1$ and $\lambda_k=k,\ k\geq 1$.

Corollary 3.4. *Assume that*

$$\sum_{k\geq 1}|\alpha_k|^p<\infty \quad \textit{and} \quad \sum_{n\geq 1}\Big(\sum_{k:2^n\leq k<2^{n+1}}|\beta_k|\Big)^q<\infty,$$

for some $1\leq p,q\leq 2$ such that $1/p+1/q=\frac{3}{2}$. Let $D(t):=\sum_{n\geq 1}\beta_n n^{it}$. Then the series $\sum_{k\geq 1}\alpha_k D(kt)$ converges in $\mathcal{S}^2$ and for λ-a.e. $t\in\mathbb{R}$.

Example 3.5. Let $\frac{1}{2}<\alpha\leq 1$. Choose $1/\alpha<p\leq 2$ and $q=2p/(3p-2)$ $(1\leq q<2)$. Let $D(t)=\sum_{n\geq 1}\beta_n n^{it}$ and assume that

$$\sum_{n\geq 1}\Big(\sum_{k:2^n\leq k<2^{n+1}}|\beta_k|\Big)^q<\infty. \tag{3-5}$$

Then the series

$$\sum_{k\geq 1}\frac{D(kt)}{k^\alpha} \tag{3-6}$$

converges almost everywhere. This extends to Dirichlet series Hartman and Wintner's result [1938] showing that the series $\Phi_\alpha(x) = \sum_{k=1}^{\infty} \psi(kx)/k^\alpha$ converges almost everywhere. Here $\psi(x) = x - [x] - \frac{1}{2} = \sum_{j=1}^{\infty} \sin 2\pi jx/j$, and $[x]$ is the integer part of x. That result is also a special case of (3-6): take $\beta_n = 1/j$ if $n = 2^j$, $j \geq 1$ and $\beta_n = 0$ elsewhere.

Remark 3.6. To our knowledge [Hartman and Wintner 1938] contains, among other results on Φ_α, the first convergence result for the series of dilates $\sum_{k=1}^{\infty} \alpha_k \psi(kx)$.

Then, we apply Theorem 3.3 with the choice $\mu_n = n,\ n \geq 1$ and $\lambda_k = k,\ k \geq 1$.

Corollary 3.7. *Assume that*

$$\sum_{k\geq 1} |\alpha_k|^p < \infty \quad \textit{and} \quad \sum_{j\geq 1} |b_j|^q < \infty$$

for some $1 \leq p, q \leq 2$ *such that* $1/p + 1/q = \frac{3}{2}$. *Let* $D(t) = \sum_{\ell\geq 1} b_\ell e^{i\ell t}$. *Then the series* $\sum_{k\geq 1} \alpha_k D(kt)$ *converges in* $\mathcal{S}^2$ *and for* λ*-a.e.* $t \in \mathbb{R}$.

Remark 3.8. Suppose that $b_j = \mathcal{O}(1/j^\alpha)$ for some $\frac{1}{2} < \alpha \leq 1$. Assume that

$$\sum_{k\geq 1} |\alpha_k|^p < \infty,$$

for some $1 \leq p < 2/(3-2\alpha)$. Then $\sum_{j\geq 1} |b_j|^q < \infty$ for q such that $1/p + 1/q = \frac{3}{2}$ and we have $1 \leq p, q \leq 2$. We deduce from Corollary 3.7 that the series $\sum_{k\geq 1} \alpha_k D(kt)$ converges in $\mathcal{S}^2$ and for λ-a.e. $t \in \mathbb{R}$. When $\frac{1}{2} < \alpha < 1$, the nearly optimal sufficient condition $\sum_{k\geq 1} |c_k|^2 \exp\{K(\log k)^{1-\alpha}/(\log\log k)^\alpha\} < \infty$ in which $K = K(\alpha)$ has been recently established in [Aistleitner et al. 2015, Theorem 2]. See also [Weber 2016, Theorem 3.1] for conditions of individual type, i.e., depending on the support of the coefficient sequence. When $\alpha = 1$, the optimal sufficient coefficient condition, namely that $\sum_{k=1}^{\infty} |\alpha_k|^2 (\log\log k)^{2+\varepsilon}$ converges for some $\varepsilon > 0$ suffices for the convergence almost everywhere, has been recently obtained by Lewko and Radziwiłł [2017, Corollary 3].

These results are clearly better. However, we note that our results are, even in the trigonometrical case, independent from these ones, and concern a larger class of trigonometrical series $D(t)$.

Proof of Theorem 3.3. Clearly, we only need to prove (3-4). Let $(\alpha_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$ be in $\ell^1(\mathbb{N})$, fixed for all the proof. Let $D(t) := \sum_{n\geq 1} \beta_n e^{i\mu_n t}$. It is enough to prove that for every $N \geq 1$,

$$\left\| \sup_{m=1}^{N} \left| \sum_{n=1}^{m} \alpha_n D(\lambda_n t) \right| \right\|_{\mathcal{S}^2} \leq C \Biggl(\sum_{n\geq 1} \Bigl(\sum_{k: n\leq \lambda_k < n+1} |\alpha_k| \Bigr)^p \Biggr)^{1/p} \Biggl(\sum_{n\geq 1} \Bigl(\sum_{k: n\leq \mu_k < n+1} |\beta_k| \Bigr)^q \Biggr)^{1/q},$$

for a constant $C > 0$ not depending on N, $(\alpha_n)_{n\geq 1}$ and $(\beta_n)_{n\geq 1}$.

We do that by interpolating (3-1) and (3-2). Define Banach spaces as follows:

$$X_1 := \Big\{ (a_n)_{n\geq 1} \in \mathbb{C}^{\mathbb{N}} : \|(a_n)_{n\geq 1}\|_{X_1} := \sum_{n\geq 1} \sum_{k: n\leq \lambda_k < n+1} |a_k| < \infty \Big\},$$

$$X_2 := \Big\{ (a_n)_{n\geq 1} \in \mathbb{C}^{\mathbb{N}} : \|(a_n)_{n\geq 1}\|_{X_2} := \Big(\sum_{n\geq 1} \Big(\sum_{k: n\leq \lambda_k < n+1} |a_k| \Big)^2 \Big)^{1/2} < \infty \Big\},$$

$$Y_1 := \Big\{ (b_n)_{n\geq 1} \in \mathbb{C}^{\mathbb{N}} : \|(b_n)_{n\geq 1}\|_{Y_1} := \sum_{n\geq 1} \sum_{k: n\leq \mu_k < n+1} |b_k| < \infty \Big\},$$

$$Y_2 := \Big\{ (b_n)_{n\geq 1} \in \mathbb{C}^{\mathbb{N}} : \|(b_n)_{n\geq 1}\|_{Y_1} := \Big(\sum_{n\geq 1} \Big(\sum_{k: n\leq \mu_k < n+1} |b_k| \Big)^2 \Big)^{1/2} < \infty \Big\}.$$

For every $t \in \mathbb{R}$, let

$$J(t) := \min \Big\{ j \in \mathbb{N} : 1 \leq j \leq N, \ \Big| \sum_{n=1}^{j} \alpha_n D(\lambda_n t) \Big| = \sup_{m=1}^{N} \Big| \sum_{n=1}^{m} \alpha_n D(\lambda_n t) \Big| \Big\}.$$

Define a linear operator T on $(X_1 + X_2) \times (Y_2 + Y_1)$ by setting

$$T((a_n)_{n\geq 1}, (b_n)_{n\geq 1}) := \sum_{k=1}^{N} \mathbf{1}_{\{k \leq J(t)\}} a_k \Big(\sum_{\ell \geq 1} b_\ell e^{i\lambda_k \mu_\ell t} \Big).$$

By Propositions 3.1 and 3.2, T is continuous from $X_1 \times Y_2$ to $\mathcal{S}^2$ and from $X_2 \times Y_1$ to $\mathcal{S}^2$.

It follows from paragraph 10.1 of [Calderón 1964] that for every $s \in [0, 1]$ there exists C_s such that, with their notation,

$$\|T((a_n)_{n\geq 1}, (b_n)_{n\geq 1})\|_{\mathcal{S}^2} \leq C_s \|(a_n)_{n\geq 1}\|_{[X_1, X_2]_s} \|(b_n)_{n\geq 1}\|_{[Y_2, Y_1]_s},$$

where

$$\|(a_n)_{n\geq 1}\|_{[X_1, X_2]_s} = \inf\{ \|f\|_{\mathcal{F}} : f \in \mathcal{F}, \ f(s) = (a_n)_{n\geq 1} \},$$

and $\mathcal{F}$ is the Banach space of continuous functions f from $\{z \in \mathbb{C} : 0 \leq \operatorname{Re} z \leq 1\}$ to $X_1 + X_2$, analytic on $\{z \in \mathbb{C} : 0 < \operatorname{Re} z < 1\}$ such that for every $t \in \mathbb{R}$, $f(it) \in X_1$ and $f(1+it) \in X_2$ with $\lim_{|t|\to+\infty} f(it) = \lim_{|t|\to+\infty} f(1+it) = 0$, endowed with the norm

$$\|f\|_{\mathcal{F}} := \max \Big(\sup_{t\in\mathbb{R}} \|f(it)\|_{X_1}, \ \sup_{t\in\mathbb{R}} \|f(1+it)\|_{X_2} \Big).$$

The norm $\|(b_n)_{n\geq 1}\|_{[Y_2, Y_1]_s}$ is defined similarly.

We shall now give an upper bound for $\|(a_n)_{n\geq 1}\|_{[X_1,X_2]_s}$. By homogeneity, we may assume that

$$\sum_{n\geq 1}\Bigl(\sum_{n\leq\lambda_k<n+1}|a_k|\Bigr)^{2/(2-s)}=1.$$

Let $\varepsilon>0$. Define an element f_ε of $\mathcal{F}$ by setting for every $z\in\mathbb{C}$ such that $0\leq \operatorname{Re} z\leq 1$, $f_\varepsilon(z)=(c_n(z))_{n\geq 1}$ where, for every $n,k\geq 1$ such that $n\leq\lambda_k<n+1$,

$$c_k(z)=\mathrm{e}^{\varepsilon(z^2-s^2)}a_k\Bigl(\sum_{n\leq\lambda_\ell<n+1}|a_\ell|\Bigr)^{(2-z)/(2-s)-1},$$

if $\sum_{n\leq\lambda_\ell<n+1}|a_\ell|\neq 0$ and $c_k(z)=0$ otherwise.

The introduction of ε here is a standard trick to ensure the assumptions

$$\lim_{|t|\to+\infty} f_\varepsilon(it)=\lim_{|t|\to+\infty} f_\varepsilon(1+it)=0.$$

Notice that $f_\varepsilon(s)=(a_n)_{n\geq 1}$. For every $t\in\mathbb{R}$,

$$\|f_\varepsilon(it)\|_{X_1}\leq\sum_{n\geq 1}\Bigl(\sum_{n\leq\lambda_k<n+1}|a_k|\Bigr)^{\frac{2}{(2-s)}}=1.$$

Similarly, for every $t\in\mathbb{R}$,

$$\|f_\varepsilon(1+it)\|_{X_2}\leq \mathrm{e}^{\varepsilon}\sum_{n\geq 1}\Bigl(\sum_{n\leq\lambda_k<n+1}|a_k|\Bigr)^{\frac{2}{(2-s)}}=\mathrm{e}^{\varepsilon}.$$

Letting $\varepsilon\to 0$, we infer that

$$\|(a_n)_{n\geq 1}\|_{[X_1,X_2]_s}\leq 1=\Biggl(\sum_{n\geq 1}\Bigl(\sum_{n\leq\lambda_k<n+1}|a_k|\Bigr)^{\frac{2}{(2-s)}}\Biggr)^{\frac{2-s}{2}}.$$

Similarly, one can prove that

$$\|(b_n)_{n\geq 1}\|_{[X_1,X_2]_s}\leq\Biggl(\sum_{n\geq 1}\Bigl(\sum_{n\leq\lambda_k<n+1}|b_k|\Bigr)^{\frac{2}{(1+s)}}\Biggr)^{\frac{1+s}{2}}.$$

Taking $s=2(1-1/p)$ yields the desired result. $\square$

4. A necessary condition for convergence almost everywhere

Hartman [1942] has proved the following result:

Theorem 4.1. *Assume that*

$$\frac{\lambda_k}{\lambda_{k-1}}\geq q>1,\quad k\geq 1. \tag{4-1}$$

Assume that the series $\sum_{k=1}^{\infty}a_k\mathrm{e}^{i\lambda_k t}$ converges for almost all real t. Then the series $\sum_{k=1}^{\infty}|a_k|^2$ converges.

The proof is similar to Zygmund's [1968, Proof of Lemma 6.5, Chapter V] (see also p. 120–122 of the 1935 edition).

Remark 4.2. The converse of Theorem 4.1 is due to Kac [1941]. If $\sum_{k=1}^{\infty}|a_k|^2$ converges, then the series $\sum_{k=1}^{\infty}a_k e^{i\lambda_k t}$ with $(\lambda_k)_{k\geq 1}$ verifying (4-1), converges for almost all real t. Kac's proof is a modification of Marcinkiewicz's. See Remark 1.1. In place of Fejér's theorem, another summation method is used. See Theorem 13 and pages 84–85 in [Titchmarsh 1948], and Theorem 21 in [Hardy and Riesz 1915].

Theorem 4.1 can be extended in the following way:

Theorem 4.3. *Let $\{\lambda_k, k \geq 1\}$ be a increasing sequence of positive reals satisfying the condition*

$$M := \sum_{\substack{k\neq\ell,\, k'\neq\ell' \\ (k,\ell)\neq(k',\ell')}} (1-|(\lambda_k-\lambda_\ell)-(\lambda_{k'}-\lambda_{\ell'})|)_+^2 < \infty. \tag{4-2}$$

Assume that

$$\lambda\Big\{\sum_k a_k e^{i\lambda_k t} \text{ converges}\Big\} > 0. \tag{4-3}$$

Then the series $\sum_{k=1}^{\infty}|a_k|^2$ converges.

Remark 4.4. By considering integers k such that $n \leq \lambda_k < n+\frac{1}{2}$, and next those such that $n+\frac{1}{2} \leq \lambda_k \leq n+1$, we observe that condition (4-2) implies that

$$\sup_n \#\{k : n \leq \lambda_k < n+1\} < \infty.$$

We give an application. Recall that a Sidon sequence is a set of integers with the property that the pairwise sums of elements are all distinct. As a corollary we get

Corollary 4.5. *Let $\{\lambda_k, k \geq 1\}$ be a Sidon sequence. Assume that (4-3) is satisfied. Then the series $\sum_{k=1}^{\infty}|a_k|^2$ converges.*

Remark 4.6. In contrast with Hadamard gap sequences, Sidon sequences may grow at most polynomially. See [Ruzsa 2001] where it is for instance proved that the sequence $\{n^5+[\xi n^4], n \geq n_0\}$ is for some real number $\xi \in [0, 1]$ and n_0 large, a Sidon sequence.

Proof of Corollary 4.5. Let $(k, \ell) \neq (k', \ell')$ with $k \neq \ell$ and $k' \neq \ell'$. As the equation $\lambda_k-\lambda_\ell = \lambda_{k'}-\lambda_{\ell'}$ means $\lambda_k+\lambda_{\ell'} = \lambda_\ell+\lambda_{k'}$, the fact that $\{\lambda_k, k \geq 1\}$ is a Sidon sequence implies that the only possible solutions are $k = k'$, $\ell' = \ell$ or $k = \ell$, $\ell' = k'$. The last one is impossible by assumption, and the first would mean that $(k, \ell) = (k', \ell')$ which is excluded. Consequently, $\lambda_k-\lambda_\ell \neq \lambda_{k'}-\lambda_{\ell'}$. Hence the sum in (4-2) is always zero. $\square$

Remark 4.7. It follows from Hartman's proof that under condition (4-1), the sequence of differences $\lambda_k - \lambda_\ell$, $k \neq \ell$ is a finite union of subsequences such that the difference of any two numbers of the same subsequence exceeds 1. These subsequences fulfill assumption (4-2) of Theorem 4.3, and thus Theorem 4.1 follows from Theorem 4.3.

Theorem 4.3 is a consequence of the following general necessary condition for almost everywhere convergence of series of functions.

Theorem 4.8. *Let $(X, \mathcal{B}, \tau)$ be a probability space. Let $\{g_k, k \geq 1\} \subset L^4(\tau)$ be a sequence of functions with $\|g_k\|_{2,\tau} = 1$, $\|g_k\|_{4,\tau} \leq K$ and satisfying the condition*

$$M := \sum_{\substack{k \neq \ell,\, k' \neq \ell' \\ (k,\ell) \neq (k',\ell')}} |\langle g_k \overline{g_\ell}, g_{k'} \overline{g_{\ell'}} \rangle_\tau|^2 < \infty. \tag{4-4}$$

Assume that

$$\tau\Big\{ \sum_k a_k g_k(t) \text{ converges} \Big\} > 0. \tag{4-5}$$

Then the series $\sum_{k=1}^{\infty} |a_k|^2$ converges.

Proof of Theorem 4.8. We use Hartman's method and the below classical generalization of Bessel's inequality.

Lemma 4.9 (Bellman–Boas inequality). *Let $x, y_1, \ldots, y_n$ be elements of an inner product space $(H, \langle \cdot, \cdot \rangle)$. Then*

$$\sum_{i=1}^{n} |\langle x, y_i \rangle|^2 \leq \|x\|^2 \Big\{ \max_{1 \leq i \leq n} \|y_i\|^2 + \Big(\sum_{1 \leq i \neq j \leq n} |\langle y_i, y_j \rangle|^2 \Big)^{1/2} \Big\}.$$

See [Bellman 1944] for instance. As

$$\Big\{ t : \sum_k a_k g_k(t) \text{ converges} \Big\} = \bigcap_{\varepsilon > 0} \bigcup_{V} \bigcap_{u > v > V} \Big\{ t : \Big| \sum_{k=v}^{u} a_k g_k(t) \Big| \leq \varepsilon \Big\},$$

by assumption it follows that for any $\varepsilon > 0$, there exists an integer V such that if

$$A := \bigcap_{u > v > V} \Big\{ \Big| \sum_{k=v}^{u} a_k g_k(t) \Big| \leq \varepsilon \Big\},$$

then

$$\tau(A) > 0. \tag{4-6}$$

Assume the series $\sum_{k \geq 1} |a_k|^2$ is divergent. We are going to prove that this will contradict (4-6).

By squaring out,

$$\int_A \Big|\sum_{k=n}^m a_k g_k(t)\Big|^2 \tau(dt) = \tau(A)\sum_{k=n}^m |a_k|^2 + \sum_{\substack{k,\ell=n\\k\neq\ell}}^m a_k\bar a_\ell \int_A g_k(t)\overline{g_\ell}(t)\tau(dt). \tag{4-7}$$

By using the Cauchy–Schwarz inequality,

$$\Big|\sum_{\substack{k,\ell=n\\k\neq\ell}}^m a_k\bar a_\ell \int_A g_k(t)\overline{g_\ell}(t)\tau(dt)\Big| \le \Big(\sum_{\substack{k,\ell=n\\k\neq\ell}}^m |a_k|^2|a_\ell|^2\Big)^{1/2}\Big(\sum_{\substack{k,\ell=n\\k\neq\ell}}^m \Big|\int_A g_k(t)\overline{g_\ell}(t)\tau(dt)\Big|^2\Big)^{1/2}.$$

Applying Lemma 4.9 to the system of vectors of $L^2_\tau(\mathbb{R})$, $\chi(A)$, $g_k(t)\overline{g_\ell}(t)$, $n \le k, \ell \le m$ gives, in view of the assumption made,

$$\begin{aligned}\sum_{\substack{k,\ell=n\\k\neq\ell}}^m \Big|\int_A g_k(t)\overline{g_\ell}(t)\tau(dt)\Big|^2 &\le \tau(A)^2\Big\{K^2 + \Big(\sum_{\substack{(k,\ell)\neq(k',\ell')\\ n\le k\neq\ell\le m\\ n\le k'\neq\ell'\le m}} |\langle g_k\overline{g_\ell}, g_{k'}\overline{g_{\ell'}}\rangle_\tau|^2\Big)^{1/2}\Big\}\\ &\le \tau(A)^2\{K^2+M^{1/2}\}.\end{aligned}$$

Letting n, m tend to infinity, it follows that the series $\sum_{k\neq\ell}\big|\int_A g_k(t)\overline{g_\ell}(t)\tau(dt)\big|^2$ converges. Consequently, for all $m > n$, $n > N$, N depending on A

$$\sum_{\substack{k,\ell=n\\k\neq\ell}}^m \Big|\int_A g_k(t)\overline{g_\ell}(t)\tau(dt)\Big|^2 \le \tau(A)^2/4.$$

There is no loss in assuming $N > V$, which we do. Therefore

$$\Big|\sum_{\substack{k,\ell=n\\k\neq\ell}}^m a_k\bar a_\ell \int_A g_k(t)\overline{g_\ell}(t)\tau(dt)\Big| \le \Big(\sum_{\substack{k,\ell=n\\k\neq\ell}}^m |a_k|^2|a_\ell|^2\Big)^{1/2}\Big(\frac{\tau(A)}{2}\Big).$$

This along with (4-7) implies

$$\int_A \Big|\sum_{k=n}^m a_k g_k(t)\Big|^2 \tau(dt) \ge \Big(\frac{\tau(A)}{2}\Big)\sum_{k=n}^m |a_k|^2, \tag{4-8}$$

for all $m > n > N$. We get

$$\Big(\frac{\tau(A)}{2}\Big)\sum_{k=n}^m |a_k|^2 \le \int_A \Big|\sum_{k=n}^m a_k g_k(t)\Big|^2 \tau(dt) \le \varepsilon^2\tau(A), \tag{4-9}$$

where for the last inequality we have used the fact $N > V$ and the definition of A.

We are now free to let m tend to infinity in (4-9), which we do. We deduce that necessarily $\tau(A) = 0$, a contradiction with (4-6). This finishes the proof. □

Proof of Theorem 4.3. Choose $\tau(dt)$ as the density function on the real line associated to $\tau(t) = (1-\cos t)/\pi t^2$. Then

$$\int_{\mathbb{R}} \tau(dt) = 1, \qquad \int_{\mathbb{R}} \mathrm{e}^{ixt}\tau(dt) = (1-|x|)_+.$$

Since τ is absolutely continuous with respect to the Lebesgue measure, (4-3) holds with τ in place of λ. Next choose $g_k(t) = \mathrm{e}^{i\lambda_k t}$. We have

$$\langle g_k\overline{g_\ell}, g_{k'}\overline{g_{\ell'}}\rangle_\tau = (1-|(\lambda_k-\lambda_\ell)-(\lambda_{k'}-\lambda_{\ell'})|)_+.$$

Condition (4-4) is thus fulfilled. Theorem 4.8 applies and we deduce that the series $\sum_{k=1}^{\infty}|a_k|^2$ converges. □

Final note

While finishing this paper, we discovered that Theorem 2.4 was proved by Guniya [1985] using a completely different method from ours. Guniya's proof makes use of Wiener's result [1926] (previously mentioned) and does not seem to provide directly a maximal inequality. Our proof is somewhat more elementary. Moreover it allows one to recover Wiener's result and provides at the same time a maximal inequality. It seems that Guniya's paper has been completely overlooked among the mathematical community. We observe in particular that Theorem 2.4 notably includes obviously Hedenmalm and Saksman's result [2003] published nearly twenty years after [Guniya 1985].

We now briefly explain Guniya's approach (see Theorem 1.2, (8) and Lemmas after and paragraph 2.10). The proof follows from the combination of several different results proved in the paper, and is based on Riemann theory of trigonometric series [Zygmund 1968, Chapter XVI-8]. Assume that the coefficients are positive. Then the series $\sum_n c_n e^{i\lambda_n x}$ converges in $\mathcal{S}^2$ to some f. Let I, J be two intervals with $|I| < 2\pi$, $|J| = 2\pi$ and $I \not\subseteq J$. Let F be represented by the term-by-term integrated Fourier series of f, and let L be a bump function of class C^5 equal to 1 on I and to 0 on $J\backslash I'$ where $I \subset I' \not\subseteq J$. Then by a theorem due to Zygmund [1968, Theorem 9.19], the partial sums of the Fourier series of f are uniformly equiconvergent on I with the partial sum of a trigonometric series $\sum_m a_m e^{imx}$. Next, if FL admits a second order derivative in the sense of distributions, say g, then the above trigonometric series is the one of g. And the a.e. convergence on I follows from Carleson's theorem. It remains to be proven that under condition (2-4), F has indeed second order Schwarz derivatives, controlled by the L^2 norm of f, which should follow from Theorem 2.2 in [Guniya 1985].

Acknowledgements

Part of that work was carried out while the first author was a member of the laboratory MICS from CentraleSupélec. The authors are grateful to Anna Rozanova-Pierrat and Vladimir Fock for translating the paper [Guniya 1985]. The second author is pleased to thank Michael Lacey for discussions on Carleson's theorem for integrals and its equivalence with Carleson's theorem for series (see Remarks 2.3) and for a proof of this equivalence. This point is actually also used in [Konyagin and Queffélec 2001/02], but the reference given ([Berkson et al. 1996]) does not however contain anything corresponding. See also [Zygmund 1968, Theorem 9.19].

References

[Aistleitner et al. 2015] C. Aistleitner, I. Berkes, K. Seip, and M. Weber, "Convergence of series of dilated functions and spectral norms of GCD matrices", *Acta Arith.* **168**:3 (2015), 221–246. MR Zbl

[Auscher and Carro 1992] P. Auscher and M. J. Carro, "On relations between operators on $\mathbf{R}^N$, $\mathbf{T}^N$ and $\mathbf{Z}^N$", *Studia Math.* **101**:2 (1992), 165–182. MR Zbl

[Bailey 2014] A. D. Bailey, "Pointwise convergence of lacunary partial sums of almost periodic Fourier series", *Proc. Amer. Math. Soc.* **142**:5 (2014), 1757–1771. MR Zbl

[Bellman 1944] R. Bellman, "Almost orthogonal series", *Bull. Amer. Math. Soc.* **50** (1944), 517–519. MR Zbl

[Berkson et al. 1996] E. Berkson, M. Paluszynski, and G. Weiss, "Transference couples and their applications to convolution operators and maximal operators", *Lecture Notes in Pure and Applied Mathematics* (1996), 69–84.

[Besicovitch 1932] A. S. Besicovitch, *Almost periodic functions*, Cambridge Univ. Press, 1932. Zbl

[Bredihina 1968] E. A. Bredihina, "Concerning A. N. Kolmogorov's theorem on lacunary partial sums of Fourier series", *Sibirsk. Mat. Ž.* **9** (1968), 456–461. In Russian; translated in *Siberian Math. J.* **9**:2 (1968), 345–349. MR

[Calderón 1964] A.-P. Calderón, "Intermediate spaces and interpolation, the complex method", *Studia Math.* **24** (1964), 113–190. MR Zbl

[Guniya 1985] N. G. Guniya, "Remarks on the convergence of Fourier series of almost-periodic Stepanov functions", *Trudy Tbiliss. Mat. Inst. Razmadze Akad. Nauk Gruzin. SSR* **76** (1985), 3–17. In Russian. MR Zbl

[Hardy and Riesz 1915] G. H. Hardy and M. Riesz, *The general theory of Dirichlet's series*, Cambridge Tracts in Mathematics and Mathematical Physics **18**, Cambridge Univ. Press, 1915. JFM

[Hartman 1942] P. Hartman, "The divergence of non-harmonic gap series", *Duke Math. J.* **9** (1942), 404–405. MR Zbl

[Hartman and Wintner 1938] P. Hartman and A. Wintner, "On certain Fourier series involving sums of divisors", *Trav. Inst. Math. Tbilisi* **3** (1938), 113–118. Zbl

[Hedenmalm and Saksman 2003] H. Hedenmalm and E. Saksman, "Carleson's convergence theorem for Dirichlet series", *Pacific J. Math.* **208**:1 (2003), 85–109. MR Zbl

[Hunt 1968] R. A. Hunt, "On the convergence of Fourier series", pp. 235–255 in *Orthogonal expansions and their continuous analogues* (Edwardsville, IL, 1967), edited by D. Tepper Haimo, Southern Illinois Univ. Press, Carbondale, IL, 1968. MR Zbl

[Jørsboe and Mejlbro 1982] O. G. Jørsboe and L. Mejlbro, *The Carleson–Hunt theorem on Fourier series*, Lecture Notes in Mathematics **911**, Springer, Berlin, 1982. MR Zbl

[Kac 1941] M. Kac, "Convergence and divergence of non-harmonic gap series", *Duke Math. J.* **8** (1941), 541–545. MR Zbl

[Konyagin and Queffélec 2001/02] S. V. Konyagin and H. Queffélec, "The translation $\frac{1}{2}$ in the theory of Dirichlet series", *Real Anal. Exchange* **27**:1 (2001/02), 155–175. MR Zbl

[Lacey 2004] M. T. Lacey, "Carleson's theorem: proof, complements, variations", *Publ. Mat.* **48**:2 (2004), 251–307. MR Zbl

[Lewko and Radziwiłł 2017] M. Lewko and M. Radziwiłł, "Refinements of Gál's theorem and applications", *Adv. Math.* **305** (2017), 280–297. MR Zbl

[Marcinkiewicz 1933] J. Marcinkiewicz, "A new proof of a theorem on Fourier series", *J. London Math. Soc.* **S1-8**:3 (1933), 179. MR Zbl

[Queffélec 1984] H. Queffélec, "Sur une estimation probabiliste liée à l'inégalité de Bohr", exposé 6, 1–21 in *Harmonic analysis: study group on translation-invariant Banach spaces*, Publ. Math. Orsay **84-1**, Univ. Paris XI, Orsay, 1984. MR Zbl

[Ruzsa 2001] I. Z. Ruzsa, "An almost polynomial Sidon sequence", *Studia Sci. Math. Hungar.* **38** (2001), 367–375. MR Zbl

[Titchmarsh 1948] E. C. Titchmarsh, *Introduction to the theory of Fourier integrals*, 2nd ed., Clarendon, Oxford, 1948. JFM

[Tornehave 1954] H. Tornehave, "On the Fourier series of Stepanov almost periodic functions", *Math. Scand.* **2** (1954), 237–242. MR Zbl

[Weber 2016] M. J. G. Weber, "An arithmetical approach to the convergence problem of series of dilated functions and its connection with the Riemann zeta function", *J. Number Theory* **162** (2016), 137–179. MR Zbl

[Wiener 1926] N. Wiener, "On the representation of functions by trigonometrical integrals", *Math. Z.* **24**:1 (1926), 575–616. MR

[Zygmund 1968] A. Zygmund, *Trigonometric series, I, II*, corrected 2nd ed., Cambridge Univ. Press, 1968. MR

Received August 11, 2016.

CHRISTOPHE CUNY
UNIVERSITÉ DE LA NOUVELLE-CALÉDONIE
NOUMÉA
NEW CALEDONIA
christophe.cuny@univ-nc.nc

MICHEL WEBER
IRMA, UMR 7501
STRASBOURG
FRANCE
michel.weber@math.unistra.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 292, No. 1, 2018
dx.doi.org/10.2140/pjm.2018.292.103

THE POSET OF RATIONAL CONES

JOSEPH GUBELADZE AND MATEUSZ MICHAŁEK

We introduce a natural partial order on the set Cones(d) of rational cones in $\mathbb{R}^d$. The poset of normal polytopes, studied by Bruns and the authors (*Discrete Comput. Geom.* 56:1 (2016), 181–215), embeds into Cones(d) via the homogenization map. The order in Cones(d) is conjecturally the inclusion order. We prove this for $d = 3$ and show a stronger version of the connectivity of Cones(d) for all d. Topological aspects of the conjecture are also discussed.

1. Introduction

Rational cones in $\mathbb{R}^d$ are important objects in toric algebraic geometry, combinatorial commutative algebra, geometric combinatorics, integer programming [Beck and Robins 2015; Bruns and Gubeladze 2009; Cox et al. 2011; Miller and Sturmfels 2005; Schrijver 1986]. The interaction of these convex objects with the integer lattice $\mathbb{Z}^d$ is governed by their *Hilbert bases* — the finite sets of indecomposable elements, notoriously difficult to characterize. General results on Hilbert bases are available only in low dimensions, e.g., see [Aguzzoli and Mundici 1994; Bouvier and Gonzalez-Sprinberg 1995; Sebő 1990]. In higher dimensions there are mostly counterexamples to conjectures, e.g., see [Bruns 2007; Bruns and Gubeladze 1999; Bruns et al. 1999]. In this paper we introduce a partial order on the set of rational cones in $\mathbb{R}^d$. It is defined in terms of the additive generation of the sets of lattice points in cones. The resulting poset Cones(d) is a structure in its own right, which captures a global picture of the interaction of $\mathbb{Z}^d$ with all cones at once. The poset NPol($d-1$) of *normal polytopes* in $\mathbb{R}^{d-1}$, introduced in [Bruns et al. 2016], monotonically embeds into Cones(d) via the *homogenization* map. But the former poset is much more difficult to analyze than Cones(d). In fact, there are maximal and nontrivial minimal normal polytopes; at present even the presence of isolated normal polytopes is not excluded [Bruns et al. 2016]. On the other extreme, we conjecture that the order in Cones(d) is just the inclusion order (Conjecture 2.6). We prove the 3-dimensional case of the conjecture (Theorem 3.2) and a stronger

Gubeladze was supported by NSF grant DMS 1301487. Michałek was supported by Polish National Science Center grant 2013/08/A/ST1/00804.

MSC2010: primary 52B20; secondary 05E99, 20M13, 52C07.

Keywords: rational cone, poset of cones, geometric realization of a poset.

version of the connectivity of Cones(d) for all d: any two cones can be connected by a sequence of $O(d)$ many elementary extensions/descents, or $O(d^2)$ many such moves if working with the full-dimensional cones (Theorem 4.1). In Section 5 we consider topological consequences of Conjecture 2.6.

1A. ***Cones.*** We consider the real vector space $\mathbb{R}^d$, consisting of d-columns, together with the integer lattice $\mathbb{Z}^d$. The standard basis vectors will be denoted by $\boldsymbol{e}_1, \dots, \boldsymbol{e}_d$, the set of nonnegative reals will be denoted by $\mathbb{R}_+$, and the set of nonnegative integers will be denoted by $\mathbb{Z}_+$.

For a subset $X \subset \mathbb{R}^d$, its *conical hull*, i.e., the set of nonnegative linear combinations of elements of X, is denoted by $\mathbb{R}_+X$. The linear span of X will be denoted by $\mathbb{R}X$. We also put $\mathsf{L}(X) = X \cap \mathbb{Z}^d$.

By a *cone* C we always mean a pointed, rational, polyhedral cone, i.e., $C = \mathbb{R}_+x_1 + \cdots + \mathbb{R}_+x_n$ for some $x_1, \dots, x_n \in \mathbb{Z}^d$ and there is no nonzero element $x \in C$ with $-x \in C$. Let $C \subset \mathbb{R}^d$ be a nonzero cone. Then there exists an affine hyperplane H, meeting C *transversally*, i.e., such that $C \cap H$ is a polytope of dimension $\dim(C) - 1$ [Bruns and Gubeladze 2009, Proposition 1.21]. The first nonzero lattice point on each 1-dimensional face of C is called an *extremal generator of* C. The additive submonoid $\mathsf{L}(C) \subset \mathbb{Z}^d$ has the smallest generating set, consisting of indecomposable elements. It is called the *Hilbert basis* of C, denoted by $\operatorname{Hilb}(C)$. The extremal generators of C belong to $\operatorname{Hilb}(C)$.

A d-cone $C \subset \mathbb{R}^d$ has a unique minimal representation as an intersection of closed half-spaces $C = \bigcap_{j=1}^n H_j^+$. The boundary hyperplanes $H_j \subset H_j^+$ intersect C in its *facets*, i.e., the codimension 1 proper faces of C. Further, for each facet $F \subset C$ there exists a unique linear function $\operatorname{ht}_F : \mathbb{R}^d \to \mathbb{R}$ which vanishes on F, is nonnegative on C, and satisfies $\operatorname{ht}_F(\mathbb{Z}^d) = \mathbb{Z}$.

A pair of cones (C, D) is a *unimodular extension* of cones if C is a facet of D, the latter has exactly one extremal generator v not in C, and $\mathsf{L}(D) = \mathsf{L}(C) + \mathbb{Z}_+v$.

A cone $C \subset \mathbb{R}^d$ is called *unimodular* if $\operatorname{Hilb}(C)$ is a part of a basis of $\mathbb{Z}^d$.

If the extremal generators of a cone C are linearly independent, then C is said to be *simplicial*.

For elements $u_1, \dots, u_d \in \mathbb{R}^d$ the matrix, whose i-th column is u_i, will be denoted by $[u_1|\cdots|u_d]$. Assume $u_1, \dots, u_d$ are linearly independent and

$$C = \mathbb{R}_+u_1 + \cdots + \mathbb{R}_+u_d.$$

Then we put

$$\begin{aligned}
\mathsf{par}(u_1, \dots, u_d) &= \{\lambda_1u_1 + \cdots + \lambda_du_d \mid 0 \le \lambda_1, \dots, \lambda_d < 1\},\\
\mathsf{Lpar}(u_1, \dots, u_d) &= \mathsf{L}(\mathsf{par}(u_1, \dots, u_d)) \setminus \{0\},\\
\operatorname{vol}(u_1, \dots, u_d) &= \operatorname{vol}\big(\mathsf{par}(u_1, \dots, u_d)\big) = |\det[u_1|\cdots|u_d]|,\\
\mu(C) &= \operatorname{vol}(u_1, \dots, u_d) \quad \text{ if the } u_i \text{ are primitive}
\end{aligned}$$

(where *primitive* means having coprime components).

A triangulation of a cone C into simplicial cones is called *unimodular* if the cones in the triangulation are unimodular, and it is called *Hilbert* if the set of extremal generators of the involved cones equals $\operatorname{Hilb}(C)$.

Proposition 1.1. (a) *Let $C \subset \mathbb{R}^d$ be a nonzero cone and $v \in \mathsf{L}(C)$ be a nonzero element in a 1-face of C. Then $\mathsf{L}(C)+\mathbb{Z}v = \mathsf{L}(C_0)+\mathbb{Z}v \cong \mathsf{L}(C_0)\times \mathbb{Z}v$ for some cone $C_0 \subset \mathbb{R}^d$ with $v \notin C_0$.*

(b) *Let $C \subset \mathbb{R}^d$ be a nonzero cone and $w \in \mathsf{L}(C)$ be an element in the relative interior of C. Then*
$$\mathsf{L}(C)+\mathbb{Z}w = \mathsf{L}(\mathbb{R}C).$$

(c) *Every nonzero cone has a unimodular triangulation.*

(d) *For every 2-cone C, its only Hilbert triangulation is unimodular.*

(e) *Every 3-dimensional cone has a unimodular Hilbert triangulation.*

The parts (a), (b), (c), (d), are standard results on cones and all five parts are proved, for instance, in [Bruns and Gubeladze 2009, Chapter 2]. The part (e) is originally due to Sebő [1990] (whose argument is reproduced in [Bruns and Gubeladze 2009, Theorem 2.78]). It was later rediscovered in the context of toric geometry in [Aguzzoli and Mundici 1994; Bouvier and Gonzalez-Sprinberg 1995], with important refinements. The existence of unimodular Hilbert triangulations fails already in dimension 4 [Bouvier and Gonzalez-Sprinberg 1995].

For a poset $(\Pi, <)$, the geometric realization of its *order (simplicial) complex* will be called the *geometric realization* of Π and denoted by $|\Pi|$. For generalities on poset topology we refer the reader to [Wachs 2007], with the caution that our posets are mostly infinite. But the "finite vs. infinite" dichotomy never plays a role in our treatment. Section 1 in Quillen's foundational work on higher algebraic K-theory [Quillen 1973] remains an indispensable source for homotopy studies of general posets (in fact, general categories).

1B. ***The poset of normal polytopes.*** A lattice polytope $P \subset \mathbb{R}^d$ (i.e., a convex polytope with vertices in $\mathbb{Z}^d$) is *normal* if for every $c \in \mathbb{N}$ and every element $x \in \mathsf{L}(cP)$ there exist $x_1, \dots, x_c \in \mathsf{L}(P)$, such that $x = x_1 + \cdots + x_c$.

The order in the poset $\operatorname{NPol}(d)$ of normal polytopes in $\mathbb{R}^d$, studied in [Bruns et al. 2016], is generated by the following elementary relations: $P < Q$ if $P \subset Q$ and $\#\mathsf{L}(Q) = \#\mathsf{L}(P) + 1$.

The poset $\operatorname{NPol}(d)$ is known to have (nontrivial) minimal and maximal elements in dimensions ≥ 4.

The *homogenization* map $P \mapsto C(P) := \mathbb{R}_+(P \times \{1\}) \subset \mathbb{R}^d$ embeds the set of lattice polytopes $P \subset \mathbb{R}^{d-1}$ into that of cones $C \subset \mathbb{R}^d$. Moreover, a lattice polytope P is normal if and only if $\operatorname{Hilb}(C(P)) = \{(x, 1) \mid x \in \mathsf{L}(P)\}$.

For a lattice d-polytope $P \subset \mathbb{R}^d$ and a facet $F \subset P$ there exists a unique affine map $\mathrm{ht}_F : \mathbb{R}^d \to \mathbb{R}$ with $\mathrm{ht}_F(P) \subset \mathbb{R}_+$ and $\mathrm{ht}_F(\mathbb{Z}^d) = \mathbb{Z}$. We have the following compatibility between the facet-height functions: the two maps $\mathrm{ht}_F(\,\cdot\,), \mathrm{ht}_{C(F)}(\,\cdot\,, 1) : \mathbb{R}^d \to \mathbb{R}$ are the same.

2. The poset Cones(d)

2A. *Elementary extensions.* For a natural number d we denote by $\mathrm{Cones}(d)$ the set of cones $C \subset \mathbb{R}^d$, made into a poset as follows: $C < D$ if and only if there exists a sequence of cones of the form

$$\begin{gathered} C = C_0 \subset \cdots \subset C_{n-1} \subset C_n = D, \\ \mathsf{L}(C_i) = \mathsf{L}(C_{i-1}) + \mathbb{Z}_+ x, \quad \text{for some } \ x \in C_i \setminus C_{i-1}, \ \ i = 1, \dots, n. \end{gathered} \tag{1}$$

When $n = 1$ we call $C \subset D$ an *elementary extension*, or *elementary descent* if read backwards. Here is an alternative characterization:

Lemma 2.1. *Let $C \subset \mathbb{R}^d$ be a nonzero cone and $v \in \mathbb{Z}^d$ be a primitive vector with $\pm v \notin C$. Assume $H \subset \mathbb{R}^d \setminus \{0\}$ is an affine hyperplane, meeting the cone $D = C + \mathbb{R}_+ v$ transversally. Put $v' = \mathbb{R}_+ v \cap H$. Then $C \subset D$ is an elementary extension in* $\mathrm{Cones}(d)$ *if and only if there exist unimodular cones $U_1, \dots, U_n \subset D$, satisfying the conditions*

(i) $v \in U_i$, $i = 1, \dots, n$,

(ii) $D = C \bigcup \left(\bigcup_{i=1}^n U_i\right)$,

(iii) $\{\mathbb{R}_+((U_i \cap H) - v')\}_{i=1}^n$ *is a triangulation of the cone* $\mathbb{R}_+((D \cap H) - v')$.

Proof. The "if" part is obvious. For the "only if" part we use (a) and (c) of Proposition 1.1 to fix a representation $\mathsf{L}(C) + \mathbb{Z}v = \mathsf{L}(D) + \mathbb{Z}v = \mathsf{L}(C_0) + \mathbb{Z}v = \mathsf{L}(C_0) \times \mathbb{Z}v$ and a unimodular triangulation $C_0 = \bigcup_{i=1}^n D_i$. Let $X \subset C$ be a finite subset, which maps bijectively to $\bigcup_{i=1}^n \mathrm{Hilb}(D_k)$ under the projection $C \to C_0$, induced by $v \mapsto 0$. Let X_i be the preimage of $\mathrm{Hilb}(D_i)$ in X. Then the cones $U_i = \mathbb{R}_+ X_i + \mathbb{R}_+ v$ satisfy (i)–(iii). □

2B. *Height* 1 *and Hilbert basis extensions.* $\mathrm{Cones}(d)$ contains many elementary extensions of two different types, making it essentially different from $\mathrm{NPol}(d)$.

Let $C \subset \mathbb{R}^d$ be a d-cone and $v \in \mathbb{Z}^d$ with $\pm v \notin C$. Denote by $\mathbb{F}^+(v)$ the set of facets of C, visible from v, i.e., $\mathrm{ht}_F(v) < 0$ for every $F \in \mathbb{F}^+(v)$. Consider the visible part of the boundary ∂C, i.e., $C^+(v) = \bigcup_{\mathbb{F}^+(v)} F$. Put $D = C + \mathbb{R}_+ v$. There is a sequence of rational numbers $0 < \lambda_1 < \lambda_2 < \dots$ with $\lambda_1 = 1/(\max(-\mathrm{ht}_F(v) \ : \ F \in \mathbb{F}^+(v)))$ and $\lim_{k\to\infty} \lambda_k = \infty$, satisfying the conditions:

$$\mathsf{L}(D \setminus C) = \bigcup_{k=1}^{\infty} \mathsf{L}(\lambda_k v + C^+(v)) \quad \text{and} \quad \mathsf{L}(\lambda_k v + C^+(v)) \neq \varnothing, \quad k = 1, 2, \dots.$$

The equality $\lambda_1 = 1$ is equivalent to the condition $\mathrm{ht}_F(v) = -1$ for all $F \in \mathbb{F}^+(v)$. In this case we say that D is a *height* 1 *extension* of C. All height 1 extensions are elementary extensions of cones but the converse is not true [Bruns et al. 2016, Theorem 4.3].

The second class of elementary extensions in $\mathrm{Cones}(d)$ are the extensions of type $C \subset D$, where $\mathbb{R}_+(\mathrm{Hilb}(D) \setminus \{v\}) \subset C$ for an extremal generator $v \in D$. We call this class the *Hilbert basis extensions* (or *descents*).

As an application of the two types of extensions, we have:

Lemma 2.2. *For every natural number $d \geq 2$,*

(a) *for every elementary extension of cones $0 \neq C < D$ there exists a cone E, such that $C < E < D$;*

(b) $\mathrm{Cones}(d)$ *has neither maximal nor minimal elements, other than the minimal element* 0.

Remark. We do not know whether 0 is the *smallest* element of $\mathrm{Cones}(d)$. If 0 were the smallest element, then the geometric realization of $\mathrm{Cones}(d)$ would be contractible; see Section 5 for topological aspects of $\mathrm{Cones}(d)$.

Proof. (a) The general case easily reduces to the full-dimensional case and then the claim follows from the observation that there is always a height 1 extension $C \subset E$ with $E \subsetneq D$. In fact, if $\{v\} = \mathrm{Hilb}(D) \setminus C$, then we can take $E = C + \mathbb{R}_+ w$ where $w \in \mathsf{L}(\lambda_1 v + C^+(v))$ with $w \neq v$ (notation as above). Obviously, $E \subset D$ is an elementary extension.

(b) One applies appropriate height 1 extensions to show that there are no maximal elements, and Hilbert basis descents to show that there are no minimal elements in $\mathrm{Cones}(d) \setminus \{0\}$. □

We formally include the extensions of type $0 \subset C,\ \dim C = 1$, in both classes of elementary extensions, discussed above.

Question 2.3. Do either the height 1 or Hilbert basis extensions generate the same poset $\mathrm{Cones}(d)$?

2C. *Distinguished subposets.* The subposet of $\mathrm{Cones}(d)$, consisting of the cones in $(\mathbb{R}^{d-1} \times \mathbb{R}_{>0}) \cup \{0\}$, will be denoted by $\mathrm{Cones}^+(d)$. The homogenization embedding $\mathrm{NPol}(d-1) \to \mathrm{Cones}^+(d)$ is a monotonic map. However, the order in $\mathrm{NPol}(d-1)$ is weaker than the one induced from $\mathrm{Cones}(d)$:

Example 2.4. In [Bruns et al. 2016, Example 4.8] we have the polytope $P \in \mathrm{NPol}(3)$ with vertices $(0, 0, 2)$, $(0, 0, 1)$, $(0, 1, 3)$, $(1, 0, 0)$, $(2, 1, 2)$, $(1, 2, 1)$. The polytope has two more lattice points: $(1, 1, 2)$, $(1, 1, 1)$. Removing either the first *or* the second vertex and taking the convex hull of the other lattice points in P yields a nonnormal polytope. However, the convex hull Q of the lattice points in P with the exception of the first two vertices is normal. We have $Q \not< P$ in $\mathrm{NPol}(3)$. Yet, using polymake [Gawrilow and Joswig 1997], one quickly finds four

Hilbert basis descents (requiring additional Hilbert basis elements at height two) $C(P) > C_1 > C_2 > C_3 > C(Q)$.

For every integer $h > 0$ we consider the poset $\mathrm{Cones}^{(h)}(d)$ of cones in $\mathrm{Cones}^{+}(d)$, satisfying $\mathrm{Hilb}(C) \subset \mathbb{R}^{d-1} \times [0, h]$ and ordered as in (1) under the additional requirement that the intermediate cones C_i are also from $\mathrm{Cones}^{(h)}(d)$.

Lemma 2.5. *For every natural $d \geq 1$,*

(a) $\mathrm{Cones}^{(1)}(d) \setminus \{0\} = \mathrm{NPol}(d-1)$;

(b) $\mathrm{Cones}^{(1)}(d) \subset \mathrm{Cones}^{(2)}(d) \subset \cdots$ *and* $\bigcup_{h=0}^{\infty} \mathrm{Cones}^{(h)}(d) = \mathrm{Cones}^{+}(d)$;

(c) $\mathrm{Pol}(d-1) \subset \mathrm{Cones}^{(d-2)}(d)$, *assuming $d \geq 3$.*

(Inclusions are those of sets, and may not represent subposets.)

Parts (a) and (b) are obvious; (c) is proved, for instance, in [Bruns and Gubeladze 2009, Theorem 2.52].

2D. *The cone conjecture.* Conjecture 2.6 is the maximal possible strengthening of the absence of extremal elements in $\mathrm{Cones}(d)$:

Conjecture 2.6. *For every d, the order in* $\mathrm{Cones}(d)$ *is the inclusion order.*

The case $d = 1$ is obvious.

When $d = 2$, the general case reduces to a pair of cones $C \subset D$ in $\mathbb{R}^2$, with $\dim D = 2$ and C a facet of D. Assume $\{v_1, \ldots, v_n\} = \mathrm{Hilb}(D)$ and $v_1 \in C$. Then, by Proposition 1.1(d), we have the following height 1 extensions:

$$C < C + \mathbb{R}_+ v_2 < \cdots < C + \mathbb{R}_+ v_2 + \cdots + \mathbb{R}_+ v_n = D.$$

In Section 3 we give a proof for $d = 3$.

In dimension 4 we have the following computational evidence.

Assume $C \subset \mathbb{R}^d$ is a cone and $v \in \mathbb{Z}^d$ with $\pm v \notin C$. We use the notation in Section 2B. In particular, $D = C + \mathbb{R}_+ v$. One introduces the *bottom-up* procedure for constructing an ascending sequence of height 1 extensions, starting with the cone C, as follows: one chooses a shortest vector $v_1 \in \mathsf{L}(\lambda_1 v + C^+(v))$, repeats the step for the pair $C_1 \subset D$ where $C_1 = C + \mathbb{R}_+ v_1$, and iterates the process. The height 1 extensions we obtain this way tend to widen the cone as much as possible at each step, as measured by the increments of the Euclidean $(d-1)$-volume of the cross sections with a prechosen affine hyperplane, transversally meeting the cone D.

A complementary approach employs Hilbert basis descents. The corresponding *top-down* procedure finds a sequence $D = D_0 > D_1 > \cdots$ of Hilbert basis descents of the form $D_{i+1} = C + \mathbb{R}_+(\mathrm{Hilb}(D_i) \setminus \{v_i\})$, at each step discarding a shortest extremal generator $v_i \in D_i \setminus C$.

Andreas Paffenholz implemented the bottom-up and top-down procedures in $\mathbb{R}^4$. The computational evidence, based on many randomly generated cones C and vectors v, supports the expectation that there are no nonterminating processes of either type, with the tendency of the bottom-up process to last longer than the top-down one.

3. Cones in $\mathbb{R}^3$

Lemma 3.1. *Let $x, y \in \mathsf{par}(u, v, w)$, and let $u, v, w \in \mathbb{R}^3$ be linearly independent vectors. Then* $\mathrm{vol}(u, x, y) < \mathrm{vol}(u, v, w)$.

Proof. We can assume $(u, v, w) = (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$. Let $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$. Then

$$\mathrm{vol}(\boldsymbol{e}_1, x, y) = \left|\det\begin{pmatrix} 1 & 0 & 0 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix}\right| = |x_2y_3 - x_3y_2| \leq \max\bigl(|x_2y_3|, |x_3y_2|\bigr) < 1. \qquad \square$$

Theorem 3.2. *The order in* $\mathrm{Cones}(d)$ *is the inclusion order for $d = 3$.*

Proof. We first prove the following basic case: for any simplicial 3-cone $D \subset \mathbb{R}^3$ and any facet $C \subset D$ we have $C < D$. This will be done by induction on $\mu(D)$ (defined in the introduction).

The case $\mu(C) = \mu(D)$ is obvious because D is a unimodular extension of C. So we can assume $\mu(C) < \mu(D)$, which is equivalent to $\mathsf{Lpar}(D) \not\subset C$.

Let v_0, v_1, w be the extremal generators of D with $v_0, v_1 \in C$. Denote by $v_0, v_1, v_2, \dots, v_k$ $(k \geq 2)$ the extremal generators of the cone

$$E = C + \mathbb{R}_+\mathsf{Lpar}(D) \subset \mathbb{R}^3.$$

We assume that the enumeration is done in the cyclic order, i.e., the cones

$$C_i = \mathbb{R}_+v_{i-1} + \mathbb{R}_+v_i \subset \mathbb{R}^3, \quad i = 1, \dots, k, k+1 \bmod (k+1)$$

are the facets of E. (Here, $C = C_1$.)

Because of the containment $\mathrm{Hilb}(D) \setminus \{w\} \subseteq E$, we have $E < D$ in the poset $\mathrm{Cones}(3)$. Further, the cone E is triangulated by the cones

$$D_i = \mathbb{R}_+v_0 + \mathbb{R}_+v_i + \mathbb{R}_+v_{i+1}, \quad i = 1, \dots, k-1.$$

By Lemma 3.1, we have the inequalities

$$\mu(D_i) < \mu(D), \quad i = 2, \dots, k.$$

Then, by the induction hypothesis, we have $C < D_1$ and

$$(D_{i-1} \cap D_i) < D_i, \quad i = 2, \dots, k-1.$$

By concatenating, we obtain the following chain in $\mathrm{Cones}(3)$:

$$C < D_1 < D_1 \cup D_2 < \cdots < D_1 \cup D_2 \cup \cdots \cup D_{k-1} = E < D.$$

This completes the proof of the basic case.

The general case easily reduces to the case of a pair of 3-cones $C \subsetneq D$ with $D = C + \mathbb{R}_+ v$, to which we apply induction on the number of facets of C visible from v. When this number is 1, the inequality $C < D$ results from the basic case. When the number of the visible facets is $k \geq 2$ then there is an intermediate cone $C \subsetneq B \subsetneq D$, satisfying the conditions

- $B = C + \mathbb{R}_+ w$ for some w;
- B has only one facet visible from v;
- there are exactly $k - 1$ facets of C, visible from w.

In fact, if $C = \bigcap_{j=1}^{l} H_j^+$ is the irreducible representation, where the indexing is in the circular order and $H_1 \cap C, \dots, H_k \cap C \subset C$ are the facets visible from v, then one can choose

$$B = \Big(\bigcap\nolimits_{j=k}^{l} H_j^+\Big) \bigcap D.$$

We are done because, by the induction hypothesis, $C < B < D$. □

4. Diameter

By the *diameter* of a subposet $X \subset \mathrm{Cones}(d)$, denoted $\mathrm{D}(X)$, we mean the supremum of the lengths of the shortest sequences $C_0 C_1 \cdots C_n$ within X, connecting any two elements C_0, C_n of X, where every two consecutive cones form an elementary extension or descent.

Consider the following subposets of $\mathrm{Cones}(d)$:

(i) $\mathrm{Cones}(d)^{\mathrm{o}}$, consisting of the d-cones in $\mathbb{R}^d$ (all quantum jumps in $\mathrm{NPol}(d-1)$ live here).

(ii) $\mathrm{Unim}(d)$, consisting of the unimodular cones in $\mathbb{R}^d$.

(iii) $\mathrm{Unim}(d)^{\mathrm{o}}$ consisting of the unimodular d-cones in $\mathbb{R}^d$.

The next theorem implies that $\mathrm{Cones}(d)$ and $\mathrm{Cones}(d)^{\mathrm{o}}$ are both connected.

Theorem 4.1. *We have*:

(a) $\mathrm{D}(\mathrm{Unim}(d)) = 2d$ *for every* $d \in \mathbb{N}$.

(b) $\mathrm{D}(\mathrm{Unim}(d)^{\mathrm{o}}) = O(d^2)$.

(c) $\mathrm{D}(\mathrm{Cones}(d)) = O(d)$.

(d) $\mathrm{D}(\mathrm{Cones}(d)^{\mathrm{o}}) = O(d^2)$.

Proof. (a) Any unimodular cone can be reached from any other unimodular cone by first removing the Hilbert basis elements of the latter, one by one, and then adding those of the former, also one at a time.

For the pairs of unimodular d-cones of type C and $-C$ there is no shorter connecting path. One should remark that this is not true for all pairs of unimodular d-cones whose intersection is 0; an example when $d=2$ is

$$\mathbb{R}_+\boldsymbol{e}_1+\mathbb{R}_+(\boldsymbol{e}_1+\boldsymbol{e}_2)<\mathbb{R}_+(-\boldsymbol{e}_1+\boldsymbol{e}_2)+\mathbb{R}_+\boldsymbol{e}_1>\mathbb{R}_+(-\boldsymbol{e}_1+\boldsymbol{e}_2)+\mathbb{R}_+\boldsymbol{e}_2.$$

(b) Let $C=\sum_{i=1}^{d}\mathbb{Z}_+v_i$ and $D=\sum_{i=1}^{d}\mathbb{Z}_+w_i$ for two bases $\{v_1,\dots,v_d\}$ and $\{w_1,\dots,w_d\}$ of $\mathbb{Z}^d$. Put $A=[v_1|\cdots|v_d]$ and $B=[w_1|\cdots|w_d]$. After renumbering of the basis elements, we can assume $\det(A)=\det(B)=1$. The special linear group $\mathrm{SL}_d(\mathbb{Z})$ is generated by the elementary matrices e_{ij}^a, i.e., the matrices with ones on the main diagonal, at most one nonzero off-diagonal entry a in the ij-spot, and zeros elsewhere. Using the equalities $(e_{ij}^a)^{-1}=e_{ij}^{-a}$, there is a representation of the form $Ae_{i_1j_1}^{a_1}\cdots e_{i_kj_k}^{a_k}=B$, where $a_1,\dots a_k,\in\mathbb{Z}$. By [Carter and Keller 1984], one can choose $k\le 36+\frac{1}{2}(3d^2-d)$. Consider the sequence of unimodular cones:

$$C_t \;=\; \text{the cone spanned by the columns of } Ae_{i_1j_1}^{a_1}\cdots e_{i_tj_t}^{a_t},\qquad 0\le t\le k.$$

(In particular, $C_0=C$). Since the multiplications by elementary matrices from the right corresponds to the elementary column transformations, for every $1\le t\le k$ the inequality $a_t>0$ yields the elementary extension $C_t<C_{t-1}$ and the inequality $a_t<0$ yields the elementary descent $C_t>C_{t-1}$.

(c), (d) For $d\le 1$ there are connecting paths of length ≤ 2. So we assume $d\ge 2$.

Pick $C\in\mathrm{Cones}(d)$. By taking unimodular extensions as needed, we can assume $\dim C=d$. We need at most $2d-1$ unimodular extensions to reach the full-dimensional case. Consequently, the parts (c) and (d) follow from the parts (a) and (b), respectively, once we show that a unimodular d-cone can be reached from C in at most $d-1$ elementary extensions/descents.

Pick arbitrarily a facet $F\subset C$ and two elements $y\in\mathsf{L}(C\setminus F)$, satisfying $\mathrm{ht}_F(y)=1$, and $x\in\mathsf{L}(\mathrm{int}(F))$, where $\mathrm{int}(F)$ is the relative interior of F. Consider the sequence of cones

$$C_k=F+\mathbb{R}_+(y-kx),\qquad k=0,1,\dots.$$

We claim that $C\subset C_k$ for all sufficiently large k.

Indeed, consider any extremal generator v of C. We have $v=\mathrm{ht}_F(v)y+v'$ for some $v'\in\mathbb{Z}^d$ with $H_F(v')=0$. By Proposition 1.1(b), $\mathsf{L}(F)+\mathbb{Z}x=\mathsf{L}(\mathbb{R}F)$. Hence $v'=-sx+z$ for some $z\in\mathsf{L}(F)$ and an integer $s\ge 0$. Consequently,

$$v=\mathrm{ht}_F(v)\Big(y-\Big\lceil\frac{s}{\mathrm{ht}_F(v)}\Big\rceil x\Big)+\mathrm{ht}_F(v)\Big(1-\Big\{\frac{s}{\mathrm{ht}_F(v)}\Big\}\Big)x+z\in C_{\lceil s/(\mathrm{ht}_F(v))\rceil}.$$

Pick $k\gg 0$ with $C\subset C_k$. Since C_k is a unimodular extension of F, we have the elementary extension $C<C_k$ in $\mathrm{Cones}(d)$.

Keeping $\mathbb{R}_+(y-kx)$ as a 1-face, we may, inductively on dimension, transform F to a unimodular $d-1$-cone using only elementary extensions and descents: one uses the fact that unimodular extensions of cones respect elementary extensions in the previous dimension. In the end, starting from C, we have reached a unimodular d-cone (in at most $d-1$ steps). □

Remark 4.2. In the proofs of Theorem 4.1(a) and (c), one does not need to descend from unimodular cones all the way to 0. The latter, not being in $\mathrm{NPol}(d-1)$, may not be desirable. It is enough to descend to 1-dimensional cones and the same argument as in the proof of Theorem 4.1(b) shows that for any pair of 1-cones in $\mathbb{R}^d$ there is an upper bound on the number of connecting elementary extensions/descents: one finds such extensions within the linear span of the pair of 1-cones. By avoiding 0 the diameter goes up by a constant, independent of d.

The proof of Theorem 4.1 does *not* imply that $\mathrm{D}(\mathrm{Cones}^+(d)) < \infty$.

5. The space of cones

Conjecture 2.6 has strong consequences for the geometric realization of $\mathrm{Cones}(d)$:

Theorem 5.1. *Assume Conjecture 2.6 holds for a natural number d. Then*:

(a) *The spaces* $|\mathrm{Cones}(d)|$, $|\mathrm{Cones}^+(d)|$, *and* $|\mathrm{Cones}^+(d) \setminus \{0\}|$ *are contractible.*

(b) $|\mathrm{Cones}(d) \setminus \{0\}|$ *is a filtered union of spaces, each containing a $(d-1)$-sphere as a strong deformation retract.*

Proof. (a) The spaces $|\mathrm{Cones}(d)|$ and $|\mathrm{Cones}^+(d)|$ are contractible because 0 is the smallest element of $\mathrm{Cones}(d)$ and $\mathrm{Cones}^+(d)$. The poset $\mathrm{Cones}^+(d) \setminus \{0\}$ is *filtering,* i.e., every finite subset has an upper bound. But the geometric realization of a filtering poset is contractible [Quillen 1973, Section 1].

(b) Let S^{d-1} be the unit $(d-1)$-sphere in $\mathbb{R}^d$, centered at the origin. Then we can think of the poset of $\mathrm{Cones}(d) \setminus \{0\}$ as the poset of intersections $C \cap S^{d-1}$, $C \in \mathrm{Cones}(d)$, ordered by inclusion. Abusing terminology, these intersections will be also called polytopes.

For two polytopal subdivisions Π_1 and Π_2 of S^{d-1} and a polytope $P \subset S^{d-1}$ we write (i) $\Pi_1 \prec \Pi_2$ if Π_2 is a subdivision of Π_1 and (ii) $P \prec \Pi_1$ if P is subdivided by polytopes in Π_1.

Fix a system of polytopal subdivisions $\{\Pi_i\}_{i=1}^{\infty}$ of S^{d-1}, such that $\Pi_i \prec \Pi_{i+1}$ for all i and every polytope $P \subset S^{d-1}$ admits i with $P \prec \Pi_i$.

For every index i, the simplicial complex $|\Pi_i|$ is a barycentric subdivision of Π_i. In particular, $|\Pi_i| \cong S^{d-1}$.

Consider the following posets:

- $\check{\Pi}_i = \{P \in \mathrm{Cones}(d) \setminus \{0\} \mid P \prec \Pi_i\}$, made into a poset by adding to the inclusion order in Π_i the new relations $Q < P$ whenever $P \in \check{\Pi}_i \setminus \Pi_i$, $Q \in \Pi_i$, $Q \subset P$; in particular, two different polytopes P and $P' \in \check{\Pi}_i \setminus \Pi_i$ are not comparable.

- The subposet $\overline{\Pi}_i = \{P \in \mathrm{Cones}(d) \setminus \{0\} \mid P \prec \Pi_i\} \subset \mathrm{Cones}(d)$; it has more relations than the poset $\check{\Pi}_i$, supported by the same set of polytopes, since for P and $P' \in \check{\Pi}_i \setminus \Pi_i$ one has $P < P'$ whenever $P \subset P'$.
- The subposets $\Pi_i(P) = \{Q \mid Q \in \Pi_i,\ Q \subset P\} \cup \{P\} \subset \mathrm{Cones}(d)$ for $P \prec \Pi_i$.

The (geometric) simplicial complex $|\check{\Pi}_i|$ is obtained from $|\Pi_i|$ by changing the contractible subcomplexes $|\Pi_i(P)|$ to pyramids over them. Any two of these pyramids either do not meet outside $|\Pi_i|$ or overlap along a pyramid from the same family. In particular, the subspace $|\Pi_i| \subset |\check{\Pi}_i|$ is a strong deformation retract. Let $F : |\check{\Pi}_i| \times [0, 1] \to |\check{\Pi}_i|$ be a corresponding homotopy.

Consider an extension of F to a homotopy

$$G : |\overline{\Pi}_i| \times [0, 1] \to |\overline{\Pi}_i|,$$

satisfying the condition that for every $t \in [0, 1]$ the map G_t is injective on $|\overline{\Pi}_i| \setminus |\check{\Pi}_i|$ and is the identity on $|\Pi_i|$. In more detail, for every chain

$$P_0 < \cdots < P_k < P_{k+1} < \cdots < P_n, \qquad P_k \in \Pi_i,\ P_{k+1} \in \check{\Pi}_i \setminus \Pi_i,$$

and every index $k < l \leq n$, the l-subsimplex $\triangle(P_0, \ldots, P_k, P_l)$ of the n-simplex $\triangle(P_0, \ldots, P_n)$ is collapsed into the k-subsimplex $\triangle(P_0, \ldots, P_k)$ by the homotopy G, while the rest of the n-simplex homeomorphically remains invariant. In particular, $G_1(\triangle(P_1, \ldots, P_n))$ is an n-disc, attached to $|\Pi_i|$ along the subdisc $\triangle(P_1, \ldots, P_k)$. Then $\operatorname{Im} G_1$ consists of $|\Pi_i|$ and the mentioned finitely many attached discs, any two of which either do not meet outside $|\Pi_i|$ or overlap along a disc from the same family.

The claim now follows because $|\Pi_i|$ is a strong deformation retract of $\operatorname{Im} G_1$. □

Remark. It is very likely that a more elaborate homotopy leads to a deformation retraction of the total space $|\mathrm{Cones}(d) \setminus \{0\}|$ to a $(d-1)$-sphere.

By Lemma 2.5(c), we have the tower of spaces

$$|\mathrm{NPol}(d-1)| = |\mathrm{Cones}^{(1)}(d) \setminus \{0\}| \subset |\mathrm{Cones}^{(2)}(d) \setminus \{0\}| \subset \cdots,$$

which, in view of Theorem 5.1, is expected to trivialize in the limit. This observation can lead to an insight into the more difficult space of normal polytopes if the trivialization occurs in a controlled way, which is an interesting question in its own right. In more detail, the group $\mathrm{Aff}_{d-1}(\mathbb{Z})$ of affine automorphisms of $\mathbb{Z}^{d-1}$ acts compatibly on the whole tower of posets

$$\mathrm{Cones}^{(1)}(d) \setminus \{0\} \subset \mathrm{Cones}^{(2)}(d) \setminus \{0\} \subset \mathrm{Cones}^{(3)}(d) \setminus \{0\} \subset \cdots$$

via the embedding

$$\mathrm{Aff}_{d-1}(\mathbb{Z}) \to \mathrm{GL}_d(\mathbb{Z}), \quad (\alpha|\beta) \mapsto \begin{pmatrix} \alpha & \beta \\ 0 & 1 \end{pmatrix}, \quad \alpha \in \mathrm{GL}_{d-1}(\mathbb{Z}),\ \beta \in \mathbb{Z}^{d-1}.$$

As a result, the homology groups of all involved geometric realizations are modules over the group ring $\mathbb{Z}[\mathrm{Aff}_{d-1}(\mathbb{Z})]$.

Question 5.2. Are the relative homology groups

$$H_i\big(|\mathrm{Cones}^{(j)}(d)\setminus\{0\}|, |\mathrm{Cones}^{(j-1)}(d)\setminus\{0\}|, \mathbb{Z}\big)$$

finitely generated $\mathbb{Z}[\mathrm{Aff}_{d-1}(\mathbb{Z})]$-modules for all i and j?

The positive answer to this question for $i=0$ (and all j), would imply that the still elusive isolated elements in $\mathrm{NPol}(d-1)$ form a highly structured family: for every j, only finitely many such isolated elements (up to unimodular equivalence) cease to be isolated when one passes from $\mathrm{Cones}^{(j-1)}(d)\setminus\{0\}$ to $\mathrm{Cones}^{(j)}(d)\setminus\{0\}$, and all isolated elements are taken out as $j\to\infty$.

Acknowledgement

We are grateful to Andreas Paffenholz for carrying out computational experiments, discussed in Section 2D.

References

[Aguzzoli and Mundici 1994] S. Aguzzoli and D. Mundici, "An algorithmic desingularization of 3-dimensional toric varieties", *Tohoku Math. J.* (2) **46**:4 (1994), 557–572. MR Zbl

[Beck and Robins 2015] M. Beck and S. Robins, *Computing the continuous discretely*, 2nd ed., Springer, New York, 2015. MR Zbl

[Bouvier and Gonzalez-Sprinberg 1995] C. Bouvier and G. Gonzalez-Sprinberg, "Système générateur minimal, diviseurs essentiels et G-désingularisations de variétés toriques", *Tohoku Math. J.* (2) **47**:1 (1995), 125–149. MR Zbl

[Bruns 2007] W. Bruns, "On the integral Carathéodory property", *Experiment. Math.* **16**:3 (2007), 359–365. MR Zbl

[Bruns and Gubeladze 1999] W. Bruns and J. Gubeladze, "Normality and covering properties of affine semigroups", *J. Reine Angew. Math.* **510** (1999), 161–178. MR Zbl

[Bruns and Gubeladze 2009] W. Bruns and J. Gubeladze, *Polytopes, rings, and K-theory*, Springer, Dordrecht, 2009. MR Zbl

[Bruns et al. 1999] W. Bruns, J. Gubeladze, M. Henk, A. Martin, and R. Weismantel, "A counterexample to an integer analogue of Carathéodory's theorem", *J. Reine Angew. Math.* **510** (1999), 179–185. MR Zbl

[Bruns et al. 2016] W. Bruns, J. Gubeladze, and M. Michałek, "Quantum jumps of normal polytopes", *Discrete Comput. Geom.* **56**:1 (2016), 181–215. MR Zbl

[Carter and Keller 1984] D. Carter and G. Keller, "Elementary expressions for unimodular matrices", *Comm. Algebra* **12**:3-4 (1984), 379–389. MR Zbl

[Cox et al. 2011] D. A. Cox, J. B. Little, and H. K. Schenck, *Toric varieties*, Graduate Studies in Mathematics **124**, American Mathematical Society, Providence, RI, 2011. MR Zbl

[Gawrilow and Joswig 1997] E. Gawrilow and M. Joswig, "Polymake", 1997, Available at http://polymake.org/doku.php.

[Miller and Sturmfels 2005] E. Miller and B. Sturmfels, *Combinatorial commutative algebra*, Graduate Texts in Mathematics **227**, Springer, New York, 2005. MR Zbl

[Quillen 1973] D. Quillen, "Higher algebraic K-theory, I", pp. 85–147 in *Algebraic K-theory, I: Higher K-theories* (Seattle, 1972), edited by H. Bass, Lecture Notes in Math. **341**, Springer, Berlin, 1973. MR Zbl

[Schrijver 1986] A. Schrijver, *Theory of linear and integer programming*, Wiley, Chichester, 1986. MR Zbl

[Sebő 1990] A. Sebő, "Hilbert bases, Carathéodory's theorem and combinatorial optimization", pp. 431–455 in *Proceedings of the 1st Integer Programming and Combinatorial Optimization Conference* (Waterloo, ON, 1990), edited by R. Kannan and W. R. Pulleyblank, Univ. Waterloo Press, 1990.

[Wachs 2007] M. L. Wachs, "Poset topology: tools and applications", pp. 497–615 in *Geometric combinatorics*, edited by E. Miller et al., IAS/Park City Math. Ser. **13**, American Mathematical Society, Providence, RI, 2007. MR Zbl

Received June 14, 2016. Revised May 29, 2017.

JOSEPH GUBELADZE
DEPARTMENT OF MATHEMATICS
SAN FRANCISCO STATE UNIVERSITY
SAN FRANCISCO, CA
UNITED STATES

soso@sfsu.edu

MATEUSZ MICHAŁEK
FREIE UNIVERSITÄT
BERLIN
GERMANY

POLISH ACADEMY OF SCIENCES
WARSAW
POLAND

and

MAX-PLANCK-INSTITUT FÜR MATHEMATIK IN DEN NATURWISSENSCHAFTEN
LEIPZIG
GERMANY

wajcha2@poczta.onet.pl

PACIFIC JOURNAL OF MATHEMATICS
Vol. 292, No. 1, 2018
dx.doi.org/10.2140/pjm.2018.292.117

DUAL MEAN MINKOWSKI MEASURES AND THE GRÜNBAUM CONJECTURE FOR AFFINE DIAMETERS

QI GUO AND GABOR TOTH

For a convex body K in a Euclidean vector space $\mathcal{X}$ of dimension n (≥ 2), we define two subarithmetic monotonic sequences $\{\sigma_{K,k}\}_{k\geq 1}$ and $\{\sigma^o_{K,k}\}_{k\geq 1}$ of functions on the interior of K. The k-th members are "mean Minkowski measures in dimension k" which are pointwise dual: $\sigma^o_{K,k}(z) = \sigma_{K^z,k}(z)$, where $z \in \operatorname{int} K$, and K^z is the dual (polar) of K with respect to z. They are measures of (anti-)symmetry of K in the following sense:

$$1 \leq \sigma_{K,k}(z), \sigma^o_{K,k}(z) \leq \frac{k+1}{2}.$$

The lower bound is attained if and only if K has a k-dimensional simplicial slice or simplicial projection. The upper bound is attained if and only if K is symmetric with respect to z. In 1953 Klee showed that the lower bound $\mathfrak{m}^*_K > n-1$ on the Minkowski measure of K implies that there are $n+1$ affine diameters meeting at a critical point $z^* \in K$. In 1963 Grünbaum conjectured the existence of such a point in the interior of any convex body (without any conditions). While this conjecture remains open (and difficult), as a byproduct of our study of the dual mean Minkowski measures, we show that

$$\frac{n}{\mathfrak{m}^*_K + 1} \leq \sigma^o_{K,n-1}(z^*)$$

always holds, and for sharp inequality Grünbaum's conjecture is valid.

1. Preliminaries and statement of results

Let $\mathcal{X}$ be an n-dimensional Euclidean vector space $(n \geq 2)$ with scalar product $\langle \cdot , \cdot \rangle$ and distance function d. We consider a *convex body* $K \subset \mathcal{X}$, a *compact* convex set in $\mathcal{X}$ with *nonempty interior*. Let ∂K denote the boundary of K. Given an interior point $z \in \operatorname{int} K$ we consider all the *chords* of K passing through z. For $x \in \partial K$, let $\lambda_K(x, z)$ denote the ratio into which z divides the chord of K starting

This work, including the stay of Toth in Suzhou, China, in January 2015, was supported by the NSF-China, No. 11671293. The authors declare that there is no conflict of interest.
MSC2010: primary 52A05, 52A20; secondary 52A41, 52B55.
Keywords: convex body, dual, Minkowski measure, affine diameter.

at x, passing through z, and ending up at the *opposite* $x^o \in \partial K$ of x (with respect to z). This defines the *distortion function* $\lambda_K : \partial K \times \operatorname{int} K \to \mathbb{R}$:

$$\lambda_K(x, z) = \frac{d(x, z)}{d(x^o, z)}, \quad x \in \partial K,\ z \in \operatorname{int} K.$$

For the involution of ∂K given by $x \mapsto x^o$ (with $(x^o)^o = x$), we have $\lambda_K(x^o, z) = 1/\lambda_K(x, z)$, $x \in \partial K$.

The (maximum) *Minkowski ratio* of K at z is defined as

$$\mathfrak{m}_K(z) = \sup_{x \in \partial K} \lambda_K(x, z) \geq 1.$$

(Due to compactness of K and continuity of the distortion function λ_K [Toth 2006, Lemma 1], the supremum is attained. This is also the case for all infima and suprema that we encounter in this paper.)

Let δK denote the (compact) space of all hyperplanes supporting K. (Associating to each $\mathcal{H} \in \delta K$ the unit normal that points inward K, say, gives rise to a topological equivalence of δK and the unit sphere $\mathcal{S} \subset \mathcal{X}$.) For $\mathcal{H} \in \delta K$, we define the ratio $\rho_K(\mathcal{H}, z) = d(\mathcal{H}, z)/d(\mathcal{H}^o, z)$, where $\mathcal{H}^o \in \delta K$ is the (unique) parallel opposite of $\mathcal{H}$ such that K is between $\mathcal{H}$ and $\mathcal{H}^o$. This gives rise to the function $\rho_K : \delta K \times \operatorname{int} K \to \mathbb{R}$. For the involution of δK given by $\mathcal{H} \mapsto \mathcal{H}^o$, $\mathcal{H} \in \delta K$, we have $\rho_K(\mathcal{H}^o, z) = 1/\rho_K(\mathcal{H}, z)$, $\mathcal{H} \in \delta K$.

It is well known that

$$\text{(1)} \qquad \mathfrak{m}_K(z) = \sup_{x \in \partial K} \lambda_K(x, z) = \sup_{\mathcal{H} \in \delta K} \rho_K(\mathcal{H}, z), \quad z \in \operatorname{int} K.$$

(See [Grünbaum 1963]. It is customary to define $\rho_K(\mathcal{H}, z)$ for a hyperplane $\mathcal{H}$ *containing* z as the ratio ≥ 1 that $\mathcal{H}$ divides the distance between the two supporting hyperplanes $\mathcal{H}', \mathcal{H}'' \in \delta K$ that are parallel to $\mathcal{H}$. In our study we need more control of the choice of the supporting hyperplane, henceforth we altered this definition accordingly. Since we are taking suprema these two definitions are equivalent.)

A technically more convenient reformulation of this second concept is as follows. Let $\operatorname{aff} = \operatorname{aff}(\mathcal{X})$ denote the $(n+1)$-dimensional vector space of affine functionals $f : \mathcal{X} \to \mathbb{R}$. We call $f \in \operatorname{aff}$ *normalized* for K if $f(K) = [0, 1]$, that is, the zero sets $\mathcal{H} = \{u \mid f(u) = 0\}$ and $\mathcal{H}^o = \{u \mid 1 - f(u) = 0\}$ are two parallel hyperplanes supporting and enclosing K. We let $\operatorname{aff}_K \subset \operatorname{aff}$ denote the (compact) subspace of affine functionals normalized for K. (Associating to each $f \in \operatorname{aff}_K$ the single zero set $\mathcal{H}$ as above gives rise to a topological equivalence of aff_K and δK. Indeed, any $\mathcal{H} \in \delta K$ and its opposite $\mathcal{H}^o$ uniquely define a normalized affine functional with the respective zero sets as above.) Note that aff_K has the obvious involution given by $f \mapsto 1 - f$, $f \in \operatorname{aff}_K$.

Using the notations above, (1) gives

$$\inf_{f\in\mathrm{aff}_K} f(z) = \inf_{f\in\mathrm{aff}_K} (1-f(z)) = \frac{1}{\sup_{\mathcal{H}\in\delta K}\rho_K(\mathcal{H},z)+1} = \frac{1}{\mathfrak{m}_K(z)+1}, \quad z\in\operatorname{int} K. \tag{2}$$

The two aspects of the Minkowski ratio above can be interpreted in terms of duality between the convex body K and its *dual* (also called *polar*) K^z with respect to the given interior point $z\in\operatorname{int}K$. (For the definition of the dual and its properties, see the next section. Note that when dealing with duality we will frequently use the *bipolar theorem* $(K^z)^z = K$ without explicit mention; [Eggleston 1958, Chapter 1.9] or [Schneider 2014, Theorem 1.6.1].)

First, as a technical tool, we will introduce and study the "musical equivalencies"

$$\flat = \flat_{K,z} : \partial K \to \mathrm{aff}_{K^z} \quad\text{and}\quad \sharp = \sharp_{K,z} : \mathrm{aff}_K \to \partial K^z.$$

(For simplicity, we will suppress the subscripts whenever no confusion arises. In Riemannian geometry the introduction of a Riemannian metric on a manifold gives rise to "musical isomorphisms" between the tangent bundle and its dual. Due to the descriptive nature of this concept and analogy we took the liberty of borrowing this term for our setting.) The musical equivalencies satisfy

$$(x^o)^\flat = 1 - x^\flat \quad\text{and}\quad (f^\sharp)^o = (1-f)^\sharp, \qquad x\in\partial K,\ f\in\mathrm{aff}\,K. \tag{3}$$

In addition, as the name suggests, they are inverses of each other:

$$\sharp_{K^z,z}\circ\flat_{K,z} = \mathrm{id}_{\partial K} \quad\text{and}\quad \flat_{K^z,z}\circ\sharp_{K,z} = \mathrm{id}_{\mathrm{aff}_K}. \tag{4}$$

These formulas (applied to the dual pair K and K^z) imply that the musical equivalencies are actually homeomorphisms of the respective spaces.

The following formulas show that the two aspects of Minkowski ratios are dual constructions applied to K and its dual K^z:

$$x^\flat(z) = \frac{1}{\lambda_K(x,z)+1}, \qquad x\in\partial K,\ z\in\operatorname{int}K, \tag{5}$$

and

$$f(z) = \frac{1}{\lambda_{K^z}(f^\sharp,z)+1}, \qquad f\in\mathrm{aff}_K,\ z\in\operatorname{int}K. \tag{6}$$

Taking the infima on the respective sets in (5)–(6) and using (2), we obtain

$$\inf_{x\in\partial K} x^\flat(z) = \frac{1}{\mathfrak{m}_K(z)+1} = \inf_{f\in\mathrm{aff}_K} f(z) = \frac{1}{\mathfrak{m}_{K^z}(z)+1}, \qquad z\in\operatorname{int}K.$$

This gives

$$\mathfrak{m}_K(z) = \mathfrak{m}_{K^z}(z), \qquad z\in\operatorname{int}K. \tag{7}$$

The *Minkowski measure* of K is defined as

$$\mathfrak{m}_K^* = \inf_{z \in \mathrm{int}\, K} \mathfrak{m}(z).$$

The set of interior points where this infimum is attained is called the *critical set*

$$K^* = \{z^* \in \mathrm{int}\, K \mid \mathfrak{m}_K(z^*) = \mathfrak{m}_K^*\}. \tag{8}$$

The critical set $K^* \subset K$ is compact and convex, and we have Klee's inequality

$$(1 \leq)\ \mathfrak{m}_K^* + \dim K^* \leq n$$

improving the classical Minkowski–Radon inequality (in which the dimension of the critical set is absent). (See [Klee 1953].) Clearly, $\mathfrak{m}_K^* = 1$ if and only if K is symmetric with respect to then unique regular point. It is also straightforward to show that the upper bound is attained for simplices. Conversely, Minkowski and Radon also proved that $\mathfrak{m}_K^* = n$ implies that K is a simplex.

For $z^* \in K^*$ critical, by (7), we have

$$\mathfrak{m}_K^* = \mathfrak{m}_K(z^*) = \mathfrak{m}_{K^{z^*}}(z^*) \geq \mathfrak{m}_{K^{z^*}}^*.$$

Whether equality holds, that is, whether $z^* \in K^*$ is also a critical point of the dual K^{z^*}, seems to be a difficult problem in general.

Recall that a chord $[x, x^o]$ of K is an *affine diameter* if there are parallel supporting hyperplanes $\mathcal{H}$ and $\mathcal{H}^o$ of K at the endpoints of the chord, that is $x \in \mathcal{H}$ and $x^o \in \mathcal{H}^o$. (For a general survey on affine diameters and related problems, see [Soltan 2005; Soltan and Nguyên 1988].) As discussed above, we describe these hyperplanes as the zero sets of a normalized affine functional $f \in \mathrm{aff}_K$, that is we have $\mathcal{H} = \{u \in \mathcal{X} \mid f(u) = 0\}$ and $\mathcal{H}^o = \{u \in \mathcal{X} \mid 1 - f(u) = 0\}$. Under the musical equivalencies, affine diameters of K correspond to affine diameters of K^z in the sense that if $[x, x^o]$ is an affine diameter of K with parallel supporting hyperplanes given by $f \in \mathrm{aff}_K$ then $[f^\sharp, (f^\sharp)^o] = [f^\sharp, (1-f)^\sharp]$ is an affine diameter of K^z with parallel supporting hyperplanes given by $x^\flat \in \mathrm{aff}_{K^z}$. (For the proof, see Section 2.)

We now introduce the sequence $\{\sigma_{K,k}\}_{k \geq 1}$ of *mean Minkowski measures* of K. (We give here a concise summary; for details, see [Toth 2004; 2006].) The *k-th measure* $\sigma_{K,k} : \mathrm{int}\, K \to \mathbb{R}$, $k \geq 1$, is a function on the interior of K defined as follows. First, a (*point*) *k-configuration of* K with respect to z is a *multiset* $\{x_0, \ldots, x_k\} \subset \partial K$ (with repetition allowed) such that the convex hull $[x_0, \ldots, x_k]$ contains z. (We use square brackets to indicate convex hull rather than "conv".) With this we define

$$\sigma_{K,k}(z) = \inf_{\{x_0,\ldots,x_k\} \in \mathfrak{C}_{K,k}(z)} \sum_{i=0}^{k} \frac{1}{\lambda_K(x_i, z) + 1}, \quad z \in \mathrm{int}\, K, \tag{9}$$

where $\mathfrak{C}_{K,k}(z)$ denotes the set of all k-configurations of K (with respect to z).

Algebraically, $\sigma_{K,k}$ is a "k-average" of the rescaled distortion, and, as we will see below, geometrically $\sigma_{K,k}(z)$ measures how far the k-dimensional slices of K across z are from a k-simplex.

A k-configuration $\{x_0, \dots, x_k\} \in \mathfrak{C}_{K,k}(z)$ at which the infimum in (9) is attained is called *minimizing*, or simply *minimal*. Since $\mathfrak{C}_{K,k}(z)$ inherits a compact topology from that of ∂K and the distortion is continuous, minimal configurations always exist. (As examples show, they are by no means unique.)

For $k = 1$, a 1-configuration of z is an opposite pair of points $\{x_0, x_1\} \subset \partial K$, $x_1 = x_0^o$. Since $\lambda_K(x_0^o, z) = 1/\lambda_K(x_0, z)$, we have $\sigma_{K,1}(z) = 1$, $z \in \operatorname{int} K$.

Since a (minimal) k-configuration can always be extended to a $(k+l)$-configuration by adding l copies of a boundary point at which the distortion $\lambda_K(\,\cdot\,, z)$ attains its maximum $\mathfrak{m}_K(z)$, we have *subarithmeticity*:

$$\sigma_{K,k+l}(z) \leq \sigma_{K,k}(z) + \frac{l}{\mathfrak{m}_K(z)+1}, \quad z \in \operatorname{int} K,\ k, l \geq 1. \tag{10}$$

By *Carathéodory's theorem*, for $k > n$, a k-configuration always contains an n-configuration. In addition, any subconfiguration of a minimal configuration is minimal, and, at the complementary configuration points, the distortion $\lambda_K(\,\cdot\,, z)$ attains its maximum $\mathfrak{m}_K(z)$. We see that the sequence $\{\sigma_{K,k}(z)\}_{k\geq 1}$ is *arithmetic with difference* $1/(\mathfrak{m}_K(z)+1)$ *from the n-th term onwards*.

For $1 \leq k \leq n$, we have

$$\sigma_{K,k}(z) = \inf_{z \in \mathcal{E} \subset \mathcal{X},\, \dim \mathcal{E} = k} \sigma_{K\cap\mathcal{E},k}(z), \quad z \in \operatorname{int} K, \tag{11}$$

where the infimum is over affine subspaces $\mathcal{E} \subset \mathcal{X}$ of dimension k which contain z. This holds because the affine span of any k-configuration $\{x_0, \dots, x_k\} \in \mathfrak{C}_{K,k}(z)$ is contained in an affine subspace $\mathcal{E}$ (containing z) of dimension k; therefore the infimum in (9) can first be taken for configurations that are contained in a specific $\mathcal{E}$, yielding $\sigma_{K\cap\mathcal{E},k}(z)$, and then for all k-dimensional affine subspaces $\mathcal{E}$ (which contain z) as in (11).

The mean Minkowski measures are *measures of symmetry* (or asymmetry for some authors) in the following sense:

$$1 \leq \sigma_{K,k}(z) \leq \frac{k+1}{2}, \quad z \in \operatorname{int} K. \tag{12}$$

(For measures of symmetry in general, see the seminal work of Grünbaum [1963].) Assuming $k \geq 2$, the upper bound is attained if and only if K is symmetric with respect to z. For the lower bound, if, for some $k \geq 1$, $\sigma_{K,k}(z) = 1$ at $z \in \operatorname{int} K$ then $k \leq n$, and K has a k-dimensional simplicial intersection across z, that is there exists a k-dimensional affine subspace $\mathcal{E} \subset \mathcal{X}$ such that $K \cap \mathcal{E}$ is a k-simplex (and consequently $\sigma_{K,k} = 1$ identically on $K \cap \mathcal{E}$).

The functions $\sigma_{K,k} : \operatorname{int} K \to \mathbb{R}$, $k \geq 1$, are continuous on $\operatorname{int} K$ and extend continuously to ∂K as

$$\lim_{d(z,\partial K)\to 0} \sigma_{K,k}(z) = 1. \tag{13}$$

The limiting behavior in (13) follows from subarithmeticity in (10) ($k = 1$ and $l = k-1$ and $\sigma_{K,1}(z) = 1$), and the lower estimate in (12). (For a different proof, see Theorem D/(b) in [Toth 2004].)

The sequence $\{\sigma_{K,k}(z)\}_{k\geq 1}$ is *superadditive*:

$$\sigma_{K,k+l}(z) - \sigma_{K,k}(z) \geq \sigma_{K,l}(z) - \sigma_{K,1}(z), \quad z \in \operatorname{int} K,\ k, l \geq 1. \tag{14}$$

In particular ($l = 1$), the sequence $\{\sigma_{K,k}(z)\}_{k\geq 1}$ is *monotonic*: $\sigma_{K,k}(z) \leq \sigma_{K,k+1}(z)$, $k \geq 1$.

Finally, note the obvious lower bound

$$\frac{k+1}{\mathfrak{m}_K(z)+1} \leq \sigma_{K,k}(z), \quad z \in \operatorname{int} K,\ k \geq 1. \tag{15}$$

The main technical tool of the present paper is the “dual construction”. Let $k \geq 1$. First, a *dual* (*or supporting*) *k-configuration* is a multiset $\{f_0, \ldots, f_n\} \subset \mathrm{aff}_K$ (repetition allowed) such that the intersection

$$\bigcap_{i=0}^{k} \{u \in \mathcal{X} \mid f_i(u) \leq 0\} = \varnothing. \tag{16}$$

With this, the *k-th dual mean Minkowski measure* $\sigma^o_{K,k} : \operatorname{int} K \to \mathbb{R}$ is defined as

$$\sigma^o_{K,k}(z) = \inf_{\{f_0,\ldots,f_k\}\in\mathfrak{C}^o_{K,k}} \sum_{i=0}^{k} f_i(z), \quad z \in \operatorname{int} K, \tag{17}$$

where $\mathfrak{C}^o_{K,k}$ denotes the set of all dual k-configurations of z.

The dual mean Minkowski measures have been introduced in [Guo and Toth 2016] along with detailed proofs of their arithmetic properties and extrema.

A dual k-configuration $\{f_0, \ldots, f_k\} \in \mathfrak{C}^o_{K,k}$ at which the infimum in (17) is attained is called *minimizing* or *minimal* for short. Since $\mathfrak{C}^o_{K,k}(z)$ inherits a compact topology from that of δK and the sum in (17) is continuous with respect to $(f_0, \ldots, f_k) \in (\mathrm{aff}_K)^{k+1}$, minimal configurations always exist.

For $k = 1$, a dual 1-configuration of any $z \in \operatorname{int} K$ is an opposite pair of affine functionals $\{f_0, f_1\} \subset \mathrm{aff}_K$, $f_1 = 1 - f_0$, and we have $\sigma^o_{K,1} = 1$ identically on $\operatorname{int} K$.

Note, by (2), the obvious lower bound

$$\frac{k+1}{\mathfrak{m}_K(z)+1} \leq \sigma^o_{K,k}(z), \quad z \in \operatorname{int} K,\ k \geq 1. \tag{18}$$

The first and most obvious property of the dual mean Minkowski measures is that, being infima of affine functions, $\sigma^o_{K,k} : \mathrm{int}\, K \to \mathbb{R}$, $k \geq 1$, are automatically *concave functions*. This is in striking contrast with the mean Minkowski measures $\sigma_{K,k} : \mathrm{int}\, K \to \mathbb{R}$, $k \geq 1$ which, albeit concave in dimension $n = 2$ (Theorem E in [Toth 2006]), for $n \geq 3$, they, in general, fail to satisfy any concavity properties. The following example illustrates this point.

Example. Let K be an n-cube, $n \geq 3$. Then the function $\sigma_{K,n-1}$ is identically 1 on the complement of the (open) cross-polytope K_0 inscribed in K (since the vertex figures provide $n-1$ dimensional simplicial intersections), but in the interior of K_0 we have $\sigma_{K,n-1} > 1$. Thus, $\sigma_{K,n-1}$ is not concave. A somewhat more involved argument shows that $\sigma_{K,n}$ is also nonconcave. (For a much more general result, see [Toth 2009, Theorem D].) As a byproduct, we see that, for the n-cube K, $n \geq 3$, $\sigma_{K,n}$ and $\sigma^o_{K,n}$ are different functions.

The following pointwise duality is the cornerstone of our study:

Theorem 1. *Let $K \subset \mathcal{X}$ be a convex body, and $z \in \mathrm{int}\, K$. For $k \geq 1$, we have*

$$\sigma^o_{K,k}(z) = \sigma_{K^z,k}(z), \tag{19}$$

where K^z is the dual of K with respect to z.

Remark. It is important to note that on the right-hand side of (19) the mean Minkowski measure has a double dependency on the point z; not only in the argument but also in forming the dual K^z. For this reason duality can only be used pointwise.

The crux of the proof of Theorem 1 (Section 3) is the equivalence

$$\{f_0, \dots, f_k\} \in \mathfrak{C}^o_{K,k} \iff \{f_0^\sharp, \dots, f_k^\sharp\} \in \mathfrak{C}_{K^z,k}(z). \tag{20}$$

As a byproduct of the proof, it will also follow that, under this equivalence, minimal configurations correspond to each other.

Pointwise duality allows the properties of the mean Minkowski measures to carry over to the dual. Replacing K with K^z in (10) and using (7) and (19), we have *subarithmeticity*:

$$\sigma^o_{K,k+l}(z) \leq \sigma^o_{K,k}(z) + \frac{l}{\mathfrak{m}_K(z)+1}, \quad z \in \mathrm{int}\, K,\ k, l \geq 1. \tag{21}$$

In addition, the sequence $\{\sigma^o_{K,k}(z)\}_{k\geq 1}$ is arithmetic with difference $1/(\mathfrak{m}_K(z)+1)$ from the n-th term onwards.

Remark. It is worth noting that the direct proof of arithmeticity (without the use of duality) beyond the dimension is an application of (the contrapositive of) *Helly's theorem* (instead of Carathéodory's): For $k > n$, any dual k-configuration (characterized by (16)) contains an n-configuration.

To state the dual version of (11), for $1 \le k \le n$, we denote by $\mathfrak{P}_k = \mathfrak{P}_{\mathcal{X},k}$ the space of all orthogonal projections $\Pi : \mathcal{X} \to \mathcal{X}$ onto k-dimensional affine subspaces $\Pi(\mathcal{X}) = \mathcal{E} \subset \mathcal{X}$. We then have

$$\sigma^o_{K,k}(z) = \inf_{\Pi \in \mathfrak{P}_k} \sigma^o_{\Pi(K),k}(\Pi(z)), \quad z \in \operatorname{int} K. \tag{22}$$

(In the infimum $\Pi(z)$ can be replaced by z if we require $z \in \Pi(\mathcal{X}) = \mathcal{E}$.)

By duality, the bounds in (12) stay the same for the dual mean Minkowski measures. To characterize the convex bodies for which the lower bound is attained is somewhat more complex (to be expounded in Section 3). We summarize these concisely in the following:

Theorem 2. *Let $K \subset \mathcal{X}$ be a convex body. For $k \ge 1$, we have*

$$1 \le \sigma^o_{K,k}(z) \le \frac{k+1}{2}, \quad z \in \operatorname{int} K. \tag{23}$$

Assuming $k \ge 2$, the upper bound in (23) is attained if and only if K is symmetric with respect to z. If, for some $k \ge 1$, $\sigma^o_{K,k}(z) = 1$ at $z \in \operatorname{int} K$ then $\sigma^o_{K,k} = 1$ identically on $\operatorname{int} K$; we have $k \le n$, and K has an orthogonal projection to a k-simplex.

The functions $\sigma^o_{K,k} : \operatorname{int} K \to \mathbb{R}$, $k \ge 1$, are continuous on $\operatorname{int} K$. As in the nondual case, by the lower bound in (23) along with subarithmeticity ($k = 1$ and $l = k-1$ in (21) with $\sigma^o_{K,1} = 1$), we have continuity up to the boundary via

$$\lim_{d(z,\partial K) \to 0} \sigma^o_{K,k}(z) = 1. \tag{24}$$

Example. Let K be a tetrahedron ($n = 3$) truncated near all four vertices (by vertex figures, say). Then $\sigma_{K,2} = 1$ identically as K has triangular intersections through any of its interior points. On the other hand, $\sigma^o_{K,2} > 1$ everywhere since K has no triangular projection. We see once again that, in general, the function $\sigma_{K,k}$ and its dual $\sigma^o_{K,k}$ are different.

Next, again by duality, we note superadditivity

$$\sigma^o_{K,k+l}(z) - \sigma^o_{K,k}(z) \ge \sigma^o_{K,l}(z) - \sigma^o_{K,1}(z), \quad z \in \operatorname{int} K,\ k, l \ge 1,$$

and, as a consequence, monotonicity: $\sigma^o_{K,k}(z) \le \sigma^o_{K,k+1}(z)$, $k \ge 1$.

Most of the properties of the dual mean Minkowski measures discussed above are consequences of the pointwise duality asserted by Theorem 2. They have, however, additional and more refined properties showing that, as measures, they are better adapted convex bodies than their nondual counterparts. Our next result asserts the striking fact that the n-th dual mean Minkowski measure can be *explicitly* calculated at the critical points of a convex body.

Theorem 3. *Let $K \subset \mathcal{X}$ be a convex body and $K^* \subset K$ its critical set. For any critical point $z^* \in K^*$, we have*

$$\sigma^o_{K,n}(z^*) = \frac{n+1}{\mathfrak{m}^*_K + 1}. \tag{25}$$

The proof of Theorem 3 (Section 3) relies heavily on Klee's delicate analysis of the critical set and the proof of his improved Minkowski–Radon inequality.

Remark. It is natural to ask if (25) holds for the n-th (nondual) mean Minkowski measure $\sigma_{K,n}$. While this remains unsolved, it seems to depend on whether a critical point $z^* \in K^*$ is also a critical point for the dual $(K)^{z^*}$ or not. For the class of *convex bodies of constant width* the answer is affirmative as follows. (For a general reference on convex bodies of constant width, see [Chakerian and Groemer 1983].) For a convex body K of constant width d, the critical set K^* is a singleton, and the unique critical point z^* is the common center of the circumcircle $\mathcal{S}_{R_K}(z^*)$ and the incircle $\mathcal{S}_{r_K}(z^*)$ with circumradius R_K and inradius r_K. The latter can be expressed in terms of the Minkowski measure as

$$R_K = \frac{\mathfrak{m}^*_K}{\mathfrak{m}^*_K + 1} d \quad \text{and} \quad r_K = \frac{1}{\mathfrak{m}^*_K + 1} d.$$

In particular, we have $R_K + r_K = d$ and

$$\mathfrak{m}^*_K = \frac{R_K}{r_K}.$$

(For these results, see [Jin and Guo 2012], and (for some) also [Bonnesen and Fenchel 1934, §63] and [Eggleston 1958, Theorem 53 and its corollary, p. 125].) Another classical fact is that $z^* \in [\partial K \cap \mathcal{S}_{R_K}(z^*)]$, so that, by Carathéodory's theorem, z^* is in the convex hull of at most $n+1$ boundary points of K on the circumcircle $\mathcal{S}_{R_K}(z^*)$. It follows that the circumcircle contains an n-configuration of z^*. Thus, for a convex body K of constant width, equality holds in (25) for the (nondual) mean Minkowski measure $\sigma_{K,n}$.

For $k = n$, an n-configuration $\{x_0, \ldots, x_n\} \in \mathfrak{C}_{K,n}(z)$, $z \in \operatorname{int} K$, is called *simplicial* if $[x_0, \ldots, x_n]$ is an n-simplex with z is in its interior. We let $\Delta_K(z) \subset \mathfrak{C}_{K,n}(z)$ denote the (noncompact) space of all simplicial configurations. (The concept of simplicial k-configurations, $1 \le k \le n$, can be defined analogously using relative interiors, but we will not need this here.) In (9) the infimum can be restricted to $\Delta_K(z)$, but a minimizing sequence of simplicial configurations may not (sub)converge. If degeneracy at the infima does not occur, that is all minimal n-configurations are simplicial then we call $z \in \operatorname{int} K$ a *regular point*. The set of regular points is denoted by $\mathcal{R}_K \subset \operatorname{int} K$.

We now turn to the dual construction (Section 4). A dual n-configuration $\{f_0, \dots, f_n\} \in \mathfrak{C}^o_{K,n}(z)$ is called *simplicial* if the intersection

$$\bigcap_{i=0}^{n} \{u \in \mathcal{X} \mid f_i(u) \geq 0\}$$

is an n-simplex. Using musical equivalences, this is equivalent to $\{f_0^\sharp, \dots, f_n^\sharp\} \in \mathfrak{C}_{K^z,n}(z)$ being simplicial. We let $\Delta^o_K(z) \subset \mathfrak{C}^o_{K,n}$ denote the space of all simplicial dual configurations. As before, in (17) the infimum can be restricted to $\Delta^o_K(z)$, but a minimizing sequence of simplicial dual configurations may not (sub)converge. If all minimal dual n-configurations are simplicial then we call $z \in \operatorname{int} K$ a *dual regular point*. The set of dual regular points is denoted by $\mathcal{R}^o_K \subset \operatorname{int} K$.

The concept of regularity meshes well with duality, and Theorem 2 gives

$$z \in \mathcal{R}^o_K \iff z \in \mathcal{R}_{K^z}, \qquad z \in \operatorname{int} K. \tag{26}$$

The significance of these concepts lie in the fact that at any regular or dual regular points $n+1$ affine diameters meet. This is closely related to *Grünbaum's conjecture*: Any convex body K has an interior point z at which $n+1$ affine diameters meet. (See [Grünbaum 1963, 6.4.3, p. 254].)

A study of subconvergence of minimizing sequences then gives the following consequence of Theorem 3:

Theorem 4. *Let $z^* \in K^* \subset K$ be as in Theorem 3. Then we have*

$$\frac{n}{\mathfrak{m}^*_K + 1} \leq \sigma^o_{K,n-1}(z^*). \tag{27}$$

If strict inequality holds then $z^* \in \mathcal{R}^o_K$ and the Grünbaum conjecture is valid for K: There are $n+1$ affine diameters that meet at z^*.

Remarks.

(A) Klee [1953] proved Grünbaum's conjecture under the condition $\mathfrak{m}_K(z^*) > n-1$. This is much more restrictive than (27) since $\sigma_{K,n-1}(z^*) \geq 1$ automatically holds.

(B) The geometric interpretation of the right-hand side in (27) follows from (22): $\sigma^o_{K,n-1}(z^*)$ is the infimum of $\sigma^o_{\Pi(K),n-1}(\Pi(z^*))$ for all projections $\Pi \in \mathfrak{P}_{K,n-1}$ of K to hyperplanes in $\mathcal{X}$.

(C) Equality holds in (27) if K is symmetric (necessarily with center z^*). In this case the Grünbaum conjecture obviously holds.

(D) Let K be a convex body of constant width. By the remark after Theorem 3, Theorem 4 holds for the (nondual) mean Minkowski measure. Whether the respective inequality is strict or not depends on the (unique) critical point

$z^* \in K^*$ being regular or not. This, in turn, depends on whether z^* is in the convex hull of boundary points of K contained in a (proper) great subsphere of the circumsphere $\mathcal{S}_{R_K}(z^*)$. Note that the construction of raising the dimension for convex bodies of constant width shows that nonregular points can well occur; see [Lachand-Robert and Oudet 2007, Theorem 6].

Example. Let $K = \{(a, b) \in \mathbb{R}^2 \mid a^2 + b^2 \leq 1,\ b \geq 0\}$ be the unit half-disk. A simple computation shows that $\mathfrak{m}_K$ attains its minimum at the (unique) critical point $z^* = (0, \sqrt{2} - 1)$. (See also [Hammer 1951].) We thus have $\mathfrak{m}^*_K = \sqrt{2}$, and, by Theorem 3, $\sigma^o_{K,2}(z^*) = 3/(\sqrt{2} + 1)$. Since $\sigma^o_{K,1} = 1$, in (27) strict inequality holds, in particular, $z^* \in \mathcal{R}^o_K$. (Note that the centroid $g(K) = (0, 4/3\pi)$ of K is different form z^*.) We claim that $\mathcal{R}^o_K = \operatorname{int}\Delta$, where $\Delta = [x_0, x_-, x_+]$ is the triangle with vertices $x_0 = (0, 1)$ and $x_\pm = (\pm 1, 0)$. Given $z = (a, b) \in \operatorname{int} K$ there may be at most three affine diameters passing through (a, b), those that also pass through x_0, x_-, and x_+. This immediately gives $\mathcal{R}^o_K \subset \operatorname{int}\Delta$. For equality, let $z = (a, b) \in \operatorname{int}\Delta$ with $a \geq 0$ (by symmetry). Define $f_0 \in \operatorname{aff}_K$ by its zero set the first axis, and let $f_\pm \in \operatorname{aff}_K$ have its zero set the tangent line to the unit circle at the opposite $x^o_\pm$ with respect to z. A simple comparison of ratios shows that $f_-(z) + f_+(z) < 1$ and $f_0(z) + f_+(z) < 1$. On the other hand, we have $1/(\mathfrak{m}_K(z) + 1) = \min(f_0(z), f_-(z))$, and we obtain

$$f_0(z) + f_-(z) + f_+(z) < 1 + \frac{1}{\mathfrak{m}_K(z) + 1}.$$

Since $\{f_0, f_-, f_+\} \in \mathfrak{C}^o_K(z)$, a dual 2-configuration, we see that z is a dual regular point. The claim follows.

A simple consideration of the affine coordinates associated to a simplex shows that the interior of a simplex consists of dual regular points only. (See Section 3.) In the other extreme it is natural to expect that the interior of a symmetric convex body does not have any dual regular points. This is indeed the case asserted by the following:

Theorem 5. *In a symmetric body K there are no dual regular points.*

Remark. The same holds for (nondual) regular points; see [Toth 2009, Theorem A]. This, however, does not imply Theorem 5 due to the fact that the duality in Theorem 2 is only pointwise.

Example. Let $K = \Delta \times I \subset \mathbb{R}^3$ be a prism, where $\Delta \subset \mathbb{R}^2$ is a triangle and $I \subset \mathbb{R}$ is a closed interval. Then there are no dual regular points in the interior of K. This shows that the converse of Theorem 5 is not true. In addition, since $\mathfrak{m}^*_\Delta = 2$, we have $\mathfrak{m}^*_K = 2$, and $\sigma_{K,2} = 1$ identically (since K has the triangular projection Δ). We see that equality holds in (27). On the other hand, through any interior points of K there are four affine diameters so that Grünbaum's conjecture holds for K.

This shows that in trying to remove the condition in (27) one needs to consider nonsymmetric convex bodies with no dual regular points. (As was pointed out by Hammer and Sobczyk [1951], K is a convex body with 1-dimensional critical set K^*. In addition, for K equality holds in Klee's inequality showing that it is sharp.)

2. Duality via the musical equivalencies

We define the dual of a convex body $K \subset \mathcal{X}$ with respect to an interior point $z \in \operatorname{int} K$ as follows.

First, let $K_0 \subset \mathcal{X}$ be a convex body with $0 \in K_0$, the origin in $\mathcal{X}$. We define the dual of K_0 with respect to 0 as

$$K_0^0 = \{u \in \mathcal{X} \mid \sup_{x \in K_0} \langle x, u\rangle \leq 1\}. \tag{28}$$

Clearly, $0 \in \operatorname{int} K_0$, and by the bipolar theorem, we have $(K_0^0)^0 = K_0$.

The general case ($z \in \operatorname{int} K$) is reduced to this by employing translations $T_v : \mathcal{X} \to \mathcal{X}$, $v \in \mathcal{X}$, where $T_v(u) = u + v$, $u \in \mathcal{X}$.

We first let $K_0 = (T_z)^{-1}(K)$ (so that the point $z \in \operatorname{int} K$ is moved to the origin $0 \in \operatorname{int} K_0$), and then define

$$K^z = T_z(K_0^0), \quad K_0 = (T_z)^{-1}(K). \tag{29}$$

Clearly, $z \in \operatorname{int} K^z$, and, by the above, we also have $(K^z)^z = K$.

The translations $T_v : \mathcal{X} \to \mathcal{X}$, $v \in \mathcal{X}$, act on the space of affine functionals $\operatorname{aff} = \operatorname{aff}(\mathcal{X})$ by $T_v^o : \operatorname{aff} \to \operatorname{aff}$, $v \in \mathcal{X}$, defined by $T_v{}^o(f) = f \circ T_v^{-1}$, $f \in \operatorname{aff}$. Using the notations above, for $z \in \operatorname{int} K$, the linear map T_z^o restricts to $T_z^o : \operatorname{aff}_{K_0} \to \operatorname{aff}_K$, $K_0 = T_z^{-1}(K)$, between the normalized affine functionals of the respective convex bodies. (Indeed, for $f_0 \in \operatorname{aff} K_0$, we have $f_0(K_0) = T_z^o(f_0)(K) = [0, 1]$.) Since, by (29), $K_0^0 = T_z^{-1}(K^z)$, we also have the restriction $T_z^o : \operatorname{aff}_{K_0^0} \to \operatorname{aff}_{K^z}$.

In this spirit, the definition of the musical equivalencies

$$\flat_{K,z} : \partial K \to \operatorname{aff}_{K^z} \quad \text{and} \quad \sharp_{K,z} : \operatorname{aff}_K \to \partial K^z$$

can be reduced to

$$\flat_{K_0,0} : \partial K_0 \to \operatorname{aff}_{K_0^0} \quad \text{and} \quad \sharp_{K_0,0} : \operatorname{aff}_{K_0} \to \partial K_0^0$$

by the formulas

$$\flat_{K,z} = T_z^o \circ \flat_{K_0,0} \circ T_z^{-1} \quad \text{and} \quad \sharp_{K,z} = T_z \circ \sharp_{K_0,0} \circ (T_z^{-1})^o. \tag{30}$$

It remains to define the musical equivalencies for K_0 with respect to $0 \in \operatorname{int} K_0$ satisfying (3)–(6). For simplicity, we now suppress the subscript 0 and set $K = K_0$ with $0 \in \operatorname{int} K$.

For $x \in \partial K$, we let $x^\flat : \mathcal{X} \to \mathbb{R}$ be the affine functional given by

$$x^\flat(u) = \frac{1}{\lambda_K(x,0)+1}(1 - \langle x, u\rangle), \quad u \in \mathcal{X}. \tag{31}$$

Evaluating this at the origin 0, (5) immediately follows.

The opposite of $x \in \partial K$ (with respect to the origin 0) is $x^o = -x/\lambda_K(x,0)$. Replacing x by x^o in (31), a simple computation gives the first formula in (3). Now a quick look at the definition of the dual K^0 in (28) shows that $x^\flat$ is normalized for K^0. We conclude that the musical map $\flat : \partial K \to \mathrm{aff}_{K^0}$ is well-defined.

For $f \in \mathrm{aff}_K$, we write $f(u) = \langle A, u\rangle + a$, $A \in \mathcal{X}$ and $a \in (0,1)$ (since f is normalized). We then define

$$f^\sharp = -\frac{A}{a}. \tag{32}$$

Since f is normalized, (28) shows that this point is on the boundary of the dual K^0. Once again, we obtain that the musical map $\sharp : \mathrm{aff}_K \to \partial K^0$ is well-defined.

Using (28) and (32) with $1-f$ in place of f, we obtain

$$(1-f)^\sharp = \frac{A}{1-a} = (f^\sharp)^o,$$

and the second formula in (3) follows. Since $-A/a$ and $A/(1-a)$ are opposites in K^0, as a byproduct, we obtain (6).

Finally, it remains to show that the musical equivalencies are inverses of each other, that is (4) holds. Indeed, combining (31) and (32), we obviously have $(x^\flat)^\sharp = x$, $x \in \partial K$, and the first relation in (4) follows. For the second, letting $f(u) = \langle A, u\rangle + a$ as above and using (6), we have $(f^\sharp)^\flat(u) = a(1 + \langle A, u\rangle/a) = f(u)$, $u \in \mathcal{X}$. The second relation in (4) also follows.

As a final preparatory step, as stated in the previous section, we need to work out the dual of an affine diameter. Let $[x, x^o] \subset K$ be an affine diameter with parallel supporting hyperplanes $\mathcal{H}, \mathcal{H}^o \in \delta K$ at both ends, that is $x \in \mathcal{H}$ and $x^o \in \mathcal{H}^o$. As above, we let $f \in \mathrm{aff}_K$ be the normalized affine functional with zero sets $\mathcal{H} = \{u \mid f(u) = 0\}$ and $\mathcal{H}^o = \{u \mid 1 - f(u) = 0\}$. We have $f(x) = 0$ and $f(x^o) = 1$. Letting $0 = z$ and $f(u) = \langle A, u\rangle + a$, $u \in \mathcal{X}$, as above, we have

$$x^\flat(f^\sharp) = \frac{1}{\lambda_K(x,0)+1}\left(1 - \left\langle x, -\frac{A}{a}\right\rangle\right) = \frac{1}{a(\lambda_K(x,0)+1)} f(x) = 0,$$

and

$$\begin{aligned} x^\flat((f^\sharp)^o) &= \frac{1}{\lambda_K(x,0)+1}\left(1 - \left\langle x, \frac{A}{1-a}\right\rangle\right) \\ &= \frac{1}{(1-a)(\lambda_K(x,0)+1)}(1 - a - \langle x, A\rangle) = 1, \end{aligned}$$

since

$$f(x^o) = \langle A, x^o \rangle + a = -\frac{1}{\lambda_K(x,0)} \langle A, x \rangle + a = 1.$$

We see that $[f^\sharp, (f^\sharp)^o]$ is an affine diameter of the dual K^0 with parallel supporting hyperplanes $x^\flat$, $(x^o)^\flat \in \delta K^0$ at the endpoints.

We conclude that the dual of an affine diameter configuration is also an affine diameter configuration.

3. Proofs of Theorems 1–3

Proof of Theorem 1. We will show that $\sigma_{K,k}(z) = \sigma^o_{K^z,k}(z)$. Since $(K^z)^z = K$, this will imply the theorem.

We first claim that, for any $\{x_0, \dots, x_k\} \subset \partial K$, we have

$$z \in [x_0, \dots, x_k] \quad \Longleftrightarrow \quad \bigcap_{i=0}^{k} \{u \in \mathcal{X} \mid x_i^\flat(u) \le 0\} = \varnothing, \tag{33}$$

where $\flat = \flat_{K,z} : \partial K \to \mathrm{aff}_{K^z}$ is the musical equivalence.

Without loss of generality, we may set $z = 0 \in \operatorname{int} K$, the origin.

First, assume that $0 \in [x_0, \dots, x_k]$, that is $\sum_{i=0}^k \lambda_i x_i = 0$ with $\sum_{i=0}^k \lambda_i = 1$, $\lambda_i \in [0,1]$, $i = 0, \dots, k$. Assume there exists $u \in \mathcal{X}$ such that $x_i^\flat(u) \le 0$, $i = 0, \dots, k$. By (31), this means that $\langle x_i, u \rangle \ge 1$, $i = 0, \dots, k$. Summing up, we obtain

$$\sum_{i=0}^k \lambda_i \langle x_i, u \rangle = \left\langle \sum_{i=0}^k \lambda_i x_i, u \right\rangle = 0 \ge \sum_{i=0}^k \lambda_i = 1,$$

a contradiction.

Conversely, assume that $0 \notin [x_0, \dots, x_k]$ so that 0 and the convex hull $[x_0, \dots, x_k]$ can be (strictly) separated by a hyperplane $\mathcal{H} \subset \mathcal{X}$. A unit normal $N \in \mathcal{X}$ of $\mathcal{H}$ then satisfies $\langle x_i, N \rangle > 0$, $i = 0, \dots, k$. For $t > 0$ large enough, we then have $\langle x_i, tN \rangle \ge 1$, $i = 0, \dots, k$. Thus, tN belongs to the intersection $\bigcap_{i=0}^k \{u \in \mathcal{X} \mid x_i^\flat(u) \le 0\}$. The converse follows.

The claim just proved can be reformulated as

$$\{x_0, \dots, x_k\} \in \mathfrak{C}_{K,k}(z) \quad \Longleftrightarrow \quad \{x_0^\flat, \dots, x_k^\flat\} \in \mathfrak{C}^o_{K^z,k}.$$

Using (5), we now calculate

$$\sigma_{K,k}(0) = \inf_{\{x_0,\dots,x_k\} \in \mathfrak{C}_{K,k}(0)} \sum_{i=0}^k \frac{1}{\lambda_K(x_i, 0) + 1} = \inf_{\{x_0^\flat,\dots,x_k^\flat\} \in \mathfrak{C}^o_{K^z,k}(0)} \sum_{i=0}^k x_i^\flat(0)$$
$$= \inf_{\{f_0,\dots,f_k\} \in \mathfrak{C}^o_{K^z,k}(0)} \sum_{i=0}^k f_i(0) = \sigma^o_{K^z,k}(0). \qquad \square$$

Remark. Dually, for $\{f_0, \dots, f_k\} \subset \mathrm{aff}_K$, we also have

$$\bigcap_{i=0}^{k} \{u \in \mathcal{X} \mid f_i(u) \leq 0\} = \varnothing \quad \Longleftrightarrow \quad z \in [f_0^\sharp, \dots, f_k^\sharp].$$

This is the same as the equivalency asserted in (20). As a byproduct of the computation above we also see that under the musical equivalencies minimal configurations correspond to each other.

We turn to the proof of (22). Given a dual k-configuration $\{f_0, \dots, f_k\} \in \mathfrak{C}^o_{K,k}$, let $\mathcal{E} \subset \mathcal{X}$ be a k-dimensional affine subspace containing the duals $f_0^\sharp, \dots, f_k^\sharp \in \mathcal{X}$ (and, by (20), also z). We have $\{f_0|_{\mathcal{E}}, \dots, f_k|_{\mathcal{E}}\} \in \mathfrak{C}^o_{\Pi(K),k}$, where $\Pi \in \mathfrak{P}_k$ is the orthogonal projection of $\mathcal{X}$ to $\mathcal{E}$. The affine functionals f_i, $i = 0, \dots, k$, are constant along (the fibers of) Π, and we also have $\sum_{i=0}^k f_i(z) = \sum_{i=0}^k (f_i|_{\mathcal{E}})(z)$. We conclude that the infimum for $\sigma^o_{K,k}(z)$ in (22) can first be taken for dual k-configurations in $\mathfrak{C}^o_{\Pi(K),k}(\Pi(z))$ for a given $\Pi \in \mathfrak{P}_k$, thus yielding $\sigma^o_{\Pi(K),k}(\Pi(z))$, and finally followed by the infimum for all $\Pi \in \mathfrak{P}_k$. The claim follows.

Proof of Theorem 2. As noted previously, the bounds in (23) follow by duality via Theorem 1.

We now consider the upper bound in (23). Let $k \geq 2$, and assume that $\sigma^o_{K,k}(z) = (k+1)/2$. Dualizing, again by Theorem 1, we have $\sigma_{K^z,k}(z) = (k+1)/2$. Hence, K^z is symmetric with respect to z. Since duality (with respect to the center) preserves symmetry, we obtain that $K = (K^z)^z$ is symmetric with respect to z.

It remains to consider the lower bound in (23). Assume that, for some $k \geq 1$, we have $\sigma^o_{K,k}(z) = 1$ at an interior point $z \in \mathrm{int}\, K$. Since $\sigma^o_{K,k}$ is a concave function on $\mathrm{int}\, K$ and, by (24), it assumes the value 1 on the boundary, we see that $\sigma^o_{K,k} = 1$ identically on K.

If $k > n$ then, by arithmeticity and (23), we have

$$1 = \sigma^o_{K,k}(z) = \sigma^o_{K,n}(z) + \frac{k-n}{\mathfrak{m}_K(z)+1} \geq 1 + \frac{k-n}{\mathfrak{m}_K(z)+1} > 1.$$

This is a contradiction. Thus $k \leq n$. (Alternatively, again by duality, $\sigma^o_{K,k}(z) = \sigma_{K^z,k}(z) = 1$ so that $k \leq n$.)

For the last statement, let the infimum in (22) be attained at an orthogonal projection $\Pi \in \mathfrak{P}_k$ (onto a k-dimensional affine subspace), so that $\sigma^o_{\Pi(K),k}(\Pi(z)) = 1$. As before, $\sigma^o_{\Pi(K),k} = 1$ identically on $\Pi(K)$. Let z^* be a critical point of $\Pi(K)$. By the obvious lower bound in (18) applied to the k-dimensional convex body $\Pi(K)$ (and z^*), we have

$$\frac{k+1}{\mathfrak{m}^*_{\Pi(K)}+1} \leq \sigma^o_{\Pi(K),k}(z^*) = 1.$$

This gives $k \le \mathfrak{m}_{\Pi(K)}(z^*)$. By the Minkowski–Radon inequality, $\mathfrak{m}^*_{\Pi(K)} \le k$, so that equality holds and $\Pi(K)$ is a k-simplex. $\square$

Example. An n-simplex $\Delta = [x_0, \dots, x_n]$ with vertices $x_0, \dots, x_n \in \mathcal{X}$ possesses a unique minimal dual n-configuration for any interior point, the *affine coordinate system* $\{f_0, \dots, f_n\} \subset \mathrm{aff}_\Delta$ associated to Δ. (For $i = 0, \dots, n$, $f_i \in \mathrm{aff}_\Delta$ is the normalized affine functional that vanishes on the i-th face $[x_0, \dots, \widehat{x_i}, \dots, x_n]$ (opposite to the vertex x_i), and $f_i(x_i) = 1$.) For $z \in \mathrm{int}\,\Delta$ with $z = \sum_{i=0}^n \lambda_i x_i$, $\sum_{i=0}^n \lambda_i = 1$, $\lambda_i \in (0, 1)$, we have $f_i(z) = \lambda_i$, $i = 0, \dots, n$. Since (16) obviously holds, we have $\sigma^o_{\Delta,n}(z) \le \sum_{i=0}^n f_i(z) = \sum_{i=0}^n \lambda_i = 1$. By (23), equality must hold. We see that $\{f_0, \dots, f_n\} \in \mathfrak{C}^o_{\Delta,n}(z)$ is the (unique) minimal dual n-configuration for all $z \in \mathrm{int}\,\Delta$. As a byproduct, we see that all interior points of an n-simplex are dual regular, that is $\mathcal{R}^o_\Delta = \mathrm{int}\,\Delta$. (The same holds for (nondual) regular points.)

Remark. The previous example can be used to show directly that if $\sigma^o_{K,n}(z) = 1$ then K is an n-simplex. This gives an alternative proof of the last part of Theorem 2 (for $\Pi(K)$ instead of K) without the recourse of the Minkowski–Radon theorem.

Assume $\sigma^o_{K,n}(z) = 1$ for some $z \in \mathrm{int}\,K$. First, any minimal dual n-configuration of z must be simplicial. Indeed, otherwise a minimal dual n-configuration would contain a proper subconfiguration, and we would have arithmeticity: $1 = \sigma^o_{K,n} = \sigma^o_{K,n-1} + 1/(\mathfrak{m}_K(z) + 1) > 1$, a contradiction. Second, let $\{f_0, \dots, f_n\} \in \Delta^o_K(z)$ be a minimal simplicial dual configuration. The corresponding n-simplex $\Delta = \bigcap_{i=0}^n \{u \in \mathcal{X} \mid f_i(u) \ge 0\}$ contains K. For each $i = 0, \dots, n$, let $\tilde{f}_i \in \mathrm{aff}_\Delta$ be the normalized affine functional such that $\{u \in \mathcal{X} \mid f_i(u) = 0\} = \{u \in \mathcal{X} \mid \tilde{f}_i(u) = 0\}$. Now, assume that K is *not* a simplex. Then $f_i(z) < \tilde{f}_i(z)$ for some $i = 0, \dots, n$. We then have $1 = \sigma^o_{K,n}(z) = \sum_{i=0}^n f_i(z) < \sum_{i=0}^n \tilde{f}_i(z) = \sigma_{\Delta,n}(z) = 1$, where the last two equalities follow from the example immediately above. This is a contradiction, so that K must be an n-simplex.

Proof of Theorem 3. We first introduce some notation. We define

$$\mathcal{M}(z) = \{x \in \partial K \mid \lambda_K(x, z) = \mathfrak{m}_K(z)\}, \quad z \in \mathrm{int}\,K,$$

where $\mathfrak{m}_K : \mathrm{int}\,K \to \mathbb{R}$ is the maximal Minkowski ratio. Clearly, $\mathcal{M}(z) \subset \partial K$ is compact, and for every $x \in \mathcal{M}(z)$, the chord $[x, x^o]$ of K is an affine diameter. (This is an elementary fact; also noted in [Klee 1953, 3.2].)

We now turn to the proof, in which we will use several results of Klee [1953]. Let $\mathcal{N}(z^*) = \mathcal{M}(z^*)^o \subset \partial K$ be the opposite set of $\mathcal{M}(z^*) \subset \partial K$ with respect to z^*. Denote by $\mathfrak{G}$ the family of closed half-spaces that intersect $\mathcal{N}(z^*)$ but disjoint from $\mathrm{int}\,K$. Clearly, for each $\mathcal{G} \in \mathfrak{G}$, the boundary $\mathcal{H} = \partial\mathcal{G}$ is a hyperplane supporting K at a point in $\mathcal{N}(z^*)$. Conversely, for any hyperplane $\mathcal{H}$ supporting K at a point in $\mathcal{N}(z^*)$, the closed half-space $\mathcal{G}$ with boundary $\mathcal{H}$ and disjoint from K belongs to $\mathfrak{G}$.

In a technical lemma, Klee [1953, 3.1] proved

$$\bigcap \mathfrak{G} = \bigcap_{\mathcal{G} \in \mathfrak{G}} \mathcal{G} = \varnothing.$$

Taking interiors, the family

$$\mathfrak{I} = \operatorname{int} \mathfrak{G} = \{\operatorname{int} \mathcal{G} \mid \mathcal{G} \in \mathfrak{G}\}$$

of open half-spaces is in Klee's terminology 0-closed. This means that, for any sequence $\{\mathcal{I}_k\}_{k \geq 1} \subset \mathfrak{G}$ which is Kuratowski convergent to a limit $\mathcal{I}$, we have $\operatorname{int} \mathcal{I} \in \mathfrak{G}$. (Note that, by definition, any Kuratowski limit is a closed set.)

We now need Klee's extension of Helly's theorem for 0-closed families: If any $n+1$ members of an 0-closed family has nonempty intersection then the interior of the intersection of all members of the family is nonempty (see [Klee 1953, 3.2]).

We apply this to our family $\mathfrak{I}$ of open half-spaces above. Since $\bigcap \mathfrak{I} = \varnothing$ (as $\bigcap \mathfrak{G} = \varnothing$) we see that there are $n+1$ open half-spaces $\mathcal{I}_0, \dots, \mathcal{I}_n \in \mathfrak{I}$ such that $\bigcap_{i=0}^{n} \mathcal{I}_k = \varnothing$.

Let $i = 0, \dots, n$. We select $x_i \in \mathcal{M}(z^*)$ such that the opposite $x_i^o \in \bar{\mathcal{I}}_i$ (with respect to z^*). Then $[x_i, x_i^o]$ is an affine diameter with $\lambda_K(x_i, z^*) = \mathfrak{m}_K(z^*) = \mathfrak{m}_K^*$. We let $f_i \in \operatorname{aff}_K$ be the (unique) normalized affine functional with zero set $\partial \mathcal{I}_i$. Since $x_i^o \in \partial \mathcal{I}_i$, we have $f_i(x_i^o) = 0$ and hence $f_i(x_i) = 1$. We calculate

$$f_i(z^*) = \frac{d(x_i^o, z^*)}{d(x_i^o, x_i)} = \frac{1}{d(x_i, z^*)/d(x_i^o, z^*) + 1} = \frac{1}{\lambda_K(x_i, z^*) + 1} = \frac{1}{\mathfrak{m}_K^* + 1}.$$

Summing up, we obtain

$$\sigma_{K,n}^o(z^*) \leq \sum_{i=0}^{n} f_i(z^*) = \frac{n+1}{\mathfrak{m}_K^* + 1}.$$

On the other hand, by (18), the right side is an obvious lower bound for $\sigma_{K,n}^o(z^*)$. □

4. Regular points and the Grünbaum conjecture

Let $K \subset \mathcal{X}$ be a convex body. Recall that $z \in \operatorname{int} K$ is a regular point if all minimal n-configurations in $\mathfrak{C}_{K,n}(z)$ are simplicial, that is they belong to $\Delta_K(z)$. Since minimal simplicial configurations do not contain any proper (necessarily minimal) subconfigurations, this condition can be conveniently reformulated in terms of the mean Minkowski measures: $z \in \operatorname{int} K$ is regular if and only if in (10) strict subarithmeticity holds:

$$\sigma_{K,n}(z) < \sigma_{K,n-1}(z) + \frac{1}{\mathfrak{m}_K(z) + 1}. \tag{34}$$

(For more details, see [Toth 2004].) Since the mean Minkowski measures are continuous, we see that the set of all regular points $\mathcal{R}_K \subset \operatorname{int} K$ is open.

Let $z \in \mathcal{R}_K$ be a regular point, and $\{x_0, \dots, x_n\} \in \Delta_K(z)$ a minimal simplicial configuration. Since z is in the interior of the n-simplex $[x_0, \dots, x_n]$, by (9), for each $i = 0, \dots, n$, the distortion $\lambda_K(\,\cdot\,, z)$ attains a *local maximum* at x_i. It is well known that at local maxima of the distortion the corresponding chord (through z) is an *affine diameter*. (See, for example [Hammer 1951] or [Toth 2006].) We conclude that, for each $i = 0, \dots, n$, the chord $[x_i, x_i^o]$ is an affine diameter. Thus, at any regular point $z \in \mathcal{R}_K$, $n+1$ affine diameters meet.

In 1963 Grünbaum conjectured that any convex body has a common point of $n+1$ affine diameters. We see that if $\mathcal{R}_K \neq \varnothing$ then we have an affirmative answer to Grünbaum's conjecture: At any regular point $n+1$ affine diameters meet.

We turn to the dual scenario. Recall that a dual n-configuration $\{f_0, \dots, f_n\} \in \mathfrak{C}^o_{K,n}(z)$ is called *simplicial* if $\{f_0^\sharp, \dots, f_n^\sharp\} \in \mathfrak{C}_{K^z,n}(z)$ is simplicial, where $\sharp = \sharp_{K,z} : \operatorname{aff}_K \to \partial K^z$ denotes the musical equivalence. As noted previously, geometrically speaking, a dual n-configuration $\{f_0, \dots, f_n\} \in \mathfrak{C}^o_{K,n}(z)$ is simplicial if and only if $\bigcap_{i=0}^n \{u \in \mathcal{X} \mid f_i(u) \geq 0\}$ is an n-simplex with z in its interior. The set of dual simplicial configurations is denoted by $\Delta^o_K(z)$. By (20), for $\{f_0, \dots, f_n\} \subset \operatorname{aff}_K$, we have

$$\{f_0, \dots, f_n\} \in \Delta^o_K(z) \quad \Longleftrightarrow \quad \{f_0^\sharp, \dots, f_n^\sharp\} \in \Delta_{K^z}(z).$$

Recall that an interior point $z \in \operatorname{int} K$ is called *dual regular* if any minimal dual n-configuration in $\mathfrak{C}^o_{K,n}(z)$ is simplicial. The set of all dual regular points is denoted by $\mathcal{R}^o_K \subset \operatorname{int} K$. As in the dual case, $z \in \mathcal{R}^o_K$ if and only if

$$\sigma^o_{K,n}(z) < \sigma^o_{K,n-1}(z) + \frac{1}{\mathfrak{m}_K(z)+1}, \tag{35}$$

in particular, $\mathcal{R}^o_K \subset \operatorname{int} K$ is open.

Now, comparing (34) and (35), Theorem 1 along with (7) gives (26).

Let $z \in \mathcal{R}^o_K$ be a dual regular point and $\{f_0, \dots, f_n\} \in \Delta^o_K(z)$ be a minimal simplicial configuration. We have $z \in \mathcal{R}_{K^z}$, and, by Theorem 1, $\{f_0^\sharp, \dots, f_n^\sharp\} \in \Delta_{K^z}(z)$ is a minimal simplicial configuration. By the discussion above, for each $i = 0, \dots, n$, the chord $[f_i^\sharp, (f_i^\sharp)^o]$ is an affine diameter of K^z. Let $\mathcal{K}_i$ and $\mathcal{K}_i^o$ be parallel hyperplanes at the endpoints of $f_i^\sharp$ and $(f_i^\sharp)^o$. Finally, let $g_i \in \operatorname{aff}_{K^z}$ be the normalized affine functional with zero sets $\mathcal{K}_i = \{u \in \mathcal{X} \mid g_i(u) = 0\}$ and $\mathcal{K}_i^o = \{u \in \mathcal{X} \mid 1 - g_i(u) = 0\}$. By the discussion at the end of Section 2, for each $i = 0, \dots, n$, the chord $[g_i^\sharp, (g_i^\sharp)^o]$ is an affine diameter of $K = (K^z)^z$, and the parallel supporting hyperplanes at the endpoints are given by the respective zero sets of the original affine functional $f_i = (f_i^\sharp)^\flat$. Letting $x_i = g_i^\sharp \in \partial K$, we see that the zero sets $\mathcal{H}_i = \{u \in \mathcal{X} \mid f_i(u) = 0\}$ and $\mathcal{H}_i^o = \{u \in \mathcal{X} \mid 1 - f_i(u) = 0\}$ are parallel supporting hyperplanes of K with affine diameters $[x_i, x_i^o] \subset K$, $i = 0, \dots, n$.

We claim that the affine diameters $[x_i, x_i^o]$, $i = 0, \dots, n$, are distinct. Assume that $[x_i, x_i^o] = [x_j, x_j^o]$ for some $i \neq j$, $i, j = 0, \dots, n$. (This means that this common affine diameter has two pairs of parallel supporting hyperplanes, $\mathcal{H}_i$, $\mathcal{H}_i^o$ and $\mathcal{H}_j$, $\mathcal{H}_j^o$.) Because $x_i = x_j$ or $x_i = x_j^o$, in the dual, we have $g_i = g_j$ or $g_i = 1 - g_j$. In particular, the affine diameters $[f_i^\sharp, (f_i^\sharp)^o]$ and $[f_j^\sharp, (f_j^\sharp)^o]$ of K^z share a single pair of parallel supporting hyperplanes, $\mathcal{K}_i = \mathcal{K}_j$, $\mathcal{K}_i^o = \mathcal{K}_j^o$, or $\mathcal{K}_i = \mathcal{K}_j^o$, $\mathcal{K}_i^o = \mathcal{K}_j$. On the other hand, in a minimal simplicial configuration of a regular point (such as $\{f_0^\sharp, \dots, f_n^\sharp\} \in \Delta_{K^z}(z)$ with $z \in \mathcal{R}_{K^z}$) two affine diameters cannot share the same parallel supporting hyperplanes since otherwise we can slide one in the respective hyperplanes (along a line segment) to the other to obtain another minimal configuration with multiple points or a pair of antipodal points. These contradict regularity.

We conclude that if $z \in \mathcal{R}_K^o$ then $n + 1$ affine diameters meet at z.

Proof of Theorem 4. Let $z^* \in K^*$ be a critical point of K. Subarithmeticity in (21) gives

$$\sigma_{K,n}^o(z^*) \leq \sigma_{K,n-1}^o(z^*) + \frac{1}{\mathfrak{m}_K^* + 1}.$$

The equality in (25) of Theorem 3 reduces this to (27), and the first statement of Theorem 4 follows. Strict inequality holds if and only if $z^* \in \mathcal{R}_K^o$, a dual regular point. By the discussion above, this implies the existence of $n + 1$ affine diameters across z^*. The second statement of Theorem 4 follows. □

Proof of Theorem 5. Let K be a symmetric convex body with center z_0. Assume that $z \in \operatorname{int} K$ is a dual regular point. Since the center z_0 is obviously not dual regular, we may assume that $z \neq z_0$. Let $\{f_0, \dots, f_n\} \in \mathfrak{C}_{K,n}^o(z)$ be a minimal configuration. Since $z \in \mathcal{R}_K^o$, this configuration is simplicial. Fix $i = 0, \dots, n$, and, for simplicity, suppress the subscript and set $f = f_i \in \operatorname{aff}_K$. By the discussion before the proof of Theorem 4, K has an affine diameter $[x, x^o] \subset K$ with supporting hyperplanes $\mathcal{H} = \{u \in \mathcal{X} \mid f(u) = 0\}$ and $\mathcal{H}^o = \{u \in \mathcal{X} \mid 1 - f(u) = 0\}$ such that $x \in \mathcal{H}$ and $x^o \in \mathcal{H}^o$. (Here the opposite is with respect to z.)

Let $A \in \partial K$ be the point at which the ray $\mathfrak{r}$ emanating from z_0 and passing through z meets the boundary of K. We claim that $[A, A^o]$ is an affine diameter of K, and, beyond A, this ray $\mathfrak{r}$ enters into the half-space $\{u \in \mathcal{X} \mid f(u) \leq 0\}$. Since $\mathfrak{r}$ is independent of $i = 0, \dots, n$, this means that the intersection in (16) is nonempty; a contradiction.

If x is on $\mathfrak{r}$ then $A = x$ and we are done. Thus we may assume that the points x, z, and z_0 are not collinear.

Let $x_0^o \in \partial K \cap \mathcal{H}^o$ be the opposite of x with respect to the center z_0. By symmetry, we have $[x^o, x_0^o] \subset \partial K \cap \mathcal{H}^o$.

Let $A_1 \in \partial K$ be the opposite of x_0^o with respect to z. Moving along the line segment $[x^o, x_0^o]$ and taking the opposites (with respect to z), we see that $A_1 \in \mathcal{H}$ since $f(z)$ is a local minimum in aff_K. Since $\mathcal{H}$ supports K, we have $[A_1, x] \subset \partial K \cap \mathcal{H}$. We now define A_k, $k \geq 1$, inductively as follows. Assume that $A_k \in \partial K$ is constructed with $[A_k, x] \subset \partial K \cap \mathcal{H}$. We take the opposite of A_k with respect to z_0 followed by the opposite with respect to z. This gives the point A_{k+1}. As before, we have $[A_{k+1}, x] \subset \partial K \cap \mathcal{H}$. The sequence $\{A_k\}_{k\geq 1}$ is actually collinear and converges to $A \in \partial K$ which then must be on $\mathcal{H}$. (In fact, an elementary argument shows that the sequence $\{d(A_k, A)\}_{k\geq 1}$ is geometric.) By construction, the chord $[A, A^o]$ is an affine diameter, where A^o is the opposite of A with respect to z. After A the ray $\mathfrak{r}$ enters the open half-space $\{u \in \mathcal{X} \mid f(u) < 0\}$. The claim follows. $\square$

Acknowledgement

The authors thank the referee for the careful reading and suggestions, which led to the improvement of the original manuscript.

References

[Bonnesen and Fenchel 1934] T. Bonnesen and W. Fenchel, *Theorie der konvexen Körper*, Ergebnisse der Math. (3) **1**, Springer, Berlin, 1934. Zbl

[Chakerian and Groemer 1983] G. D. Chakerian and H. Groemer, "Convex bodies of constant width", pp. 49–96 in *Convexity and its applications*, edited by P. M. Gruber and J. M. Wills, Birkhäuser, Basel, 1983. MR Zbl

[Eggleston 1958] H. G. Eggleston, *Convexity*, Cambridge Tracts in Mathematics and Mathematical Physics **47**, Cambridge Univ. Press, 1958. MR Zbl

[Grünbaum 1963] B. Grünbaum, "Measures of symmetry for convex sets", pp. 233–270 in *Convexity*, edited by V. Klee, Proc. Sympos. Pure Math. **7**, American Mathematical Society, Providence, RI, 1963. MR Zbl

[Guo and Toth 2016] Q. Guo and G. Toth, "Dual mean Minkowski measures of symmetry for convex bodies", *Sci. China Math.* **59**:7 (2016), 1383–1394. MR Zbl

[Hammer 1951] P. C. Hammer, "Convex bodies associated with a convex body", *Proc. Amer. Math. Soc.* **2** (1951), 781–793. MR Zbl

[Hammer and Sobczyk 1951] P. C. Hammer and A. Sobczyk, "Critical points of a convex body", *Bull. Amer. Math. Soc.* **57** (1951), 127, Abstract 112.

[Jin and Guo 2012] H. Jin and Q. Guo, "Asymmetry of convex bodies of constant width", *Discrete Comput. Geom.* **47**:2 (2012), 415–423. MR Zbl

[Klee 1953] V. Klee, "The critical set of a convex body", *Amer. J. Math.* **75**:1 (1953), 178–188. MR Zbl

[Lachand-Robert and Oudet 2007] T. Lachand-Robert and E. Oudet, "Bodies of constant width in arbitrary dimension", *Math. Nachr.* **280**:7 (2007), 740–750. MR Zbl

[Schneider 2014] R. Schneider, *Convex bodies: the Brunn–Minkowski theory*, 2nd ed., Encyclopedia of Mathematics and Its Applications **151**, Cambridge Univ. Press, 2014. MR Zbl

[Soltan 2005] V. Soltan, "Affine diameters of convex bodies: a survey", *Expo. Math.* **23**:1 (2005), 47–63. MR Zbl

[Soltan and Nguyên 1988] V. Soltan and M. H. Nguyên, "On the Grünbaum problem on affine diameters", *Soobshch. Akad. Nauk Gruzin. SSR* **132**:1 (1988), 33–35. In Russian. MR Zbl

[Toth 2004] G. Toth, "Simplicial intersections of a convex set and moduli for spherical minimal immersions", *Michigan Math. J.* **52**:2 (2004), 341–359. MR Zbl

[Toth 2006] G. Toth, "On the shape of the moduli of spherical minimal immersions", *Trans. Amer. Math. Soc.* **358**:6 (2006), 2425–2446. MR Zbl

[Toth 2009] G. Toth, "On the structure of convex sets with symmetries", *Geom. Dedicata* **143** (2009), 69–80. MR Zbl

Received March 14, 2016. Revised June 9, 2017.

QI GUO
DEPARTMENT OF MATHEMATICS
SUZHOU UNIVERSITY OF SCIENCE AND TECHNOLOGY
SUZHOU
CHINA
guoqi@mail.usts.edu.cn

GABOR TOTH
DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
CAMDEN, NJ
UNITED STATES
gtoth@camden.rutgers.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 292, No. 1, 2018
dx.doi.org/10.2140/pjm.2018.292.139

BORDERED FLOER HOMOLOGY OF (2, 2n)-TORUS LINK COMPLEMENT

JAEPIL LEE

We compute the bordered Floer homology $\widehat{CFDD}$ of the $(2, 2n)$-torus link complement and discuss assorted examples and type-*DD* structure homotopy equivalence.

1. Introduction

In recent years, Heegaard Floer theory has fascinated many low-dimensional topologists. Developed by P. Ozsváth and Z. Szábo, Heegaard Floer invariants of closed three-manifolds led to a breakthrough in low dimensional topology. These invariants were recently shown to be equivalent to three-dimensional Seiberg–Witten Floer homology by Kutluhan, Lee and Taubes [Kutluhan et al. 2011]. They were also proven to be equivalent to contact homology by Colin, Ghiggini and Honda [Colin et al. 2011]; this equivalence had initially motivated Oszváth and Szabó's constructions. Moreover, Heegaard Floer theory turned out to be useful in defining knot and link invariants; see [Ozsváth and Szabó 2004a; 2008a; Rasmussen 2002]. These invariants are now known as *knot Floer homology* and *link Floer homology*. In particular, knot Floer homology and Heegaard Floer homology of a three-manifold obtained by integral surgery on knot turned out to be closely related; see [Rasmussen 2002; Ozsváth and Szabó 2008b]. For the link surgery case, the relation was discovered but appeared more complicated than the knot case; see [Manolescu and Ozsváth 2010].

MSC2010: 53D40, 57M25.
Keywords: bordered Floer homology, knots and links.

More recently, Lipshitz, Ozsváth and Thurston extended the theory to three-manifolds with nonempty boundary. *Bordered Floer homology*, first introduced in [Lipshitz et al. 2008], consists of two different modules: $\widehat{CFD}$ and $\widehat{CFA}$. The homotopy type of each module is a topological invariant of a three-manifold with *connected* boundary equipped with a framing (a diffeomorphism to a model surface). The bordered theory is a powerful tool thanks to the pairing theorem: one can recover the Heegaard Floer homology of a closed 3-manifold decomposed into two pieces by taking "$\mathcal{A}_\infty$-tensor product" of $\widehat{CFA}$ of the first piece and $\widehat{CFD}$ of the second piece.

Moreover, Lipshitz, Ozsváth and Thurston [Lipshitz et al. 2015] have generalized bordered Floer homology to *doubly bordered Floer homology*. As the name suggests, this is an invariant associated to a three-manifold with two boundary components; we get three different types of bimodules, $\widehat{CFDA}$, $\widehat{CFDD}$, and $\widehat{CFAA}$. These bimodules are orginally invented to compute the bordered Floer homology of three-manifold with different framings. However, the doubly bordered Floer homology also provides an elegant algorithm to the compute Heegaard Floer homology of a closed three-manifold [Lipshitz et al. 2014], independent of the previously known combinatoric approach [Sarkar and Wang 2010].

In this paper, we give a calculation of $\widehat{CFDD}(S^3 \backslash \nu(L))$, where L is $(2, 2n)$-torus link. For a number of reasons, we mainly focus on the $\widehat{CFDD}$ module. First, it is the easiest bimodule to compute since it does not involve any $\mathcal{A}_\infty$-structure. Second, it is always possible to convert the $\widehat{CFDD}$ module to $\widehat{CFDA}$ or $\widehat{CFAA}$, by attaching the $\widehat{CFAA}(\mathbb{I})$ module to the left or right side of the $\widehat{CFDD}$ module. In Section 2, we collect the necessary background and notation. The actual calculation is in Section 3; the answer is shown in Proposition 3.9. (See also Figure 6 for a $(2, 6)$-torus link case.) The simplified version of the answer is in Figure 8. We work with a specific Heegaard diagram in order to find the generators and differentials of the module explicitly. However, only a few of the differentials can be obtained by the direct examination of their domains; for the remaining differentials, we have to exploit the A_∞-structure of $\widehat{CFAA}$. In Section 4, we give several applications of the pairing formula, recovering some known Floer homologies from our calculation, to illustrate and check the result.

Some other calculations of bordered invariants for manifolds with disconnected boundaries were recently obtained by Jonathan Hanselman [2016]. Hanselman computes the $\widehat{CFD}$-type trimodule associated to the trivial S^1-bundle over a pair of pants, and uses this, together with certain features of the bordered theory, to recover Heegaard Floer invariants of all graph manifolds. In principle, our results can also be obtained via Hanselman's approach (although he does not perform this calculation); however, our calculations are based on a more direct examination of the $(2, 2n)$ link complement, and thus it is perhaps more useful for understanding the bordered theory of more general link complements.

2. Background on doubly bordered Floer theory

We will assume that the reader is familiar with bordered Floer homology of a single boundary case. If not, we suggest the reader refer to [Lipshitz et al. 2011] for a brief introduction to the topic. In this section, we list the definitions and important results that will be used in the rest of the paper.

Algebraic preliminaries. Throughout this paper, we will use $\mathbb{F}$ to refer to $\mathbb{F}_2$.

We first begin with the algebra associated to a boundary surface of a three-manifold. In a handle decomposition of a genus g surface Σ^g, the zero-handle D of Σ^g has $2g$ marked points $\boldsymbol{a}$ on $\partial D = Z$ equipped with a two-to-one matching M between the points so that each one-handle is attached to a pair of matched points. We also fix a point z on Z away from $\boldsymbol{a}$. This set of data is called a *pointed matched circle* and denoted by

$$\mathcal{Z} = \{Z, \boldsymbol{a}, M, z\}.$$

$F(\mathcal{Z})$ denotes the surface obtained by the data and $D \subset F(\mathcal{Z})$ is called a *preferred disk*. The bordered Floer package associates a dg algebra to $\mathcal{Z}$, which will be a *strands algebra*, and denoted by $\mathcal{A}(\mathcal{Z})$.

Since we will be studying the torus boundary case, from now on we will assume that the genus g of the boundary surface equals one. In this case, $\mathcal{A}(\mathcal{Z})$ is $\mathbb{F}_2$-vector space generated by Reeb chords ρ_I, $I \in \{1, 2, 3, 12, 23, 123\}$ and two idempotents ι_1 and ι_2 such that $\iota_1 + \iota_2 = 1$. The multiplication rule between Reeb chords follows the concatenation rule of labels of chords; that is, $\rho_I \cdot \rho_J = \rho_{IJ}$ where $I, J \in \{1, 2, 3, 12, 23, 123\}$ and IJ is the concatenation of I and J. (If IJ is not in that set, then their product is zero.) For idempotents, $\iota_1 \rho_I = \rho_I$ if I starts with 1 or 3, $\rho_I \iota_1 = \rho_I$ if I ends with 2, $\iota_2 \rho_I = \rho_I$ if I starts with 2, and $\rho_I \iota_2 = \rho_I$ if I ends with 1 or 3. We let $\mathcal{I} \subset \mathcal{A}(\mathcal{Z})$ denote the subalgebra generated by idempotents ι_1 and ι_2. This strands algebra is called a *torus algebra*. A detailed description can be found in [Lipshitz et al. 2008, Chapter 3].

Next, we will study a (right) $\mathcal{A}_\infty$*-module* and a (left) *type-D module*. For an $\mathcal{A}_\infty$-algebra (A, μ_i), an $\mathcal{A}_\infty$-module is a $\mathbb{F}$-module M, equipped with maps

$$m_i : M \otimes A^{\otimes(i-1)} \to M$$

satisfying compatibility relations

$$\begin{aligned} 0 = &\sum_{i+j=n+1} m_i\big(m_j(\boldsymbol{x} \otimes a_1 \otimes \cdots \otimes a_{j-1}) \otimes \cdots \otimes a_{n-1}\big) \\ &+ \sum_{i+j=n+1} \sum_{l=1}^{n-j} m_i\big(\boldsymbol{x} \otimes a_1 \otimes \cdots \otimes a_{l-1} \otimes \mu_j(a_l \otimes \cdots \otimes a_{l+j-1}) \otimes \cdots \otimes a_{n-1}\big), \end{aligned}$$

for all $i \geq 1$. An $\mathcal{A}_\infty$-module is *strictly unital* if

$$m_2(\boldsymbol{x} \otimes 1) = \boldsymbol{x} \quad \text{and} \quad m_i(\boldsymbol{x} \otimes a_1 \otimes \cdots \otimes a_{i-1}) = 0 \text{ for } i > 2 \text{ and some } a_j = 1.$$

In bordered Floer theory, an $\mathcal{A}_\infty$-module is called a *type-A* module.

For a dg-algebra (A, μ_1, μ_2), a type-D module is a $\mathbb{F}$-module equipped with a map $\delta^1 : N \to A \otimes N$, satisfying the compatibility relation

$$0 = (\mu_2 \otimes \mathbb{I}_N) \circ (\mathbb{I}_A \otimes \delta^1) \circ \delta^1 + (\mu_1 \otimes \mathbb{I}_N) \circ \delta^1.$$

These modules are generalized to the following bimodules, namely a *type-AA bimodule* and a *type-DD bimodule*. In this paper, we will be mainly studying these bimodules.

Definition 2.1. Let A and B be $\mathcal{A}_\infty$-algebras over $\mathbb{F}$ equipped with $\mathcal{A}_\infty$-maps $\{\mu_i^A\}_{i>0}$ and $\{\mu_i^B\}_{i>0}$, respectively. A *right-right $\mathcal{A}_\infty$-bimodule* of *type-AA bimodule* $M_{\mathcal{A},\mathcal{B}}$ over A and B consists of a right-right $(\mathbb{F}, \mathbb{F})$-bimodule M and maps

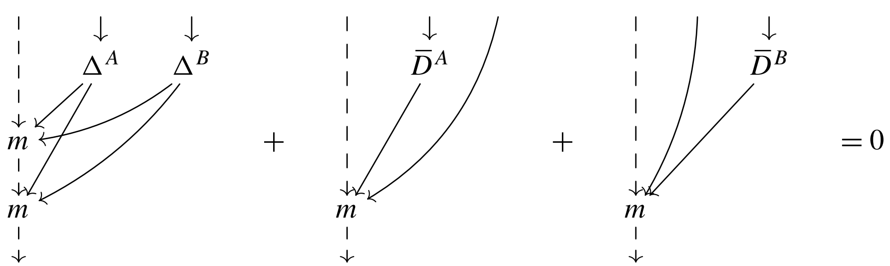
$$m_{1,i,j} : M \otimes A^{\otimes i} \otimes B^{\otimes j} \to M$$

such that the following compatibility condition holds.

$$\begin{aligned}
0 = \sum_{\substack{k+l=i\\ \lambda+\eta=j}} & m_{1,k,\lambda}\big(m_{1,l,\eta}(\boldsymbol{x}, a_1\otimes\cdots\otimes a_l, b_1\otimes\cdots\otimes b_\eta), a_{l+1}\otimes\cdots\otimes a_i, b_{\eta+1}\otimes\cdots\otimes b_j\big)\\
&+ \sum_{k+l=i+1} \sum_{n=1}^{i-l+1} m_{1,k,j}\big(\boldsymbol{x}, a_1 \otimes \cdots \otimes a_{n-1} \otimes \mu_l^A(a_n \otimes \cdots \otimes a_{n+l-1}) \otimes \cdots \otimes a_i,\\
&\qquad\qquad b_1 \otimes \cdots \otimes b_j\big)\\
&+ \sum_{\lambda+\eta=j+1} \sum_{n=1}^{j-\eta+1} m_{1,i,\lambda}\big(\boldsymbol{x}, a_1 \otimes \cdots \otimes a_i,\\
&\qquad\qquad b_1 \otimes \cdots \otimes b_{n-1} \otimes \mu_\eta^B(b_n \otimes \cdots \otimes b_{n+l-1}) \otimes \cdots \otimes b_j\big)
\end{aligned}$$

for all $i \geq 0$ and $j \geq 0$.

By writing $m = \sum_{i,j} m_{1,i,j}$, the compatibility condition can be drawn as the diagram below.



The dashed line above represents a module element, and the regular line represents an element from tensor algebra $\mathcal{T}^*A$ and $\mathcal{T}^*B$. The map $\Delta^A : \mathcal{T}^*A \to \mathcal{T}^*A \otimes \mathcal{T}^*A$ represents the canonical comultiplication

$$\Delta^A(a_1 \otimes \cdots \otimes a_n) = \sum_{m=0}^{n} (a_1 \otimes \cdots \otimes a_m) \otimes (a_{m+1} \otimes \cdots \otimes a_n),$$

and $\overline{D}^A : \mathcal{T}^*A \to \mathcal{T}^*A$ is defined as

$$\overline{D}^A(a_1 \otimes \cdots \otimes a_n) = \sum_{j=1}^{n} \sum_{l=1}^{n-j+1} a_1 \otimes \cdots \otimes \mu_j^A(a_l \otimes \cdots \otimes a_{l+j-1}) \otimes \cdots \otimes a_n.$$

Δ^B and $\overline{D}^B$ are defined similarly.

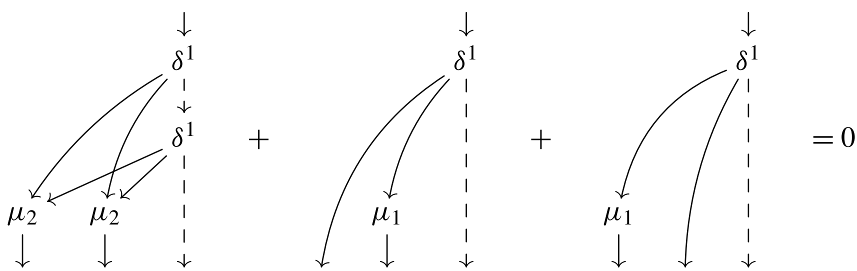
Definition 2.2. Let A and B be $\mathcal{A}_\infty$-algebras over $\mathbb{F}$. A *left-left type-DD bimodule* ${}^{A,B}M$ over A and B consists of left-left ($\mathbb{F}$, $\mathbb{F}$)-bimodule M and maps

$$\delta^1 : M \to A \otimes B \otimes M$$

satisfying the following compatibility condition.

$$((\mu_2^L, \mu_2^R) \otimes \mathbb{I}_M) \circ ((\mathbb{I}_A, \mathbb{I}_B) \otimes \delta^1) \circ \delta^1 + ((\mu_1^L, \mathbb{I}_B) \otimes \mathbb{I}_M) \circ \delta^1 + ((\mathbb{I}_A, \mu_1^R) \otimes \mathbb{I}_M) \circ \delta^1 = 0.$$

Again, the compatibility condition is drawn as the diagram below.



Heegaard diagram of the bordered three-manifold. A *bordered three-manifold* is a quadruple $(Y_1, \Delta_1, z_1, \psi_1)$, where Y_1 is a three-manifold with boundary, Δ_1 is a disk in ∂Y_1, z_1 is a point in $\partial \Delta_1$, and $\psi_1 : (F(\mathcal{Z}), D, z) \to (\partial Y_1, \Delta_1, z_1)$ is a parametrization of boundary. That is , ψ is a homeomorphism from $F(\mathcal{Z})$ to ∂Y_1 sending D to Δ_1 and z to z_1.

To describe a bordered three-manifold, we use a *bordered Heegaard diagram.*

Definition 2.3. A *bordered Heegaard diagram* is a quadruple $\mathcal{H} = (\bar{\sigma}, \overline{\boldsymbol{\alpha}}, \boldsymbol{\beta}, z)$ consisting of

- a compact, oriented surface $\bar{\sigma}$ of genus g with a single boundary component;
- a g-tuple of disjoint circles $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_g\}$ in the interior of $\bar{\sigma}$;

- a $g+k$-tuple of disjoint curves $\overline{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^c \cup \boldsymbol{\alpha}^a$ in $\bar{\sigma}$, where $\boldsymbol{\alpha}^c = \{\alpha_1^c, \dots, \alpha_{g-k}^c\}$ is a set of circles in the interior of $\bar{\sigma}$, and $\boldsymbol{\alpha}^a = \{\alpha_1^a, \dots, \alpha_{2k}^a\}$ is a set of arcs whose boundaries are in $\partial\bar{\sigma}$;
- a point z in $\partial\bar{\sigma}$, away from the boundaries of arcs in $\boldsymbol{\alpha}^a$,

such that $\bar{\sigma}\backslash\overline{\boldsymbol{\alpha}}$ and $\bar{\sigma}\backslash\boldsymbol{\beta}$ are connected, and $\overline{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta}$ intersect transversally.

We construct a bordered three-manifold from a bordered Heegaard diagram $\mathcal{H}$ in the following manner. First, we obtain a three-manifold with boundary $Y(\mathcal{H})$ by thickening $\bar{\sigma} \times [0, 1]$ and attaching a three-dimensional two-handle to each $\alpha_i^c \times \{0\} \times \bar{\sigma}$ and a three-dimensional two-handle to each $\beta_i \times \{1\} \times \bar{\sigma}$. The boundary of the resulting manifold is a genus k surface, and the surface is decomposed into a disk D and a genus k surface with a single boundary by $\partial\bar{\sigma} \times \{1\}$. Then, we get a bordered three-manifold $(Y(\mathcal{H}), D, z, \psi)$, where ψ is determined by $\boldsymbol{\alpha}^a$, which is considered as parametrization data of the surface.

A bordered Floer package defines a type-D module $\widehat{CFD}(\mathcal{H})$ and a type-A module $\widehat{CFA}(\mathcal{H})$ from a bordered Heegaard diagram $\mathcal{H}$, which are well defined up to quasi-isomorphism. Each module has a generating set $\mathfrak{S}(\mathcal{H})$, whose element $\boldsymbol{x} = \{x_1, \dots, x_g\}$ is a g-tuple of points in $\bar{\sigma}$ such that

- exactly one x_i lies on each β-circle,
- exactly one x_i lies on each α-circle and
- at most one x_i lies on each α-arc.

To compute nontrivial differentials for the Floer theory, we will need to compute holomophic curves in $\bar{\sigma} \times I_s \times \mathbb{R}_t$, where $I_s = [0, 1]$ with parameter s and $\mathbb{R}_t$ is $\mathbb{R}$ with parameter t. We will consider curves whose boundaries are on $\overline{\boldsymbol{\alpha}} \times \{1\} \times \mathbb{R}_t$ and $\boldsymbol{\beta} \times \{0\} \times \mathbb{R}_t$, asymtotic to $\boldsymbol{x} \times I_s$ and $\boldsymbol{y} \times I_s$ at $t = \pm\infty$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathfrak{S}(\mathcal{H})$. Each of the curves carries a relative homology class in the relative homology group

$$H_2\big(\bar{\sigma} \times I_s \times [-\infty, +\infty],\ ((\overline{\boldsymbol{\alpha}} \times \{1\} \cup \boldsymbol{\beta} \times \{0\} \cup ((\partial\bar{\sigma}\backslash z) \times I_s)) \times [-\infty, +\infty]) \\ \cup ((\boldsymbol{x} \times I_s \times \{-\infty\}) \cup (\boldsymbol{y} \times I_s \times \{+\infty\}))\big).$$

We write $\pi_2(\boldsymbol{x}, \boldsymbol{y})$ as the set of these relative homology classes.

Note that for $B \in \pi_2(\boldsymbol{x}, \boldsymbol{y})$, projecting B onto $\bar{\sigma}$ gives an element in $H_2(\bar{\sigma}, \overline{\boldsymbol{\alpha}} \cup \boldsymbol{\beta} \cup \partial\bar{\sigma})$. This is a linear combination of components of $\bar{\sigma}\backslash(\overline{\boldsymbol{\alpha}} \cup \boldsymbol{\beta})$. This linear combination will be called *domain*. Typically a domain is written as a linear combination of regions (connected subset of $\bar{\sigma}\backslash(\overline{\boldsymbol{\alpha}} \cup \boldsymbol{\beta})$). In particular, if any $B \in \pi_2(\boldsymbol{x}, \boldsymbol{y})$ is meeting $(\partial\bar{\sigma}\backslash z) \times I_s \times [-\infty, +\infty]$, then it can be interpreted as the corresponding domain being adjacent to the boundary of $\bar{\sigma}$, and that gives a sequence of Reeb chords $\boldsymbol{\rho} = (\rho_1, \dots, \rho_n)$. We call $(B, \boldsymbol{\rho})$ a *compatible pair*.

There is an operation $* : \pi_2(\boldsymbol{x}, \boldsymbol{y}) \times \pi_2(\boldsymbol{y}, \boldsymbol{z}) \to \pi_2(\boldsymbol{x}, \boldsymbol{z})$, defined by concatenating two homology classes in the t factor. In particular, if $\pi_2(\boldsymbol{x}, \boldsymbol{y})$ is nonempty, then the action of $\pi_2(\boldsymbol{x},\boldsymbol{x})$ on $\pi_2(\boldsymbol{x},\boldsymbol{y})$ is free and transitive. The domain of the element in $\pi_2(\boldsymbol{x},\boldsymbol{x})$ is called *periodic domain*. In addition, $\pi_2^{\partial}(\boldsymbol{x},\boldsymbol{x})$ denotes a set of periodic domains not adjacent to the boundary. An element in $\pi_2^{\partial}(\boldsymbol{x},\boldsymbol{x})$ is a *provincial periodic domain*, and if every provincial periodic domain of a Heegaard diagram has both positive and negative coefficients, then the Heegaard diagram is called *provincially admissible*.

It is worth mentioning that

- if any $B\in\pi_2(\boldsymbol{x},\boldsymbol{y})$ represents a holomorphic curve, then all the coefficients of the domain of B must be nonnegative, and
- the operation $*$ of two classes corresponds to the sum of the respective domains.

We sometimes blur the distinction between homology classes and their domains if it does not cause confusion.

We define $\widehat{CFD}(\mathcal{H})$ as the following. Let $X(\mathcal{H})$ be the $\mathbb{F}$-module generated by $\mathfrak{S}(\mathcal{H})$ equipped with an action of $\mathcal{I}\subset\mathcal{A}=\mathcal{A}(-\mathcal{Z})$ (the negative sign means the algebra obtained from the pointed matched circle has an orientation opposite from the induced orientation of $\mathcal{H}$) such that for any idempotent $\iota\in\mathcal{I}$,

$$\iota\otimes\boldsymbol{x}:=\begin{cases}\boldsymbol{x} & \text{if the arc corresponding to } \iota \text{ is not occupied by } \boldsymbol{x},\\ 0 & \text{otherwise.}\end{cases}$$

Then $\widehat{CFD}(\mathcal{H}):=\mathcal{A}\otimes_{\mathcal{I}}X(\mathcal{H})$. Its differential δ^1 is defined as

$$\delta^1(\boldsymbol{x}):=\sum_{\boldsymbol{y}\in\mathfrak{S}(\mathcal{H})}\sum_{B\in\pi_2(\boldsymbol{x},\boldsymbol{y})}a^B_{\boldsymbol{x},\boldsymbol{y}}\cdot\boldsymbol{y},$$

where

$$a^B_{\boldsymbol{x},\boldsymbol{y}}:=\sum_{\{\boldsymbol{\rho}\mid\operatorname{ind}(B,\boldsymbol{\rho})=1\}}\#(\mathcal{M}^B(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\rho}))a(-\boldsymbol{\rho}).$$

Here, $\mathcal{M}^B(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\rho})$ denotes the moduli space of holomorphic curves of B representing the compatible pair $(B,\boldsymbol{\rho})$, and $\operatorname{ind}(B,\boldsymbol{\rho})$ the expected dimension of the moduli space. In addition, for $\boldsymbol{\rho}=\{\rho_1,\ldots,\rho_n\}$ a sequence of Reeb chords, $a(-\boldsymbol{\rho})$ be the product $a(-\rho_1)\cdots a(-\rho_n)\in\mathcal{A}$. (Again, the negative sign means that the orientation of the boundary $\partial\bar{\sigma}$ is opposite from the induced orientation.)

The differential δ^1 may not be well defined. In fact, there may be infinitely many homology classes in $\pi_2(\boldsymbol{x},\boldsymbol{y})$ if there is a periodic domain representing a holomorphic curve. To prevent this, we will work on a Heegaard diagram such that every periodic domain has both positive and negative coefficients. Such diagram is called *admissible*, and it is shown in [Lipshitz et al. 2008, Proposition 4.25] that every Heegaard diagram is isotopic to an admissible Heegaard diagram. (In fact, the provincial admissibility also ensures the sum is finite since the concatenation of nonprovincial periodic domains of holomorphic curves produces an algebra element that equals zero.)

The definition of $\widehat{CFA}(\mathcal{H})$ is similar. $\widehat{CFA}(\mathcal{H})$ is a $\mathbb{F}$-module generated by $\mathfrak{S}(\mathcal{H})$, equipped with an action of $\mathcal{I} \subset \mathcal{A}(\mathcal{Z})$ such that

$$\boldsymbol{x} \otimes \iota := \begin{cases} \boldsymbol{x} & \text{if the arc corresponding to } \iota \text{ is occupied by } \boldsymbol{x}, \\ 0 & \text{otherwise.} \end{cases}$$

$\widehat{CFA}(\mathcal{H})$ is $\mathbb{F}$-module $X(\mathcal{H})$ generated by $\mathfrak{S}$, equipped with the $\mathcal{A}_\infty$-module maps

$$m_{i+1} : X(\mathcal{H}) \otimes \underbrace{\mathcal{A}(\mathcal{Z}) \otimes \cdots \otimes \mathcal{A}(\mathcal{Z})}_{i\text{-times}} \to X(\mathcal{H})$$

such that

$$m_{n+1}(\boldsymbol{x}, \rho_1, \dots, \rho_n) := \sum_{\boldsymbol{y} \in \mathfrak{S}(\mathcal{H})} \sum_{\substack{B \in \pi_2(\boldsymbol{x}, \boldsymbol{y}) \\ \operatorname{ind}(B, \boldsymbol{\rho}) = 1}} \#(\mathcal{M}^B(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\rho}))\boldsymbol{y},$$

$$m_2(\boldsymbol{x}, 1) := \boldsymbol{x},$$

$$m_{n+1}(\boldsymbol{x}, \dots, 1, \dots) := 0, \quad n > 1.$$

Although these modules are defined via a specific Heegaad diagram $\mathcal{H}$, it turns out the homotopy type of these modules are well defined. Thus, they are modules defined on bordered three-manifold (with single boundary).

Doubly bordered three-manifold. The bordered three-manifold is easily extended to a three-manifold with two boundary components. A *doubly bordered three-manifold* has the following data; $(Y_{12}, \Delta_1, \Delta_2, z_1, z_2, \psi_1, \psi_2, \gamma)$. Y_{12} is an oriented three-manifold with boundary $F(\mathcal{Z}_1) \amalg F(\mathcal{Z}_2)$, Δ_i is a preferred disk of surface $F(\mathcal{Z}_i)$, z_i is a point on $\partial\Delta_i$, and ψ_i is a parametrization of $F(\mathcal{Z}_i)$, $i = 1, 2$. Moreover, γ is an arc connecting z_1 and z_2, equipped with a framing pointing into Δ_i.

A doubly bordered three-manifold can be realized by a Heegaard diagram with two boundaries, namely *arced bordered Heegaard diagram with two boundaries*.

Definition 2.4. An *arced bordered Heegaard diagram $\mathcal{H}$ with two boundaries* is a tuple $(\Sigma, \overline{\boldsymbol{\alpha}}, \boldsymbol{\beta}, z)$ satisfying:

- $\bar{\sigma}$ is a compact, genus g surface with two boundary components $\partial_L\bar{\sigma}$ and $\partial_R\bar{\sigma}$;
- $\boldsymbol{\beta}$ is g-tuple of pairwise disjoint curves in the interior of $\bar{\sigma}$;
- $\overline{\boldsymbol{\alpha}} = \{\boldsymbol{\alpha}^{a,L} = \{\alpha_1^{a,L}, \dots, \alpha_{2l}^{a,L}\}, \boldsymbol{\alpha}^{a,R} = \{\alpha_1^{a,R}, \dots, \alpha_{2r}^{a,R}\}, \boldsymbol{\alpha}^c = \{\alpha_1^c, \dots, \alpha_{g-l-r}^c\}\}$, is a collection of pairwise disjoint embedded arcs with boundary on $\partial_L\bar{\sigma}$ (the $\alpha_i^{a,L}$), arcs with boundary on $\partial_R\bar{\sigma}$ (the $\alpha_i^{a,R}$), and circles (the α_i^c) in the interior of $\bar{\sigma}$;
- z is a path in $\bar{\sigma} \backslash (\overline{\boldsymbol{\alpha}} \cup \boldsymbol{\beta})$ between $\partial_L\bar{\sigma}$ and $\partial_R\bar{\sigma}$,

such that $\overline{\boldsymbol{\alpha}}$ intersects $\boldsymbol{\beta}$ transversely, and $\bar{\sigma} \backslash \overline{\boldsymbol{\alpha}}$ and $\bar{\sigma} \backslash \boldsymbol{\beta}$ are connected.

Since there are two boundaries, we have two pointed matched circles. These are

$$\mathcal{Z}_L(\mathcal{H}) = (\partial_L\bar{\sigma}, \boldsymbol{\alpha}^{a,L} \cap \partial_L\bar{\sigma}, M_L, \boldsymbol{z} \cap \partial_L\bar{\sigma}),$$
$$\mathcal{Z}_R(\mathcal{H}) = (\partial_R\bar{\sigma}, \boldsymbol{\alpha}^{a,R} \cap \partial_R\bar{\sigma}, M_R, \boldsymbol{z} \cap \partial_R\bar{\sigma}).$$

The construction of a doubly bordered three-manifold is similar to the construction of a single boundary case. For an arced bordered Heegaard diagram $\mathcal{H}$, cut open the diagram along the arc $\boldsymbol{z}$. The resulting diagram is a bordered Heegaard diagram with a single boundary, which will be written as $\mathcal{H}_{dr}$. Then, construct a bordered three-manifold $Y(\mathcal{H}_{dr})$. The boundary of $Y(\mathcal{H}_{dr})$ is a surface that can be decomposed as a connected sum $F(\mathcal{Z}_L)\#F(\mathcal{Z}_R)$. Finally, attach a three-dimensional two-handle along the connect sum annulus.

The three-manifold $Y(\mathcal{H}) := Y(\mathcal{H}_{dr}) \cup \{\text{two-handle}\}$ has the following properties.

- It has two boundary surfaces $F(\mathcal{Z}_L)$ and $F(\mathcal{Z}_R)$ with parametrization given by $\boldsymbol{\alpha}^{a,L}$ and $\boldsymbol{\alpha}^{a,R}$, respectively.
- Each boundary surface has a preferred disk bounded by $\partial_L\bar{\sigma}$ or $\partial_R\bar{\sigma}$.
- Cutting open the Heegaard diagram $\mathcal{H}$ would result in two arcs $\boldsymbol{z}^+$ and $\boldsymbol{z}^-$ on the deleted neighborhood of $\boldsymbol{z}$. Then, the arc $\boldsymbol{z}^+$, thought as a subset of the boundary of $Y(\mathcal{H}_{dr})$, is the framed arc in $Y(\mathcal{H})$ connecting z_1 and z_2.

For an arced Heegaard diagram $\mathcal{H}$, the type-*DD* bimodule $\widehat{CFDD}(\mathcal{H})$ is defined almost the same as in $\widehat{CFD}$. $\widehat{CFDD}(\mathcal{H})$ is a left-left $\mathbb{F}$-$\mathbb{F}$-module generated by $\mathfrak{S}(\mathcal{H}_{dr})$, equipped with two left actions of $\mathcal{I}_L \subset \mathcal{A}_L := \mathcal{A}(-\mathcal{Z}_L)$ and $\mathcal{I}_R \subset \mathcal{A}_R := \mathcal{A}(-\mathcal{Z}_R)$ such that for $\iota_L \in \mathcal{I}_L$ and $\iota_R \in \mathcal{I}_R$,

$$\iota_L \otimes \iota_R \otimes \boldsymbol{x} := \begin{cases} \boldsymbol{x} & \text{if the arc corresponding to } \iota_L \text{ and } \iota_R \text{ are not occupied by } \boldsymbol{x}, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\widehat{CFDD}(\mathcal{H}) = \mathcal{A}_L \otimes \mathcal{A}_R \otimes \mathfrak{S}(\mathcal{H}_{dr})$ with the differential

$$\delta^1(\boldsymbol{x}) := \sum_{\boldsymbol{y} \in \mathfrak{S}(\mathcal{H}_{dr})} \sum_{B \in \pi_2(\boldsymbol{x},\boldsymbol{y})} a^B_{\boldsymbol{x},\boldsymbol{y}} \cdot \boldsymbol{y},$$

where

$$a^B_{\boldsymbol{x},\boldsymbol{y}} := \sum_{\substack{\boldsymbol{\rho}^L, \boldsymbol{\rho}^R \\ \operatorname{ind}(B,\boldsymbol{\rho}^L,\boldsymbol{\rho}^R)=1}} \#(\mathcal{M}^B(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\rho}^L,\boldsymbol{\rho}^R)) a(-\boldsymbol{\rho}^L) \otimes a(-\boldsymbol{\rho}^R).$$

Similarly, a type-*AA* bimodule $\widehat{CFAA}(\mathcal{H})$ is defined by a right-right $\mathbb{F}$-$\mathbb{F}$ bimodule generated by $\mathfrak{S}(\mathcal{H}_{dr})$ with right-right actions of idempotents.

$$\boldsymbol{x} \otimes \iota_L \otimes \iota_R := \begin{cases} \boldsymbol{x} & \text{if the arc corresponding to } \iota_L \text{ and } \iota_R \text{ are occupied by } \boldsymbol{x}, \\ 0 & \text{otherwise.} \end{cases}$$

The type-AA module maps are

$$m_{n+m+1}(\boldsymbol{x}, \rho_1^L, \ldots, \rho_n^L, \rho_1^R, \ldots, \rho_m^R) := \sum_{\boldsymbol{y} \in \mathfrak{S}(\mathcal{H})} \sum_{\substack{B \in \pi_2(\boldsymbol{x}, \boldsymbol{y}) \\ \operatorname{ind}(B, \boldsymbol{\rho}^L, \boldsymbol{\rho}^R) = 1}} \#(\mathcal{M}^B(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\rho}^L, \boldsymbol{\rho}^R))\boldsymbol{y}.$$

Lastly, the expected dimension of the moduli space of $\mathcal{M}^B(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\rho}^L, \boldsymbol{\rho}^R)$, or $\operatorname{ind}(B, \boldsymbol{\rho}^L, \boldsymbol{\rho}^R)$ is computed by the formula below.

$$\operatorname{ind}(B, \boldsymbol{\rho}^L, \boldsymbol{\rho}^R) = e(B) + n_{\boldsymbol{x}}(B) + n_{\boldsymbol{y}}(B) + |\boldsymbol{\rho}^L| + |\boldsymbol{\rho}^R| + \iota(\boldsymbol{\rho}^L) + \iota(\boldsymbol{\rho}^R),$$

where $e(B)$ is *Euler measure*, $n_{\boldsymbol{x}}(B)$ sum of average of local multiplicities surrounding generator $\boldsymbol{x}$, $|\boldsymbol{\rho}^L|$ number of Reeb chords in the sequence $\boldsymbol{\rho}^L$, and $\iota(\boldsymbol{\rho}^L)$ *linking number* of sequence $\boldsymbol{\rho}^L$. In particular, if $(B, \boldsymbol{\rho}^L, \boldsymbol{\rho}^R)$ is a provincial domain, then the above formula reduces to

$$\operatorname{ind}(B, \boldsymbol{\rho}^L, \boldsymbol{\rho}^R) = e(B) + n_{\boldsymbol{x}}(B) + n_{\boldsymbol{y}}(B). \tag{1}$$

See [Lipshitz et al. 2008, Definition 5.11] for a detailed explanation.

Pairing theorem. The type-A module and type-D modules can be paired, which results in the classical Heegaard Floer homology of a closed three-manifold. The original pairing theorem is given in [Lipshitz et al. 2008, Theorem 1.3]. For any two three-manifolds Y_1 and Y_2 with $\partial Y_1 = F(\mathcal{Z}) = -\partial Y_2$,

$$\widehat{CFA}(Y_1) \widetilde{\otimes}_{\mathcal{A}(\mathcal{Z})} \widehat{CFD}(Y_2) \cong \widehat{CF}(Y_1 \cup_{F(\mathcal{Z})} Y_2)$$

where $\widetilde{\otimes}$ denotes the derived tensor product. The bimodule version of the pairing theorem is also given in [Lipshitz et al. 2015]. If Y_{12} is a doubly bordered three-manifold with boundary $F(\mathcal{Z}_1) \amalg F(\mathcal{Z}_2)$ and Y_1 is a bordered three-manifold with boundary $F(\mathcal{Z}_1)$, then

$$\widehat{CFD}(Y_1 \cup_{F(\mathcal{Z}_1)} Y_{12}) \cong \widehat{CFA}(Y_1) \widetilde{\otimes}_{\mathcal{A}(\mathcal{Z}_1)} \widehat{CFDD}(Y_{12}).$$

There exists many other variations of the pairing theorem. Interested readers should refer to [Lipshitz et al. 2015].

3. Computation of the bordered Floer bimodule of the (2, 2n)-torus link

Schubert normal form and diagram of 2-bridge link complements. As we will mainly focus on 2-bridge links, it is useful to mention *Schubert normal form* of a 2-bridge links (or knots). Let p be an even positive integer and q be an integer such that $0 < q < p$ and $\gcd(p, q) = 1$. Let us consider a circle with $2p$ marked point on its boundary. Choose a point and label it a_0. Label the other points $a_1, \ldots, a_{2p-1}$ in a clockwise direction. Then, connect a_i and a_{2p-i} with a straight

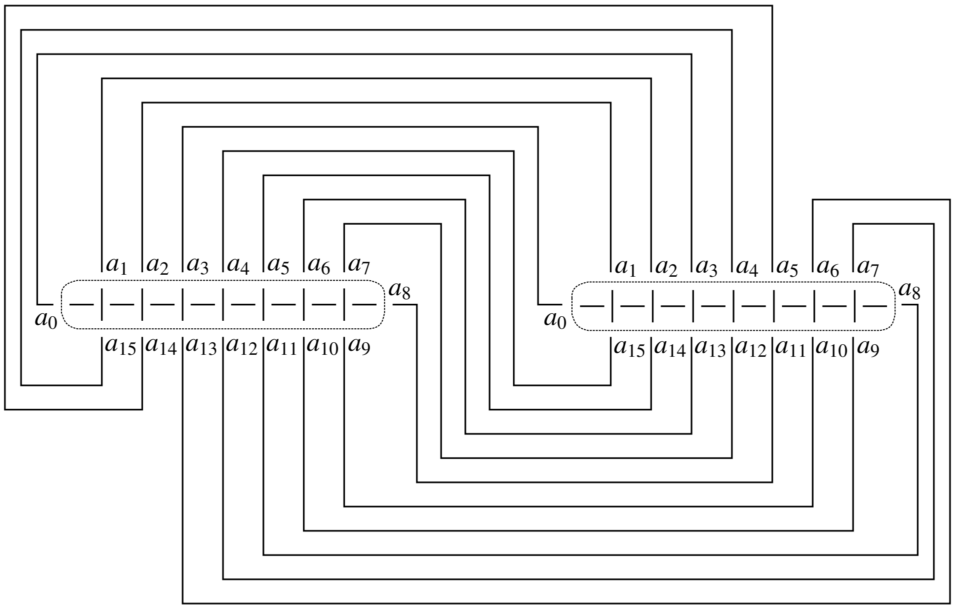


Figure 1. Schubert normal form of the $S(8, 3)$-link, or $L5a1$ in Thistlethwaite's notation.

line, $i = 1, \dots, p-1$. Finally, connect a_0 and a_p with an underbridge, a straight line that crosses below all of the other straight lines.

Now consider two copies of such circle. Draw arcs between these two circles so that each arc is connecting a_i on the left circle to a_{q-i} on the right circle (the labeling is modulo $2p$). These arcs should not intersect any of the straight lines and arcs. The resulting diagram gives a link that we denote $S(p, q)$. The diagram is called *Schubert normal form* of the link. See Figure 1. By construction, the diagram $S(p, q)$ has exactly two component since every even-labelled point on the right is connected to the odd-labelled point on the left (and the odd-labelled point on the right to the even-labelled point on the left). In particular, $S(2n, 1)$ is the $(2, 2n)$-torus link. More detailed description, especially about the Schubert normal form of 2-bridge knot can be found in [Rasmussen 2002, Chapter 2].

Heegaard diagram of* 2*-bridge link complement. Recall that a 2-bridge link L is a link in S^3 that admits a link diagram with two maxima and two minima. Let B_1 and B_2 be small neighborhoods of those two maxima. Consider

$$(S^3 \backslash \nu L) \backslash (B_1 \cup B_2).$$

Drilling a tunnel connecting B_1 and B_2 gives a three-manifold Y with single boundary, and the boundary is a genus 2 surface. Also, the longitudes λ_L and λ_R of the

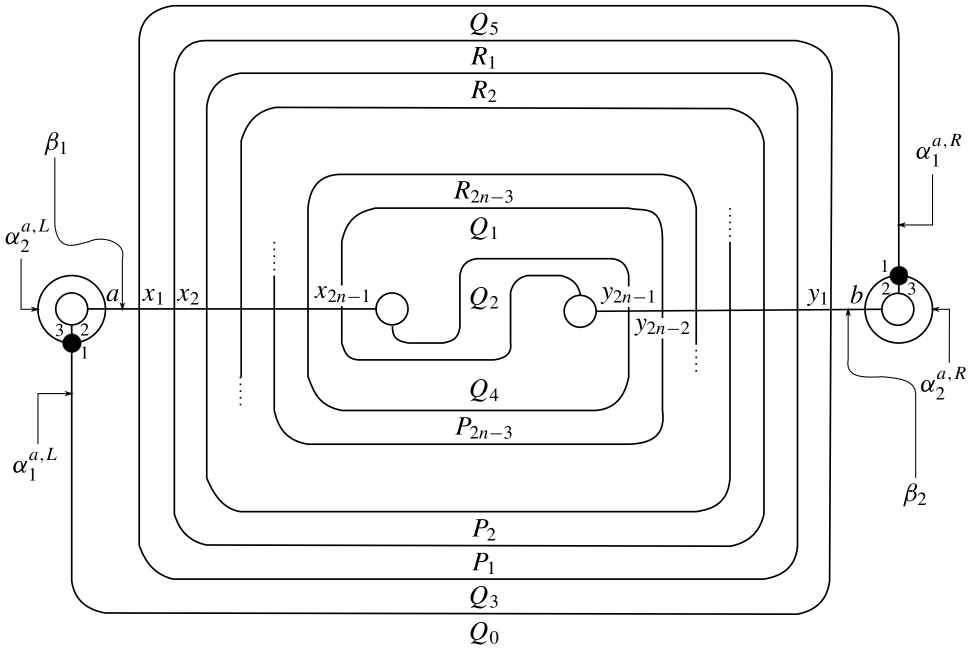


Figure 2. A general diagram of the $(2, 2n)$-torus link. The domain Q_0 has a framed arc. The orientation on the boundaries is opposite from the usual "right-hand" orientation.

left and right components of L are considered as curves on $\partial(\nu L)$; therefore the longitudes are also curves on the boundary of the drilled three-manifold.

The resulting manifold can be viewed as a handlebody with one zero-handle and two one-handles attached to it. To get a bordered Heegaard diagram, we will apply the following procedures on the boundary of the three-manifold. First, apply an isotopy of the boundary surface so that the longitudes have the Schubert normal form. Then, draw two circles β_1 and β_2 on the boundary surface so that they are parallel to the core of the one-handles on the boundary of the one-handles. Next, draw the meridians μ_L and μ_R on the belt sphere of each one-handle. Finally, make two punctures at the two intersections of meridians and longitudes and relabel λ_L to $\alpha_1^{a,L}$ and μ_L to $\alpha_2^{a,L}$ (respectively, λ_R to $\alpha_1^{a,R}$ and μ_R to $\alpha_2^{a,R}$).

In particular, if L equals the $(2, 2n)$-torus link, then we can draw an arc z on the surface connecting two punctures so that the arc is not intersecting $\overline{\boldsymbol{\alpha}}$ or $\boldsymbol{\beta}$ curves. The resulting diagram of the $(2, 2n)$-torus link complement is given in Figure 2.

Remark 3.1. Readers should be aware that connecting the left and right punctures with an (framed) arc is not always possible. In fact, a domain that is adjacent to both punctures does not exist except for the $(2, 2n)$-torus link case. To fix this, choose

μ_L or μ_R and apply a finger move on the chosen meridian along the longitude so that the resulting puncture is on the domain that is adjacent to the other puncture.

Computation of the type-DD module differential. Now, we compute $\widehat{CFDD}(\mathcal{H})$, where $\mathcal{H}$ is the Heegaard diagram of the $(2, 2n)$-torus link complement constructed in the previous section.

First, we will see whether the diagram $\mathcal{H}$ is provincially admissible. Second, we will investigate the genus-zero rectangular domains that cause a nontrivial differential. Then, using the result as a building block, we will consider domains of higher genus and the moduli space of homolorphic curves of the domains. The differentials associated to the higher genus domains are computed by $\mathcal{A}_\infty$-relations, dualizing $\widehat{CFDD}$-bimodule to $\widehat{CFAA}$-bimodule.

Periodic domain. First, we investigate periodic domains $\pi_2(\boldsymbol{x}, \boldsymbol{x})$. It is well known that $\pi_2(\boldsymbol{x}, \boldsymbol{x}) \cong H_2(Y(\mathcal{H}), \partial Y(\mathcal{H})) \cong \mathbb{Z} \oplus \mathbb{Z}$ by the Mayer–Vietoris sequence. Thus, there are two linearly independent periodic domains in the diagram. The proof can be found in [Lipshitz 2006, Lemma 2.6.1] or [Lipshitz et al. 2008, Lemma 4.18]. In their proof, they use the isomorphism

$$\boldsymbol{\pi}_2(\boldsymbol{x}, \boldsymbol{x}) \cong H_2\big(\Sigma' \times [0, 1], (\overline{\boldsymbol{\alpha}} \times \{1\}) \cup (\boldsymbol{\beta} \times \{0\})\big),$$

where $\Sigma' = (\bar{\sigma}/\partial\bar{\sigma})\backslash\{z\}$. The isomorphism given above is proved by investigating the long exact sequence of pair $\big(\Sigma' \times [0, 1], (\overline{\boldsymbol{\alpha}} \times \{1\}) \cup (\boldsymbol{\beta} \times \{0\})\big)$. That is,

$$\begin{aligned} \cdots \to \underbrace{H_2(\Sigma' \times [0, 1])}_{\cong 0} \to H_2\big(\Sigma' \times [0, 1], (\overline{\boldsymbol{\alpha}} \times \{1\}) \cup (\boldsymbol{\beta} \times \{0\})\big) \\ \to H_1\big((\overline{\boldsymbol{\alpha}} \times \{1\}) \cup (\boldsymbol{\beta} \times \{0\})\big) \to H_1(\Sigma'). \end{aligned}$$

Thus, the periodic domain $\boldsymbol{\pi}_2(\boldsymbol{x}, \boldsymbol{x}) \cong \ker\big(H_1(\overline{\boldsymbol{\alpha}}/\partial\overline{\boldsymbol{\alpha}}) \oplus H_1(\boldsymbol{\beta}) \to H_1(\bar{\sigma}/\partial\overline{\Sigma})\big)$. This isomorphism enables us to find periodic domains from a bordered Heegaard diagram by choosing combinations of $\overline{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta}$ curves such that the sum of their image in $H_1(\bar{\sigma}/\partial\bar{\sigma})$ equals zero. We briefly describe how to find the periodic domain from such combinations. Explicitly, first choose any orientation on the longitude $\alpha_1^{a,L}$ ($\alpha_1^{a,R}$, respectively). This induces the orientation of β_1 (β_2, respectively) as follows. For example, if the orientation of $\alpha_1^{a,L}$ is in a counterclockwise direction, then the orientation of β_1 is from right to left in the diagram. Then, we impose the coefficient zero to the outermost region that contains the framed arc. Starting from the outermost region, we give coefficients to regions adjacent to it according to the following rule. Suppose we have two adjacent regions A and B such that the coefficient of A equals l and the coefficient of B is not determined. If we can reach region B from region A by crossing a curve of multiplicity k from right to left (notion of "left" and "right" is justified since we have orientation of curves), we give the region B coefficient $k+l$; otherwise we give coefficient $-k+l$. If we can give coefficients to all regions consistently in this way, then the orientations given to curves $\overline{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta}$ is boundary in $H_1(\bar{\sigma}/\partial\bar{\sigma})$.

Since there are two possible choices of orientations of longitudes up to sign, we find two generators of $\pi_2(\boldsymbol{x},\boldsymbol{x})$. Then the periodic domains are

$$Q_3+Q_5+\sum_{i=1}^{2n-3}(i+1)(P_i+R_i)+(n+2)(Q_1+Q_4)+(n+3)Q_2$$

and

$$Q_3-Q_5+\sum_{i=1}^{2n-3}\frac{1+(-1)^i}{2}(P_i-R_i)+Q_4-Q_1.$$

The two generators are shown in Figure 3.

Thus, this diagram is provincially admissible; in fact, there is no provincial periodic domain here.

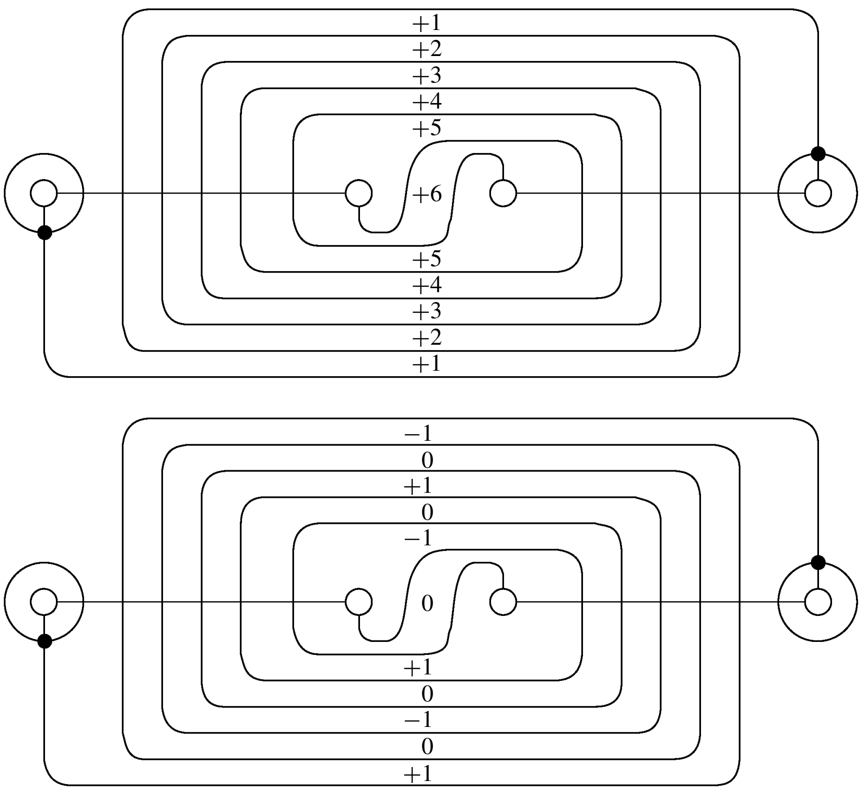


Figure 3. These two diagrams represents the two generators of the periodic domain $\pi_2(\boldsymbol{x},\boldsymbol{x})$, where the black dots represent left and right punctures.

Generators. According to the labeling given in the diagram, there are $2n^2 + 2n$ generators which are classified into four groups.

$$\begin{cases} \boldsymbol{x}_i \boldsymbol{y}_j & \text{where } i \text{ and } j \text{ have the same parity,} \\ \boldsymbol{a}\boldsymbol{y}_i & \text{where } i \text{ is even,} \\ \boldsymbol{x}_i \boldsymbol{b} & \text{where } i \text{ is even,} \\ \boldsymbol{a}\boldsymbol{y}_i, \boldsymbol{x}_j \boldsymbol{b} & \text{where } i \text{ and } j \text{ are odd.} \end{cases}$$

From now on, we will disregard generators of the last group for the following reason. The main purpose of the bordered Floer homology is to compute the Heegaard Floer homology of a three-manifold obtained by gluing along the boundaries of two three-manifolds with homeomorphic boundaries. In the link complement case, we glue the link complement and solid tori. Typically, a bordered Heegaard diagram of a solid torus is a genus one surface with a puncture, equipped with $\boldsymbol{\beta} = \{\beta_1\}$ and $\overline{\boldsymbol{\alpha}} = \{\alpha_1^a, \alpha_2^a\}$. In particular, these α_i^a arcs are glued to $\alpha_j^{a,L}$ or $\alpha_i^{a,R}$ of the doubly bordered diagram of the link complement, and every generator of the diagram of the solid torus is occupying exactly one α arc. Therefore, after pairing two diagrams of the solid tori to both sides of the diagram of the link complement, the generators of the last kind cannot appear in the generator set of the resulting diagram.

The differential $\delta^1 : \mathfrak{S}(\mathcal{H}) \to \mathcal{A}(-\partial_L \bar{\sigma}) \otimes \mathcal{A}(-\partial_R \bar{\sigma}) \otimes \mathfrak{S}(\mathcal{H})$ maps a generator $\boldsymbol{x} \in \mathfrak{S}(\mathcal{H})$ to $\sum \rho_I \otimes \sigma_J \otimes \boldsymbol{y}$, where $I, J \in \{\phi, 1, 2, 3, 12, 23, 123\}$. Here, ρ_I means an algebra element that comes from the left boundary strands algebra and σ_J, the right strands algebra. To investigate δ^1 actions on generators, it is convenient to classify the resulting terms by their strands algebra elements.

Algebra element 1. We begin by finding all provincial domains and show that only rectangular domains contribute to the differential δ^1.

Lemma 3.2. *Let $(B, \boldsymbol{\rho})$ be a compatible pair with $ind(B, \boldsymbol{\rho}) = 1$. If B is a nonrectangular domain, then $\boldsymbol{\rho}$ is nonempty.*

Proof. Suppose there is a nonrectangular provincial domain B that has a nontrivial contribution to differential δ^1. Then B must be a linear combination of R_i and P_j. See Figure 2. The region covered by B must be connected, otherwise the number of corners of B will be more than four. If a domain has more than four corners then it cannot represent proper differential because the each of two generators of the differential consists of two points. Therefore, B must be an annulus. Next, we claim that the number of the corners of B must be two. This claim is justified by considering the number of corners of different types. Since the number of corners of any domain should not exceed four, there are only five possibilities, each having i 90° corners and $4 - i$ 270° corners ($i = 0, 1, 2, 3, 4$). Since the domain was assumed to be provincial, it must be a combination of the regions $P_1, \ldots, P_{2n-3}$ and $R_1, \ldots, R_{2n-3}$. Considering the index formula $e(B) + n_{\boldsymbol{x}}(B) + n_{\boldsymbol{y}}(B)$ of (1), the indices of the first three cases cannot be one. Likewise, we can easily rule out the last case. The fourth case does not exist due to the following reason; because the shape of the domain is an annulus, the 270° corner must be on the boundary of the domain. Then, the other boundary must have two 90° corners. If not, i.e., if one boundary component has all three 90° corners, then there cannot be a holomorphic involution interchanging inner and outer boundaries. See [Ozsváth and Szabó 2004b, Lemma 9.4]. Thus, one boundary has two 90° corners and the other boundary has one 90° corner and one 270° corner. In particular, the boundary that has two 90° corners should consist of one $\overline{\boldsymbol{\alpha}}$ curve and one $\boldsymbol{\beta}$ curve, and the intersections have to be 90°. However, such a boundary cannot be obtained by any combination of the domains in Figure 2. □

In Figure 2, regions $P_1, \dots, P_{2n-3}$ and $R_1, \dots, R_{2n-3}$ are the only ones not adjacent to the boundaries. Thus, rectangular domains obtained by combining these regions are the only provincial domains. For $i = 1, \dots, 2n-3$ and $l \le (2n-5-i)/2$, the combinations have the form

$$P_i + \sum_{k=0}^{l} R_{i+1+2k} + P_{i+2+2k}, \quad R_i + \sum_{k=0}^{l} P_{i+1+2k} + R_{i+2+2k},$$
$$P_i + \sum_{k=0}^{l} P_{i+1+2k} + P_{i+2+2k}, \quad R_i + \sum_{k=0}^{l} R_{i+1+2k} + R_{i+2+2k}.$$

All of these domains are rectangular with four corners and each of these domains admits a unique holomorphic representative (up to translation) by the Riemann mapping theorem. The labellings of the corners tell which generators are involved. For example, the domain P_i has four corners $\boldsymbol{x}_i$, $\boldsymbol{x}_{i+1}$, $\boldsymbol{y}_{i+1}$ and $\boldsymbol{y}_{i+2}$; due to the orientation convention, this domain contributes to a differential from $\boldsymbol{x}_i \boldsymbol{y}_{i+2}$ to $\boldsymbol{x}_{i+1} \boldsymbol{y}_{i+1}$. We can write the terms with algebra element 1 obtained by taking differential of $\boldsymbol{x}_i \boldsymbol{y}_j$:

$$\boldsymbol{x}_i \boldsymbol{y}_j \mapsto \begin{cases} \boldsymbol{x}_{j-1} \boldsymbol{y}_{i+1} + \boldsymbol{x}_{i+1} \boldsymbol{y}_{j-1} & \text{if } j-i > 2, \\ \boldsymbol{x}_{j+1} \boldsymbol{y}_{i-1} + \boldsymbol{x}_{i-1} \boldsymbol{y}_{j+1} & \text{if } i-j > 2, \\ \boldsymbol{x}_{i+1} \boldsymbol{y}_{j-1} & \text{if } j-i = 2, \\ \boldsymbol{x}_{i-1} \boldsymbol{y}_{j+1} & \text{if } i-j = 2, \\ 0 & \text{if } i = j. \end{cases} \tag{2}$$

Algebra elements ρ_1 *and* σ_1. First, consider the algebra element ρ_1. Domain Q_3 is adjacent to the Reeb chords of algebra element ρ_1. Note that if the multiplicity of the domain Q_3 is greater than 1, then it cannot contribute to the nontrivial differential. (If so, then it will produce the algebra element $\rho_1 \cdot \rho_1$, which equals zero.) We list the possible domains that result in nontrivial differentials:

$$Q_3+\sum_{k=0}^{l} P_{1+2k}+P_{2+2k} \quad \text{and} \quad Q_3+\sum_{k=0}^{l} R_{1+2k}+P_{2+2k},$$

where $l \le n-2$.

All such domains are rectangular domains containing Q_3. These domains are all quadrilateral, and the dimension and the modulo two count of the moduli spaces are obvious. The differentials obtained from these domains are listed below.

$$\boldsymbol{a}\boldsymbol{y}_{2k} \mapsto \begin{cases} \rho_1 \otimes (\boldsymbol{x}_1\boldsymbol{y}_{2k-1}+\boldsymbol{x}_{2k-1}\boldsymbol{y}_1) & \text{if } k \neq 1, \\ \rho_1 \otimes \boldsymbol{x}_1\boldsymbol{y}_1 & \text{otherwise.} \end{cases}$$

Differentials involving σ_1 can be found in a parallel manner, by using the symmetry of the diagram.

$$\boldsymbol{x}_{2k}\boldsymbol{b} \mapsto \begin{cases} \sigma_1 \otimes (\boldsymbol{x}_{2k-1}\boldsymbol{y}_1+\boldsymbol{x}_1\boldsymbol{y}_{2k-1}) & \text{if } k \neq 1, \\ \sigma_1 \otimes \boldsymbol{x}_1\boldsymbol{y}_1 & \text{otherwise.} \end{cases}$$

Algebra elements ρ_3 and σ_3. Similarly, domains adjacent to ρ_3 are all listed

$$Q_1+\sum_{k=0}^{l} R_{2n-2k-3}+R_{2n-2k-4} \quad \text{and} \quad Q_1+\sum_{k=0}^{l} P_{2n-2k-3}+R_{2n-2k-4},$$

where $l \le n-2$.

Domains adjacent to σ_3 are similar. We get the differentials below:

$$\boldsymbol{a}\boldsymbol{y}_{2k} \mapsto \begin{cases} \rho_3 \otimes (\boldsymbol{x}_{2k+1}\boldsymbol{y}_{2n-1}+\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2k+1}) & \text{if } k \neq n-1, \\ \rho_3 \otimes \boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1} & \text{otherwise;} \end{cases}$$

$$\boldsymbol{x}_{2k}\boldsymbol{b} \mapsto \begin{cases} \sigma_3 \otimes (\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2k+1}+\boldsymbol{x}_{2k+1}\boldsymbol{y}_{2n-1}) & \text{if } k \neq n-1, \\ \sigma_3 \otimes \boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1} & \text{otherwise.} \end{cases}$$

Algebra element $\rho_2 \otimes \sigma_2$. The domain Q_2 adjacent to ρ_2 is adjacent to σ_2 as well. So, this is the one and only domain where the algebra element $\rho_2 \otimes \sigma_2$ occurs. Thus, we have $\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1} \mapsto \rho_2 \otimes \sigma_2 \otimes \boldsymbol{ab}$.

Algebra elements $\rho_3 \otimes \sigma_1$ and $\rho_1 \otimes \sigma_3$. There are two domains which contribute to $\rho_3 \otimes \sigma_1$; those are $Q_1+R_1+R_2+\cdots+R_{2n-3}+Q_5$ and $Q_1+P_1+R_2+P_3+R_4+\cdots+R_{2n-4}+P_{2n-3}+Q_5$. This gives $\boldsymbol{ab} \mapsto \rho_3 \otimes \sigma_1 \otimes (\boldsymbol{x}_1\boldsymbol{y}_{2n-1}+\boldsymbol{x}_{2n-1}\boldsymbol{y}_1)$. Again, using the symmetry of the diagram, $\boldsymbol{ab} \mapsto \rho_1 \otimes \sigma_3 \otimes (\boldsymbol{x}_1\boldsymbol{y}_{2n-1}+\boldsymbol{x}_{2n-1}\boldsymbol{y}_1)$.

Now, we will work on differentials whose domains are nonrectangular. To find holomorphic curves of such domains, we will consider $\widehat{CFAA}(\mathcal{H},0)$ so that we can use the $\mathcal{A}_\infty$-structure of it and ensure the existence of holomorphic curves and their count (modulo two).

Algebra element containing ρ_{12}. To take advantage of the $\mathcal{A}_\infty$-structure of $\widehat{CFAA}$, the orientation of two boundaries of the Heegaard diagram has to be reversed. We let $\bar{\rho}_I$ denote (respectively, $\bar{\sigma}_I$) the algebra element of the left strands algebra $\mathcal{A}(\mathcal{Z})$ (respectively, the right strands algebra); that is, an orientation reversing diffeomorphism $R : -S^1\backslash\{z\} \to S^1\backslash\{z\}$ induces a map $R_* : \mathcal{A}(-\mathcal{Z}) \to \mathcal{A}(\mathcal{Z})$ that maps $R_*(\rho_1) = \bar{\rho}_3$, $R_*(\rho_2) = \bar{\rho}_2$, $R_*(\rho_3) = \bar{\rho}_1$, and so on. The right boundary is similar.

Returning to $\widehat{CFDD}$, the domains contributing to ρ_{12} must contain Q_2 and Q_3. Clearly $Q_2 + Q_3$ has more than four corners, so we will consider $Q_2 + Q_3 + Q_4$ instead to get the domain of four corners. This domain possibly contributes to the differential from $\boldsymbol{x}_{2n-2}\boldsymbol{y}_2$ to $\boldsymbol{x}_1\boldsymbol{y}_1$. The only possible Maslov index one interpretation is $\mathcal{M}(\boldsymbol{x}_{2n-2}\boldsymbol{y}_2, \boldsymbol{x}_1\boldsymbol{y}_1; \bar{\rho}_{23}, \bar{\sigma}_{12})$ (there can be cuts between $\bar{\rho}_2$ and $\bar{\rho}_3$, and $\bar{\sigma}_1$ and $\bar{\sigma}_2$, but these cuts will increase the Maslov index by one). Under the interpretation, the domain is an annulus with one boundary consisting of two segments of $\boldsymbol{\alpha}$ curves and two segments of $\boldsymbol{\beta}$ curves, and another boundary consisting of $\boldsymbol{\alpha}$ curve only. In the sense of [Ozsváth and Szabó 2004b, Lemma 9.4], such an annulus cannot allow a holomorphic involution that interchanges one boundary with another, carrying $\boldsymbol{\alpha}$ curves to $\boldsymbol{\alpha}$ curves and $\boldsymbol{\beta}$ curves to $\boldsymbol{\beta}$ curves. Thus, the moduli space $\mathcal{M}(\boldsymbol{x}_{2n-2}\boldsymbol{y}_2, \boldsymbol{x}_1\boldsymbol{y}_1; \bar{\rho}_{23}, \bar{\sigma}_{12})$ cannot give a nontrivial differential. Domains such as $Q_2 + Q_3 + Q_4 + P_1 + P_2$ or $Q_2 + Q_3 + Q_4 + R_1 + P_2$ can be considered similarly to $Q_2 + Q_3 + Q_4$. In fact, they do not contribute to the nontrivial differential as long as the shape of the domain is topologically equivalent to $Q_2 + Q_3 + Q_4$.

There are two domains possibly giving a nontrivial differential; they are $Q_2 + Q_3 + P_1 + \cdots + P_{2n-3} + Q_4$ and $Q_2 + Q_3 + R_1 + P_2 + \cdots + R_{2n-3} + Q_4$. We will consider the domain $Q_2 + Q_3 + P_1 + \cdots + P_{2n-3} + Q_4$ first. It has three interpretations. Each of the interpretations comes from the choice of cuts made on the boundary of the domain. Cuts are allowed where the domain has 270° or 180° corners, or a point on the boundary intersecting $\boldsymbol{\alpha}$ curve. (Detailed discussions of domains and their cuts can be found in [Lipshitz et al. 2008, Chapter 6]. An interested reader may also want to see [Lipshitz et al. 2009] for more examples.) Thus, the domain $Q_2 + Q_3 + P_1 + \cdots + P_{2n-3} + Q_4$ has two points that possibly allow cuts; a point between ρ_1 and ρ_2, and a point between σ_2 and σ_3. Of course, it may not have any cuts at all. We list the moduli spaces of these interpretations as below:

- $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2n-1}, \boldsymbol{a}\boldsymbol{y}_1; \bar{\rho}_3, \bar{\rho}_2, \bar{\sigma}_{12})$,
- $\mathcal{M}(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \boldsymbol{x}_{2n-1}\boldsymbol{y}_1; \bar{\rho}_{23}, \bar{\sigma}_2, \bar{\sigma}_1)$,
- $\mathcal{M}(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \boldsymbol{x}_{2k-1}\boldsymbol{y}_1; \bar{\rho}_{23}, \bar{\sigma}_{12})$.

First, we will consider $\mathcal{M}(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \boldsymbol{x}_{2k-1}\boldsymbol{y}_1; \bar{\rho}_{23}, \bar{\sigma}_{12})$.

Lemma 3.3. *The modulo two count of the moduli space*

$$\mathcal{M}(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \boldsymbol{x}_{2k-1}\boldsymbol{y}_1; \bar{\rho}_{23}, \bar{\sigma}_{12})$$

is zero.

Proof. We will compute the signed number of the moduli space by considering the following $\mathcal{A}_\infty$-compatibility condition.

$$\begin{aligned}0 = m(m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}), \bar{\rho}_2, \bar{\rho}_3, \bar{\sigma}_{12}) + m(m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \bar{\rho}_2), \bar{\rho}_3, \bar{\sigma}_{12})\\ + m(m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \bar{\sigma}_{12}), \bar{\rho}_2, \bar{\rho}_3) + m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \mu(\bar{\rho}_2, \bar{\rho}_3), \bar{\sigma}_{12})\\ + m(m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \bar{\rho}_2, \bar{\sigma}_{12}), \bar{\rho}_3) + m(m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \bar{\rho}_2, \bar{\rho}_3, \bar{\sigma}_{12})).\end{aligned}$$

The right-hand side of the equation above consists of six terms. The second term vanishes because $m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \bar{\rho}_2)$ does not have the algebra element $\bar{\sigma}_2$ (note that domain Q_2 is adjacent to $\bar{\rho}_2$ and $\bar{\sigma}_2$). Similarly, the third term vanishes since $m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \bar{\sigma}_{12})$ has $\bar{\sigma}_{12}$ as its input but lacks $\bar{\rho}_2$. The last term also vanishes because the Maslov index is not one. Replacing $\mu(\bar{\rho}_2, \bar{\rho}_3) = \bar{\rho}_{23}$ and $m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}) = \boldsymbol{x}_{2n-2}\boldsymbol{y}_{2k} + \boldsymbol{x}_{2k}\boldsymbol{y}_{2n-2}$ (equation (2)), the above equation is reduced as follows.

$$\begin{aligned}0 = m(\boldsymbol{x}_{2n-2}\boldsymbol{y}_{2k}, \bar{\rho}_2, \bar{\rho}_3, \bar{\sigma}_{12}) + m(\boldsymbol{x}_{2k}\boldsymbol{y}_{2n-2}, \bar{\rho}_2, \bar{\rho}_3, \bar{\sigma}_{12})\\ + m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \bar{\rho}_{23}, \bar{\sigma}_{12}) + m(m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \bar{\rho}_2, \bar{\sigma}_{12}), \bar{\rho}_3).\end{aligned}$$

The first term on the right-hand side corresponds to the moduli space

$$\mathcal{M}(\boldsymbol{x}_{2n-2}\boldsymbol{y}_{2k}, \boldsymbol{x}_{2k-1}\boldsymbol{y}_1; \bar{\rho}_2, \bar{\rho}_3, \bar{\sigma}_{12}),$$

whose Maslov index is not one. The second vanishes because any domain containing $Q_2 + Q_3 + Q_4$ cannot have corners that contain $\boldsymbol{x}_{2k}$ and $\boldsymbol{y}_{2n-2}$. The last term also vanishes because the moduli space $\mathcal{M}(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \boldsymbol{a}\boldsymbol{y}_{2k}; \bar{\rho}_2, \bar{\sigma}_{12})$ has no holomorphic representative since the domain is an annulus and does not allow holomorphic involution, so $m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \bar{\rho}_2, \bar{\sigma}_{12}) = 0$. Hence, we have $m(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \bar{\rho}_{23}, \bar{\sigma}_{12}) = 0$ and $\#\mathcal{M}(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \boldsymbol{x}_{2k-1}\boldsymbol{y}_1; \bar{\rho}_{23}, \bar{\sigma}_{12}) = 0$ modulo two. □

The second interpretation is $\mathcal{M}(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \boldsymbol{x}_{2n-1}\boldsymbol{y}_1; \bar{\rho}_{23}, \bar{\sigma}_2, \bar{\sigma}_1)$. The domain is an annulus; each boundary consists of one $\boldsymbol{\alpha}$ curve segment and one $\boldsymbol{\beta}$ curve segment. The modulo two count of the moduli space can be computed by a similar computation above.

Lemma 3.4. *The modulo two count of the moduli space*

$$\mathcal{M}(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \boldsymbol{x}_{2n-1}\boldsymbol{y}_1; \bar{\rho}_{23}, \bar{\sigma}_2, \bar{\sigma}_1)$$

is one.

Proof. Again, we will consider the $\mathcal{A}_\infty$-compatibility relation as below:

$$\begin{aligned}0 &= m^2(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar\rho_2, \bar\rho_3, \bar\sigma_2, \bar\sigma_1)\\ &= m(m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}), \bar\rho_2, \bar\rho_3, \bar\sigma_2, \bar\sigma_1) + m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \mu(\bar\rho_2, \bar\rho_3), \bar\sigma_2, \bar\sigma_1)\\ &\quad + m(m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar\rho_2, \bar\sigma_2), \bar\rho_3, \bar\sigma_1) + m(m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar\rho_2, \bar\rho_3), \bar\sigma_2, \bar\sigma_1)\\ &\quad + m(m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar\sigma_2, \bar\sigma_1), \bar\rho_2, \bar\rho_3) + m(m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar\rho_2, \bar\rho_3, \bar\sigma_2), \bar\sigma_1)\\ &\quad + m(m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar\rho_2, \bar\sigma_2, \bar\sigma_1), \bar\rho_3) + m(m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar\rho_2, \bar\rho_3, \bar\sigma_2, \bar\sigma_1)).\end{aligned}$$

We have $m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}) = 0$ since there is no provincial domain connecting $\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}$, so the first term on the right-hand side vanishes. The fourth term also vanishes because $m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar\rho_2, \bar\rho_3) = 0$ (domain Q_2 is adjacent to both $\bar\rho_2$ and $\bar\sigma_2$). For the same reason, the fifth term vanishes.

In the sixth term, $m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar\rho_2, \bar\rho_3, \bar\sigma_2)$ does not represent a domain with four corners. Recall that a domain that involves $\bar\rho_2$ and $\bar\rho_3$ must have $\bar\sigma_1$. Thus, the sixth term vanishes. Similarly, the seventh term also vanishes. We have $m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar\rho_2, \bar\rho_3, \bar\sigma_2, \bar\sigma_1) = 0$ when considering the Maslov index.

Then the above compatibility relation is reduced to

$$m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar\rho_{23}, \bar\sigma_2, \bar\sigma_1) + m(m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar\rho_2, \bar\sigma_2), \bar\rho_3, \bar\sigma_1) = 0$$

The second term on the left-hand side equals $\boldsymbol{x}_{2n-1}\boldsymbol{y}_1 + \boldsymbol{x}_1\boldsymbol{y}_{2n-1}$. This implies modulo two count of the moduli spaces $\mathcal{M}(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \boldsymbol{x}_{2n-1}\boldsymbol{y}_1; \bar\rho_{23}, \bar\sigma_2, \bar\sigma_1)$ and $\mathcal{M}(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \boldsymbol{x}_1\boldsymbol{y}_{2n-1}; \bar\rho_{23}, \bar\sigma_2, \bar\sigma_1)$ equal one. □

However, idempotents of the type-*DD* module prohibit a nontrivial differential from moduli spaces considered above. Explicitly,

$$\begin{aligned}\delta^1(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}) &= \rho_{12}\otimes\sigma_{23}\otimes\boldsymbol{x}_{2n-1}\boldsymbol{y}_1 + \cdots\\ &= \rho_{12}\iota_1\otimes\sigma_{23}\otimes\boldsymbol{x}_{2n-1}\boldsymbol{y}_1 + \cdots\\ &= \rho_{12}\otimes\sigma_{23}\otimes\iota_1\boldsymbol{x}_{2n-1}\boldsymbol{y}_1 + \cdots.\end{aligned}$$

Recall that $\iota_1\boldsymbol{x}_{2n-1}\boldsymbol{y}_1 = 0$ since $\boldsymbol{x}_{2n-1}\boldsymbol{y}_1$ occupies $\alpha_1^{a,L}$ and the idempotent ι_1 also occupies the same α-arc.

The third interpretation is $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2n-1}, \boldsymbol{a}\boldsymbol{y}_1; \bar\rho_3, \bar\rho_2, \bar\sigma_{12})$. This is again an annulus and one of its boundaries has two $\boldsymbol{\alpha}$ curve segments and two $\boldsymbol{\beta}$ curve segments, thus it cannot give a nontrivial differential either.

Next, we will consider domain $Q_2 + Q_3 + R_1 + P_2 + \cdots + R_{2n-3} + Q_4$. Possible cuts may arise from a point between σ_2 and σ_3. The possible interpretations are

- $\mathcal{M}(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \boldsymbol{x}_1\boldsymbol{y}_{2n-1}; \bar\rho_{23}, \bar\sigma_2, \bar\sigma_1)$,
- $\mathcal{M}(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2k-1}, \boldsymbol{x}_1\boldsymbol{y}_{2k-1}; \bar\rho_{23}, \bar\sigma_{12})$.

By the above lemma, the modulo two count of the first moduli space is one, but because of idempotents, it cannot give a nontrivial contribution to the differential. The second moduli space has modulo two count zero by a similar computation in Lemma 3.3 or Lemma 3.4.

Algebra element containing ρ_{23}. Roughly speaking, the domains that possibly contribute to the algebra element ρ_{23} are obtained by adding regions to the domain $Q_1 + Q_2$ so that the resulting domain has at most four corners.

We will consider these domains by classifying them into three cases.

Case 1. We will first consider the following annular domains:

(3) $Q_1 + Q_2,$

(4) $Q_1 + Q_2 + Q_4 + R_{2n-3} + P_{2n-3} + R_{2n-4},$

(5) $Q_1+Q_2+Q_4+\sum_{k=0}^{l} R_{2n-2k-3}+P_{2n-2k-3}+R_{2n-2k-4}+P_{2n-2k-4}+R_{2n-2l-3},$

where $1 \le l \le n-2$. (Basically, these domains are obtained by adding an even number of regions to the top and bottom of $Q_1 + Q_2$.)

We will first consider the domain $Q_1 + Q_2$. The domain can be interpreted as $\mathcal{M}(\boldsymbol{ay}_{2n-2}, \boldsymbol{ab}; \bar{\rho}_{12}, \bar{\sigma}_2)$. Again, the modulo two count of the moduli space can be computed by using the $\mathcal{A}_\infty$-relation of $m^2(\boldsymbol{ay}_{2n-2}, \bar{\rho}_1, \bar{\rho}_2, \bar{\sigma}_2)$. Recall that $m(\boldsymbol{ay}_{2n-2}, \bar{\rho}_1) = \boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}$ and $m(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}, \bar{\rho}_2, \bar{\sigma}_2) = \boldsymbol{ab}$ since the associated domains are rectangular.

$$
\begin{aligned}
0 &= m(m(\boldsymbol{ay}_{2n-2}, \bar{\rho}_1), \bar{\rho}_2, \bar{\sigma}_2) + m(\boldsymbol{ay}_{2n-2}, (\bar{\rho}_1, \bar{\rho}_2), \bar{\sigma}_2) \\
&\qquad\qquad\qquad\qquad\qquad\qquad + m(m(\boldsymbol{ay}_{2n-2}, \bar{\sigma}_2), \bar{\rho}_1, \bar{\rho}_2) \\
&= \boldsymbol{ab} + m(\boldsymbol{ay}_{2n-2}, \bar{\rho}_{12}, \bar{\sigma}_2) + m(m(\boldsymbol{ay}_{2n-2}, \bar{\sigma}_2), \bar{\rho}_1, \bar{\rho}_2).
\end{aligned}
$$

The last term on the right-hand side equals zero because $m(\boldsymbol{ay}_{2n-2}, \bar{\sigma}_2) = 0$ (domain Q_2 is adjacent to Reeb chords $\bar{\rho}_2$ and $\bar{\sigma}_2$). This implies $m(\boldsymbol{ay}_{2n-2}, \bar{\rho}_{12}, \bar{\sigma}_2) = \boldsymbol{ab}$, hence $\#\mathcal{M}(\boldsymbol{ay}_{2n-2}, \boldsymbol{ab}; \bar{\rho}_{12}, \bar{\sigma}_2) = 1$.

Remark 3.5. An annulus domain of this kind (i.e., an outside boundary consisting of both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ curves and an inside boundary of $\boldsymbol{\alpha}$ curve only, including a cut on the inside boundary) always admits a holomorphic representative; since we are free to choose the length of the cut starting from the point so that the annulus admits a biholomorphic involution of it, again in the sense of [Ozsváth and Szabó 2004b, Lemma 9.4].

The moduli space $\mathcal{M}(\boldsymbol{ay}_{2n-2}, \boldsymbol{ab}; \bar{\rho}_{12}, \bar{\sigma}_2) = \mathcal{M}(\boldsymbol{ay}_{2n-2}, \boldsymbol{ab}; \rho_{23}, \sigma_2)$ corresponding to $\rho_{23} \otimes \sigma_2 \otimes \boldsymbol{ab}$ term occurs in $\delta^1(\boldsymbol{ay}_{2n-2})$ in $\widehat{CFDD}$. However, the right-hand side is zero because of the idempotents.

Likewise, the second and third domains (see (4) and (5) on the previous page) allow the following interpretations:

- $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2j}, \boldsymbol{a}\boldsymbol{y}_{2j+2}; \bar{\rho}_{12}, \bar{\sigma}_2, \bar{\sigma}_1)$,
- $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2j}, \boldsymbol{a}\boldsymbol{y}_{2j+2}; \bar{\rho}_{12}, \bar{\sigma}_{12})$.

First, $\#\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2j}, \boldsymbol{a}\boldsymbol{y}_{2j+2}; \bar{\rho}_{12}, \bar{\sigma}_2, \bar{\sigma}_1) = 1$ modulo two for reasons similar to those described in Remark 3.5. These contribute to the differential between generators $\boldsymbol{a}\boldsymbol{y}_{2j}$ and $\boldsymbol{a}\boldsymbol{y}_{2j+2}$ with an algebra element containing ρ_{23}, but all going to zero because of idempotents. (Similarly, $\#\mathcal{M}(\boldsymbol{a}\boldsymbol{b}, \boldsymbol{a}\boldsymbol{y}_2; \bar{\rho}_{12}, \bar{\sigma}_3, \bar{\sigma}_2, \bar{\sigma}_1) = 1$, but it does not affect the differential because of idempotents.)

Second, $\#\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2j}, \boldsymbol{a}\boldsymbol{y}_{2j+2}; \bar{\rho}_{12}, \bar{\sigma}_{12}) = 0$ modulo two. It can be proved by considering the following $\mathcal{A}_\infty$-relation:

$$\begin{aligned}0 &= m(m(\boldsymbol{a}\boldsymbol{y}_{2j}, \bar{\rho}_{12}, \bar{\sigma}_1, \bar{\sigma}_2)) + m(m(\boldsymbol{a}\boldsymbol{y}_{2j}, \bar{\sigma}_1), \bar{\rho}_{12}, \bar{\sigma}_2)\\ &\quad + m(m(\boldsymbol{a}\boldsymbol{y}_{2j}, \bar{\rho}_{12}), \bar{\sigma}_1, \bar{\sigma}_2) + m(m(\boldsymbol{a}\boldsymbol{y}_{2j}, \bar{\rho}_{12}, \bar{\sigma}_1), \bar{\sigma}_2) + m(\boldsymbol{a}\boldsymbol{y}_{2j}, \bar{\rho}_{12}, (\bar{\sigma}_1, \bar{\sigma}_2)).\end{aligned}$$

The term $m(\boldsymbol{a}\boldsymbol{y}_{2j}, \bar{\rho}_{12}, \bar{\sigma}_1, \bar{\sigma}_2) = 0$ since Maslov index is not one. We have that $m(\boldsymbol{a}\boldsymbol{y}_{2j}, \bar{\rho}_{12})$ and $m(\boldsymbol{a}\boldsymbol{y}_{2j}, \bar{\rho}_{12}, \bar{\sigma}_1)$ equal zero because $\bar{\sigma}_2$ was not involved and there is no such domain corresponding to these interpretations. From the diagram, it is clear that $m(\boldsymbol{a}\boldsymbol{y}_{2j}, \bar{\sigma}_1) = 0$. Thus, the last term $m(\boldsymbol{a}\boldsymbol{y}_{2j}, \bar{\rho}_{12}, (\bar{\sigma}_1, \bar{\sigma}_2)) = m(\boldsymbol{a}\boldsymbol{y}_{2j}, \bar{\rho}_{12}, \bar{\sigma}_{12})$ equals zero, too.

Case 2. Next, we will consider the following domains:

$$Q_1 + Q_2 + \sum_{k=0}^{l} P_{2n-2k-3} + R_{2n-2k-4} + P_{2n-2l-5},$$

$$Q_1 + Q_2 + Q_4 + \sum_{k=0}^{l} P_{2n-2k-3} + R_{2n-2k-4} + P_{2n-2l-5} + \sum_{k=0}^{m} R_{2n-2k-3} + P_{2n-2k-4} + R_{2n-2m-5},$$

$$Q_1 + Q_2 + Q_4 + \sum_{k=0}^{l} P_{2n-2k-3} + R_{2n-2k-4} + \sum_{k=0}^{m} R_{2n-2k-3} + P_{2n-2k-4}, \quad \text{and}$$

$$Q_1 + Q_2 + Q_4 + Q_5 + \sum_{k=0}^{n-3} P_{2n-2k-3} + R_{2n-2k-4} + \sum_{k=0}^{m} R_{2n-2k-3} + P_{2n-2k-4},$$

where $0 \le l, m \le n-3$. These domains are obtained by adding a topologically rectangular domain containing $Q_1 + Q_2$ and another rectangular domain containing Q_4.

The first domain can have a cut at a point between ρ_2 and ρ_3. The interpretation

$$\mathcal{M}(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}, \boldsymbol{x}_{2k}\boldsymbol{b}; \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_2)$$

is essentially a rectangle so modulo two count of the corresponding moduli space is one. The second domain can have cuts at two different points; a point between

ρ_2 and ρ_3, and a point between σ_2 and σ_3. Considering the interpretation that has only one cut, the domain is an annulus with one of its boundary consisting of two $\boldsymbol{\alpha}$ curve segments and two $\boldsymbol{\beta}$ curve segments, which does not allow any holomorphic representative. If the interpretation has both of the cuts, then it is also a rectangular domain of the moduli space $\mathcal{M}(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2l-1}, \boldsymbol{x}_{2k}\boldsymbol{y}_{2l}; \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_2, \bar{\sigma}_1)$. Dualizing them, they yield a nontrivial differential of algebra elements $\rho_{23} \otimes \sigma_2$ and $\rho_{23} \otimes \sigma_{23}$ for the type-D structure map δ^1 in $\widehat{CFDD}$.

Remark 3.6. Both of the domains considered above have interpretations without any cut. However, those interpretations do not have a holomorphic representative. For example, we can see that the modulo two count of the moduli space $\mathcal{M}(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2l-1}, \boldsymbol{x}_{2k}\boldsymbol{y}_{2l}; \bar{\rho}_{12}, \bar{\sigma}_{12})$ equals zero by considering an $\mathcal{A}_\infty$-relation similar to that discussed in Lemmas 3.3 and 3.4.

The third domain has almost the same interpretation; the only meaningful one is

$$\mathcal{M}(\boldsymbol{x}_{2l}\boldsymbol{y}_{2k}, \boldsymbol{x}_{2l+1}\boldsymbol{y}_{2k+1}; \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_2, \bar{\sigma}_1).$$

Again, this interpretation is rectangular and modulo two count of the moduli space is one.

The last domain has two interpretations with Maslov index one. They are

$$\mathcal{M}(\boldsymbol{x}_{2l}\boldsymbol{b}, \boldsymbol{x}_{2l+1}\boldsymbol{y}_1; \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_3, \bar{\sigma}_2, \bar{\sigma}_1)$$

and

$$\mathcal{M}(\boldsymbol{x}_{2l}\boldsymbol{b}, \boldsymbol{x}_{2l+1}\boldsymbol{y}_1; \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_{123}).$$

The first interpretation is clearly a rectangle. However, the second one is topologically a punctured torus. To count the signed number of the moduli space, we investigate the $\mathcal{A}_\infty$-relation $m^2(\boldsymbol{x}_{2l}\boldsymbol{b}, \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_{12}, \bar{\sigma}_3) = 0$.

Lemma 3.7. *The modulo two count of the moduli space*

$$\mathcal{M}(\boldsymbol{x}_{2l}\boldsymbol{b}, \boldsymbol{x}_{2l+1}\boldsymbol{y}_1; \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_{123})$$

is one.

Proof. Disregarding all terms that equal zero, the relation is reduced to

$$m(\boldsymbol{x}_{2l}\boldsymbol{b}, \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_{123}) + m(m(\boldsymbol{x}_{2l}\boldsymbol{b}, \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_{12}), \bar{\sigma}_3) = 0.$$

$m(\boldsymbol{x}_{2l}\boldsymbol{b}, \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_{12}) = \boldsymbol{x}_{2l+2}\boldsymbol{b}$ because the corresponding domain is an annulus as in Remark 3.5. Thus, the relation is reduced to

$$m(\boldsymbol{x}_{2l}\boldsymbol{b}, \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_{123}) + m(\boldsymbol{x}_{2l+2}\boldsymbol{b}, \bar{\sigma}_3) = 0.$$

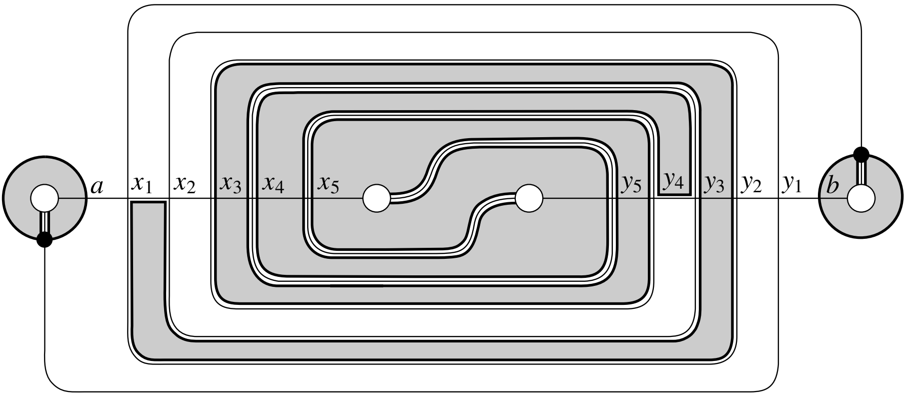


Figure 4. A diagram of the $(2, 6)$-torus link complement. The shaded region is a domain obtained by adding a rectangular domain to Q_2. This domain corresponds to a differential from $\boldsymbol{x}_1\boldsymbol{y}_3$ to $\boldsymbol{x}_2\boldsymbol{y}_2$. Cutting along the bold curve on the boundary of the domain, the domain turns out to be rectangular.

The second term of the right-hand side is clearly $\boldsymbol{x}_{2l+1}\boldsymbol{y}_1 + \boldsymbol{x}_1\boldsymbol{y}_{2l+1}$. This implies modulo two count of the moduli space

$$\mathcal{M}(\boldsymbol{x}_{2l}\boldsymbol{b}, \boldsymbol{x}_{2l+1}\boldsymbol{y}_1; \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_{123})$$

equals one. $\square$

Therefore, the two interpretations of the last domain result in the two same terms $\rho_{23} \otimes \sigma_{123} \otimes \boldsymbol{x}_{2l+1}\boldsymbol{y}_1$ in the $\widehat{CFDD}$ module; so they do not contribute to the differential.

Remark 3.8. Lemma 3.7 also proves that modulo two count of

$$\mathcal{M}(\boldsymbol{x}_{2l}\boldsymbol{b}, \boldsymbol{x}_1\boldsymbol{y}_{2l+1}; \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_{123})$$

also equals one.

Case 3. Domains that possibly contribute to a differential with an algebra element containing ρ_{23} are obtained by adding domains to the top of $Q_1 + Q_2$. That is, we add $2j - 1$ domains, $j = 1, \ldots, n-1$ on the top and the resulting domain is $R_{2n-2j-1} + \cdots + R_{2n-3} + Q_1 + Q_2$. The only possible interpretation is $\mathcal{M}(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-2j-1}, \boldsymbol{x}_{2n-2j}\boldsymbol{b}; \bar{\rho}_{12}, \bar{\sigma}_2)$. It does not allow any holomorphic representative because the domain does not allow any holomorphic involution interchanging two boundaries.

Likewise, we shall consider domains obtained by adding domains to Q_2 on the top and bottom. Consider a domain

$$Q_1 + Q_2 + Q_4 + (R_{2n-k} + \cdots + R_{2n-3}) + (P_{2n-l} + \cdots + P_{2n-3}).$$

The domain is obtained by adding $k-2$ domains on the top and $l-2$ domains on the bottom of $Q_1+Q_2+Q_4$ (k and l should have the same parity). If $k=l$, then the resulting domain is the domain that we have considered in Case 2. (See the bottom right of Figure 5.) If $k \neq l$, then three interpretations are possible. The first is $\mathcal{M}(\boldsymbol{x}_{2n-l-2}\boldsymbol{y}_{2n-k-2}, \boldsymbol{x}_{2n-k-1}\boldsymbol{y}_{2n-l-1}; \bar{\rho}_{12}, \bar{\sigma}_{12})$. This is a genus two domain, and modulo two count of this moduli space is zero by a similar reason given in Lemma 3.3. The second and third interpretations are

$$\mathcal{M}(\boldsymbol{x}_{2n-l-2}\boldsymbol{y}_{2n-k-2}, \boldsymbol{x}_{2n-k-1}\boldsymbol{y}_{2n-l-1}; \bar{\rho}_{12}, \bar{\sigma}_2, \bar{\sigma}_1),$$
$$\mathcal{M}(\boldsymbol{x}_{2n-l-2}\boldsymbol{y}_{2n-k-2}, \boldsymbol{x}_{2n-k-1}\boldsymbol{y}_{2n-l-1}; \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_{12}).$$

These are both annular interpretations, and they do not have any holomorphic representative because they do not allow a holomorphic involution.

Lastly, if $k=2n-2$, then, the domain contains Q_5. Then this domain has the two interpretations

$$\mathcal{M}(\boldsymbol{x}_{2l-2}\boldsymbol{b}, \boldsymbol{x}_1\boldsymbol{y}_{2l-1}; \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_{123}) \quad \text{and} \quad \mathcal{M}(\boldsymbol{x}_{2l-2}\boldsymbol{b}, \boldsymbol{x}_1\boldsymbol{y}_{2l-1}; \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_3, \bar{\sigma}_{12}).$$

The signed number of the moduli space of the first interpretation was proved to be one modulo two by Lemma 3.7. The signed number of the second interpretation is not one because it does not allow a holomorphic involution either.

To sum up, the differentials that give the algebra element containing ρ_{23} are listed below:

- $\boldsymbol{x}_i\boldsymbol{y}_j \mapsto \rho_{23}\otimes\sigma_{23}\otimes\boldsymbol{x}_{i+1}\boldsymbol{y}_{j+1}$ if $i, j \neq 2n-1$,
- $\boldsymbol{x}_i\boldsymbol{y}_{2n-1} \mapsto \rho_{23}\otimes\sigma_2\otimes\boldsymbol{x}_{i+1}\boldsymbol{b}$ if $j=2n-1$ and $i=1,3,\dots,2n-3$,
- $\boldsymbol{x}_i\boldsymbol{b} \mapsto \rho_{23}\otimes\sigma_{123}\otimes\boldsymbol{x}_1\boldsymbol{y}_{i+1}$.

Algebra element contains ρ_{123}. Domains that possibly contribute to the algebra element ρ_{123} are listed below:

$$(Q_1+Q_2+Q_3+Q_4)+(R_1+P_2+R_3+P_4+\cdots+R_{2n-5}+P_{2n-4})+R_{2n-3};$$
$$(Q_1+Q_2+Q_3+Q_4)+(P_1+\cdots+P_{2n-3});$$
$$(Q_1+Q_2+Q_3+Q_4)+\sum_{k=0}^{l}(R_{2k+1}+P_{2k+2}) + \sum_{k=0}^{n-4-l}(R_{2n-2k-3}+P_{2n-2k-3}+R_{2n-2k-4}+P_{2n-2k-4})+R_{2l+3};$$
$$(Q_1+Q_2+Q_3+Q_4)+\sum_{k=0}^{l}(P_{2k+1}+P_{2k+2}) + \sum_{k=0}^{n-4-l}(R_{2n-2k-3}+P_{2n-2k-3}+R_{2n-2k-4}+P_{2n-2k-4})+P_{2l+3}; \quad \text{and}$$
$$Q_1+\cdots+Q_5+P_1+\cdots+P_{2n-3}+R_1+\cdots+R_{2n-3},$$

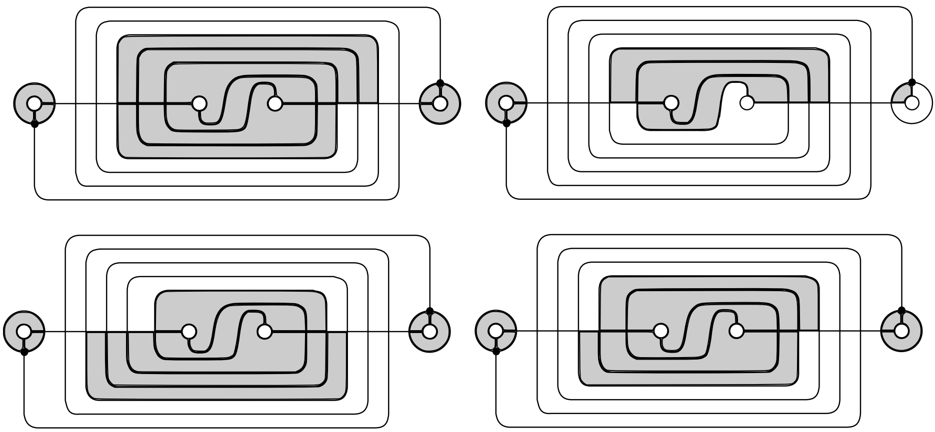


Figure 5. Examples of obtaining nonrectangular domains of the (2, 6)-torus link. Top left can be interpreted as an annular domain, but it cannot give a nontrivial differential due to idempotents. Top right is obtained by adding a domain to Q_2 on the top, but its only possible interpretation does not allow any holomorphic representative. Bottom left and bottom right are obtained by adding domains to Q_2 on the top and bottom. If the number of regions attached on the top is not equal to the number of regions attached on the bottom, it has two interpretations; and they do not allow a holomorphic representative either (bottom left). If two numbers are equal, then the domain gives a nontrivial differential. (This case was previously considered. See Figure 4.)

where $1 \le l \le n-3$.

Each of these domains are obtained by adding a rectangular domain containing a region adjacent to ρ_1 to the annular domain listed in the algebra element containing ρ_{23}.

We investigate the first domain. As before, we list all possible interpretations:

- $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2n-2}, \boldsymbol{x}_1\boldsymbol{y}_{2n-1}; \bar{\rho}_{123}, \bar{\sigma}_2, \bar{\sigma}_1)$,
- $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2n-2}, \boldsymbol{x}_1\boldsymbol{y}_{2n-1}; \bar{\rho}_3, \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_2, \bar{\sigma}_1)$,
- $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2n-2}, \boldsymbol{x}_1\boldsymbol{y}_{2n-1}; \bar{\rho}_{23}, \bar{\rho}_1, \bar{\sigma}_2, \bar{\sigma}_1)$,
- $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2n-2}, \boldsymbol{x}_1\boldsymbol{y}_{2n-1}; \bar{\rho}_3, \bar{\rho}_{12}, \bar{\sigma}_2, \bar{\sigma}_1)$.

The third interpretation is an annulus whose outer boundary has two $\boldsymbol{\alpha}$ curve segments and two $\boldsymbol{\beta}$ curve segments; thus it does not have a holomorphic representative. The fourth interpretation cannot give a nontrivial contribution either because of the $\mathcal{A}_\infty$-module compatibility relation. On the other hand, the second interpretation is a rectangular one; it allows a holomorphic representative and its modulo two count of the moduli space is one. The first interpretation also has a moduli space with modulo two count one by the same analysis in Lemma 3.7. Again, the first and second interpretations will result in the same term after in $\widehat{CFDD}$ module. The sum of these two terms equals zero, so this domain actually has no contribution after all.

The second domain has two interpretations; $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2n-2}, \boldsymbol{x}_{2n-1}\boldsymbol{y}_1; \bar{\rho}_{123}, \bar{\sigma}_2, \bar{\sigma}_1)$ and $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2n-2}, \boldsymbol{x}_{2n-1}\boldsymbol{y}_1; \bar{\rho}_3, \bar{\rho}_{12}, \bar{\sigma}_2, \bar{\sigma}_1)$. The first interpretation was considered in the above computation, and the second interpretation is an annulus whose outer boundary consists of two $\boldsymbol{\alpha}$ curve segments and two $\boldsymbol{\beta}$ curve segments, so there is no holomorphic representative.

Similarly, the other domains (except for the last) give Whitney disks, and the moduli spaces corresponding to the domains are $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2j}, \boldsymbol{x}_1\boldsymbol{y}_{2j+1}; \bar{\rho}_{123}, \bar{\sigma}_2, \bar{\sigma}_1)$, $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2j}, \boldsymbol{x}_1\boldsymbol{y}_{2j+1}; \bar{\rho}_3, \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_2, \bar{\sigma}_1)$ and $\mathcal{M}(\boldsymbol{a}\boldsymbol{y}_{2j}, \boldsymbol{x}_{2j+1}\boldsymbol{y}_1; \bar{\rho}_{123}, \bar{\sigma}_2, \bar{\sigma}_1)$. The signed number of each of these moduli spaces is one modulo two.

The moduli space of the last domain $Q_1 + \cdots + Q_5 + P_1 + \cdots + P_{2n-3} + Q_1 + \cdots + Q_{2n-3}$ can be interpreted in four ways. The first is $\mathcal{M}(\boldsymbol{a}\boldsymbol{b}, \boldsymbol{x}_1\boldsymbol{y}_1; \bar{\rho}_{123}, \bar{\sigma}_{123})$ whose Maslov index is different from one. The second possible interpretation is

$$\mathcal{M}(\boldsymbol{a}\boldsymbol{b}, \boldsymbol{x}_1\boldsymbol{y}_1; \bar{\rho}_{123}, \bar{\sigma}_3, \bar{\sigma}_2, \bar{\sigma}_1)$$

$\mathcal{A}_\infty$-relation of $m^2(\boldsymbol{a}\boldsymbol{b}, \bar{\rho}_{12}, \bar{\rho}_3, \bar{\sigma}_3, \bar{\sigma}_2, \bar{\sigma}_1)$ gives $m(\boldsymbol{a}\boldsymbol{b}, \bar{\rho}_{123}, \bar{\sigma}_3, \bar{\sigma}_2, \bar{\sigma}_1) = \boldsymbol{x}_1\boldsymbol{y}_1$, by considering $m(\boldsymbol{a}\boldsymbol{b}, \bar{\rho}_{12}, \bar{\sigma}_3, \bar{\sigma}_2, \bar{\sigma}_1) = \boldsymbol{a}\boldsymbol{y}_2$ and $m(\boldsymbol{a}\boldsymbol{y}_2, \bar{\rho}_3) = \boldsymbol{x}_1\boldsymbol{y}_1$. Thus, the modulo two count of the moduli space is one. The third interpretation

$$\mathcal{M}(\boldsymbol{a}\boldsymbol{b}, \boldsymbol{x}_1\boldsymbol{y}_1; \bar{\rho}_3, \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_{123})$$

can be done precisely in the same way. The last interpretation is

$$\mathcal{M}(\boldsymbol{a}\boldsymbol{b}, \boldsymbol{x}_1\boldsymbol{y}_1; \bar{\rho}_3, \bar{\rho}_2, \bar{\rho}_1, \bar{\sigma}_3, \bar{\sigma}_2, \bar{\sigma}_1)$$

The existence of a holomorphic curve and its modulo two count is quite clear from the diagram; the domain is essentially rectangular in this interpretation.

It is worth mentioning that there are three moduli spaces contributing to $\rho_{123}\sigma_{123} \otimes \boldsymbol{x}_1\boldsymbol{y}_1$ term in $\delta^1(\boldsymbol{a}\boldsymbol{b})$.

To sum up, we have the following nontrivial differentials of the algebra element containing ρ_{123}:

- $\boldsymbol{a}\boldsymbol{y}_{2k} \mapsto \rho_{123} \otimes \sigma_{23} \otimes \boldsymbol{x}_{2k+1}\boldsymbol{y}_1$,
- $\boldsymbol{a}\boldsymbol{b} \mapsto \rho_{123} \otimes \sigma_{123} \otimes \boldsymbol{x}_1\boldsymbol{y}_1$.

For the algebra elements containing σ_{12}, σ_{23} and σ_{123}, those differentials can be computed in a parallel manner by taking advantage of the symmetry of the diagram.

We close this section by summarizing the computation:

Proposition 3.9. *Let $\mathcal{H}$ be a bordered Heegaard diagram of $(2,2n)$-torus link complement in S^3 as in Figure 2. Then, $\widehat{CFDD}(\mathcal{H})$ has the following generators*:

- $\boldsymbol{x}_i\boldsymbol{y}_j$, *where* $1 \le i, j \le 2n-1$ *and* $i = j$ *modulo two*,
- $\boldsymbol{ab}$,
- $\boldsymbol{a}\boldsymbol{y}_k$, *where* $k = 2, 4, \dots, 2n-2$,
- $\boldsymbol{x}_k\boldsymbol{b}$, *where* $k = 2, 4, \dots, 2n-2$.

The map $\delta^1 : \mathfrak{S}(\mathcal{H}) \to \mathcal{A}(-\mathcal{Z}_L) \otimes \mathcal{A}(-\mathcal{Z}_R) \otimes \mathfrak{S}(\mathcal{H})$ *is computed in the following way.*

- *For* $\boldsymbol{x}_i\boldsymbol{y}_j$, *if* $i, j \neq 2n-1$,

$$\boldsymbol{x}_i\boldsymbol{y}_j \mapsto \begin{cases} \boldsymbol{x}_{j-1}\boldsymbol{y}_{i+1} + \boldsymbol{x}_{i+1}\boldsymbol{y}_{j-1} + \rho_{23}\sigma_{23}\boldsymbol{x}_{i+1}\boldsymbol{y}_{j+1} & \text{if } j-i > 2,\\ \boldsymbol{x}_{j+1}\boldsymbol{y}_{i-1} + \boldsymbol{x}_{i-1}\boldsymbol{y}_{j+1} + \rho_{23}\sigma_{23}\boldsymbol{x}_{i+1}\boldsymbol{y}_{j+1} & \text{if } i-j > 2,\\ \boldsymbol{x}_{i+1}\boldsymbol{y}_{j-1} + \rho_{23}\sigma_{23}\boldsymbol{x}_{i+1}\boldsymbol{y}_{j+1} & \text{if } j-i = 2,\\ \boldsymbol{x}_{i-1}\boldsymbol{y}_{j+1} + \rho_{23}\sigma_{23}\boldsymbol{x}_{i+1}\boldsymbol{y}_{j+1} & \text{if } i-j = 2,\\ \rho_{23}\sigma_{23}\boldsymbol{x}_{i+1}\boldsymbol{y}_{j+1} & \text{if } i = j. \end{cases}$$

- *If* $j = 2n-1$ *and* $i = 1, 3, \dots, 2n-3$,

$$\boldsymbol{x}_i\boldsymbol{y}_j \mapsto \begin{cases} \boldsymbol{x}_{j-1}\boldsymbol{y}_{i+1} + \boldsymbol{x}_{i+1}\boldsymbol{y}_{j-1} + \rho_{23}\sigma_2\boldsymbol{x}_{i+1}\boldsymbol{b} & \text{if } j-i > 2,\\ \boldsymbol{x}_{i+1}\boldsymbol{y}_{j-1} + \rho_{23}\sigma_2\boldsymbol{x}_{i+1}\boldsymbol{b} & \text{if } i = 2n-3. \end{cases}$$

- *If* $i = 2n-1$ *and* $j = 1, 3, \dots, 2n-3$,

$$\boldsymbol{x}_i\boldsymbol{y}_j \mapsto \begin{cases} \boldsymbol{x}_{j+1}\boldsymbol{y}_{i-1} + \boldsymbol{x}_{i-1}\boldsymbol{y}_{j+1} + \rho_2\sigma_{23}\boldsymbol{a}\boldsymbol{y}_{j+1} & \text{if } i-j > 2,\\ \boldsymbol{x}_{j+1}\boldsymbol{y}_{i-1} + \rho_2\sigma_{23}\boldsymbol{a}\boldsymbol{y}_{j+1} & \text{if } j = 2n-3. \end{cases}$$

- $\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1} \mapsto \rho_2\sigma_2\boldsymbol{ab}$.
- $\boldsymbol{a}\boldsymbol{y}_j \mapsto$

$$\begin{cases} \rho_1\boldsymbol{x}_1\boldsymbol{y}_1 + \rho_3(\boldsymbol{x}_{2n-1}\boldsymbol{y}_3 + \boldsymbol{x}_3\boldsymbol{y}_{2n-1}) + \rho_{123}\sigma_{23}\boldsymbol{x}_3\boldsymbol{y}_1 & \text{if } j = 2,\\ \rho_1(\boldsymbol{x}_1\boldsymbol{y}_{2n-3} + \boldsymbol{x}_{2n-3}\boldsymbol{y}_1) + \rho_3\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1} + \rho_{123}\sigma_{23}\boldsymbol{x}_{2n-1}\boldsymbol{y}_1 & \text{if } j = 2n-2,\\ \rho_1(\boldsymbol{x}_1\boldsymbol{y}_{j-1} + \boldsymbol{x}_{j-1}\boldsymbol{y}_1) + \rho_3(\boldsymbol{x}_{2n-1}\boldsymbol{y}_{j+1} + \boldsymbol{x}_{j+1}\boldsymbol{y}_{2n-1}) + \rho_{123}\sigma_{23}\boldsymbol{x}_{j+1}\boldsymbol{y}_1 & \\ & \text{otherwise.} \end{cases}$$

- $\boldsymbol{x}_i\boldsymbol{b} \mapsto$

$$\begin{cases} \sigma_1\boldsymbol{x}_1\boldsymbol{y}_1 + \sigma_3(\boldsymbol{x}_3\boldsymbol{y}_{2n-1} + \boldsymbol{x}_{2n-1}\boldsymbol{y}_3) + \rho_{23}\sigma_{123}\boldsymbol{x}_1\boldsymbol{y}_3 & \text{if } i = 2,\\ \sigma_1(\boldsymbol{x}_{2n-3}\boldsymbol{y}_1 + \boldsymbol{x}_1\boldsymbol{y}_{2n-3}) + \sigma_3\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1} + \rho_{23}\sigma_{123}\boldsymbol{x}_1\boldsymbol{y}_{2n-1} & \text{if } i = 2n-2,\\ \sigma_1(\boldsymbol{x}_{i-1}\boldsymbol{y}_1 + \boldsymbol{x}_1\boldsymbol{y}_{i-1}) + \sigma_3(\boldsymbol{x}_{i+1}\boldsymbol{y}_{2n-1} + \boldsymbol{x}_{2n-1}\boldsymbol{y}_{i+1}) + \rho_{23}\sigma_{123}\boldsymbol{x}_1\boldsymbol{y}_{i+1} & \\ & \text{otherwise.} \end{cases}$$

- $\boldsymbol{ab} \mapsto (\rho_1\sigma_3 + \rho_3\sigma_1)(\boldsymbol{x}_1\boldsymbol{y}_{2n-1} + \boldsymbol{x}_{2n-1}\boldsymbol{y}_1) + \rho_{123}\sigma_{123}\boldsymbol{x}_1\boldsymbol{y}_1$.

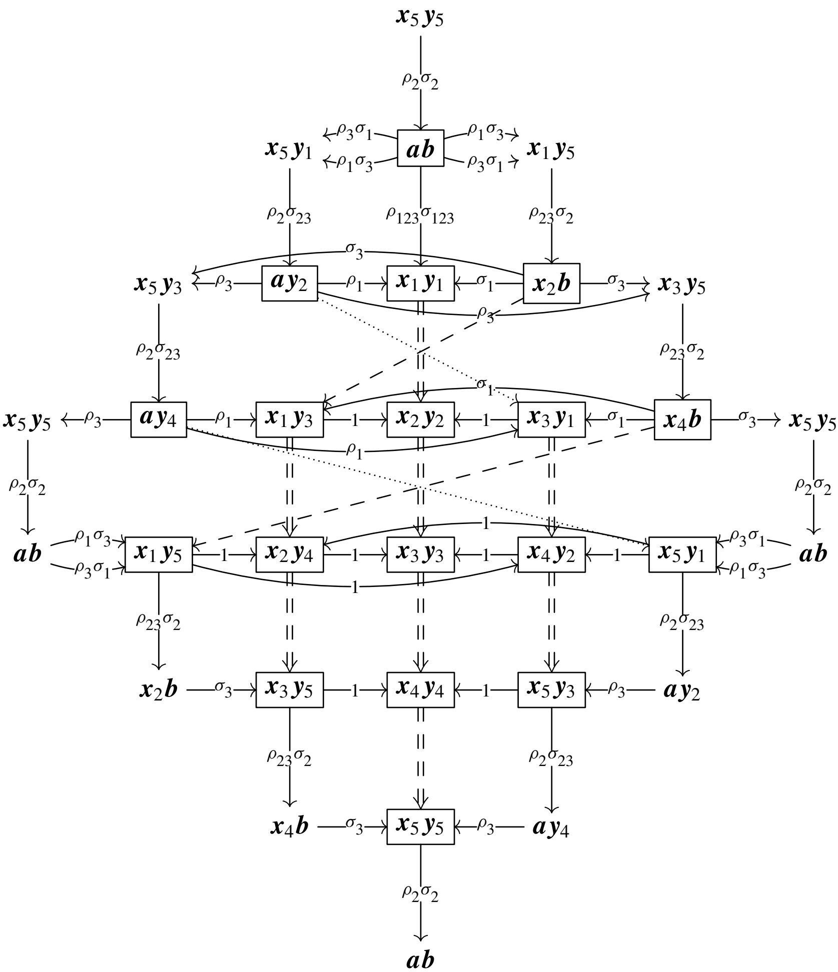


Figure 6. A diagram of the $(2, 6)$-torus link complement. The arrows emanating from a generator in the box show the resulting terms of the differential of the generator. The dashed arrow represents algebra element $\rho_{23}\sigma_{123}$, the dotted arrow $\rho_{123}\sigma_{23}$, and the doubly dashed arrow $\rho_{23}\sigma_{23}$. Other algebra elements are written on the arrows.

4. Examples

In this section, we will relate our result to the known calculation for knot complements and closed 3-manifolds. These examples show how to use the algebraic structure of the pairing theorem given in [Lipshitz et al. 2008].

Derived tensor product of bimodule. The pairing of modules associated to a single boundary three-manifold is well studied in [Lipshitz et al. 2008]. In this section, we will be using the pairing theorem of doubly bordered cases. There are many versions of the pairing theorem depending on the types of bimodules [Lipshitz et al. 2015, Theorem 2], but for our purpose, the pairing of a type-A module and type-DD module will suffice.

The pairing of bimodules associated to double bordered three-manifold is also similar to the single boundary case; the only difference is the framed arc $\boldsymbol{z}$. If we glue a doubly bordered diagram and a single boundary diagram together, we match the marked point z from the single boundary diagram with one end of the framed arc $\boldsymbol{z}$. After pairing, the framed arc is reduced to a marked point on the other side of the boundary (when pairing two doubly bordered diagrams, then we connect the two framed arcs). In our example, we will be mainly interested in a type-D structure obtained by the derived tensor product $\widehat{CFA}(\mathcal{H}_1)\widetilde{\otimes}_{\mathcal{A}(\mathcal{Z})}\widehat{CFDD}(\mathcal{H}_2)$, where a single boundary diagram $\mathcal{H}_1$ is glued on the right side of a doubly bordered diagram $\mathcal{H}_2$. The resulting type-D structure map $(\delta')^1$ is

$$(\delta')^1 = \sum_{k=1}^{\infty}((m_R)_{k+1} \otimes \mu_L \otimes \mathbb{I}_{\widehat{CFDD}})(\boldsymbol{x} \otimes \delta^k(\boldsymbol{y}))$$

where $\boldsymbol{x} \in \mathfrak{S}(\mathcal{H}_1)$ and $\boldsymbol{y} \in \mathfrak{S}(\mathcal{H}_2)$.

Infinity-surgery on right component of link. First, we will consider an ∞-surgery on the right component of the $(2, 2n)$-torus link complement. Since the longitudes $\alpha_1^{a,L}$ and $\alpha_1^{a,R}$ of the left and right components are passing through β_1 and β_2 respectively, the ∞-surgery on the right components gives an unknot complement with framing $(n-1)$. We compute $\widehat{CFD}$ of the unknot complement as follows.

Let $\mathcal{H}_{(2,2n)}$ be a doubly bordered diagram of the $(2, 2n)$-torus link complement, and $\mathcal{H}_\infty$ be a single bordered diagram of a solid torus with ∞-framing. Then, the generator set $\mathfrak{S}(\mathcal{H}_\infty \cup_\partial \mathcal{H}_{(2,2n)})$ consists of $\boldsymbol{w} \otimes \boldsymbol{ab}$ and $\boldsymbol{w} \otimes \boldsymbol{x}_{2k}\boldsymbol{b}$, $k = 1, \dots, n-1$.

Computing $\widehat{CFA}(\mathcal{H}_\infty)$ is easy; that is,

$$m_{k+3}(\boldsymbol{w}, \sigma_3, \underbrace{\sigma_{23}, \dots, \sigma_{23}}_{k\text{-times}}, \sigma_2) = \boldsymbol{w}.$$

Now, we shall consider the type-D structure of $\widehat{CFDD}(\mathcal{H}_{(2,2n)})$. We omit the terms which do not appear after taking box tensor product with $\widehat{CFA}(\mathcal{H}_\infty)$; thus, they have no contribution in computing $\widehat{CFA}(\mathcal{H}_\infty)\widetilde{\otimes}_{\mathcal{A}(\mathcal{Z})}\widehat{CFDD}(\mathcal{H}_{(2,2n)})$.

$$\begin{aligned}
\delta^2(\boldsymbol{ab}) &= (\rho_1 \otimes \rho_{23}) \otimes (\sigma_3 \otimes \sigma_2) \otimes \boldsymbol{x}_2\boldsymbol{b} + \cdots, \\
\delta^2(\boldsymbol{x}_{2k}\boldsymbol{b}) &= (\rho_{23}) \otimes (\sigma_3 \otimes \sigma_2) \otimes \boldsymbol{x}_{2k+2}\boldsymbol{b} + \cdots \qquad \text{for } k = 1, \dots, n-2, \\
\delta^2(\boldsymbol{x}_{2n-2}\boldsymbol{b}) &= (\rho_2) \otimes (\sigma_3 \otimes \sigma_2) \otimes \boldsymbol{ab} + \cdots.
\end{aligned}$$

Thus, the type-D structure $(\delta')^1$ is

$$(\delta')^1(\boldsymbol{w}\otimes\boldsymbol{ab}) = \mu(\rho_1\otimes\rho_{23})\otimes m_3(\boldsymbol{w},\sigma_3,\sigma_2)\otimes\boldsymbol{x}_2\boldsymbol{b} = \rho_{123}\otimes\boldsymbol{w}\otimes\boldsymbol{x}_2\boldsymbol{b},$$

$$(\delta')^1(\boldsymbol{w}\otimes\boldsymbol{x}_{2k}\boldsymbol{b}) = \mu(\rho_{23})\otimes m_3(\boldsymbol{w},\sigma_3,\sigma_2)\otimes\boldsymbol{x}_{2k+2}\boldsymbol{b} = \rho_{23}\otimes\boldsymbol{w}\otimes\boldsymbol{x}_{2k+2}\boldsymbol{b} \quad \text{for } k = 1,\dots,n-2,$$

$$(\delta')^1(\boldsymbol{w}\otimes\boldsymbol{x}_{2n-2}\boldsymbol{b}) = \mu(\rho_2)\otimes m_3(\boldsymbol{w},\sigma_3,\sigma_2)\otimes\boldsymbol{ab} = \rho_2\otimes\boldsymbol{w}\otimes\boldsymbol{ab}.$$

Compare this result with [Hom 2011, Example 2.2].

Knot complement of trefoil. Consider the (2, 4)-torus link complement. If we glue the right component with a solid torus of framing $+2$, then the resulting manifold will be diffeomorphic to a trefoil complement (after handleslide and blowing down the +1 unknot component). A type-D structure

$$(N_1, (\delta_1)^1) := \widehat{CFA}(\mathcal{H}_{+2})\widetilde{\otimes}_{\mathcal{A}(\mathcal{Z})}\widehat{CFDD}(\mathcal{H}_{(2,4)})$$

computes as

$$\begin{tikzcd}
\boldsymbol{p}_1\otimes\boldsymbol{ab} \arrow[ddrr, dashed] & \boldsymbol{q}\otimes\boldsymbol{x}_3\boldsymbol{y}_3 \arrow[l, "\rho_2"'] & \boldsymbol{q}\otimes\boldsymbol{a}\boldsymbol{y}_2 \arrow[l, "\rho_3"'] \arrow[d, "\rho_1"] \\
 & & \boldsymbol{q}\otimes\boldsymbol{x}_1\boldsymbol{y}_1 \\
 & & \boldsymbol{p}_2\otimes\boldsymbol{ab} \arrow[u, "\rho_{123}"']
\end{tikzcd}$$

The dashed line is called an *unstable chain*, where

$$\begin{tikzcd}
\cdots \arrow[r] & \boldsymbol{p}_1\otimes\boldsymbol{ab} \arrow[r, "\rho_{123}"'] & \boldsymbol{p}_2\otimes\boldsymbol{x}_2\boldsymbol{b} \arrow[r, "\rho_{23}"'] \arrow[d, "\rho_{23}"'] & \boldsymbol{q}\otimes\boldsymbol{x}_1\boldsymbol{y}_3 \arrow[r, "\rho_{23}"'] \arrow[d, "1"'] & \boldsymbol{p}_1\otimes\boldsymbol{x}_2\boldsymbol{b} \arrow[r, "\rho_2"'] & \boldsymbol{p}_2\otimes\boldsymbol{ab} \arrow[r] & \cdots \\
 & & \boldsymbol{q}\otimes\boldsymbol{x}_3\boldsymbol{y}_1 \arrow[r, "1"'] & \boldsymbol{q}\otimes\boldsymbol{x}_2\boldsymbol{y}_2 & & &
\end{tikzcd}$$

We claim that the chain complex described above is homotopy equivalent to a complex $(N_2, (\delta_2)^1)$, which is identical to the complex above except for the unstable complex that has been replaced by

$$\cdots \to \boldsymbol{p}_1\otimes\boldsymbol{ab} \xrightarrow[\rho_{123}]{} \boldsymbol{p}_2\otimes\boldsymbol{x}_2\boldsymbol{b} \xrightarrow[\rho_{23}]{} \boldsymbol{q}\otimes\boldsymbol{x}_1\boldsymbol{y}_3 \xrightarrow[\rho_{23}]{} \boldsymbol{p}_1\otimes\boldsymbol{x}_2\boldsymbol{b} \xrightarrow[\rho_2]{} \boldsymbol{p}_2\otimes\boldsymbol{ab} \to \cdots$$

Define a map $\pi : N_1 \to N_2$ such that $\pi(\boldsymbol{q}\otimes\boldsymbol{x}_3\boldsymbol{y}_1) = 0$, $\pi(\boldsymbol{q}\otimes\boldsymbol{x}_2\boldsymbol{y}_2) = 0$, and otherwise identity. We also define a map $\iota : N_2 \to N_1$ as an inclusion. Then, $\pi\circ\iota = \mathbb{I}_{N_2}$ is obvious. In addition, a homotopy equivalence $H : N_1 \to N_1$ is given as

$$H(x) := \begin{cases} \boldsymbol{q} \otimes \boldsymbol{x}_3 \boldsymbol{y}_1 & \text{if } x = \boldsymbol{q} \otimes \boldsymbol{x}_2 \boldsymbol{y}_2, \\ \boldsymbol{q} \otimes \boldsymbol{x}_1 \boldsymbol{y}_3 + \boldsymbol{q} \otimes \boldsymbol{x}_3 \boldsymbol{y}_1 & \text{if } x = \boldsymbol{q} \otimes \boldsymbol{x}_1 \boldsymbol{y}_3, \\ \boldsymbol{p}_2 \otimes \boldsymbol{x}_2 \boldsymbol{b} & \text{if } x = \boldsymbol{p}_2 \otimes \boldsymbol{x}_2 \boldsymbol{b}, \\ 0 & \text{otherwise}, \end{cases}$$

which extends to a $\mathcal{A}(T)$-equivariant map. It is clear that $\iota \circ \pi = (\delta_1)^1 \circ H + H \circ (\delta_1)^1$.

Remark 4.1. Compare this result with [Lipshitz et al. 2008, Section 11.5], from which they spelled out an algorithm to recover $\widehat{CFD}(S^3 \backslash \nu K)$ from CFK^-. According to their notation, the length of the unstable chain is 3 (the number of generators between two outermost ones). This length is closely related to the framing of the knot complement and concordance invariant $\tau(K)$; see [Lipshitz et al. 2008, equation (11.18)]. In our case, the framing of the left component of the link was originally -1, but a handleslide procedure has added +4 and therefore the framing is 3. Since $\tau(\text{Trefoil}) = 1$ is less than the framing, the length of the unstable chain agrees with the framing. Interested readers will find the precise description of the relation between $\tau(K)$ and the unstable chain in [Lipshitz et al. 2008, Theorem A.11].

An integral surgery on Hopf link. Hopf link is $(2, 2)$-torus link. If n_1 and n_2 are two positive integers such that $n_1 n_2 \neq 1$, then (n_1, n_2)-surgery on Hopf link produces the lens space $L(n_1 n_2 - 1, n_1)$. The Heegaard Floer homology of the lens space has $n_1 n_2 - 1$ generators whose differentials equal zero.

The diagram of the Hopf link complement is easy. In addition, $\alpha_1^{a,L}$ and $\alpha_1^{a,R}$ do not intersect β_1 and β_2, respectively; therefore pairing the diagram with $\mathcal{H}_{n_1}^L$ and $\mathcal{H}_{n_2}^R$ will give a closed Heegaard diagram of the lens space $L(n_1 n_2 - 1, n_1)$. The $\mathcal{A}_\infty$-relation of $\widehat{CFA}(\mathcal{H}_m)$ is as follows (see Figure 7):

$$m(q, \rho_2) = p_1,$$

$$m(p_i, \rho_3, \underbrace{\rho_{23}, \ldots, \rho_{23}}_{j \text{ times}}, \rho_2) = p_{i+j+1},$$

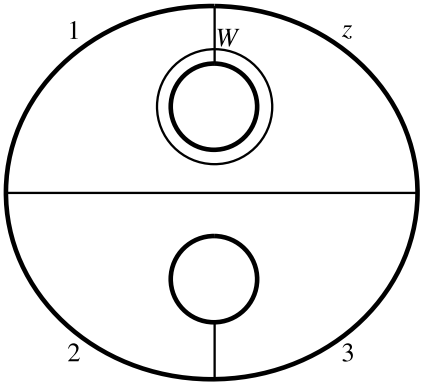
$$m(p_m, \rho_3, \rho_2, \rho_1) = q.$$

$\widehat{CFDD}(S^3 \backslash \nu(\text{Hopf link}))$ has two generators $\boldsymbol{ab}$ and $\boldsymbol{x}_1 \boldsymbol{y}_1$. Its type-$D$ structure is given below:

$$\delta^1(\boldsymbol{ab}) = (\rho_1 \otimes \sigma_3 + \rho_3 \otimes \sigma_1 + \rho_{123} \otimes \sigma_{123}) \otimes \boldsymbol{x}_1 \boldsymbol{y}_1,$$

$$\delta^1(\boldsymbol{x}_1 \boldsymbol{y}_1) = \rho_2 \otimes \sigma_2 \otimes \boldsymbol{ab}.$$

Remark 4.2. See [Lipshitz et al. 2015, Proposition 10.1]. Note that Hopf link complement is $T^2 \times [0, 1]$ and it is exactly an identity module described there.



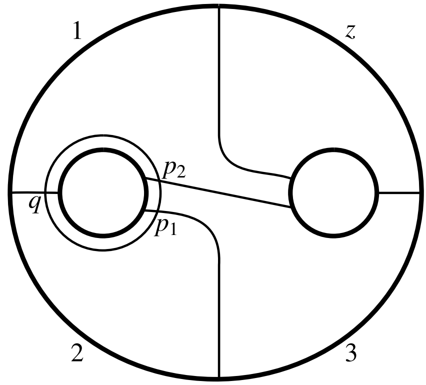


Figure 7. The diagram $\mathcal{H}_\infty$ on the left shows ∞-surgery on the right component of the link. The diagram $\mathcal{H}_{+2}$ on the right is $+2$-surgery on the right component. The $\mathcal{A}_\infty$-relation of $\widehat{CFA}(\mathcal{H}_{+2})$ is given as $m(q,\sigma_2)=p_1$, $m(p_1,\sigma_3,\sigma_2)=p_2$, and $m(p_2,\sigma_3,\sigma_2,\sigma_1)=q$.

Let p_i^L and $q^L(p_j^R$ and q^R, respectively) be the generators of the bordered Heegaard diagram $\mathcal{H}_{n_1}^L$ attached to the left ($\mathcal{H}_{n_2}^R$ attached to the right, respectively). Then, $\widehat{CFA}(\mathcal{H}_{n_1}^L)\widetilde{\otimes}_{\mathcal{A}(\mathcal{Z})}\widehat{CFA}(\mathcal{H}_{n_2}^R)\widetilde{\otimes}_{\mathcal{A}(\mathcal{Z})}\widehat{CFDD}(S^3\backslash\nu(\text{Hopf link}))$ has the following n_1n_2+1 generators:

$$\begin{aligned}p_i^L\otimes p_j^R\otimes \boldsymbol{ab},\quad & i=1,\dots,n_1 \text{ and } j=1,\dots,n_2,\\ q^L\otimes q^R\otimes \boldsymbol{x}_1\boldsymbol{y}_1.&\end{aligned}$$

The only nontrivial differential is

$$\partial(q^L\otimes q^R\otimes \boldsymbol{x}_1\boldsymbol{y}_1)=m(q^L,\rho_2)\otimes m(q^R,\sigma_2)\otimes \boldsymbol{ab}=p_1^L\otimes p_1^R\otimes \boldsymbol{ab}.$$

Thus, the homology of $\widehat{CFA}(\mathcal{H}_{n_1}^L)\widetilde{\otimes}_{\mathcal{A}(\mathcal{Z})}\widehat{CFA}(\mathcal{H}_{n_2}^R)\widetilde{\otimes}_{\mathcal{A}(\mathcal{Z})}\widehat{CFDD}(S^3\backslash\nu(\text{Hopf link}))$ has n_1n_2-1 generators as expected.

5. Homotopy equivalence

In this section, we streamline the type-*DD* structure computed in Section 3 to a type-*DD* structure that does not involve any differential with the algebra element 1.

Proposition 5.1. *The type-DD structure of the link complement of the* $(2,2n)$*-torus link complement, where* $n\geq 3$, *has the same homotopy type as the complex given in Figure 8.*

Proof. Let (M,δ^1) denote the type-*DD* structure computed in Proposition 3.9 and $(N,(\delta^1)')$ the type-*DD* structure given as Figure 8. More specifically, the map $(\delta^1)'$

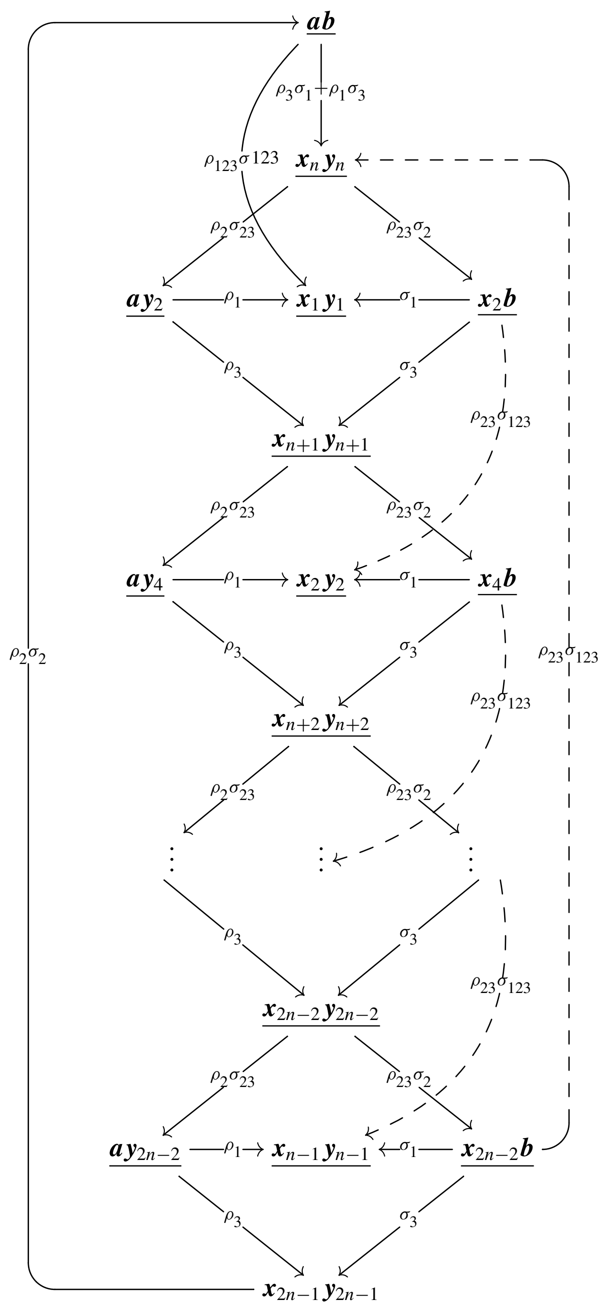


Figure 8. Simplified diagram of $\widehat{CFDD}$ of the $(2, 2n)$-torus link complement, $n \geq 3$. Note that the dashed arrows can be changed to the arrows in Figure 9.

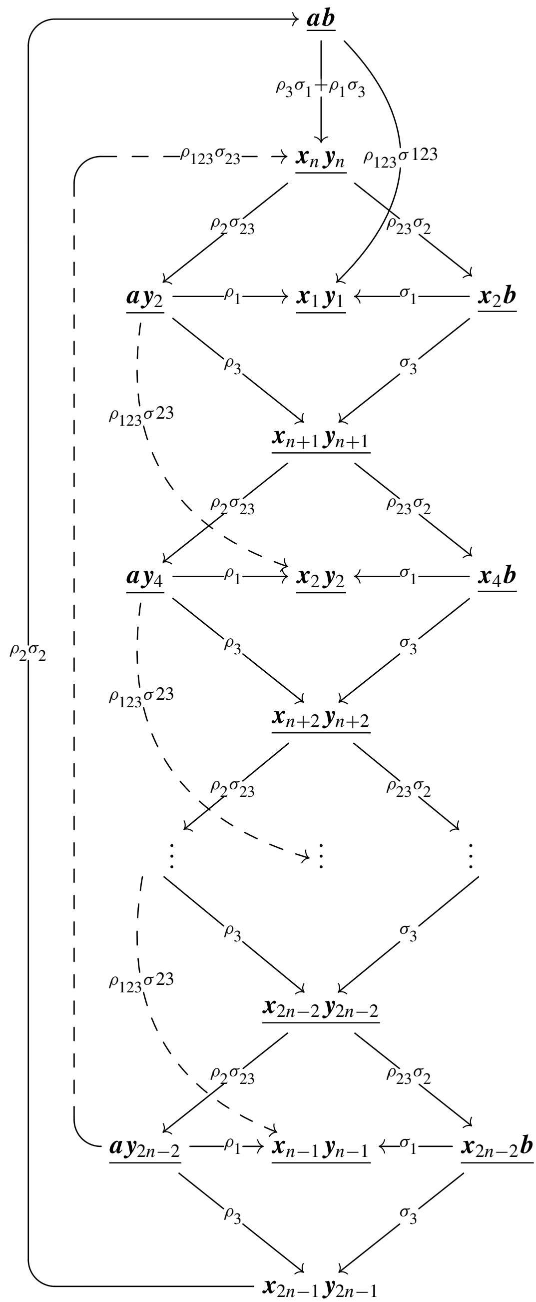


Figure 9. Another type-*DD* structure homotopy equivalent to the original type-*DD* structure. The differential represented by the dashed line can be changed to the differential in Figure 8, too.

has the following differentials:

$$\begin{aligned}
&\underline{\boldsymbol{ab}} \mapsto \rho_{123}\sigma_{123} \otimes \underline{\boldsymbol{x}_1\boldsymbol{y}_1} + (\rho_1\sigma_3 + \rho_3\sigma_1) \otimes \underline{\boldsymbol{x}_n\boldsymbol{y}_n}, && \\
&\underline{\boldsymbol{a}\boldsymbol{y}_{2k}} \mapsto \rho_1 \otimes \underline{\boldsymbol{x}_k\boldsymbol{y}_k} + \rho_3 \otimes \underline{\boldsymbol{x}_{n+k}\boldsymbol{y}_{n+k}} && \text{if } k = 1, \dots, n-1, \\
&\underline{\boldsymbol{x}_{2k}\boldsymbol{b}} \mapsto \sigma_1 \otimes \underline{\boldsymbol{x}_k\boldsymbol{y}_k} + \sigma_3 \otimes \underline{\boldsymbol{x}_{n+k}\boldsymbol{y}_{n+k}} + \rho_{23}\sigma_{123} \otimes \underline{\boldsymbol{x}_{k+1}\boldsymbol{y}_{k+1}} && \text{if } k = 1, \dots, n-1, \\
&\underline{\boldsymbol{x}_k\boldsymbol{y}_k} \mapsto 0 && \text{if } k = 1, \dots, n-1, \\
&\underline{\boldsymbol{x}_k\boldsymbol{y}_k} \mapsto \rho_2\sigma_{23} \otimes \underline{\boldsymbol{a}\boldsymbol{y}_{2(k-n+1)}} + \rho_{23}\sigma_2 \otimes \underline{\boldsymbol{x}_{2(k-n+1)}\boldsymbol{b}} && \text{if } k = n, \dots, 2n-2, \\
&\underline{\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}} \mapsto \rho_2\sigma_2 \otimes \underline{\boldsymbol{ab}}. &&
\end{aligned}$$

We shall now define type-*DD* structure maps $F : M \to \mathcal{A}(-\mathcal{Z}_L) \otimes \mathcal{A}(-\mathcal{Z}_R) \otimes N$ and $G : N \to \mathcal{A}(-\mathcal{Z}_L) \otimes \mathcal{A}(-\mathcal{Z}_R) \otimes M$. First, the map F is defined as below.

$$\begin{aligned}
F(\boldsymbol{ab}) &= \underline{\boldsymbol{ab}}, \\
F(\boldsymbol{a}\boldsymbol{y}_{2k}) &= \underline{\boldsymbol{a}\boldsymbol{y}_{2k}}, \\
F(\boldsymbol{x}_{2k}\boldsymbol{b}) &= \underline{\boldsymbol{x}_{2k}\boldsymbol{b}}, \\
F(\boldsymbol{x}_1\boldsymbol{y}_{2k-1}) &= \underline{\boldsymbol{x}_k\boldsymbol{y}_k} \quad \text{for } k = 1, \dots, n, \\
F(\boldsymbol{x}_{2k-1}\boldsymbol{y}_{2n-1}) &= \underline{\boldsymbol{x}_{k+n-1}\boldsymbol{y}_{k+n-1}} \quad \text{for } k = 1, \dots, n, \\
F(\boldsymbol{x}_{2k}\boldsymbol{y}_{2n-2}) &= \rho_2\sigma_{23} \otimes \underline{\boldsymbol{a}\boldsymbol{y}_{2k}} \quad \text{for } k = 1, \dots, n-1,
\end{aligned}$$

and zero otherwise.

The map G is defined as follows:

$$\begin{aligned}
G(\underline{\boldsymbol{ab}}) &= \boldsymbol{ab}, \\
G(\underline{\boldsymbol{a}\boldsymbol{y}_{2k}}) &= \boldsymbol{a}\boldsymbol{y}_{2k}, \\
G(\underline{\boldsymbol{x}_{2k}\boldsymbol{b}}) &= \boldsymbol{x}_{2k}\boldsymbol{b}, \\
G(\underline{\boldsymbol{x}_1\boldsymbol{y}_1}) &= \boldsymbol{x}_1\boldsymbol{y}_1 + \rho_{23}\sigma_{23} \otimes \boldsymbol{x}_3\boldsymbol{y}_1, \\
G(\underline{\boldsymbol{x}_k\boldsymbol{y}_k}) &= \boldsymbol{x}_1\boldsymbol{y}_{2k-1} + \boldsymbol{x}_{2k-1}\boldsymbol{y}_1 + \rho_{23}\sigma_{23} \otimes \boldsymbol{x}_{2k+1}\boldsymbol{y}_1 \quad \text{for } k = 2, \dots, n-1, \\
G(\underline{\boldsymbol{x}_k\boldsymbol{y}_k}) &= \boldsymbol{x}_{2k-2n+1}\boldsymbol{y}_{2n-1} + \boldsymbol{x}_{2n-1}\boldsymbol{y}_{2k-2n+1} \quad \text{for } k = n, \dots, 2n-2, \\
G(\underline{\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}}) &= \boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}.
\end{aligned}$$

These maps are easily seen satisfying the compatibility condition spelled out in [Lipshitz et al. 2015, Definition 2.2.55]. Then, the composition of two maps $F \circ G : N \to N$ is the identity map. Another composition $G \circ F$ is homotopic to identity by introducing the seemingly complicated map $H : M \to \mathcal{A}(-\mathcal{Z}_L) \otimes \mathcal{A}(-\mathcal{Z}_R) \otimes M$. For the generators of M listed below, the map H is defined as

$$\begin{aligned}
H(\boldsymbol{ab}) &= 0, \\
H(\boldsymbol{a}\boldsymbol{y}_{2k}) &= \rho_3 \otimes (\boldsymbol{x}_{2k+1}\boldsymbol{y}_{2n-1} + \boldsymbol{x}_{2n-1}\boldsymbol{y}_{2k+1}) \quad \text{for } k = 1, \dots, n-2,
\end{aligned}$$

$$H(\boldsymbol{a}\boldsymbol{y}_{2n-2}) = \rho_3 \otimes \boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1},$$
$$H(\boldsymbol{x}_{2k}\boldsymbol{b}) = \sigma_3 \otimes (\boldsymbol{x}_{2n-1}\boldsymbol{y}_{2k+1} + \boldsymbol{x}_{2k+1}\boldsymbol{y}_{2n-1}) \quad \text{for } k = 1, \dots, n-2,$$
$$H(\boldsymbol{x}_{2n-2}\boldsymbol{b}) = \sigma_3 \otimes \boldsymbol{x}_{2n-1}\boldsymbol{y}_{2n-1}.$$

Now, we need to define $H(\boldsymbol{x}_i\boldsymbol{y}_j)$. Before giving the definition, we will introduce the new notation $\boldsymbol{x}\boldsymbol{y}(k, l) \in M$ for simplicity:

$$\boldsymbol{x}\boldsymbol{y}(i, j) := \begin{cases} \boldsymbol{x}_i\boldsymbol{y}_j + \boldsymbol{x}_j\boldsymbol{y}_i & \text{if } i \neq j, \\ \boldsymbol{x}_i\boldsymbol{y}_j & \text{if } i = j. \end{cases}$$

Case 1, if $i < j$:

$$H(\boldsymbol{x}_i\boldsymbol{y}_j) = \begin{cases} \boldsymbol{x}\boldsymbol{y}(i+1, j-1) & \text{if } i = 1 \text{ or } j = 2n-1, \\ \boldsymbol{x}\boldsymbol{y}(i+1, j-1) + \boldsymbol{x}_{j+1}\boldsymbol{y}_{i-1} & \text{otherwise.} \end{cases}$$

Case 2, if $i > j$:

$$H(\boldsymbol{x}_i\boldsymbol{y}_j) = \begin{cases} \boldsymbol{x}\boldsymbol{y}(i-1, j+1) + \rho_{23}\sigma_{23} \otimes \boldsymbol{x}\boldsymbol{y}(i+1, j+1) & \text{if } j = 1 \text{ and } 3 \leq i \leq 2n-3, \\ \boldsymbol{x}\boldsymbol{y}(i-1, j+1) & \text{otherwise.} \end{cases}$$

Case 3, if $i = j$:

$$H(\boldsymbol{x}_i\boldsymbol{y}_j) = \begin{cases} \rho_{23}\sigma_{23} \otimes \boldsymbol{x}_2\boldsymbol{y}_2 & \text{if } i = j = 1, \\ 0 & \text{if } i = j = 2n-1, \\ \boldsymbol{x}_{i+1}\boldsymbol{y}_{j-1} & \text{otherwise.} \end{cases}$$

It is easy to verify that the above map satisfies $G \circ F + \mathbb{I}_M = \delta^1 \circ H + H \circ \delta^1$. $\square$

Remark 5.2. The symmetry of Figure 6 seems to be lost after removing the differentials of the algebra element 1 since the differentials of algebra element $\rho_{23}\sigma_{123}$ are between $\underline{\boldsymbol{x}_{2k}\boldsymbol{b}}$ and $\overline{\boldsymbol{x}_{k+1}\boldsymbol{y}_{k+1}}$. This phenomenon is caused because we set the map F such that the bottom right corner of the original type-*DD* structure "collapses." If we set F to collapse the top left corner of the original diagram, then the resulting complex will look like Figure 9.

Acknowledgements

My thanks go to Robert Lipshitz and Peter Ozsváth for helpful conversations and advice, and to Adam Levine for commenting on the final version of this paper. I am very grateful to my advisor, Olga Plamenevskaya, for her enormous patience and encouragement.

References

[Colin et al. 2011] V. Colin, P. Ghiggini, and K. Honda, "Equivalence of Heegaard Floer homology and embedded contact homology via open book decompositions", *Proc. Natl. Acad. Sci. USA* **108**:20 (2011), 8100–8105. MR Zbl

[Hanselman 2016] J. Hanselman, "Bordered Heegaard Floer homology and graph manifolds", *Algebr. Geom. Topol.* **16**:6 (2016), 3103–3166. MR Zbl

[Hom 2011] J. Hom, *Heegaard Floer invariants and cabling*, Ph.D. thesis, University of Pennsylvania, 2011, available at http://search.proquest.com/docview/878890913.

[Kutluhan et al. 2011] C. Kutluhan, Y.-J. Lee, and C. H. Taubes, "HF = HM, I: Heegaard Floer homology and Seiberg–Witten Floer homology", preprint, 2011. arXiv

[Lipshitz 2006] R. Lipshitz, *A Heegaard–Floer invariant of bordered three-manifolds*, Ph.D. thesis, Stanford University, 2006, available at https://search.proquest.com/docview/304976553.

[Lipshitz et al. 2008] R. Lipshitz, P. Ozsváth, and D. Thurston, "Bordered Heegaard Floer homology: invariance and pairing", preprint, 2008. arXiv

[Lipshitz et al. 2009] R. Lipshitz, P. Ozsváth, and D. Thurston, "Slicing planar grid diagrams: a gentle introduction to bordered Heegaard Floer homology", pp. 91–119 in *Proceedings of Gökova Geometry-Topology Conference 2008*, edited by S. Akbulut et al., Gökova Geometry/Topology Conference, Gökova, 2009. MR Zbl

[Lipshitz et al. 2011] R. Lipshitz, P. S. Ozsváth, and D. P. Thurston, "Tour of bordered Floer theory", *Proc. Natl. Acad. Sci. USA* **108**:20 (2011), 8085–8092. MR Zbl

[Lipshitz et al. 2014] R. Lipshitz, P. S. Ozsváth, and D. P. Thurston, "Computing $\widehat{HF}$ by factoring mapping classes", *Geom. Topol.* **18**:5 (2014), 2547–2681. MR Zbl

[Lipshitz et al. 2015] R. Lipshitz, P. S. Ozsváth, and D. P. Thurston, "Bimodules in bordered Heegaard Floer homology", *Geom. Topol.* **19**:2 (2015), 525–724. MR Zbl

[Manolescu and Ozsváth 2010] C. Manolescu and P. Ozsváth, "Heegaard Floer homology and integer surgeries on links", preprint, 2010. arXiv

[Ozsváth and Szabó 2004a] P. Ozsváth and Z. Szabó, "Holomorphic disks and knot invariants", *Adv. Math.* **186**:1 (2004), 58–116. MR Zbl

[Ozsváth and Szabó 2004b] P. Ozsváth and Z. Szabó, "Holomorphic disks and topological invariants for closed three-manifolds", *Ann. of Math.* (2) **159**:3 (2004), 1027–1158. MR Zbl

[Ozsváth and Szabó 2008a] P. Ozsváth and Z. Szabó, "Holomorphic disks, link invariants and the multi-variable Alexander polynomial", *Algebr. Geom. Topol.* **8**:2 (2008), 615–692. MR Zbl

[Ozsváth and Szabó 2008b] P. S. Ozsváth and Z. Szabó, "Knot Floer homology and integer surgeries", *Algebr. Geom. Topol.* **8**:1 (2008), 101–153. MR Zbl

[Rasmussen 2002] J. A. Rasmussen, "Floer homology of surgeries on two-bridge knots", *Algebr. Geom. Topol.* **2** (2002), 757–789. MR Zbl

[Sarkar and Wang 2010] S. Sarkar and J. Wang, "An algorithm for computing some Heegaard Floer homologies", *Ann. of Math.* (2) **171**:2 (2010), 1213–1236. MR Zbl

Received April 13, 2015. Revised February 14, 2017.

JAEPIL LEE
MATHEMATICS DEPARTMENT
YONSEI UNIVERSITY
SEOUL
SOUTH KOREA
jaepil@yonsei.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 292, No. 1, 2018
dx.doi.org/10.2140/pjm.2018.292.177

A FEYNMAN–KAC FORMULA FOR DIFFERENTIAL FORMS ON MANIFOLDS WITH BOUNDARY AND GEOMETRIC APPLICATIONS

LEVI LOPES DE LIMA

We establish a Feynman–Kac-type formula for differential forms satisfying absolute boundary conditions on Riemannian manifolds with boundary and of bounded geometry. We use this to construct L^2-harmonic forms out of bounded ones on the universal cover of a compact Riemannian manifold whose geometry displays a positivity property expressed in terms of a certain stochastic average of the Weitzenböck operator R_p acting on p-forms and the second fundamental form of the boundary. This extends previous work by Elworthy, Li and Rosenberg on closed manifolds to this more general setting. As an application we find a new obstruction to the existence of metrics with positive R_2 (in particular, positive isotropic curvature) and 2-convex boundary. We also discuss a version of the Feynman–Kac formula for spinors under suitable boundary conditions and use this to prove a semigroup domination result for the corresponding Dirac Laplacian under a mean convexity assumption.

1. Introduction

A celebrated result by Gromov [1971] says that an open manifold carries both positively and negatively curved metrics. Thus, in any such manifold there is enough room to interpolate between two rather distinct types of geometries. In contrast, no such flexibility is available in the context of closed manifolds. For instance, it already follows from Hadamard and Bonnet–Myers theorems from basic Riemannian Geometry that a closed manifold which carries a metric with nonpositive sectional curvature does not carry a metric with positive Ricci curvature.

Our interest here lies in another manifestation of this "exclusion principle" for closed manifolds due to Elworthy, Li and Rosenberg [Elworthy et al. 1998]. Relying heavily on stochastic methods, these authors put forward an elegant refinement of the

Research partially suported by CNPq grant 311258/2014-0 and FUNCAP/CNPq/PRONEX grant 00068.01.00/15.

MSC2010: primary 53C21, 53C27; secondary 58J35, 58J65.

Keywords: Feynman–Kac formula, absolute boundary conditions, heat flow on differential forms, Brownian motion on manifolds.

famous Bochner technique with far-reaching consequences. For example, they prove that a sufficiently negatively pinched closed manifold does not carry a metric whose Weitzenböck operator acting on 2-forms is even allowed to be negative in a region of small volume, an improvement which definitely makes the obstruction unapproachable by the classical reasoning [Rosenberg 1997]. We focus here on extending this kind of geometric obstruction to compact manifolds with boundary (∂-manifolds, for short). When pursuing this goal we should have in mind that balls carry a huge variety of metrics as illustrated by geodesic balls in an arbitrary Riemannian manifold. These simple examples also show that the boundary can always be chosen convex just by taking the radius sufficiently small. Thus, even if we insist on having the boundary appropriately convex in both metrics, some topological assumption on the underlying manifold must be imposed. Our purpose is to present results in this direction which qualify as natural extensions of those in [Elworthy et al. 1998].

We now introduce the notation needed to state our main results. If N is a Riemannian ∂-manifold of dimension n, the Weitzenböck decomposition reads

$$\Delta_q = \Delta_q^B + R_q,$$

where $\Delta_q = dd^\star + d^\star d$ is the Hodge Laplacian acting on q-forms, $1 \le q \le n-1$, $d^\star = \pm \star d \star$ is the codifferential, $\star$ is the Hodge star operator, Δ_q^B is the Bochner Laplacian and R_q, the Weitzenböck curvature operator, depends linearly on the curvature tensor, albeit in a rather complicated way. Recall that $R_1 = \mathrm{Ric}$, and since $\star R_p = R_{n-p}\star$, this also determines R_{n-1}, but in general the structure of R_q, $2 \le q \le n-2$, is notoriously hard to grasp. To these invariants we attach the functions $r_{(q)} : N \to \mathbb{R}$, $r_{(q)}(x) = \inf_{|\omega|=1} \langle R_q(x)\omega, \omega\rangle$, the least eigenvalue of $R_q(x)$. We also consider the principal curvatures $\rho_1, \dots, \rho_{n-1}$ of ∂N computed with respect to the inward unit normal vector field. For each $x \in \partial N$ and $q = 1, \dots, n-1$, define

$$\rho_{(q)}(x) = \inf_{1 \le i_1 < \cdots < i_q \le n-1} \rho_{i_1}(x) + \cdots + \rho_{i_q}(x),$$

the sum of the q smallest principal curvatures at x. We say that ∂N is q-convex if $\underline{\rho}_{(q)} := \inf_{x \in \partial M} \rho_{(q)}(x) > 0$. Note that q-convexity implies $(q+1)$-convexity. Also, N is said to be convex if $\underline{\rho}_{(1)} \ge 0$ everywhere. Finally, recall that a Riemannian metric h on a manifold is κ-negatively pinched if its sectional curvature satisfies $-1 \le K_{\mathrm{sec}}(h) < -\kappa < 0$.

Stochastic notions make their entrance in the theory by means of the following considerations. Let N be a Riemannian ∂-manifold. In case N is noncompact we always assume that the underlying metric h is complete and the triple $(N, \partial N, h)$ has bounded geometry in the sense of [Schick 1996; 1998; 2001]. We then consider reflecting Brownian motion $\{x^t\}$ on N starting at some $x^0 \in N$; see Section 5 for a (necessarily brief) description of this diffusion process. Let $\alpha : N \to \mathbb{R}$ and $\beta : \partial N \to \mathbb{R}$ be C^1 functions. Adapting a classical definition to our setting, we say that the pair (α, β) is strongly stochastically positive (s.s.p.) if

$$\limsup_{t\to+\infty} \frac{1}{t} \sup_{x^0\in K} \log \mathbb{E}_{x^0}\biggl(\exp\biggl(-\frac{1}{2}\int_0^t \alpha(x^s)\,ds - \int_0^t \beta(x^s)\,dl^s\biggr)\biggr) < 0,$$

for any $K \subset N$ compact, where l^t is the boundary local time associated to $\{x^t\}$. This is certainly the case if both α and β have strictly positive lower bounds but the point we would like to emphasize here is that, at least if N is compact, it might well happen with the functions being positive except possibly in regions of small volume, given that the definition involves expectation with respect to the underlying diffusion.

Similarly to [Elworthy et al. 1998], our main results provide examples of ∂-manifolds for which there holds an exclusion principle involving the various notions of curvature appearing above. From now on we always assume that $n \geq 4$ and set $\kappa_p = p^2/(n-p-1)^2$.

Theorem 1.1. *Let M be a compact ∂-manifold with infinite fundamental group. Assume also that M satisfies $H^p(M;\mathbb{R}) \neq 0$, where $2 \leq p < (n-1)/2$. If M carries a convex κ_p-negatively pinched metric then it does not carry a metric with both $(r_{(p\pm1)}, \rho_{(p-1)})$ s.s.p.*

Our next result, which handles the least possible value for the degree p, has a somewhat more satisfactory statement.

Theorem 1.2. *Let M be a compact manifold with nonamenable fundamental group. If M carries a convex κ_1-negatively pinched metric then it does not carry a metric with $(r_{(2)}, \rho_{(2)})$ s.s.p.*

Remark 1.1. These results correspond respectively to Corollary 2.1 and Theorem 2.3 in [Elworthy et al. 1998]. We point out that our assumptions on the fundamental group are natural in the sense that they are automatically satisfied there. As mentioned above, balls are obvious counterexamples to our results if the topological assumptions are removed. Also, the manifold $\mathbb{S}^1 \times \mathbb{D}^{n-1}$ shows that merely assuming that the fundamental group is infinite does not suffice in Theorem 1.2; see Remark 1.5 below. On the other hand, it is not clear whether the convexity hypothesis with respect to the negatively curved metric can be relaxed somehow.

Using Theorem 1.2 we can exhibit an interesting family of compact ∂-manifolds for which a natural class of metrics is excluded.

Theorem 1.3. *If X is a closed hyperbolic manifold of dimension, $l \geq 2$ then its product with a disk $\mathbb{D}^m$ does not carry a metric with $(r_{(2)}, \rho_{(2)})$ s.s.p.*

Proof. Write $X = \mathbb{H}^l/\Gamma$ as the quotient of hyperbolic space $\mathbb{H}^l$ by a (necessarily nonamenable) group Γ of hyperbolic motions. Embed $\mathbb{H}^l$ as a totally geodesic submanifold of $\mathbb{H}^{l+m}$ and let $\widetilde{M} \subset \mathbb{H}^{l+m}$ be a tubular neighborhood of $\mathbb{H}^l$ of constant radius. Extend the Γ-action to $\widetilde{M}$ in the obvious manner and observe that, since $\widetilde{M}$

is convex, $M = \widetilde{M}/\Gamma = X \times \mathbb{D}^m$ with the induced hyperbolic metric is convex as well. Thus, Theorem 1.2 applies. □

Remark 1.2. Theorem 1.3 provides a geometric obstruction to the existence of metrics with $(r_{(2)}, \rho_{(2)})$ s.s.p. Notice that if the second Betti number of X vanishes, the obstruction can not be detected by the classical version of the Bochner technique for ∂-manifolds [Yano 1970, Chapter 8] even if we assume strict positivity of $(r_{(2)}, \rho_{(2)})$.

Remark 1.3. A larger class of manifolds for which the conclusion of Theorem 1.3 obviously holds is formed by tubular neighborhoods of closed embedded totally geodesic submanifolds in a given hyperbolic manifold.

Corollary 1.1. *Under the conditions of Theorem 1.3, assume that $n = l + m$ is even. Then $X \times \mathbb{D}^m$ does not carry a metric with positive isotropic curvature and* 2*-convex boundary.*

Proof. For even-dimensional manifolds it is shown in [Micallef and Wang 1993] that positive isotropic curvature implies $R_2 > 0$. □

Remark 1.4. Since the computation in [Micallef and Wang 1993] expresses R_2 as a sum of isotropic curvatures, in Corollary 1.1 we can even relax the condition on the metric to allow the invariants to be negative in a region of small volume.

Remark 1.5. The standard product metric on $\mathbb{S}^1 \times \mathbb{S}^{n-1}$ is known to have positive isotropic curvature. It is easy to check that if $r < \pi/2$ the boundary of the tubular neighborhood $U_r \subset \mathbb{S}^1 \times \mathbb{S}^{n-1}$ of radius r of the circle factor is 2-convex. Thus, the conclusion of Corollary 1.1 does not hold for $U_r = \mathbb{S}^1 \times \mathbb{D}^{n-1}$. Notice that U_r carries a convex hyperbolic metric since its universal cover $\widetilde{U}_r = \mathbb{R} \times \mathbb{D}^{n-1}$ is diffeomorphic to a tubular neighborhood of a geodesic in $\mathbb{H}^n$. The problem here is that the fundamental group is abelian, hence amenable, and the argument leading to Theorem 1.2 breaks down. This also can be understood in stochastic terms. In effect, the proof of Theorem 1.2 shows that, under the given conditions, Brownian motion on the universal cover is transient, while recurrence certainly occurs in $\widetilde{U}_r$; see Remark 5.1. In this respect it would be interesting to investigate if the conclusion of Theorem 1.3 holds in case X is flat or, more generally, has nonpositive sectional curvature.

Remark 1.6. Compact ∂-manifolds with positive isotropic curvature have deserved a lot of attention in recent years. An important result by Fraser [2002] says that such a ∂-manifold is contractible if it is simply connected and its boundary is connected and 2-convex. The proof combines index estimates for minimal surfaces and a variant of the Sachs–Uhlenbeck theory adapted to this setting. However, as the examples in Remark 1.5 attest, this geometric condition is compatible with an infinite fundamental group. With no assumption on the fundamental group or on the topology of the boundary, the techniques in [Fraser 2002] still imply that all the (absolute and relative) homotopy groups vanish in the range $2 \leq i \leq n/2$. Moreover, it is shown in [Chen and Fraser 2010] that the fundamental group of the boundary injects into the fundamental group of the manifold. However, if we take $m \geq l+2$ it is easy to check that none of these homotopical obstructions rules out the metrics in Corollary 1.1. We point out that a conjecture in [Fraser 2002] asserts that a closed, embedded 2-convex hypersurface in a manifold with positive isotropic curvature is either $\mathbb{S}^n$ or a connected sum of finitely many copies of $\mathbb{S}^1 \times \mathbb{S}^{n-1}$. Since the fundamental group of a closed hyperbolic manifold is neither infinite cyclic nor a free product, Corollary 1.1 provides further support to the conjecture.

This paper is partly inspired by the beautiful work by Elworthy, Rosenberg and Li [Elworthy et al. 1998]. Their ideas are used in Section 2 to construct L^2 harmonic forms on the universal cover of certain compact ∂-manifolds starting from bounded ones. This is precisely where stochastic techniques come into play and a crucial ingredient at this point is a Feynman–Kac-type formula for differential forms in higher degree meeting absolute boundary conditions. In order not to interrupt the exposition, this technical result is established in the final Section 5 following ideas in [Hsu 2002a], where the case of 1-forms is treated; see also [Airault 1976; Ikeda and Watanabe 1989] for previous contributions. To illustrate the flexibility of the method we also discuss a similar formula for spinors evolving under the heat semigroup generated by the Dirac Laplacian on a spin^c ∂-manifold under suitable boundary conditions. Another important ingredient in the argument is a Donnelly–Xavier-type eigenvalue estimate described in Section 3, whose proof uses both the convexity and the assumption that the fundamental group is infinite. Combined with Schick's [1996; 1998] L^2 Hodge–de Rham theory this allows us to prove a vanishing result for the relevant L^2 cohomology group. Finally, the proofs of the main applications (Theorem 1.1 and 1.2 above) are presented in Section 4.

2. From bounded to L^2-harmonic forms

We consider a complete Riemannian ∂-manifold N with volume element dN and boundary ∂N oriented by an inward unit normal vector field ν. As always we assume that the triple $(N, \partial N, h)$ has bounded geometry in the sense of [Schick 1996; 1998; 2001]. For us the case of interest occurs when $N = \widetilde{M}$, the universal cover of a compact ∂-manifold (M, g) and $h = \tilde{g}$, the lifted metric. Recall that a q-form ω on N satisfies absolute boundary conditions if

$$\nu \lrcorner \omega = 0, \quad \nu \lrcorner d\omega = 0 \tag{2-1}$$

along ∂N. Equivalently,

$$\omega_{\mathrm{nor}} = 0, \quad (d\omega)_{\mathrm{nor}} = 0, \tag{2-2}$$

where $\omega = \omega_{\tan} + \nu \wedge \omega_{\text{nor}}$ is the natural decomposition of ω in its tangential and normal components. Here, we identify ν to its dual 1-form in the standard manner. For simplicity we say that ω is absolute if any of these conditions is satisfied. Notice that for $q = 0$ this means that the given function satisfies Neumann boundary condition.

For $t > 0$ let $P_t = e^{-t\Delta_q^{\text{abs}}/2}$ be the corresponding heat kernel acting on forms. Thus, for any absolute q-form $\omega_0 \in L^2 \cap L^\infty$, $\omega_t = P_t\omega_0$ is a solution to the initial-boundary value problem

$$\frac{\partial \omega_t}{\partial t} + \frac{1}{2}\Delta_q^{\text{abs}}\omega_t = 0, \quad \lim_{t\to 0} \omega_t = \omega_0, \quad \nu \lrcorner \omega_t = 0, \quad \nu \lrcorner d\omega_t = 0. \tag{2-3}$$

Moreover, the long term behavior of the flow is determined by the space of absolute L^2-harmonic q-forms on (N, h) in the sense that

$$P = \lim_{t\to+\infty} P_t \tag{2-4}$$

exists and defines the orthogonal projection onto this space. Proofs of these facts follow from standard spectral theory and the elliptic machinery developed in [Schick 1996; 1998].

A key ingredient in our approach is a Feynman–Kac-type representation of any solution ω_t as above in terms of Brownian motion in N. This is well known to hold in the boundaryless case [Elworthy 1988; Hsu 2002b; Güneysu 2010; Malliavin 1974; Stroock 2000]. However, as pointed out in [Hsu 2002a], where the case $q = 1$ is discussed in detail, extra difficulties appear when trying to establish a similar result in the presence of a boundary. In Section 5 we explain how the method in [Hsu 2002a] can be adapted to establish a Feynman–Kac formula for solutions of (2-3), regardless of the value of q; see Theorem 5.2. For the moment we need an immediate consequence of this formula, namely, the useful estimate

$$|\omega_t(x^0)| \le \mathbb{E}_{x^0}\left(|\omega_0(x^t)| \exp\left(-\frac{1}{2}\int_0^t r_{(q)}(x^s)\, ds - \int_0^t \rho_{(q)}(x^s)\, dl^s\right)\right), \tag{2-5}$$

where $\{x^t\}$ is reflecting Brownian motion on N starting at x^0 and l^t is the associated boundary local time. The remarkable feature of (2-5) is that the geometric quantities $r_{(q)}$ and $\rho_{(q)}$ play entirely similar roles in stochastically controlling the solution in the long run. Now we put this estimate to good use and establish a central result in this work; compare to [Elworthy et al. 1998, Lemma 2.1].

Proposition 2.1. *Let $P = \lim_{t\to+\infty} P_t$,*

$$\theta_q(x^0) = \int_0^{+\infty} \mathbb{E}_{x^0}\left(\exp\left(-\frac{1}{2}\int_0^t r_{(q)}(x^s)\, ds - \int_0^t \rho_{(q)}(x^s)\, dl^s\right)\right) dt,$$

and take compactly supported p-forms ϕ and ψ with $\psi_{\mathrm{nor}} = 0$ along ∂N and $\phi = 0$ in a neighborhood of ∂N. If $2 \le p \le n-2$,

$$\left|\int_N \langle P\phi - \phi, \psi\rangle \, dN\right| \le \tfrac{1}{2}\big(\sup\nolimits_{x^0 \in \operatorname{supp}\phi} \theta_{p+1}(x^0)\big)|d\psi|_\infty |d\phi|_1 + \tfrac{1}{2}\big(\sup\nolimits_{x^0 \in \operatorname{supp}\phi} \theta_{p-1}(x^0)\big)|d^\star\psi|_\infty |d^\star\phi|_1.$$

If $p = 1$ we have instead

$$\left|\int_N \langle P\phi - \phi, \psi\rangle \, dN\right| \le \tfrac{1}{2}\big(\sup\nolimits_{x^0 \in \operatorname{supp}\phi} \theta_2(x^0)\big)|d\psi|_\infty |d\phi|_1 + \tfrac{1}{2}\sup\nolimits_{x^0 \in \operatorname{supp}\phi}\left|\int_0^{+\infty} (P_\tau d^\star\phi)(x^0)\, d\tau\right| |d^\star\psi|_\infty.$$

Proof. We have

$$\begin{aligned}\int_N \langle P\phi - \phi, \psi\rangle \, dN &= \lim_{t\to+\infty} \int_N \langle P_t\phi - P_0\phi, \psi\rangle \, dN \\ &= \lim_{t\to+\infty} \int_0^t \int_N \langle \partial_\tau P_\tau\phi, \psi\rangle \, dN\, d\tau \\ &= -\frac{1}{2}\int_0^{+\infty}\int_N \langle \Delta_p^{\mathrm{abs}} P_\tau \phi, \psi\rangle\, dN\, d\tau \\ &= -\frac{1}{2}\int_0^{+\infty}\int_N \langle d P_\tau\, d^\star\phi, \psi\rangle\, dN\, d\tau - \frac{1}{2}\int_0^{+\infty}\int_N \langle d^\star P_\tau\, d\phi, \psi\rangle\, dN\, d\tau.\end{aligned}$$

We now recall Green's formula: if $\alpha \wedge \star\beta$ is compactly supported then

$$\int_N \langle d\alpha, \beta\rangle \, dN = \int_N \langle \alpha, d^\star\beta\rangle\, dN + \int_{\partial N} \alpha_{\tan} \wedge \star\beta_{\mathrm{nor}}.$$

Since $(P_\tau d\phi)_{\mathrm{nor}} = 0$ this leads to

$$\int_N \langle P\phi - \phi, \psi\rangle \, dN = -\frac{1}{2}\int_0^{+\infty}\int_N \langle P_\tau\, d^\star\phi, d^\star\psi\rangle\, dN\, d\tau - \frac{1}{2}\int_0^{+\infty}\int_N \langle P_\tau\, d\phi, d\psi\rangle\, dN\, d\tau.$$

The result now follows by applying (2-5) to $\omega_\tau = P_\tau\, d^\star\psi$ and $\omega_\tau = P_\tau\, d\psi$. □

From this we derive the existence of absolute L^2-harmonic p-forms from bounded ones under appropriate positivity assumptions; compare to [Elworthy et al. 1998, Theorem 2.1]. In the following we denote by $H^q_{(2),\mathrm{abs}}(N, h)$ the q-th L^2 absolute cohomology group of (N, h). We refer to [Schick 1996; 1998] for the definition and basic properties of these invariants, including the corresponding L^2 Hodge–de Rham theory.

Proposition 2.2. *Let (N, h) and p be as above. Assume that both $\sup_{x^0\in K} \theta_{p+1}(x^0)$ and $\sup_{x^0\in K} \theta_{p-1}(x^0)$ are finite if $2 \le p \le n-2$ and that both $\sup_{x^0\in K} \theta_2(x^0)$ and $\sup_{x^0\in K} \left|\int_0^{+\infty} (P_\tau\, d^\star\phi)(x^0)\, d\tau\right|$ are finite if $p = 1$, where $K \subset N$ is any compact. Then N carries a nontrivial absolute L^2-harmonic p-form whenever it carries a nontrivial absolute bounded harmonic p-form. In particular, $H^p_{(2),\mathrm{abs}}(N, h)$ is nontrivial.*

Proof. Let ψ be a nontrivial absolute bounded harmonic p-form. Consider a Gaffney-type cutoff sequence $\{h_n\}$, i.e., each function h_n satisfies $0 \le h_n \le 1$, $|\nabla h_n| \le 1/n$, $h_n \to 1$ and $\partial h_n/\partial\nu = 0$ [Gaffney 1959] and set $\psi_n = h_n\psi$, so that each ψ_n is a compactly supported absolute form. Also, $\psi_n \to \psi$ and since $d\psi_n = dh_n \wedge \psi$ and $d^*\psi_n = -\nabla h_n \lrcorner \psi$ we see that $|d\psi_n|_\infty + |d^\star\psi_n|_\infty \to 0$ as $n \to +\infty$. Applying Proposition 2.1 with ψ replaced by ψ_n and sending $n \to +\infty$ we see that

$$\int_N \langle P\phi - \phi, \psi\rangle\, dN = 0.$$

If no nontrivial absolute L^2-harmonic p-form exists then $P\phi = 0$ for any ϕ and hence $\psi = 0$, a contradiction. The last assertion follows from the L^2 Hodge–de Rham theory in [Schick 1996; 1998]. □

Remark 2.1. Implicit in the discussion above is the well-known fact that the bounded geometry assumption implies that reflecting Brownian motion x^t is nonexplosive. For the sake of completeness we include here the well-known argument. We first observe that the geometric assumption implies that both $\underline{r}_{(1)}$ and $\underline{\rho}_{(1)}$ are finite. Let ξ and η be compactly supported functions on N with $\partial\xi/\partial\nu = 0$ along ∂N and $\eta = 0$ in a neighborhood of ∂N. Proceeding as above we find that

$$\int_N (P_t\xi - \xi)\eta\, dN = -\frac{1}{2}\int_0^t \int_N \langle P_\tau\, d\xi, d\eta\rangle\, dN\, d\tau, \qquad t > 0.$$

Using (2-5) with $\omega = d\xi$ we get

$$\left|\int_N (P_t\xi - \xi)\eta\, dN\right| \le \frac{1}{2}|d\xi|_\infty |d\eta|_1 \sup_{0\le\tau\le t} e^{-\tau \underline{r}_{(1)}/2 - \underline{\rho}_{(1)}\int_0^\tau dl^s}.$$

Again applying Gaffney's trick, i.e., replacing ξ by ξ_n approaching $\mathbf{1}$, the function identically equal to 1, and satisfying $|d\xi_n|_\infty \to 0$ as $n \to +\infty$, we conclude that $P_t\mathbf{1} = \mathbf{1}$. The result follows.

3. A Donnelly–Xavier-type estimate for ∂-manifolds

In this section we present a Donnelly–Xavier-type estimate for the universal cover of κ-negatively pinched ∂-manifolds which implies the vanishing of certain absolute L^2 cohomology groups. This extends to this setting a sharp result for boundaryless manifolds obtained in [Elworthy and Rosenberg 1993], which by its turn improves on the original result in [Donnelly and Xavier 1984]. The exact analogue for ∂-manifolds of the estimate in that work, hence with a tighter pinching, appears in [Schick 1996]; see Remark 3.1 below. Our proof adapts a computation in [Ballmann and Brüning 2001, Section 5], where the sharp result for boundaryless manifolds is also achieved, and relies on a rather general integral formula.

Proposition 3.1. *Let (N, h) be a ∂-manifold, $f : N \to \mathbb{R}$ a C^2 function and $\{\mu_i\}_{i=1}^n$ the eigenvalues of the Hessian operator of f. If $p \geq 1$ then for any compactly supported p-form ω in N,*

$$\begin{aligned}\int_N (\langle d\omega, \nabla f \wedge \omega\rangle + \langle d^\star\omega, \nabla f \lrcorner \omega\rangle)\, dN \\ = \int_N \Bigl(\sum_i \mu_i |e_i \lrcorner \omega|^2 + \tfrac{1}{2}|\omega|^2 \Delta_0 f\Bigr)\, dN - \int_{\partial N} \langle \nabla f \lrcorner \omega, \nu \lrcorner \omega\rangle\, d\partial N \\ - \frac{1}{2}\int_{\partial N} |\omega|^2 \langle \nabla f, \nu\rangle\, d\partial N,\end{aligned}$$

where $\{e_i\}$ is a local orthonormal frame diagonalizing the Hessian of f and ν is the inward unit normal vector field along ∂N.

Proof. Consider the vector field V defined by $\langle V, W\rangle = \langle \nabla f \lrcorner \omega, W \lrcorner \omega\rangle$, for any W. A computation in [Ballmann and Brüning 2001, Section 5] gives

$$\operatorname{div} V = \sum_i \mu_i |e_i \lrcorner \omega|^2 - \langle d\omega, \nabla f \wedge \omega\rangle - \langle d^\star\omega, \nabla f \lrcorner \omega\rangle + \langle \nabla_{\nabla f}\omega, \omega\rangle.$$

Integrating by parts we obtain

$$\begin{aligned}\int_N (\langle d\omega, \nabla f \wedge \omega\rangle + \langle d^\star\omega, \nabla f \lrcorner \omega\rangle)\, dN \\ = \int_N \Bigl(\sum_i \mu_i |e_i \lrcorner \omega|^2 + \langle \nabla_{\nabla f}\omega, \omega\rangle\Bigr)\, dN - \int_{\partial N} \langle \nabla f \lrcorner \omega, \nu \lrcorner \omega\rangle\, d\partial N.\end{aligned}$$

We thus obtain, as required:

$$\begin{aligned}\int_N \Bigl(\langle \nabla_{\nabla f}\omega, \omega\rangle - \tfrac{1}{2}|\omega|^2 \Delta_0 f\Bigr)\, dN \\ = \frac{1}{2}\int_N (\langle \nabla f, \nabla|\omega|^2\rangle - |\omega|^2 \Delta_0 f)\, dN \\ = \frac{1}{2}\int_N \operatorname{div}(|\omega|^2 \nabla f)\, dN = -\frac{1}{2}\int_{\partial N} |\omega|^2 \langle \nabla f, \nu\rangle\, d\partial N. \quad \square\end{aligned}$$

We can now present a version of the Donnelly–Xavier-type estimate that suffices for our purposes.

Proposition 3.2. *Let (M, g) be a compact and convex ∂-manifold with infinite fundamental group and assume that g satisfies $-1 \le K_{\sec}(g) \le -\kappa < 0$. If $p \ge 1$ then for any compactly supported p-form ω in $\widetilde{M}$ satisfying $\nu \lrcorner \omega = 0$ along $\partial \widetilde{M}$,*

$$|d\omega|_2 + |d^\star \omega|_2 \ge \tfrac{1}{2}((n-p-1)\sqrt{\kappa} - p)|\omega|_2. \tag{3-1}$$

Proof. Convexity implies that any $x \in M \setminus \partial M$ and $y \in M$ can be joined by a minimizing geodesic segment lying in the interior of M (except possibly for y). The same holds in $\widetilde{M}$ with the segment now being unique. Thus, for any $x \in \widetilde{M} \setminus \partial \widetilde{M}$ the Riemannian distance d_x to x is well-defined. Notice that $\langle \nabla d_x, \nu\rangle \le 0$ along $\partial \widetilde{N}$, $|\nabla d_x| = 1$ and $\Delta_0 \, d_x = -\sum_i \mu_i$, where we may assume that $\mu_1 = 0$. Thus, using the boundary condition $\nu \lrcorner \omega = 0$ and Proposition 3.1 with $f = d_x$ we obtain

$$|\omega|_2(|d\omega|_2 + |d^\star\omega|_2) \ge \int_N \left(\sum_i \mu_i |e_i \lrcorner \omega|^2 - \frac{1}{2}|\omega|^2 \sum_i \mu_i \right) d\widetilde{M}.$$

Expand $\omega = \sum_I \omega_I e_I$, where $I = \{i_1 < \cdots < i_p\}$ and $e_I = e_{i_1} \wedge \cdots \wedge e_{i_p}$. Since $\sum_i \mu_i |e_i \lrcorner e_I|^2 = \sum_{i \in I} \mu_i$, the right-hand side equals

$$\frac{1}{2} \int_N \sum_{i,I} \left(\sum_{i \notin I} \eta_i - \sum_{i \in I} \eta_i \right) |\omega_I|^2 \, d\widetilde{M},$$

where $\eta_i = -\mu_i$ are the principal curvatures of the geodesic ball centered at x. Thus, by standard comparison theory this is bounded from below by

$$\frac{1}{2} \int_N ((n-p-1)\sqrt{\kappa} \coth \sqrt{\kappa}\, d_x - p \coth d_x)|\omega|^2 \, d\widetilde{M}.$$

Now observe that $\widetilde{M}$ has infinite diameter because $\pi_1(M)$ is infinite. Hence, we can find a sequence $\{x_i\} \subset \widetilde{M}$ so that $d_{x_i}(y) \to +\infty$ uniformly in $y \in \operatorname{supp} \omega$. By taking $x = x_i$ and passing to the limit we obtain the desired result. □

Remark 3.1. Notice (3-1) is meaningful only if $\kappa > \kappa_p$, which forces $\kappa_p < 1$, that is, $2p < n-1$. We note that Schick [1996] proved that under the conditions above

$$|d\omega|_2 + |d^\star\omega|_2 \ge \tfrac{1}{2}((n-1)\sqrt{\kappa} - 2p)|\omega|_2.$$

This only makes sense if $\kappa > \kappa'_p := 4p^2/(n-1)^2$, which again forces $2p < n-1$, but notice that (3-1) gives a better pinching constant if $1 \le p < (n-1)/2$. It is observed in the same work that

$$|d\eta|_2 + |d^\star\eta|_2 \ge \tfrac{1}{2}((n-1)\sqrt{\kappa} - 2(n-p))|\eta|_2,$$

for any p-form η satisfying $\nu \wedge \eta = 0$ along $\partial \widetilde{M}$. Taking $p = n$ and using Hodge duality, we find that

$$|d\varphi|_2 \ge \tfrac{1}{2}(n-1)\sqrt{\kappa}|\varphi|_2, \tag{3-2}$$

for any compactly supported function φ satisfying the Neumann boundary condition. In other words, (3-1) holds for $p = 0$ as well. This transplants to our setting a famous estimate by McKean [1970]. Observe however that the assumption on the fundamental group is essential in (3-2) as the first Neumann eigenvalue of geodesic balls in hyperbolic space converges to zero as the radius goes to infinity [Chavel 1984]. Thus, (3-2) illustrates a situation where a topological condition on a compact ∂-manifold poses spectral constraints on its universal cover.

With these estimates at hand it is rather straightforward to establish vanishing theorems for L^2-harmonic forms. For this we consider (M, g) as in Proposition 3.2 and define the absolute Hodge Laplacian Δ_p^{abs} on $\widetilde{M}$ with domain $\mathcal{D}(\Delta_p^{\mathrm{abs}}) = \{\omega \in H^2(\wedge^p T^*\widetilde{M});\, \omega_{\mathrm{nor}} = 0,\, (d\omega)_{\mathrm{nor}} = 0\}$. Let $\lambda_p^{\mathrm{abs}}(\tilde{g}) = \inf \operatorname{Spec}(\Delta_p^{\mathrm{abs}})$. The spectral argument in [Schick 1996, Section 6] then provides, under the conditions of Proposition 3.2, the lower bound

$$\lambda_p^{\mathrm{abs}}(\tilde{g}) \geq \tfrac{1}{4}((n-p-1)\sqrt{\kappa} - p)^2. \tag{3-3}$$

We remark that the proof in [Schick 1996] uses induction in p starting at $p = 0$, which corresponds to (3-2). Here we use this to prove the following vanishing result.

Proposition 3.3. *Let (M, g) be a compact and convex ∂-manifold with infinite fundamental group and assume that g is κ_p-negatively pinched where $2 \leq 2p < n-1$. Then $\lambda_p^{\mathrm{abs}}(\tilde{g}) > 0$ and $(\widetilde{M}, \tilde{g})$ carries no nontrivial absolute L^2-harmonic p-form. Hence, $H^p_{(2),\mathrm{abs}}(\widetilde{M}, \tilde{g})$ vanishes.*

Proof. The assumptions imply that $\kappa_p < 1$, so we can find $\kappa_p < \kappa < 1$ such that $-1 \leq K_{\mathrm{sec}}(\tilde{g}) \leq -\kappa$. The result follows from (3-3) and the L^2 Hodge–de Rham theory in [Schick 1996; 1998]. □

4. The proofs of Theorems 1.1 and 1.2

Here we prove the main results of this work. Notice that if $(r_{(q)}, \rho_{(q)})$ is s.s.p. then

$$\sup_{x^0 \in K} \theta_q(x^0) < +\infty \quad \text{for any } K. \tag{4-1}$$

Also, if (α, β) is s.s.p. then $(\bar{\alpha}, \bar{\beta})$ is s.s.p. as well for any $\bar{\alpha} \geq \alpha$ and $\bar{\beta} \geq \beta$.

Proof of Theorem 1.1. If M is convex with respect to a κ_p-negatively pinched metric g_- then $H^p_{(2),\mathrm{abs}}(\widetilde{M}, \tilde{g}_-)$ vanishes by Proposition 3.3. On the other hand, by standard Hodge theory for compact ∂-manifolds [Taylor 2011], any nontrivial class in $H^p(M; \mathbb{R})$ can be represented by a nontrivial absolute harmonic p-form with respect to any metric g_+ on M. The lift of this form to $(\widetilde{M}, \tilde{g}_+)$ defines a nontrivial absolute harmonic p-form which is uniformly bounded. Now, if g_+ has both $(r_{(p\pm 1)}, \rho_{(p-1)})$ s.s.p. then the corresponding invariants of $\tilde{g}_+$ are s.s.p. as well, since the property is preserved by passage to covers; see Remark 5.1. In particular, (4-1) holds with $q = p \pm 1$. Thus we may apply Proposition 2.2 to conclude that $H^p_{(2),\mathrm{abs}}(\widetilde{M}, \tilde{g}_+) \neq \{0\}$. Since $H^p_{(2),\mathrm{abs}}(\widetilde{M}, \cdot\,)$ is a quasi-isometric invariant of the metric [Schick 1996] we obtain a contradiction which completes the proof. □

We now consider Theorem 1.2. For its proof we need an extension of a well-known result in [Lyons and Sullivan 1984] to our setting.

Proposition 4.1. *If (M, g) is a compact ∂-manifold and $\pi_1(M)$ is nonamenable then $(\widetilde{M}, \tilde{g})$ carries a nonconstant bounded absolute harmonic function.*

Proof. The argument in [Lyons and Sullivan 1984, Section 5] carries over to our case. More precisely, using the Neumann heat kernel we construct a natural $\pi_1(M)$-equivariant projection from $L^\infty_{\mathrm{abs}}(\widetilde{M})$, the space of absolute bounded functions, onto $\mathcal{H}^\infty_{\mathrm{abs}}(\widetilde{M}, \tilde{g})$, the space of bounded absolute harmonic functions. Also, there exists a $\pi_1(M)$-equivariant injection $l^\infty(\pi_1(M)) \hookrightarrow L^\infty_{\mathrm{abs}}(\widetilde{M})$. Hence, if $\mathcal{H}^\infty_{\mathrm{abs}}(\widetilde{M}, \tilde{g}) = \mathbb{R}$ the composition $l^\infty(\pi_1(M)) \to \mathbb{R}$ defines an invariant mean. □

Proof of Theorem 1.2. If M carries a metric g_- which is κ_1-negatively curved, then $H^1_{(2),\mathrm{abs}}(\widetilde{M}, \tilde{g}_-)$ vanishes. On the other hand, by Proposition 4.1, for any metric g_+ on M, $(\widetilde{M}, \tilde{g}_+)$ carries a nonconstant bounded absolute harmonic function, say f. This implies that reflecting Brownian motion in $(\widetilde{M}, \tilde{g}_+)$ is transient and in particular

$$\sup_{x^0 \in K} \int_0^{+\infty} (P_t\, d^\star \phi)(x^0)\, dt < +\infty,$$

for any $K \subset \widetilde{M}$ and compactly supported 1-form ϕ as in Proposition 2.1; see [Grigor'yan 1999, Theorem 5.1]. Assuming that g_- is such that the corresponding pair $(r_{(2)}, \rho_{(2)})$ is s.s.p. we can apply Proposition 2.2 because $\psi = df$ is a bounded absolute harmonic 1-form; see Lemma 4.1 below. Hence, $H^1_{(2),\mathrm{abs}}(\widetilde{M}, \tilde{g}_+) \neq \{0\}$ and we get a contradiction. Thus, the proof of Theorem 1.2 is complete as soon as the next lemma is established. □

Lemma 4.1. *If f is a uniformly bounded absolute function as above then the absolute harmonic 1-form $\phi = df$ is uniformly bounded as well.*

Proof. Assume that $|f| \leq K$. The Bismut–Elworthy–Li formula in [Elworthy and Li 1994, Theorem 3.1] holds for our reflecting Brownian motion x^t. Hence, if $v^0 \in T_{x^0}\widetilde{M}$ and $P_t = e^{-t\Delta_0^{\mathrm{abs}}/2}$ then

$$d(P_t f)_{x^0}(v^0) = \frac{1}{t}\mathbb{E}_{x^0}\left(f(x^t) \int_0^t \langle v^s, dx^s \rangle_{x^s} \right), \qquad t > 0,$$

where v^t is defined by (5-4) below. Since f is harmonic, $P_t f = f$. Thus,

$$|df_{x^0}(v^0)| \leq \frac{|v^0|}{t} \sup\nolimits_{\widetilde{M}} f \int_0^t ds \leq K|v^0|,$$

as desired. □

5. A Feynman–Kac formula on ∂-manifolds

In this final section we explain how the method put forward in [Airault 1976; Hsu 2002a] can be adapted to prove a Feynman–Kac-type formula for q-forms on ∂-manifolds. As an illustration of the flexibility of the method we also include a similar formula for spinors evolving by the heat semigroup of the Dirac Laplacian on spin^c ∂-manifolds. These results are presented in the second and third subsections, respectively, after some preparatory material in the first subsection.

The Eells–Elworthy–Malliavin approach. Let (N, h) be a Riemannian ∂-manifold of dimension n. As in Section 2 we assume that $(N, \partial N, h)$ has bounded geometry. Let $\pi : P_{\mathrm{O}_n}(N) \to N$ be the orthonormal frame bundle of N. This is a principal bundle with structural group O_n, the orthogonal group in dimension n. Any orthogonal representation $\zeta : \mathrm{O}_n \to \mathrm{End}(V)$ gives rise to the associated vector bundle $\mathcal{E}_\zeta = P_{\mathrm{O}_n}(N) \times_\zeta V$, which comes endowed with a natural metric and compatible connection derived from h and its Levi-Civita connection ∇. Moreover, any section $\sigma \in \Gamma(\mathcal{E}_\zeta)$ can be identified to its lift $\sigma^\dagger : P_{\mathrm{O}_n}(N) \to V$, which is ζ-equivariant in the sense that $\sigma^\dagger(ug) = \zeta(g^{-1})(\sigma^\dagger(u))$, $u \in P_{\mathrm{O}_n}(N)$, $g \in \mathrm{O}_n$. Also, we recall that in terms of lifts, covariant derivation essentially corresponds to Lie differentiation along horizontal tangent vectors.

Any bundle $\mathcal{E}_\zeta$ as above comes equipped with a second order elliptic operator $\Delta^B = -\operatorname{tr}_h \nabla^2 : \Gamma(\mathcal{E}_\zeta) \to \Gamma(\mathcal{E}_\zeta)$, the Bochner Laplacian. Here, ∇^2 is the standard Hessian operator acting on sections. Given an algebraic (zero order) self-adjoint map $\mathcal{R} \in \Gamma(\mathrm{End}(\mathcal{E}_\zeta))$ we can form the elliptic operator

$$\Delta = \Delta^B + \mathcal{R}$$

acting on $\Gamma(\mathcal{E}_\zeta)$. Standard results [Eichhorn 2007; Schick 1996; 1998] imply that the heat semigroup $P_t = e^{-t\Delta/2}$ has the property that, for any $\sigma_0 \in L^2 \cap L^\infty$, $\sigma_t = P_t\sigma_0$ solves the heat equation

$$\frac{\partial \sigma_t}{\partial t} + \frac{1}{2}\Delta\sigma_t = 0, \qquad \lim_{t\to 0} \sigma_t = \sigma_0, \tag{5-1}$$

where we eventually impose elliptic boundary conditions in case $\partial N \neq \varnothing$.

An important question concerning us here is whether the solutions of (5-1) admit a stochastic representation in terms of Brownian motion on N. If $\partial N = \varnothing$ this problem admits a very elegant solution in great generality and a Feynman–Kac formula is available [Elworthy 1988; Güneysu 2010; Hsu 1999; 2002b; Malliavin 1974; Stroock 2000]. Moreover, this representation permits us to estimate the solutions in terms of the overall expectation of $\mathcal{R}$ with respect to the diffusion process; see (5-5)–(5-6) below. However, in the presence of a boundary it is well known that the problem is much harder to handle; see [Hsu 2002a].

Let us assume that N has a nonempty boundary endowed with an inward unit normal field ν. We first briefly recall how reflecting Brownian motion is defined on N. We take for granted that Brownian motion $\{b^t\}$ on $\mathbb{R}^n$ is defined. This is the diffusion process which has half the standard Laplacian $\sum_i \partial_i^2$ as generator. To transplant this to N we make use of the so-called Eells–Elworthy–Malliavin approach [Elworthy 1988; Eells and Elworthy 1971; Hsu 1999; 2002b; Stroock 2000]. Note that any $u \in P_{\mathrm{O}_n}(N)$ defines an isometry $u : \mathbb{R}^n \to T_x N,\ x = \pi(u)$. Also, the Levi-Civita connection on TN lifts to an Ehresmann connection on $P_{\mathrm{O}_n}(M)$ which determines fundamental horizontal vector fields H_i, $i = 1, \dots, n$. As explained in [Hsu 2002b, Chapter 2], these elementary remarks naturally lead to an identification of semimartingales on $\mathbb{R}^n$, horizontal semimartingales on $P_{\mathrm{O}_n}(M)$ and semimartingales on M. Thus, on $P_{\mathrm{O}_n}(N)$ we may consider the stochastic differential equation

$$du^t = \sum_{i=1}^{n} H_i(u^t) \circ db_i^t + \nu^\dagger(u^t)\, dl^t, \tag{5-2}$$

which has a unique solution $\{u^t\}$ starting at any initial frame u^0. This is a horizontal reflecting Brownian motion on $P_{\mathrm{O}_n}(N)$ and its projection $x^t = \pi u^t$ defines reflecting Brownian motion on N starting at $x^0 = \pi u^0$. Moreover, l^t is the associated boundary local time. Notice that x^t satisfies

$$dx^t = \sum_{i=1}^{n} X_i(x^t) \circ db_i^t + \nu(x^t)\, dl^t, \qquad X_i = \pi_* H_i, \tag{5-3}$$

so that if F^t is the corresponding stochastic flow, i.e., $x^t = F^t(x^0)$, then $v^t = dF^t_{x^0}(v^0)$, $v^0 \in T_{x^0}N$, satisfies the derivative equation

$$dv^t = \sum_{i=1}^{n} (\nabla X_i)(v^t) \circ db_i^t + (\nabla \nu)(v^t)\, dl^t. \tag{5-4}$$

Remark 5.1. Due to the obvious functorial character of this construction it is not hard to obtain highly desirable properties of Brownian motion. For instance, if the manifold splits as an isometric product of two other manifolds then its Brownian motion is the product of Brownian motions on the factors. In particular, if $N = X \times Y$, where Y is a compact ∂-manifold, then Brownian motion in N is transient if and only if the same happens to X. Also, if $\widetilde{N} \to N$ is a normal Riemannian covering then Brownian motion in $\widetilde{N}$ projects down to Brownian motion in N. From this it is obvious that a pair (α, β) on $(N, \partial N)$ is s.s.p. if and only if its lift $(\tilde{\alpha}, \tilde{\beta})$ on $(\widetilde{N}, \partial\widetilde{N})$ is s.s.p. as well.

We now describe how this formalism leads to an elegant approach to Feynman–Kac-type formulas. Let $\mathcal{A} \in \Gamma(\mathrm{End}(\mathcal{E}_\zeta|_{\partial N}))$ be a pointwise self-adjoint map.

In practice, $\mathcal{A}$ relates to the zero order piece of the given boundary conditions. In analogy with the boundaryless case, Itô's calculus suggests considering the multiplicative functional $M^t \in \mathrm{End}(V)$ satisfying

$$dM^t + M^t\big(\tfrac{1}{2}\mathcal{R}^\dagger\,dt + \mathcal{A}^\dagger\,dl^t\big) = 0, \quad M^0 = I.$$

Standard results imply that a solution exists along each path u^t. We now apply Itô's formula to the process $M^t\sigma^\dagger(T-t,u^t)$, $0 \le t \le T$, where σ is a (time-dependent) section of $\mathcal{E}_\zeta$. With the help of (5-2) we obtain

$$\begin{aligned} dM^t\sigma^\dagger(T-t,u^t) = [M^t\mathcal{L}_H\sigma^\dagger(T-t,u^t), db^t] - M^t L^\dagger\sigma^\dagger(T-t,u^t)\,dt \\ +M^t(\mathcal{L}_{\nu^\dagger} - \mathcal{A}^\dagger)\sigma^\dagger(T-t,u^t)\,dl^t, \end{aligned}$$

where $\mathcal{L}$ is the Lie derivative,

$$[M^t\mathcal{L}_H\sigma^\dagger(T-t,u^t), db_t]_i = \sum_{j=1}^{\dim V}\sum_{k=1}^{n} M^t_{ij}\mathcal{L}_{H_k}\sigma^\dagger_j(T-t,u^t)\,db^t_k,$$

and

$$L^\dagger = \frac{\partial}{\partial t} + \frac{1}{2}(\Delta_B^\dagger + \mathcal{R}^\dagger)$$

is the lifted heat operator, with $\Delta_B^\dagger = -\sum_k \mathcal{L}^2_{H_k}$ being the horizontal Bochner Laplacian. Notice that in case $\partial N = \varnothing$ and σ satisfies (5-1) the computation gives

$$dM^t\sigma^\dagger(T-t,u^t) = [M^t\mathcal{L}_H\sigma^\dagger(T-t,u^t), db^t],$$

which characterizes $M^t\sigma^\dagger(T-t,u^t)$ as a martingale. Equating the expectations of this process at $t=0$ and $t=T$ yields the celebrated Feynman–Kac formula

$$\sigma^\dagger(t,u^0) = \mathbb{E}_{u^0}(M^t\sigma^\dagger(0,u^t)), \tag{5-5}$$

where $dM^t = -M^t\mathcal{R}^\dagger\,dt/2$ [Elworthy 1988; Hsu 1999; 2002b; Güneysu 2010; Stroock 2000]. From this we easily obtain the well-known estimate

$$|\sigma(t,x^0)| \le \mathbb{E}_{x^0}\bigg(|\sigma(0,x^t)|\exp\bigg(-\frac{1}{2}\int_0^t \underline{\mathcal{R}}(x^s)\,ds\bigg)\bigg), \tag{5-6}$$

where $\underline{\mathcal{R}}(x)$ is the least eigenvalue of $\mathcal{R}(x)$. However, if $\partial N \neq \varnothing$ the calculation merely says that $\{M^t\}$ is the multiplicative functional associated with the operator L under boundary conditions

$$(\nabla_\nu - \mathcal{A})\sigma = 0. \tag{5-7}$$

As we shall see below through examples, (5-7) is too stringent to encompass boundary conditions commonly occurring in applications, which usually are of mixed type.

The Feynman–Kac formula for absolute differential forms. It turns out that natural elliptic boundary conditions do not quite fit into the prescription in (5-7). Hence, the formalism in the previous subsection does not apply as presented. We illustrate this issue by considering the case $\zeta = \wedge^q \mu_n^*$, where μ_n is the standard representation of O_n on $\mathbb{R}^n$, so that $\mathcal{E}_\zeta$ is the bundle of q-forms over N. In this case, $\mathcal{A}$ is explicitly described in terms of the second fundamental form of ∂N but degeneracies occur due to the splitting of forms into tangential and normal components which is inherent to absolute boundary conditions.

The splitting is determined by the "fermionic relation" $\nu\lrcorner\nu\wedge+\nu\wedge\nu\lrcorner=I$, which induces an orthogonal decomposition

$$\wedge^q T^*N|_{\partial N} = \mathrm{Ran}(\nu\lrcorner\nu\wedge) \oplus \mathrm{Ran}(\nu\wedge\nu\lrcorner),$$

and we denote by $\Pi_{\tan}$ and Π_{nor} the orthogonal projections onto the factors. As is clear from the notation, these maps project onto the space of tangential and normal q-forms, respectively.

Let

$$A : T\partial N \to T\partial N, \qquad AX = -\nabla_X \nu,$$

be the second fundamental form of ∂N, which we extend to $TN|_{\partial N}$ by declaring that $A\nu = 0$. This induces the pointwise self-adjoint map $\mathcal{A}_q \in \mathrm{End}(\wedge^q T^*N|_{\partial N})$,

$$(\mathcal{A}_q\omega)(X_1, \ldots, X_q) = \sum_i \omega(X_1, \ldots, AX_i, \ldots, X_q).$$

Notice that $\Pi_{\mathrm{nor}}\mathcal{A}_q\omega = 0$, that is, $\mathcal{A}_q\omega$ only has tangential components. In order to determine the tangential coefficients of $\mathcal{A}_q\omega$ we fix an orthonormal frame $\{e_1, \ldots, e_{n-1}\}$ in $T\partial N$ which is principal at $x \in \partial N$ in the sense that $Ae_i = \rho_i e_i$. We then find that, at x,

$$(\mathcal{A}_q\omega)(e_{i_1}, \ldots, e_{i_q}) = \left(\sum_{j=1}^{q} \rho_{i_j}\right)\omega(e_{i_1}, \ldots, e_{i_q}). \tag{5-8}$$

The next result is inspired by [Hsu 2002a, Lemma 4.1]; see also [Yano 1970; Donnelly and Li 1982] for similar computations.

Proposition 5.1. *A q-form ω is absolute if and only if its lift $\omega^\dagger$ satisfies*

$$\Pi^\dagger_{\mathrm{nor}}\omega^\dagger = 0 \quad \text{and} \quad \Pi^\dagger_{\tan}(\mathcal{L}_{\nu^\dagger} - \mathcal{A}^\dagger_q)\omega^\dagger = 0 \qquad \textit{on } \partial P_{\mathrm{O}_n}(N). \tag{5-9}$$

Proof. We work downstairs on ∂N and drop the dagger from the notation. First, $\omega_{\mathrm{nor}} = 0$ means that $\omega = \omega_{\tan} + \nu\wedge\omega_{\mathrm{nor}} = \omega_{\tan}$, that is, $\Pi^\dagger_{\mathrm{nor}}\omega^\dagger = 0$. On the other hand, in terms of the principal frame $\{e_i\}$ above,

$$\begin{aligned}
\nu \lrcorner d\omega(e_{i_1},\dots,e_{i_q}) &= d\omega(\nu,e_{i_1},\dots,e_{i_q})\\
&= \nu(\omega(e_{i_1},\dots,e_{i_q})) + \sum_j (-1)^j e_{i_j}(\omega(\nu,e_{i_1},\dots,\widehat{e_{i_j}},\dots,e_{i_q}))\\
&\qquad + \sum_j (-1)^j \omega([\nu,e_{i_j}],e_{i_1},\dots,\widehat{e_{i_j}},\dots,e_{i_q})\\
&\qquad\quad + \sum_{1\le j<k} (-1)^{j+k}\omega([e_{i_j},e_{i_k}],\nu,e_{i_1},\dots,\widehat{e_{i_j}},\dots,\widehat{e_{i_k}},\dots,e_{i_q})\\
&= \nu(\omega(e_{i_1},\dots,e_{i_q})) + \sum_j (-1)^j e_{i_j}((\nu\lrcorner\omega)(e_{i_1},\dots,\widehat{e_{i_j}},\dots,e_{i_q}))\\
&\qquad - \sum_j \omega(e_{i_1},\dots,[\nu,e_{i_j}],\dots,e_{i_q})\\
&= \nu(\omega(e_{i_1},\dots,e_{i_q})) - \sum_j \omega(e_{i_1},\dots,\nabla_\nu e_{i_j},\dots,e_{i_q})\\
&\qquad - \Big(\sum_j \rho_{i_j}\Big)\omega(e_{i_1},\dots,e_{i_q}),
\end{aligned}$$

where we used that $[e_{i_j}, e_{i_k}] = 0$, certainly a justifiable assumption, and $\nu\lrcorner\omega = 0$. But

$$\nu(\omega(e_{i_1},\dots,e_{i_q})) = (\nabla_\nu\omega)(e_{i_1},\dots,e_{i_q}) + \sum_j \omega(e_{i_1},\dots,\nabla_\nu e_{i_j},\dots,e_{i_q}),$$

so we obtain

$$\nu\lrcorner d\omega(e_{i_1},\dots,e_{i_q}) = \Big(\nabla_\nu - \sum_j \rho_{i_j}\Big)\omega(e_{i_1},\dots,e_{i_q}).$$

The results follow in view of (5-8). □

This proposition shows that absolute boundary conditions are of mixed type, namely, they are Dirichlet in normal directions and Neumann in tangential directions. This should be compared with (5-7), which is of pure Neumann type. This confirms that Itô's calculus is insensitive to the projections defining absolute boundary conditions. To remedy this we proceed as in [Hsu 2002a]. We can write the boundary condition as the superposition of two independent components, namely,

$$\Pi^\dagger_{\tan}(\mathcal{L}_{\nu^\dagger} - \mathcal{A}^\dagger_q)\omega^\dagger - \Pi^\dagger_{\mathrm{nor}}\omega^\dagger = 0.$$

The key idea, which goes back to [Airault 1976], is to fix $\epsilon > 0$ and replace $\Pi^\dagger_{\tan}$ by $\Pi^\dagger_{\tan} + \epsilon I$ above, so the condition becomes

$$(\mathcal{L}_{\nu^\dagger} - (\mathcal{A}^\dagger_q + \epsilon^{-1}\Pi^\dagger_{\mathrm{nor}}))\omega^\dagger = 0,$$

which in a sense is the best we can reach in terms of resemblance to (5-7). The next step is to solve for $\mathcal{M}_\epsilon^t \in \mathrm{End}(\wedge^q \mathbb{R}^n)$ in

$$d\mathcal{M}_\epsilon^t + \mathcal{M}_\epsilon^t\big(\tfrac{1}{2}R_q^\dagger(u^t)\,dt + (\mathcal{A}_q^\dagger(u^t) + \epsilon^{-1}\Pi_{\mathrm{nor}}^\dagger(u^t))\,dl^t\big) = 0, \quad \mathcal{M}_\epsilon^0 = I. \tag{5-10}$$

Proposition 5.2. *For all $\epsilon > 0$ such that $\epsilon^{-1} \geq \underline{\rho}_{(q)}$ we have*

$$|\mathcal{M}_\epsilon^t| \leq \exp\biggl(-\frac{1}{2}\int_0^t r_{(q)}(x^s)\,ds - \int_0^t \rho_{(q)}(x^s)\,dl^s\biggr), \quad t > 0. \tag{5-11}$$

Proof. The same as in [Hsu 2002a, Lemma 3.1], once we take into account that, as is clear from (5-8), the sums $\sum_{j=1}^q \rho_{i_j}$ are the eigenvalues of $\Pi_{\tan}\mathcal{A}_q$. □

The following convergence result provides the crucial input in the argument.

Theorem 5.1. *As $\epsilon \to 0$, $\mathcal{M}_\epsilon^t$ converges to a multiplicative functional $\mathcal{M}^t$ in the sense that $\lim_{\epsilon\to 0}\mathbb{E}|\mathcal{M}_\epsilon^t - \mathcal{M}^t|^2 = 0$. Moreover, $\mathcal{M}^t\Pi_{\mathrm{nor}}^\dagger(u) = 0$ whenever $u \in \partial P_{O_n}(N)$.*

Proof. The rather technical proof of this result for $q = 1$ is presented in detail in [Hsu 2002a]. Fortunately, with the formalism above in place, it is not hard to check that the proof of the general case follows along the lines of the original argument. More precisely, in that work the letters P and Q denote normal and tangential projection, respectively. If we replace these symbols by Π_{nor} and $\Pi_{\tan}$, the proof there works here with only minor modifications. Therefore, it is omitted. □

We now have all the ingredients needed to prove the Feynman–Kac-type formula for differential forms.

Theorem 5.2. *Let ω_0 be an absolute L^2 q-form on N as above. If $P_t = e^{-t\Delta_q^{\mathrm{abs}}/2}$ is the corresponding heat semigroup, so that $\omega_t = P_t\omega_0$ provides the solution to*

$$\frac{\partial\omega_t}{\partial t} + \frac{1}{2}\Delta_q^{\mathrm{abs}}\omega_t = 0, \quad \lim_{t\to 0}\omega_t = \omega_0, \quad \nu\lrcorner\omega_t = 0, \quad \nu\lrcorner d\omega_t = 0, \tag{5-12}$$

then the following Feynman–Kac formula holds:

$$\omega_t^\dagger(u^0) = \mathbb{E}_{u^0}(\mathcal{M}^t\omega_0^\dagger(u^t)), \tag{5-13}$$

where u_t is the horizontal reflecting Brownian motion starting at u_0. Consequently,

$$|\omega_t(x^0)| \leq \mathbb{E}_{x^0}\biggl(|\omega_0(x^t)|\exp\biggl(-\frac{1}{2}\int_0^t r_{(q)}(x^s)\,ds - \int_0^t \rho_{(q)}(x^s)\,dl^s\biggr)\biggr), \tag{5-14}$$

where $x^t = \pi u^t$.

Proof. Itô's formula and (5-2) yield

$$\begin{aligned} d\mathcal{M}_\epsilon^t\omega_{T-t}^\dagger(u^t) = [\mathcal{M}_\epsilon^t\mathcal{L}_H\omega_{T-t}^\dagger(u^t), db^t] - \mathcal{M}_\epsilon^t L^\dagger\omega_{T-t}^\dagger(u^t)\,dt \\ + \mathcal{M}_\epsilon^t(\mathcal{L}_{\nu^\dagger} - \mathcal{A}^\dagger - \epsilon^{-1}\Pi_{\mathrm{nor}}^\dagger)\omega_{T-t}^\dagger(u^t)\,dl^t. \end{aligned}$$

If ω_t is a solution of (5-12) then the second term on the right-hand side drops out. Moreover, by Proposition 5.1 the same happens to the term involving ϵ^{-1}. Sending $\epsilon \to 0$ we end up with

$$d\mathcal{M}^t\omega^\dagger_{T-t}(u^t) = [\mathcal{M}^t\mathcal{L}_H\omega^\dagger_{T-t}(u^t), db^t] + \mathcal{M}^t\Pi^\dagger_{\tan}(\mathcal{L}_{\nu^\dagger} - \mathcal{A}^\dagger)\omega^\dagger_{T-t}(u^t)\, dl^t,$$

where the insertion of $\Pi^\dagger_{\tan}$ in the last term is legitimate due to the last assertion in Theorem 5.1. By Proposition 5.1 this actually reduces to

$$d\mathcal{M}^t\omega^\dagger_{T-t}(u^t) = [\mathcal{M}^t\mathcal{L}_H\omega^\dagger_{T-t}(u^t), db^t],$$

which shows that $\mathcal{M}^t\omega^\dagger_{T-t}(u^t)$ is a martingale. Thus, (5-13) follows by equating the expectations at $t = 0$ and $t = T$. Finally, (5-14) follows from (5-11). □

The estimate (5-14) has many interesting consequences. We illustrate its usefulness by mentioning a semigroup domination result which can be proved as in [Elworthy and Rosenberg 1988, Theorem 3A]; see also [Bérard 1990; Donnelly and Li 1982; Elworthy 1988; Hsu 1999; 2002b] for similar results.

Theorem 5.3. *Let $(N, \partial N, h)$ be as above and assume that $\rho_{(q)} \geq 0$ for some $1 \leq q \leq n-1$. Then there holds*

$$|e^{-t\Delta^{\mathrm{abs}}_q/2}(x, y)| \leq \binom{n}{q} e^{-\underline{r}_{(q)}t/2} e^{-t\Delta^{\mathrm{abs}}_0/2}(x, y), \quad x, y \in N,\ t > 0,$$

where $\underline{r}_{(q)} = \inf_{x\in N} r_{(q)}(x)$. In particular, if $\lambda^{\mathrm{abs}}_0(h) + \underline{r}_{(q)} \geq 0$ and $r_{(q)} > \underline{r}_{(q)}$ somewhere then N carries no nontrivial absolute L^2-harmonic q-form.

The Feynman–Kac formula for spinors. Let N be a spin^c ∂-manifold [Friedrich 2000]. As usual we assume that $(N, \partial N, h)$ has bounded geometry. Let $\mathbb{S}N = P_{\mathrm{Spin}^c_n}(N) \times_\zeta V$ be the spin^c bundle of N, where ζ is the complex spin representation. Recall that $P_{\mathrm{Spin}^c_n}(N)$ is a Spin^c principal bundle double covering $P_{\mathrm{SO}_n}(N) \times P_{\mathrm{U}_1}(N)$, where $P_{\mathrm{U}_1}(N)$ is the U_1 principal bundle associated to the auxiliary complex line bundle $\mathcal{F}$. After fixing a unitary connection C on $\mathcal{F}$, the Levi-Civita connection on TN induces a metric connection on $\mathbb{S}N$, still denoted ∇. The corresponding Dirac operator $D : \Gamma(\mathbb{S}N) \to \Gamma(\mathbb{S}N)$ is locally given by

$$D\psi = \sum_{i=1}^n \gamma(e_i)\nabla_{e_i}\psi, \quad \psi \in \Gamma(\mathbb{S}N),$$

where $\{e_i\}_{i=1}^n$ is a local orthonormal frame and $\gamma : \mathrm{Cl}(TN) \to \mathrm{End}(\mathbb{S}N)$ is the Clifford product. The Dirac Laplacian operator is

$$D^2\psi = \Delta_B\psi + \mathfrak{R}\psi, \tag{5-15}$$

where
$$\mathfrak{R}\psi = \frac{R}{4}\psi + \frac{1}{2}\gamma(i\Omega).$$
Here, R is the scalar curvature of h and $i\Omega$ is the curvature 2-form of C.

The spin^c bundle $\mathbb{S}N|_{\partial N}$, obtained by restricting $\mathbb{S}N$ to ∂N, becomes a Dirac bundle if its Clifford product is

$$\gamma^{\intercal}(X)\psi = \gamma(X)\gamma(\nu)\psi, \quad X \in \Gamma(T\partial N), \quad \psi \in \Gamma(\mathbb{S}N|_{\partial N}),$$

and its connection is

$$\nabla^{\intercal}_X\psi = \nabla_X\psi - \tfrac{1}{2}\gamma^{\intercal}(AX)\psi, \tag{5-16}$$

where as usual $A = -\nabla\nu$ is the second fundamental form of ∂N; see [Nakad and Roth 2013]. The associated Dirac operator $D^{\intercal} : \Gamma(\mathbb{S}N|_{\partial N}) \to \Gamma(\mathbb{S}N|_{\partial N})$ is

$$D^{\intercal}\psi = \sum_{j=1}^{n-1}\gamma^{\intercal}(e_j)\nabla^{\intercal}_{e_j}\psi,$$

where the frame has been adapted so that $e_n = \nu$. Imposing that $Ae_j = \rho_j e_j$, where ρ_j are the principal curvatures of ∂N, a direct computation shows that

$$D^{\intercal}\psi = \frac{K}{2}\psi + \sum_{j=1}^{n-1}\gamma(e_j)\nabla_{e_j}\psi,$$

where $K = \operatorname{tr} A$ is the mean curvature. It follows that this tangential Dirac operator enters into the boundary decomposition of D, namely,

$$-\gamma(\nu)D = \nabla_\nu + D^{\intercal} - \frac{K}{2}, \tag{5-17}$$

which by its turn appears in Green's formula for the Dirac Laplacian

$$\int_N \langle D^2\psi, \xi\rangle\, dN = \int_N \langle D\psi, D\xi\rangle\, dN - \int_{\partial N}\langle\gamma(\nu)D\psi, \xi\rangle\, d\partial N, \tag{5-18}$$

where ψ and ξ are compactly supported. Also, since $\gamma^{\intercal}(e_j)\gamma(\nu) = -\gamma(\nu)\gamma^{\intercal}(e_j)$ and $\nabla^{\intercal}_{e_j}\gamma(\nu) = \gamma(\nu)\nabla^{\intercal}_{e_j}$, we see that

$$D^{\intercal}\gamma(\nu) = -\gamma(\nu)D^{\intercal}. \tag{5-19}$$

Now fix a nontrivial orthogonal projection $\Pi \in \Gamma(\mathrm{End}(\mathbb{S}N|_{\partial N}))$ and set $\Pi_+ = \Pi$ and $\Pi_- = I - \Pi$. It is clear from (5-17) – (5-18) that any of the boundary conditions

$$\Pi_\pm\psi = 0, \quad \Pi_\mp\Big(\nabla_\nu + D^{\intercal} - \frac{K}{2}\Big)\psi = 0, \tag{5-20}$$

turns the Dirac Laplacian D^2 into a formally self-adjoint operator. The next definition isolates a notion of compatibility between the tangential Dirac operator and the projections which will allow us to get rid of the middle term in the second condition above.

Definition 5.1. We say that the tangential Dirac operator $D^\intercal$ intertwines the projections if $\Pi_\pm D^\intercal = D^\intercal \Pi_\mp$.

Remark 5.2. If $D^\intercal$ intertwines the projections then $\Pi_\pm D^\intercal \Pi_\pm = D^\intercal \Pi_\mp \Pi_\pm = 0$. Equivalently, $\langle D^\intercal \Pi_\pm \psi, \Pi_\pm \xi\rangle = 0$ for any spinors ψ and ξ.

Proposition 5.3. *Under the conditions above assume further that $D^\intercal$ intertwines the projections as in Definition 5.1. Then a spinor $\psi \in \Gamma(\mathbb{S}N|_{\partial N})$ satisfies the boundary conditions (5-20) if and only if its lift $\psi^\dagger : P_{\mathrm{Spin}_n^c}(N) \to V$ satisfies*

$$\Pi_\pm^\dagger \psi^\dagger = 0 \quad \text{and} \quad \Pi_\mp^\dagger \Big(\mathcal{L}_{\nu^\dagger} - \frac{K^\dagger}{2}\Big)\psi^\dagger = 0 \quad \text{on} \quad \partial P_{\mathrm{Spin}_n^c}(N). \tag{5-21}$$

Proof. Obvious in view of (5-20) and Remark 5.2. □

We can now proceed exactly as in the previous subsection. We assume that (5-21) gives rise to a self-adjoint elliptic realization of D^2 and we denote by $e^{-tD^2/2}$ the corresponding heat semigroup [Grubb 2003]. We lift everything in sight to $P_{\mathrm{Spin}_n^c}(N)$ and consider there the functional $\mathcal{M}_\epsilon^t$ defined by

$$d\mathcal{M}_\epsilon^t + \mathcal{M}_\epsilon^t \Big(\frac{1}{2}\mathfrak{R}^\dagger(u^t)dt + \Big(\frac{1}{2}K^\dagger(u^t) + \epsilon^{-1}\Pi_+^\dagger(u^t)\Big)\, dl^t\Big) = 0, \quad \mathcal{M}_\epsilon^0 = I.$$

The limiting functional $\mathcal{M}^t$, whose existence is guaranteed by the analogue of Theorem 5.1, appears in the corresponding Feynman–Kac formula.

Theorem 5.4. *Let $\psi_0 \in \Gamma(\mathbb{S}N)$ be a spinor satisfying any of the boundary conditions (5-20), where we assume $D^\intercal$ intertwines the projections as in Definition 5.1. If $\psi_t = e^{-tD^2/2}\psi_0$ is the solution to*

$$\frac{\partial \psi_t}{\partial t} + \frac{1}{2}D^2\psi_t = 0, \quad \lim_{t\to 0}\psi_t = \psi_0, \quad \Pi_\pm \psi_t = 0, \quad \Pi_\mp\Big(\nabla_\nu - \frac{K}{2}\Big)\psi_t = 0, \tag{5-22}$$

then the following Feynman–Kac formula holds:

$$\psi_t^\dagger(u^0) = \mathbb{E}_{u^0}(\mathcal{M}_t \psi_0^\dagger(u^t)), \tag{5-23}$$

where u^t is the horizontal reflecting Brownian motion on $P_{\mathrm{Spin}_n^c}(N)$ starting at u^0. As a consequence,

$$|\psi_t(x^0)| \le \mathbb{E}_{x^0}\Big(|\psi_0(x^t)| \exp\Big(-\frac{1}{2}\int_0^t \mathfrak{r}(x^s)\, ds - \frac{1}{2}\int_0^t K(x^s)\, dl^s\Big)\Big), \tag{5-24}$$

where $\mathfrak{r}(x) = \inf_{|\psi|=1}\langle \mathfrak{R}(x)\psi, \psi\rangle$.

Proof. The same as in Theorem 5.2. □

It is worthwhile to state the analogue of Theorem 5.3 for spinors.

Theorem 5.5. *Let (N, h) be a* spin^c *∂-manifold as above and assume that $K \geq 0$ along ∂N. Let $e^{-tD^2/2}$ be the heat semigroup of the Dirac Laplacian acting on spinors subject to boundary conditions as in Theorem 5.4. Then*

$$|e^{-tD^2/2}(x, y)| \leq 2^{[n/2]+1} e^{-\underline{\mathfrak{r}}t/2} e^{-t\Delta_0^{\mathrm{abs}}/2}(x, y), \quad x, y \in N, t > 0,$$

where $\underline{\mathfrak{r}} = \inf_{x \in N} \mathfrak{r}(x)$. In particular, if $\lambda_0^{\mathrm{abs}}(h) + \underline{\mathfrak{r}} \geq 0$ and $\mathfrak{r} > \underline{\mathfrak{r}}$ somewhere then N carries no nontrivial L^2-harmonic spinor satisfying the given boundary conditions.

We now discuss a couple of examples of local boundary conditions for spinors to which Theorem 5.4 applies.

Example 5.1. (Chirality boundary condition) A chirality operator on a spin^c ∂-manifold $(N, \partial N)$ is an orthogonal and parallel involution $Q \in \Gamma(\mathrm{End}(\mathbb{S}N))$ which anticommutes with the Clifford product with any tangent vector. Examples include the Clifford product with the complex volume element in an even-dimensional spin manifold and with the timelike unit normal to an immersed spacelike hypersurface in a Lorentzian spin manifold. It is easy to check that $D^\intercal Q = QD^\intercal$ and $D^\intercal\gamma(\nu) = -\gamma(\nu)D^\intercal$. Given any such Q define the boundary chirality operator $\hat{Q} = \gamma(\nu)Q \in \Gamma(\mathrm{End}(\mathbb{S}N)|_{\partial N})$, which is still an orthogonal and parallel involution with associated projections given by

$$\Pi_\pm = \tfrac{1}{2}(I \mp \hat{Q}). \tag{5-25}$$

Since $D^\intercal\hat{Q} = D^\intercal\gamma(\nu)Q = -\gamma(\nu)QD^\intercal = -\hat{Q}D^\intercal$, we conclude that $D^\intercal\Pi_\pm = \Pi_\mp D^\intercal$, that is, $D^\intercal$ intertwines the projections. Thus, Theorem 5.4 applies to the self-adjoint elliptic realization of D^2 under this boundary condition.

Example 5.2 (MIT bag boundary condition). This time we choose $\hat{Q} = i\gamma(\nu)$, an involution which clearly satisfies $D^\intercal\hat{Q} = -\hat{Q}D^\intercal$. Thus, $D^\intercal$ intertwines the projections exactly as in the previous example and Theorem 5.4 again applies to the self-adjoint elliptic realization of D^2 under this boundary condition.

Remark 5.3. For the sake of comparison, it is instructive to examine how absolute and relative boundary conditions for differential forms fit into the framework developed in this subsection. In particular, this helps to clarify the role played by Proposition 5.1 and its analogue for relative forms. Recall that $\wedge^\bullet T^*N$ has the structure of a Clifford module if we define the Clifford product by tangent vectors as $\gamma(v) = v \wedge - v \lrcorner$. The corresponding Dirac operator is $D = d + d^\star$, so that $D^2 = \Delta$, the Hodge Laplacian. If ω is a q-form then we know that along ∂N,

$$\omega = \omega_{\tan} + \nu \wedge \omega_{\mathrm{nor}} = \Pi_{\tan}\omega + \Pi_{\mathrm{nor}}\omega.$$

Instead of (5-16) we now have

$$\nabla_X^\intercal = \nabla_X^{\partial N} + \nu \wedge A(X)\lrcorner.$$

A direct computation then shows that, with respect to the splitting above, the boundary decomposition of D is

$$-\gamma(\nu)D\begin{pmatrix}\omega_{\mathrm{tan}}\\ \omega_{\mathrm{nor}}\end{pmatrix}=\begin{pmatrix}\nabla_\nu\omega_{\mathrm{tan}}\\ \nabla_\nu\omega_{\mathrm{nor}}\end{pmatrix}-\begin{pmatrix}\mathcal{A}_q^{\mathrm{tan}} & D_{\partial N}\\ D_{\partial N} & \mathcal{A}_{q-1}^{\mathrm{nor}}\end{pmatrix}\begin{pmatrix}\omega_{\mathrm{tan}}\\ \omega_{\mathrm{nor}}\end{pmatrix},$$

where $D_{\partial N}=d_{\partial N}+d^*_{\partial N}$ and in terms of a principal frame,

$$\mathcal{A}_q^{\mathrm{tan,nor}}=\sum_j\rho_j\Pi_{e_j}^{\mathrm{tan,nor}},$$

with $\Pi_v^{\mathrm{tan}}=v\wedge v\lrcorner$ and $\Pi_v^{\mathrm{nor}}=v\lrcorner v\wedge$. If $\omega_{\mathrm{nor}}=0$ then $\mathcal{A}_q^{\mathrm{tan}}\omega=\mathcal{A}_q\omega$ and the boundary integral in Green's formula for the Hodge Laplacian is

$$\int_{\partial N}(\langle\nabla_\nu\omega_{\mathrm{tan}},\omega_{\mathrm{tan}}\rangle-\langle\mathcal{A}_q\omega_{\mathrm{tan}},\omega_{\mathrm{tan}}\rangle-\langle D_{\partial N}\omega_{\mathrm{tan}},\omega_{\mathrm{tan}}\rangle)\,d\partial N.$$

However, the last term vanishes because the forms involved in the inner product have different parities. Thus, the right boundary conditions are

$$\Pi_{\mathrm{nor}}\omega=0,\quad \Pi_{\mathrm{tan}}(\nabla_\nu-\mathcal{A}_q)\omega=0. \tag{5-26}$$

Proposition 5.1 then shows that (5-26) defines absolute boundary conditions for the Hodge Laplacian. Similarly, if $\omega_{\mathrm{tan}}=0$ then $\omega=\nu\wedge\omega_{\mathrm{nor}}$ and $\mathcal{A}_{q-1}^{\mathrm{nor}}\omega_{\mathrm{nor}}=\star\mathcal{A}_{n-q}\star\omega_{\mathrm{nor}}$, where here $\star$ is the Hodge star operator of ∂N. This time the boundary integral is

$$\int_{\partial N}(\langle\nabla_\nu\omega_{\mathrm{nor}},\omega_{\mathrm{nor}}\rangle-\langle\star\mathcal{A}_{n-q}\star\omega_{\mathrm{nor}},\omega_{\mathrm{nor}}\rangle-\langle D_{\partial N}\omega_{\mathrm{nor}},\omega_{\mathrm{nor}}\rangle)\,d\partial N.$$

Again, the last term drops out and the correct boundary conditions are

$$\Pi_{\mathrm{tan}}\omega=0,\quad \Pi_{\mathrm{nor}}(\nabla_\nu-\star\mathcal{A}_{n-q}\star)\omega=0. \tag{5-27}$$

As in Proposition 5.1 we compute that

$$\begin{aligned}(\nabla_\nu-\star\mathcal{A}_{n-q}\star)\omega(\nu,e_{i_1},\dots,e_{i_{q-1}})&=(\nu\wedge d^\star\omega)(\nu,e_{i_1},\dots,e_{i_{q-1}})\\&=(\nu\lrcorner\nu\wedge d^\star\omega)(e_{i_1},\dots,e_{i_{q-1}})\\&=(\Pi_{\mathrm{tan}}\,d^\star\omega)(e_{i_1},\dots,e_{i_{q-1}})\end{aligned}$$

so that (5-27) can be rewritten as

$$\omega_{\mathrm{tan}}=0,\quad (d^\star\omega)_{\mathrm{tan}}=0.$$

This is exactly how relative boundary conditions for the Hodge Laplacian are defined [Taylor 2011]. We thus see that for differential forms the cancellations leading to the correct boundary conditions are caused by the fact that $D_{\partial N}$ clearly intertwines the projections onto the spaces of even and odd degree forms; compare to Definition 5.1.

References

[Airault 1976] H. Airault, "Perturbations singulières et solutions stochastiques de problèmes de D. Neumann-Spencer", *J. Math. Pures Appl.* (9) **55**:3 (1976), 233–267. MR Zbl

[Ballmann and Brüning 2001] W. Ballmann and J. Brüning, "On the spectral theory of manifolds with cusps", *J. Math. Pures Appl.* (9) **80**:6 (2001), 593–625. MR Zbl

[Bérard 1990] P. Bérard, "A note on Bochner type theorems for complete manifolds", *Manuscripta Math.* **69**:3 (1990), 261–266. MR Zbl

[Chavel 1984] I. Chavel, *Eigenvalues in Riemannian geometry*, Pure and Applied Mathematics **115**, Academic Press, Orlando, FL, 1984. MR Zbl

[Chen and Fraser 2010] J. Chen and A. Fraser, "On stable minimal disks in manifolds with nonnegative isotropic curvature", *J. Reine Angew. Math.* **643** (2010), 21–37. MR Zbl

[Donnelly and Li 1982] H. Donnelly and P. Li, "Lower bounds for the eigenvalues of Riemannian manifolds", *Michigan Math. J.* **29**:2 (1982), 149–161. MR Zbl

[Donnelly and Xavier 1984] H. Donnelly and F. Xavier, "On the differential form spectrum of negatively curved Riemannian manifolds", *Amer. J. Math.* **106**:1 (1984), 169–185. MR Zbl

[Eells and Elworthy 1971] J. Eells and K. D. Elworthy, "Wiener integration on certain manifolds", pp. 67–94 in *Problems in non-linear analysis* (Varenna, Italy, 1970), edited by G. Prodi, Centro Internazionale Matematico Estivo (CIME) **4**, Edizioni Cremonese, Rome, 1971. MR Zbl

[Eichhorn 2007] J. Eichhorn, *Global analysis on open manifolds*, Nova Science, New York, 2007. MR Zbl

[Elworthy 1988] D. Elworthy, "Geometric aspects of diffusions on manifolds", pp. 277–425 in *École d'Été de Probabilités de Saint-Flour XV-XVII* (Saint-Flour, 1985-1987), edited by P. L. Hennequin, Lecture Notes in Math. **1362**, Springer, Berlin, 1988. MR Zbl

[Elworthy and Li 1994] K. D. Elworthy and X.-M. Li, "Formulae for the derivatives of heat semigroups", *J. Funct. Anal.* **125**:1 (1994), 252–286. MR Zbl

[Elworthy and Rosenberg 1988] K. D. Elworthy and S. Rosenberg, "Generalized Bochner theorems and the spectrum of complete manifolds", *Acta Appl. Math.* **12**:1 (1988), 1–33. MR Zbl

[Elworthy and Rosenberg 1993] K. D. Elworthy and S. Rosenberg, "The Witten Laplacian on negatively curved simply connected manifolds", *Tokyo J. Math.* **16**:2 (1993), 513–524. MR Zbl

[Elworthy et al. 1998] K. D. Elworthy, X.-M. Li, and S. Rosenberg, "Bounded and L^2 harmonic forms on universal covers", *Geom. Funct. Anal.* **8**:2 (1998), 283–303. MR Zbl

[Fraser 2002] A. M. Fraser, "Minimal disks and two-convex hypersurfaces", *Amer. J. Math.* **124**:3 (2002), 483–493. MR Zbl

[Friedrich 2000] T. Friedrich, *Dirac operators in Riemannian geometry*, Graduate Studies in Mathematics **25**, American Mathematical Society, Providence, RI, 2000. MR Zbl

[Gaffney 1959] M. P. Gaffney, "The conservation property of the heat equation on Riemannian manifolds", *Comm. Pure Appl. Math.* **12** (1959), 1–11. MR Zbl

[Grigor'yan 1999] A. Grigor'yan, "Analytic and geometric background of recurrence and non-explosion of the Brownian motion on Riemannian manifolds", *Bull. Amer. Math. Soc.* (*N.S.*) **36**:2 (1999), 135–249. MR Zbl

[Gromov 1971] M. L. Gromov, "A topological technique for the construction of solutions of differential equations and inequalities", pp. 221–225 in *Actes du Congrès International des Mathématiciens, II* (Nice, 1970), Gauthier-Villars, Paris, 1971. MR Zbl

[Grubb 2003] G. Grubb, "Spectral boundary conditions for generalizations of Laplace and Dirac operators", *Comm. Math. Phys.* **240**:1-2 (2003), 243–280. MR Zbl

[Güneysu 2010] B. Güneysu, "The Feynman–Kac formula for Schrödinger operators on vector bundles over complete manifolds", *J. Geom. Phys.* **60**:12 (2010), 1997–2010. MR Zbl

[Hsu 1999] E. P. Hsu, "Analysis on path and loop spaces", pp. 277–347 in *Probability theory and applications* (Princeton, 1996), edited by E. P. Hsu and S. R. S. Varadhan, IAS/Park City Math. Ser. **6**, American Mathematical Society, Providence, RI, 1999. MR Zbl

[Hsu 2002a] E. P. Hsu, "Multiplicative functional for the heat equation on manifolds with boundary", *Michigan Math. J.* **50**:2 (2002), 351–367. MR Zbl

[Hsu 2002b] E. P. Hsu, *Stochastic analysis on manifolds*, Graduate Studies in Mathematics **38**, American Mathematical Society, Providence, RI, 2002. MR Zbl

[Ikeda and Watanabe 1989] N. Ikeda and S. Watanabe, *Stochastic differential equations and diffusion processes*, 2nd ed., North-Holland Mathematical Library **24**, North-Holland, Amsterdam, 1989. MR Zbl

[Lyons and Sullivan 1984] T. Lyons and D. Sullivan, "Function theory, random paths and covering spaces", *J. Differential Geom.* **19**:2 (1984), 299–323. MR Zbl

[Malliavin 1974] P. Malliavin, "Formules de la moyenne, calcul de perturbations et théorèmes d'annulation pour les formes harmoniques", *J. Functional Analysis* **17**:3 (1974), 274–291. MR Zbl

[McKean 1970] H. P. McKean, "An upper bound to the spectrum of Δ on a manifold of negative curvature", *J. Differential Geometry* **4** (1970), 359–366. MR Zbl

[Micallef and Wang 1993] M. J. Micallef and M. Y. Wang, "Metrics with nonnegative isotropic curvature", *Duke Math. J.* **72**:3 (1993), 649–672. MR Zbl

[Nakad and Roth 2013] R. Nakad and J. Roth, "The Spin^c Dirac operator on hypersurfaces and applications", *Differential Geom. Appl.* **31**:1 (2013), 93–103. MR Zbl

[Rosenberg 1997] S. Rosenberg, *The Laplacian on a Riemannian manifold*, London Mathematical Society Student Texts **31**, Cambridge Univ. Press, 1997. MR Zbl

[Schick 1996] T. Schick, *Analysis on ∂-manifolds of bounded geometry, Hodge–de Rham isomorphism and L^2-index theorem*, dissertation, Johannes Gutenberg-Universität Mainz, 1996, Available at http://tinyurl.com/schickprom. Zbl

[Schick 1998] T. Schick, "Analysis and geometry of boundary-manifolds of bounded geometry", preprint, 1998. arXiv

[Schick 2001] T. Schick, "Manifolds with boundary and of bounded geometry", *Math. Nachr.* **223** (2001), 103–120. MR Zbl

[Stroock 2000] D. W. Stroock, *An introduction to the analysis of paths on a Riemannian manifold*, Mathematical Surveys and Monographs **74**, American Mathematical Society, Providence, RI, 2000. MR Zbl

[Taylor 2011] M. E. Taylor, *Partial differential equations, I: Basic theory*, 2nd ed., Applied Mathematical Sciences **115**, Springer, New York, 2011. MR Zbl

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

Received June 19, 2016. Revised May 4, 2017.

LEVI LOPES DE LIMA
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DO CEARÁ (UFC)
FORTALEZA
BRAZIL
levi@mat.ufc.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 292, No. 1, 2018
dx.doi.org/10.2140/pjm.2018.292.203

ORE'S THEOREM ON CYCLIC SUBFACTOR PLANAR ALGEBRAS AND BEYOND

SEBASTIEN PALCOUX

Ore proved that a finite group is cyclic if and only if its subgroup lattice is distributive. Now, since every subgroup of a cyclic group is normal, we call a subfactor planar algebra cyclic if all its biprojections are normal and form a distributive lattice. The main result generalizes one side of Ore's theorem and shows that a cyclic subfactor is singly generated in the sense that there is a minimal 2-box projection generating the identity biprojection. We conjecture that this result holds without assuming the biprojections to be normal, and we show that this is true for small lattices. We finally exhibit a dual version of another theorem of Ore and a nontrivial upper bound for the minimal number of irreducible components for a faithful complex representation of a finite group.

1. Introduction

Vaughan Jones [1983] proved that the set of possible values for the index $|M:N|$ of a subfactor $(N \subseteq M)$ is

$$\left\{4cos^2\left(\frac{\pi}{n}\right) \mid n \geq 3\right\} \sqcup [4, \infty].$$

We observe that it is the disjoint union of a discrete series and a continuous series. Moreover, $|M:N| = |M:P|\cdot|P:N|$ for a given intermediate subfactor $N \subseteq P \subseteq M$, therefore by applying a kind of Eratosthenes sieve, we get that a subfactor with

MSC2010: primary 46L37; secondary 06D10, 16W30, 20C15, 20D30.
Keywords: von Neumann algebra, subfactor, planar algebra, biprojection, distributive lattice, finite group, representation.

an index in the discrete series or in the interval $(4, 8)$, except the countable set of numbers composed of numbers in the discrete series, can't have a nontrivial intermediate subfactor. A subfactor without nontrivial intermediate subfactor is called maximal [Bisch 1994]. For example, any subfactor of index in $(4, 3+\sqrt{5})$ is maximal; A_∞ excepted there are exactly 19 irreducible subfactor planar algebras for this interval (see [Jones et al. 2014; Afzaly et al. 2015]). The first example is the Haagerup subfactor [Peters 2010]. Thanks to Galois correspondence [Nakamura and Takeda 1960], a finite group subfactor, $(R^G \subseteq R)$ or $(R \subseteq R \rtimes G)$, is maximal if and only if it is a prime order cyclic group subfactor (i.e., $G = \mathbb{Z}/p$ with p prime). Thus we can say that the maximal subfactors are an extension of the prime numbers.

Question 1.1. *What could be the extension of the natural numbers*?

To answer this question, we need to find a natural class of subfactors, that we will call the "cyclic subfactors", satisfying the following properties:

(1) Every maximal subfactor is cyclic.

(2) A finite group subfactor is cyclic if and only if the group is cyclic.

An old and little-known theorem published in 1938 by the Norwegian mathematician Øystein Ore states that:

Theorem 1.2 [Ore 1938]. *A finite group G is cyclic if and only if its subgroup lattice $\mathcal{L}(G)$ is distributive.*

Firstly, the intermediate subfactor lattice of a maximal subfactor is obviously distributive. Next, by Galois correspondence, the intermediate subfactor lattice of a finite group subfactor is exactly the subgroup lattice (or its reversal) of the group; but distributivity is invariant under reversal, so (1) and (2) hold by Ore's theorem. Now an abelian group, and a fortiori a cyclic group, admits only normal subgroups; but T. Teruya [1998] generalized the notion of normal subgroup by the notion of normal intermediate subfactor, so:

Definition 1.3. A finite index irreducible subfactor is cyclic if all its intermediate subfactors are normal and form a distributive lattice.

Note that an irreducible finite index subfactor $(N \subseteq M)$ admits a finite lattice $\mathcal{L}(N \subseteq M)$ of intermediate subfactors by [Watatani 1996], as for the subgroup lattice of a finite group. Moreover, a finite group subfactor remembers the group by [Jones 1980]. Section 4A exhibits some examples of cyclic subfactors: of course the cyclic group subfactors and the (irreducible finite index) maximal subfactors; moreover, up to equivalence, exactly 23279 among 34503 inclusions of groups of index < 30, give a cyclic subfactor. The class of cyclic subfactors is stable under dual, intermediate, free composition and certain tensor products. Now the natural problem concerning cyclic subfactors is to understand in what sense they are "singly generated". To answer this question, we extend the following theorem of Ore.

Theorem 1.4 [Ore 1938]. *If an interval of finite groups* $[H, G]$ *is distributive, then there exists* $g \in G$ *such that* $\langle H, g\rangle = G$.

Theorem 1.5. *An irreducible subfactor planar algebra whose biprojections are central and form a distributive lattice, has a minimal* 2*-box projection generating the identity biprojection* (*i.e., w-cyclic subfactor*).

But "normal" means "bicentral", so a cyclic subfactor planar algebra is w-cyclic. The converse is false, a group subfactor ($R^G \subseteq R$) is cyclic if and only if G is cyclic, and is w-cyclic if and only if G is linearly primitive (consider $G = S_3$). That's why we have chosen the name w-cyclic (i.e., weakly cyclic). We conjecture that Theorem 1.5 holds without the assumption that the biprojections are central.

Conjecture 1.6. *An irreducible subfactor planar algebra with a distributive biprojection lattice is w-cyclic.*

This is true if the lattice has less than 32 elements (and so, at index < 32). Now the group-theoretic reformulation of Conjecture 1.6 for the planar algebra $\mathcal{P}(R^G \subseteq R^H)$, gives a dual version of Theorem 1.4.

Conjecture 1.7. *If the interval of finite groups* $[H, G]$ *is distributive then there exists an irreducible complex representation* V *of* G *such that* $G_{(V^H)} = H$.

In general, we deduce a nontrivial upper bound for the minimal number of minimal central projections generating the identity biprojection. For $\mathcal{P}(R^G \subseteq R)$, this gives a nontrivial upper bound for the minimal number of irreducible components for a faithful complex representation of G. This is a bridge linking combinatorics and representations in the theory of finite groups. This paper is a short version of [Palcoux 2015].

2. Ore's theorem on finite groups

2A. ***Basics in lattice theory.*** A lattice $(L, \wedge, \vee)$ is a poset L in which every two elements a, b have a unique supremum (or join) $a \vee b$, and a unique infimum (or meet) $a \wedge b$. Let G be a finite group. The set of subgroups $K \subseteq G$ forms a lattice, denoted by $\mathcal{L}(G)$, ordered by $\subseteq$, with $K_1 \vee K_2 = \langle K_1, K_2\rangle$ and $K_1 \wedge K_2 = K_1 \cap K_2$. A sublattice of $(L, \wedge, \vee)$ is a subset $L' \subseteq L$ such that $(L', \wedge, \vee)$ is also a lattice. Consider $a, b \in L$ with $a \le b$, then the interval $[a, b]$ is the sublattice $\{c \in L \mid a \le c \le b\}$. Any finite lattice admits a minimum and a maximum, denoted by $\hat{0}$ and $\hat{1}$. An atom is a minimal element in $L \setminus \{\hat{0}\}$ and a coatom is a maximal element in $L \setminus \{\hat{1}\}$. The top interval of a finite lattice L is the interval $[t, \hat{1}]$, with t the meet of all the coatoms. The height of a finite lattice L is the greatest length of a (strict) chain. A lattice is distributive if the join and meet operations distribute over each other.

Remark 2.1. Distributivity is stable under taking a sublattice, reversal, direct product and concatenation.

A distributive lattice is called boolean if any element b admits a unique complement $b^{\complement}$ (i.e., $b \wedge b^{\complement} = \hat{0}$ and $b \vee b^{\complement} = \hat{1}$). The subset lattice of $\{1, 2, \ldots, n\}$, with union and intersection, is called the boolean lattice $\mathcal{B}_n$ of rank n. Any finite boolean lattice is isomorphic to some $\mathcal{B}_n$.

Lemma 2.2. *The top interval of a finite distributive lattice is boolean.*

Proof. See [Stanley 2012, items a–i, pages 254–255] which use Birkhoff's representation theorem, which states a finite lattice is distributive if and only if it embeds into some $\mathcal{B}_n$. □

A lattice with a boolean top interval will be called *top boolean* (and its reversal, *bottom boolean*). See [Stanley 2012] for more details on lattice basics.

2B. *Ore's theorem on distributive intervals of finite groups.* Øystein Ore [1938, Theorem 4, page 267] proved the following result.

Theorem 2.3. *A finite group G is cyclic if and only if its subgroup lattice $\mathcal{L}(G)$ is distributive.*

Proof. ($\Leftarrow$): This is just a particular case of Theorem 2.5 with $H = \{e\}$.

($\Rightarrow$): A finite cyclic group $G = \mathbb{Z}/n$ has exactly one subgroup of order d, denoted by $\mathbb{Z}/d$, for every divisor d of n. Now $\mathbb{Z}/d_1 \vee \mathbb{Z}/d_2 = \mathbb{Z}/\mathrm{lcm}(d_1, d_2)$ and $\mathbb{Z}/d_1 \wedge \mathbb{Z}/d_2 = \mathbb{Z}/\mathrm{gcd}(d_1, d_2)$, but the lcm and gcd distribute other each over, so the result follows. □

Definition 2.4. An interval of finite groups $[H, G]$ is said to be H-cyclic if there is $g \in G$ such that $\langle H, g\rangle = G$. Note that $\langle H, g\rangle = \langle Hg\rangle$.

Ore extended one side of Theorem 2.3 to the interval of finite groups [Ore 1938, Theorem 7] for which we will give our own proof (which is a group-theoretic reformulation of the proof of Theorem 4.26):

Theorem 2.5. *A distributive interval $[H, G]$ is H-cyclic.*

Proof. The proof follows from the claims below and Lemma 2.2.

<u>Claim:</u> Let M be a maximal subgroup of G. Then $[M, G]$ is M-cyclic.

Proof of claim. For $g \in G$ with $g \notin M$, we have $\langle M, g\rangle = G$ by maximality. □

<u>Claim:</u> A boolean interval $[H, G]$ is H-cyclic.

Proof of claim. Let M be a coatom in $[H, G]$, and $M^{\complement}$ be its complement. By the previous claim and induction on the height of the lattice, we can assume $[H, M]$ and $[H, M^{\complement}]$ both to be H-cyclic, i.e., there are $a, b \in G$ such that $\langle H, a\rangle = M$ and $\langle H, b\rangle = M^{\complement}$. For $g = ab$, $\langle H, a, g\rangle = \langle H, g, b\rangle = \langle H, a, b\rangle = M \vee M^{\complement} = G$, since $a = gb^{-1}$ and $b = a^{-1}g$. Now, $\langle H, g\rangle = \langle H, g\rangle \vee H = \langle H, g\rangle \vee (M \wedge M^{\complement})$ but by distributivity $\langle H, g\rangle \vee (M \wedge M^{\complement}) = (\langle H, g\rangle \vee M) \wedge (\langle H, g\rangle \vee M^{\complement})$. So $\langle H, g\rangle = \langle H, a, g\rangle \wedge \langle H, g, b\rangle = G$. The result follows. □

Claim: $[H, G]$ is H-cyclic if its top interval $[K, G]$ is K-cyclic.

Proof of claim. Consider $g \in G$ with $\langle K, g\rangle = G$. For any coatom $M \in [H, G]$, we have $K \subseteq M$ by definition, and so $g \notin M$, then a fortiori $\langle H, g\rangle \not\subseteq M$. It follows that $\langle H, g\rangle = G$. □

3. Subfactor planar algebras and biprojections

For the notions of subfactor, subfactor planar algebra and basic properties, we refer to [Jones and Sunder 1997; Jones 1999; Kodiyalam and Sunder 2004]. See also [Palcoux 2015, Section 3] for a short introduction. A subfactor planar algebra is of finite index by definition.

3A. ***Basics on the 2-box space.*** Let $(N \subseteq M)$ be a finite index irreducible subfactor. The n-box spaces $\mathcal{P}_{n,+}$ and $\mathcal{P}_{n,-}$ of the planar algebra $\mathcal{P} = \mathcal{P}(N \subseteq M)$, are $N' \cap M_{n-1}$ and $M' \cap M_n$. Let $R(a)$ be the range projection of $a \in \mathcal{P}_{2,+}$. We define the relations $a \preceq b$ by $R(a) \leq R(b)$, and $a \sim b$ by $R(a) = R(b)$. Let $e_1 := e_N^M$ and $\mathrm{id} := e_M^M$ be the Jones and the identity projections in $\mathcal{P}_{2,+}$. Note that $\mathrm{tr}(e_1) = |M : N|^{-1} = \delta^{-2}$ and $\mathrm{tr}(\mathrm{id}) = 1$. Let $\mathcal{F} : \mathcal{P}_{2,\pm} \to \mathcal{P}_{2,\mp}$ be the Fourier transform (90° rotation), $\bar{a} := \mathcal{F}(\mathcal{F}(a))$ be the contragredient of $a \in \mathcal{P}_{2,\pm}$, and $a * b = \mathcal{F}(\mathcal{F}^{-1}(a) \cdot \mathcal{F}^{-1}(b))$ be the coproduct of $a, b \in \mathcal{P}_{2,\pm}$.

Lemma 3.1. *Let a, b, c, d be positive operators of $\mathcal{P}_{2,+}$. Then*

(1) *$a * b$ is also positive,*

(2) *$[a \preceq b$ and $c \preceq d] \Rightarrow a * c \preceq b * d$,*

(3) *$a \preceq b \Rightarrow \langle a\rangle \leq \langle b\rangle$,*

(4) *$a \sim b \Rightarrow \langle a\rangle = \langle b\rangle$.*

Proof. This is precisely [Liu 2016, Theorem 4.1 and Lemma 4.8] for (1) and (2). Next, if $a \preceq b$, then by (2), for any integer k, $a^{*k} \preceq b^{*k}$, and hence for all n,

$$\sum_{k=1}^{n} a^{*k} \preceq \sum_{k=1}^{n} b^{*k},$$

so $\langle a\rangle \leq \langle b\rangle$ by Definition 3.8. Finally, (4) is immediate from (3). □

The next lemma follows by irreducibility (i.e., $\mathcal{P}_{1,+} = \mathbb{C}$).

Lemma 3.2. *Let $p, q \in \mathcal{P}_{2,+}$ be projections. Then*

$$e_1 \preceq p * \bar{q} \Leftrightarrow pq \neq 0.$$

Note that if $p \in \mathcal{P}_{2,+}$ is a projection then $\bar{p}$ is also a projection.

Lemma 3.3. *Let $a, b, c \in \mathcal{P}_{2,+}$ be projections with $c \preceq a * b$. Then there exist $a' \preceq c * \bar{b}$ and $b' \preceq \bar{a} * c$ such that a', b' are projections and $aa', bb' \neq 0$.*

Proof. The proof follows from Lemmas 3.1 and 3.2, and

$$e_1 \preceq c * \bar{c} \preceq (a * b) * \bar{c} = a * (b * \bar{c}).$$

We can also apply [Liu 2016, Lemma 4.10]. □

3B. ***On the biprojections.***

Definition 3.4 [Liu 2016, Definition 2.14]. A biprojection is a projection $b \in \mathcal{P}_{2,\pm}$ with $\mathcal{F}(b)$ a multiple of a projection.

Note that $e_1 = e_N^M$ and $\mathrm{id} = e_M^M$ are biprojections.

Theorem 3.5 [Bisch 1994, page 212]. *A projection b is a biprojection if and only if it is the Jones projection e_K^M of an intermediate subfactor $N \subseteq K \subseteq M$.*

Therefore the set of biprojections is a lattice of the form $[e_1, \mathrm{id}]$.

Theorem 3.6. *An operator b is a biprojection if and only if*

$$e_1 \leq b = b^2 = b^\star = \bar{b} = \lambda b * b, \text{ with } \lambda^{-1} = \delta \,\mathrm{tr}(b).$$

Proof. See [Landau 2002, items 0–3, page 191] and [Liu 2016, Theorem 4.12]. □

Lemma 3.7. *Consider $a_1, a_2, b \in \mathcal{P}_{2,+}$ with b a biprojection. Then*

$$(b \cdot a_1 \cdot b) * (b \cdot a_2 \cdot b) = b \cdot (a_1 * (b \cdot a_2 \cdot b)) \cdot b = b \cdot ((b \cdot a_1 \cdot b) * a_2) \cdot b,$$
$$(b * a_1 * b) \cdot (b * a_2 * b) = b * (a_1 \cdot (b * a_2 * b)) * b = b * ((b * a_1 * b) \cdot a_2) * b.$$

Proof. By exchange relations [Landau 2002] for b and $\mathcal{F}(b)$. □

Definition 3.8. Consider $a \in \mathcal{P}_{2,+}$ positive, and let p_n be the range projection of $\sum_{k=1}^n a^{*k}$. By finiteness, there exists N such that for all $m \geq N$, $p_m = p_N$, which is a biprojection [Liu 2016, Lemma 4.14], denoted $\langle a \rangle$, called the biprojection generated by a. It is the smallest biprojection $b \succeq a$. For S a finite set of positive operators, let $\langle S \rangle$ be the biprojection $\left\langle \sum_{s \in S} s \right\rangle$, it is the smallest biprojection b such that $b \succeq s$, for all $s \in S$.

3C. ***Intermediate planar algebras and 2-box spaces.*** Let $N \subseteq K \subseteq M$ be an intermediate subfactor. The planar algebras $\mathcal{P}(N \subseteq K)$ and $\mathcal{P}(K \subseteq M)$ can be derived from $\mathcal{P}(N \subseteq M)$, see [Bakshi 2016; Landau 1998].

Theorem 3.9. *Consider the intermediate subfactors*

$$N \subseteq P \subseteq K \subseteq Q \subseteq M.$$

Then there are two isomorphisms of von Neumann algebras

$$l_K : \mathcal{P}_{2,+}(N \subseteq K) \to e_K^M \mathcal{P}_{2,+}(N \subseteq M) e_K^M,$$
$$r_K : \mathcal{P}_{2,+}(K \subseteq M) \to e_K^M * \mathcal{P}_{2,+}(N \subseteq M) * e_K^M,$$

for the usual $+$, $\times$ *and* $()^\star$, *such that*

$$l_K(e_P^K) = e_P^M \quad \textit{and} \quad r_K(e_Q^M) = e_Q^M.$$

Moreover, the coproduct $*$ *is also preserved by these maps, but up to a multiplicative constant,* $|M:K|^{1/2}$ *for* l_K *and* $|K:N|^{-1/2}$ *for* r_K. *Then, for all* $m \in \{l_K^{\pm 1}, r_K^{\pm 1}\}$, *and for all* $a_i > 0$ *in the domain of* m, $m(a_i) > 0$ *and*

$$\langle m(a_1), \dots, m(a_n) \rangle = m(\langle a_1, \dots, a_n \rangle).$$

Proof. This is clear from [Bakshi 2016] or [Landau 1998], using Lemma 3.7. □

Notations 3.10. Let $b_1 \le b \le b_2$ be the biprojections $e_P^M \le e_K^M \le e_Q^M$. We define $l_b := l_K$ and $r_b := r_K$; also $\mathcal{P}(b_1, b_2) := \mathcal{P}(P \subseteq Q)$ and

$$|b_2 : b_1| := \operatorname{tr}(b_2)/\operatorname{tr}(b_1) = |Q:P|.$$

4. Ore's theorem on subfactor planar algebras

4A. ***The cyclic subfactor planar algebras.*** In this subsection, we define the class of cyclic subfactor planar algebras, we show that it contains plenty of examples, and we prove that it is stable under dual, intermediate, free composition and certain tensor products. Let $\mathcal{P}$ be an irreducible subfactor planar algebra.

Definition 4.1 [Teruya 1998]. A biprojection b is normal if it is bicentral (that is, if b and $\mathcal{F}(b)$ are central).

Definition 4.2. An irreducible subfactor planar algebra is said to be

- distributive if its biprojection lattice is distributive,
- Dedekind if all its biprojections are normal,
- cyclic if it is both Dedekind and distributive.

Moreover, we call a subfactor cyclic if its planar algebra is cyclic.

Examples 4.3. A group subfactor is cyclic if and only if the group is cyclic; every maximal subfactor is cyclic, in particular every 2-supertransitive subfactor, as the Haagerup subfactor [Asaeda and Haagerup 1999; Izumi 2001; Peters 2010], is cyclic. Up to equivalence, exactly 23279 among 34503 inclusions of groups of index < 30, give a cyclic subfactor (more than 65%).

Definition 4.4. Let G be a finite group and H a subgroup. The core H_G is the largest normal subgroup of G contained in H. The subgroup H is called core-free if $H_G = \{1\}$; in this case the interval $[H, G]$ is also called core-free. Two intervals of finite groups $[A, B]$ and $[C, D]$ are called equivalent if there is a group isomorphism $\phi : B/A_B \to D/C_D$ such that $\phi(A/A_B) = C/C_D$.

Remark 4.5. A finite group subfactor remembers the group [Jones 1980], but a finite group-subgroup subfactor does not remember the equivalence class of the interval in general. A counterexample was found by V. S. Sunder and V. Kodiyalam [2000], the intervals $[\langle(1234)\rangle, S_4]$ and $[\langle(12)(34)\rangle, S_4]$ are not equivalent whereas their corresponding subfactors are isomorphic; but thanks to the complete characterization by M. Izumi [2002], it remembers the interval in the maximal case, because the intersection of a core-free maximal subgroup with an abelian normal subgroup is trivial.

Theorem 4.6. *The free composition of irreducible finite index subfactors has no extra intermediate.*

Proof. See [Liu 2016, Theorem 2.22]. □

Corollary 4.7. *The class of finite index irreducible cyclic subfactors is stable under free composition.*

Proof. By Theorem 4.6, the intermediate subfactor lattice of a free composition is the concatenation of the lattice of the two components (see also Remark 2.1). By Theorem 3.9 and Lemma 3.7, the biprojections remain normal. □

The following theorem was proved in the 2-supertransitive case by Y. Watatani [1996, Proposition 5.1]. The general case was conjectured by the author, but specified and proved after a discussion with F. Xu.

Theorem 4.8. *Let* $(N_i \subset M_i)$, $i = 1, 2$, *be irreducible finite index subfactors. Then*

$$\mathcal{L}(N_1 \subset M_1) \times \mathcal{L}(N_2 \subset M_2) \subsetneq \mathcal{L}(N_1 \otimes N_2 \subset M_1 \otimes M_2)$$

if and only if there are intermediate subfactors $N_i \subseteq P_i \subset Q_i \subseteq M_i$, $i = 1, 2$, *such that* $(P_i \subset Q_i)$ *is of depth* 2 *and isomorphic to* $(R^{\mathbb{A}_i} \subset R)$, *with* $\mathbb{A}_2 \simeq \mathbb{A}_1^{\mathrm{cop}}$ *being the* (*very simple*) *Kac algebra* $\mathbb{A}_1$ *with the opposite coproduct.*

Proof. Consider the intermediate subfactors

$$N_1 \otimes N_2 \subseteq P_1 \otimes P_2 \subset R \subset Q_1 \otimes Q_2 \subseteq M_1 \otimes M_2$$

with R not of tensor product form, $P_1 \otimes P_2$ and $Q_1 \otimes Q_2$ the closest (below and above, respectively) to R among those of tensor product form. Now using [Xu 2013, Proposition 3.5(2)], $(P_i \subseteq Q_i)$, $i = 1, 2$, are of depth 2, their corresponding Kac algebras, $\mathbb{A}_i$, $i = 1, 2$, are very simple and $\mathbb{A}_2 \simeq \mathbb{A}_1^{\mathrm{cop}}$ [Xu 2013, Definition 3.6 and Proposition 3.10]. The converse is given by [Xu 2013, Theorem 3.14]. □

Remark 4.9. By Theorem 4.8 and Remark 2.1, the class of (finite index irreducible) cyclic subfactors is stable under certain tensor products (i.e., if there is no cop-isomorphic intermediate of depth 2), and by Theorem 3.9 and Lemma 3.7, the biprojections remain normal.

Lemma 4.10. *If a subfactor is cyclic then the intermediate and dual subfactors are also cyclic.*

Proof. The proof follows from Remark 2.1, Theorem 3.9 and Lemma 3.7. □

A subfactor as $(R \subseteq R \rtimes G)$ or $(R^G \subseteq R)$ is called a “group subfactor”. Then, the following lemma justifies the choice of the word “cyclic”.

Lemma 4.11. *A cyclic “group subfactor” is a “cyclic group” subfactor.*

Proof. By Galois correspondence, if a “group subfactor” is cyclic then the subgroup lattice is distributive, and so the group is cyclic by Ore’s Theorem 2.3. The normal biprojections of a group subfactor corresponds to the normal subgroups [Teruya 1998], but every subgroup of a cyclic group is normal. □

Problem 4.12. Is a depth 2 irreducible finite index cyclic subfactor, a cyclic group subfactor?

The answer could be “no” because the following fusion ring (first reported in [Palcoux 2013]), the first known to be simple integral and nontrivial, *could be* the Grothendieck ring of a “maximal” Kac algebra of dimension 210 and type $(1, 5, 5, 5, 6, 7, 7)$.

$$\begin{matrix}1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&1&0&0\\0&0&0&0&0&1&0\\0&0&0&0&0&0&1\end{matrix}\quad
\begin{matrix}0&1&0&0&0&0&0\\1&1&0&1&0&1&1\\0&0&1&0&1&1&1\\0&1&0&0&1&1&1\\0&0&1&1&1&1&1\\0&1&1&1&1&1&1\\0&1&1&1&1&1&1\end{matrix}\quad
\begin{matrix}0&0&1&0&0&0&0\\0&0&1&0&1&1&1\\1&1&1&0&0&1&1\\0&0&0&1&1&1&1\\0&1&0&1&1&1&1\\0&1&1&1&1&1&1\\0&1&1&1&1&1&1\end{matrix}\quad
\begin{matrix}0&0&0&1&0&0&0\\0&1&0&0&1&1&1\\0&0&0&1&1&1&1\\1&0&1&1&0&1&1\\0&1&1&0&1&1&1\\0&1&1&1&1&1&1\\0&1&1&1&1&1&1\end{matrix}\quad
\begin{matrix}0&0&0&0&1&0&0\\0&0&1&1&1&1&1\\0&1&0&1&1&1&1\\0&1&1&0&1&1&1\\1&1&1&1&1&1&1\\0&1&1&1&1&2&1\\0&1&1&1&1&1&2\end{matrix}\quad
\begin{matrix}0&0&0&0&0&1&0\\0&1&1&1&1&1&1\\0&1&1&1&1&1&1\\0&1&1&1&1&1&1\\0&1&1&1&1&2&1\\1&1&1&1&2&1&2\\0&1&1&1&1&2&2\end{matrix}\quad
\begin{matrix}0&0&0&0&0&0&1\\0&1&1&1&1&1&1\\0&1&1&1&1&1&1\\0&1&1&1&1&1&1\\0&1&1&1&1&1&2\\0&1&1&1&1&2&2\\1&1&1&1&2&2&1\end{matrix}$$

4B. ***The w-cyclic subfactor planar algebras.*** Let $\mathcal{P}$ be an irreducible subfactor planar algebra.

Theorem 4.13. *Let $p \in \mathcal{P}_{2,+}$ be a minimal central projection. Then there exists a minimal projection $v \le p$ such that $\langle v \rangle = \langle p \rangle$.*

Proof. If p is a minimal projection, then the theorem clearly holds. Else, let $b_1, \dots, b_n$ be the coatoms of $[e_1, \langle p \rangle]$ (n is finite by [Watatani 1996]). If $p \not\preceq \sum_{i=1}^n b_i$, then there exists $u \le p$, a minimal projection such that $u \not\preceq b_i$ for all i, so that $\langle u \rangle = \langle p \rangle$. If not, $p \preceq \sum_{i=1}^n b_i$ (with $n > 1$, otherwise $p \le b_1$ and $\langle p \rangle \le b_1$, a contradiction). Let $E_i = \mathrm{range}(b_i)$ and $F = \mathrm{range}(p)$, then $F = \sum_i E_i \cap F$ (because p is a minimal central projection) with $1 < n < \infty$ and $E_i \cap F \subsetneq F$ for all i (otherwise there exists i with $p \le b_i$, a contradiction), so $\dim(E_i \cap F) < \dim(F)$ and there exists $U \subseteq F$, a one-dimensional subspace such that $U \not\subseteq E_i \cap F$ for all i, and so a fortiori $U \not\subseteq E_i$ for all i. It follows that $u = p_U \le p$ is a minimal projection such that $\langle u \rangle = \langle p \rangle$. □

Thanks to Theorem 4.13, we can give the following definition:

Definition 4.14. The planar algebra $\mathcal{P}$ is weakly cyclic (or w-cyclic) if it satisfies one of the following equivalent assertions:

- There exists a minimal projection $u \in \mathcal{P}_{2,+}$ such that $\langle u \rangle = \mathrm{id}$.
- There exists a minimal central projection $p \in \mathcal{P}_{2,+}$ such that $\langle p \rangle = \mathrm{id}$.

We call a subfactor w-cyclic if its planar algebra is w-cyclic.

The following remark justifies the choice of the word "w-cyclic".

Remark 4.15. By Corollary 6.12, a finite group subfactor $(R^G \subset R)$ is w-cyclic if and only if G is linearly primitive, which is strictly weaker than cyclic (see for example S_3), nevertheless the notion of w-cyclic is a singly generated notion in the sense that "there is a minimal projection generating the identity biprojection". We can also see the weakness of this assumption by the fact that the minimal projection does not necessarily generate a basis for the set of positive operators, but just the support of it, i.e., the identity.

Question 4.16. *Is a cyclic subfactor planar algebra w-cyclic?*
The answer is "yes" by Theorem 4.27.

Let $\mathcal{P} = \mathcal{P}(N \subseteq M)$ be an irreducible subfactor planar algebra. Take an intermediate subfactor $N \subseteq K \subseteq M$ and its biprojection $b = e^M_K$.

Lemma 4.17. *Let $\mathcal{A}$ be a $\star$-subalgebra of $\mathcal{P}_{2,+}$. Then any element $x \in \mathcal{A}$ is positive in $\mathcal{A}$ if and only if it is positive in $\mathcal{P}_{2,+}$.*

Proof. If x is positive in $\mathcal{A}$, it is of the form $aa^\star$, with $a \in \mathcal{A}$, but $a \in \mathcal{P}_{2,+}$ also, so x is positive in $\mathcal{P}_{2,+}$. Conversely, if x is positive in $\mathcal{P}_{2,+}$ then

$$\langle xy | y \rangle = \mathrm{tr}(y^\star x y) \geq 0,$$

for any $y \in \mathcal{P}_{2,+}$, so in particular, for any $y \in \mathcal{A}$, which means x is positive in $\mathcal{A}$. $\square$

Note that Lemma 4.17 will be applied to $\mathcal{A} = b\mathcal{P}_{2,+}b$ or $b * \mathcal{P}_{2,+} * b$.

Proposition 4.18. *The planar algebra $\mathcal{P}(e_1, b)$ is w-cyclic if and only if there is a minimal projection $u \in \mathcal{P}_{2,+}$ such that $\langle u \rangle = b$.*

Proof. The planar algebra $\mathcal{P}(N \subseteq K)$ is w-cyclic if and only if there is a minimal projection $x \in \mathcal{P}_{2,+}(N \subseteq K)$ such that $\langle x \rangle = e^K_K$, if and only if $l_K(\langle x \rangle) = l_K(e^K_K)$, if and only if $\langle u \rangle = e^M_K$ (by Theorem 3.9), with $u = l_K(x)$ a minimal projection in $e^M_K \mathcal{P}_{2,+} e^M_K$ and in $\mathcal{P}_{2,+}$. $\square$

Lemma 4.19. *For any minimal projection $x \in \mathcal{P}_{2,+}(b, \mathrm{id})$, $r_b(x)$ is positive and for any minimal projection $v \preceq r_b(x)$, there is $\lambda > 0$ such that $b * v * b = \lambda r_b(x)$.*

Proof. Firstly, x is positive, so by Theorem 3.9, $r_b(x)$ is also positive. For any minimal projection $v \preceq r_b(x)$, we have $b * v * b \preceq r_b(x)$, because

$$b * v * b \preceq b * r_b(x) * b = b * b * u * b * b \sim b * u * b = r_b(x),$$

by Lemma 3.1(2) and with $u \in \mathcal{P}_{2,+}$. Now by Lemma 3.1(1), $b * v * b > 0$, so $r_b^{-1}(b * v * b) > 0$ also, and by Theorem 3.9,

$$r_b^{-1}(b * v * b) \preceq x.$$

But x is a minimal projection, so by positivity, there exists $\lambda > 0$ such that

$$r_b^{-1}(b * v * b) = \lambda x.$$

It follows that $b * v * b = \lambda r_b(x)$. □

Lemma 4.20. *Consider $v \in \mathcal{P}_{2,+}$ positive. Then $\langle b * v * b\rangle = \langle b, v\rangle$.*

Proof. Firstly, by Definition 3.8, $b * v * b \preceq \langle b, v\rangle$, so $\langle b * v * b\rangle \le \langle b, v\rangle$, by Lemma 3.1(3). Next $e_1 \le b$ and $x * e_1 = e_1 * x = \delta^{-1}x$, so

$$v = \delta^2 e_1 * v * e_1 \preceq b * v * b.$$

Moreover by Theorem 3.6, $\bar{v} \preceq \langle b * v * b\rangle$, but by Lemma 3.2,

$$\bar{v} * b * v * b \succeq \bar{v} * e_1 * v * b \sim \bar{v} * v * b \succeq e_1 * b \sim b.$$

Then $b, v \le \langle b * v * b\rangle$, so we also have $\langle b, v\rangle \le \langle b * v * b\rangle$. □

Proposition 4.21. *The planar algebra $\mathcal{P}(b, \mathrm{id})$ is w-cyclic if and only if there is a minimal projection $v \in \mathcal{P}_{2,+}$ such that $\langle b, v\rangle = \mathrm{id}$ and $r_b^{-1}(b * v * b)$ is a positive multiple of a minimal projection.*

Proof. The planar algebra $\mathcal{P}(K \subseteq M)$ is w-cyclic if and only if there is a minimal projection $x \in \mathcal{P}_{2,+}(K \subseteq M)$ such that $\langle x\rangle = e_M^M$, if and only if $r_K(\langle x\rangle) = r_K(e_M^M)$, if and only if $\langle r_K(x)\rangle = e_M^M$ by Theorem 3.9. The result follows by Lemmas 4.19 and 4.20. □

4C. *The main result.*

Let $\mathcal{P}$ be an irreducible subfactor planar algebra.

Lemma 4.22. *A maximal subfactor planar algebra is w-cyclic.*

Proof. By maximality $\langle u\rangle = \mathrm{id}$ for any minimal projection $u \neq e_1$. □

Definition 4.23. The top intermediate subfactor planar algebra is the intermediate associated to the top interval of the biprojection lattice.

Lemma 4.24. *An irreducible subfactor planar algebra is w-cyclic if its top intermediate is so.*

Proof. Let $b_1, \dots, b_n$ be the coatoms in $[e_1, \mathrm{id}]$ and $t = \bigwedge_{i=1}^n b_i$. By assumption and Proposition 4.21, there is a minimal projection $v \in \mathcal{P}_{2,+}$ with $\langle t, v\rangle = \mathrm{id}$. If there exists i such that $v \le b_i$, then $\langle t, v\rangle \le b_i$, a contradiction. So $v \not\le b_i$ for all i, and then $\langle v\rangle = \mathrm{id}$. □

Definition 4.25. Let $h(\mathcal{P})$ be the height of the biprojection lattice $[e_1, \mathrm{id}]$. Note that $h(\mathcal{P}) < \infty$ because the index is finite.

Theorem 4.26. *If the biprojections in $\mathcal{P}_{2,+}$ are central and form a distributive lattice, then $\mathcal{P}$ is w-cyclic.*

Proof. By Lemma 4.10, we can make an induction on $h(\mathcal{P})$. If $h(\mathcal{P}) = 1$, then we apply Lemma 4.22. Now suppose that the theorem holds for $h(\mathcal{P}) < n$, we will prove it for $h(\mathcal{P}) = n \geq 2$. By Lemmas 2.2 and 4.24, we can assume the biprojection lattice to be boolean. For b in the open interval (e_1, id), its complement $b^{\complement}$ (see Section 2A) is also in (e_1, id). By induction and Proposition 4.18, there are minimal projections u, v such that $b = \langle u \rangle$ and $b^{\complement} = \langle v \rangle$. Take any minimal projection $c \preceq u * v$, then

$$\langle c \rangle = \langle c \rangle \vee e_1 = \langle c \rangle \vee (b \wedge b^{\complement}) = \langle c \rangle \vee (\langle u \rangle \wedge \langle v \rangle),$$

so *by distributivity*

$$\langle c \rangle = (\langle c \rangle \vee \langle u \rangle) \wedge (\langle c \rangle \vee \langle v \rangle) = \langle c, u \rangle \wedge \langle c, v \rangle.$$

Then by Lemma 3.3, $\langle c \rangle = \langle u', c, v \rangle \wedge \langle u, c, v' \rangle$ with u', v' minimal projections and uu', $vv' \neq 0$, so in particular the central support $Z(u') = Z(u)$ and $Z(v') = Z(v)$. Now by assumption, every biprojection is central, so $u \leq Z(u') \leq \langle u', c, v \rangle$ and $v \leq Z(v') \leq \langle u, c, v' \rangle$, so $\langle c \rangle = \mathrm{id}$. □

Theorem 4.27. *A cyclic subfactor planar algebra is w-cyclic.*

Proof. This is immediate by Theorem 4.26 because a normal biprojection is by definition bicentral, so a fortiori central. □

5. Extension for small distributive lattices

We extend Theorem 4.26 without assuming the biprojections to be central, but for distributive lattices with less than 32 elements. Because the top lattice of a distributive lattice is boolean (Lemma 2.2), we can reduce the proof to $\mathcal{B}_n$ with $n < 5$.

Definition 5.1. An irreducible subfactor planar algebra is said to be boolean (or $\mathcal{B}_n$) if its biprojection lattice is boolean (of rank n).

Proposition 5.2. *An irreducible subfactor planar algebra such that the coatoms $b_1, \dots, b_n \in [e_1, \mathrm{id}]$ satisfy $\sum_i \frac{1}{|\mathrm{id} : b_i|} \leq 1$, is w-cyclic.*

Proof. Firstly, by Lemmas 4.22 and 4.24, we can assume that $n > 1$. By definition, $|\mathrm{id} : b_i| = \mathrm{tr}(\mathrm{id})/\mathrm{tr}(b_i)$ so by assumption $\sum_i \mathrm{tr}(b_i) \leq \mathrm{tr}(\mathrm{id})$. If $\sum_i b_i \sim \mathrm{id}$ then $\sum_i b_i \geq \mathrm{id}$, but $\sum_i \mathrm{tr}(b_i) \leq \mathrm{tr}(\mathrm{id})$ so $\sum_i b_i = \mathrm{id}$. Now $e_1 \leq b_i$ for all i, therefore $ne_1 \leq \sum_i b_i = \mathrm{id}$, contradiction with $n > 1$. So $\sum_i b_i \prec id$, which implies the existence of a minimal projection $u \not\leq b_i$ for all i, which means that $\langle u \rangle = \mathrm{id}$. □

Remark 5.3. The converse is false, $(R \subset R \rtimes \mathbb{Z}/30)$ is a counterexample, because $\frac{1}{2} + \frac{1}{3} + \frac{1}{5} = \frac{31}{30} > 1$.

Corollary 5.4. *An irreducible subfactor planar algebra with at most two coatoms in* $[e_1, \mathrm{id}]$ *is w-cyclic.*

Proof. We have $\sum_i \frac{1}{|\mathrm{id} : b_i|} \le \frac{1}{2} + \frac{1}{2}$, and the result follows by Proposition 5.2. $\square$

Examples 5.5. Every $\mathcal{B}_2$ subfactor planar algebra is w-cyclic.

$$\begin{tikzcd}[arrows=dash] & \mathrm{id} & \\ b_1 \arrow[ur] & & b_2 \arrow[ul] \\ & e_1 \arrow[ul] \arrow[ur] & \end{tikzcd}$$

Lemma 5.6. *Let* $u, v \in \mathcal{P}_{2,+}$ *be minimal projections. If* $v \not\preceq \langle u \rangle$ *then there exist minimal projections* $c \preceq u * v$ *and* $w \preceq \bar{u} * c$ *such that* $w \not\preceq \langle u \rangle$.

Proof. Assume that for all $c \preceq u * v$ and for all $w \preceq \bar{u} * c$, we have $w \le \langle u \rangle$. Now there are minimal projections $(c_i)_i$ and $(w_{i,j})_{i,j}$ such that $u * v \sim \sum_i c_i$ and $\bar{u} * c_i \sim \sum_j w_{i,j}$. It follows that $u * v \sim \sum_{i,j} w_{i,j} \preceq \langle u \rangle$, but

$$v \sim e_1 * v \preceq (\bar{u} * u) * v = \bar{u} * (u * v) \preceq \langle u \rangle,$$

which is in contradiction with $v \not\preceq \langle u \rangle$. $\square$

For the distributive case, we can upgrade Proposition 5.2 as follows:

Theorem 5.7. *A distributive subfactor planar algebra with coatoms* $b_1, \dots, b_n \in [e_1, \mathrm{id}]$ *satisfying* $\sum_i \frac{1}{|\mathrm{id} : b_i|} \le 2$, *is w-cyclic.*

Proof. By Lemmas 2.2 and 4.24, we can assume the subfactor planar algebra to be boolean. If $K := \bigwedge_{i,j,i \neq j} (b_i \wedge b_j)^{\perp} \neq 0$, then consider $u \le K$ a minimal projection, and $Z(u)$ its central support. If $\langle Z(u) \rangle = \mathrm{id}$, then we are okay. Otherwise there exists i such that $\langle u \rangle = \langle Z(u) \rangle = b_i$. But $b_i^{\complement}$ is an atom in $[e_1, \mathrm{id}]$, so there is a minimal projection v such that $b_i^{\complement} = \langle v \rangle$. Recall that $b_i \wedge b_i^{\complement} = e_1$, so $v \not\preceq \langle u \rangle$, and by Lemma 5.6, there are minimal projections $c \preceq u * v$ and $w \preceq \bar{u} * c$ such that $w \not\preceq \langle u \rangle$ (and $\langle u, w \rangle = \mathrm{id}$ by maximality). By Lemma 3.3, there exists $u' \preceq c * \bar{u}$ with $Z(u') = Z(u)$ and $u' \not\le u$, but $u \le K$ so $u' \not\preceq b_i \wedge b_j$ for all $j \neq i$, and now $u' \le Z(u) \le b_i$, so $\langle u' \rangle = b_i$. Using distributivity (as for Theorem 4.26) we conclude

$$\langle c \rangle = \langle u, c \rangle \wedge \langle c, v \rangle \ge \langle u, w \rangle \wedge \langle u', v \rangle = \mathrm{id} \wedge \mathrm{id} = \mathrm{id}\,.$$

Otherwise $K = 0$, but $(b_i \wedge b_j)^{\perp} \ge b_j^{\perp}$ for all i, so $\bigwedge_{j \neq i} b_j^{\perp} = 0$ for all i. Let $p_1, \dots, p_r$ be the minimal central projections. Then $b_i = \bigoplus_{s=1}^{r} p_{i,s}$ with $p_{i,s} \le p_s$ and $p_{i,1} = p_1 = e_1$. Now $b_i^{\perp} = \bigoplus_{s=1}^{r} (p_s - p_{i,s})$, so by assumption,

$$0 = \bigwedge_{j \neq i} \bigoplus_{s=1}^{r} (p_s - p_{j,s}) = \bigoplus_{s=1}^{r} \bigwedge_{j \neq i} (p_s - p_{j,s}), \quad \text{for all } i.$$

It follows that $p_s = \bigvee_{j\neq i} p_{j,s}$ for all i and s, so $\mathrm{tr}(p_s) \le \sum_{j\neq i} \mathrm{tr}(p_{j,s})$. Now if there exists s such that $p_{i,s} < p_s$ for all i, then $\langle p_s \rangle = id$, which is okay; otherwise for all s, there exists i with $p_{i,s} = p_s$, but $\sum_{j\neq i} \mathrm{tr}(p_{j,s}) \ge \mathrm{tr}(p_s)$, so $\sum_j tr(p_{j,s}) \ge 2\,\mathrm{tr}(p_s)$. Then

$$\sum_i \mathrm{tr}(b_i) \ge n \cdot \mathrm{tr}(e_1) + 2\sum_{s\neq 1} \mathrm{tr}(p_s) = 2\,\mathrm{tr}(\mathrm{id}) + (n-2)\,\mathrm{tr}(e_1).$$

Now $|\mathrm{id} : b_i| = \mathrm{tr}(\mathrm{id})/\,\mathrm{tr}(b_i)$, so

$$\sum_i \frac{1}{|\mathrm{id} : b_i|} \ge 2 + \frac{n-2}{|\mathrm{id} : e_1|}$$

which contradicts the assumption, because we can assume $n > 2$ by Corollary 5.4. The result follows. □

Remark 5.8. The converse is false because there exist w-cyclic distributive subfactor planar algebras with $\sum_i (1/|\mathrm{id} : b_i|) > 2$. For example, the subfactor $(R \rtimes S_2^n \subset R \rtimes S_3^n)$ is w-cyclic and $\mathcal{B}_n$, but $\sum_i (1/|\mathrm{id} : b_i|) = \frac{n}{3}$.

Corollary 5.9. *Every $\mathcal{B}_n$ subfactor planar algebra with $|\mathrm{id} : b| \ge \frac{n}{2}$, for any coatom $b \in [e_1, \mathrm{id}]$, is w-cyclic. Then for all $n \le 4$, any $\mathcal{B}_n$ subfactor planar algebra is w-cyclic.*

Proof. By assumption (following the notations of Theorem 5.7)

$$\sum_i \frac{1}{|\mathrm{id} : b_i|} \le \sum_i \frac{2}{n} = 2.$$

But $|\mathrm{id} : b| \ge 2$, so any $n \le 4$ works. □

Corollary 5.10. *A distributive subfactor planar algebra having less than* 32 *biprojections* (*or of index* < 32), *is w-cyclic.*

Proof. In this case, the top of $[e_1, \mathrm{id}]$ is boolean of rank $n < 5$, because $32 = 2^5$; the result follows by Lemma 4.24 and Corollary 5.9. □

Conjecture 5.11. *A distributive subfactor planar algebra is w-cyclic.*

By Lemmas 2.2 and 4.24, we can reduce Conjecture 5.11 to the boolean case, and then extend it to the top boolean case.

Remark 5.12. The converse of Conjecture 5.11 is false, because the group S_3 is linearly primitive but not cyclic (see Corollary 6.12).

Problem 5.13. What is the natural additional assumption (A) such that $\mathcal{P}$ is distributive if and only if it is w-cyclic and satisfies (A)?

Assuming Conjecture 5.11 and using Remark 2.1, we get:

Conjecture 5.14. *For any distributive subfactor planar algebra $\mathcal{P}$ and any biprojection $b \in \mathcal{P}_{2,+}$, the planar algebras $\mathcal{P}(e_1, b)$, $\mathcal{P}(b, \mathrm{id})$ and their duals are w-cyclic.*

Remark 5.15. The converse is false because the interval $[S_2, S_4]$, proposed by Zhengwei Liu, gives a counterexample.

Remark 5.16. A cyclic subfactor planar algebra satisfies Conjecture 5.14 (thanks to Theorem 4.27 and Lemma 4.10).

Problem 5.17. Is a Dedekind subfactor planar algebra $\mathcal{P}$ distributive if and only if for any biprojection $b \in \mathcal{P}_{2,+}$, the planar algebras $\mathcal{P}(e_1, b)$, $\mathcal{P}(b, \mathrm{id})$ and their duals are w-cyclic?

6. Applications

6A. ***A nontrivial upper bound.*** For any irreducible subfactor planar algebra $\mathcal{P}$, we exhibit a nontrivial upper bound for the minimal number of minimal 2-box projections generating the identity biprojection. We will use the notations of Section 3C.

Lemma 6.1. *Let $b' < b$ be biprojections. If $\mathcal{P}(b', b)$ is w-cyclic, then there is a minimal projection $u \in \mathcal{P}_{2,+}$ such that $\langle b', u\rangle = b$.*

Proof. Consider the isomorphisms of von Neumann algebras

$$l_b : \mathcal{P}_{2,+}(e_1, b) \to b\mathcal{P}_{2,+}b$$

and, with $a = l_b^{-1}(b')$,

$$r_a : \mathcal{P}_{2,+}(b', b) \to a * \mathcal{P}_{2,+}(e_1, b) * a.$$

Then, by assumption, the planar algebra $\mathcal{P}(b', b)$ is w-cyclic, so by Proposition 4.21, there exists a minimal projection $u' \in \mathcal{P}_{2,+}(e_1, b)$ such that

$$\langle a, u'\rangle = l_b^{-1}(b).$$

Then by applying the map l_b and Theorem 3.9, we get

$$b = \langle l_b(a), l_b(u')\rangle = \langle b', u\rangle$$

with $u = l_b(u')$ a minimal projection in $b\mathcal{P}_{2,+}b$, so in $\mathcal{P}_{2,+}$. □

Assuming Conjecture 5.11 and using Lemma 6.1, we get a nontrivial upper bound:

Conjecture 6.2. *The minimal number r of minimal projections generating the identity biprojection* (*i.e.*, $\langle u_1, \dots, u_r\rangle = \mathrm{id}$) *is at most the minimal length ℓ for an ordered chain of biprojections*

$$e_1 = b_0 < b_1 < \cdots < b_\ell = id$$

such that $[b_i, b_{i+1}]$ is distributive (*or better*, *top boolean*).

Remark 6.3. We can deduce some theorems from Conjecture 6.2, by adding some assumptions to $[b_i, b_{i+1}]$, according to Theorems 4.26 or 5.7.

Remark 6.4. Let $(N \subset M)$ be any irreducible finite index subfactor. We can deduce a nontrivial upper bound for the minimal number of (algebraic) irreducible sub-N-N-bimodules of M, generating M as a von Neumann algebra.

6B. ***Back to the finite groups theory.*** As applications, we get a dual version of Theorem 2.5, and for any finite group G, we get a nontrivial upper bound for the minimal number of irreducible components for a faithful complex representation. The action of G on the hyperfinite II_1 factor R is always assumed to be outer.

Theorem 6.5 [Burnside 1911, § 226]. *A complex representation V of a finite group G is faithful if and only if any irreducible complex representation W is equivalent to a subrepresentation of $V^{\otimes n}$, for some $n \geq 0$.*

Definition 6.6. A group G is linearly primitive if it admits a faithful irreducible complex representation.

Definition 6.7. Let W be a representation of a group G, K be a subgroup of G, and X be a subspace of W. Let the *fixed-point subspace* be

$$W^K := \{w \in W \mid kw = w, \text{ for all } k \in K\}$$

and the *pointwise stabilizer subgroup*

$$G_{(X)} := \{g \in G \mid gx = x, \text{ for all } x \in X\}.$$

Definition 6.8. An interval $[H, G]$ is said to be linearly primitive if there is an irreducible complex representation V of G with $G_{(V^H)} = H$.

The group G is linearly primitive if and only if the interval $[\{e\}, G]$ is.

Lemma 6.9. *Let H be a core-free subgroup of G. Then G is linearly primitive if $[H, G]$ is so.*

Proof. Take V as above. Now, $V^H \subset V$ so $G_{(V)} \subset G_{(V^H)}$, but $\ker(\pi_V) = G_{(V)}$, so it follows that $\ker(\pi_V) \subset H$, but H is a core-free subgroup of G, and $\ker(\pi_V)$ is a normal subgroup of G, so $\ker(\pi_V) = \{e\}$, which means that V is faithful on G, i.e., G is linearly primitive. □

Lemma 6.10. *Letting $p_x \in \mathcal{P}_{2,+}(R^G \subseteq R)$ be a minimal projection on the one-dimensional subspace $\mathbb{C}x$ and H a subgroup of G, then*

$$p_x \leq b_H := |H|^{-1} \sum\nolimits_{h \in H} \pi_V(h) \Leftrightarrow H \subset G_x.$$

Proof. If $p_x \leq b_H$ then $b_H x = x$ and for every $h \in H$ we have that

$$\pi_V(h)x = \pi_V(h)[b_H x] = [\pi_V(h) \cdot b_H]x = b_H x = x$$

which means that $h \in G_x$, and so $H \subset G_x$. Conversely, if $H \subset G_x$ (i.e., $\pi_V(h)x = x$ for every $h \in H$) then $b_H x = x$, which means that $p_x \leq b_H$. □

Theorem 6.11. *Let $[H, G]$ be an interval of finite groups. Then*

- $(R \rtimes H \subseteq R \rtimes G)$ *is w-cyclic if and only if* $[H, G]$ *is* H*-cyclic.*
- $(R^G \subseteq R^H)$ *is w-cyclic if and only if* $[H, G]$ *is linearly primitive.*

Proof. By Proposition 4.21, $(R \rtimes H \subseteq R \rtimes G)$ is w-cyclic if and only if there exists a minimal projection u in

$$\mathcal{P}_{2,+}(R \subseteq R \rtimes G) \simeq \bigoplus_{g \in G} \mathbb{C} e_g \simeq \mathbb{C}^G$$

such that $\langle b, u \rangle = \mathrm{id}$, with $b = e^{R \rtimes G}_{R \rtimes H}$, and $r_b^{-1}(b * u * b)$ is a minimal projection if and only if there exists $g \in G$ such that $\langle H, g \rangle = G$, because u is of the form e_g and $Hg'H = HgH$ for all $g' \in HgH$.

By Proposition 4.18, $(R^G \subseteq R^H)$ is w-cyclic if and only if there exists a minimal projection u in

$$\mathcal{P}_{2,+}(R^G \subseteq R) \simeq \bigoplus_{V_i \text{ irr}} \mathrm{End}(V_i) \simeq \mathbb{C}G$$

such that $\langle u \rangle = e^R_{R^H}$, if and only if, by Lemma 6.10, $H = G_x$ with $u = p_x$ the projection on $\mathbb{C}x \subseteq V_i$ (with $Z(p_x) = p_{V_i}$). Note that $H \subset G_{(V_i^H)} \subset G_x$ so $H = G_{(V_i^H)}$. □

Corollary 6.12. *The subfactor* $(R^G \subseteq R)$ (*respectively*, $(R \subseteq R \rtimes G)$) *is w-cyclic if and only if G is linearly primitive* (*respectively*, *cyclic*).

Examples 6.13. The subfactors $(R^{S_4} \subset R^{S_2})$, its dual and $(R^{S_3} \subset R)$, are w-cyclic, but $(R \subset R \rtimes S_3)$ and $(R^{S_4} \subset R^{\langle (1,2)(3,4) \rangle})$ are not.

By Theorem 6.11, the group-theoretic reformulation of Conjecture 5.11 on $(R^G \subseteq R^H)$ is the following dual version of Theorem 2.5.

Conjecture 6.14. *Let $[H, G]$ be a distributive interval of finite groups. Then there exists an irreducible complex representation V of G such that $G_{(V^H)} = H$.*

If, moreover, H is core-free, then G is linearly primitive (Lemma 6.9).

Problem 6.15. Is a finite group G linearly primitive if and only if there is a core-free subgroup H such that the interval $[H, G]$ is bottom boolean?

By Theorem 6.5, Conjecture 6.2 on $\mathcal{P}(R^G \subseteq R)$ reformulates as follows:

Conjecture 6.16. *The minimal number of irreducible components for a faithful complex representation of a finite group G is at most the minimal length ℓ for an ordered chain of subgroups*

$$\{e\} = H_0 < H_1 < \cdots < H_\ell = G$$

such that $[H_i, H_{i+1}]$ *is distributive* (*or better*, *bottom boolean*).

This provides a bridge linking combinatorics and representations in the theory of finite groups.

Remark 6.17. We can upgrade Conjecture 6.16 by taking for H_0 any core-free subgroup of H_1, instead of just $\{e\}$; we can also deduce some theorems, by adding some assumptions to $[H_i, H_{i+1}]$, according to the group-theoretic reformulation of Theorems 4.27 or 5.7. Note that a normal biprojection in $\mathcal{P}(R^G \subseteq R^H)$ is given by a subgroup $K \in [H, G]$ with $HgK = KgH$ for all $g \in G$, see [Teruya 1998, Proposition 3.3].

Remark 6.18. We can also formulate results for finite quantum groups (i.e., finite-dimensional Kac algebras), where the biprojections correspond to the left coideal $\star$-subalgebras, see [Izumi et al. 1998, Theorem 4.4].

Acknowledgments

This work is supported by the Institute of Mathematical Sciences, Chennai. The author is grateful to Vaughan Jones, Dietmar Bisch, Scott Morrison and David Evans for their recommendation for a postdoc at the IMSc. Thanks to my hosts V.S. Sunder and Vijay Kodiyalam, and to Zhengwei Liu, Feng Xu, Keshab Chandra Bakshi and Mamta Balodi, for useful advice and fruitful exchanges.

References

[Afzaly et al. 2015] N. Afzaly, S. Morrison, and D. Penneys, "The classification of subfactors with index at most $5\frac{1}{4}$", preprint, 2015. arXiv

[Asaeda and Haagerup 1999] M. Asaeda and U. Haagerup, "Exotic subfactors of finite depth with Jones indices $(5+\sqrt{13})/2$ and $(5+\sqrt{17})/2$", *Comm. Math. Phys.* **202**:1 (1999), 1–63. MR Zbl

[Bakshi 2016] K. C. Bakshi, "Intermediate planar algebra revisited", preprint, 2016. arXiv

[Bisch 1994] D. Bisch, "A note on intermediate subfactors", *Pacific J. Math.* **163**:2 (1994), 201–216. MR Zbl

[Burnside 1911] W. Burnside, *Theory of groups of finite order*, 2nd ed., Cambridge Univ. Press, 1911. Zbl

[Izumi 2001] M. Izumi, "The structure of sectors associated with Longo–Rehren inclusions, II: Examples", *Rev. Math. Phys.* **13**:5 (2001), 603–674. MR Zbl

[Izumi 2002] M. Izumi, "Characterization of isomorphic group-subgroup subfactors", *Int. Math. Res. Not.* **2002**:34 (2002), 1791–1803. MR Zbl

[Izumi et al. 1998] M. Izumi, R. Longo, and S. Popa, "A Galois correspondence for compact groups of automorphisms of von Neumann algebras with a generalization to Kac algebras", *J. Funct. Anal.* **155**:1 (1998), 25–63. MR Zbl

[Jones 1980] V. F. R. Jones, *Actions of finite groups on the hyperfinite type* II_1 *factor*, Mem. Amer. Math. Soc. **237**, American Mathematical Society, Providence, RI, 1980. MR Zbl

[Jones 1983] V. F. R. Jones, "Index for subfactors", *Invent. Math.* **72**:1 (1983), 1–25. MR Zbl

[Jones 1999] V. F. R. Jones, "Planar algebras, I", 1999. To appear in *New Zealand J. Math.* Zbl arXiv

[Jones and Sunder 1997] V. F. R. Jones and V. S. Sunder, *Introduction to subfactors*, London Mathematical Society Lecture Note Series **234**, Cambridge Univ. Press, 1997. MR Zbl

[Jones et al. 2014] V. F. R. Jones, S. Morrison, and N. Snyder, "The classification of subfactors of index at most 5", *Bull. Amer. Math. Soc.* (*N.S.*) **51**:2 (2014), 277–327. MR Zbl

[Kodiyalam and Sunder 2000] V. Kodiyalam and V. S. Sunder, "The subgroup-subfactor", *Math. Scand.* **86**:1 (2000), 45–74. MR Zbl

[Kodiyalam and Sunder 2004] V. Kodiyalam and V. S. Sunder, "On Jones' planar algebras", *J. Knot Theory Ramifications* **13**:2 (2004), 219–247. MR Zbl

[Landau 1998] Z. A. Landau, *Intermediate subfactors*, Ph.D. thesis, University of California, Berkeley, 1998, Available at https://search.proquest.com/docview/304420808.

[Landau 2002] Z. A. Landau, "Exchange relation planar algebras", *Geom. Dedicata* **95** (2002), 183–214. MR

[Liu 2016] Z. Liu, "Exchange relation planar algebras of small rank", *Trans. Amer. Math. Soc.* **368**:12 (2016), 8303–8348. MR Zbl

[Nakamura and Takeda 1960] M. Nakamura and Z. Takeda, "On the fundamental theorem of the Galois theory for finite factors", *Proc. Japan Acad.* **36** (1960), 313–318. MR

[Ore 1938] O. Ore, "Structures and group theory, II", *Duke Math. J.* **4**:2 (1938), 247–269. MR Zbl

[Palcoux 2013] S. Palcoux, "Non-'weakly group theoretical' integral fusion categories?", message board post, 2013, Available at http://mathoverflow.net/q/132866.

[Palcoux 2015] S. Palcoux, "Ore's theorem for cyclic subfactor planar algebras and applications", preprint, 2015. arXiv

[Peters 2010] E. Peters, "A planar algebra construction of the Haagerup subfactor", *Internat. J. Math.* **21**:8 (2010), 987–1045. MR Zbl

[Stanley 2012] R. P. Stanley, *Enumerative combinatorics, I*, 2nd ed., Cambridge Studies in Advanced Mathematics **49**, Cambridge Univ. Press, 2012. MR Zbl

[Teruya 1998] T. Teruya, "Normal intermediate subfactors", *J. Math. Soc. Japan* **50**:2 (1998), 469–490. MR Zbl

[Watatani 1996] Y. Watatani, "Lattices of intermediate subfactors", *J. Funct. Anal.* **140**:2 (1996), 312–334. MR Zbl

[Xu 2013] F. Xu, "On a problem about tensor products of subfactors", *Adv. Math.* **246** (2013), 128–143. MR Zbl

Received February 5, 2017. Revised July 26, 2017.

SEBASTIEN PALCOUX
INSTITUTE OF MATHEMATICAL SCIENCES
CHENNAI 600113
INDIA
sebastienpalcoux@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 292, No. 1, 2018
dx.doi.org/10.2140/pjm.2018.292.223

DIVISIBILITY OF BINOMIAL COEFFICIENTS AND GENERATION OF ALTERNATING GROUPS

JOHN SHARESHIAN AND RUSS WOODROOFE

We examine an elementary problem on prime divisibility of binomial coefficients. Our problem is motivated by several related questions on alternating groups.

1. Introduction

We will discuss several closely related problems. The first is an elementary problem concerning divisibility of binomial coefficients by primes. Consider the following condition that a positive integer n might satisfy:

(1) There exist primes p and r such that if $1 \leq k \leq n-1$, then the binomial coefficient $\binom{n}{k}$ is divisible by at least one of p or r.

Question 1.1. *Does Condition (1) hold for all positive integers n?*

We were led to ask Question 1.1 by a problem on the alternating groups. Indeed, we consider several related group-theoretic conditions on a positive integer n:

(2) There exist primes p and r such that if $H < A_n$ is a proper subgroup, then the index $[A_n : H]$ is divisible by at least one of p or r.

(2') There exist primes p and r such that if P is a Sylow p-subgroup and R a Sylow r-subgroup of A_n, then $\langle P, R\rangle = A_n$.

(3) There exist a prime p and a conjugacy class D in A_n consisting of elements of prime power order, such that if P is a Sylow p-subgroup of A_n and $d \in D$, then $\langle P, d\rangle = A_n$.

(4) There exist conjugacy classes C and D in A_n, both consisting of elements of prime power order, such that if $(c, d) \in C \times D$, then $\langle c, d\rangle = A_n$.

(5) There exist conjugacy classes C and D in A_n, both consisting of elements of prime order, such that if $(c, d) \in C \times D$, then $\langle c, d\rangle = A_n$.

The first author was supported in part by NSF Grants DMS-0902142 and DMS-1202337.
MSC2010: 05A10, 05E15, 11B65, 20B30.
Keywords: alternating group, binomial coefficients, generation, prime density.

If we wish to specify one or both of the primes, then we may say that n satisfies Condition (1) with p, or that n satisfies Condition (1) with p and r. We'll use similar language for the other conditions.

Conditions (2) and (2') are equivalent, and each condition in the above list implies the previous condition. That is, for any positive integer n the following chain of implications holds, where the primes p and r may be held fixed.

$$(5) \Longrightarrow (4) \Longrightarrow (3) \Longrightarrow (2') \Longleftrightarrow (2) \Longrightarrow (1). \tag{1-1}$$

See also Theorem 1.3 below.

All implications in (1-1) are completely trivial or immediate from the definition of a Sylow subgroup, with the exception of the implication (2) $\Longrightarrow$ (1). This implication follows since A_n has subgroups of index $\binom{n}{k}$ for each $0 \leq k \leq n$. (The stabilizer in A_n of a k-subset of $[n]$ is such a subgroup.)

There are infinitely many positive integers n that do not satisfy Condition (5). However, the set of such integers is rather sparse, and likely very sparse. See Proposition 1.6 and Theorem 1.5 below. We are not aware of any integer n for which Conditions (1)–(4) fail to hold. In addition to Question 1.1, we will consider the following.

Questions 1.2–1.4. *Do Conditions (2)–(4) hold for all positive integers n?*

1A. ***Motivations and related questions.*** Question 1.1 fits into a line of inquiry going back to Kummer [1852] on the distribution of binomial coefficients that are divisible by a given prime. The remaining conditions and questions arose from our work and that of others on generation of finite simple groups. Recall that the classification of finite simple groups tells us that every simple group is isomorphic to one of the following: an alternating group A_n with $n \geq 5$, a cyclic group of prime order, a group of Lie type, or one of twenty six sporadic groups. Conditions analogous to Conditions (2)–(5) are known or conjectured for sporadic and Lie type groups.

We became interested in these problems via Question 1.2. In [Shareshian and Woodroofe 2016], we define a group G to be *universally* (p, r)*-generated* if $G = \langle P, R \rangle$ for any Sylow p-subgroup P and Sylow r-subgroup R. (Compare with Condition (2')!) We say G is *universally* $(2, *)$*-generated* if there is some prime p such that G is universally $(2, p)$-generated. We showed the following.

Theorem 1.2 [Shareshian and Woodroofe 2016]. *If G is a finite simple group that is abelian, of Lie type, or sporadic, then G is universally* $(2, *)$*-generated.*

We used Theorem 1.2, along with fixed-point theorems of Smith [1941] and Oliver [1975], to show that the order complex of the coset poset of any finite group is noncontractible.

In light of Theorem 1.2, it is natural to ask whether A_n is universally $(2, *)$-generated for every n — that is, whether every n satisfies Condition (2) with 2. This is not the case. The first failure of universal $(2, *)$-generation is at $n = 7$. It may be easier to understand the second failure, at 15, since $n = 15$ does not even satisfy Condition (1) with 2. Question 1.2 naturally suggests itself. We will further discuss the case $p = 2$ below in Section 1C.

We found that similar conditions had been examined earlier. The general problem of generation by elements selected from fixed conjugacy classes has been more broadly studied under the name of "invariable generation". See for example [Dixon 1992; Kantor et al. 2011; Detomi and Lucchini 2015; Eberhard et al. 2017]. Dolfi, Guralnick, Herzog and Praeger first asked Question 1.4 in [Dolfi et al. 2012, Section 6]. These authors conjecture that the analogue of Condition (5) holds for all but finitely many simple groups of Lie type, but point out that the corresponding statement for alternating groups occasionally fails.

Condition (3) interpolates naturally between Conditions (2) and (4). Although they do not ask Question 1.3, Damian and Lucchini [2007] show that an analogue of Condition (3) holds for many sporadic simple groups and groups of Lie type. Indeed, they show that many simple groups are generated by a Sylow 2-subgroup P together with any element of a certain conjugacy class consisting of elements of prime order.

1B. ***Results for arbitrary primes.*** Our first result adds an additional implication to the list in (1-1).

Theorem 1.3. *Let p and r be primes. If the positive integer n is not a prime power, then Conditions (1) and (2) are equivalent for n with p and r.*

The case where n is a prime power is not difficult.

Proposition 1.4. *If n is a power of the prime p, then*

(A) *n satisfies Condition (3) with a Sylow 2-subgroup unless $n = 7$, and*

(B) *n satisfies Condition (4) with p.*

In particular, it follows from Theorem 1.3 and Proposition 1.4 that Questions 1.1 and 1.2 are equivalent. We remark that the requirement $n \neq 7$ in Proposition 1.4(A) is necessary, as $n = 7$ satisfies Condition (1), but not Condition (2), with the prime 2.

While Questions 1.1–1.4 are still open, we have amassed a large collection of integers for which the answers are "yes". The *asymptotic density* [Niven et al. 1991] of a set S of positive integers is defined to be

$$\liminf_{M\to\infty} \frac{|S \cap [M]|}{M}.$$

Dolfi, Guralnick, Herzog and Praeger [2012] remark that Condition (5) appears likely to hold with asymptotic density 1. We show the following:

Theorem 1.5. *Let α be the asymptotic density of the set of positive integers n that satisfy Condition (5), and let ρ denote the Dickman–de Bruijn function (see for example [Granville 2008]). We have*

(A) $\alpha \geq 1 - \rho(20) > 1 - 10^{-28}$, *and*

(B) *if either the Riemann hypothesis or the Cramér conjecture holds, then $\alpha = 1$.*

The authors also claim in [Dolfi et al. 2012] that Condition (5) fails for infinitely many values of n, and that the smallest n for which Condition (5) fails is 210. We will see that the first claim is true, but the second is not.

Proposition 1.6. *For any $a \geq 3$, the integer $n = 2^a$ fails to satisfy Condition (5).*

Theorem 1.5 suggests a positive answer to Questions 1.1–1.4 for all but a vanishingly sparse set of large integers. We have also examined many small integers with the aid of a computer, verifying the following.

Proposition 1.7. *Every $n \leq 1{,}000{,}000{,}000$ satisfies Condition (2).*

The key tool in the proofs of both Theorem 1.5 and Proposition 1.7 is the following sieve lemma.

Lemma 1.8 (sieve lemma). *Let $n \geq 9$ be an integer. Let p and r be primes, and let a and b be positive integers.*

(A) *If n is not a prime power, p^a divides n, and $r^b < n < r^b + p^a$, then n satisfies Condition (2) with p and r.*

(B) *If p divides n and $r + 2 < n < r + p$, then n satisfies Condition (5) with p and r.*

Theorem 1.5 follows from combining Lemma 1.8 with known results on prime gaps and smooth numbers. We also use Lemma 1.8 to do much of the work in verifying Proposition 1.7.

For those integers not handled by Lemma 1.8(A), Theorem 1.3 tells us that it suffices to check divisibility of binomial coefficients. In particular, we can avoid making any computations in large alternating groups. We do not know how to avoid such computations for Condition (4). The slow speed of these computations is the main obstacle to a computational verification of Condition (4) for those values of n not addressed by Lemma 1.8.

1C. ***Results for $p = 2$.*** We return now to the case where one of the primes in Condition (2) is 2. Theorem 1.2 suggests this case as being particularly worthy of attention, and Proposition 1.4 gives infinitely many values of n for which Condition (2) holds with 2.

However, there are also infinitely many positive integers n that do not even satisfy Condition (1) with 2. By a theorem of Kummer (see Lemma 3.1 below), if

$n = 2^a - 1$ for some positive integer a, then $\binom{n}{k}$ is odd for all $1 \le k \le n-1$. (Indeed, a similar statement holds for any prime p. In the language of group-actions, this says that any Sylow p-subgroup of S_{p^a-1} stabilizes a set of every possible size k with $1 < k < p^a - 1$.) Kummer's theorem also implies that there is no prime dividing every nontrivial $\binom{n}{k}$ unless n is a prime power. There are infinitely many n of the form $2^a - 1$ that are not prime powers.

Using techniques similar to those for Proposition 1.7, we computationally verify the following.

Proposition 1.9. *About* 86.7% *of the positive integers* $n \le 1{,}000{,}000$ *satisfy Condition* (2) *with* 2.

1D. ***Organization.*** We begin in Section 2 by giving necessary background on maximal subgroups of alternating groups. In Section 3 we state the well-known theorem of Kummer on prime divisibility of binomial coefficients, and prove an analogue on prime divisibility of the number of equipartitions of a set. We use these results in Section 4 to prove Theorem 1.3, Propositions 1.4 and 1.6, and Lemma 1.8. We also verify that Condition (4) holds for all small alternating groups. We apply Lemma 1.8 to prove Theorem 1.5 in Section 5. We describe our computational verification of Propositions 1.7 and 1.9 in Section 6.

2. Preliminaries

In this section we discuss necessary background on alternating and symmetric groups. Readers familiar with basic facts about permutation groups can safely skip this section.

In order to show that the index of every subgroup of the alternating group A_n is divisible by either p or r, it suffices to show the same for every maximal subgroup. The maximal subgroups of A_n are well-understood, as we now review. Additional background can be found in [Dixon and Mortimer 1996], see also [Liebeck et al. 1987].

We say that a subgroup $H \le A_n$ is transitive or primitive if the action of H on $[n]$ satisfies the same property. That is, H is *transitive* if for every $i, j \in [n]$, there is some $\sigma \in H$ such that $i \cdot \sigma = j$. A transitive subgroup H is *imprimitive* if there is a proper partition π of $[n]$ into sets of size greater than one, such that the parts of π are permuted by the action of H. If H is transitive and not imprimitive, then it is *primitive*. Clearly, every subgroup is either intransitive, imprimitive, or primitive. We examine maximal subgroups of A_n according to this trichotomy.

An intransitive subgroup H is maximal in the (sub)poset of intransitive subgroups of A_n if and only if H is the stabilizer in A_n of some nonempty proper subset $X \subset [n]$. As A_n sits naturally in S_n, it is illuminating to also consider the stabilizer H^+ in S_n of X. Then $H = H^+ \cap A_n$. It is clear that $H^+ \cong S_{|X|} \times S_{n-|X|}$. If $|X| = k$, then it follows either from this isomorphism or the orbit-stabilizer theorem that

$$[A_n : H] = [S_n : H^+] = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k}.$$

Every imprimitive subgroup of A_n stabilizes a partition of $[n]$. It follows easily that a subgroup H is maximal in the (sub)poset of imprimitive subgroups of A_n if and only if H is the stabilizer of a partition of $[n]$ into n/d parts of size d for some nontrivial proper divisor d of n. As in the intransitive case, we also consider the stabilizer H^+ of the same partition in the action by S_n. Then H^+ is isomorphic to the wreath product $S_d \wr S_{n/d}$. Since $H = H^+ \cap A_n$ (and $H^+ \not\leq A_n$), we see that

$$[A_n : H] = [S_n : H^+] = \frac{n!}{(d!)^{n/d} \cdot (n/d)!}.$$

By either the orbit-stabilizer theorem or an elementary counting argument, $[A_n : H]$ counts the number of partitions of $[n]$ into n/d equal-sized parts.

The index of a primitive proper subgroup of A_n is typically divisible by every prime smaller than n. See Theorem 4.1 and the discussion following for a precise statement.

3. Kummer's theorem and an analogue

3A. ***Kummer's theorem.*** We make considerable use of the following result of Kummer. The most useful case of the lemma for us will be that where $a = 1$. See also [Granville 1997] for an overview of related results.

Lemma 3.1 (Kummer's theorem [1852, pp. 115–116]). *Let k and n be integers with $0 \leq k \leq n$. If a is a positive integer, then p^a divides $\binom{n}{k}$ if and only if at least a carries are needed when adding k and $n-k$ in base p.*

3B. ***An analogue for the number of equipartitions.*** Lemma 3.1 completely describes the prime divisibility of indices of intransitive maximal subgroups of A_n. Lemma 3.2 below provides a weaker but similarly useful characterization regarding indices of imprimitive subgroups. Throughout this section, if d is a nontrivial proper divisor of the positive integer n, then we will write $I_{n,d}$ for the number of equipartitions of n into parts of size d. Thus,

$$I_{n,d} = \frac{n!}{(d!)^{n/d} \cdot (n/d)!}.$$

Lemma 3.2. *Let n be a positive integer, d be a nontrivial proper divisor of n, and p be a prime. Then p divides $I_{n,d}$ if and only if*

(1) *at least one carry is necessary when adding n/d copies of d in base p, and*

(2) *d is not a power of p.*

Proof. It is straightforward to show by elementary arguments that

$$I_{n,d} = \frac{1}{(n/d)!} \cdot \prod_{j=1}^{n/d} \binom{jd}{d} = \prod_{j=1}^{n/d} \frac{1}{j} \binom{jd}{d} = \prod_{j=1}^{n/d} \binom{jd-1}{d-1}. \tag{3-1}$$

Our strategy is to use Lemma 3.1 to examine divisibility of the terms in these products.

Case 1 ($n/d < p$). In this case p does not divide $(n/d)!$, and the first condition of the hypothesis implies the second. From (3-1) we thus see that p divides $I_{n,d}$ if and only if p divides $\binom{jd}{d}$ for some $1 \le j \le n/d$. The claim for this case then follows from Lemma 3.1.

Case 2 ($n/d \ge p$). In this case a carry is always necessary when adding n/d copies of d, so we need only consider the second condition of the hypothesis.

If the base p expansion of d has at least 2 nonzero places, then there are at least 2 carries when adding d to $pd - d$, as the base p expansion of pd is obtained by shifting that of d to the left by one place. It follows that p^2 divides $\binom{pd}{d}$, hence that p divides $\binom{pd-1}{d-1} = \frac{1}{p}\binom{pd}{d}$. By (3-1), p divides $I_{n,d}$.

Otherwise, we have $d = kp^a$ for some $1 \le k < p$. Then the base p expansion of $(jd-1)-(d-1) = (j-1)kp^a$ vanishes below the a-th place. Also, the base p expansion of $d-1$ is $(k-1)p^a + \sum_{i=0}^{a-1}(p-1)p^i$. As the latter vanishes above the a-th place, this place is the only possible location for a carry in adding $d-1$ and $jd-1$. If $k=1$, then the a-th place of $d-1$ is 0, so no carry occurs and for no j does p divide $\binom{jd-1}{d-1}$. If $k > 1$, then a carry occurs at the a-th place for values of j such that $(j-1)\cdot k \equiv p-1 \mod p$. (Such a $j < n/d$ exists since $\mathbb{Z}/p\mathbb{Z}$ is a field.) □

Remark 3.3. After submission of the paper, we became aware that a slightly different (from Lemma 3.2) characterization of prime divisibility of $I_{n,d}$ appears as [Thompson 1966, Lemma 2].

Corollary 3.4. *Let n and b be positive integers and r be a prime, such that $n/2 < r^b \le n$. If d is a nontrivial proper divisor of n which is not a power of r, then r divides $I_{n,d}$.*

Proof. Since $r^b > n/2$, there is a 1 in the b-th place of the base r expansion of n. On the other hand, $d \le n/2$. Hence, the base r expansion of d has a 0 in the b-th place. It follows that there is at least one carry when we sum n/d copies of d. Lemma 3.2 then gives that r divides $I_{n,d}$ unless d is a power of r. □

One can indeed extract from (3-1) the highest power of p dividing $I_{n,d}$, but we will not need to do so.

4. Proofs of the sieve lemma and other tools

In this section we prove several results that we will use as tools in the sections that follow, including Theorem 1.3, Lemma 1.8, and Propositions 1.4 and 1.6.

4A. ***Proof of Theorem 1.3.*** Suppose n satisfies Condition (2) with p and r. As described in Section 2, the maximal intransitive subgroups of A_n are stabilizers of k-subsets of $[n]$, and have index $\binom{n}{k}$ in A_n. Hence, n also satisfies Condition (1) with p and r. See the discussion following (1-1).

Thus, in order to prove Theorem 1.3, it suffices to show that if n satisfies Condition (1) with p and r, then the index of every primitive or imprimitive maximal subgroup is divisible by at least one of p or r.

For the primitive case, we use the following version of a classic theorem of permutation group theory due to Jordan.

Theorem 4.1 [Jordan 1875; Dixon and Mortimer 1996, Section 3.3]. *Let $n \geq 9$ and let H be a primitive subgroup of A_n:*

(1) *If $p \leq n-3$ is a prime and H contains a p-cycle, then $H = A_n$.*

(2) *If H contains the product of two transpositions, then $H = A_n$.*

The next lemma follows quickly.

Lemma 4.2. *Let p be a prime. If $n \geq 9$ and $p \leq n-3$, then p divides the index of every primitive proper subgroup of A_n.*

Proof. If p is odd, then every Sylow p-subgroup of A_n contains a p-cycle. Similarly, every Sylow 2-subgroup of A_n contains an element that is the product of two transpositions. In either case, Theorem 4.1 gives that no primitive proper subgroup of A_n contains any Sylow p-subgroup of A_n. □

Since Lemma 4.2 only applies when $n \geq 9$, we pause to handle the situation when $n < 9$. The only integer less than 9 that is not a prime power is 6, and the equivalence of Conditions (1) and (2) for $n = 6$ is obtained by direct inspection (see Table 1 below).

Now assume as above that $n \geq 9$ satisfies Condition (1) with p and r. Since n is not a prime power, we see from Lemma 3.1 that p and r must be distinct, hence one must be smaller than $n-2$. As $n \geq 9$, it follows from Lemma 4.2 that the index of every primitive proper subgroup is divisible by at least one of p or r, as desired.

We now handle the imprimitive case, using Lemma 3.2. Let d be a divisor of n. We notice that if p divides $\binom{n}{d}$, then adding $n-d$ and d in base p requires a carry (by Lemma 3.1). It follows immediately from Lemma 3.2 that the index $n!/((d!)^{n/d} \cdot (n/d)!)$ of an imprimitive maximal subgroup is divisible by either p or r, except possibly if d is a power of p or r.

Suppose that d is a power of p, and that p^a is the highest power of p dividing n. Then Lemma 3.1 shows that $\binom{n}{p^a}$ is not divisible by p, hence it is divisible by r. Adding n/p^a copies of p^a in base r therefore requires a carry. Since $d \le p^a$, adding n/d copies of d in base r will also require a carry. Therefore, $n!/((d!)^{n/d} \cdot (n/d)!)$ is divisible by r, as desired. The case where d is a power of r is handled similarly.

4B. ***Proof of Lemma 1.8(A).*** Kummer's theorem (Lemma 3.1) gives us the following.

Lemma 4.3. *Let n be a positive integer and let p and r be distinct primes. If there are positive integers a and b such that $p^a \mid n$ and $r^b < n < p^a + r^b$, then for $0 < k < n$ at least one of p, r divides $\binom{n}{k}$.*

Proof. Notice that since $p^a > n - r^b$, either $k < p^a$ or else $k > n - r^b$. We assume without loss of generality that $k \le n/2$.

Let $k = \sum k_i p^i$ and $n = \sum n_i p^i$ respectively be the base p expansions of k and n. As $p^a \mid n$, therefore $n_i = 0$ for $i < a$. When $k < p^a$, then $k_j = 0$ for all $j \ge a$. Since $k \ne 0$, there is a carry when adding k and $n - k$ in base p. It follows from Lemma 3.1 that $p \mid \binom{n}{k}$.

When $k > n - r^b$, we notice that $k \le n/2 < r^b$, and therefore both k and $n - k$ are between $n - r^b$ and r^b. In particular, the b-th place of the base r expansion of both k and $n - k$ has a 0. Since $n/2 < r^b < n$, the b-th place of the base r expansion of n has a 1. It follows that there is a carry when adding k and $n - k$, hence by Lemma 3.1 that $r \mid \binom{n}{k}$. $\square$

Lemma 1.8 follows from Lemma 4.3 and Theorem 1.3.

4C. ***Proof of Lemma 1.8(B).*** Let $x \in A_n$ have cycle type $p^{n/p}$, that is, let x be the product of n/p pairwise disjoint p-cycles. (Since $p \ne 2$, a p-cycle is an even permutation.) Let $y \in A_n$ be an r-cycle. We take C to be the conjugacy class containing x, and D to be the conjugacy class containing y. Since we chose (x, y) arbitrarily from $C \times D$, it is enough to show $\langle x, y\rangle = A_n$, that is, that $\langle x, y\rangle$ is not contained in a maximal subgroup of any of the three types discussed in Section 2.

Since $r < n - 2$, it is immediate from Theorem 4.1 that $\langle y\rangle$ is contained in no maximal primitive subgroup.

If p is a proper divisor of n, we see that $p \le n/2$ and hence that $r > n - p \ge n/2$. It is then immediate by Corollary 3.4 that $\langle y\rangle$ is contained in no imprimitive maximal subgroup. Otherwise, if $n = p$, then A_n has no imprimitive maximal subgroups.

It remains to show that $\langle x, y\rangle$ is transitive in the natural action on $[n]$. Since y acts transitively on an r-set $Y \subseteq [n]$, it suffices to show that every $i \in [n]$ can be moved into Y by x. But i is permuted in a p-cycle by x, and since $r + p > n$, some element of this p-cycle must be in Y, as desired.

4D. ***Proof of Proposition 1.4.*** Direct inspection verifies the proposition for $n \leq 8$. See Table 1 below. We assume henceforth that $n \geq 9$.

We first verify part (B). By the Bertrand–Chebyshev theorem [Niven et al. 1991, Theorem 8.7] there is a prime r with $n/2 < r < n-2$. We let x be any r-cycle, and notice that $\langle x \rangle$ is a Sylow r-subgroup. Then r divides the index of any imprimitive or primitive maximal subgroup by Corollary 3.4 and Lemma 4.2 respectively.

We now take y to be any n-cycle in the case where $n = p^a$ is odd, or the product of any two disjoint 2^{a-1}-cycles in the case where $n = 2^a$ is even. In the former case, $\langle y \rangle$ is transitive. In the latter case, as $r > 2^{a-1}$, we see that $\langle x, y \rangle$ is transitive. In either case, $\langle x, y \rangle$ is contained in no intransitive maximal subgroup, hence

$$\langle x, y \rangle = A_n.$$

Since conjugation fixes cycle type, part (B) follows.

It remains to verify (A). In the case where n is even, it follows from part (B). Otherwise, we take y to be any n-cycle. Then $\langle y \rangle$ is transitive, while Lemmas 3.2 and 4.2 give that no imprimitive or primitive maximal subgroup contains a Sylow 2-subgroup. It follows that

$$\langle y, P \rangle = A_n$$

for any Sylow 2-subgroup P, completing the proof of part (A).

4E. ***Proof of Proposition 1.6.*** Let C and D be as in Condition (5). We will find $(c, d) \in C \times D$ such that $\langle c, d \rangle \neq A_n$.

Since A_n is transitive, if D does not consist of derangements then we may find an element d of D fixing n. The same holds for C. If c and d both fix n, then $\langle c, d \rangle$ is intransitive, hence a proper subgroup of A_n. This reduces us to the situation where one conjugacy class (without loss of generality C) consists of derangements.

Since $n = 2^a$, derangements of prime order in A_n are fixed-point-free involutions. It is straightforward to verify that the fixed-point-free involutions of A_n form a single conjugacy class. Thus, C consists of all fixed-point-free involutions in A_n.

Since a Sylow 2-subgroup of A_n intersects every conjugacy class of involutions nontrivially, we see that D must consist of elements of odd prime order p. For any $d \in D$, every orbit of $\langle d \rangle$ is of size 1 or p. If $\langle d \rangle$ has more than two orbits, then let O_1 and O_2 be orbits. Now there is some $c \in C$ such that $O_1 \cup O_2$ is the union of the supports of 2-cycles in the disjoint cycle decomposition of c. The subgroup $\langle c, d \rangle$ is thus intransitive.

It remains only to consider the case where $\langle d \rangle$ has exactly two orbits. As $n = 2^a$, so d is a p-cycle fixing exactly one point. Now $n = p + 1$, and so by the Sylow theorems the subgroups of order p in A_n form a single conjugacy class. Thus, it suffices to find a proper subgroup of A_n that contains both a fixed-point-free involution c and an element d of order p.

n	maximal subgroup indices	Condition (4) conjugacy class representatives
5	5, 6, 10	(1 2 3), (1 2 3 4 5)
6	6, 10, 15	(1 2 3 4)(5 6), (1 2 3 4 5)
7	7, 15, 21, 35	(1 2 3 4 5), (1 2 3 4 5 6 7)
8	8, 15, 28, 35, 56	(1 2 3 4)(5 6 7 8), (1 2 3 4 5)

Table 1. Indices of maximal subgroups and generating conjugacy class representatives for A_n, $5 \leq n \leq 8$.

Consider the transitive action of $\mathrm{PSL}_2(p)$ on the set $\mathcal{S}$ of 1-dimensional subspaces of $\mathbb{F}_p^2$. Since $|\mathcal{S}| = n$ and $\mathrm{PSL}_2(p)$ is simple, we obtain from the group action a subgroup $H \cong \mathrm{PSL}_2(p)$ of A_n. Then $|H| = (p \cdot (p^2 - 1))/2$, and by the orbit-stabilizer theorem, the stabilizer of any point has order $(p \cdot (p-1))/2 = p \cdot (2^{a-1} - 1)$. In particular, the subgroup H contains elements of order p and order 2, and no element of order 2 in H fixes any point.

Remark 4.4. Powers of 2 satisfy Condition (4) by Proposition 1.4.

4F. ***Very small alternating groups.*** As Lemma 1.8 does not apply when $n \leq 8$, we examine small n separately. The solvable alternating groups (where $n < 5$) all trivially satisfy Condition (5). For $5 \leq n \leq 8$, we present in Table 1 the indices of maximal subgroups of A_n, together with representatives for generating conjugacy classes as in Condition (4). This list is easy to produce either by GAP [2012], or else by hand (using well-known facts about primitive groups of small degree).

For $n = 5$ or 7, these representatives are of prime order, so 5 and 7 satisfy Condition (5). Proposition 1.6 tells us that 8 fails Condition (5), and a similar argument or GAP computation shows that 6 also fails Condition (5).

5. Asymptotic density

In this section, we use part (B) of Lemma 1.8 to prove Theorem 1.5.

Lemma 1.8 tells us that n satisfies Condition (5) unless both the largest prime divisor p of n and the largest prime r that is less than $n - 2$ are small relative to n. This allows us to apply known and conjectured results about prime gaps, which we combine with known results about numbers without large prime divisors ("smooth numbers").

We will use the following notation:

- We will denote the k-th smallest prime number by p_k. For example, $p_1 = 2$ and $p_2 = 3$.
- For a real number $x > 2$, we will denote by $r(x)$ the largest prime that is no larger than x.

- For positive real numbers x, y, we will denote by $\Psi(x, y)$ the number of positive integers no larger than x which have no prime factor larger than y.

Our strategy is to show that if p is the largest prime divisor of n, then asymptotically $r(n) + p$ is frequently greater than n. We remark that $r(n) \geq n - 2$ only on a set of asymptotic density 0, so we may treat the $r + 2 < n$ condition of Lemma 1.8 as reading $r \leq n$ for the purpose of asymptotic density arguments.

We will require several tools from number theory, as we will describe below. See [Granville 2008] for further background on (5-2) and (5-3), and [Granville 1995] for background and history on (5-4) and (5-5).

5A. ***Proof of Theorem 1.5(A).*** Jia [1996] showed that, for any $\epsilon > 0$, there is a prime on the interval $[n, n + n^{1/20+\epsilon}]$ for all n excluding a set of asymptotic density 0. It follows by routine manipulation that

$$n - r(n) < n^{1/20} \quad \text{except on a set of asymptotic density 0.} \tag{5-1}$$

See [Harman 2007, Chapter 9] for further discussion of results of this type.

Dickman [1930] showed that

$$\lim_{x\to\infty} \frac{\Psi(x, x^{1/u})}{x} = \rho(u) \quad \text{for any fixed } u, \tag{5-2}$$

where ρ denotes the so-called *Dickman–de Bruijn function*, that is, the solution to the differential equation $u\rho'(u) + \rho(u - 1) = 0$.

By combining (5-1) and (5-2) with Lemma 1.8, we see that the desired asymptotic density α satisfies

$$\alpha \geq 1 - \rho(20),$$

as desired. Consulting the table of values for ρ in [Granville 2008, Table 2], we see that $\rho(20) \cong 2.462 \cdot 10^{-29} < 10^{-28}$.

5B. ***Proof of Theorem 1.5(B).*** Rankin [1938] showed that

$$\lim_{x\to\infty} \frac{\Psi(x, \log^b x)}{x} = 0, \quad \text{for any } b > 1. \tag{5-3}$$

Taking $b = 3$ in (5-3), we see that the set of integers n with no prime factor larger than $\log^3 n$ has asymptotic density 0.

The Cramér conjecture [1936, (4)] says that there is a constant C such that

$$p_{k+1} - p_k \leq C \log^2 p_k \quad \text{for all } k. \tag{5-4}$$

In the same paper, Cramér [1936, Theorem II] showed the Riemann hypothesis to imply that

$$\lim_{x\to\infty} \frac{1}{x} \cdot \sum_{\substack{p_k \le x, \\ p_{k+1}-p_k \ge \log^3 p_k}} (p_{k+1} - p_k) = 0. \tag{5-5}$$

Thus, if either the Cramér conjecture or the Riemann hypothesis hold, then

$$n - r(n) \le \log^3 r(n) \le \log^3 n \tag{5-6}$$

except on a set of asymptotic density zero. Theorem 1.5(B) follows upon combining (5-3) with $b = 3$, (5-6), and Lemma 1.8.

6. Computational results

In this section we describe the verification by computer of Proposition 1.7.

Our program iterates through the integers, beginning with $n = 9$. We factor each integer into primes. If n is a prime power, then n satisfies Condition (3) and hence Condition (2) by Proposition 1.4. In this case, we store $n = r^b$ as the largest prime power known so far in the computation. Otherwise, we find the largest prime power p^a dividing n. The program then checks whether $p^a + r^b$ is greater than n, where r^b is the largest prime power found so far. If so, then n satisfies Condition (2) with p and r by Lemma 1.8. This sieving method succeeds for all but 14,638 of the integers in the interval from 9 to 1,000,000,000. For these remaining integers, the program checks directly which indices of intransitive and imprimitive subgroups are divisible by p (using Lemmas 3.1 and 3.2), and searches for a prime r dividing those that are not. This second method works for all but 22 of the remaining 14,638 integers. For these 22 integers we perform a similar search, using divisors of n other than p. See Table 2 for the results of this search.

Running this program out to $n = 1{,}000{,}000{,}000$ on a 2012 MacBook Pro with the GAP computer algebra system [GAP 2012] takes around 2 weeks. This computation verifies Proposition 1.7.

We approach checking which values of n satisfy Condition (2) with the prime 2 in a similar fashion. When we apply Lemma 1.8, we look for a pair $p^a + r^b > n > r^b$ (where $p^a \mid n$) as before, but now we require $2 \in \{p, r\}$. This technique gives a positive answer for about 45.7% of the first 1,000,000 integers $n \ge 9$. The remaining values of n require significantly more computation, and as a result we did not examine values of n beyond 1,000,000.

Running the program to check Condition (2) with the prime 2 out to $n = 1{,}000{,}000$ takes around a day on a 2012 MacBook Pro. This computation verifies Proposition 1.9. More precisely, 867,247 of the integers between 9 and 1,000,0000 satisfy Condition (2) with 2. The histogram in Figure 6.1 shows the density of

n	p^a	Condition (2) prime pairs
$31{,}416 = 2^3 \cdot 3 \cdot 7 \cdot 11 \cdot 17$	17^1	(2, 7853)
$46{,}800 = 2^4 \cdot 3^2 \cdot 5^2 \cdot 13$	5^2	(2, 149)
$195{,}624 = 2^3 \cdot 3^2 \cdot 11 \cdot 13 \cdot 19$	19^1	(2, 3)
$5{,}504{,}490 = 2 \cdot 3^3 \cdot 5 \cdot 19 \cdot 29 \cdot 37$	37^1	(3, 5)
$7{,}458{,}780 = 2^2 \cdot 3 \cdot 5 \cdot 7^2 \cdot 43 \cdot 59$	59^1	(2, 276251)
$9{,}968{,}112 = 2^4 \cdot 3^2 \cdot 7 \cdot 11 \cdot 29 \cdot 31$	31^1	(2, 3)
$12{,}387{,}600 = 2^4 \cdot 3^3 \cdot 5^2 \cdot 31 \cdot 37$	37^1	(2, 3)
$105{,}666{,}600 = 2^3 \cdot 3 \cdot 5^2 \cdot 13 \cdot 19 \cdot 23 \cdot 31$	31^1	(2, 5)
$115{,}690{,}848 = 2^5 \cdot 3 \cdot 7 \cdot 13 \cdot 17 \cdot 19 \cdot 41$	41^1	(2, 3)
$130{,}559{,}352 = 2^3 \cdot 3 \cdot 7 \cdot 11 \cdot 31 \cdot 43 \cdot 53$	53^1	(2, 112843)
$146{,}187{,}444 = 2^2 \cdot 3 \cdot 13 \cdot 19 \cdot 31 \cdot 37 \cdot 43$	43^1	(2, 31)
$225{,}613{,}050 = 2 \cdot 3 \cdot 5^2 \cdot 13 \cdot 37 \cdot 53 \cdot 59$	59^1	(2, 516277)
$275{,}172{,}996 = 2^2 \cdot 3 \cdot 7 \cdot 29 \cdot 37 \cdot 43 \cdot 71$	71^1	(2, 567367)
$282{,}429{,}840 = 2^4 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 17 \cdot 29 \cdot 31$	31^1	(2, 29)
$300{,}688{,}752 = 2^4 \cdot 3 \cdot 7 \cdot 13 \cdot 23 \cdot 41 \cdot 73$	73^1	(2, 11)
$539{,}509{,}620 = 2^2 \cdot 3 \cdot 5 \cdot 13 \cdot 17 \cdot 23 \cdot 29 \cdot 61$	61^1	(2, 1201)
$653{,}426{,}796 = 2^2 \cdot 3 \cdot 11 \cdot 19 \cdot 43 \cdot 73 \cdot 83$	83^1	(2, 73)
$696{,}595{,}536 = 2^4 \cdot 3^2 \cdot 7 \cdot 13 \cdot 17 \cdot 53 \cdot 59$	59^1	(2, 13)
$784{,}474{,}592 = 2^5 \cdot 11 \cdot 29 \cdot 31 \cdot 37 \cdot 67$	67^1	(2, 29)
$798{,}772{,}578 = 2 \cdot 3 \cdot 19 \cdot 29 \cdot 41 \cdot 71 \cdot 83$	83^1	(2, 563)
$815{,}224{,}800 = 2^5 \cdot 3 \cdot 5^2 \cdot 13 \cdot 17 \cdot 29 \cdot 53$	53^1	(2, 87013)
$851{,}716{,}320 = 2^5 \cdot 3 \cdot 5 \cdot 7 \cdot 13 \cdot 17 \cdot 31 \cdot 37$	37^1	(2, 31)

Table 2. The values of $n \le 1{,}000{,}000{,}000$ together with their maximal prime power divisors p^a, such that n does not satisfy Condition (2) with p. Each such n satisfies Condition (2) with either 2 or 3.

those n which do not satisfy Condition (2) with 2. We remark that this histogram appears to show that the failing values are concentrated towards the values of n slightly preceding integers that are divisible by a high power of 2.

Source code and output for all computer programs discussed in this section are available in the online supplement as ancillary files. They are also currently available from Woodroofe's web page. A list of the values of $n \le 1{,}000{,}000$ such that n does not satisfy Condition (2) with the prime 2 can be found in the same places.

Acknowledgements

We thank Andrew Granville and Bob Guralnick for their thoughtful remarks. A comment by Ben Green led to a significant improvement in the bound given in Theorem 1.5(A).

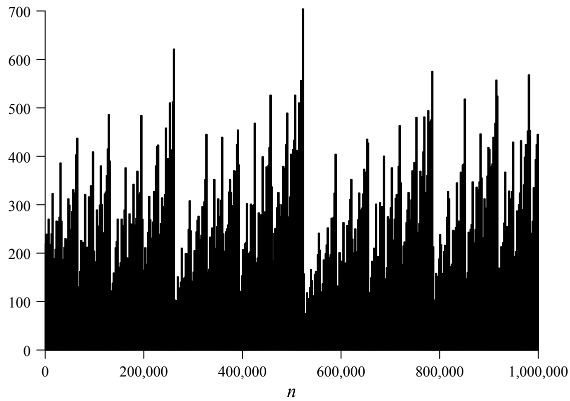


Figure 6.1. A histogram, for $n = 9$ to 1,000,000 in bins of size 2500, showing the density of integers that do not meet Condition (2) with the prime 2.

References

[Cramér 1936] H. Cramér, "On the order of magnitude of the difference between consecutive prime numbers", *Acta Arith.* **2**:1 (1936), 23–46. Zbl

[Damian and Lucchini 2007] E. Damian and A. Lucchini, "The probabilistic zeta function of finite simple groups", *J. Algebra* **313**:2 (2007), 957–971. MR Zbl

[Detomi and Lucchini 2015] E. Detomi and A. Lucchini, "Invariable generation with elements of coprime prime-power orders", *J. Algebra* **423** (2015), 683–701. MR Zbl

[Dickman 1930] K. Dickman, "On the frequency of numbers containing prime factors of a certain relative magnitude", *Ark. Mat. Astr. Fys.* **22A**:10 (1930), 1–14. JFM

[Dixon 1992] J. D. Dixon, "Random sets which invariably generate the symmetric group", *Discrete Math.* **105**:1-3 (1992), 25–39. MR Zbl

[Dixon and Mortimer 1996] J. D. Dixon and B. Mortimer, *Permutation groups*, Graduate Texts in Mathematics **163**, Springer, New York, 1996. MR Zbl

[Dolfi et al. 2012] S. Dolfi, R. M. Guralnick, M. Herzog, and C. E. Praeger, "A new solvability criterion for finite groups", *J. Lond. Math. Soc.* (2) **85**:2 (2012), 269–281. MR Zbl

[Eberhard et al. 2017] S. Eberhard, K. Ford, and B. Green, "Invariable generation of the symmetric group", *Duke Math. J.* **166**:8 (2017), 1573–1590. MR Zbl

[GAP 2012] "GAP – Groups, Algorithms, and Programming", version 4.5.5, 2012, available at http://www.gap-system.org/Releases/4.5.5.html.

[Granville 1995] A. Granville, "Harald Cramér and the distribution of prime numbers", *Scand. Actuar. J.* 1 (1995), 12–28. MR Zbl

[Granville 1997] A. Granville, "Arithmetic properties of binomial coefficients, I: Binomial coefficients modulo prime powers", pp. 253–276 in *Organic mathematics* (Burnaby, BC, 1995), edited by J. Borwein et al., CMS Conf. Proc. **20**, American Mathematical Society, Providence, RI, 1997. MR Zbl

[Granville 2008] A. Granville, "Smooth numbers: computational number theory and beyond", pp. 267–323 in *Algorithmic number theory: lattices, number fields, curves and cryptography*, edited by

J. P. Buhler and P. Stevenhagen, Math. Sci. Res. Inst. Publ. **44**, Cambridge Univ. Press, 2008. MR Zbl

[Harman 2007] G. Harman, *Prime-detecting sieves*, London Mathematical Society Monographs Series **33**, Princeton Univ. Press, 2007. MR Zbl

[Jia 1996] C. Jia, "Almost all short intervals containing prime numbers", *Acta Arith.* **76**:1 (1996), 21–84. MR Zbl

[Jordan 1875] C. Jordan, "Sur la limite du degré des groupes primitifs qui contiennent une substitution donnée", *J. Reine Angew. Math.* **79** (1875), 248–258. MR JFM

[Kantor et al. 2011] W. M. Kantor, A. Lubotzky, and A. Shalev, "Invariable generation and the Chebotarev invariant of a finite group", *J. Algebra* **348** (2011), 302–314. MR Zbl

[Kummer 1852] E. E. Kummer, "Über die Ergänzungssätze zu den allgemeinen Reciprocitätsgesetzen", *J. Reine Angew. Math.* **44** (1852), 93–146. MR Zbl

[Liebeck et al. 1987] M. W. Liebeck, C. E. Praeger, and J. Saxl, "A classification of the maximal subgroups of the finite alternating and symmetric groups", *J. Algebra* **111**:2 (1987), 365–383. MR Zbl

[Niven et al. 1991] I. Niven, H. S. Zuckerman, and H. L. Montgomery, *An introduction to the theory of numbers*, 5th ed., Wiley, New York, 1991. MR Zbl

[Oliver 1975] R. Oliver, "Fixed-point sets of group actions on finite acyclic complexes", *Comment. Math. Helv.* **50** (1975), 155–177. MR Zbl

[Rankin 1938] R. A. Rankin, "The difference between consecutive prime numbers", *J. London Math. Soc.* **S1-13**:4 (1938), 242–247. MR Zbl

[Shareshian and Woodroofe 2016] J. Shareshian and R. Woodroofe, "Order complexes of coset posets of finite groups are not contractible", *Adv. Math.* **291** (2016), 758–773. MR Zbl

[Smith 1941] P. Smith, "Fixed-point theorems for periodic transformations", *Amer. J. Math.* **63** (1941), 1–8. MR Zbl

[Thompson 1966] J. G. Thompson, "Hall subgroups of the symmetric groups", *J. Combinatorial Theory* **1** (1966), 271–279. MR Zbl

Received July 25, 2016. Revised May 12, 2017.

JOHN SHARESHIAN
DEPARTMENT OF MATHEMATICS
WASHINGTON UNIVERSITY IN ST. LOUIS
ST. LOUIS, MO
UNITED STATES
shareshi@math.wustl.edu

RUSS WOODROOFE
DEPARTMENT OF MATHEMATICS & STATISTICS
MISSISSIPPI STATE UNIVERSITY
STARKVILLE, MS
UNITED STATES

and

UNIVERSITY OF PRIMORSKA
UP FAMNIT
KOPER, SLOVENIA
russ.woodroofe@famnit.upr.si

PACIFIC JOURNAL OF MATHEMATICS
Vol. 292, No. 1, 2018
dx.doi.org/10.2140/pjm.2018.292.239

ON RATIONAL POINTS OF CERTAIN AFFINE HYPERSURFACES

ALEXANDER S. SIVATSKI

Let F be a field with char $F \neq 2$, let $a_1, \dots, a_n \in F^*$, and let $f \in F[y]$ be a monic polynomial of degree $2m$. Let further S be an affine hypersurface over F determined by the equation $f(y) = \sum_{i=1}^n a_i x_i^2$. In the first part of the paper we prove a certain version of Springer's theorem. Namely, we show that if the form $\psi \simeq \langle 1, -a_1, \dots, -a_n \rangle$ is anisotropic and S has an L-rational point for some odd-degree extension L/F, then S has an L-rational point for some odd-degree extension L/F with $[L : F] \leq m$, and the last inequality is strict in general.

In the second part we consider the case where the polynomial f is quartic. We show that the surface S has a rational point if and only if the quadratic form $\psi \perp \langle -x, g(x) \rangle$ is isotropic over $F(x)$, where $g(x) \in F[x]$ is a certain polynomial of degree at most 3, whose coefficients are expressed in a polynomial way via the coefficients of f.

In the third part we describe all Pfister forms that belong to the Witt kernel $W(F(C)/F)$, where C is the plane nonsingular curve determined by the equation $y^2 = a_4x^4 + a_2x^2 + a_1x + a_0$. In the case where the u-invariant of F is at most 10, we describe generators of the ideal $W(F(C)/F)$.

Introduction

Let F be a field of characteristic different from 2. We investigate some properties of the affine hypersurface S determined by the equation $f(y) = \sum_{i=1}^n a_i x_i^2$, where $a_i \in F^*$ and f is a monic polynomial of degree $2m$. In Section 1, we prove a version of Springer's theorem for S (Proposition 1.1). In particular, we show that if $m = 2$ (i.e., the polynomial f is quartic), and S has a K-rational point for some odd-degree extension K/F, then S has an F-rational point. Sections 2 and 3 can be considered as generalizations of some results in [Haile and Han 2007; Shick 1994]. Namely, for the affine hyperelliptic curve C with the equation $f(y) = ax^2$ over a field F, where $a \in F^* \setminus F^{*2}$ and $f(y)$ is a quartic polynomial, two questions have been investigated in [Haile and Han 2007]. First, it has been shown that existence

MSC2010: primary 11E04; secondary 11E81.
Keywords: quadratic form, Springer's theorem, Brauer group, Pfister form, field extension.

of a rational point on C is equivalent to triviality of a certain quaternion algebra over a certain quadratic extension of the rational function field $F(x)$. It is easy to see that this is equivalent to isotropicity of some 4-dimensional quadratic form over $F(x)$. In Proposition 2.1 we obtain a similar criterion for the affine hypersurface $S : f(y) = \sum_{i=1}^{n} a_i x_i^2$, where $a_i \in F^*$, the form $\langle 1, -a_1, \dots, -a_n\rangle$ is anisotropic, and f is a monic quartic polynomial. This proves independently of Section 1 that existence of a rational point over any odd-degree field extension K/F implies existence of a rational point of S over the field F itself.

Another result in [Haile and Han 2007; Shick 1994] is a computation of the relative Brauer group $\mathrm{Br}(F(C)/F)$, where C is the affine hyperelliptic curve above. Obviously, this is equivalent to description of all 2-fold Pfister forms π over F such that $\pi_{F(C)} = 0$. Section 3 is devoted to investigation of the Witt kernel $W(F(C)/F)$. Applying an invertible change of variables, we may assume that the curve C is determined by the equation $y^2 = a_4x^4 + a_2x^2 + a_1x + a_0$, where $a_i \in F$, $a_4 \neq 0$. We will also assume that C is nonsingular, for the opposite case is trivial. Let $e \in F$. Set

$$d(e) = -\det \begin{pmatrix} a_4 & 0 & \frac{1}{2}(a_2 - e) \\ 0 & e & \frac{1}{2}a_1 \\ \frac{1}{2}(a_2 - e) & \frac{1}{2}a_1 & a_0 \end{pmatrix}.$$

In Proposition 3.1 we show that if $0 \neq Q \in \mathrm{Br}(F(C)/F)$, then either $Q = (a_4, e)$, where $e \neq 0$, $d(e) \in F^{*2} \cup \{0\}$, or $a_1 = 0$ and $Q = (a_4, a_2^2 - 4a_0a_4)$. Conversely, any quaternion algebra of the types above belongs to $\mathrm{Br}(F(C)/F)$.

Proposition 3.1 is not new, but we give it for the convenience of the reader, and because we need its proof a bit later in Proposition 3.2. In fact, the original proof of Proposition 3.1, which is very similar to ours, is given in [Shick 1994]. However, in Proposition 3.2 and Corollary 3.3 we describe *all* Pfister forms π (not necessarily 2-fold) over F such that $\pi_{F(C)} = 0$. More precisely, if $\pi_{F(C)} = 0$, then either π is divisible by a 2-fold Pfister form ρ such that $\rho_{F(C)} = 0$, or there exist $e, r \in F$, $e \neq 0$, $r^2 - d(e) \neq 0$ such that $\langle\langle a_4, e, r^2 - d(e)\rangle\rangle \subset \pi$. Conversely, $\langle\langle a_4, e, r^2 - d(e)\rangle\rangle \in W(F(C)/F)$ for any $e, r \in F$, $e \neq 0$, $r^2 - d(e) \neq 0$. If the u-invariant of F is at most 10, this is sufficient for the computation of the Witt kernel $W(F(C)/F)$.

A few words about the notation. Throughout all the fields have characteristic different from 2. By a form we always mean a quadratic form over a field. For $a_1, \dots a_n \in F^*$ we denote the Pfister form $\langle 1, -a_1\rangle \otimes \cdots \otimes \langle 1, -a_n\rangle$ as $\langle\langle a_1, \dots, a_n\rangle\rangle$ (take notice of signs!), and $D(\varphi)$ is the set of all nonzero values of the form φ. If the form φ is considered as an element of the Witt ring $W(F)$, then $\dim \varphi$ denotes the dimension of the anisotropic part of φ.

If φ is a regular form over the field F, $\dim \varphi \geq 3$, then by $F(\varphi)$ we denote the function field of the corresponding projective quadric.

Slightly abusing notation, we often identify a form with its symmetric matrix.

1. A version of Springer's theorem

The well-known Springer's theorem claims that if K/F is an odd-degree field extension, and a projective quadric X has a rational point over K, then it has a rational point over F. Below we give an affine version of this theorem for certain hypersurfaces.

Proposition 1.1. *Let F be a field, let $a_1, \dots, a_n \in F^*$, and let $f \in F[y]$ be a monic polynomial of degree $2m$. Let $S = S(f, a_1, \dots, a_n)$ be the affine hypersurface over F determined by the equation $f(y) = \sum_{i=1}^n a_i x_i^2$. Suppose that S has a K-rational point for some odd-degree extension K/F.*

(1) *If the form $\langle a_1, \dots, a_n\rangle$ is anisotropic, then S has an L-rational point for some odd-degree extension L/F with $[L : F] \le 2m - 1$.*

(2) *If the form $\langle 1, -a_1, \dots, -a_n\rangle$ is anisotropic, then S has an L-rational point for some odd-degree extension L/F with $[L : F] \le m$, and the last inequality is strict in general. In particular, if $m = 2$, i.e., f is a quartic polynomial, then S has an F-rational point.*

(3) *If the form $\langle a_1, \dots, a_n\rangle$ is isotropic, then S has an F-rational point.*

Proof. (1)–(2) Assume the form $\langle a_1, \dots, a_n\rangle$ is anisotropic, and K/F is an odd-degree field extension. Suppose $[K : F] \ge s$, where $s = m + 1$ if the form $\langle 1, -a_1, \dots, -a_n\rangle$ is anisotropic, and $s = 2m+1$ otherwise. Let $f(\alpha) = \sum_{i=1}^n a_i \beta_i^2$ for some $\alpha, \beta_i \in K$. It suffices to find an odd-degree field extension L/F with $[L : F] < [K : F]$ such that S has a rational L-point. Since $[K : F(\alpha)]$ is odd, we get by Springer's theorem, applied to the extension $K/F(\alpha)$, that the form $\langle a_1, \dots, a_n, -f(\alpha)\rangle$ is isotropic over $F(\alpha)$. Hence we may assume that $\beta_i \in F(\alpha)$ for each i. We may assume also that $[F(\alpha) : F] \ge s$, for otherwise there is nothing to be proved. Let g be the minimal polynomial of α. In particular, $\deg g = [F(\alpha) : F] \ge s$. Let $\beta_i = p_i(\alpha)$, where $p_i \in F[x]$, $\deg p_i \le \deg g - 1$. Also $\deg f = 2m \le 2(s - 1) \le 2(\deg g - 1)$. We have

$$\sum_{i=1}^n a_i p_i^2 - f = gh \quad \text{for some } h \in F[x], \quad \text{and} \quad \deg\Bigl(\sum_{i=1}^n a_i p_i^2 - f\Bigr) \le 2(\deg g - 1).$$

If $\deg\bigl(\sum_{i=1}^n a_i p_i^2 - f\bigr)$ is even, then $\deg h$ is odd, and

$$\deg h \le 2(\deg g - 1) - \deg g = \deg g - 2 = [F(\alpha) : F] - 2 \le [K : F] - 2.$$

Hence S has an L-rational point, where $L = F[x]/p(x)$, and p is an arbitrary odd-degree prime divisor of h. Moreover, $[L : F] < [K : F]$.

If $\deg\bigl(\sum_{i=1}^n a_i p_i^2 - f\bigr)$ is odd, or $h = 0$, then, since f is monic of even degree, the form $\langle 1, -a_1, \dots, -a_n\rangle$ is isotropic. Hence $s = 2m + 1$, and $\deg\bigl(\sum_{i=1}^n a_i p_i^2\bigr) = \deg f = 2m$. Therefore, in this case $h = 0$, and so S has an F-rational point.

Now let us show that in the inequality $[L : F] \leq m$ in the second part of Proposition 1.1, the number m cannot be replaced by a smaller number, provided we consider all fields F and all odd-degree extensions K/F. Consider two cases:

Case (a): m is odd. Let F be a field such that there exists an irreducible polynomial p of degree m over F. Consider the equation

$$p(y)^2 = \sum_{i=1}^{n} a_i x_i^2.$$

Clearly, it has a solution over the field $K = F[y]/p(y)$ with $x_1 = \cdots = x_n = 0$. Suppose that L/F is an odd-degree extension, α, $\beta_i \in L$, and $p(\alpha)^2 = \sum_{i=1}^{n} a_i \beta_i^2$. Since the form $\langle 1, -a_1, \ldots, -a_n\rangle$ is anisotropic, we get by Springer's theorem applied to the odd-degree extension K/F that $p(\alpha) = \beta_1 = \cdots = \beta_n = 0$. Hence $m = \deg p = [F(\alpha) : F] \leq [L : F]$.

Case (b): m is even. Let k be a field, let $F = k((t))$ be the Laurent series field, and let the hypersurface S be determined by the equation $(y^{m-1}+t)(y^{m+1}+t) = \sum_{i=1}^{n} a_i x_i^2$. Let L/F be an odd-degree extension, $[L : F] \leq m - 3$. Obviously, the field L is complete with respect to a discrete valuation v such that $1 \leq v(t) \leq m-3$. It is easy to show that $(\alpha^{m-1} + t)(\alpha^{m+1} + t) \in L^{*2}$ for any $\alpha \in L$. Therefore, by Springer's theorem

$$(\alpha^{m-1} + t)(\alpha^{m+1} + t) \neq \sum_{i=1}^{n} a_i \beta_i^2 \quad \text{for any } \beta_i \in L.$$

(3) This is obvious, since any element of F is a value of the form $\langle a_1, \ldots, a_n\rangle$. $\square$

Remark 1.2. The hypothesis that the form $\langle 1, -a_1, \ldots, -a_n\rangle$ is anisotropic is essential in the second part of Proposition 1.1, at least for $m = 2$. Indeed, consider the equation $y^4 + 2 = x^2$ over $\mathbb{Q}$. Let $L = F(\delta)$, where δ is a root of the irreducible polynomial $p(u) = 2u^3 - u^2 + 2$. Obviously, $x = \delta^2 - \delta,\ y = \delta$ is a solution of the equation in question over L.

Let us prove now that this equation has no solution over $\mathbb{Q}$. It suffices to show that if $x, y, z \in \mathbb{Z}$, and $y^4+2z^4 = x^2$, then $z = 0$. Assume the contrary, so we may suppose that $y^4 + 2z^4 = x^2,\ z > 0$ and z is as small as possible. In particular, y and z are coprime; hence y is odd. Over $\mathbb{Q}(\sqrt{-2})$ we have $(y^2+z^2\sqrt{-2})(y^2-z^2\sqrt{-2}) = x^2$, and it is easy to see that the numbers $y^2 + z^2\sqrt{-2}$ and $y^2 - z^2\sqrt{-2}$ are coprime in the Euclidean ring $\mathbb{Z}[\sqrt{-2}]$. Since the group of units of the ring $\mathbb{Z}[\sqrt{-2}]$ consists of 1 and -1, we get that $y^2 + z^2\sqrt{-2} = \pm(u + v\sqrt{-2})^2$ for some $u, v \in \mathbb{Z},\ v > 0$. If $y^2 + z^2\sqrt{-2} = -(u + v\sqrt{-2})^2$, then $y^2 = 2v^2 - u^2,\ z^2 = -2uv$. The equality $y^2 = 2v^2 - u^2$ implies that u and v are odd. But then, clearly, the equality $z^2 = -2uv$ is impossible.

Thus $y^2 + z^2\sqrt{-2} = (u + v\sqrt{-2})^2$, which means that $y^2 = u^2 - 2v^2,\ z^2 = 2uv$. In particular, u is odd. Since $(u - y)(u + y) = 2v^2$, and the numbers $\frac{1}{2}(u - y)$,

$\frac{1}{2}(u+y)$ are, obviously, coprime, we may assume, changing if needed the sign of y, that $\frac{1}{2}(u-y)=t^2$, $\frac{1}{2}(u+y)=2s^2$ for some coprime $s, t>0$. Therefore, we have

$$\begin{cases} u=2s^2+t^2, \\ y=2s^2-t^2, \\ v=2st; \end{cases}$$

hence $z^2=2uv=4st(2s^2+t^2)$, and so $s=\alpha^2$, $t=\beta^2$, $2s^2+t^2=\gamma^2$, which implies $\beta^4+2\alpha^4=\gamma^2$ for some positive integers α, β, γ. Moreover, obviously,

$$0<\alpha=\sqrt{s}<\sqrt{v}<z,$$

a contradiction to the minimality of z.

In fact, there are similar counterexamples for any characteristic. Namely, let k be a field, t indeterminate, and $F=k(t)$. By an argument similar to the one for the equation $y^4+2=x^2$ over $\mathbb{Q}$, one can easily show that the equation $y^4-t=x^2$ has no solution in F. On the other hand, $x=\alpha^2-\alpha$, $y=\alpha$ is a solution of the same equation over the field $F(\alpha)$, where α is a root of the polynomial $p(u)=2u^3-u^2-t$.

However, we do not know if there exists a counterexample for each finite field, and for each number field.

Proposition 1.3. *Let F be a field, $a_1, \dots, a_n \in F^*$, and the form $\langle 1, -a_1, \dots, -a_n\rangle$ be isotropic. Let further $f \in F[y]$ be a monic polynomial of degree $2m$, where m is not divisible by* char F. *Then the hypersurface $S=S(f, a_1, \dots, a_n)$ has an L-rational point for some odd-degree field extension L/F with $[L:F] \le 2m-1$.*

Proof. Since $1 \in D(\langle a_1, \dots, a_n\rangle)$, we may assume that $n=1$ and $a_1=1$. Replacing if needed y by $y+c$, where $c \in F^*$, we may assume that the coefficient a at y^{2m-1} of the polynomial $f(y)$ is nonzero. Then setting $x=z+y^m$, one can see that the equation $f(y)=x^2$ is equivalent to the equation $ay^{2m-1}+\sum_{i=0}^{2m-2} p_i(z)y^i=0$, where $p_i(z) \in F[z]$. It is clear that the last equation has a required point. □

Remark 1.4. We do not know whether Proposition 1.3 remains valid if m is divisible by char F.

Another natural question is whether the inequality $[L:F] \le 2m-1$ in the first part of Proposition 1.1 is strict for each m. In view of Remark 1.2 it is strict for $m=2$.

2. A criterion for existence of rational points for certain affine hypersurfaces

We give a criterion in the language of quadratic forms for the existence of a rational point for the hypersurface S in the case where $m=2$ (the polynomial f is quartic) and the form $\langle 1, -a_1, \dots, -a_n\rangle$ is anisotropic. The main ingredient in the sequel is the strong form of the Cassels–Pfister theorem [Pfister 1995, Chapter 1, Generalization 2.3 of Theorem 2.2], which reads as follows:

Theorem. *Let $\varphi(x_1, \dots, x_n) = \sum_{1 \le i,j \le n} l_{ij}(t) x_i x_j$ be an anisotropic form over $F(t)$, where $l_{ij}(t) \in F[t]$, and $\deg l_{ij}(t) \le 1$. Suppose $f \in F[t] \cap D(\varphi)$. Then there exist polynomials $p_i \in F[t]$ such that $f = \varphi(p_1, \dots, p_n)$.*

In the following statement, using the theorem above, we get a criterion for existence of rational points for the hypersurface S in the case of a quartic polynomial f.

Proposition 2.1. *Let F be a field, $a_1, \dots, a_n \in F^*$, and $u_1, u_2, u_3 \in F$. Suppose that the form $\psi \simeq \langle 1, -a_1, \dots, -a_n \rangle$ is anisotropic. Then the following two conditions are equivalent*:

(1) $-u_3^2 x^3 + u_2 x^2 + u_1 x + 1 \in D(\psi \perp \langle -x \rangle)$, *i.e., the form*

$$\psi \perp \langle -x, u_3^2 x^3 - u_2 x^2 - u_1 x - 1 \rangle$$

is isotropic over $F(x)$.

(2) *The affine hypersurface S determined by the equation*

$$y^4 + 2u_1 y^2 - 8u_3 y + u_1^2 - 4u_2 = \sum_{i=1}^{n} a_i x_i^2$$

has a rational point.

Moreover, *if*, *in contrast the form ψ is isotropic*, *and $u_3 \neq 0$*, *then both conditions necessarily hold. If the form ψ is isotropic*, *and $u_3 = 0$*, *then condition* (1) *necessarily holds*, *but in general condition* (2) *does not.*

Proof. (1) $\Longrightarrow$ (2): Obviously, the form $\psi \perp \langle -x \rangle$ is anisotropic. By the strong form of the Cassels–Pfister theorem

$$-u_3^2 x^3 + u_2 x^2 + u_1 x + 1 \in D(\psi \perp \langle -x \rangle)$$

if and only if

$$-u_3^2 x^3 + u_2 x^2 + u_1 x + 1 = p_0^2 - a_1 p_1^2 - \cdots - a_n p_n^2 - x p_{n+1}^2$$

for some $p_i \in F[x]$. Since the form ψ is anisotropic, we get $p_i(x) = \alpha_i x + \beta_i$ for each i, where $\alpha_i, \beta_i \in F$. Moreover, $\alpha_{n+1}^2 = u_3^2$; hence we may assume that $\alpha_{n+1} = u_3$. Therefore, α_i, β_i satisfy the equations

$$(*) \qquad \begin{cases} \alpha_0^2 - a_1 \alpha_1^2 - \cdots - a_n \alpha_n^2 - 2u_3 \beta_{n+1} = u_2, \\ 2\alpha_0 \beta_0 - 2a_1 \alpha_1 \beta_1 - \cdots - 2a_n \alpha_n \beta_n - \beta_{n+1}^2 = u_1, \\ \beta_0^2 - a_1 \beta_1^2 - \cdots - a_n \beta_n^2 = 1. \end{cases}$$

Let $\boldsymbol{u} = (\alpha_0, \alpha_1, \dots, \alpha_n)$ and $\boldsymbol{v} = (\beta_0, \beta_1, \dots, \beta_n)$. Obviously, the system $(*)$ is equivalent to the system

$$(**) \qquad \begin{cases} \psi(\boldsymbol{u}) = u_2 + 2u_3 \beta_{n+1}, \\ \psi(\boldsymbol{u}, \boldsymbol{v}) = \frac{1}{2}(u_1 + \beta_{n+1}^2), \\ \psi(\boldsymbol{v}) = 1. \end{cases}$$

If the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent, then the system (∗∗) implies

$$\det\begin{pmatrix} u_2+2\alpha_{n+1}\beta_{n+1} & \frac{1}{2}(u_1+\beta_{n+1}^2) \\ \frac{1}{2}(u_1+\beta_{n+1}^2) & 1 \end{pmatrix} = u_2+2u_3\beta_{n+1}-\tfrac{1}{4}(u_1+\beta_{n+1}^2)^2=0.$$

Hence S has a rational point $x_i=0,\ y=\beta_{n+1}$.

If the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly independent, then the 2-dimensional form τ with the matrix

$$\begin{pmatrix} u_2+2u_3\beta_{n+1} & \frac{1}{2}(u_1+\beta_{n+1}^2) \\ \frac{1}{2}(u_1+\beta_{n+1}^2) & 1 \end{pmatrix}$$

is a subform of ψ with the underlying linear space generated by the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. Obviously,

$$\tau \simeq \bigl\langle 1,\, u_2+2\alpha_{n+1}\beta_{n+1}-\tfrac{1}{4}(u_1+\beta_{n+1}^2)^2\bigr\rangle.$$

Therefore,

$$-u_2-2u_3\beta_{n+1}+\tfrac{1}{4}(u_1+\beta_{n+1}^2)^2 \in D(\langle a_1,\dots,a_n\rangle),$$

which means that $(u_1+\beta_{n+1}^2)^2-8u_3\beta_{n+1}-4u_2=\sum_{i=1}^n a_i x_i^2$ for some $x_i\in F$, and we are done.

(2) $\Longrightarrow$ (1): Assume that S has a rational point, say, $y=\beta_{n+1},\ x_i=c_i$. If $c_1=\dots=c_n=0$, then $u_2+2\alpha_{n+1}\beta_{n+1}-\frac{1}{4}(u_1+\beta_{n+1}^2)^2=0$. Put

$$\begin{cases} \alpha_1=\dots=\alpha_n=0, \\ \alpha_0=\frac{1}{2}(u_1+\beta_{n+1}^2), \\ \beta_0=1, \\ \beta_1=\dots=\beta_n=0. \end{cases}$$

Since the elements α_i, β_i satisfy the system (∗), we get $-u_3^2x^3+u_2x^2+u_1x+1\in D(\psi\perp\langle -x\rangle)$.

If at least one of c_i is not zero, then, since the form $\langle a_1,\dots,a_n\rangle$ is anisotropic,

$$-\det\begin{pmatrix} u_2+2u_3\beta_{n+1} & \frac{1}{2}(u_1+\beta_{n+1}^2) \\ \frac{1}{2}(u_1+\beta_{n+1}^2) & 1 \end{pmatrix} \in D(\langle a_1,\dots,a_n\rangle),$$

or, equivalently, the form with the matrix

$$\begin{pmatrix} u_2+2u_3\beta_{n+1} & \frac{1}{2}(u_1+\beta_{n+1}^2) \\ \frac{1}{2}(u_1+\beta_{n+1}^2) & 1 \end{pmatrix}$$

is a subform of the form ψ. In other words, there are linearly independent vectors $\boldsymbol{u}=(\alpha_0,\alpha_1,\dots,\alpha_n)$, $\boldsymbol{v}=(\beta_0,\beta_1,\dots,\beta_n)$ such that the system (∗∗) holds. Hence in this case we have $-u_3^2x^3+u_2x^2+u_1x+1\in D(\psi\perp\langle -x\rangle)$ as well.

If the form ψ is isotropic, then, obviously,

$$-u_3^2x^3+u_2x^2+u_1x+1\in D(\psi)\subset D(\psi\perp\langle -x\rangle).$$

Assume $u_3\neq 0$. We may suppose $a_1=1$, and put $y=-u_2/(2u_3)$, $x_1=u_2^2/(4u_3^2)+u_1$, $x_2=\cdots=x_n=0$.

Finally, if ψ is isotropic, and $u_3=0$, then Remark 1.2 shows that in general S does not always have a rational point. □

In the following example we show how Proposition 2.1 can be applied to construct elements from $\mathrm{Br}(F(S)/F)$ in the case $n=1$.

Example 2.2. Let $n=1$ and $a\in F^*\setminus F^{*2}$. Proposition 2.1 claims that

$$-u_3^2x^3+u_2x^2+u_1x+1\in D(\langle 1,-a,-x\rangle)$$

if and only if the equation

$$az^2=y^4+2u_1y^2-8u_3y+u_1^2-4u_2$$

has a solution over F, or, equivalently, multiplying by $4a$, and setting $t=2az$, if and only if the equation

$$t^2=4ay^4+8au_1y^2-32au_3y+4a(u_1^2-4u_2)$$

has a solution over F. Let us set

$$a_4=4a,\quad a_2=8au_1,\quad a_1=-32au_3,\quad a_0=4a(u_1^2-4u_2)$$

(here the meaning of the elements a_i is different from the previous one). Hence we get that the equation $t^2=a_4y^4+a_2y^2+a_1y+a_0$ has a solution over F if and only if

$$-\left(\frac{a_1}{32a}\right)^2x^3+\frac{4a(a_2/(8a))^2-a_0}{16a}x^2+\frac{a_2}{8a}x+1\in D(\langle 1,-a,-x\rangle).$$

A straightforward computation shows that the last condition is equivalent to

$$z^3-2a_2z^2+(a_2^2-4a_0a_4)z+a_1^2a_4\in D(\langle z,a_4,-a_4z\rangle),$$

where $z=-4a_4/x$, which means that $(a_4,z)_{F(z)(\sqrt{g(z)})}=0$, where

$$g(z)=z^3-2a_2z^2+(a_2^2-4a_0a_4)z+a_1^2a_4.$$

This is the result of [Haile and Han 2007, Propositions 5 and 17], originally obtained by means of algebraic geometry and cohomology groups.

Further, if $(a_4,z)_{F(z)(\sqrt{g(z)})}=0$, by the evaluating argument we get $(a,e)=0$ if $g(e)\in F^{*2}$ and $e\neq 0$. Therefore, $(a,e)\in\mathrm{Br}(F(S)/F)$ for each $e\in F^*$ such that $g(e)\in F^{*2}$.

Note also that

$$z^3 - 2a_2z^2 + (a_2^2 - 4a_0a_4)z + a_1^2a_4 = -4\det\begin{pmatrix} a_4 & 0 & \frac{1}{2}(a_2 - z) \\ 0 & z & \frac{1}{2}a_1 \\ \frac{1}{2}(a_2 - z) & \frac{1}{2}a_1 & a_0 \end{pmatrix}.$$

Later, in Proposition 3.1 we will see why this determinant is involved here.

Example 2.3. Suppose S has the equation $(y^2 - b)^2 = \sum_{i=1}^n a_i x_i^2$, where $b \in F^*$. Then, since the form $\psi \simeq \langle 1, -a_1, \dots, -a_n \rangle$ is anisotropic, it is easy to see that the surface S has a rational point if and only if $b \in F^{*2}$. On the other hand, in this case $u_1 = -b$, $u_2 = u_3 = 0$. Hence Proposition 2.1 claims that S has a rational point if and only if the form $\langle 1, -a_1, \dots, -a_n, -x, bx - 1 \rangle$ is isotropic. By Brumer's theorem [1978] this is the case if and only if the forms $\langle 1, -a_1, \dots, -a_n, 0, -1 \rangle$ and $\langle 0, 0, \dots, 0, -1, b \rangle$ have a common nontrivial zero. It is easy to verify independently that this is equivalent to $b \in F^{*2}$.

In the algebraic theory of quadratic forms over fields, there are many results concerning splitting of forms by the function field of a quadric. In the following statements (Corollaries 2.4–2.7) we consider the similar questions for the hypersurface S from Proposition 2.1. In particular, we assume that $a_1, \dots, a_n \in F^*$, and the form $\langle 1, -a_1, \dots, -a_n \rangle$ is anisotropic.

Let $W(k)$ be the Witt group of a field k. It is well known, see, for example, [Scharlau 1985], that the sequence of abelian groups

$$0 \to W(k) \xrightarrow{\text{res}} W(k(t)) \xrightarrow{\coprod \partial_p} \coprod_{p \in \mathbb{A}_k^1} W(k_p) \to 0$$

is split exact. We consider here a point $p \in \mathbb{A}_k^1$ as a monic irreducible polynomial over k. We denote by $k_p = k[t]/p$ the corresponding residue field and by $\partial_p : W(k(t)) \to W(k_p)$ the residue homomorphism well defined by the rule

$$\partial_p(\langle f \rangle) = \begin{cases} 0 & \text{if } v_p(f) = 0, \\ \langle \overline{fp^{-1}} \rangle & \text{if } v_p(f) = 1. \end{cases}$$

There is a splitting map $W(k(t)) \to W(k)$ defined by the rule $\langle f \rangle \to \langle l(f) \rangle$, where $l(f)$ is the leading coefficient of the polynomial $f \in k[t]$.

Corollary 2.4. *In the notation of Proposition 2.1, assume that the hypersurface S has no F-rational point, $n = 1$, and φ is a 3-dimensional form over F. Then S has no $F(\varphi)$-rational point.*

Proof. Let π be the 2-fold Pfister form corresponding to φ. We may assume that $\pi \neq 0$. Suppose that S determined by the equation

$$az^2 = y^4 + 2u_1y^2 - 8u_3y + u_1^2 - 4u_2$$

has an $F(\varphi)$-rational point. In view of Example 2.2 we have $\langle\langle a, z\rangle\rangle_{F(\pi)(\sqrt{g(z)})} = 0$, where $g(z) = z^3 - 2a_2z^2 + (a_2^2 - 4a_0a_4)z + a_1^2a_4$. Then, since S has no rational point, i.e., $\langle\langle a, z\rangle\rangle_{F(\sqrt{g(z)})} \neq 0$, we get $\langle\langle a, z\rangle\rangle = \langle\langle g(z)\rangle\rangle\tau + \pi$ for some $\tau \in W(F(z))$. Therefore,

$$0 = l(\langle\langle a, z\rangle\rangle) - l(\langle\langle g(z)\rangle\rangle\tau) = l(\pi) = \pi,$$

a contradiction. □

Corollary 2.5. *Assume S has no F-rational point, $n = 2$, and φ is a 4-dimensional anisotropic form over F, $\operatorname{disc}\varphi = d \neq 1$. The following conditions are equivalent:*

(1) *S has an $F(\varphi)$-rational point.*

(2) *$u_3^2x^3 - u_2x^2 - u_1x - 1 = h(x)q(x)$, where $h, q \in F[x]$, $\deg h \leq 1$, $\deg q = 2$, q is monic and irreducible, $-\overline{a_1a_2x} \in F_q^{*2}$, $\operatorname{disc} q = d$, and φ is similar to the form $\langle 1, -a_1, -a_2, a_1a_2d\rangle$.*

Proof. (2) $\Longrightarrow$ (1): Consider first the case $u_3 \neq 0$. Since $-\overline{a_1a_2x} \in F_q^{*2}$, we have $N_{F_q/F}(\bar{x}) \in F^{*2}$; hence $q(x) = x^2 + cx + b^2$ for some $c, b \in F$, $b \neq 0$. Therefore, $h(x) = u_3^2x - b^{-2}$, in particular, $h \in D(\langle -1, x\rangle)$. Hence $u_3^2x^3 - u_2x^2 - u_1x - 1 = h(x)q(x) \in D(\langle -q, qx\rangle)$. It follows that

(2-1) $\langle 1, -a_1, -a_2, -x, hq\rangle \subset \langle 1, -a_1, -a_2, -x, -q, qx\rangle \subset \langle 1, -a_1, -a_2, -x\rangle\langle\langle q\rangle\rangle$.

On the other hand,

(2-2) $\langle 1, -a_1, -a_2, -x\rangle\langle\langle q\rangle\rangle = \langle\langle a_1, a_2, q\rangle\rangle + \langle -a_1a_2, -x\rangle\langle\langle q\rangle\rangle = \langle\langle a_1, a_2, q\rangle\rangle$,

as $-\overline{a_1a_2x} \in F_q^{*2}$.

Finally, $\varphi\langle\langle q\rangle\rangle \sim \langle 1, -a_1, -a_2, a_1a_2d\rangle\langle\langle q\rangle\rangle = \langle\langle a_1, a_2, q\rangle\rangle$, since $\langle\langle d, q\rangle\rangle = 0$. We conclude that $\langle\langle a_1, a_2, q\rangle\rangle_{F(x)(\varphi)} = 0$. In view of (2-1) and (2-2), the form

$$\langle 1, -a_1, -a_2, -x, hq\rangle_{F(x)(\varphi)} = \langle 1, -a_1, -a_2, -x, u_3^2x^3 - u_2x^2 - u_1x - 1\rangle_{F(x)(\varphi)}$$

is isotropic, which implies by Proposition 2.1 that S has an $F(\varphi)$-rational point.

The case $u_3 = 0$ is similar. In this case

$$-u_2x^2 - u_1x - 1 = -u_2q = -u_2(x^2 + cx + b^2);$$

hence $u_2 \in F^{*2}$, and obviously,

$$\langle 1, -a_1, -a_2, -x, -u_2q\rangle \subset \langle 1, -a_1, -a_2, -x\rangle\langle\langle q\rangle\rangle.$$

Now we can finish the proof as in the case $u_3 \neq 0$.

(1) $\Longrightarrow$ (2): Assume that S has an $F(\varphi)$-rational point. Then by Proposition 2.1 the form $\Phi \simeq \langle 1, -a_1, -a_2, -x, u_3^2x^3 - u_2x^2 - u_1x - 1\rangle$ is anisotropic over $F(x)$, but isotropic over $F(x)(\varphi)$. Consider two possible cases:

Case (a): $\operatorname{ind}(\Phi) = 4$. Then by [Hoffmann 1995] there exists a squarefree $p \in F[x]$ such that $p\varphi \subset \Phi$. Comparing the determinants we get

$$\Phi \simeq p\varphi \perp \langle -a_1a_2 \operatorname{disc}(\varphi)x(u_3^2x^3 - u_2x^2 - u_1x - 1)\rangle. \tag{2-3}$$

Note that p is not divisible by x, for otherwise (2-3) would imply $\dim \partial_x(\Phi) \geq 3$, a contradiction. Comparing the residues at x of the left-hand and the right-hand parts of (2-3), we get $a_1a_2 \operatorname{disc}(\varphi) = -1$; hence

$$\Phi \simeq p\varphi \perp \langle x(u_3^2x^3 - u_2x^2 - u_1x - 1)\rangle. \tag{2-4}$$

Applying the "leading coefficient" homomorphism $l : W(F(x)) \to W(F)$ to both sides of (2-4), we get

$$\langle 1, -a_1, -a_2, -1, 1\rangle \simeq l(p)\varphi \perp \langle 1\rangle$$

if $u_3 \neq 0$, or

$$\langle 1, -a_1, -a_2, -1, -u_2\rangle \simeq l(p)\varphi \perp \langle -u_2\rangle$$

if $u_3 = 0$ (if $u_3 = 0$, then it easily follows that $u_2 \neq 0$). Hence in any case $l(p)\varphi \simeq \langle 1, -a_1, -a_2, -1\rangle$, so φ is isotropic, a contradiction.

Case (b): $\operatorname{ind}(\Phi) = 2$. Then Φ is a Pfister neighbor of some anisotropic 3-fold Pfister form π over $F(x)$, say, $\pi \simeq \Phi \perp \sigma$. Since $\pi_{F(x)(\varphi)}$ is isotropic (or, equivalently, hyperbolic), $\pi \simeq \langle\langle a_1, a_2, P\rangle\rangle$ for some squarefree $P \in F[x]$. We claim that P does not have any odd-degree irreducible divisor p. Indeed, otherwise, taking into account that $\pi \simeq \varphi\langle\langle h(x)\rangle\rangle$ for some $h(x) \in F[x]$ [Wadsworth 1975], we get that $\langle\langle a_1, a_2\rangle\rangle_{F_p} = \partial_p(\langle\langle a_1, a_2, P\rangle\rangle)$ either equals 0 or is similar to φ_{F_p}. But since $\langle 1, -a_1, -a_2\rangle$ is anisotropic, $\operatorname{disc}(\varphi) \neq 1$, and $\deg p$ is odd, both cases are impossible.

Furthermore, if $s \neq x$ is a monic irreducible divisor of P, which is not a divisor of $u_3^2x^3 - u_2x^2 - u_1x - 1$, then

$$\langle\langle a_1, a_2\rangle\rangle \sim \partial_s(\pi) = \partial_s(\Phi) + \partial_s(\sigma) = \partial_s(\sigma).$$

Since $\dim \partial_s(\sigma) \leq 3$, we get $\langle\langle a_1, a_2\rangle\rangle_{F_s} = 0$; hence $\langle\langle a_1, a_2, s\rangle\rangle = 0$, and so we can replace P by P/s.

Thus, we may assume that P divides $u_3^2x^3 - u_2x^2 - u_1x - 1$, and P is an irreducible quadratic polynomial. Therefore, $u_3^2x^3 - u_2x^2 - u_1x - 1 = hq$, where $h, q \in F[x]$, $\deg h \leq 1$ ($\deg h = 0$ if and only if $u_3 = 0$), $\deg q = 2$, and q is monic irreducible. Obviously, $P = \lambda q$ for some $\lambda \in F^*$. We have $\dim l(\Phi) \leq 3$; hence $\dim l(\pi) \leq 3 + \dim l(\sigma) \leq 6$, which implies that $\dim l(\pi) = 0$. Therefore, we can replace P by q, so $\pi \simeq \langle\langle a_1, a_2, q\rangle\rangle$. In particular, $\langle\langle a_1, a_2\rangle\rangle_{F_q} \neq 0$. Since $\langle 1, -a_1, -a_2, -x\rangle$ is a subform of π, we have $\langle 1, -a_1, -a_2, -x\rangle\langle\langle R\rangle\rangle \simeq \langle\langle a_1, a_2, q\rangle\rangle$

for some squarefree $R \in F[x]$ [Wadsworth 1975]. In other words,

$$\begin{cases} \langle\langle a_1, a_2, q\rangle\rangle = \langle\langle a_1, a_2, R\rangle\rangle, \\ \langle\langle -a_1 a_2 x, R\rangle\rangle = 0. \end{cases} \tag{2-5}$$

From the first equality of (2-5) we get that q divides R, since $\partial_q(\langle\langle a_1, a_2, q\rangle\rangle) = \langle\langle a_1, a_2\rangle\rangle_{F_q} \neq 0$. Therefore,

$$\bar{1} = \partial_q(\langle\langle -a_1 a_2 x, R\rangle\rangle) = -\overline{a_1 a_2 x} \in F_q^*/{F_q}^{*2}.$$

Hence $N_{F_q/F}(x) \in F^{*2}$, and $q(x) = x^2 + cx + b^2$ for some $c, b \in F$, $b \neq 0$. Further, since $\langle\langle a_1, a_2, q\rangle\rangle_{F(x)(\varphi)} = 0$, we have $\langle\langle a_1, a_2, q\rangle\rangle \sim \varphi\langle\langle T\rangle\rangle$ for some $T \in F[x]$. This implies that q divides T, and $\varphi_{F_q} \sim \langle\langle a_1, a_2\rangle\rangle_{F_q}$, i.e., $\varphi_{F(\sqrt{\operatorname{disc} q})} \sim \langle\langle a_1, a_2\rangle\rangle_{F(\sqrt{\operatorname{disc} q})}$. Therefore, $\operatorname{disc}\varphi = \operatorname{disc} q = d$. Finally, by [Wadsworth 1975] we get that $\varphi \sim \langle 1, -a_1, -a_2, a_1 a_2 d\rangle$. The verification of the implication (1) $\Longrightarrow$ (2) is done. □

Corollary 2.6. *Assume S has no F-rational point, $n = 2$, and φ is a 5-dimensional anisotropic form over F. Then S has no $F(\varphi)$-rational point.*

Proof. Let $\sigma \subset \varphi$ be a 4-dimensional subform of $\varphi_{F(t)}$, which does not satisfy condition (2) in Corollary 2.5 (with replacement of the ground field F by $F(t)$). Then S has no $F(t)(\sigma)$-rational points; hence S has no $F(t)(\varphi)$-rational points. □

Recall that u-invariant of the field k is the maximum of dimensions of anisotropic forms over k.

Corollary 2.7. *In the notation of Proposition 2.1, assume that the hypersurface S has no F-rational point*:

1. *If $n = 1$, then there exists a field extension L/F such that S_L has no rational point, L does not have an odd-degree field extension, and $u(L) = 2$. In particular*, $\operatorname{cd}_2 L = 1$.
2. *If $n = 2$, then there exists a field extension L/F such that S_L has no rational point, L does not have an odd-degree field extension, and $u(L) = 4$. In particular*, $\operatorname{cd}_2 L = 2$.

Proof. (1) By Proposition 1.1 and Corollary 2.4 the field L can be constructed by subsequent splitting of all 2-fold Pfister forms and passing to a maximal odd-degree extension; see, for instance, [Elman et al. 2008, Theorem 38.4]. Clearly, $u(L) = 2$.

(2) Similar to (1), the field L can be constructed by subsequent splitting of all 5-dimensional forms and passing to a maximal odd-degree extension. □

Corollary 2.8. *In the notation of Proposition 2.1, the following conditions are equivalent*:

1. *The polynomial $f(y) = y^4 + 2u_1 y^2 - 8u_3 y + u_1^2 - 4u_2$ has a root in F.*
2. *Let $p(x)$ be any monic polynomial divisor of $g(x) = -u_3^2 x^3 + u_2 x^2 + u_1 x + 1$ such that $v_p(-u_3^2 x^3 + u_2 x^2 + u_1 x + 1)$ is odd. Then $\bar{x}$ is a square in the field F_p.*

Proof. (1) $\Longrightarrow$ (2): Since the polynomial $f(y)$ has a root α in F, the affine curve $f(y) = tx^2$ has a rational point, namely $(0, \alpha)$, over the Laurent series field $F((t))$. Hence by Proposition 2.1 the form $\langle 1, -t, -x, -g(x)\rangle$ is isotropic, which implies that the form $\langle 1, -x, -g(x)\rangle$ is isotropic as well. This means that the Pfister form $\langle\langle x, g(x)\rangle\rangle$ is trivial. Then $\bar{x}$ is a square in the field F_p.

(2) $\Longrightarrow$ (1): In view of the exact sequence for $W(F(x))$ the Pfister form $\langle\langle x, g(x)\rangle\rangle$ is trivial; hence the form $\langle 1, -x, -g(x)\rangle$ is isotropic. By Proposition 2.1 the affine curve $f(y) = tx^2$ has a rational point over $F((t))$, say (x_0, y_0). Suppose f has no root in F. Let v be the discrete F-valuation on $F((t))$ such that $v(t) = 1$. Obviously, $v(tx_0^2)$ is odd, but $v(y_0^4 + 2u_1y_0^2 - 8u_3y_0 + u_1^2 - 4u_2)$ is even, a contradiction. $\square$

3. On the Witt kernel $W(F(C)/F)$ for the plane curve C with the equation $y^2 = a_4x^4 + a_2x^2 + a_1x + a_0$

If C is a nonsingular algebraic curve over the field F with a rational point $p \in C$ and the function field $F(C)$, then the composition of the restriction map $W(F) \to W(F(C))$ and the *first* residue map $\partial_p : W(F(C)) \to W(F)$ is the identity; hence $W(F(C)/F) = 0$. More generally, applying Springer's theorem, it is easy to see that $W(F(C)/F) = 0$ if C has a point of odd degree. In the opposite case the computation of $W(F(C)/F)$ can hardly be done in general. In this section we describe all Pfister forms from the ideal $W(F(C)/F)$, with C being the affine plane curve determined by the equation $y^2 = f(x)$, where $f(x) = a_4x^4 + a_2x^2 + a_1x + a_0 \in F[x]$ is a squarefree quartic polynomial, $a_4 \neq 0$. Obviously, the last equation is equivalent to the equation $y^4 + 2u_1y^2 - 8u_3y + u_1^2 - 4u_2 = ax^2$, $a \neq 0$, under an invertible change of the coefficients. As a consequence, we compute $W(F(C)/F)$ if the u-invariant of the field F is at most 10.

The description of 2-fold Pfister forms in $W(F(C)/F)$, or, equivalently quaternion algebras in $\mathrm{Br}(F(C)/F)$, was made in [Shick 1994; Haile and Han 2007] correspondingly. The proof of Proposition 3.1 below, is, in fact, very similar to that in [Shick 1994, Theorem 9], but we give it here for the sake of completeness, and because we need it in Proposition 3.2.

Let $e \in F$. Set

$$M = \begin{pmatrix} a_4 & 0 & \frac{1}{2}(a_2 - e) \\ 0 & e & \frac{1}{2}a_1 \\ \frac{1}{2}(a_2 - e) & \frac{1}{2}a_1 & a_0 \end{pmatrix},$$

and $d(e) = -\det(M)$.

Proposition 3.1. *Assume that* $0 \neq Q \in \mathrm{Br}(F(C)/F)$. *Then either* $Q = (a_4, e)$, *where* $e \neq 0$, $d(e) \in F^{*2} \cup \{0\}$, *or* $a_1 = 0$ *and* $Q = (a_4, a_2^2 - 4a_0a_4)$. *Conversely, any quaternion algebra of the types above belongs to* $\mathrm{Br}(F(C)/F)$.

Proof. Let $0 \neq Q \in \mathrm{Br}(F(C)/F)$, let π be the 2-fold Pfister form corresponding to Q, and let $-\varphi$ be the pure subform of π, i.e., $\pi \simeq \langle 1 \rangle \perp -\varphi$. Let V be the underlying vector space of φ. Assume that $Q_{F(C)} = 0$. Since φ is anisotropic, by the Cassels–Pfister theorem there exist $v_0, v_1, v_2 \in V$ such that $\varphi(x^2 v_2 + x v_1 + v_0) = f(x)$. Comparing the coefficients on the left-hand and the right-hand sides of the last equality, we get the system

$$(\star) \qquad \begin{cases} \varphi(v_2, v_2) = a_4, \\ \varphi(v_1, v_2) = 0, \\ \varphi(v_1, v_1) + 2\varphi(v_0, v_2) = a_2, \\ \varphi(v_0, v_1) = \frac{1}{2} a_1, \\ \varphi(v_0, v_0) = a_0. \end{cases}$$

If $d(e) \neq 0$, then M is the matrix of φ with respect to the basis (v_2, v_1, v_0), and so $d(e) \in F^{*2}$.

If $e \neq 0$, then $\langle a_4, e \rangle$ is a regular subform of the form $\varphi|_{\langle v_0, v_1, v_2 \rangle}$. Since $\det \varphi = -1$, we get $\varphi \simeq \langle a_4, e, -a_4 e \rangle$, which implies $\pi \simeq \langle\langle a_4, e \rangle\rangle$ and $Q = (a_4, e)$. If $e = 0$, then

$$M = \begin{pmatrix} a_4 & 0 & \frac{1}{2} a_2 \\ 0 & 0 & \frac{1}{2} a_1 \\ \frac{1}{2} a_2 & \frac{1}{2} a_1 & a_0 \end{pmatrix}.$$

If additionally $a_1 \neq 0$, then

$$\mathbb{H} = \begin{pmatrix} 0 & \frac{1}{2} a_1 \\ \frac{1}{2} a_1 & a_0 \end{pmatrix}$$

is a regular subform of φ; hence $Q = 0$, a contradiction. If $e = a_1 = 0$, then, since f is squarefree, $a_2^2 - 4a_0 a_4 \neq 0$. Hence

$$\begin{pmatrix} a_4 & \frac{1}{2} a_2 \\ \frac{1}{2} a_2 & a_0 \end{pmatrix}$$

is a regular subform of φ, so $\varphi \simeq \langle a_4, -a_4(a_2^2 - 4a_0 a_4), a_2^2 - 4a_0 a_4 \rangle$, $\pi \simeq \langle\langle a_4, a_2^2 - 4a_0 a_4 \rangle\rangle$, and $Q = (a_4, a_2^2 - 4a_0 a_4)$.

Conversely, assume that $d(e) \in F^{*2}$, $e \neq 0$. Consider the form φ with the matrix

$$M = \begin{pmatrix} a_4 & 0 & \frac{1}{2}(a_2 - e) \\ 0 & e & \frac{1}{2} a_1 \\ \frac{1}{2}(a_2 - e) & \frac{1}{2} a_1 & a_0 \end{pmatrix}$$

with respect to a certain basis v_2, v_1, v_0. Then $\varphi \simeq \langle a_4, e, -a_4 e \rangle$. Hence system $(\star)$ implies

$$f = \varphi(x^2 v_2 + x v_1 + v_0) \in D(\varphi) = D(\langle a_4, e, -a_4 e \rangle),$$

so $(a_4, e)_{F(C)} = 0$. Assume now that $d(e) = 0$, $e \neq 0$. Then φ is degenerate, and $\langle a_4, e\rangle$ is a regular subform of φ; hence $\varphi \simeq \langle a_4, e, 0\rangle$. Therefore, $f \in D(\langle a_4, e, 0\rangle) = D(\langle a_4, e\rangle)$, so again $(a_4, e)_{F(C)} = 0$.

Finally, if $a_1 = 0$, then

$$f(x) = a_4x^4 + a_2x^2 + a_0 = a_4\left(x^2 + \frac{a_2}{2a_4}\right)^2 + \left(a_0 - \frac{a_2^2}{4a_4}\right) \in D\left\langle a_4, a_0 - \frac{a_2^2}{4a_4}\right\rangle,$$

so $(a_4, a_2^2 - 4a_0a_4)_{F(C)} = (a_4, a_0 - a_2^2/(4a_4))_{F(C)} = 0$. □

Let $n \geq 3$. Let $P_n(f)$ be the set of n-fold Pfister forms π over F such that $\pi_{F(C)} = 0$, where f and C are as in Proposition 3.1. We say that $\pi \in P_n(f)$ is standard if $\rho \subset \pi$ for some $\rho \in P_2(f)$. Otherwise we say that $\pi \in P_n(f)$ is nonstandard.

Proposition 3.2. *Assume $n \geq 3$, $\pi \in P_n(f)$ is nonstandard, and $d(e)$ has the same meaning as in Proposition 3.1. Then there exist $e, r \in F$, $e \neq 0$, $r^2 - d(e) \neq 0$, such that $\langle\langle a_4, e, r^2 - d(e)\rangle\rangle \subset \pi$. Moreover, $\langle\langle a_4, e, r^2 - d(e)\rangle\rangle \in P_3(f)$ for any $e, r \in F$, $e \neq 0$, $r^2 - d(e) \neq 0$.*

Proof. Assume that $\pi \in P_n(f)$, or, equivalently, $f \in D(-\pi')$. Then the proof of Proposition 3.1 shows there is some $e \in F$ such that one of the following cases holds:

(1) $e \neq 0$, $d(e) \neq 0$,

$$\begin{pmatrix} a_4 & 0 & \frac{1}{2}(a_2 - e) \\ 0 & e & \frac{1}{2}a_1 \\ \frac{1}{2}(a_2 - e) & \frac{1}{2}a_1 & a_0 \end{pmatrix} \subset -\pi',$$

where π' is the pure subform of π.

(2) $e \neq 0$, $d(e) = 0$, $\langle a_4, e\rangle \subset -\pi'$.

(3) $a_1 = 0$,

$$\begin{pmatrix} a_4 & \frac{1}{2}a_2 \\ \frac{1}{2}a_2 & a_0 \end{pmatrix} \subset -\pi'.$$

In the second case, $\langle\langle a_4, e\rangle\rangle \subset \pi$ and $d(e) = 0$. In the third case $\langle\langle a_4, a_2^2 - 4a_0a_4\rangle\rangle \subset \pi$. In both cases, π is standard.

In the first case, $\langle a_4, e, -a_4ed(e)\rangle \subset -\pi'$. Set $\tau \simeq \langle 1, -a_4, -e, a_4ed(e)\rangle$. Hence $\tau \subset \pi$, which implies $\pi_{F(\tau)} = 0$. By [Fitzgerald 1983, Corollary 1.5] there is a 3-fold Pfister form ρ such that $\tau \subset \rho \subset \pi$. In particular, by [Wadsworth 1975] there is $s \in F^*$ such that $\rho \simeq \tau \otimes \langle\langle s\rangle\rangle$. Since $\rho \in I^3(F)$, we have $\langle\langle d(e), s\rangle\rangle = 0$; i.e., $\langle\langle s\rangle\rangle \simeq \langle\langle r^2 - d(e)\rangle\rangle$ for some $r \in F$. Therefore, $\rho \simeq \langle\langle a_4, e, r^2 - d(e)\rangle\rangle$.

Conversely, let $\delta \simeq \langle\langle a_4, e, r^2 - d(e)\rangle\rangle \neq 0$ for some $e, r \in F$, $e \neq 0$, $r^2 - d(e) \neq 0$. In particular, $d(e) \neq 0$. Then $\delta \simeq \tau \otimes \langle\langle r^2 - d(e)\rangle\rangle$, where $\tau \simeq \langle 1, -a_4, -e, a_4ed(e)\rangle$

as earlier. The form $\langle a_4, e, -a_4ed(e)\rangle \subset -\delta'$ is isomorphic to the form φ with the matrix

$$M_\varphi = \begin{pmatrix} a_4 & 0 & \frac{1}{2}(a_2-e) \\ 0 & e & \frac{1}{2}a_1 \\ \frac{1}{2}(a_2-e) & \frac{1}{2}a_1 & a_0 \end{pmatrix}$$

with respect to a certain basis v_2, v_1, v_0, which implies that $f = \varphi(x^2v_2+xv_1+v_0) \in D(-\delta')$. Therefore, $\delta_{F(C)} = 0$, and we are done. Certainly, δ is not necessarily nonstandard. □

Corollary 3.3. *Let $\pi \in P_n(f)$, $n \geq 3$. Then there are $s_1, \dots, s_{n-3} \in F^*$ and $\rho \in P_3(f)$ such that $\pi \simeq \rho \otimes \langle\langle s_1, \dots, s_{n-3}\rangle\rangle$.*

Proof. This follows at once from the definition of standard Pfister forms and Proposition 3.2. □

If the u-invariant of F is small enough, then one can give a complete description of the ideal $W(F(C)/F)$.

Proposition 3.4. *Let F be a field with $u(F) \leq 10$ (for instance, F is the function field of a 3-dimensional variety over an algebraically closed field). Then any element of $W(F(C)/F)$ is a sum of an element from $P_2(f)$ and an element from $P_3(f)$.*

Proof. Let $\varphi \in W(F(C)/F)$. Since $\operatorname{disc}(\varphi)_{F(x)(\sqrt{f(x)})} = 1$, $a_4 \neq 0$, and $f(x) = a_4x^4+a_2x^2+a_1x+a_0$ is squarefree, we have $\operatorname{disc}(\varphi) = 1$. Since $C(\varphi)_{F(x)(\sqrt{f(x)})} = 0$, we get that $C(\varphi)$ is a quaternion. Let $\pi \in P_2(f)$ be a 2-fold Pfister form associated with $C(\varphi)$. If $\pi = 0$, then $C(\varphi) = 0$. Since $\dim(\varphi) \leq 10$, a result of Pfister implies that $\varphi \in I^3(F)$ [Scharlau 1985, Chapter 2, Theorem 14.4] (also this follows from Merkurjev's theorem, but we do not need this profound result here). Since $u(F) \leq 10$, it follows that φ is a 3-fold Pfister form [Lam 2005, Chapter XII, Proposition 2.8].

If $\pi \neq 0$, then similarly $\varphi - \pi \in I^3(F)$; hence $\varphi = \pi + (\varphi - \pi)$ is a sum of a 2-fold Pfister form and a 3-fold one from $W(F(C)/F)$. □

Open Question. Is the ideal $W(F(C)/F)$ generated by 2-fold and 3-fold Pfister forms in general?

A natural question arises as to whether nonstandard Pfister forms exist. The following statement shows that this is really the case.

Proposition 3.5. *Let $f(x) = a_4x^4 + a_2x^2 + a_1x + a_0$ be a squarefree polynomial over a field k. Let C be the curve with the equation $y^2 = f(x)$. The following conditions are equivalent:*

(1) *The curve C has no rational point over k.*

(2) *There exists a field extension F/k with a nonstandard 3-fold Pfister form over F for the curve C_F.*

(3) *There exist a field extension K/k such that* $\mathrm{cd}_2 K = 1$, *and a nonstandard 3-fold Pfister form over the rational function field $F = K(u, v)$ for the curve C_F. Moreover*, *in this example* $\mathrm{Br}(F(C)/F) = 0$.

Proof. (2) $\Longrightarrow$ (1): This is obvious, since if C had a k-rational point, then $W(F(C)/F)$ would be trivial for any field extension F/k.

(3) $\Longrightarrow$ (2): This is also obvious.

(1) $\Longrightarrow$ (3): In view of Corollary 2.7, there is a field extension K/k such that $\mathrm{cd}_2 K = 1$ and C has no K-rational point. Set $F = K(u, v)$ and consider the Pfister form $\pi \simeq \langle\langle a_4, u, v^2 - d(u)\rangle\rangle \in W(F(C)/F)$. Since C has no K-rational point, we get by Example 2.2 that $\partial_{v^2-d(u)}(\pi) = \langle\langle a_4, u\rangle\rangle_{K(u)(\sqrt{d(u)})} \neq 0$. Therefore, $\pi \neq 0$. Now to check that π is nonstandard, it suffices to show that $\mathrm{Br}(F(C)/F) = 0$. Since ${}_2\mathrm{Br}(K) = 0$, this is a direct consequence of the following:

Lemma 3.6. *The restriction map* $\mathrm{Br}(L(C)/L) \to \mathrm{Br}(L(u)(C)/L(u))$ *is an isomorphism for any field extension L/k.*

Proof. Obviously, the map in question is injective. By Proposition 3.1 any element of $\mathrm{Br}(L(u)(C)/L(u))$ equals $(a_4, p(u))$ for some $p \in L[u]$. Let q be a prime divisor of p. We have

$$\bar{a}_4 = \partial_q(a_4, p) \in \ker(L_q^*/L_q^{*2} \to L_q(C)^*/L_q(C)^{*2}).$$

Since $f(x) = a_4x^4 + a_2x^2 + a_1x + a_0$ is squarefree, $a_4 \in L_q^{*2}$; that is, $\partial_q(a_4, p) = \bar{1}$. Therefore, $(a_4, p) \in \mathrm{Br}(L(C)/L)$, so the lemma is proven, which completes also the proof of the implication (1) $\Longrightarrow$ (3). □

References

[Brumer 1978] A. Brumer, "Remarques sur les couples de formes quadratiques", *C. R. Acad. Sci. Paris Sér. A-B* **286**:16 (1978), 679–681. MR Zbl

[Elman et al. 2008] R. Elman, N. Karpenko, and A. Merkurjev, *The algebraic and geometric theory of quadratic forms*, American Mathematical Society Colloquium Publications **56**, American Mathematical Society, Providence, RI, 2008. MR Zbl

[Fitzgerald 1983] R. W. Fitzgerald, "Witt kernels of function field extensions", *Pacific J. Math.* **109**:1 (1983), 89–106. MR Zbl

[Haile and Han 2007] D. Haile and I. Han, "On an algebra determined by a quartic curve of genus one", *J. Algebra* **313**:2 (2007), 811–823. MR Zbl

[Hoffmann 1995] D. W. Hoffmann, "Isotropy of 5-dimensional quadratic forms over the function field of a quadric", pp. 217–225 in *K-theory and algebraic geometry: connections with quadratic forms and division algebras* (Santa Barbara, CA, 1992), edited by B. Jacob and A. Rosenberg, Proc. Sympos. Pure Math. **58**, American Mathematical Society, Providence, RI, 1995. MR Zbl

[Lam 2005] T. Y. Lam, *Introduction to quadratic forms over fields*, Graduate Studies in Mathematics **67**, American Mathematical Society, Providence, RI, 2005. MR Zbl

[Pfister 1995] A. Pfister, *Quadratic forms with applications to algebraic geometry and topology*, London Mathematical Society Lecture Note Series **217**, Cambridge Univ. Press, 1995. MR Zbl

[Scharlau 1985] W. Scharlau, *Quadratic and Hermitian forms*, Grundlehren der Math. Wissenschaften **270**, Springer, Berlin, 1985. MR Zbl

[Shick 1994] J. Shick, "On Witt-kernels of function fields of curves", pp. 389–398 in *Recent advances in real algebraic geometry and quadratic forms* (Berkeley, CA, 1990/1991; San Francisco, 1991), edited by W. B. Jacob et al., Contemp. Math. **155**, American Mathematical Society, Providence, RI, 1994. MR Zbl

[Wadsworth 1975] A. R. Wadsworth, "Similarity of quadratic forms and isomorphism of their function fields", *Trans. Amer. Math. Soc.* **208** (1975), 352–358. MR Zbl

Received January 7, 2017. Revised February 13, 2017.

ALEXANDER S. SIVATSKI
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
NATAL
BRAZIL
alexander.sivatski@gmail.com

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 292 No. 1 January 2018
