Pacific Journal of Mathematics

Volume 272 No. 1 November 2014

PACIFIC JOURNAL OF MATHEMATICS
msp.org/pjm

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

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

The subscription price for 2014 is US \$410/year for the electronic version, and \$535/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

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

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

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.272.1

NONCONCORDANT LINKS WITH HOMOLOGY COBORDANT ZERO-FRAMED SURGERY MANIFOLDS

JAE CHOON CHA AND MARK POWELL

We use topological surgery theory to give sufficient conditions for the zero-framed surgery manifold of a 3-component link to be homology cobordant to the zero-framed surgery on the Borromean rings (also known as the 3-torus) via a topological homology cobordism preserving the free homotopy classes of the meridians.

This enables us to give examples of 3-component links with unknotted components and vanishing pairwise linking numbers, such that any two of these links have homology cobordant zero-surgeries in the above sense, but the zero-surgery manifolds are not homeomorphic. Moreover, the links are not concordant to one another, and in fact they can be chosen to be height h but not height $h+1$ symmetric grope concordant, for each h which is at least three.

1. Introduction

It is well known that the study of homology cobordism of 3-manifolds is essential for understanding the concordance of knots and links: homology cobordism of the *exteriors* of links in S^3 is equivalent to concordance in a homology $S^3 \times I$, and an additional mild normal generation condition for π_1 is equivalent to topological concordance in $S^3 \times I$ (this also holds modulo the 4-dimensional Poincaré conjecture in the smooth case).

We recall the definitions: two m-component links L_0 and L_1 in S^3 are said to be *topologically* (respectively *smoothly*) *concordant* if there exist m locally flat (respectively smoothly embedded) disjoint annuli in $S^3 \times [0, 1]$ cobounded by components of $L_0 \times \{0\}$ and $-L_1 \times \{1\}$. Two 3-manifolds M_0 and M_1 bordered by a 2-manifold Σ, that is, endowed with a marking $\mu_i : \Sigma \xrightarrow{\cong} \partial M_i$, are *topologically* (respectively *smoothly*) *homology cobordant* if there is a topological (respectively smooth) 4-manifold W with

$$\partial W = M_0 \sqcup -M_1 \sqcup \Sigma \times [0, 1]/(\mu_0(x) \sim x \times \{0\}, \mu_1(x) \sim x \times \{1\}, x \in \Sigma),$$

MSC2010: 57M25, 57N70.
Keywords: homology cobordism, zero-framed surgery, topological surgery, link concordance, symmetric grope concordance.

such that the inclusions $M_i \to W$ ($i = 0, 1$) induce isomorphisms on integral homology groups. In this paper links are oriented, and link exteriors are always bordered by $\bigsqcup_m S^1 \times S^1$ under the zero framing.

In high dimensions, concordance classification results were obtained by studying homology surgery, with the aim of surgeries being to produce a homology cobordism of the exteriors (for example, see [Cappell and Shaneson 1974; 1980; Le Dimet 1988]). On the other hand, for knots and links in dimension three, the *zero-surgery manifolds* and their 4-dimensional homology cobordisms have been extensively used in the literature in order to understand the structure peculiar to low dimensions, especially in the topological category. Recall that performing zero-framed surgery on a link in S^3 yields a closed 3-manifold, called the zero-surgery manifold.

The classical invariants such as the knot signature and Levine's algebraic knot concordance class [Levine 1969a; 1969b] are obtained from the zero-surgery manifold of a knot, via the Blanchfield form. Also, higher-order knot concordance obstructions, such as Casson–Gordon invariants [Casson and Gordon 1978; 1986], and Cochran–Orr–Teichner L^2-signatures [Cochran et al. 2003] are obtained from the zero-surgery manifold (often together with the homology class of the meridian).

A natural interesting question is whether the homology cobordism class of a zero-surgery manifold determines the concordance class of a knot or link or if it determines the homology cobordism class of the exterior.

In this paper we show, in a strong sense involving homotopy of meridians, that the answer is negative for a large class of *links* satisfying a certain nonvanishing condition on Milnor's $\bar{\mu}$-invariants, even in the framework of symmetric grope and Whitney tower generalisations of concordance and homology cobordism in the sense of [Cochran et al. 2003; Cha 2014]. Also, we employ topological surgery in dimension 4 to give a new construction of homology cobordisms of zero-surgery manifolds. Next we state our main theorems, after which we will discuss these aspects further.

Theorem 1.1. *Suppose $h \geq 3$. Then there are infinitely many 3-component links $L_0, L_1, \ldots,$ with vanishing pairwise linking numbers and with unknotted components, satisfying the following for any $i \neq j$:*

(1) *The zero-surgery manifolds M_{L_i} and M_{L_j} are not homeomorphic.*

(2) *There is a topological homology cobordism between M_{L_i} and M_{L_j} in which the k-th meridians of L_i and L_j are freely homotopic for each $k = 1, 2, 3$.*

(3) *The links L_i and L_j are height h but not height $h+1$ symmetric grope concordant. In particular, L_i and L_j are not concordant.*

For a definition of height h symmetric grope concordance, see Definition 4.2. Our links are obtained from the Borromean rings by performing a satellite construction along a curve lying in the kernel of the map $\pi_1(S^3 \setminus L) \to \pi_1(M_L)$ induced by inclusion.

As a counterpoint to Theorem 1.1, we show that there are infinite families of links with the same nonvanishing Milnor invariants, with homeomorphic zero-surgery manifolds preserving the homotopy classes of the meridians, but which are not concordant.

The Milnor invariant of an m-component link associated to a multi-index $I = i_1 i_2 \cdots i_r$ with $i_j \in \{1, \dots, m\}$, as defined in [Milnor 1957], will be denoted by $\bar{\mu}_L(I)$. We denote its length by $|I| := r$. Define $k(m) := \lfloor \log_2(m-1) \rfloor$.

Theorem 1.2. *Let I be a multi-index with nonrepeating indices with length $m \geq 2$. For any $h \geq k(m)+2$ there are infinitely many m-component links $L_0, L_1, \dots,$ with unknotted components, satisfying the following*:

(1) *The L_i have identical $\bar{\mu}$-invariants, $\bar{\mu}_{L_i}(I) = 1$, and $\bar{\mu}_{L_i}(J) = 0$ for $|J| < |I|$.*

(2) *There is a homeomorphism between the zero-surgery manifolds M_{L_i} and M_{L_j} which preserves the homotopy classes of the meridians.*

(3) *The links L_i and L_j are height h but not height $h+1$ symmetric grope concordant. In particular, L_i and L_j are not concordant.*

The case when $m \geq 3$ should be compared with Theorem 1.1 since then the links L_i have vanishing pairwise linking numbers. To construct such links we start with certain iterated Bing doubles constructed using T. Cochran's algorithm, which realise the Milnor invariant required. We then perform satellite operations which affect the concordance class of the link but do not change the homeomorphism type of the zero-surgery manifold.

We remark that we could also phrase Theorems 1.1 and 1.2 in terms of symmetric Whitney tower concordance instead of grope concordance.

In the three subsections below, we discuss some features of Theorem 1.1, regarding (i) the use of topological surgery in dimension 4, (ii) link concordance versus zero-surgery homology cobordism, and (iii) link exteriors and the homology surgery approach.

1A. ***Topological surgery for 4-dimensional homology cobordism.*** An interesting aspect of the proof of Theorem 1.1 is that we employ topological surgery in dimension 4 to give a sufficient condition for certain zero-surgery manifolds of 3-component links to be homology cobordant. It is well known that topological surgery in dimension 4 is useful for obtaining homology cobordisms (and consequently concordances), although the current state of the art in terms of "good" groups, for which the π_1-null disc lemma is known, is still insufficient for the general case. M. Freedman showed that knots of Alexander polynomial one are concordant to the unknot [Freedman and Quinn 1990, Theorem 11.7B]. J. Davis

[2006] extended the program to show that 2-component links with Alexander polynomial one are concordant to the Hopf link. These two cases use topological surgery over fundamental groups $\mathbb{Z}$ and $\mathbb{Z}^2$, respectively. Due to the rarity of good groups for 4-dimensional topological surgery, there are not many other situations where such positive results on knot and link concordance can currently be proven. As another case, S. Friedl and P. Teichner [2005] found sufficient conditions for a knot to be homotopy ribbon, and in particular slice, with a certain ribbon group $\mathbb{Z} \ltimes \mathbb{Z}\left[\frac{1}{2}\right]$.

We give another instance of the utility of topological surgery for constructing homology cobordisms, using the group $\mathbb{Z}^3$, which is manageable from the point of view of topological surgery in dimension 4. Indeed, our sufficient condition for zero-surgery manifolds to be homology cobordant focuses on the Borromean rings as a base link. The zero-surgery manifold M_{Bor} of the Borromean rings is the 3-torus $T^3 = S^1 \times S^1 \times S^1$, whose fundamental group is $\mathbb{Z}^3$.

To state our result, we with the following notation: Let

$$\Lambda := \mathbb{Z}[\mathbb{Z}^3] = \mathbb{Z}[t_1^{\pm 1}, t_2^{\pm 1}, t_3^{\pm 1}].$$

Denote the zero-surgery manifold of a link L by M_L as before. For a 3-component link L with vanishing pairwise linking numbers, there is a canonical homotopy class of maps $f_L : M_L \to M_{\mathrm{Bor}} = T^3$ which send the homotopy class of the i-th meridian of L to that of the Borromean rings, namely the i-th circle factor of T^3. After choosing an identification of $\pi_1(T^3) = \mathbb{Z}^3$, we can use this to define the Λ-coefficient homology $H_1(M_L; \Lambda)$. We say that a map $f : M_L \to T^3$ is a Λ-*homology equivalence* if f is homotopic to f_L and f induces isomorphisms on $H_*(-; \Lambda)$.

Theorem 1.3. *Suppose L is a 3-component link whose components have trivial Arf invariants and there exists a Λ-homology equivalence $M_L \to T^3$. Then there is a homology cobordism W between M_L and $T^3 = M_{\mathrm{Bor}}$ for which the inclusion-induced maps $\pi_1(M_L) \to \pi_1(W) \xleftarrow{\cong} \pi_1(T^3)$ are such that the composition from left to right takes meridians to meridians.*

1B. ***Link concordance versus zero-surgery homology cobordism.*** We review the general question of whether links with homology cobordant zero-surgery manifolds are concordant. The answer to the basic question is easily seen to be no, once one knows of a result of C. Livingston [1983] that there are knots not concordant to their reverses. Note that a knot and its reverse have the same zero-surgery manifold. This leads us to consider some additional conditions on the homology cobordism, involving the meridians. In what follows, meridians are always positively oriented.

First, observe that the exteriors of two links are homology cobordant if and only if the zero-framed meridians cobound framed annuli disjointly embedded in a homology cobordism of the zero-surgery manifolds. (For the if direction, note that the exterior of the framed annuli is a homology cobordism of the link exteriors.)

In particular, it holds if two links (or knots) are concordant.

Regarding the knot case, in [Cochran et al. 2013], T. Cochran, B. Franklin, M. Hedden and P. Horn considered homology cobordisms of zero-surgery manifolds in which the meridians are *homologous*: in the smooth category, they showed that the existence of such a homology cobordism is insufficient for knots to be concordant. In the topological case this is still left unknown.

Concerning a stronger *homotopy* analogue, the following is unknown in both the smooth and topological cases:

Question 1.4. If there is a homology cobordism of zero-surgery manifolds of two knots in which the meridians are homotopic, are the knots concordant? Or concordant in a homology $S^3 \times I$?

For the link case, results in the literature give nonconcordant examples whose zero-surgery manifolds admit a homology cobordism with homotopic meridians. As a generic example in the topological category, consider a 2-component link with linking number one. The zero-surgery manifold is a homology 3-sphere, which bounds a contractible topological 4-manifold by [Freedman and Quinn 1990, Corollary 9.3C]. Taking the connected sum of such 4-manifolds, one obtains the following: *the zero-surgery manifolds of any two linking number one 2-component links cobound a simply connected topological homology cobordism.* Note that in this case the meridians are automatically homotopic. There are many linking number one 2-component links which are not concordant, as can be detected, for example, by the multivariable Alexander polynomial [Kawauchi 1978; Nakagawa 1978]. For related in-depth study, the reader is referred to, for instance, [Cha and Ko 1999; Friedl and Powell 2011; Cha 2014]. With our respective coauthors, we gave nonconcordant linking number one links with two unknotted components, for which abelian invariants such as the multivariable Alexander polynomial are unable to obstruct them from being concordant.

There are other examples which have knotted components: in [Cochran et al. 2013, end of Section 1], the authors discuss 2-component linking number zero links with homeomorphic zero-surgery manifolds which have nonconcordant (knotted) components. These links are obviously not concordant, and it can be seen that the homeomorphisms preserve meridians up to homotopy.

By contrast with the above examples, our links have unknotted components and vanishing pairwise linking numbers. Another feature exhibited by the links of Theorems 1.1 and 1.2 is that the entire subtlety of symmetric grope concordance of links can occur, within a single homology cobordism/homeomorphism class of the zero-surgeries, even modulo local knot tying.

We remark that all the links of Theorems 1.1 and 1.2 lie in the same “k-solv-equivalence class” for all k in the sense of [Cochran and Kim 2008, Definition 2.5].

1C. ***Link exteriors and the homology surgery approach.*** Our results serve to underline the philosophy that when investigating the relative problem of whether two links are concordant, and neither of them are the unlink, one should consider obstructions to homology cobordism of the link exteriors viewed as bordered manifolds, rather than to homology cobordism of the zero-surgery manifolds, even in low dimensions. This was implemented in, for example, [Kawauchi 1978; Nakagawa 1978; Cha 2014] (see also [Friedl and Powell 2011] for a related approach).

Although we stated our results in terms of grope concordance of links in Theorems 1.1 and 1.2 given above, in fact we show more: the link exteriors are far from being homology cobordant, as measured in terms of Whitney towers. A more detailed discussion is given in Section 5. For the purpose of distinguishing exteriors, we use the amenable Cheeger–Gromov ρ-invariant technology for bordered 3-manifolds (particularly for link exteriors) developed in [Cha 2014], generalising applications of ρ-invariants to concordance and homology cobordism in [Cochran et al. 2003; 2009; Cha and Orr 2012].

We will now discuss our results from the viewpoint of the homology surgery approach to link concordance classification, initiated by S. Cappell and J. Shaneson [1974; 1980] and implemented in high dimensions by J. Le Dimet [1988] using P. Vogel's homology localisation of spaces [1978]. The strategy consists of two parts. Consider the problem of comparing two given link exteriors. First we decide whether the exteriors have the same "Poincaré type", which roughly means that they have homotopy-equivalent Vogel homology localisations. If so, there is a common finite target space, into which the exteriors are mapped by homology equivalences rel. boundary. Once this is the case, a surgery problem is defined, and one can try to decide whether homology surgery gives a homology cobordism of the exteriors. The first step is obstructed by homotopy invariants (including Milnor $\bar{\mu}$-invariants in low dimension). The failure of the second step is measured by surgery obstructions, which are not yet fully formulated in low dimension (even modulo the fact that 4-dimensional surgery might not work), since the fundamental group plays a more sophisticated central rôle; see [Powell 2012] for the beginning of an algebraic surgery approach to this problem in the context of knot slicing.

Our examples illustrate that for many Poincaré types, namely those in Theorems 1.1 and 1.2, we get a rich theory of surgery obstructions within each Poincaré type, which is invisible via zero-surgery manifolds. We remark that for our links L_i in Theorems 1.1 and 1.2, there is a homology equivalence of the exterior of each L_i into that of a fixed one, say L_1, since we use satellite constructions (see Section 4). It follows that the exteriors have the same Poincaré type in the above sense. In this paper, (parts of the not yet fully formulated) homology surgery obstructions in dimension 4 have their incarnation in Theorem 5.2, the Amenable Signature Theorem.

Organisation of the paper. In Section 2, we explore the implications of the hypothesis that a homology equivalence $f : M_L \to T^3$ as in Theorem 1.3 exists, and we prove Theorem 1.3 in Section 3. In Section 4, we construct links with a given Milnor invariant with nonrepeating indices, and perform satellite operations on the links to construct the links of Theorems 1.1 and 1.2, which are height h symmetric grope concordant. In Section 5, we show that none of these links are height $h+1$ grope concordant to one another.

2. Homology type of zero-surgery manifolds and the 3-torus

This section discusses the hypotheses of Theorem 1.3. We begin the section by briefly reminding the reader who is familiar with Kirby calculus of a nice way to see the following fact.

Lemma 2.1. *The zero-surgery manifold of the Borromean rings is homeomorphic to the* 3*-torus.*

Proof. Place dots on two components of the Borromean rings and a zero near the other. Each component of the Borromean rings is a commutator in the meridians of the other two components, so this is a Kirby diagram for $T^2 \times D^2$, whose boundary is T^3. The 1-handles (dotted circles) can be replaced with zero-framed 2-handles without changing the boundary. □

In the following proposition we expand on the meaning and implications of the condition in Theorem 1.3. Denote the exterior of a link L by $X_L := S^3 \smallsetminus \nu L$ as before.

Proposition 2.2. *Suppose that L is a* 3*-component link. Then the following are equivalent*:

(1) *There is a Λ-homology equivalence $f : M_L \to T^3$.*
(2) *The preferred longitudes generate the link module $H_1(X_L;\, \Lambda)$.*
(3) *The pairwise linking numbers of L vanish and $H_1(M_L;\, \Lambda) = 0$.*

Furthermore, (*any of*) *the above conditions imply that L has multivariable Alexander polynomial $\Delta_L = (t_1 - 1)(t_2 - 1)(t_3 - 1)$, and this implies that the Milnor invariant $\bar{\mu}_L(123)$ is equal to ± 1.*

Proof. First we will observe that (2) and (3) are equivalent. Longitudes of L represent elements in $H_1(X_L;\, \Lambda) \cong \pi_1(X_L)^{(1)}/\pi_1(X_L)^{(2)}$ if and only if they are zero in $H_1(X_L;\, \mathbb{Z}) \cong \mathbb{Z}^3$; that is, the pairwise linking numbers are zero. If this is the case, $H_1(M_L;\, \Lambda)$ is isomorphic to $H_1(X_L;\, \Lambda)/\langle\text{longitudes}\rangle$, since M_L is obtained by attaching three 2-handles to E_L along the longitudes and then attaching three 3-handles along the boundary. It follows that longitudes generate $H_1(X_L;\, \Lambda)$ if and only if $H_1(M_L;\, \Lambda) = 0$.

Suppose (1) holds. Denote the meridians of L by μ_i ($i = 1, 2, 3$) and the linking number of the i-th and j-th components by ℓ_{ij}. The i-th longitude λ_i, which is homologous to $\sum_{j\neq i} \ell_{ij}\mu_i$, is zero in $H_1(M_L; \mathbb{Z}) \cong H_1(T^3; \mathbb{Z})$. Since $\{f_*([\mu_i])\}$ forms a basis of $H_1(T^3; \mathbb{Z}) \cong \mathbb{Z}^3$, it follows by linear independence that $\ell_{ij} = 0$ for any i and j. Also, $H_1(M_L; \Lambda) \cong H_1(T^3; \Lambda) = 0$. This shows that (3) holds.

Suppose (3) holds. Start with a map $g : \partial X_L = \bigsqcup_3 S^1 \times S^1 \to T^3$ that sends μ_i to the i-th S^1 factor and λ_i to a point. Observe that $g_* : H_1(\partial X_L; \mathbb{Z}) \to H_1(T^3; \mathbb{Z})$ factors through the inclusion-induced map $i_* : H_1(\partial X_L; \mathbb{Z}) \to H_1(M_L; \mathbb{Z})$ and the identifications $H_1(M_L; \mathbb{Z}) \xrightarrow{\cong} \mathbb{Z}^3 \xleftarrow{\cong} H_1(T^3; \mathbb{Z})$; this follows from the fact that $H_1(\partial X_L; \mathbb{Z}) \cong \mathbb{Z}^6$ is generated by the μ_i and λ_i and that both g_* and i_* are quotient maps, with their kernels generated by the λ_i. Since T^3 is a $K(\mathbb{Z}^3, 1)$, elementary obstruction theory shows that g extends to a map $f : M_L \to T^3$.

Consider the universal coefficient spectral sequence (see, e.g., [Levine 1977, Theorem 2.3]) $E^2_{p,q} = \operatorname{Ext}^p_\Lambda(H_q(M_L; \Lambda), \Lambda) \Rightarrow H^n(M_L; \Lambda)$. We have $E^2_{0,1} = 0$ since $H_1(M_L; \Lambda) = 0$, and $E^2_{1,0} = \operatorname{Ext}^1_\Lambda(\mathbb{Z}, \Lambda) = H^1(T^3; \Lambda) = 0$. It follows that $H^1(M_L; \Lambda) = 0$. By duality, $H_2(M_L; \Lambda) = 0$. Also, $H_3(M_L; \Lambda) = 0$ since the $\mathbb{Z}^3$-cover of M_L is noncompact. Since $H_0(M_L; \Lambda) \cong \mathbb{Z} \cong H_0(T^3; \Lambda)$ and $H_i(T^3; \Lambda) = 0$ for $i > 0$, it follows that f is a Λ-homology equivalence. This completes the proof of the equivalence of (1), (2) and (3).

Suppose (1), (2) and (3) hold. Recall that the scalar multiplication of a loop by t_i in the module $H_1(X_L; \Lambda)$ is defined to be conjugation by μ_i. Since λ_i and μ_i commute, we have $(t_i - 1)\lambda_i = 0$ in $H_1(X_L; \Lambda)$. From this and (2), it follows that there is an epimorphism of $A := \bigoplus_{i=1}^3 \Lambda/\langle t_i - 1\rangle$ onto $H_1(X_L; \Lambda)$. Since the zero-th elementary ideal of A is the principal ideal generated by $(t_1 - 1)(t_2 - 1)(t_3 - 1)$, it follows that Δ_L is a factor of $(t_1 - 1)(t_2 - 1)(t_3 - 1)$. We now invoke the Torres condition (see, e.g., [Kawauchi 1996, Theorem 7.4.1]):

$$\Delta_L(1, t_2, t_3) = (t_2^{\ell_{12}} t_3^{\ell_{13}} - 1)\Delta_{L'}(t_2, t_3),$$

where L' is the sublink of L with the first component deleted and ℓ_{ij} is the pairwise linking number. Since $\ell_{ij} = 0$ by (3), we have $\Delta_L(1, t_2, t_3) = 0$. It follows that $t_1 - 1$ is a factor of Δ_L. Similarly $t_2 - 1$ and $t_3 - 1$ are factors. Therefore $\Delta_L(t_1, t_2, t_3) = (t_1 - 1)(t_2 - 1)(t_3 - 1)$.

To show the last part, suppose that $\Delta_L(t_1, t_2, t_3) = (t_1 - 1)(t_2 - 1)(t_3 - 1)$. By [Kawauchi 1996, Proposition 7.3.14], the single-variable Alexander polynomial $\Delta_L(t)$ of L is given by

$$\Delta_L(t) = (t - 1)\Delta_L(t, t, t) = (t - 1)^4 \doteq ((\sqrt{t})^{-1} - \sqrt{t})^4.$$

It follows that L has Conway polynomial $\nabla_L(z) = z^4$, by the standard substitution $z = (\sqrt{t})^{-1} - \sqrt{t}$. Cochran [1985, Theorem 5.1] identified the coefficient of z^4

in $\nabla_L(z)$ with $(\mu_L(123))^2$ for 3-component links with pairwise linking number zero. Applying this to our case, it follows that $\bar{\mu}_L(123) = \pm 1$. □

3. Construction of homology cobordisms using topological surgery

This section gives the proof of Theorem 1.3. The proof will use surgery theory, and will parallel the proof given by Davis [2006] (see also [Hillman 2002, Section 7.6]). We will provide some details in order to fill in where the treatment in [Davis 2006] was terse, and to convince ourselves that the analogous arguments work in the case of interest.

For the convenience of the reader we restate Theorem 1.3 here.

Theorem 1.3. *Suppose L is a 3-component link whose components have trivial Arf invariants and there exists a Λ-homology equivalence $M_L \to T^3$. Then there is a homology cobordism W between M_L and $T^3 = M_{\mathrm{Bor}}$ for which the inclusion-induced maps $\pi_1(M_L) \to \pi_1(W) \xleftarrow{\simeq} \pi_1(T^3)$ are such that the composition from left to right takes meridians to meridians.*

Remark 3.2. It is an interesting question to determine whether there are extra conditions which can be imposed in order to see that the Arf invariants of the components are forced to vanish by the homological assumptions. In the cases of knots and 2-component links with Alexander polynomial one, the Arf invariants of the components are automatically trivial. For knots, $\Delta_K(-1)$ computes the Arf invariant, by [Levine 1966]. For 2-component links one observes that $\Delta_L(t, 1)$ and $\Delta_L(1, t)$ give the Alexander polynomials of the components, by the Torres condition, and then applies Levine's theorem. These arguments do not seem to extend to the 3-component case of current interest.

The proof of Theorem 1.3 will occupy the rest of this section. In order to produce a homology cobordism, we will first show that there is a normal cobordism between normal maps $f : M_L \to T^3$ and $\mathrm{Id} : T^3 \to T^3$. Interestingly, we can work with smooth manifolds in order to establish the existence of a normal cobordism. This will make arguments which invoke tangent bundles and transversality easier to digest. Only at the end of the proof of Theorem 1.3, where we take connected sums with the E_8-manifold, and where we claim that the vanishing of a surgery obstruction implies that surgery can be done, do we need to leave the realm of smooth manifolds.

Definition 3.3. Let X be an n-dimensional manifold with a vector bundle $\nu \to X$. A degree-one normal map (F, b) over X is an n-manifold M with a map $F : M \to X$ which induces an isomorphism $F_* : H_n(M; \mathbb{Z}) \xrightarrow{\simeq} H_n(X; \mathbb{Z})$, together with a stable trivialisation $b : TM \oplus F^*\nu \oplus \varepsilon^l \cong \varepsilon^k$.

A degree-one normal cobordism (J, e) between normal maps $(F : M \to X, b)$ and $(G : N \to X, c)$ is an $(n+1)$-dimensional cobordism between M and N with a map $J : W \to X \times I$ extending $F : M \to X \times \{0\}$ and $G : M \to X \times \{1\}$, which induces an isomorphism

$$J_* : H_{n+1}(W, \partial W; \mathbb{Z}) \xrightarrow{\simeq} H_{n+1}(X \times I, X \times \{0, 1\}; \mathbb{Z}),$$

together with a stable trivialisation $e : TW \oplus J^*(\nu \times I) \oplus \varepsilon^{l'} \cong \varepsilon^{k'}$.

For us, let $X = T^3$, and let ν be its tangent bundle. We fix a framing on the stable tangent bundle of the target torus T^3 once and for all. Note that this canonically determines a trivialisation of the tangent bundle of $F^*\nu$, for any map $F : M \to X$, by the following diagram, in which the bottom composition is the constant map, denoted $*$, and the top composition is the pull back $F^*\nu$. The middle composition is the induced framing.

$$\begin{tikzcd}
M \times \{0\} \arrow[r, "F\times \mathrm{Id}"] \arrow[d] & T^3 \times \{0\} \arrow[r, "\nu"] \arrow[d] & \mathrm{BO}(n) \arrow[d] \\
M \times I \arrow[r, "F\times \mathrm{Id}"] & T^3 \times I \arrow[r] & \mathrm{BO} \\
M \times \{1\} \arrow[r, "F\times \mathrm{Id}"] \arrow[u] & T^3 \times \{1\} \arrow[r, "*"] \arrow[u] & \mathrm{BO}(n) \arrow[u]
\end{tikzcd}$$

A framing of the tangent bundle of the domain therefore determines a normal map.

Lemma 3.4. *Let L be a link whose components have trivial Arf invariants, and let $f : M_L \to T^3$ be a degree-one normal map which induces a $\mathbb{Z}$-homology isomorphism and which maps a chosen meridian μ_i to the i-th S^1 factor of T^3 for $i = 1, 2, 3$. We can make a homotopy of f and choose a framing on M_L so that $f : M_L \to T^3$ and $\mathrm{Id} : T^3 \to T^3$ are degree-one normal cobordant.*

Proof. We need to show that we can choose a framing on M_L such that the disjoint union $M_L \sqcup -T^3$ represent the trivial element of $\Omega_3^{\mathrm{fr}}(T^3)$. We compute this bordism group:

$$\widetilde{\Omega}_3^{\mathrm{fr}}(T^3) \cong \widetilde{\Omega}_4^{\mathrm{fr}}(\Sigma T^3) \cong \widetilde{\Omega}_4^{\mathrm{fr}}(S^2 \vee S^2 \vee S^2 \vee S^3 \vee S^3 \vee S^3 \vee S^4),$$

with this last isomorphism induced by a homotopy equivalence of spaces. There is a copy of S^{i+1} for each i-cell of T^3, for $i = 1, 2, 3$. To see this homotopy equivalence, we need to see that the attaching maps of the cells are null-homotopic. The suspension of the 1-skeleton of T^3 is $S^2 \vee S^2 \vee S^2$. The Hilton–Milnor theorem [Hilton 1955, Theorem A] computes the homotopy groups of a wedge of spheres.

The attaching maps for the 2-cells of T^3 become the attaching maps for the 3-cells of ΣT^3, namely maps in

$$\pi_2(S^2 \vee S^2 \vee S^2) \cong \bigoplus_3 \pi_2(S^2) \cong \bigoplus_3 \mathbb{Z},$$

where the first isomorphism is by the Hilton–Milnor theorem. The commutator attaching maps become trivial in the abelian $\pi_2(S^2)$. Therefore the 3-skeleton of ΣT^3 is a wedge $S^2 \vee S^2 \vee S^2 \vee S^3 \vee S^3 \vee S^3$. The attaching map for the 3-cell of T^3 becomes the attaching map for the 4-cell of ΣT^3, a map in

$$\pi_3(S^2 \vee S^2 \vee S^2 \vee S^3 \vee S^3 \vee S^3) \cong \bigoplus_{1 \leq i \leq 3} \pi_3(S^3) \oplus \bigoplus_3 \pi_3(S^2) \oplus \bigoplus_{1 \leq i < j \leq 3} \pi_3(S^3),$$

again by the Hilton–Milnor theorem, where the last three $\pi_3(S^3)$ summands include into $\pi_3(S^2 \vee S^2 \vee S^2 \vee S^3 \vee S^3 \vee S^3)$ via composition with the Whitehead product: let $f_i : S^2 \to S_i^2$ be a generator of $\pi_2(S_i^2)$, where S_i^2 is the i-th S^2 component in the wedge. Then the Whitehead product is the homotopy class of the map $[f_i, f_j] \in \pi_3(S_i^2 \vee S_j^2)$, which is the attaching map for the 4-cell in a standard cellular decomposition of $S^2 \times S^2$. Since $\pi_2(S^1) \cong \pi_2(S^1 \vee S^1) \cong 0$, the summands associated to the S^2 components of the wedge do not arise from a suspension. The summands associated to the S^3 components are null-homotopic since the 3-cell of T^3 is attached to each 2-cell twice, once on either side. This completes the explanation of the claimed homotopy equivalence:

$$\Sigma T^3 \simeq S^2 \vee S^2 \vee S^2 \vee S^3 \vee S^3 \vee S^3 \vee S^4.$$

By Mayer–Vietoris, the bordism group $\widetilde{\Omega}_4^{\text{fr}}(S^2 \vee S^2 \vee S^2 \vee S^3 \vee S^3 \vee S^3 \vee S^4)$ is a direct sum

$$\begin{aligned} \bigoplus_3 \widetilde{\Omega}_4^{\text{fr}}(S^2) \oplus \bigoplus_3 \widetilde{\Omega}_4^{\text{fr}}(S^3) \oplus \widetilde{\Omega}_4^{\text{fr}}(S^4) &\cong \bigoplus_3 \widetilde{\Omega}_2^{\text{fr}}(S^0) \oplus \bigoplus_3 \widetilde{\Omega}_1^{\text{fr}}(S^0) \oplus \widetilde{\Omega}_0^{\text{fr}}(S^0) \\ &\cong \bigoplus_3 \Omega_2^{\text{fr}} \oplus \bigoplus_3 \Omega_1^{\text{fr}} \oplus \Omega_0^{\text{fr}} \\ &\cong \bigoplus_3 \mathbb{Z}_2 \oplus \bigoplus_3 \mathbb{Z}_2 \oplus \mathbb{Z}. \end{aligned}$$

Therefore

$$\Omega_3^{\text{fr}}(T^3) \cong \Omega_3^{\text{fr}} \oplus \bigoplus_3 \mathbb{Z}_2 \oplus \bigoplus_3 \mathbb{Z}_2 \oplus \mathbb{Z} \cong \mathbb{Z}_{24} \oplus \bigoplus_3 \mathbb{Z}_2 \oplus \bigoplus_3 \mathbb{Z}_2 \oplus \mathbb{Z}.$$

The isomorphism is given as follows. Let

$$\mathrm{pr}_i : T^3 = S^1 \times S^1 \times S^1 \longrightarrow S^1$$

be given by projection onto the i-th factor. Similarly, let

$$\mathrm{qr}_i : T^3 = S^1 \times S^1 \times S^1 \longrightarrow S^1 \times S^1$$

be given by forgetting the i-th factor. Let $F : M \to T^3$ be an element of $\Omega_3^{\mathrm{fr}}(T^3)$. Making all maps transverse to a point, we obtain an 8-tuple

$$\begin{aligned}\big([M], (\mathrm{pr}_1 \circ F)^{-1}(*), (\mathrm{pr}_2 \circ F)^{-1}(*), (\mathrm{pr}_3 \circ F)^{-1}(*),\\ (\mathrm{qr}_1 \circ F)^{-1}(*), (\mathrm{qr}_2 \circ F)^{-1}(*), (\mathrm{qr}_3 \circ F)^{-1}(*), F^{-1}(*)\big)\\ \in \Omega_3^{\mathrm{fr}} \oplus \bigoplus_3 \Omega_2^{\mathrm{fr}} \oplus \bigoplus_3 \Omega_1^{\mathrm{fr}} \oplus \Omega_0^{\mathrm{fr}} \cong \mathbb{Z}_{24} \oplus \bigoplus_3 \mathbb{Z}_2 \oplus \bigoplus_3 \mathbb{Z}_2 \oplus \mathbb{Z}.\end{aligned}$$

We consider each of the summands in turn.

By choosing the appropriate orientation on M_L and making the degree-one normal maps transverse to a point, one can arrange for the disjoint union $f^{-1}(*) \sqcup -\mathrm{Id}^{-1}(*)$ to be equal to $\{\mathrm{pt}\} \sqcup -\{\mathrm{pt}\} = 0 \in \Omega_0^{\mathrm{fr}}$.

As observed in [Davis 2006, proof of the lemma], we can change the framing so that the elements of Ω_1^{fr} agree. First, we change the framing on each of three chosen meridians μ_i to the link components L_i.

Orientable k-plane vector bundles over S^1 are classified by homotopy classes of maps $[S^1, \mathrm{BSO}(k)]$. Consider the exact sequence

$$\pi_2(\mathrm{BSO}) \longrightarrow \pi_2(\mathrm{BSO}, \mathrm{BSO}(k)) \longrightarrow \pi_1(\mathrm{BSO}(k)) \xrightarrow{\gamma} \pi_1(\mathrm{BSO}).$$

A stably trivial vector bundle over S^1 gives us an element of $\ker(\gamma)$. A choice of trivialisation of the vector bundle gives us a null homotopy and therefore an element of $\pi_2(\mathrm{BSO}, \mathrm{BSO}(k))$. The possible choices of stable trivialisations, or framings, are indexed by $\pi_2(\mathrm{BSO}) \cong \pi_1(\mathrm{SO}) \cong \mathbb{Z}_2$.

We can therefore, if necessary, change the framing on each μ_i to be the bounding framing using an element of $\pi_1(\mathrm{SO}(2))$ which maps to the nontrivial element of $\pi_1(\mathrm{SO})$. Use the element of $\pi_1(\mathrm{SO}(2))$ to change the framing on the normal bundle of μ_i in M_L. We claim that these changes in the framing can be extended to the whole of M_L. To see this, we argue as follows. The dual of the inclusion map $H^1(M_L; \mathbb{Z}) \to H^1(\mu_i; \mathbb{Z})$ is surjective, since each $[\mu_i]$ is a generator of $H_1(M_L; \mathbb{Z})$. The change of framing map $\mu_i \to \mathrm{SO}(2)$ represents a homotopy class of maps in $[\mu_i, S^1]$ and therefore an element of $H^1(\mu_i; \mathbb{Z})$. Since this pulls back to an element of $H^1(M_L; \mathbb{Z})$, which can in turn produce a map $M_L \to \mathrm{SO}(2)$, the change of framing map can be extended as claimed.

Let $N_i \subset M_L$ be the submanifolds given by $(\mathrm{qr}_i \circ f)^{-1}(*)$, after perturbing f to make $\mathrm{qr}_i \circ f$ transverse to a point. As the inverse image of the i-th S^1 factor of T^3 (e.g., $f^{-1}(S^1 \times \{*\} \times \{*\})$), N_i is a collection of circles. After a homotopy of f, it can be arranged, by the assumption on f, that N_i is a single meridian μ_i, which has the bounding framing and therefore represents the zero element in Ω_1^{fr}. To make this arrangement, it suffices to be able to remove circles N_i whose image in T^3 is null-homologous. But in T^3, a null-homologous curve is also null-homotopic. Therefore we can make a homotopy of f so that N_i misses $S^1 \times \{*\} \times \{*\}$.

After another homotopy, the inverse image $(\mathrm{pr}_i \circ f)^{-1}(*)$ can be arranged to be a capped-off Seifert surface $F_i \cup D^2$, where F_i is a Seifert surface for L_i (possibly with closed connected components). To see this, we again use our assumption that f sends the i-th meridian μ_i to the i-th circle. This assumption enables us to homotope f so that $\mathrm{pr}_i \circ f$ sends a regular neighbourhood $\mu_i \times D^2$ to S^1 by projection onto the first factor. Then the inverse image is as desired. A homotopy of f preserves the framed bordism class of $(\mathrm{pr}_i \circ f)^{-1}(*)$, and the class $[F_i \cup D^2] \in \Omega_2^{\mathrm{fr}}$ is determined by the Arf invariant of L_i. By hypothesis, this vanishes.

Finally, again following [Davis 2006] (see also [Freedman and Quinn 1990, proof of Lemma 11.6B]), the framing can be altered in the neighbourhood of a point to change the element $[M] \in \Omega_3^{\mathrm{fr}}$ to the trivial element. We recall the definition of the J-homomorphism $J : \pi_3(\mathrm{SO}) \to \pi_3^S \cong \Omega_3^{\mathrm{fr}}$, for the convenience of the reader, where π_k^S is the k-th stable homotopy group of spheres. (Incidentally, $\pi_3(\mathrm{SO}) \cong \mathbb{Z}$ and $\pi_3^S \cong \mathbb{Z}_{24}$.) Given $\theta : S^3 \to \mathrm{SO}$, choose a k sufficiently large so that we can represent θ by a map $\theta : S^3 \to \mathrm{SO}(k)$. We proceed to construct a map $(J(\theta) : S^{k+3} \to S^k) \in \pi_3^S$. So:

$$S^{k+3} = S^3 \times D^k \cup_{S^3 \times D^{k-1}} D^4 \times S^{k-1}.$$

Define a map

$$\begin{aligned} j(\theta) : S^3 \times D^k &\to S^3 \times D^k \\ (x, y) &\mapsto (x, \theta(x)(y)), \end{aligned}$$

since $\theta(x) \in \mathrm{SO}(k)$ acts on D^k by identifying D^k with the unit ball in $\mathbb{R}^k$. This map extends to a homeomorphism $j(\theta)$ of $S^3 \times D^k \cup_{S^3 \times D^{k-1}} D^4 \times S^{k-1}$. Form the composition

$$\begin{aligned} S^{k+3} = S^3 \times D^k \cup_{S^3 \times D^{k-1}} D^4 \times S^{k-1} &\xrightarrow{j(\theta)} S^3 \times D^k \cup_{S^3 \times D^{k-1}} D^4 \times S^{k-1} \\ &\xrightarrow{\mathrm{col}} S^3 \times S^k \xrightarrow{\mathrm{proj}_1} S^k, \end{aligned}$$

where col is the collapse map which squashes $D^4 \times S^{k-1}$ and proj_1 is the projection onto the first factor. This gives an element of π_3^S, which is the image of θ under $J : \pi_3(\mathrm{SO}) \to \pi_3^S \cong \Omega_3^{\mathrm{fr}}$.

This J-homomorphism is onto [Adams 1966, Example 7.17], so that composing the framing in a neighbourhood D^3 of a point with the choice of map $\theta \in \pi_3(\mathrm{SO}) = [(D^3, \partial D^3), (\mathrm{SO}, *)]$ such that $-J(\theta) = [M] \in \Omega_3^{\mathrm{fr}}$ changes the class in Ω_3^{fr} as desired.

This shows the existence of a normal cobordism W'. To see that this is of degree one, note that the map to T^3 which extends over W' can be used to define a map to $T^3 \times I$, by defining a map $g : W' \to I$ such that $g(M_L) = \{0\}$ and $g(T^3) = \{1\}$. Now consider the commutative diagram

$$\begin{array}{ccc} H_4(W', \partial W'; \mathbb{Z}) & \longrightarrow & H_3(\partial W'; \mathbb{Z}) \\ \downarrow & & \downarrow \\ H_4(T^3 \times I, T^3 \times \{0, 1\}; \mathbb{Z}) & \longrightarrow & H_3(T^3 \times \{0, 1\}; \mathbb{Z}) \end{array}$$

Going right, then down, the fundamental class $[W', \partial W']$ maps to

$$(1, -1) \in H_3(T^3 \times \{0, 1\}; \mathbb{Z}) \cong \mathbb{Z} \oplus \mathbb{Z}.$$

By commutativity, the relative fundamental class $[W', \partial W']$ must map to a generator of $H_4(T^3 \times I, T^3 \times \{0, 1\}; \mathbb{Z})$. □

A Λ-homology equivalence is also an integral homology equivalence, by the following argument. By definition (see above the statement of Theorem 1.3), a Λ-homology equivalence induces an isomorphism on $H_1(-; \mathbb{Z})$. By duality, we also have an isomorphism on $H_2(-; \mathbb{Z})$. It remains to see that $f : M_L \to T^3$ is a degree-one map. The assumption that

$$f_* : H_*(M_L; \Lambda) \xrightarrow{\simeq} H_*(T^3; \Lambda)$$

is an isomorphism implies that the relative homology vanishes: $H_*(T^3, M_L; \Lambda) \cong 0$. The universal coefficient spectral sequence then implies that $H_*(T^3, M_L; \mathbb{Z}) \cong 0$ since all the E^2 terms $\operatorname{Tor}_p^{\Lambda}(H_q(T^3, M_L; \Lambda), \mathbb{Z})$ vanish. Therefore a Λ-homology equivalence as in Theorem 1.3 is a degree-one map.

Lemma 3.4 then establishes the existence of a choice of stable framing b on M_L such that there is a degree-one normal cobordism

$$(F' : W' \longrightarrow T^3 \times I, e')$$

between $(f : M_L \to T^3, b)$ and $(\mathrm{Id} : T^3 \to T^3, c)$. Choosing such a framing, we proceed to apply surgery theory to alter W' into a homology cobordism. Davis' observation [2006] was that the framing on W' is not an intrinsic part of the concordance problem, but rather necessary additional data which is required in order to be able to apply surgery theory. Without the information provided by the self-intersection form, it is not possible to obtain algebraic sufficient conditions which ensure that surgery can be performed. Nevertheless, as we shall see, there is a certain amount of freedom in the choice of framing data.

Before giving the proof of Theorem 1.3, we first give the definition of the Wall even-dimensional surgery obstruction groups, which we will use in the proof.

Definition 3.5 [Wall 1999, Chapter 5]. Let A be a ring with involution. A $(-1)^k$-*Hermitian sesquilinear quadratic form* on a free based A-module M is a $(-1)^k$-Hermitian sesquilinear form $\lambda : M \times M \to A$ together with a quadratic enhancement. A *quadratic enhancement* of a form $\lambda : M \times M \to A$ is a function $\mu : M \to A/\{a - (-1)^k \bar{a} \mid a \in A\}$ such that

(1) $\lambda(x, x) = \mu(x) + \overline{\mu(x)}$,

(2) $\mu(x + y) - \mu(x) - \mu(y) = \lambda(x, y)$,

(3) $\mu(ax) = a\mu(x)\bar{a}$,

for all $x, y \in M$ and for all $a \in A$.

A hyperbolic quadratic form is a direct sum of standard hyperbolic forms, where the standard hyperbolic form (H, χ, ν) is given by

$$\left(A \oplus A, \begin{pmatrix} 0 & 1 \\ (-1)^k & 0 \end{pmatrix}, \nu((1, 0)^T) = 0 = \nu((0, 1)^T)\right).$$

The even-dimensional surgery obstruction group $L_{2k}(A)$ is defined to be the Witt group of nonsingular $(-1)^k$-Hermitian sesquilinear quadratic forms on free based A-modules, where addition in the Witt group is by direct sum, and the equivalence class of the hyperbolic forms is the identity element, where the equivalence relation is as follows. Quadratic forms (M, λ, μ) and (M', λ', μ') are said to be equivalent if there are hyperbolic forms (H, χ, ν) and (H', χ', ν') such that there is an isomorphism of forms $(M \oplus H, \lambda \oplus \chi, \mu \oplus \nu) \cong (M' \oplus H', \lambda' \oplus \chi', \mu' \oplus \nu')$. This completes the definition of $L_{2k}(A)$.

For us, A will be the group ring $\mathbb{Z}[\pi]$ of some group π; initially π will be $\mathbb{Z}^3$, so that we take $A = \mathbb{Z}[\mathbb{Z}^3] = \Lambda$. We omit the definition of the odd-dimensional L-groups since they will only play a peripheral rôle in the proof of Theorem 1.3.

Proof of Theorem 1.3. First, do surgery below the middle dimension [ibid., Chapter 1] on (W', F', e') to create a normal cobordism $(F : W \to T^3 \times I, e)$ which is 2-connected, i.e., W is connected and $\pi_1(W) \cong \pi_1(T^3) \cong \mathbb{Z}^3$. The induced map $F_* : \pi_2(W) \to \pi_2(T^3 \times I)$ is automatically surjective since T^3 is aspherical.

The Wall surgery obstruction [ibid., Chapter 5] of the normal cobordism $(F : W \to T^3 \times I, e)$ is now defined in $L_4(\mathbb{Z}[\mathbb{Z}^3])$ to be given by the intersection form

$$\lambda_{W'} : H_2(W'; \Lambda) \times H_2(W'; \Lambda) \longrightarrow \Lambda,$$

together with the quadratic enhancement

$$\mu : H_2(W; \Lambda) \longrightarrow \mathbb{Z}[\mathbb{Z}^3]/\{a = \bar{a} \mid a \in \mathbb{Z}[\mathbb{Z}^3]\}$$

defined by counting the self-intersections of an immersion of a sphere $S^2 \looparrowright W$ representing an element of $H_2(W; \Lambda) \cong \pi_2(W)$, where the regular homotopy class of the immersion is fixed by the framing e to be the unique class of immersions for which the induced trivialisation of TS^2 extends over the null-homotopy of S^2 in T^3.

The fact that the homology of the boundary $H_j(M_L;\Lambda) \cong H_j(T^3;\Lambda)$ vanishes for $j = 1, 2$, is used crucially here to see that the intersection form λ_W is nonsingular, as observed by the surgeon in the "dialogue" of [Davis 2006].

By [Wall 1999, Proposition 13B.8], which is based on Shaneson's formula $L_n(\mathbb{Z}[\pi \times \mathbb{Z}]) \cong L_n(\mathbb{Z}[\pi]) \oplus L_{n-1}(\mathbb{Z}[\pi])$, when π has trivial Whitehead group [Shaneson 1969] we have that

$$\begin{aligned} L_4(\mathbb{Z}[\mathbb{Z}^3]) &\cong \bigoplus_{i=0}^{3} \binom{3}{i} L_{4-i}(\mathbb{Z}) \cong L_4(\mathbb{Z}) \oplus \bigoplus_3 L_3(\mathbb{Z}) \oplus \bigoplus_3 L_2(\mathbb{Z}) \oplus L_1(\mathbb{Z}) \\ &\cong L_0(\mathbb{Z}) \oplus \bigoplus_3 L_2(\mathbb{Z}), \end{aligned}$$

where the last isomorphism is by periodicity of the L-groups and the fact that the odd-dimensional simply connected L-groups vanish. The even-dimensional simply connected L-groups $L_{2k}(\mathbb{Z})$ are computed [Kervaire and Milnor 1963], when $k = 0 \bmod 2$, as

$$\begin{aligned} L_0(\mathbb{Z}) &\xrightarrow{\simeq} \mathbb{Z} \\ (M, \lambda, \mu) &\mapsto \sigma(\mathbb{R} \otimes_{\mathbb{Z}} M, \mathrm{Id} \otimes \lambda)/8, \end{aligned}$$

while for the dimensions where $k = 1 \bmod 2$ they are computed via

$$\begin{aligned} L_2(\mathbb{Z}) &\xrightarrow{\simeq} \mathbb{Z}_2 \\ (M, \lambda, \mu) &\mapsto \mathrm{Arf}(\mathbb{Z}_2 \otimes_{\mathbb{Z}} M, \mathrm{Id} \otimes \lambda, \mathrm{Id} \otimes \mu). \end{aligned}$$

We need to see that we can make further alterations to W in order to make the surgery obstruction vanish.

First, we take the connected sum with $-\sigma(W)/8$ copies of the E_8 manifold, namely the simply connected 4-manifold which is constructed by plumbing disc bundles $D^2 \times D^2$ according to the E_8 lattice. It turns out that the boundary of the resulting 4-manifold is the Poincaré homology sphere. One then caps off with the contractible topological 4-manifold whose boundary is the Poincaré homology sphere [Freedman and Quinn 1990, Corollary 9.3C]. This produces the E_8 manifold, a closed topological 4-manifold. It has a nonsingular intersection form, with a quadratic enhancement induced from a normal map to S^4, and its signature is 8. By a negative copy of this 4-manifold we of course mean the same manifold but with the opposite choice of orientation. By making such a modification to W, we obtain a new normal map, which by abuse of notation we again denote by (W, F, e), for which the obstruction in $L_0(\mathbb{Z})$ is trivial. Note that W still has fundamental group $\mathbb{Z}^3$ since $\pi_1(E_8 \text{ manifold}) \cong \{1\}$, and moreover ∂W is unchanged.

Next, we may need to alter W again, so that the three Arf invariant obstructions in $L_2(\mathbb{Z})$ vanish. For $i = 1, 2, 3$, define maps

$$\mathrm{qr}_i : T^3 \times I = S^1 \times S^1 \times S^1 \times I \longrightarrow S^1 \times S^1$$

which forget the i-th S^1 factor and the I factor. Perform a homotopy of F to ensure that $\mathrm{qr}_i \circ F$ is transverse to $* \in S^1 \times S^1$, and such that

$$F^{-1}(S^1 \times \{*\} \times \{*\} \times \{\partial I\}) \longrightarrow S^1 \times \{*\} \times \{*\} \times \{\partial I\}$$

is a homotopy equivalence (and similarly with the $*$ terms moved appropriately for $i = 2, 3$). This homotopy equivalence was already arranged in the proof of Lemma 3.4, when we saw that the elements of Ω_1^{fr} can be removed. Let S_i be the surfaces $(\mathrm{qr}_i \circ F)^{-1}(*)$; each surface has boundary ∂S_i given by the meridian μ_i and the corresponding S^1 factor of T^3.

Let $\mathrm{pr}_i : T^3 \times I = S^1 \times S^1 \times S^1 \times I \to S^1 \times I$ be the map which remembers the i-th S^1 factor and the I factor. Making F transverse to a point, $(\mathrm{pr}_i \circ F)^{-1}(*)$ is a surface $\Sigma_i \subset W$. Since $F(S_i \cap \Sigma_i)$ is a single point and F is of degree one, we can assume that S_i and Σ_i intersect in a single point. By choosing different points in the I factor, we can ensure that the Σ_i are all distinct.

Now, as in [Davis 2006], for each i with nonzero-surgery obstruction in the corresponding $L_2(\mathbb{Z})$ summand of $L_4(\mathbb{Z}[\mathbb{Z}^3])$, remove a neighbourhood $\Sigma_i \times D^2$ of Σ_i and replace it with $\Sigma_i \times \mathrm{cl}(S^1 \times S^1 \setminus D^2)$. That is, replace the D^2 factor with a punctured torus, but define the framing on the torus to be the framing which yields Arf invariant one, that is, the Lie framing on both S^1 factors. Since Σ_i is dual to S_i, this adds one to the Arf invariant of the element of $L_2(\mathbb{Z})$ represented by S_i, and so changes the Arf invariant one summands to having Arf invariant zero.

After these alterations we have a normal map $(G' : V' \to T^3 \times I, k')$, with vanishing surgery obstruction. Since the fundamental group $\mathbb{Z}^3$ is *good* in the sense of Freedman (polycyclic groups are good [Freedman and Quinn 1990, Theorem 5.1A]), the surgery sequence is exact in the topological category — see [ibid., Theorem 11.3A]. We can therefore find embedded two-spheres representing a half-basis for $\pi_2(G')$, perform surgery, and obtain a topological 4-manifold V which is homotopy equivalent to $T^3 \times I$; in particular, V is a homology cobordism between M_L and T^3.

Moreover, the following diagram commutes:

$$\begin{tikzcd}
\pi_1(M_L) \arrow[r] \arrow[d, "f_*"'] & \pi_1(V) \arrow[d, "\cong"] & \pi_1(T^3) \arrow[l] \arrow[d, "\mathrm{Id}"', "\cong"] \\
\pi_1(T^3) \arrow[r, "\cong"] & \pi_1(T^3 \times I) & \pi_1(T^3) \arrow[l, "\cong"']
\end{tikzcd}$$

Since the meridians μ_i of L are mapped to standard generators of $\pi_1(T^3)$, an easy diagram chase shows that the homotopy classes of the meridians are preserved in the homology cobordism V. $\square$

4. Construction of links and grope concordance

In this section we give constructions of certain links with a given Milnor invariant, and construct grope concordances, using the methods of [Cochran 1990] and [Cha 2014].

4A. ***Iterated Bing doubles with a prescribed Milnor invariant.*** Let I be a multi-index with nonrepeating indices with length $m := |I| \geq 2$. We describe a rooted binary tree $T(m)$ associated to $m \geq 2$, which has m leaves: the right subtree of the root just consists of a single vertex, and the left subtree $T^\dagger(m)$ is the complete binary tree of height $h(m) := \lceil \log_2(m-1) \rceil$ with the rightmost $2(m - h(m) - 1)$ pairs of leaves (and edges ending at these) removed. (By convention, a binary tree is always embedded in a plane with the root on the top.) That is, $T^\dagger(m)$ is a minimal height binary tree with $m-1$ leaves. For example, $T(m)$ for $m = 7$ is shown in Figure 1.

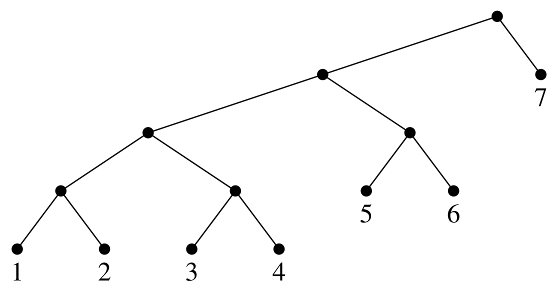

Figure 1. The tree $T(m)$ for $m = 7$, labelled with $I = 1234567$.

Following the proof of [Cochran 1990, Theorem 7.2], a rooted binary tree T describes a link with components corresponding to the leaves of T. First, a complete binary tree of height one is associated to a Hopf link. If T is obtained from T' by attaching two new leaves to a leaf v of T', then the link associated to T is obtained from that of T by Bing doubling the component corresponding to v.

Consider the link described by the tree $T(m)$. Labelling the leaves of $T(m)$ from left to right with the multi-index I (see Figure 1 for $I = 1234567$), the components of the link are ordered. We denote this ordered link by L_I. Then, by [Cochran 1990, Theorem 8.1], the link L_I has $\bar{\mu}_{L_I}(I) = \pm 1$ and $\bar{\mu}_{L_I}(I') = 0$ for $|I'| < |I|$.

4B. ***Satellite construction and grope concordance of links.*** To construct links which are grope concordant, we employ the method of [Cha 2014, Section 4]. We begin by giving the definition of grope concordance. The use of gropes in this context first appeared in [Cochran et al. 2003].

Definition 4.1 [Freedman and Teichner 1995]. A *grope* is a pair (2-complex, base circle) of a certain type described below. A grope has a height $h \in \mathbb{N}$. For $h = 1$ a grope is precisely a compact oriented surface Σ with a single boundary component which is the base circle. A grope of height $h + 1$ is defined inductively as follows: let $\{\alpha_i \mid i = 1, \dots, 2 \cdot \text{genus}\}$ be a standard symplectic basis of circles for Σ. Then a grope of height $h + 1$ is formed by attaching gropes of height h to each α_i along the base circles.

An annular grope is defined by replacing the bottom stage surface by a surface with two boundary components.

Definition 4.2 [Cha 2014, Definition 2.16]. Two m-component links L and L' in S^3 are *height n grope concordant* if there are m framed annular gropes G_i of height n, $i = 1, \dots, m$, disjointly embedded in $S^3 \times [0, 1]$, with the boundary of G_i the zero-framed i-th component of $L_i \subset S^3 \times \{0\}$ and $-L'_i \subset S^3 \times \{1\}$.

As mentioned in the introduction, we could also phrase our theorems in terms of Whitney towers, but for simplicity of exposition we stick to gropes. See [ibid., Section 2] for an exposition on gropes, Whitney towers, and n-solvable cobordisms (our Section 5 also contains a limited discussion of n-solvable cobordisms).

We recall that a *capped grope of height k* is a grope of height k together with 2-discs attached along each of the standard symplectic basis curves of the top-layer surfaces. The attached 2-discs are called *caps*, and the grope itself is called the *body*. We always assume that a capped grope embedded in a 4-manifold is framed.

We denote the exterior of a link L by X_L. If L is a link in S^3, η is an unknotted circle in S^3 disjoint from L, and K is a knot, then we denote the satellite link of L with axis η and companion K by $L(\eta, K)$; this is the image of L under the homeomorphism $X_\eta \cup_\partial X_K \xrightarrow{\approx} S^3$, where the gluing identifies the longitude of η with the meridian of K, and vice versa.

Following [ibid., Definition 4.2], we call (L, η) a *satellite configuration of height k* if L is a link in S^3, η is an unknotted circle in S^3 disjoint from L, and the 0-linking parallel of η in $X_\eta = X_\eta \times \{0\}$ bounds a capped grope of height k embedded in $X_\eta \times [0, 1]$ with body disjoint from $L \times [0, 1]$. The caps should be embedded in $X_\eta \times [0, 1]$ but may intersect $L \times [0, 1]$.

Lemma 4.3, stated below, describes how iterated satellite constructions using satellite configurations give us grope concordant links. The setup is as follows. Fix n. (To obtain Theorems 1.1 and 1.2, set $h = n + 2$.) Suppose that (L_0, η) is a satellite configuration of height $k \le n$. (Later we will use the link L_I described above as L_0.) Suppose that (K_i, α_i) is a satellite configuration of height one, with K_i a slice knot, for $i = 0, \dots, n - k - 1$. Let J_0^j be the connected sum of N_j copies of the knot described in [Cochran and Teichner 2007, Figure 3.6], where $\{N_j\}$ is an increasing sequence of integers which will be specified later. (Indeed, these will be given in terms of the Cheeger–Gromov bound on the ρ-invariants and, for the links of Theorem 1.1, in terms of the Kneser–Haken bound on the number of disjoint nonparallel incompressible surfaces. See Section 5, just before the proof of Theorem 5.3, and Section 4D, just before Lemma 4.7.) Define $J_i^j := K_{i-1}(\alpha_{i-1}, J_{i-1}^j)$ inductively for $i = 1, \ldots, n-k$. Finally define $L_j := L_0(\eta, J_{n-k}^j)$.

Lemma 4.3 [Cha 2014, Proposition 4.7]. *The link L_j is height $n+2$ grope concordant to L_0 for all j.*

Proof. The same as the proof of [loc. cit.], except that L_0 replaces the Hopf link in the last sentence. □

The following observation on the satellite construction is useful.

Lemma 4.4. *If $L' = L(\eta, K)$ is obtained from L by a satellite construction, then L and L' have the same Milnor $\bar{\mu}$-invariants.*

Proof. It is well known that a satellite construction $L' = L(\eta, K)$ comes with an integral homology equivalence $f : (X_{L'}, \partial X_{L'}) \to (X_L, \partial X_L)$ which restricts to a homeomorphism on the boundary preserving longitudes and meridians (see, e.g., [Cha 2010, proof of Proposition 4.8; Cha and Orr 2013, Lemma 5.3]). As in [Cha et al. 2012, Lemma 2.1], by [Stallings 1965] it follows that f induces an isomorphism $\pi_1(X_L)/\pi_1(X_L)_q \cong \pi_1(X_{L'})/\pi_1(X_{L'})_q$ that preserves the classes of meridians and longitudes for any q, and consequently L and L' have identical $\bar{\mu}$-invariants. □

4C. ***Satellite configuration of iterated Bing doubles.*** Now we consider again the link L_I described in Section 4A. Recall that $k(m) := \lfloor \log_2(m-1) \rfloor$, where $m = |I|$. Let η be the zero-framed longitude of the component of L_I labelled with m, namely the component of the original Hopf link that is never Bing doubled in the construction of L_I.

Lemma 4.5. (1) *The pair (L_I, η) is a satellite configuration of height $k(m)$.*

(2) *The curve η is nonzero in $\pi_1(L_I)/\pi_1(L_I)_m$.*

(3) *For any knot K, the link $L_I(\eta, K)$ has zero-surgery manifold homeomorphic to the zero-surgery manifold of L_I.*

We remark that Lemma 4.5(2) will be used in Section 5.

Proof. Denote $L := L_I$ for this proof.

(1) We go back to the construction of L, and construct the grope as we construct L. We begin with the Hopf link (i.e., $m = 2$), and the curve η as a longitude of L_2.

We also begin with a thickened cap $D^2 \times [-1, 1]$, such that $\partial D^2 \times \{0\} = \eta$. This intersects the other component of the Hopf link in a single point.

Every time a component K is Bing doubled in the construction of L, we arrange that one of the clasps lies in $D^2 \times [-1, 1]$, and then replace the thickened cap that intersected K with a genus-one capped surface with a single boundary component, whose body surface misses the new Bing doubled components, and such that each cap intersects one of the two new components. See Figure 2, which is somewhat reminiscent of a figure in [Freedman and Quinn 1990, Chapter 2.1].

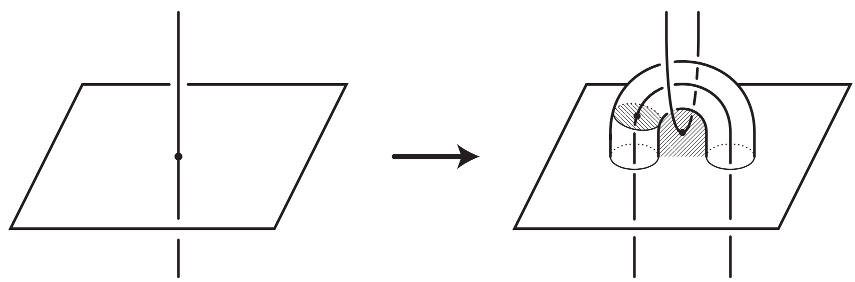

Figure 2. Replacing a cap with a capped surface.

Since a complete binary tree of height $k(m)$ can be embedded in $T(m)$, we obtain a symmetric embedded capped grope of the required height, with the body lying in the link exterior X_L and the caps intersecting the link transversely.

(2) The nonvanishing of the Milnor invariant $\bar{\mu}_L(I)$ implies that all of the longitudes of L are nontrivial in $\pi_1(X_L)/\pi_1(X_L)_{|I|}$.

(3) A Kirby diagram for the 3-manifold M_L given by zero-framed surgery on L can be produced by putting a 0 next to every component of L. If we perform a satellite construction with pattern K and with η as axis, this is equivalent to tying all the strands of L which intersect a disc D, whose boundary is η, in the knot K, with framing zero. In other words, replace the trivial string link in $D \times [0, 1]$ with the string link obtained by taking suitably many parallel copies of K.

But we can make a crossing change of these parallel copies of K at will, by performing handle slides, sliding the parallel strands over the zero-framed 2-handle attached along the component parallel to η. This gives a Kirby presentation of a homeomorphic 3-manifold.

By making sufficiently many such crossing changes/handle slides, all the parallel strands which the satellite construction ties in the knot K can be unknotted, recovering the link L. Thus the zero-surgery manifolds of the satellite link and the original link are homeomorphic. It is easy to see that the homotopy classes of the meridians of L are preserved under such homeomorphisms. □

Now, let $n \geq k(|I|) = k(m)$. Let L_j be the links obtained by the construction just before Lemma 4.3, using our (L_I, η) as (L_0, η), and using the Stevedore satellite configuration described in [Cha 2014, Figure 6], which for the reader's convenience is shown in Figure 3, as the (K_i, α_i). Then by Lemma 4.3 and Lemma 4.5(1), the links L_j are height $n+2$ grope concordant to the link $L_0 = L_I$.

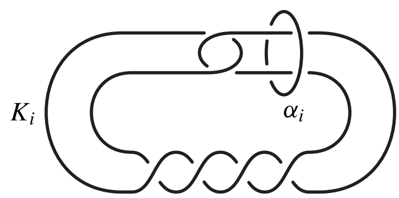

Figure 3. Stevedore satellite configuration (K_i, α_i).

Lemma 4.4 shows that the links L_j satisfy Theorem 1.2(1). They also satisfy Theorem 1.2(2) by Lemma 4.5(3). We have also proved, in Lemma 4.3, the first part of Theorem 1.2(3): the links L_j are mutually height $n+2$ grope concordant. The second part of Theorem 1.2(3), namely the failure of the links to be pairwise height $n+3$ grope concordant, will be shown in Section 5.

4D. ***Examples with nonhomeomorphic zero-surgery manifolds.*** In order to produce examples satisfying Theorem 1.1(1), we alter the construction of Sections 4B and 4C to give examples with nonhomeomorphic zero-surgery manifolds. We consider the case of $m = 3$ and $I = 123$ only. Then the link $L := L_I$ described in Section 4A is the Borromean rings. Let η be the simple closed curve in $S^3 \smallsetminus L$ shown in Figure 4; x, y, and z denote the components of L.

The pair (L, η) also has two of the properties stated in Lemma 4.5, for $m = 3$:

Lemma 4.6. (1) *The pair (L, η) is a satellite configuration of height one.*

(2) *In $\pi_1(X_L)$, $\eta = [x, y][[x, y], x]$, where x, y, and z are the Wirtinger generators corresponding to the dotted arcs in Figure 4. Also, η is nontrivial in $\pi_1(X_L)/\pi_1(X_L)_3$.*

Here $[a, b]$ denotes the commutator $aba^{-1}b^{-1}$.

Proof. (1) Tubing the obvious disc bounded by η along the components of L that intersect it, we obtain a genus-two surface V with boundary η which is shown in Figure 5. This is the body of the desired capped grope. The whole capped grope is the body taken together with the four caps shown in Figure 5 as shaded discs.

(2) The claim that $\eta = [x, y][[x, y], x]$ follows from a straightforward computation in terms of the Wirtinger generators, reading undercrossings of η starting from the dot on η in Figure 4. Since L has vanishing linking number, due to Milnor [1957]

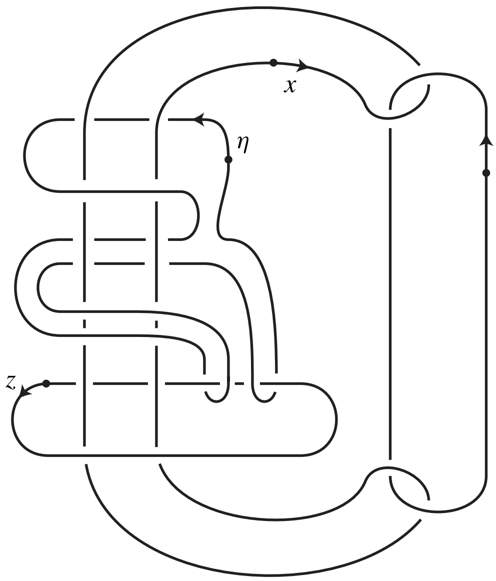

Figure 4. A satellite configuration on the Borromean rings.

(see also [Stallings 1965]), $\pi_1(X_L)/\pi_1(X_L)_3$ is isomorphic to F/F_3, where F is the free group generated by x, y, and z. Consequently, $[[x, y], x] \in \pi_1(X_L)_3$ and $[x, y] \notin \pi_1(X_L)_3$. From this the second conclusion follows. $\square$

As in Section 4C, we apply the construction described just before Lemma 4.3, using our (L, η) as the seed link (L_0, η) and using the Stevedore satellite configuration described in [Cha 2014, Figure 6] (see our Figure 3) as (K_i, α_i) for $i = 0, \ldots, n-2$ as above. Let the resulting links be the L_j. Then by Lemma 4.3, the L_j are height

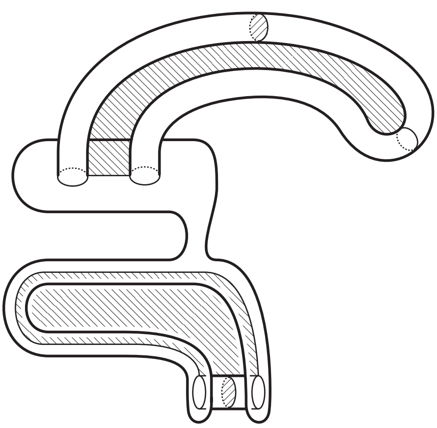

Figure 5. The capped grope bounded by η.

$n+2$ grope concordant to the Borromean rings L, so these satisfy the first part of Theorem 1.1(3). The second part of Theorem 1.1(3), on the failure of the links to be pairwise height $n+3$ grope concordant, will be shown in Section 5.

Furthermore, the links L_j satisfies the hypothesis of Theorem 1.3. First note that since our satellite operation does not change the knot type of the components, L_j has unknotted components. In particular, the Arf invariants of the components vanish. Recall from the proof of Lemma 4.4 that there is a homology equivalence $f : X_{L_j} \to X_{L_0}$ obtained from the satellite construction $L_j = L_0(\eta, J^j_{n-1})$; indeed, f is obtained by gluing the identity map of $X_{L_0 \sqcup \eta}$ with the standard homology equivalence

$$(X_{J^j_{n-1}}, \partial X_{J^j_{n-1}}) \longrightarrow (S^1 \times D^2, S^1 \times S^1)$$

along $S^1 \times S^1$. Since our curve $\eta \subset S^3 - L_0$ lies in the commutator subgroup of $\pi_1(S^3 - L_0)$, f is indeed a Λ-homology equivalence $X_{L_j} \to X_{L_0}$, by a Mayer–Vietoris argument. Filling it in with 3 solid tori, we obtain a Λ-homology equivalence $M_{L_j} \to T^3 = M_{L_0}$ as desired. Therefore, by applying Theorem 1.3, it follows that the links L_j satisfy Theorem 1.1(2). We need to confirm that the L_j satisfy Theorem 1.1(1), namely, that the M_{L_j} are not homeomorphic. The underlying idea is as follows. Recall that L_j is defined by a satellite construction, starting with a knot J^j_0. In many cases, the JSJ pieces of the exterior of J^j_0 become parts of the JSJ decomposition of M_{L_j}, so that the M_{L_j} have distinct JSJ decompositions. Since a complete proof of this seems to require complicated arguments (a technical issue is that an essential torus might not be parallel to a JSJ torus, because of Seifert fibred pieces), we will present a simpler argument using only the number of incompressible tori; this is enough for our purpose.

We need the following. The Kneser–Haken finiteness theorem [Haken 1961] states that for each 3-manifold M, there is a bound, say $C_{KH}(M)$, on the number of disjoint pairwise nonparallel incompressible surfaces that can be embedded in M. Recall that the knot J^j_0 used in the construction of the link L_j is a connected sum of N_j knots, where $\{N_j\}$ was an increasing sequence to be specified (see the paragraph before Lemma 4.3). Here is the first requirement on the N_j: we choose the N_j inductively in such a way that $N_j > \max\{C_{KH}(M_{L_k}) \mid k = 0, 1, \dots, j-1\}$.

Lemma 4.7. *The zero-surgery manifolds M_{L_i} and M_{L_j} are not homeomorphic for $i \neq j$.*

Proof. Recall that $M_{L_0} = M_L$ is the 3-torus T^3. Consider $Y := M_L \smallsetminus \nu(\eta)$, where $\nu(\eta)$ is an open tubular neighbourhood of η. For notational convenience, denote the exterior of J^j_{n-1} by $X := X_{J^j_{n-1}}$. The 3-manifold M_{L_j} is obtained by glueing Y and X along their boundaries. Let $T = \partial Y = \partial X$ be the common boundary torus. Note that M_{L_0} can also be described in the same way, using $J^0_{n-1} :=$ unknot; in this case, the torus T is compressible in M_{L_0} since X is a solid torus.

Claim. *For $j \geq 1$, the torus T is incompressible in Y.*

Using the claim, we will show that the 3-manifolds M_{L_j} are not pairwise homeomorphic. Suppose $j \geq 1$. Since the knot J^j_{n-1} is obtained from an iterated satellite construction with the first-stage knot J^j_0 a connected sum of N_j nontrivial knots, the exterior X of J^j_{n-1} has at least N_j incompressible tori, including the boundary T. Since $M_{L_j} = Y \cup_T X$ and T is incompressible in Y, it follows that there are N_j nonparallel incompressible tori in M_{L_j}. For any $k < j$, since $N_j > C_{KH}(M_{L_k})$, it follows that M_{L_j} is not homeomorphic to M_{L_k}.

Now, to complete the proof, we will verify the claim. If there is an essential curve on T which bounds a disc in Y, then it must be a zero-linking longitude, say η', of η, since the meridian of η is a generator of $H_1(Y \smallsetminus \eta) = \mathbb{Z}^4$. By the following lemma, we have a contradiction. □

Lemma 4.8. *The class of η' is nontrivial in the fundamental group $\pi_1(Y \smallsetminus \eta)$.*

Proof. We consider a Wirtinger presentation of $\pi_1(Y \smallsetminus \eta)$ given as follows: it has 24 generators, denoted by $x_1, \dots, x_{24}$, associated to arcs in Figure 4. Here $(x_1, \dots, x_{10})$, (x_{11}, x_{12}), $(x_{13}, \dots, x_{16})$, and $(x_{17}, \dots, x_{24})$ are those associated to the arcs of the components x, y, z, and η, respectively. In each component, the arc with a dot on it is the first one, and other arcs are ordered along the orientation. There are 27 relators:

$$\begin{gathered}
x_1x_{11}\bar{x}_1\bar{x}_{12},\ x_{11}x_1\bar{x}_{11}\bar{x}_2,\ x_2x_{18}\bar{x}_2\bar{x}_{19},\ x_{19}x_2\bar{x}_{19}\bar{x}_3,\ x_3x_{20}\bar{x}_3\bar{x}_{21},\\
x_3x_{23}\bar{x}_3\bar{x}_{22},\ x_{22}x_4\bar{x}_{22}\bar{x}_3,\ x_{21}x_4\bar{x}_{21}\bar{x}_5,\ x_5x_{16}\bar{x}_5\bar{x}_{13},\ x_{13}x_5\bar{x}_{13}\bar{x}_6,\\
x_{11}x_7\bar{x}_{11}\bar{x}_6,\ x_7x_{11}\bar{x}_7\bar{x}_{12},\ x_{13}x_8\bar{x}_{13}\bar{x}_7,\ x_8x_{16}\bar{x}_8\bar{x}_{15},\ x_{21}x_9\bar{x}_{21}\bar{x}_8,\\
x_{22}x_9\bar{x}_{22}\bar{x}_{10},\ x_{10}x_{23}\bar{x}_{10}\bar{x}_{24},\ x_{10}x_{20}\bar{x}_{10}\bar{x}_{19},\ x_{19}x_1\bar{x}_{19}\bar{x}_{10},\ x_1x_{18}\bar{x}_1\bar{x}_{17},\\
x_{15}x_{22}\bar{x}_{15}\bar{x}_{21},\ x_{22}x_{15}\bar{x}_{22}\bar{x}_{14},\ x_{24}x_{13}\bar{x}_{24}\bar{x}_{14},\ x_{13}x_{24}\bar{x}_{13}\bar{x}_{17},\\
\bar{x}_{11}\bar{x}_{19}x_{22}\bar{x}_{21}\bar{x}_{13}x_{11}x_{13}x_{21}\bar{x}_{22}x_{19},\ \bar{x}_1x_7,\ \bar{x}_{24}x_{22}x_8\bar{x}_5.
\end{gathered}$$

Indeed, the first 24 are the standard Wirtinger relators for the 4-component link $L \sqcup \eta$ (thus one of these may be omitted), and the last 3 relators arise from the zero-surgery performed along L. It is straightforward to read off the curve η':

$$\eta' = x_1\bar{x}_2x_{10}\bar{x}_3x_{15}x_3\bar{x}_{10}\bar{x}_{13}.$$

We define a representation $\rho : \pi_1(Y \smallsetminus \eta) \to \mathrm{SL}(2, \mathbb{Z}_5)$ by mapping the above 24 generators, respectively, to:

$$\begin{gathered}
\begin{bmatrix} 0 & 4 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 1 & 4 \end{bmatrix}, \begin{bmatrix} 0 & 4 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} 4 & 4 \\ 0 & 4 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 1 & 4 \end{bmatrix}, \begin{bmatrix} 0 & 4 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 4 & 4 \end{bmatrix},\\
\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 4 & 0 \end{bmatrix}, \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 4 & 3 \\ 2 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 2 \\ 2 & 0 \end{bmatrix}, \begin{bmatrix} 4 & 2 \\ 3 & 3 \end{bmatrix},\\
\begin{bmatrix} 0 & 1 \\ 4 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 4 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 3 & 4 \\ 4 & 4 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 4 & 0 \end{bmatrix}, \begin{bmatrix} 4 & 4 \\ 4 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 4 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.
\end{gathered}$$

It can be verified that all the relators are sent to the identity, by a straightforward computation. (We found the representation ρ using a computer program.) Also, we have that

$$\rho(\eta') = \begin{bmatrix} 3 & 1 \\ 4 & 0 \end{bmatrix}$$

is not the identity. This completes the proof. □

5. Grope concordance and amenable signatures

In this section we show that the links described in Sections 4C and 4D are not height $n+3$ grope concordant by using amenable signature obstructions from [Cha 2014]. In fact, the amenable signatures we use are obstructions to being n-solvably cobordant, which is a relative analogue for manifolds with boundary, or bordered manifolds, of the notion of n-solvability of [Cochran et al. 2003]. For our purpose it suffices to consider the case of link exteriors; an *n-solvable cobordism* between the exteriors X and X' of two links with the same number of components is a 4-manifold W with $\partial W = X \cup_\partial -X'$ satisfying the conditions described in [Cha 2014, Definition 2.8], where the boundary tori of X and X' are identified along the zero framing. Since we do not use the defining condition right now, instead of spelling it out here, we begin with its relationship to grope concordance. The following theorem originates from [Cochran et al. 2003, Theorem 8.11], and was given in our context in [Cha 2014].

Theorem 5.1 [ibid., Theorems 2.16 and 2.13, and Remark 2.11]. *If two links are height $n+2$ grope concordant, then their exteriors are n-solvably cobordant as bordered 3-manifolds.*

As our key ingredient to detect nonsolvably cobordant 3-manifolds and therefore non-grope-concordant links, we will use the Amenable Signature Theorem, which was first introduced in [Cha and Orr 2012] for homology cobordism of closed 3-manifolds and then generalised to n-solvable cobordisms of bordered 3-manifolds in [Cha 2014]. We state a special case which will be sufficient for our purpose. For a closed 3-manifold M and a homomorphism $\phi : \pi_1(M) \to G$, denote the von Neumann–Cheeger–Gromov ρ-invariant by $\rho^{(2)}(M, \phi) \in \mathbb{R}$. See, e.g., [Cochran et al. 2003, Section 5] as well as [Chang and Weinberger 2003; Harvey 2008; Cha 2008; Cha and Orr 2012] for definitions and useful properties of $\rho^{(2)}(M, \phi)$. Precise references for the properties that we need will be recalled as we go along.

Theorem 5.2 (A special case of [Cha 2014, Amenable Signature Theorem 3.2]). *Suppose W is an $(n+1)$-solvable cobordism between two bordered 3-manifolds X and X', and G admits a subnormal series*

$$G = G_0 \supset G_1 \supset \cdots \supset G_n \supset G_{n+1} = \{e\}$$

with each quotient G_i/G_{i+1} torsion-free abelian. Then $\rho^{(2)}(X \cup_\partial -X', \phi) = 0$ for any $\phi : \pi_1(X \cup_\partial -X') \to G$ which factors through $\pi_1(W)$.

Recall that in our construction of the links L_j, the knot J_0^j was the connected sum of N_j copies of Cochran and Teichner's knot, say J. Now we proceed to specify the integers N_j. Denote by $\rho^{(2)}(K) := \int_{S^1} \sigma_K(\omega)\, d\omega$ the integral of the Levine–Tristram signature function over the circle normalised to length one. We have $\rho^{(2)}(J_0^j) = N_j \rho^{(2)}(J) = 4N_j/3$ by additivity under connected sum and [Cochran and Teichner 2007, Lemma 4.5]. Due to Cheeger and Gromov [1985], for any closed 3-manifold Y there is a constant $C_Y > 0$ such that $|\rho^{(2)}(Y, \psi)| < C_Y$ for any ψ. From now on we abbreviate $\ell := n - k(m)$. Define

$$R := C_{X_{L_0} \cup_\partial -X_{L_0}} + 2 \sum_{i=0}^{\ell-1} C_{M_{K_i}}.$$

We choose the large integers N_j inductively in such a way that

$$N_j > 3R/4 + \max\{N_k \mid k < j\}.$$

Then we have

$$\rho^{(2)}(J_0^j) > R + \rho^{(2)}(J_0^k)$$

whenever $j > k$. For Theorem 1.1, we make these choices so that the condition in the preamble to Lemma 4.7 relating to the Kneser–Haken bound is simultaneously satisfied.

Now we start the proof that our links L_j are not height $n+3$ grope concordant to one another. Let X and X' be the exteriors of L_j and L_k, respectively. To distinguish them in the notation, we denote the axis curve η in X by η_j, and we denote the corresponding axis curve in X' by η_k.

Recall that $m = |I|$ and that $k(m) = \lfloor \log_2(m-1) \rfloor$. Also note that $k(m) + 1 = \lceil \log_2(m) \rceil$. By Theorem 5.1, it suffices to show the following:

Theorem 5.3. *For $n \geq k(m)$, the bordered 3-manifolds X and X' are not $(n+1)$-solvably cobordant when $j \neq k$.*

By Theorem 5.1, it then follows that our links L_j and L_k are not height $n+3$ grope concordant when $j \neq k$.

Proof. The proof proceeds almost identically to that of [Cha 2014, Theorem 4.8], which combines the Amenable Signature Theorem of that reference with a higher-order Blanchfield duality argument for a certain 4-dimensional cobordism introduced in [Cochran et al. 2009] (see our W_0 below). So we will give an outline for our case and discuss differences from [Cha 2014, Theorem 4.8].

Suppose W is an $(n+1)$-solvable cobordism with $\partial W = X \cup_\partial -X'$. Similarly to [ibid., Section 4.3] (see the paragraph entitled “Cobordism associated to an iterated satellite construction"), we consider a cobordism V with

$$\partial V = M_{J_0^j} \sqcup -M_{J_0^k} \sqcup M_{K_0} \sqcup -M'_{K_0} \sqcup \cdots \sqcup M_{K_{\ell-1}} \sqcup -M'_{K_{\ell-1}} \\ \sqcup (X_{L_0} \cup_\partial -X_{L_0}) \sqcup -(X \cup_\partial -X')$$

which is built by stacking cobordisms associated to satellite constructions [Cochran et al. 2009, p. 1429], where M'_{K_i} is a copy of M_{K_i}, and then construct a cobordism W_0 with

$$\partial W_0 = M_{J_0^j} \sqcup -M_{J_0^k} \sqcup M_{K_0} \sqcup -M'_{K_0} \sqcup \cdots \sqcup M_{K_{\ell-1}} \sqcup -M'_{K_{\ell-1}} \sqcup (X_{L_0} \cup_\partial -X_{L_0})$$

by attaching V to W along $X \cup_\partial -X'$. We omit the detailed construction of V and W_0 but state a couple of useful facts which can be verified as in [Cha 2014, Section 4.3]. Let $\{\mathcal{P}^r G\}$ be the rational derived series of a group G, i.e., $\mathcal{P}^0 G := G$ and $\mathcal{P}^{r+1} G$ is the kernel of $\mathcal{P}^r G \to H_1(\mathcal{P}^r G; \mathbb{Q})$. Let ϕ_0 be the quotient map $\pi_1(W_0) \to G := \pi_1(W_0)/\mathcal{P}^{n+1}\pi_1(W_0)$. Also we denote by ϕ_0 the restrictions of ϕ_0 to the components of ∂W_0 and to $W \subset W_0$, as an abuse of notation. Then we have the following facts:

(1) $$\rho^{(2)}(M_{J_0^j}, \phi_0) - \rho^{(2)}(M_{J_0^k}, \phi_0) + \rho^{(2)}(X_{L_0} \cup_\partial -X_{L_0}, \phi_0) \\ + \sum_{i=0}^{\ell-1} \rho^{(2)}(M_{K_i}, \phi_0) - \sum_{i=0}^{\ell-1} \rho^{(2)}(M'_{K_i}, \phi_0) = \rho^{(2)}(X \cup_\partial -X', \phi_0).$$

(2) The image of the meridian of J_0^j in $M_{J_0^j} \subset \partial W_0$ under ϕ_0 is a nontrivial element in the torsion-free abelian subgroup $\mathcal{P}^n\pi_1(W)/\mathcal{P}^{n+1}\pi_1(W)$ of G. Similarly for k instead of j.

The proof of (1) is completely identical to that given in [Cha 2014, Section 4.3] (see the paragraphs entitled "Cobordism associated to an iterated satellite construction" and "Applications of Amenable Signature Theorem"): briefly, the $\rho^{(2)}$-invariant of ∂W_0, which is the left-hand side of (1), is equal to the L^2-signature defect of $W_0 = V \cup_{X \cup_\partial -X'} W$ (this is a standard fact from index theory, or can be taken as the definition of $\rho^{(2)}$). It turns out that V has no contribution to the L^2-signature defect, by [Cochran et al. 2009, Lemma 2.4]. So the left-hand side of (1) is equal to the L^2-signature defect of W, which is the $\rho^{(2)}$-invariant of ∂W, namely the right-hand side of (1).

The proof of (2) is almost identical to that given in [Cha 2014, Theorem 4.10]. Only the following change is required: in the initial step of the inductive argument in that result, it was shown that the image of (a parallel copy of) $\eta \subset X \subset \partial W$ is nontrivial under the quotient map $\pi_1(W) \to \pi_1(W)/\mathcal{P}^2\pi_1(W)$ (see the fourth paragraph of the proof) using a Blanchfield duality argument.

In our case, instead we use Lemma 5.4 below, which is a generalisation of [Cha et al. 2012, Lemma 3.5], to show that the image of η is nontrivial in the quotient $\pi_1(W)/\mathcal{P}^{(k(m)+1)}\pi_1(W)$. The argument used in Lemma 5.4 is essentially an application of Dwyer's theorem.

Lemma 5.4. *If W is an n-solvable cobordism between two link exteriors (or, more generally, bordered 3-manifolds) X and X', then the inclusions induce isomorphisms*

$$\pi_1(X)/\pi_1(X)_q \cong \pi_1(W)/\pi_1(W)_q \cong \pi_1(X')/\pi_1(X')_q$$

for $q \le 2^n + 1$.

Proof. Recall Dwyer's theorem [1975]: if $f : X \to Y$ induces an isomorphism $H_1(X;\mathbb{Z}) \cong H_1(Y;\mathbb{Z})$ and an epimorphism

$$H_2(X;\mathbb{Z}) \longrightarrow H_2(Y;\mathbb{Z})/\operatorname{Im}\{H_2(Y;\mathbb{Z}[\pi_1(W)/\pi_1(W)_q]) \to H_2(Y;\mathbb{Z})\},$$

then f induces an isomorphism $\pi_1(X)_q/\pi_1(X)_{q+1} \cong \pi_1(Y)_q/\pi_1(Y)_{q+1}$.

In our case, by the definition of an n-solvable cobordism [Cha 2014, Definition 2.8], we have $H_1(X;\mathbb{Z}) \cong H_1(W;\mathbb{Z}) \cong H_1(X';\mathbb{Z})$. Also, by the same definition, there are elements $\ell_1, \dots, \ell_r, d_1, \dots, d_r$ lying in $H_2(W;\mathbb{Z}[\pi_1(W)/\pi_1(W)^{(n)}])$ such that the images of ℓ_i and d_j generate $H_2(W;\mathbb{Z})$. Since $\pi_1(W)^{(n)}$ is contained in $\pi_1(W)_{2^n}$, the H_2 condition of Dwyer's theorem is satisfied. Therefore, it follows that

$$\pi_1(X)_q/\pi_1(X)_{q+1} \cong \pi_1(W)_q/\pi_1(W)_{q+1} \cong \pi_1(X')_q/\pi_1(X')_{q+1}$$

for $q \le 2^n$ by Dwyer's theorem. From this the desired conclusion follows by the five lemma. ☐

Recall that Lemma 4.5(2) implies that $\eta \subset X$ represents a nontrivial element in $\pi_1(X)/\pi_1(X)_m$. Since the above isomorphisms preserve longitudes (and meridians), $\eta_j \subset X$ represents a nontrivial element in $\pi_1(W)/\pi_1(W)_m$. Since L_j has vanishing Milnor invariants of length less than $|I| = m$, we have $\pi_1(X)/\pi_1(X)_m \cong F/F_m$, where F is the free group with rank m, by [Milnor 1957, Theorem 4]. Consequently $\pi_1(W)/\pi_1(W)_m$ is torsion-free.

We note that for any group π, we have $\pi^{(k(q)+1)} = \pi^{(\lceil \log_2(q) \rceil)} \subseteq \pi_q$. Therefore there is a quotient map $\pi_1(W)/\pi_1(W)^{(k(m)+1)} \to \pi_1(W)/\pi_1(W)_m$, and this map factors through $\pi_1(W)/\mathcal{P}^{(k(m)+1)}\pi_1(W)$ by the definition of $\mathcal{P}^{(k(m)+1)}$ and the fact that the codomain is torsion-free. Since η_j is nontrivial in $\pi_1(W)/\pi_1(W)_m$, η_j is also nontrivial in $\pi_1(W)/\mathcal{P}^{(k(m)+1)}\pi_1(W)$. By replacing j with k and X with X' we obtain the corresponding fact for η_k in X'.

To complete the proof of Theorem 5.3, we proceed as in [Cha 2014, Section 4.3]. Observe that for the normal subgroups $G_i := \mathcal{P}^i\pi_1(W_0)/\mathcal{P}^{n+1}\pi_1(W_0)$ of our G, the quotient G_i/G_{i+1} is torsion-free abelian. So by Amenable Signature Theorem 5.2 we have $\rho^{(2)}(X \cup_\partial -X') = 0$. Since the curve η_j represents a nontrivial element in a torsion-free abelian normal subgroup of G, the image of $\pi_1(M_{J_0^j})$ in G under ϕ_0 is the infinite cyclic group. By L^2-induction (see, e.g., [Cheeger and Gromov 1985, page 8(2.3); Cochran et al. 2003, Proposition 5.13]) and [Cochran et al. 2004, Proposition 5.1], we have $\rho^{(2)}(M_{J_0^j}, \phi_0) = \rho^{(2)}(J_0^j)$, and similarly for J_0^k. Now, combining these two facts with (1), we obtain

$$\rho^{(2)}(J_0^j) - \rho^{(2)}(J_0^k) + \rho^{(2)}(X_{L_0} \cup_\partial -X_{L_0}, \phi_0) + \sum_{i=0}^{\ell-1} \rho^{(2)}(M_{K_i}, \phi_0) - \sum_{i=0}^{\ell-1} \rho^{(2)}(M'_{K_i}, \phi_0) = 0. \tag{3}$$

Recall that

$$\left|\rho^{(2)}(X_{L_0} \cup_\partial -X_{L_0}, \phi_0) + \sum_{i=0}^{\ell-1} \rho^{(2)}(M_{K_i}, \phi_0) - \sum_{i=0}^{\ell-1} \rho^{(2)}(M'_{K_i}, \phi_0)\right| < R := C_{X_{L_0} \cup_\partial -X_{L_0}} + 2\sum_{i=0}^{\ell-1} C_{M_{K_i}},$$

and in the preamble to Theorem 5.3, we chose N_j so that $\left|\rho^{(2)}(J_0^k) - \rho^{(2)}(J_0^j)\right| > R$ whenever $k \neq j$. Therefore (3) implies that $j = k$. Thus the existence of the $(n+1)$-solvable cobordism W implies that $j = k$, which is the contrapositive of the desired statement. $\square$

Acknowledgements

We would like to thank Stefan Friedl for many conversations: our first examples were motivated by joint work with him. We also thank Jim Davis for discussions relating to the proof of his theorem on Hopf links, and Prudence Heck, Charles Livingston, Kent Orr and Vladimir Turaev for helpful discussions. Finally we thank the referee for valuable comments and for being impressively expeditious.

Part of this work was completed while the first author was a visitor at Indiana University in Bloomington and the second author was a visitor at the Max Planck Institute for Mathematics in Bonn. We would like to thank these institutions for their hospitality.

The first author was partially supported by National Research Foundation of Korea grants 2013067043 and 2013053914. The second author gratefully acknowledges an AMS Simons travel grant which aided his travel to Bonn.

References

[Adams 1966] J. F. Adams, "On the groups $J(X)$, IV", *Topology* **5** (1966), 21–71. MR 33 #6628 Zbl 0145.19902

[Cappell and Shaneson 1974] S. E. Cappell and J. L. Shaneson, "The codimension two placement problem and homology equivalent manifolds", *Ann. of Math.* (2) **99** (1974), 277–348. MR 49 #3978 Zbl 0279.57011

[Cappell and Shaneson 1980] S. E. Cappell and J. L. Shaneson, "Link cobordism", *Comment. Math. Helv.* **55**:1 (1980), 20–49. MR 81j:57011 Zbl 0444.57008

[Casson and Gordon 1978] A. J. Casson and C. M. Gordon, "On slice knots in dimension three", pp. 39–53 in *Algebraic and geometric topology* (Stanford, CA, 1976), vol. 2, edited by R. J. Milgram, Proc. Sympos. Pure Math. **32**, Amer. Math. Soc., Providence, RI, 1978. MR 81g:57003 Zbl 0394.57008

[Casson and Gordon 1986] A. J. Casson and C. M. Gordon, "Cobordism of classical knots", pp. 181–199 in *À la recherche de la topologie perdue*, edited by L. Guillou and A. Marin, Progr. Math. **62**, Birkhäuser, Boston, 1986. MR 900252

[Cha 2008] J. C. Cha, "Topological minimal genus and L^2-signatures", *Algebr. Geom. Topol.* **8**:2 (2008), 885–909. MR 2010f:57024 Zbl 1162.57016

[Cha 2010] J. C. Cha, "Link concordance, homology cobordism, and Hirzebruch-type defects from iterated p-covers", *J. Eur. Math. Soc.* (*JEMS*) **12**:3 (2010), 555–610. MR 2011f:57003 Zbl 1195.57012

[Cha 2014] J. C. Cha, "Symmetric Whitney tower cobordism for bordered 3-manifolds and links", *Trans. Amer. Math. Soc.* **366**:6 (2014), 3241–3273. MR 3180746

[Cha and Ko 1999] J. C. Cha and K. H. Ko, "On equivariant slice knots", *Proc. Amer. Math. Soc.* **127**:7 (1999), 2175–2182. MR 2000a:57006 Zbl 0959.57003

[Cha and Orr 2012] J. C. Cha and K. E. Orr, "L^2-signatures, homology localization, and amenable groups", *Comm. Pure Appl. Math.* **65**:6 (2012), 790–832. MR 2903800 Zbl 1256.57018

[Cha and Orr 2013] J. C. Cha and K. E. Orr, "Hidden torsion, 3-manifolds, and homology cobordism", *J. Topol.* **6**:2 (2013), 490–512. MR 3065182

[Cha et al. 2012] J. C. Cha, S. Friedl, and M. Powell, "Concordance of links with identical Alexander invariants", preprint, 2012. arXiv 1212.2924

[Chang and Weinberger 2003] S. Chang and S. Weinberger, "On invariants of Hirzebruch and Cheeger–Gromov", *Geom. Topol.* **7** (2003), 311–319. MR 2004c:57052 Zbl 1037.57028

[Cheeger and Gromov 1985] J. Cheeger and M. Gromov, "Bounds on the von Neumann dimension of L^2-cohomology and the Gauss–Bonnet theorem for open manifolds", *J. Differential Geom.* **21**:1 (1985), 1–34. MR 87d:58136 Zbl 0614.53034

[Cochran 1985] T. D. Cochran, "Concordance invariance of coefficients of Conway's link polynomial", *Invent. Math.* **82**:3 (1985), 527–541. MR 87c:57002 Zbl 0589.57005

[Cochran 1990] T. D. Cochran, "Derivatives of links: Milnor's concordance invariants and Massey's products", *Mem. Amer. Math. Soc.* **84**:427 (1990). MR 91c:57005 Zbl 0705.57003

[Cochran and Kim 2008] T. D. Cochran and T. Kim, "Higher-order Alexander invariants and filtrations of the knot concordance group", *Trans. Amer. Math. Soc.* **360**:3 (2008), 1407–1441. MR 2008m:57008 Zbl 1132.57005

[Cochran and Teichner 2007] T. D. Cochran and P. Teichner, "Knot concordance and von Neumann ρ-invariants", *Duke Math. J.* **137**:2 (2007), 337–379. MR 2008f:57005 Zbl 1186.57006

[Cochran et al. 2003] T. D. Cochran, K. E. Orr, and P. Teichner, "Knot concordance, Whitney towers and L^2-signatures", *Ann. of Math.* (2) **157**:2 (2003), 433–519. MR 2004i:57003 Zbl 1044.57001

[Cochran et al. 2004] T. D. Cochran, K. E. Orr, and P. Teichner, "Structure in the classical knot concordance group", *Comment. Math. Helv.* **79**:1 (2004), 105–123. MR 2004k:57005 Zbl 1061.57008

[Cochran et al. 2009] T. D. Cochran, S. Harvey, and C. Leidy, "Knot concordance and higher-order Blanchfield duality", *Geom. Topol.* **13**:3 (2009), 1419–1482. MR 2009m:57006 Zbl 1175.57004

[Cochran et al. 2013] T. D. Cochran, B. D. Franklin, M. Hedden, and P. D. Horn, "Knot concordance and homology cobordism", *Proc. Amer. Math. Soc.* **141**:6 (2013), 2193–2208. MR 3034445 Zbl 1276.57007 arXiv 1102.5730

[Davis 2006] J. F. Davis, "A two component link with Alexander polynomial one is concordant to the Hopf link", *Math. Proc. Cambridge Philos. Soc.* **140**:2 (2006), 265–268. MR 2006k:57010 Zbl 1090.57005

[Dwyer 1975] W. G. Dwyer, "Homology, Massey products and maps between groups", *J. Pure Appl. Algebra* **6**:2 (1975), 177–190. MR 52 #6710 Zbl 0338.20057

[Freedman and Quinn 1990] M. H. Freedman and F. Quinn, *Topology of* 4*-manifolds*, Princeton Mathematical Series **39**, Princeton University Press, 1990. MR 94b:57021 Zbl 0705.57001

[Freedman and Teichner 1995] M. H. Freedman and P. Teichner, "4-manifold topology, I: Subexponential groups", *Invent. Math.* **122**:3 (1995), 509–529. MR 96k:57015 Zbl 0857.57017

[Friedl and Powell 2011] S. Friedl and M. Powell, "Links not concordant to the Hopf link", *Math. Proc. Cambridge Philos. Soc.* (2011). To appear. arXiv 1105.2773

[Friedl and Teichner 2005] S. Friedl and P. Teichner, "New topologically slice knots", *Geom. Topol.* **9** (2005), 2129–2158. MR 2007b:57007 Zbl 1120.57004

[Haken 1961] W. Haken, "Ein Verfahren zur Aufspaltung einer 3-Mannigfaltigkeit in irreduzible 3-Mannigfaltigkeiten", *Math. Z.* **76** (1961), 427–467. MR 25 #4519c Zbl 0111.18803

[Harvey 2008] S. L. Harvey, "Homology cobordism invariants and the Cochran–Orr–Teichner filtration of the link concordance group", *Geom. Topol.* **12**:1 (2008), 387–430. MR 2009d:57045 Zbl 1157.57006

[Hillman 2002] J. Hillman, *Algebraic invariants of links*, Series on Knots and Everything **32**, World Scientific, River Edge, NJ, 2002. MR 2003k:57014 Zbl 1007.57001

[Hilton 1955] P. J. Hilton, "On the homotopy groups of the union of spheres", *J. London Math. Soc.* **30** (1955), 154–172. MR 16,847d Zbl 0064.17301

[Kawauchi 1978] A. Kawauchi, "On the Alexander polynomials of cobordant links", *Osaka J. Math.* **15**:1 (1978), 151–159. MR 58 #7599 Zbl 0401.57013

[Kawauchi 1996] A. Kawauchi, *A survey of knot theory*, Birkhäuser, Basel, 1996. MR 97k:57011 Zbl 0861.57001

[Kervaire and Milnor 1963] M. A. Kervaire and J. W. Milnor, "Groups of homotopy spheres, I", *Ann. of Math.* (2) **77** (1963), 504–537. MR 26 #5584 Zbl 0115.40505

[Le Dimet 1988] J.-Y. Le Dimet, *Cobordisme d'enlacements de disques*, Mém. Soc. Math. France (N.S.) **32**, Soc. Math. France, Paris, 1988. MR 90e:57046 Zbl 0666.57015

[Levine 1966] J. Levine, "Polynomial invariants of knots of codimension two", *Ann. of Math.* (2) **84** (1966), 537–554. MR 34 #808 Zbl 0196.55905

[Levine 1969a] J. Levine, "Invariants of knot cobordism", *Invent. Math.* **8** (1969), 98–110, 355. MR 40 #6563 Zbl 0179.52401

[Levine 1969b] J. Levine, "Knot cobordism groups in codimension two", *Comment. Math. Helv.* **44** (1969), 229–244. MR 39 #7618 Zbl 0176.22101

[Levine 1977] J. Levine, "Knot modules, I", *Trans. Amer. Math. Soc.* **229** (1977), 1–50. MR 57 #1503 Zbl 0653.57012

[Livingston 1983] C. Livingston, "Knots which are not concordant to their reverses", *Quart. J. Math. Oxford Ser.* (2) **34**:135 (1983), 323–328. MR 85d:57005 Zbl 0537.57003

[Milnor 1957] J. Milnor, "Isotopy of links", pp. 280–306 in *Algebraic geometry and topology: a symposium in honor of S. Lefschetz*, edited by R. H. Fox et al., Princeton Math. Series **12**, Princeton University Press, 1957. MR 19,1070c Zbl 0080.16901

[Nakagawa 1978] Y. Nakagawa, "On the Alexander polynomials of slice links", *Osaka J. Math.* **15**:1 (1978), 161–182. MR 58 #7600 Zbl 0401.57014

[Powell 2012] M. Powell, "A second order algebraic knot concordance group", *Algebr. Geom. Topol.* **12**:2 (2012), 685–751. MR 2914616 Zbl 1244.57023

[Shaneson 1969] J. L. Shaneson, "Wall's surgery obstruction groups for $G \times Z$", *Ann. of Math.* (2) **90** (1969), 296–334. MR 39 #7614 Zbl 0182.57303

[Stallings 1965] J. Stallings, "Homology and central series of groups", *J. Algebra* **2** (1965), 170–181. MR 31 #232 Zbl 0135.05201

[Vogel 1978] P. Vogel, "Localization of spaces with respect to a class of maps", preprint, Université de Nantes, 1978, Available at http://www.maths.ed.ac.uk/~aar/papers/vogelnantes.pdf.

[Wall 1999] C. T. C. Wall, *Surgery on compact manifolds*, 2nd ed., edited by A. A. Ranicki, Math. Surveys Monogr. **69**, Amer. Math. Soc., Providence, RI, 1999. MR 2000a:57089 Zbl 0935.57003

Received October 28, 2013. Revised December 6, 2013.

JAE CHOON CHA
DEPARTMENT OF MATHEMATICS
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY
GYUNGBUK
POHANG 790–784
SOUTH KOREA

jccha@postech.ac.kr

and

SCHOOL OF MATHEMATICS
KOREA INSTITUTE FOR ADVANCED STUDY
SEOUL 130–722
SOUTH KOREA

MARK POWELL
DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
RAWLES HALL
831 EAST 3RD STREET
BLOOMINGTON, IN 47405
UNITED STATES

macp@indiana.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.272.35

CERTAIN SELF-HOMOTOPY EQUIVALENCES ON WEDGE PRODUCTS OF MOORE SPACES

HO WON CHOI AND KEE YOUNG LEE

For a based 1-connected finite CW-complex X, let $\mathcal{E}(X)$ denote the group of homotopy classes of self-homotopy equivalences on X, and $\mathcal{E}_\sharp^{\dim+r}(X)$ the subgroup of $\mathcal{E}(X)$ of homotopy classes of self-homotopy equivalences on X that induce the identity homomorphism on the homotopy groups of X in dimensions $\leq \dim X + r$. For two given Moore spaces $M_1 = M(Z_q, n+1)$ and $M_2 = M(Z_p, n)$ with $n \geq 5$, we investigate the subsets of $[M_1, M_2]$ and $[M_2, M_1]$ consisting of homotopy classes of maps that induce the trivial homomorphism between the homotopy groups of M_1 and those of M_2 in dimensions $\leq \dim X + r$. Using the results of this investigation, we completely determine the subgroups $\mathcal{E}_\sharp^{\dim+r}\big(M(Z_q, n+1) \vee M(Z_p, n)\big)$, where p and q are positive integers, for $n \geq 5$ and $r = 0, 1$.

1. Introduction

If X and Y are based topological spaces, let $[X, Y]$ denote the set of homotopy classes of based maps from X to Y, let $\mathcal{E}(X)$ denote the subset of $[X, X]$ that consists of homotopy classes of self-homotopy equivalences of X and let $\mathcal{E}_\sharp^{\dim+r}(X)$ denote the set of homotopy classes of self-homotopy equivalences that induce the identity on the homotopy groups of X in dimensions at most $\dim X + r$. Then, $\mathcal{E}(X)$ is a group with a group operation given by the composition of homotopy classes, and $\mathcal{E}_\sharp^{\dim+r}(X)$ is a subgroup of $\mathcal{E}(X)$. The group $\mathcal{E}(X)$ and certain natural subgroups including $\mathcal{E}_\sharp^{\dim+r}(X)$ are fundamental objects in homotopy theory and have been studied extensively. For a survey of the known results and applications of $\mathcal{E}(X)$, see [Arkowitz 1990].

When G is an abelian group, we let $M(G, n)$ denote the Moore space, that is, the space with G as a single nonvanishing homology group at n-level. Also, in this case, $M(G, n)$ is a simply connected space. We note that if $n \geq 3$, then $M(G, n)$ is characterized by

$$\widetilde{H}_i(M(G, n)) \cong \begin{cases} G & \text{if } i = n, \\ 0 & \text{if } i \neq n. \end{cases}$$

This work was supported by a Korea University Grant.
MSC2010: primary 55P10, 55Q05, 55Q20; secondary 55Q40, 55Q52.
Keywords: self-homotopy equivalence, Moore space, homotopy group.

Let $C(G, n)$ denote the co-Moore space of type (G, n) defined by

$$\widetilde{H}^i(C(G, n)) \cong \begin{cases} G & \text{if } i = n, \\ 0 & \text{if } i \neq n. \end{cases}$$

If G is a finitely generated abelian group and $G = F \oplus T$, where F is a free abelian group of rank r and T is a finite group, then $M(G, n) = M(F, n) \vee M(T, n)$ and $C(G, n) = M(F, n) \vee M(T, n-1)$ for $n \geq 3$.

Arkowitz and Maruyama [1998] showed that $\mathcal{E}_\sharp^{\dim}(M(G, n)) \cong \oplus^{(r+s)s} Z_2$ and $\mathcal{E}_\sharp^{\dim+1}(M(G, n)) = 1$ for $n > 3$, where r is the rank of G and s is the number of 2-torsion summands in G. Moreover, they completely determined $\mathcal{E}_\sharp^{\dim}(C(G, n))$ for $n \geq 3$ by means of 2×2 matrices, where G is a finitely generated abelian group.

Jeong [2010] computed the groups $\mathcal{E}_\sharp^{\dim}(Y)$ for $Y = M(Z_p, n+1) \vee M(Z_p, n)$, $n \geq 5$ as follows:

$$\mathcal{E}_\sharp^{\dim}(Y) \cong \begin{cases} Z_p & \text{if } p \text{ is odd}, \\ Z_2 \oplus Z_2 & \text{if } p \equiv 2 \pmod 4, \\ Z_2 \oplus Z_2 \oplus Z_2 & \text{if } p \equiv 0 \pmod 4. \end{cases}$$

In this paper we study the self-homotopy equivalences on the wedge product $X = M(Z_q, n+1) \vee M(Z_p, n)$ for $n \geq 5$, where p and q are positive integers. For two given Moore spaces $M_1 = M(Z_q, n+1)$ and $M_2 = M(Z_p, n)$, we compute $[M_1, M_2]$ and $[M_2, M_1]$ and find their generators. Moreover, we investigate the subset of $[M_1, M_2]$ or $[M_2, M_1]$ that consists of elements whose induced homomorphisms are trivial between the homotopy groups of M_1 and those of M_2 in dimensions at most $\dim X + r$ with $r = 0, 1$. Using these results, we completely determine the groups $\mathcal{E}_\sharp^{\dim+r}(X)$ for $r = 0, 1$. As a result, we obtain Table 1 and the following:

$$\mathcal{E}_\sharp^{\dim+1}(X) \cong \begin{cases} 1 & \text{if } q \text{ is odd or } p \text{ is odd } (d = 1), \\ Z_d & \text{if } q \text{ is odd or } p \text{ is odd } (d \neq 1), \\ Z_{d/2} \oplus Z_2 & \text{if } p \equiv 0 \pmod 4 \text{ and } (p, 24) = 4 \text{ or } 12 \ (d \neq 1), \\ Z_{d/2} & \text{if } p \equiv 0 \pmod 4 \text{ and } (p, 24) = 8 \text{ or } 24 \ (d \neq 1), \\ Z_{d/2} & \text{if } q \equiv 2,\ p \equiv 2 \pmod 4, \\ Z_{d/2} \oplus Z_2 & \text{if } q \equiv 0,\ p \equiv 2 \pmod 4, \end{cases}$$

where d is the greatest common divisor of p and q.

The space X is neither a Moore space nor a co-Moore space but is characterized by finite homology groups and cohomology groups. That is,

$$\widetilde{H}_i(X) \cong \begin{cases} Z_p & \text{if } i = n, \\ Z_q & \text{if } i = n+1, \\ 0 & \text{otherwise}, \end{cases}$$

	q is odd $d=1$	q is odd $d\neq 1$	$q\equiv 2 \pmod 4$	$q\equiv 0 \pmod 4$
p is odd ($d=1$)	1	$\cdot$	Z_2	Z_2
p is odd ($d\neq 1$)	$\cdot$	Z_d	$Z_2\oplus Z_d$	$Z_2\oplus Z_d$
$p\equiv 2 \pmod 4$	1	Z_d	$Z_2\oplus Z_{d/2}\oplus Z_2$	$Z_2\oplus Z_{d/2}\oplus Z_4$
$p\equiv 0 \pmod 4$	1	Z_d	$Z_2\oplus Z_{d/2}\oplus Z_2\oplus Z_2$	$Z_2\oplus Z_{d/2}\oplus Z_2\oplus Z_2$

Table 1. Isomorphism class of the groups $\mathcal{E}_\sharp^{\dim}(X)$.

and

$$\widetilde{H}^i(X,\pi)\cong\begin{cases}\mathrm{Hom}(Z_p,\pi) & \text{if } i=n,\\ \mathrm{Ext}(Z_p,\pi)\oplus\mathrm{Hom}(Z_q,\pi) & \text{if } i=n+1,\\ \mathrm{Ext}(Z_q,\pi) & \text{if } i=n+2,\\ 0 & \text{otherwise.}\end{cases}$$

From this perspective, X is an interesting space for studying self-homotopy equivalences.

Throughout this paper, all topological spaces are based and have the based homotopy type of a finite l-connected CW-complex. All maps and homotopies will preserve base points. For the spaces X and Y, we denote by $[X, Y]$ the set of homotopy classes of maps from X to Y. We do not distinguish between the notation of a map $X\to Y$ and that of its homotopy class in $[X, Y]$. If a group G is generated by a set $\{a_1,\dots,a_n\}$, then we denote the group by $G\{a_1,\dots,a_2\}$ or $G=\langle a_1,\dots,a_n\rangle$.

2. Preliminaries

Let X be a space. Then, we denote by SX the suspension of X and by S^nX the iterated suspension defined by $S^nX=S(S^{n-1}X)$. Let $f:A\to B$ be a map and let $C_f=B\cup_f CA$ be the mapping cone of f. Then, we have a Puppe sequence [1958] for f,

$$A\xrightarrow{f}B\xrightarrow{i}C_f\xrightarrow{\pi}SA\xrightarrow{Sf}SB\xrightarrow{Si}SC_f\xrightarrow{S\pi}S^2A\xrightarrow{S^2f}S^2B\longrightarrow\cdots,$$

such that the following sequence is exact for any space X:

$$\cdots\longrightarrow[SC_f,X]\xrightarrow{S\pi^*}[SB,X]\xrightarrow{Sf^*}[SA,X]\xrightarrow{\pi^*}[C_f,X]\xrightarrow{i^*}[B,X]\xrightarrow{f^*}[A,X],$$

where S^nf is a suspension map induced by f.

If A is m-connected and B is n-connected, then we have the following exact sequence for any CW-complex Y with dimension at most $m+n$ as a dual sequence of the above sequence [Blakers and Massey 1952]:

$$[Y, A] \xrightarrow{f_*} [Y, B] \xrightarrow{i_*} [Y, C_f] \xrightarrow{\pi_*} [Y, SA] \xrightarrow{Sf_*} [Y, SB] \longrightarrow \cdots .$$

Both sequences will be called *the exact sequences associated with the cofibration*

$$B \to C_f \to SA.$$

Proposition 2.1 [Arkowitz and Maruyama 1998]. *If X is $(k-1)$-connected, Y is $(l-1)$-connected, $k, l \geq 2$ and* $\dim P \leq k+l-1$, *then the projections $X \vee Y \to X$ and $X \vee Y \to Y$ induce a bijection*

$$[P, X \vee Y] \to [P, X] \oplus [P, Y].$$

Proposition 2.1 is a consequence of [Spanier 1966, p. 405] since the inclusion $X \vee Y \to X \times Y$ is a $(k+l-1)$-equivalence.

Next, we consider abelian groups G_1 and G_2 and Moore spaces $M_1 = M(G_1, n_1)$ and $M_2 = M(G_2, n_2)$. Let $X = M_1 \vee M_2$. We denote by $i_j : M_j \to X$ the inclusion and by $p_j : X \to M_j$ the projection, where $j = 1, 2$. If $f : X \to X$, then we define $f_{jk} : M_k \to M_j$ by $f_{jk} = p_j f i_k$ for $j, k = 1, 2$.

If $f : X \to Y$ is a map, then $f_{\sharp n} : \pi_n(X) \to \pi_n(Y)$ denotes the induced homomorphism in dimension n.

Proposition 2.2 [Arkowitz and Maruyama 1998]. *The function θ that assigns to each $f \in [X, X]$ the 2×2 matrix*

$$\theta(f) = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix},$$

where $f_{jk} \in [M_k, M_j]$, is a bijection. In addition:

(1) $\theta(f+g) = \theta(f) + \theta(g)$, *so θ is an isomorphism* $[X, X] \to \bigoplus_{j,k=1,2} [M_k, M_j]$.

(2) $\theta(fg) = \theta(f)\theta(g)$, *where fg denotes composition in $[X, X]$ and $\theta(f)\theta(g)$ denotes matrix multiplication.*

(3) *If $\alpha_r : \pi_r(M_1) \oplus \pi_r(M_2) \to \pi_r(M_1 \vee M_2)$ is the homomorphism induced by the inclusions and $\beta_r : \pi_r(M_1 \vee M_2) \to \pi_r(M_1) \oplus \pi_r(M_2)$ the homomorphism induced by the projections respectively, then*

$$\beta_r f_{\sharp r} \alpha_r(x, y) = \big(f_{11\,\sharp r}(x) + f_{12\,\sharp r}(y),\, f_{21\,\sharp r}(x) + f_{22\,\sharp r}(y)\big)$$

for $x \in \pi_r(M_1)$ and $y \in \pi_r(M_2)$.

Proposition 2.3 [Araki and Toda 1965]. (1) $\pi_n(M(Z_q, n)) \cong Z_q$ *for all q.*

(2) $\pi_{n+1}(M(Z_q, n)) \cong \begin{cases} 0 & \text{if } q \text{ is odd}, \\ Z_2 & \text{if } q \text{ is even}. \end{cases}$

(3) $\pi_{n+2}(M(Z_q, n)) \cong \begin{cases} 0 & \textit{if } q \textit{ is odd}, \\ Z_4 & \textit{if } q \equiv 2 \pmod 4, \\ Z_2 \oplus Z_2 & \textit{if } q \equiv 0 \pmod 4). \end{cases}$

(4) $\pi_{n+3}(M(Z_q, n)) \cong \begin{cases} Z_{(q,24)} & \textit{if } q \textit{ is odd}, \\ Z_{(q,24)} \oplus Z_2 & \textit{if } q \textit{ is even}. \end{cases}$

The generators of $[S^{n+i}, S^n]$ can be summarized thus [Toda 1962]:

	$i < 0$	$i = 0$	$i = 1$	$i = 2$	$i = 3$	$i = 4, 5$
$[S^{n+i}, S^n]$	0	Z	Z_2	Z_2	Z_{24}	0
Generator		ι	η	η^2	ν	0

Proposition 2.4 [Araki and Toda 1965].

(1) $[M(Z_q, n), S^n] \cong \begin{cases} 0 & \textit{if } q \textit{ is odd}, \\ Z_2 & \textit{if } q \textit{ is even}. \end{cases}$

(2) $[M(Z_q, n+1), S^n] \cong \begin{cases} 0 & \textit{if } q \textit{ is odd}, \\ Z_4 & \textit{if } q \equiv 2 \pmod 4), \\ Z_2 \oplus Z_2 & \textit{if } q \equiv 0 \pmod 4). \end{cases}$

Proposition 2.5 [Arkowitz and Maruyama 1998]. *For the Moore space $X = M(G, n)$:*

(1) $\mathcal{E}_\sharp^{\dim}(X) \cong \oplus^{(r+s)s} Z_2$, *where r is the rank of G and s is the number of 2-torsion summands in G.*

(2) $\mathcal{E}_\sharp^{\dim+1}(X) \cong 1$ *if $n > 3$.*

Proposition 2.6 (universal coefficient theorem for homotopy groups with coefficients [Hilton 1965]). *There is an exact sequence*

$$0 \to \mathrm{Ext}(G, \pi_{n+1}(X)) \to \pi_n(G; X) \to \mathrm{Hom}(G, \pi_n(X)) \to 0,$$

where $\pi_n(G; X)$, the n-th homotopy group of X with coefficients in G, is given by $\pi_n(G; X) = [M(G, n), X]$, where $M(G, n)$ is a Moore space.

3. Generators of the sets of homotopy classes on Moore spaces

In this section, we find generators of homotopy groups of Moore spaces and the sets of homotopy classes between two Moore spaces. Let

$$M_1 = M(Z_q, n+1) = S^{n+1} \cup_q e^{n+2} \quad \text{and} \quad M_2 = M(Z_p, n) = S^n \cup_p e^{n+1},$$

with $p, q \geq 1$. Then, there are two mapping cone sequences

$$S^{n+1} \xrightarrow{q\iota_1} S^{n+1} \xrightarrow{i_1} S^{n+1} \cup_q e^{n+2} \xrightarrow{\pi_1} S^{n+2} \xrightarrow{q\iota_1} S^{n+2}$$

and

$$S^n \xrightarrow{p_{\iota_2}} S^n \xrightarrow{i_2} S^n \cup_p e^{n+1} \xrightarrow{\pi_2} S^{n+1} \xrightarrow{p_{\iota_2}} S^{n+1},$$

where p_{ι_2} and q_{ι_1} are maps with degree p and q respectively.

Remark 3.1. We find generators of $\pi_m(M(Z_r, n))$, for $n \le m \le n+2$.

Recall that $\pi_n(M(Z_r, n)) \cong Z_r$. From the mapping cone sequence

$$S^n \xrightarrow{r_\iota} S^n \xrightarrow{i} M(Z_r, n) \xrightarrow{\pi} S^{n+1} \xrightarrow{r_\iota} S^{n+1},$$

we obtain the long exact sequence

$$\pi_n(S^n) \xrightarrow{r_{\iota\sharp}} \pi_n(S^n) \xrightarrow{i_\sharp} \pi_n(M(Z_r, n)) \xrightarrow{\pi_\sharp} \pi_n(S^{n+1}) \xrightarrow{r_{\iota\sharp}} \pi_n(S^{n+1}).$$

By the results in [Toda 1962], we have the sequence

$$Z\{\iota\} \xrightarrow{r_{\iota\sharp}} Z\{\iota\} \xrightarrow{i_\sharp} \pi_n(M(Z_r, n)) \longrightarrow 0,$$

so $i_\sharp$ is surjective. Thus, $\pi_n(M(Z_r, n)) \cong Z\{\iota\}/\operatorname{Im}(r_{\iota\sharp})$. Let $i_\sharp(\iota) = i$. Then, we can take i as a generator of $\pi_n(M(Z_r, n))$.

Next, we find a generator of $\pi_{n+1}(M(Z_r, n))$. There are two cases according to the parity of the positive integer r. If r is odd, then $\pi_{n+1}(M(Z_r, n))$ is trivial. If r is even, then we can take $i_\sharp(\eta)$ as a generator of $\pi_{n+1}(M(Z_r, n))$, where η is the generator of $\pi_{n+1}(S^n)$.

Finally, we find a generator of $\pi_{n+2}(M(Z_r, n))$. Consider the exact sequence

$$\pi_{n+2}(S^n) \xrightarrow{r_{\iota\sharp}} \pi_{n+2}(S^n) \xrightarrow{i_\sharp} \pi_{n+2}(M(Z_r, n)) \xrightarrow{\pi_\sharp} \pi_{n+2}(S^{n+1}) \xrightarrow{r_{\iota\sharp}} \pi_{n+2}(S^{n+1}).$$

Then by the results in [Toda 1962], we have the exact sequence

$$Z_2\{\eta^2\} \xrightarrow{r_{\iota\sharp}} Z_2\{\eta^2\} \xrightarrow{i_\sharp} \pi_{n+2}(M(Z_r, n)) \xrightarrow{\pi_\sharp} Z_2\{\eta\} \xrightarrow{q_{\iota\sharp}} Z_2\{\eta\}.$$

Since r is an even number, we obtain the exact sequence

$$0 \longrightarrow Z_2\{\eta^2\} \xrightarrow{i_\sharp} \pi_{n+2}(M(Z_r, n)) \xrightarrow{\pi_\sharp} Z_2\{\eta\} \longrightarrow 0.$$

If $r \equiv 2 \pmod 4$, then $\pi_{n+2}(M(Z_r, n)) \cong Z_4\{\bar{\eta}\}$ such that $i_\sharp(\eta^2) = 2\bar{\eta}$ and $\pi_\sharp(\bar{\eta}) = \eta$. On the other hand, if $r \equiv 0 \pmod 4$, then $\pi_{n+2}(M(Z_r, n)) \cong Z_2 \oplus Z_2\{\eta_1, \eta_2\}$ such that $i_\sharp(\eta^2) = \eta_1$ and $\pi_\sharp(\eta_2) = \eta$.

By Remark 3.1, it follows that

$$\pi_{n+1}(M_1) \cong Z_q\{i_1\}, \qquad \pi_n(M_2) \cong Z_p\{i_2\},$$
$$\pi_{n+2}(M_1) \cong Z_2\{i_{1\sharp}(\eta)\}, \qquad \pi_{n+1}(M_2) \cong Z_2\{i_{2\sharp}(\eta)\}.$$

Moreover, $\pi_{n+2}(M_2) \cong Z_4\{\bar{\eta}\}$ or $\pi_{n+2}(M_2) \cong Z_2 \oplus Z_2\{\eta_1, \eta_2\}$.

Lemma 3.2. *Let p and q be positive integers and (p,q) be the greatest common divisor of p and q. Consequently, if $(p,q)=d\neq 1$, then $[M_2, M_1]\cong Z_d\{\pi_2^*(i_1)\}$ and if $(p,q)=1$, then $[M_2, M_1]\cong 0$.*

Proof. Consider the mapping cone sequence of M_2,

$$S^n \xrightarrow{p\iota_2} S^n \xrightarrow{i_2} S^n\cup_p e^{n+1} \xrightarrow{\pi_2} S^{n+1} \xrightarrow{p\iota_2} S^{n+1}.$$

This sequence induces the following exact sequence:

$$\pi_{n+1}(M_1) \xrightarrow{p\iota_2 *} \pi_{n+1}(M_1) \xrightarrow{\pi_2 *} [M_2, M_1] \xrightarrow{i_2 *} \pi_n(M_1) \xrightarrow{p\iota_2 *} \pi_n(M_1).$$

Since $\pi_{n+1}(M_1)\cong Z_q\{i_1\}$ and $\pi_n(M_1)\cong 0$, the exact sequence above becomes

$$Z_q\{i_1\} \xrightarrow{p_{\iota_2}^*} Z_q\{i_1\} \xrightarrow{\pi_2^*} [M_2, M_1] \longrightarrow 0.$$

If $(p,q)=1$, the first $p_{\iota_2}^*$ is an isomorphism, so $[M_2, M_1]\cong 0$. Let $(p,q)=d\neq 1$. Then, since π_2^* is surjective and $p_{\iota_2}^*(i_1)=pi_1$, we have

$$[M_2, M_1]=\operatorname{im}\pi_2^*\cong Z_q\{i_1\}/\operatorname{im}p_{\iota_2}^*\cong Z_d\{\pi_2^*(i_1)\}.$$ □

Lemma 3.3. *If p or q is odd, then $[M_1, M_2]\cong 0$.*

Proof. Consider the mapping cone sequence of M_1,

$$S^{n+1} \xrightarrow{q_{\iota_1}} S^{n+1} \xrightarrow{i_1} S^{n+1}\cup_q e^{n+2} \xrightarrow{\pi_1} S^{n+2} \xrightarrow{q_{\iota_1}} S^{n+2}.$$

Then, we have the exact sequence

$$\pi_{n+2}(M_2) \xrightarrow{q_{\iota_1}^*} \pi_{n+2}(M_2) \xrightarrow{\pi_1^*} [M_1, M_2] \xrightarrow{i_1^*} \pi_{n+1}(M_2) \xrightarrow{q_{\iota_1}^*} \pi_{n+1}(M_2).$$

Let $p\equiv 2 \pmod 4$ and let q be odd. Then, since $\pi_{n+1}(M_2)\cong Z_2$ and $\pi_{n+2}(M_2)\cong Z_4$, we have the sequence

$$Z_4 \xrightarrow{q_{\iota_1}^*} Z_4 \xrightarrow{\pi_1^*} [M_1, M_2] \xrightarrow{i_1^*} Z_2 \xrightarrow{q_{\iota_1}^*} Z_2.$$

Furthermore, since $(q,4)=1$ and $(q,2)=1$, each $q_{\iota_1}^*$ is an isomorphism. Thus we have the exact sequence

$$0\to [M_1, M_2]\to 0.$$

Therefore, $[M_1, M_2]\cong 0$.

In the case where $p\equiv 0 \pmod 4$ and q is odd, we can give a similar proof.

Next, let p be odd. Since $\pi_{n+1}(M_2)$ and $\pi_{n+2}(M_2)$ are trivial groups, so is $[M_1, M_2]$ by exactness. □

Let p and q be even. From the exact sequences associated with the cofibrations $S^{n+1} \to M_1 \to S^{n+2}$ and $S^n \to M_2 \to S^{n+1}$, we have the following commutative diagram with exact rows and columns:

$$\begin{tikzcd}
{[S^{n+2},S^n]} \arrow[r, "q_{\iota_1}^*"] \arrow[d, "p_{\iota_2 *}"] & {[S^{n+2},S^n]} \arrow[r, "\pi_1^*"] \arrow[d, "p_{\iota_2 *}"] & {[M_1,S^n]} \arrow[r, "i_1^*"] \arrow[d, "p_{\iota_2 *}"] & {[S^{n+1},S^n]} \arrow[r, "q_{\iota_1}^*"] \arrow[d, "p_{\iota_2 *}"] & {[S^{n+1},S^n]} \arrow[d, "p_{\iota_2 *}"] \\
{[S^{n+2},S^n]} \arrow[r, "q_{\iota_1}^*"] \arrow[d, "i_{2*}"] & {[S^{n+2},S^n]} \arrow[r, "\pi_1^*"] \arrow[d, "i_{2*}"] & {[M_1,S^n]} \arrow[r, "i_1^*"] \arrow[d, "i_{2*}"] & {[S^{n+1},S^n]} \arrow[r, "q_{\iota_1}^*"] \arrow[d, "i_{2*}"] & {[S^{n+1},S^n]} \arrow[d, "i_{2*}"] \\
{[S^{n+2},M_2]} \arrow[r, "q_{\iota_1}^*"] \arrow[d, "\pi_{2*}"] & {[S^{n+2},M_2]} \arrow[r, "\pi_1^*"] \arrow[d, "\pi_{2*}"] & {[M_1,M_2]} \arrow[r, "i_1^*"] \arrow[d, "\pi_{2*}"] & {[S^{n+1},M_2]} \arrow[r, "q_{\iota_1}^*"] \arrow[d, "\pi_{2*}"] & {[S^{n+1},M_2]} \arrow[d, "\pi_{2*}"] \\
{[S^{n+2},S^{n+1}]} \arrow[r, "q_{\iota_1}^*"] \arrow[d, "p_{\iota_2 *}"] & {[S^{n+2},S^{n+1}]} \arrow[r, "\pi_1^*"] \arrow[d, "p_{\iota_2 *}"] & {[M_1,S^{n+1}]} \arrow[r, "i_1^*"] \arrow[d, "p_{\iota_2 *}"] & {[S^{n+1},S^{n+1}]} \arrow[r, "q_{\iota_1}^*"] \arrow[d, "p_{\iota_2 *}"] & {[S^{n+1},S^{n+1}]} \arrow[d, "p_{\iota_2 *}"] \\
{[S^{n+2},S^{n+1}]} \arrow[r, "q_{\iota_1}^*"] & {[S^{n+2},S^{n+1}]} \arrow[r, "\pi_1^*"] & {[M_1,S^{n+1}]} \arrow[r, "i_1^*"] & {[S^{n+1},S^{n+1}]} \arrow[r, "q_{\iota_1}^*"] & {[S^{n+1},S^{n+1}]}
\end{tikzcd}$$

Lemma 3.4. *Let* $(p,q) \neq 1$. *Then, if either* $p \equiv 0 \pmod 4$ *and* $q \equiv 2 \pmod 4$ *or* $p \equiv 2 \pmod 4$ *and* $q \equiv 0 \pmod 4$, *we have* $[M_1, M_2] \cong Z_4 \oplus Z_2$.

Proof. Suppose that $p \equiv 0 \pmod 4$ and $q \equiv 2 \pmod 4$. With the results in [Araki and Toda 1965], we obtain the following diagram from the above diagram:

$$\begin{tikzcd}
 & & 0 \arrow[d] & & \\
 & & Z_4 \arrow[d, "i_{2*}"] & & \\
0 \arrow[r] & Z_2 \oplus Z_2 \arrow[r, "\pi_1^*"] & {[M_1, M_2]} \arrow[r, "i_1^*"] \arrow[d, "\pi_{2*}"] & Z_2 \arrow[r] & 0 \\
 & & Z_2 \arrow[d] & & \\
 & & 0 & &
\end{tikzcd}$$

Thus, $[M_1, M_2]$ is isomorphic to one of three groups: Z_8, $Z_4 \oplus Z_2$ or $Z_2 \oplus Z_2 \oplus Z_2$. Since i_{2*} is injective, $[M_1, M_2]$ has an element of order 4. However, $Z_2 \oplus Z_2 \oplus Z_2$ does not have an element of order 4. Since π_1^* is injective, $[M_1, M_2]$ has a subgroup which is not cyclic. It follows that $[M_1, M_2] \neq Z_8$. Therefore, $[M_1, M_2] \cong Z_4 \oplus Z_2$.

Now, let $p \equiv 2 \pmod 4$ and $q \equiv 0 \pmod 4$. With the results in [Araki and Toda 1965], we obtain the following diagram from the above commutative diagram:

$$\begin{tikzcd}
 & & 0 \arrow[d] & & \\
 & & Z_2 \oplus Z_2 \arrow[d, "i_{2*}"] & & \\
0 \arrow[r] & Z_4 \arrow[r, "\pi_1^*"] & {[M_1, M_2]} \arrow[r, "i_1^*"] \arrow[d, "\pi_{2*}"] & Z_2 \arrow[r] & 0 \\
 & & Z_2 \arrow[d] & & \\
 & & 0 & &
\end{tikzcd}$$

Thus, $[M_1, M_2]$ is isomorphic to one of the three groups: Z_8, $Z_4 \oplus Z_2$ or $Z_2 \oplus Z_2 \oplus Z_2$. Since π_1^* is injective, $[M_1, M_2]$ has an element of order 4. However, $Z_2 \oplus Z_2 \oplus Z_2$ does not have an element of order 4. Since i_{2*} is injective, $[M_1, M_2]$ has a subgroup which is not cyclic. It follows that $[M_1, M_2] \neq Z_8$. Thus, $[M_1, M_2] \cong Z_4 \oplus Z_2$. □

By Lemma 3.4, $[M_1, M_2] \cong Z_4 \oplus Z_2$. However, $[M_1, M_2]$ has different generators under different conditions. Here we determine the generators.

If $p \equiv 0 \pmod 4$ and $q \equiv 2 \pmod 4$, then $[M_1, M_2] \cong Z_4 \oplus Z_2\{\alpha, \pi_1^*(\eta_2)\}$, where $\pi_1^*(\eta_1) = 2\alpha$ and $i_1^*(\alpha) = i_{2\sharp}(\eta)$.

If $p \equiv 2 \pmod 4$ and $q \equiv 0 \pmod 4$, then $[M_1, M_2] \cong Z_4 \oplus Z_2\{\pi_1^*(\bar{\eta}), \beta\}$, where $i_1^*(\beta) = i_{2\sharp}(\eta)$.

For a given homomorphism $h : G_1 \to G_2$, we have from Proposition 2.6 the commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & \mathrm{Ext}(G_2, \pi_{n+1}(X)) \arrow[r] \arrow[d, "\bar{h}^\sharp"] & \pi_n(G_2; X) \arrow[r] \arrow[d, "h^*"] & \mathrm{Hom}(G_2, \pi_n(X)) \arrow[r] \arrow[d, "h^\sharp"] & 0 \\
0 \arrow[r] & \mathrm{Ext}(G_1, \pi_{n+1}(X)) \arrow[r] & \pi_n(G_1; X) \arrow[r] & \mathrm{Hom}(G_1, \pi_n(X)) \arrow[r] & 0
\end{tikzcd}$$

where $\bar{h}^\sharp$ and $h^\sharp$ are induced by h and h^* is associated with h. This shows that the nonuniqueness of h^* is substantially limited. The measure of choice is bounded by the group

$$\mathrm{Hom}\big(\mathrm{Hom}(G_2, \pi_n(X)), \mathrm{Ext}(G_1, \pi_{n+1}(X))\big).$$

Lemma 3.5. *If $(p, q) = d \neq 1$, we have*

$$[M_1, M_2] \cong \begin{cases} Z_2 \oplus Z_2 & \text{if } p \equiv 2 \text{ and } q \equiv 2 \pmod 4, \\ Z_2 \oplus Z_2 \oplus Z_2 & \text{if } p \equiv 0 \text{ and } q \equiv 0 \pmod 4. \end{cases}$$

Proof. Suppose that $p \equiv 2 \pmod 4$ and $q \equiv 2 \pmod 4$. By the universal coefficient theorem for homotopy groups with coefficients, we have the short exact sequence

$$0 \to \mathrm{Ext}(Z_q, Z_4) \to [M_1, M_2] \to \mathrm{Hom}(Z_q, Z_2) \to 0.$$

Since $\mathrm{Ext}(Z_q, Z_4) \cong Z_{(q,4)} \cong Z_2$ and $\mathrm{Hom}(Z_q, Z_2) = Z_{(q,2)} = Z_2$, this sequence becomes

$$0 \to Z_2 \to [M_1, M_2] \to Z_2 \to 0.$$

Let $M_3 = M(Z_p, n+1)$. By the universal coefficient theorem for homotopy with coefficients, we have the sequence

$$0 \to \mathrm{Ext}(Z_p, Z_4) \to [M_3, M_2] \to \mathrm{Hom}(Z_p, Z_2) \to 0.$$

Similarly, this sequence becomes

$$0 \to Z_2 \to [M_3, M_2] \to Z_2 \to 0.$$

We may assume that $q \geq p$. Let $q = kd$ and $p = ld$, where $(k, l) = 1$. Then both k and l are odd. We define $h : Z_q \to Z_p$ by $h(\bar{1}) = \bar{l}$ with $\bar{s} = s + rZ \in Z_r$. Then, $\mathrm{im}(h)$ is congruent to Z_d in Z_p and h is a nontrivial homomorphism since $(q, p) = d \neq 1$. Thus, we have the commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & Z_2 \arrow[r] \arrow[d, "\bar{h}^\sharp"] & {[M_3, M_2]} \arrow[r] \arrow[d, "h^*"] & Z_2 \arrow[r] \arrow[d, "h^\sharp"] & 0 \\
0 \arrow[r] & Z_2 \arrow[r] & {[M_1, M_2]} \arrow[r] & Z_2 \arrow[r] & 0
\end{tikzcd}$$

where $\bar{h}^\sharp : \mathrm{Ext}(Z_p, Z_4) \to \mathrm{Ext}(Z_q, Z_4)$ and $h^\sharp : \mathrm{Hom}(Z_p, Z_2) \to \mathrm{Hom}(Z_q, Z_2)$ are induced by h.

To show that $h^\sharp : \mathrm{Hom}(Z_p, Z_2) \to \mathrm{Hom}(Z_q, Z_2)$ is an isomorphism, it is sufficient to show that $h^\sharp$ is nontrivial. Let α be an nonzero element in $\mathrm{Hom}(Z_p, Z_2)$ such that $\alpha(\bar{1}) = \bar{1}$. Since $h^\sharp(\alpha) = \alpha \circ h \in \mathrm{Hom}(Z_q, Z_2)$ and $\alpha \circ h(\bar{1}) = \alpha(\bar{l}) = \bar{l} = \bar{1}$, where l is odd, it follows that $h^\sharp(\alpha)$ is a nontrivial homomorphism.

Next, we show that $\bar{h}^\sharp : \mathrm{Ext}(Z_p, Z_4) \to \mathrm{Ext}(Z_q, Z_4)$ is an isomorphism. Consider the resolutions of Z_q and Z_p. Then we have following commutative diagram:

$$\begin{tikzcd}
0 \arrow[r] & Z \arrow[r, "q"] \arrow[d, "h_1"] & Z \arrow[r, "\beta"] \arrow[d, "h_2"] & Z_q \arrow[r] \arrow[d, "h"] & 0 \\
0 \arrow[r] & Z \arrow[r, "p"] & Z \arrow[r, "\beta'"] & Z_p \arrow[r] & 0
\end{tikzcd}$$

See [Gray 1975, Lemma 25.3]. Now, we give precise definitions of the maps h_1, h_2 and $h^\sharp$. Since $\bar{l} = h(\bar{1}) = h \circ \beta(1) = \beta'(h_2(1))$, we have h_2 given by $h_2(1) = l$. Moreover, we can obtain h_1 using h_2. Since $p \circ h_1 = h_2 \circ q$, we have

$ph_1(1) = h_2(q) = qh_2(1) = dkl = pk$. Thus, h_1 is given by $h_1(1) = k$. If we consider the three homomorphisms $h^\sharp$, $h_1^\sharp$ and $h_2^\sharp$ induced by h, h_1 and h_2 respectively, we have the following commutative diagram:

$$\begin{tikzcd}
0 \arrow[r] & \mathrm{Hom}(Z_q, Z_4) \arrow[r, "\beta^*"] & \mathrm{Hom}(Z, Z_4) \arrow[r, "q^*"] & \mathrm{Hom}(Z, Z_4) \\
0 \arrow[r] & \mathrm{Hom}(Z_p, Z_4) \arrow[r, "\beta'^*"] \arrow[u, "h^\sharp"] & \mathrm{Hom}(Z, Z_4) \arrow[r, "p^*"] \arrow[u, "h_2^\sharp"] & \mathrm{Hom}(Z, Z_4) \arrow[u, "h_1^\sharp", "\cong"']
\end{tikzcd}$$

Next, we show that $h_1^\sharp$ is an isomorphism. We choose a generator α of $\mathrm{Hom}(Z, Z_4)$ such that $\alpha(1) = \bar{1}$. Then $h_1^\sharp(\alpha)(1) = (\alpha \circ h_1)(1) = \alpha(k) \neq 0 \pmod 2$ since k is odd. Therefore, $h_1^\sharp(\alpha)$ is a generator of $\mathrm{Hom}(Z, Z_4)$. Thus, $h_1^\sharp$ is an isomorphism.

By using $h_1^\sharp$, we determine the homomorphism $\bar{h}^\sharp : \mathrm{Ext}(Z_p, Z_4) \to \mathrm{Ext}(Z_q, Z_4)$. Since $q \equiv p \equiv 2 \pmod 4$ and

$$\mathrm{Ext}(Z_p, Z_4) = \mathrm{Hom}(Z, Z_4)/\operatorname{im}(p^*) \quad \text{and} \quad \mathrm{Ext}(Z_q, Z_4) = \mathrm{Hom}(Z, Z_4)/\operatorname{im}(q^*),$$

we have

$$\mathrm{Ext}(Z_p, Z_4) = \langle \alpha + \{2\alpha\} \rangle \quad \text{and} \quad \mathrm{Ext}(Z_q, Z_4) = \langle \alpha + \{2\alpha\} \rangle.$$

By well-known facts of homological algebra, $\bar{h}^\sharp : \mathrm{Ext}(Z_p, Z_4) \to \mathrm{Ext}(Z_q, Z_4)$ is given by $\bar{h}^\sharp(\alpha + \{2\alpha\}) = \alpha \circ h_1 + \{2\alpha\} \neq 0$. Therefore, $\bar{h}^\sharp$ is nontrivial. Thus, $\bar{h}^\sharp$ is an isomorphism.

By the five lemma, $h^* : [M_1, M_2] \to [M_3, M_2]$ is an isomorphism. From [Araki and Toda 1965], we have $[M_3, M_2] \cong Z_2 \oplus Z_2$. Therefore, $[M_1, M_2] \cong Z_2 \oplus Z_2$.

Next, we suppose that $q \equiv 0$ and $p \equiv 0 \pmod 4$.

From [Araki and Toda 1965] and the commutative diagram above Lemma 3.4, we obtain the following commutative diagram:

$$\begin{tikzcd}
 & & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & Z_2 \arrow[r, "\pi_1^*"] \arrow[d, "i_{2*}"] & Z_2 \oplus Z_2 \arrow[r, shift left, "i_1^*"] \arrow[d, "i_{2*}"] & Z_2 \arrow[r] \arrow[l, shift left, "r"] \arrow[d, shift left, "i_{2*}"] & 0 \\
0 \arrow[r] & Z_2 \oplus Z_2 \arrow[r, "\pi_1^*"] & {[M_1, M_2]} \arrow[r, "i_1^*"] \arrow[d, "\pi_{2*}"] & Z_2 \arrow[r] \arrow[u, shift left, "\theta"] & 0 \\
 & & Z_2 \arrow[d] & & \\
 & & 0 & &
\end{tikzcd}$$

Since the second row is a split exact sequence, there exists $r : [S^{n+1}, S^n] \to [M_1, S^n]$ such that $i_1^* \circ r = \mathrm{id}_{[S^{n+1}, S^n]}$. Moreover, since the third i_{2*} is an isomorphism, there exists $\theta : [S^{n+1}, M_2] \to [S^{n+1}, S^n]$ such that $\theta \circ i_{2*} = \mathrm{id}_{[S^{n+1}, S^n]}$ and $i_{2*} \circ \theta = \mathrm{id}_{[S^{n+1}, M_2]}$.

We define the map $k : [S^{n+1}, M_2] \to [M_1, M_2]$ by $k = i_{2*} \circ r \circ \theta$. Then, we have

$$\begin{aligned} i_1^* \circ k &= i_1^* \circ i_{2*} \circ r \circ \theta \\ &= i_{2*} \circ i_1^* \circ r \circ \theta \\ &= i_{2*} \circ \mathrm{id}_{[S^{n+1}, S^n]} \circ \theta \\ &= i_{2*} \circ \theta = \mathrm{id}_{[S^{n+1}, M_2]}. \end{aligned}$$

Therefore, the third row is a split exact sequence. Hence,

$$[M_1, M_2] \cong Z_2 \oplus Z_2 \oplus Z_2. \qquad \square$$

Now, we determine the generators of $[M_1, M_2]$ when either $p \equiv 2 \pmod 4$ and $q \equiv 2 \pmod 4$ or $p \equiv 0 \pmod 4$ and $q \equiv 0 \pmod 4$.

Let $p \equiv 2 \pmod 4$ and $q \equiv 2 \pmod 4$. By using the Puppe exact sequence, we have the following exact sequence:

$$\pi_{n+2}(M_2) \xrightarrow{q_{\iota_1}^*} \pi_{n+2}(M_2) \xrightarrow{\pi_1^*} [M_1, M_2] \xrightarrow{i_1^*} \pi_{n+1}(M_2) \xrightarrow{p_{\iota_1}^*} \pi_{n+1}(M_2).$$

By exactness, we obtain the exact sequence

$$0 \longrightarrow Z_2 \xrightarrow{\pi_1^*} [M_1, M_2] \xrightarrow{i_1^*} Z_2 \longrightarrow 0.$$

Thus, $[M_1, M_2] \cong Z_2 \oplus Z_2\{\pi_1^*(\bar{\eta}), \beta\}$, where $i_1^*(\beta) = i_{2\sharp}(\eta)$.

Next, we let $p \equiv 0 \pmod 4$ and $q \equiv 0 \pmod 4$. By a similar method we obtain $[M_1, M_2] \cong Z_2 \oplus Z_2 \oplus Z_2 \{\pi_1^*(\eta_1), \pi_1^*(\eta_2), \alpha\}$, where $i_1^*(\alpha) = i_{2\sharp}(\eta)$.

Remark 3.6. Here we determine the generators of $\pi_{n+3}(M(Z_q, n))$. By using the mapping cone sequence of the Moore space

$$S^n \xrightarrow{q_\iota} S^n \xrightarrow{i} M(Z_q, n) \xrightarrow{\pi} S^{n+1} \xrightarrow{q_\iota} S^{n+1},$$

we obtain a long exact sequence

$$\pi_{n+3}(S^n) \xrightarrow{q_{\iota\sharp}} \pi_{n+3}(S^n) \xrightarrow{i_\sharp} \pi_{n+3}(M(Z_q, n)) \xrightarrow{\pi_\sharp} \pi_{n+3}(S^{n+1}) \xrightarrow{q_{\iota\sharp}} \pi_{n+3}(S^{n+1}).$$

From the work by Toda [1962], we have

$$Z_{24}\{\nu\} \xrightarrow{q_{\iota\sharp}} Z_{24}\{\nu\} \xrightarrow{i_\sharp} \pi_{n+3}(M(Z_q, n)) \xrightarrow{\pi_\sharp} Z_2\{\eta^2\} \xrightarrow{q_{\iota\sharp}} Z_2\{\eta^2\}.$$

Thus, if q is odd, then $\pi_{n+3}(M(Z_q, n)) \cong Z_{(q,24)}\{i_\sharp(\nu)\}$, and if q is even, then $\pi_{n+3}(M(Z_q, n)) \cong Z_{(q,24)} \oplus Z_2\{i_\sharp(\nu), \overline{\eta^2}\}$ where $\pi_\sharp(\overline{\eta^2}) = \eta^2$.

Based on Remarks 3.1 and 3.6, we obtain for M_1 the table

	q odd	$q \equiv 2 \pmod 4$	$q \equiv 0 \pmod 4$
$\pi_{n+3}(M_1)$	0	Z_4	$Z_2 \oplus Z_2$
Generator		$\hat{\eta}$	η_3, η_4
Relation		$i_{1\sharp}(\eta^2) = 2\hat{\eta}, \pi_{1\sharp}(\hat{\eta}) = \eta$	$i_{1\sharp}(\eta^2) = \eta_3, \pi_{1\sharp}(\eta_4) = \eta$

while for M_2 we obtain

	p odd	$p \equiv 2 \pmod 4$	$p \equiv 0 \pmod 4$
$\pi_{n+3}(M_2)$	$Z_{(p,24)}$	$Z_{(p,24)} \oplus Z_2$	$Z_{(p,24)} \oplus Z_2$
Generator	$i_{2\sharp}(\nu)$	$i_{2\sharp}(\nu), \overline{\eta^2}$	$i_{2\sharp}(\nu), \overline{\eta^2}$
Relation		$\pi_{2\sharp}(\overline{\eta^2}) = \eta^2$	$\pi_{2\sharp}(\overline{\eta^2}) = \eta^2$

By Lemmas 3.4 and 3.5, we have the following table, where $\pi_1^*(\eta_1) = 2\alpha$, $i_1^*(\alpha) = i_{2\sharp}(\eta)$ and $i_1^*(\beta) = i_{2\sharp}(\eta)$:

	$[M_1, M_2]$	Generator
either q odd or p odd	0	
$q \equiv 2, p \equiv 0 \pmod 4$	$Z_4 \oplus Z_2$	$\alpha, \pi_1^*(\eta_2)$
$q \equiv 0, p \equiv 2 \pmod 4$	$Z_4 \oplus Z_2$	$\pi_1^*(\bar{\eta}), \beta$
$q \equiv p \equiv 2 \pmod 4$	$Z_2 \oplus Z_2$	$\pi_1^*(\bar{\eta}), \beta$
$q \equiv p \equiv 0 \pmod 4$	$Z_2 \oplus Z_2 \oplus Z_2$	$\pi_1^*(\eta_1), \pi_1^*(\eta_2), \alpha$

4. Computation of $\mathcal{E}_\sharp^{\dim+r}(M(Z_q, n+1) \vee M(Z_p, n))$ for $r = 0, 1$

In this section, we compute $\mathcal{E}_\sharp^{\dim+r}(M_1 \vee M_2)$, where $M_1 = M(Z_q, n+1) = S^{n+1} \cup_q e^{n+2}$ and $M_2 = M(Z_p, n) = S^n \cup_p e^{n+1}$ with $p, q \geq 1$. In [Jeong 2010], these groups were computed in the case of $p = q$. However, we compute those groups in the general case, that is, $p \neq q$ and $r = 0, 1$. Throughout this section we assume that $X = M_1 \vee M_2$. Note that $\pi_{n+k}(M_1 \vee M_2) \cong \pi_{n+k}(M_1) \oplus \pi_{n+k}(M_2)$ for $k \leq n$ by Proposition 2.1. Moreover, from Proposition 2.2, we can identify $f \in [X, X]$ with the 2×2 matrix

$$\theta(f) = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix},$$

where $f_{11} \in [M_1, M_1]$, $f_{12} \in [M_2, M_1]$, $f_{21} \in [M_1, M_2]$, and $f_{22} \in [M_1, M_1]$.

Lemma 4.1. *Let $f \in [X, X]$ be given by*

$$f = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}.$$

Then $f \in \mathcal{E}(X)$ *if and only if* $f_{11} \in \mathcal{E}(M_1)$ *and* $f_{22} \in \mathcal{E}(M_2)$. *Additionally, if* $f \in \mathcal{E}_\sharp^{\dim}(X)$, *then* $f_{22} = 1$.

Proof. Let us denote by $h_{*n} : H_n(U) \to H_n(V)$ the induced homomorphism on the homology group from $h : U \to V$. Then, $f \in \mathcal{E}(X)$ if and only if f_* is an isomorphism if and only if f_{11*n+1} and f_{22*n} are isomorphisms if and only if $f_{11} \in \mathcal{E}(M_1)$ and $f_{22} \in \mathcal{E}(M_2)$. For the proof of the second statement, see [Jeong 2010, Lemma 3.3]. □

Let us denote by $g_{\sharp s} : \pi_s(U) \to \pi_s(V)$ the homomorphism induced by $g : U \to V$. It is clear from Lemma 4.1 that if $f \in \mathcal{E}(X)$, then $f_{\sharp(n+k)} : \pi_{n+k}(X) \to \pi_{n+k}(X)$ is given by

$$f_{\sharp(n+k)} = \begin{pmatrix} f_{11\sharp(n+k)} & f_{12\sharp(n+k)} \\ f_{21\sharp(n+k)} & f_{22\sharp(n+k)} \end{pmatrix},$$

where $f_{11\sharp(n+k)}$ and $f_{22\sharp(n+k)}$ are isomorphisms and $k \leq n$.

Lemma 4.2. *If* $f \in \mathcal{E}(X)$ *and either* q *is odd or* p *is odd, then* $f_{12\sharp k} = 0$ *for* $k = 1, 2, \dots, n+2$.

Proof. Since M_1 is n-connected, we have $\pi_k(M_1) = 0$ for $k = 1, 2, \dots, n$. Thus it is sufficient to show that $f_{12\sharp k} = 0$ for $k = n+1, n+2$.

If p is odd, then $\pi_{n+1}(M_2)$ and $\pi_{n+2}(M_2)$ are trivial groups. Thus, $f_{12\sharp(n+1)} = f_{12\sharp(n+2)} = 0$.

Suppose that q is odd, p is even and $(p, q) = d \neq 1$. Then, $\pi_{n+1}(M_2) \cong Z_2\{i_{2\sharp}(\eta)\}$. Since $[M_2, M_1] \cong Z_d\{\pi_2^*(i_1)\}$, we have $f_{12\sharp n+1} = t\pi_2^*(i_1)_\sharp$ for some integer t such that $1 \leq t \leq d$. Thus, we have

$$f_{12\sharp(n+1)}(i_{2\sharp}(\eta)) = t\pi_2^*(i_1)(i_{2\sharp}(\eta)) = t(i_1 \circ \pi_2 \circ i_2 \circ \eta) = 0$$

because $\pi_2 \circ i_2$ is homotopic to a constant map. Hence, $f_{12\sharp(n+1)} = 0$. If $d = 1$, $[M_2, M_1] = 0$ and it is trivial.

For $k = n+2$, we are done since $\pi_{n+2}(M_1) = 0$. □

Here we introduce certain generators and elements of $[M_1, M_1]$ and $\mathcal{E}_\sharp^{\dim+r}(M_1)$ for $r = -1, 0, 1$ as described in [Jeong 2010].

Remark 4.3. Let $M_1 = M(Z_q, n+1)$ be a Moore space with q is even. By Proposition 2.5, $\mathcal{E}_\sharp^{\dim}(M_1) \cong Z_2$ and $\mathcal{E}_\sharp^{\dim+1}(M_1) = 1$. In this remark, we describe the generator of $\mathcal{E}_\sharp^{\dim}(M_1)$ explicitly.

Consider the mapping cone sequence

$$S^{n+1} \xrightarrow{q_{\iota_1}} S^{n+1} \xrightarrow{i_1} S^{n+1} \cup_q e^{n+2} \xrightarrow{\pi_1} S^{n+2} \xrightarrow{q_{\iota_1}} S^{n+2}.$$

Then, we have the following exact sequence:

$$\pi_{n+2}(M_1) \xrightarrow{q_{\iota_1}^*} \pi_{n+2}(M_1) \xrightarrow{\pi_1^*} [M_1, M_1] \xrightarrow{i_1^*} \pi_{n+1}(M_1) \xrightarrow{q_{\iota_1}^*} \pi_{n+1}(M_1).$$

Since $\pi_{n+2}(M_1) \cong Z_2\{i_1\eta\}$ and $\pi_{n+1}(M_1) \cong Z_q\{1\}$, we have the short exact sequence

$$0 \longrightarrow Z_2\{i_1\eta\} \xrightarrow{\pi_1^*} [M_1, M_1] \xrightarrow{i_1^*} Z_q\{1\} \longrightarrow 0\ .$$

By [Araki and Toda 1965, Theorem 4.1],

$$[M_1, M_1] \cong \begin{cases} Z_{2q}\{1\} & \text{if } q \equiv 2 \pmod 4, \\ Z_q \oplus Z_2\{1, i_1 \circ \eta \circ \pi_1\} & \text{if } q \equiv 0 \pmod 4, \end{cases}$$

and

$$\pi_1^*(i_1 \circ \eta) = i_1 \circ \eta \circ \pi_1 \in [M_1, M_1].$$

Let $i_1 \circ \eta \circ \pi_1 = \epsilon$. Then, ϵ has order 2 and $1 + \epsilon \in [M_1, M_1]$. Since $n \geq 5$, we have that $1 + \epsilon$ is a suspension map. Thus,

$$(1+\epsilon) \circ (1+\epsilon) \simeq 1 \circ (1+\epsilon) + \epsilon \circ (1+\epsilon) = 1 + \epsilon + \epsilon + \epsilon \circ \epsilon = 1 + 2\epsilon + \epsilon^2.$$

If $q \equiv 2 \pmod 4$, then $i_1 \circ \eta \circ \pi_1 = q1$ and $\epsilon^2 = i_1 \circ \eta \circ \pi_1 \circ i_1 \circ \eta \circ \pi_1$. Since $\pi_1 \circ i_1 = 0$ and ϵ has order 2, we have $2\epsilon = 0$ and $\epsilon^2 = 0$. Thus, $(1+\epsilon) \circ (1+\epsilon) \simeq 1$ and $1 + \epsilon \in \mathcal{E}(M_1)$.

Since each $\alpha \in \pi_{n+r}(M_1)$ is a suspension map, for $r = 1, 2, 3$, we have

$$(1+\epsilon)_\sharp(\alpha) = \alpha + \epsilon \circ \alpha.$$

Since $\pi_{n+1}(M_1) \cong Z_q\{i_1\}$ and $\epsilon_\sharp(i_1) = i_1 \circ \eta \circ \pi_1 \circ i_1 = 0$, we have $1 + \epsilon \in \mathcal{E}_\sharp^{\dim -1}(M_1)$.

Since $\pi_{n+2}(M_1) \cong Z_2\{i_{1\sharp}(\eta)\}$ and $\epsilon_\sharp(i_{1\sharp}(\eta)) = i_1 \circ \eta \circ \pi_1 \circ i_1 \circ \eta = 0$, we have $1 + \epsilon \in \mathcal{E}_\sharp^{\dim}(M_1)$.

Since $\pi_{n+3}(M_1) \cong Z_4\{\hat{\eta}\}$ and

$$\epsilon_\sharp(\hat{\eta}) = i_1 \circ \eta \circ \pi_1 \circ \hat{\eta} = i_1 \circ \eta \circ \eta = i_1 \circ \eta^2 = 2\hat{\eta} \neq 0,$$

we have $1 + \epsilon \notin \mathcal{E}_\sharp^{\dim+1}(M_1)$.

We obtain similar results in the case of $q \equiv 0 \pmod 4$.

Theorem 4.4. *If $X = M_1 \vee M_2$ and $(p, q) = 1$, then*

$$\mathcal{E}_\sharp^{\dim}(X) \cong \begin{cases} 1 & \textit{if } q \textit{ is odd}, \\ Z_2 & \textit{if } q \textit{ is even and } p \textit{ is odd}. \end{cases}$$

Proof. Let $(q, p) = 1$. Then, either q or p is odd. By Lemmas 3.2 and 3.3, we have $[M_2, M_1] = 0$ and $[M_1, M_2] = 0$.

If q is odd, then $\mathcal{E}_\sharp^{\dim}(M_1) = 1$ and $\mathcal{E}_\sharp^{\dim}(M_2) = 1$ by Proposition 2.5 and Lemma 4.1. Therefore $\mathcal{E}_\sharp^{\dim}(X) = 1$.

If p is odd and q is even, then $\mathcal{E}_\sharp^{\dim}(M_1) \cong Z_2\{1+\epsilon\}$ and $\mathcal{E}_\sharp^{\dim}(M_2) = 1$ by Proposition 2.5, Lemma 4.1, and Remark 4.3. Thus, we have

$$\mathcal{E}_\sharp^{\dim}(X) \cong \left\{ \begin{pmatrix} 1+\epsilon & 0 \\ 0 & 1 \end{pmatrix} \middle| \ \epsilon \in Z_2 \ \{i_1\eta\pi_1\} \right\},$$

where η is the generator of $\pi_{n+2}(S^{n+1})$. □

Theorem 4.5. *If $X = M_1 \vee M_2$ and $(p, q) = d \neq 1$, then*

$$\mathcal{E}_\sharp^{\dim}(X) \cong \begin{cases} Z_d & \text{if } q \text{ is odd,} \\ Z_2 \oplus Z_d & \text{if } q \text{ is even and } p \text{ is odd.} \end{cases}$$

Proof. By Lemmas 3.2 and 3.3, we have $[M_2, M_1] \cong Z_d\{\pi_2^*(i_1)\}$ and $[M_1, M_2] = 0$. Moreover, $f_{12\sharp k} = 0$ for $k = 1, 2, \dots, n+2$ by Lemma 4.2.

Thus, if q is odd, then we have

$$\mathcal{E}_\sharp^{\dim}(X) \cong \left\{ \begin{pmatrix} 1 & f_{12} \\ 0 & 1 \end{pmatrix} \middle| \ f_{12} \in Z_d \ \{\pi_2^*(i_1)\} \right\},$$

but if q is even and p is odd, then we have

$$\mathcal{E}_\sharp^{\dim}(X) \cong \left\{ \begin{pmatrix} 1+\epsilon & f_{12} \\ 0 & 1 \end{pmatrix} \middle| \ f_{12} \in Z_d \ \{\pi_2^*(i_1)\}, \epsilon \in Z_2 \ \{i_1\eta\pi_1\} \right\}. \qquad \square$$

Let f_{12} be an element of $[M_2, M_1] \cong Z_d\{\pi_2^*(i_1)\}$, Then $f_{12} = s\pi_2^*(i_1)$ for $1 \le s \le d$.

Lemma 4.6. *For $f = \left(\begin{smallmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{smallmatrix}\right) \in \mathcal{E}(X)$, let p and q be even. Then, $f_{12\sharp k} = 0$ for $k = 1, 2, \dots, n+1$.*

Proof. Since M_1 is n-connected, $\pi_k(M_1) = 0$ for $k = 1, 2, \dots, n$. Thus, it is sufficient to show that $f_{12\sharp(n+1)} = 0$. Since $[M_2, M_1] \cong Z_d\{\pi_2^*(i_1)\}$ by Lemma 3.2 and f_{12} belongs to $[M_2, M_1]$, we have $f_{12} = s\pi_2^*(i_1)$ for some $1 \le s \le d$. Moreover, $\pi_{n+1}(M_2) \cong Z_2\{i_{2\sharp}(\eta)\}$ by Remark 3.1. Thus, we have

$$f_{12\sharp(n+1)}(i_{2\sharp}(\eta)) = s\pi_2^*(i_1)(i_{2\sharp}(\eta)) = s(i_1 \circ \pi_2 \circ i_2 \circ \eta) = 0$$

since $\pi_2 \circ i_2$ is homotopic to the constant map. □

Lemma 4.7. *Let p and q be even and $f_{12} = s\pi_2^*(i_1)$ be an element of $[M_2, M_1] \cong Z_d\{\pi_2^*(i_1)\}$ for $1 \le s \le d$. Then, $f_{12\sharp(n+2)} \neq 0$ if s is odd, and $f_{12\sharp(n+2)} = 0$ if s is even.*

Proof. First, we note that $\pi_{n+2}(M_1) \cong Z_2\{i_{\sharp 1}(\eta)\}$.

Suppose that $p \equiv 0 \pmod 4$. Since $\pi_{n+2}(M_2) \cong Z_2 \oplus Z_2\{\eta_1, \eta_2\}$, we have

$$\pi_2^*(i_1) = \pi_2^*(i_1)(\eta_1) = \pi_2^*(i_1)(i_{2\sharp}(\eta^2)) = i_1 \circ \pi_2 \circ i_2 \circ \eta^2 = 0$$

and

$$\pi_2^*(i_1)(\eta_2) = i_1 \circ \pi_2 \circ \eta_2 = i_1 \circ \eta \neq 0.$$

Thus, $f_{12\sharp(n+2)}(\eta_1)=0$ for all f_{12}. Moreover, if $s=2l$ for some $1\le l\le d/2$, then

$$s\pi_2^*(i_1)(\eta_2)=si_1\circ\pi_2\circ\eta_2=2li_1\circ\eta=0.$$

Therefore, each element in $\langle 2\pi_2^*(i_1)\rangle\cong Z_{d/2}$ induces the trivial homomorphism on $\pi_{n+2}(M_2)$. However, if $s=2l+1$ for some $0\le l<d/2-1$, then

$$s\pi_2^*(i_1)(\eta_2)=si_1\circ\pi_2\circ\eta_2=(2l+1)i_1\circ\eta=i_1\circ\eta\ne 0.$$

Thus, if f_{12} does not belong to $\langle 2\pi_2^*(i_1)\rangle\cong Z_{d/2}$, then $f_{12\sharp(n+2)}\ne 0$.

Suppose that $p\equiv 2 \pmod 4$. Since $\pi_{n+2}(M_2)\cong Z_4\{\bar\eta\}$, we have

$$\pi_2^*(i_1)_\sharp(\bar\eta)=i_1\circ\pi_2\circ\bar\eta=i_1\circ\eta=i_{1\sharp}(\eta)\ne 0.$$

If $s=2k$ for some $1\le l\le d/2$, then

$$s\pi_2^*(i_1)_\sharp(\bar\eta)=si_1\circ\pi_2\circ\bar\eta=si_1\circ\eta=2li_{1\sharp}(\eta)=0.$$

Thus, each element in $\langle 2\pi_2^*(i_1)\rangle\cong Z_{d/2}$ induces the trivial homomorphism on $n+2$.

However, if $s=2l+1$ for some $0\le l\le d/2-1$, then

$$s\pi_2^*(i_1)_\sharp(\bar\eta)=si_1\circ\pi_2\circ\bar\eta=si_1\circ\eta=(2l+1)i_{1\sharp}(\eta)=i_{1\sharp}(\eta)\ne 0.$$

Thus if f_{12} does not belong to $\langle 2\pi_2^*(i_1)\rangle\cong Z_{d/2}$, then $f_{12\sharp(n+2)}\ne 0$. □

Theorem 4.8. *Let p and q be even and let $X=M_1\vee M_2$. Then if $(p,q)=d\ne 1$, we have*

$$\mathcal{E}_\sharp^{\dim}(X)\cong\begin{cases} Z_2\oplus Z_{d/2}\oplus Z_2\oplus Z_2 & \text{if } q\equiv 2,\ p\equiv 0 \pmod 4,\\ Z_2\oplus Z_{d/2}\oplus Z_4 & \text{if } q\equiv 0,\ p\equiv 2 \pmod 4,\\ Z_2\oplus Z_{d/2}\oplus Z_2 & \text{if } q\equiv 2,\ p\equiv 2 \pmod 4,\\ Z_2\oplus Z_{d/2}\oplus Z_2\oplus Z_2 & \text{if } q\equiv 0,\ p\equiv 0 \pmod 4.\end{cases}$$

Proof. By Proposition 2.5, $\mathcal{E}_\sharp^{\dim}(M_1)\cong Z_2$ and $\mathcal{E}_\sharp^{\dim}(M_2)=1$. By Lemma 4.6, for each $f=\left(\begin{smallmatrix} f_{11} & f_{12}\\ f_{21} & f_{22}\end{smallmatrix}\right)\in\mathcal{E}(X)$, we have $f_{12\sharp k}=0$ for $k=1,2,\dots,n+1$. By Lemma 4.7, each element in $\langle 2\pi_2^*(i_1)\rangle\cong Z_{d/2}$ induces the trivial homomorphism on $\pi_{n+2}(M_2)$. Furthermore, if f_{12} does not belong to $\langle 2\pi_2^*(i_1)\rangle\cong Z_{d/2}$, then $f_{12\sharp(n+2)}\ne 0$. Thus, it is sufficient to investigate $f_{21\sharp n}$, $f_{21\sharp(n+1)}$ and $f_{21\sharp(n+2)}$.

Case 1. Let $q\equiv 2 \pmod 4$ and $p\equiv 0 \pmod 4$. From Lemma 3.4, we obtain $[M_1,M_2]\cong Z_4\oplus Z_2\{\alpha,\pi_1^*(\eta_2)\}$, where $\pi_1^*(\eta_1)=2\alpha$ and $i_1^*(\alpha)=i_{2\sharp}(\eta)$.

Since M_1 is n-connected, $\pi_n(M_1)=0$. Thus, $f_{21\sharp n}=0$.

Since $\pi_{n+1}(M_1)\cong Z_q\{i_1\}$, we have $\pi_1^*(\eta_2)_\sharp(i_1)=\eta_2\circ\pi_1\circ i_1=0$.

Conversely, since $\pi_{n+1}(M_2)\cong Z_2\{i_{2\sharp}(\eta)\}$ and $\alpha_\sharp(i_1)=\alpha\circ i_1=i_{2\sharp}(\eta)\ne 0$, we have $(2\alpha)_\sharp=0$ and $(3\alpha)_\sharp\ne 0$. Moreover, since $\pi_{n+2}(M_1)\cong Z_2\{i_{1\sharp}(\eta)\}$, we have $\pi_1^*(\eta_2)_\sharp(i_{1\sharp}(\eta))=\eta_2\circ\pi_1\circ i_1\circ\eta=0$. Hence, $\left(\begin{smallmatrix} 1 & 0\\ f_{21} & 1\end{smallmatrix}\right)$ belongs to $\mathcal{E}_\sharp^{\dim}(X)$ if $f_{21}\in Z_2\oplus Z_2\{2\alpha,\pi_1^*(\eta_2)\}$.

Therefore,

$$\mathcal{E}_{\sharp}^{\dim}(X) \cong \left\{ \begin{pmatrix} 1+\epsilon & f_{12} \\ f_{21} & 1 \end{pmatrix} \;\middle|\; f_{12} \in \langle 2\pi_2^*(i_1)\rangle,\ f_{21} \in \langle 2\alpha\rangle \oplus \langle \pi_1^*(\eta_2)\rangle \right\},$$

where $\epsilon \in \langle i_1\eta\pi_1\rangle$.

Case 2. Let $q \equiv 0 \pmod 4$ and $p \equiv 2 \pmod 4$. From Lemma 3.4, we obtain $[M_1, M_2] \cong Z_4 \oplus Z_2\{\pi_1^*(\overline{\eta}), \beta\}$, where $i_1^*(\beta) = i_{2\sharp}(\eta)$.

Since $\pi_n(M_1) = 0$, we have $f_{21\sharp n} = 0$. However, since $\pi_{n+1}(M_1) \cong Z_q\{i_1\}$ and $\pi_{n+1}(M_2) \cong Z_2\{i_{2\sharp}(\eta)\}$, we have $\pi_1^*(\overline{\eta})_\sharp(i_1) = \overline{\eta}\circ\pi_1\circ i_1 = 0$, but $\beta_\sharp(i_1) = \beta\circ i_1 = i_{2\sharp}(\eta) \neq 0$.

For the generator $\pi_1^*(\overline{\eta})$ of $[M_1, M_2] \cong Z_4 \oplus Z_2\{\pi_1^*(\overline{\eta}), \beta\}$ and the generator $i_{1\sharp}(\eta)$ of $\pi_{n+2}(M_1) \cong Z_2\{i_{1\sharp}(\eta)\}$, we have $\pi_1^*(\overline{\eta})_\sharp(i_{1\sharp}(\eta)) = \overline{\eta}\circ\pi_1\circ i_{1\sharp}(\eta) = 0$.

Hence, $\left(\begin{smallmatrix} 1 & 0 \\ f_{21} & 1 \end{smallmatrix}\right)$ belongs to $\mathcal{E}_{\sharp}^{\dim}(X)$ if $f_{21} \in \langle \pi_1^*(\overline{\eta})\rangle$.

Therefore,

$$\mathcal{E}_{\sharp}^{\dim}(X) \cong \left\{ \begin{pmatrix} 1+\epsilon & f_{12} \\ f_{21} & 1 \end{pmatrix} \;\middle|\; f_{12} \in \langle 2\pi_2^*(i_1)\rangle,\ f_{21} \in \langle \pi_1^*(\overline{\eta})\rangle \right\},$$

where $\epsilon \in \langle i_1\eta\pi_1\rangle$.

Case 3. Let $q \equiv 2 \pmod 4$ and $p \equiv 2 \pmod 4$. From Lemma 3.5, we obtain $[M_1, M_2] \cong Z_2 \oplus Z_2\{\pi_1^*(\overline{\eta}), \beta\}$, where $i_1^*(\beta) = i_{2\sharp}(\eta)$.

First, we recall that $f_{21\sharp n} = 0$ since $\pi_n(M_1) = 0$.

Since $\pi_{n+1}(M_1) \cong Z_q\{i_1\}$ and $\pi_{n+1}(M_2) \cong Z_2\{i_{2\sharp}(\eta)\}$, we have $\pi_1^*(\overline{\eta})_\sharp(i_1) = \overline{\eta}\circ\pi_1\circ i_1 = 0$, but $\beta_\sharp(i_1) = \beta\circ i_1 = i_{2\sharp}(\eta) \neq 0$. Moreover, since $\pi_{n+2}(M_1) \cong Z_2\{i_{1\sharp}(\eta)\}$, we have $\pi_1^*(\overline{\eta})_\sharp(i_{1\sharp}(\eta)) = \overline{\eta}\circ\pi_1\circ i_1\circ\eta = 0$.

Hence, if $f_{21} \in \langle \pi_1^*(\overline{\eta})\rangle$, then $\left(\begin{smallmatrix} 1 & 0 \\ f_{21} & 1 \end{smallmatrix}\right)$ belongs to $\mathcal{E}_{\sharp}^{\dim}(X)$. However, if $f_{21} \in \langle\beta\rangle$, this cannot be the case. Therefore,

$$\mathcal{E}_{\sharp}^{\dim}(X) \cong \left\{ \begin{pmatrix} 1+\epsilon & f_{12} \\ f_{21} & 1 \end{pmatrix} \;\middle|\; f_{12} \in \langle 2\pi_2^*(i_1)\rangle,\ f_{21} \in \langle \pi_1^*(\overline{\eta})\rangle \right\},$$

where $\epsilon \in \langle i_1\eta\pi_1\rangle$.

Case 4. Let $q \equiv 0 \pmod 4$ and $p \equiv 0 \pmod 4$. From Lemma 3.5, we obtain $[M_1, M_2] \cong Z_2 \oplus Z_2 \oplus Z_2\{\pi_1^*(\eta_1), \pi_1^*(\eta_2), \alpha\}$, where $i_1^*(\alpha) = i_{2\sharp}(\eta)$. First, we note that $f_{21\sharp n} = 0$ since $\pi_n(M_1) = 0$.

Since $\pi_{n+1}(M_1) \cong Z_q\{i_1\}$ and $\pi_{n+1}(M_2) \cong Z_2\{i_{2\sharp}(\eta)\}$, we have $\pi_1^*(\eta_1)_\sharp(i_1) = \eta_1\circ\pi_1\circ i_1 = 0$ and $\pi_1^*(\eta_2)_\sharp(i_1) = \eta_2\circ\pi_1\circ i_1 = 0$, but $\alpha_\sharp(i_1) = \alpha\circ i_1 = i_{2\sharp}(\eta) \neq 0$. Also, since $\pi_{n+2}(M_1) \cong Z_2\{i_{1\sharp}(\eta)\}$, we have $\pi_1^*(\eta_1)_\sharp(i_{1\sharp}(\eta)) = \eta_1\circ\pi_1\circ i_{1\sharp}(\eta) = 0$ and $\pi_1^*(\eta_2)_\sharp(i_{1\sharp}(\eta)) = \eta_2\circ\pi_1\circ i_{1\sharp}(\eta) = 0$.

Hence, if $f_{21} \in \langle \pi_1^*(\eta_1)\rangle \oplus \langle \pi_1^*(\eta_2)\rangle$, then $\left(\begin{smallmatrix} 1 & 0 \\ f_{21} & 1 \end{smallmatrix}\right)$ belongs to $\mathcal{E}_{\sharp}^{\dim}(X)$. However, if $f_{21} \in \langle\alpha\rangle$, this cannot be the case. Therefore,

$$\mathcal{E}_\sharp^{\dim}(X) \cong \left\{ \begin{pmatrix} 1+\epsilon & f_{12} \\ f_{21} & 1 \end{pmatrix} \,\middle|\, f_{12} \in \langle 2\pi_2^*(i_1)\rangle,\ f_{21} \in \pi_1^*(\langle \eta_1\rangle \oplus \langle \eta_2\rangle) \right\},$$

where $\epsilon \in \langle i_1 \eta \pi_1\rangle$. □

From Theorems 4.4–4.8, we obtain Table 1 (see page 37).

Theorem 4.9. *Let $X = M_1 \vee M_2$, $n \geq 5$ and $(q, p) = d$. Then we have*

$$\mathcal{E}_\sharp^{\dim+1}(X) \cong \begin{cases} 1 & \textit{if } q \textit{ is odd or } p \textit{ is odd } (d=1), \\ Z_d & \textit{if } q \textit{ is odd or } p \textit{ is odd } (d \neq 1), \\ Z_{d/2} \oplus Z_2 & \textit{if } p \equiv 0 \pmod 4 \textit{ and } (p, 24) = 4 \textit{ or } 12 (d \neq 1), \\ Z_{d/2} & \textit{if } p \equiv 0 \pmod 4 \textit{ and } (p, 24) = 8 \textit{ or } 24 (d \neq 1), \\ Z_{d/2} & \textit{if } q \equiv 2,\ p \equiv 2 \pmod 4, \\ Z_{d/2} \oplus Z_2 & \textit{if } q \equiv 0,\ p \equiv 2 \pmod 4. \end{cases}$$

Proof. By virtue of Remark 4.3, Theorem 4.4 and the fact that $\mathcal{E}_\sharp^{\dim+1}(X) \subseteq \mathcal{E}_\sharp^{\dim}(X)$, we have $\mathcal{E}_\sharp^{\dim+1}(X) = 1$ if $(p, q) = 1$.

By Proposition 2.5, we have $\mathcal{E}_\sharp^{\dim+1}(M_1) = 1$. Thus, it is sufficient to identify $f_{12\sharp(n+3)}$ and $f_{21\sharp(n+3)}$. First, we note that $[M_2, M_1] \cong Z_d\{\pi_2^*(i_1)\}$ by Lemma 3.2.

Case 1. Suppose that q is odd or p is odd and $(p, q) = d \neq 1$. Since $[M_1, M_2] = 0$ by Lemma 3.3, we only investigate $f_{12\sharp(n+3)}$.

If q is odd, $f_{12\sharp(n+3)} = 0$ since $\pi_{n+3}(M_1) = 0$. If q is even and p is odd, $\pi_{n+3}(M_2) \cong Z_{(p,24)}\{i_{2\sharp}(\nu)\}$. Since

$$\pi_2^*(i_1)_\sharp(i_{2\sharp}(\nu)) = i_1 \circ \pi_2 \circ i_2 \circ \nu = 0,$$

we have $f_{12\sharp(n+3)} = 0$ for each $f_{12} \in [M_2, M_1]$. Therefore,

$$\mathcal{E}_\sharp^{\dim+1}(X) \cong \left\{ \begin{pmatrix} 1 & f_{12} \\ 0 & 1 \end{pmatrix} \,\middle|\, f_{12} \in \langle \pi_2^*(i_1)\rangle \right\}.$$

Case 2. Suppose that $q \equiv 2 \pmod 4$ and $p \equiv 0 \pmod 4$. First, we note that

$$\pi_{n+3}(M_2) \cong Z_{(p,24)} \oplus Z_2\ \{i_{2\sharp}(\nu), \overline{\eta^2}\}$$

and that $\pi_{n+3}(M_1) \cong Z_4\{\hat{\eta}\}$ by Proposition 2.3. Let $f_{12} = s\pi_2^*(i_1)$. If $s = 2l$ for some $1 \leq l \leq d/2$, then

$$s\pi_2^*(i_1)_\sharp(\overline{\eta^2}) = 2l\pi_2^*(i_1)_\sharp(\overline{\eta^2}) = 4l\hat{\eta} = 0$$

since

$$\pi_2^*(i_1)_\sharp(i_{2\sharp}(\nu)) = i_1 \circ \pi_2 \circ i_2 \circ \nu = 0$$

and

$$\pi_2^*(i_1)_\sharp(\overline{\eta^2}) = i_1 \circ \pi_2 \circ \overline{\eta^2} = i_{1\sharp}(\eta^2) = 2\hat{\eta} \neq 0 \in \pi_{n+3}(M_1) \cong Z_4.$$

Further, if $s = 2l + 1$ for some $0 \le l \le d/2 - 1$, then

$$s\pi_2^*(i_1)_\sharp(\overline{\eta^2}) = (2l+1)\pi_2^*(i_1)_\sharp(\overline{\eta^2}) = 4l\hat{\eta} + 2\hat{\eta} = 2\hat{\eta} \neq 0.$$

Thus, each $f_{12} \in \langle 2\pi_2^*(i_1)\rangle \cong Z_{d/2}$ induces the trivial homomorphism on $\pi_{n+3}(M_2)$. However, if f_{12} does not belong to $\langle 2\pi_2^*(i_1)\rangle$, then $f_{12\sharp(n+3)} \neq 0$.

Let us investigate $f_{21\sharp(n+3)}$. Note that $[M_1, M_2] \cong Z_4 \oplus Z_2\{\alpha, \pi_1^*(\eta_2)\}$ and $\pi_{n+3}(M_1) \cong Z_4\{\hat{\eta}\}$ with $\pi_1^*(\eta_1) = 2\alpha$, $i_1^*(\alpha) = i_{2\sharp}(\eta)$, $i_{1\sharp}(\eta^2) = 2\hat{\eta}$ and $\pi_{1\sharp}(\hat{\eta}) = \eta$. Since $\pi_{2\sharp}(\eta_2 \circ \eta) = \eta^2$, we have

$$\pi_1^*(\eta_2)_\sharp(\hat{\eta}) = \eta_2 \circ \pi_1 \circ \hat{\eta} = \eta_2 \circ \eta \neq 0.$$

Moreover, since $\eta^3 = 4\nu$ [Toda 1962, (5.5)], we have

$$2\alpha_\sharp(\hat{\eta}) = 2\alpha \circ \hat{\eta} = \eta_1 \circ \pi_1 \circ \hat{\eta} = \eta_1 \circ \eta = i_{1\sharp}(\eta^2) \circ \eta = i_2 \circ \eta^3 = 4i_{2\sharp}(\nu).$$

Therefore, $\alpha_\sharp(\hat{\eta}) = 2i_{2\sharp}(\nu)$. Since $(p, 24)$ is a multiple of 4, we have $\alpha_\sharp(\hat{\eta}) = 2i_{2\sharp}(\nu) \neq 0$ and $3\alpha_\sharp(\hat{\eta}) = 6i_{2\sharp}(\nu) \neq 0$.

Since ν is 2-primary, if $(p, 24) = 4$ or $(p, 24) = 12$, then $2\alpha_\sharp(\hat{\eta}) = 0$, and if $(p, 24) = 8$ or $(p, 24) = 24$, then $2\alpha_\sharp(\hat{\eta}) \neq 0$. Thus, each $f_{21} \in \langle 2\alpha\rangle$ induces the trivial homomorphism on $\pi_{n+3}(M_1)$ provided that $(p, 24) = 4$ or $(p, 24) = 12$.

Therefore, if $(p, 24) = 4$ or $(p, 24) = 12$, we have

$$\mathcal{E}_\sharp^{\dim+1}(X) \cong \left\{ \begin{pmatrix} 1 & f_{12} \\ f_{21} & 1 \end{pmatrix} \,\middle|\, f_{12} \in \langle 2\pi_2^*(i_1)\rangle,\ f_{21} \in \langle 2\alpha\rangle \right\},$$

and if $(p, 24) = 8$ or 24, we have

$$\mathcal{E}_\sharp^{\dim+1}(X) \cong \left\{ \begin{pmatrix} 1 & f_{12} \\ 1 & 1 \end{pmatrix} \,\middle|\, f_{12} \in \langle 2\pi_2^*(i_1)\rangle \right\}.$$

Case 3. Suppose that $q \equiv 0 \pmod 4$ and $p \equiv 2 \pmod 4$. We note that

$$\pi_{n+3}(M_2) \cong Z_{(p,24)} \oplus Z_2\{i_{2\sharp}(\nu), \overline{\eta^2}\},$$
$$\pi_{n+3}(M_1) \cong Z_2 \oplus Z_2\{\eta_3, \eta_4\}$$

and $[M_1, M_2] \cong Z_4 \oplus Z_2\{\pi_1^*(\bar{\eta}), \beta\}$. First, we investigate $f_{12\sharp(n+3)}$. Let $f_{12} = s\pi_2^*(i_1) \in [M_2, M_1] \cong Z_d\{\pi_2^*(i_1)\}$. Then, we have

$$\pi_2^*(i_1)_\sharp(i_{2\sharp}(\nu)) = i_1 \circ \pi_2 \circ i_2 \circ \nu = 0$$

and

$$\pi_2^*(i_1)_\sharp(\overline{\eta^2}) = i_1 \circ \pi_2 \circ \overline{\eta^2} = i_1 \circ \eta^2 \neq 0.$$

If $s = 2l$ for some $1 \le l \le d/2$, then $2l\pi_2^*(i_1)_\sharp(\overline{\eta^2}) = 2li_1 \circ \eta^2 = 0$, because $i_1 \circ \eta^2 = \eta_3 \in \pi_{n+3}(M_1)$. However, if $s = 2l + 1$ for some $0 \le l \le d/2 - 1$, then $(2l+1)\pi_2^*(i_1)_\sharp(\overline{\eta^2}) = (2k+1)i_1 \circ \eta^2 = i_1 \circ \eta^2 \neq 0$.

Thus, any $f_{12} \in \langle 2\pi_2^*(i_1)\rangle \cong Z_{d/2}$ induces the trivial homomorphism on $\pi_{n+3}(M_2)$. However, for $f_{12} \notin \langle 2\pi_2^*(i_1)\rangle$, we have $f_{12\sharp n+3} \neq 0$.

Next, we investigate $f_{21\sharp(n+3)}$. Because $[M_1, M_2] \cong Z_4 \oplus Z_2\{\pi_1^*(\overline{\eta}), \beta\}$ and $\beta_{\sharp(n+2)} \neq 0$, we check only the generators $\pi_1^*(\overline{\eta})$. For η_3, we have

$$\pi_1^*(\overline{\eta})_\sharp(\eta_3) = \overline{\eta} \circ \pi_1 \circ \eta_3 = \overline{\eta} \circ \pi_1 \circ i_{1\sharp}(\eta^2) = 0.$$

For η_4, we have

$$\pi_1^*(\overline{\eta})_\sharp(\eta_4) = \overline{\eta} \circ \pi_1 \circ \eta_4 = \overline{\eta} \circ \eta \neq 0$$

since $\pi_{2\sharp}(\overline{\eta} \circ \eta) = \eta^2 \neq 0$.

However, $2\pi_1^*(\overline{\eta})_\sharp(\eta_4) = \overline{\eta} \circ \pi_1 \circ 2\eta_4 = 0$.

Thus, every $f_{21} \in \langle 2\pi_1^*(\overline{\eta})\rangle$ induces the trivial homomorphism on $n+3$.

Therefore, we have

$$\mathcal{E}_\sharp^{\dim+1}(X) \cong \left\{ \begin{pmatrix} 1 & f_{12} \\ f_{21} & 1 \end{pmatrix} \;\middle|\; f_{12} \in \langle 2\pi_2^*(i_1)\rangle,\ f_{21} \in \langle 2\pi_1^*(\overline{\eta})\rangle \right\}.$$

Case 4. Suppose that $q \equiv 2 \pmod 4$ and $p \equiv 2 \pmod 4$. Note that $\pi_{n+3}(M_2) \cong Z_{(p,24)} \oplus Z_2\{i_{2\sharp}(\nu), \overline{\eta^2}\}$ and $\pi_{n+3}(M_1) \cong Z_4\{\hat{\eta}\}$. First, we investigate $f_{12\sharp(n+3)}$. For the generator $\pi_2^*(i_1)$ of $[M_2, M_1]$, we have

$$\pi_2^*(i_1)_\sharp(i_{2\sharp}(\nu)) = i_1 \circ \pi_2 \circ i_2 \circ \nu = 0$$

and

$$\pi_2^*(i_1)_\sharp(\overline{\eta^2}) = i_1 \circ \pi_2 \circ \overline{\eta^2} = i_1 \circ \eta^2 = 2\hat{\eta} \neq 0.$$

Let $f_{12} = s\pi_2^*(i_1)$. If $s = 2l$ for $1 \leq l \leq d/2$, then $s\pi_2^*(i_1)_\sharp(\overline{\eta^2}) = 4l\hat{\eta} = 0$, and if $s = 2l+1$ for $0 \leq l \leq d/2 - 1$, then $s\pi_2^*(i_1)_\sharp(\overline{\eta^2}) = (4l+2)\hat{\eta} = 2\hat{\eta} \neq 0$.

Thus, each $f_{12} \in \langle 2\pi_2^*(i_1)\rangle \cong Z_{d/2}$ induces the trivial homomorphism on $n+3$. However, for $f_{12} \notin \langle 2\pi_2^*(i_1)\rangle$, we have $f_{12\sharp(n+3)} \neq 0$.

Next, we investigate $f_{21\sharp(n+3)}$. Note that $[M_1, M_2] \cong Z_2 \oplus Z_2\{\pi_1^*(\overline{\eta}), \beta\}$. Since $\beta_{\sharp n+2} \neq 0$, we consider only the generator $\pi_1^*(\overline{\eta})$.

Since $\pi_{2\sharp}(\overline{\eta} \circ \eta) = \pi_2 \circ \overline{\eta} \circ \eta = \eta^2 \neq 0$, we have $\pi_1^*(\overline{\eta})_\sharp(\hat{\eta}) = \overline{\eta} \circ \pi_1 \circ \hat{\eta} = \overline{\eta} \circ \eta \neq 0$. Therefore, no f_{21} induces a trivial homomorphism.

Thus, we have

$$\mathcal{E}_\sharp^{\dim+1}(X) \cong \left\{ \begin{pmatrix} 1 & f_{12} \\ 1 & 1 \end{pmatrix} \;\middle|\; f_{12} \in \langle 2\pi_2^*(i_1)\rangle \right\}.$$

Case 5. Suppose that $q \equiv 0 \pmod 4$ and $p \equiv 0 \pmod 4$. Note that $\pi_{n+3}(M_2) \cong Z_{(p,24)} \oplus Z_2\{i_{2\sharp}(\nu), \overline{\eta^2}\}$ and $\pi_{n+3}(M_1) \cong Z_2 \oplus Z_2\{\eta_3, \eta_4\}$.

First, we investigate $f_{12\sharp(n+3)}$. For the generator $\pi_2^*(i_1)$ of $[M_2, M_1]$, we have

$$\pi_2^*(i_1)_\sharp(i_{2\sharp}(\nu)) = i_1 \circ \pi_2 \circ i_2 \circ \nu = 0$$

and

$$\pi_2^*(i_1)_\sharp(\overline{\eta^2}) = i_1 \circ \pi_2 \circ \overline{\eta^2} = i_1 \circ \eta^2 \neq 0.$$

Let $f_{12}=s\pi_2^*(i_1)$. If $s=2l$ for $1\le l\le d/2$, then

$$s\pi_2^*(i_1)_\sharp(\overline{\eta^2})=2li_1\circ\eta^2=l2\eta_3=0.$$

However, if $s=2l+1$ for $0\le l\le d/2-1$, then

$$s\pi_2^*(i_1)_\sharp(\overline{\eta^2})=(2l+1)i_1\circ\eta^2=\eta_3\neq 0.$$

Thus, each $f_{12}\in\langle 2\pi_2^*(i_1)\rangle\cong Z_{d/2}$ induces the trivial homomorphism on $n+3$. However, for $f_{12}\notin\langle 2\pi_2^*(i_1)\rangle$, we have $f_{12\sharp(n+3)}\neq 0$.

Next, we consider $f_{21\sharp(n+3)}$. Note that

$$[M_1,M_2]\cong Z_2\oplus Z_2\oplus Z_2\ \{\ \pi_1^*(\eta_1),\pi_1^*(\eta_2),\alpha\}.$$

Since $\alpha_{\sharp(n+2)}=0$, we consider only the generators $\pi_1^*(\eta_1)$ and $\pi_1^*(\eta_2)$. For $\pi_1^*(\eta_1)$, we have

$$\pi_1^*(\eta_1)_\sharp(\eta_3)=\eta_1\circ\pi_1\circ\eta_3=\eta_1\circ\pi_1\circ i_1\eta^2=0$$

and

$$\pi_1^*(\eta_1)_\sharp(\eta_4)=\eta_1\circ\pi_1\circ\eta_4=\eta_1\circ\eta=i_{2\sharp}(\eta^2)\circ\eta=4i_{2\sharp}(\nu).$$

Thus, if $(p,24)=4$ or $(p,24)=12$, then $\pi_1^*(\eta_1)_\sharp(\eta_4)=4i_{1\sharp}(\nu)=0$, and if $(p,24)=8$ or $(p,24)=24$, then $\pi_1^*(\eta_1)_\sharp(\eta_4)=4i_{1\sharp}(\nu)\neq 0$.

Since $\pi_{2\sharp}(\eta_2\circ\eta)=\eta^2$, we have $\pi_1^*(\eta_2)_\sharp(\eta_4)=\eta_2\circ\pi_1\circ\eta_4=\eta_2\circ\eta\neq 0$.

Therefore, if $(p,24)=4$ or $(p,24)=12$, we have

$$\mathcal{E}_\sharp^{\dim+1}(X)\cong\left\{\begin{pmatrix}1 & f_{12}\\ f_{21} & 1\end{pmatrix}\,\middle|\, f_{12}\in\langle 2\pi_2^*(i_1)\rangle,\ f_{21}\in\langle\pi_2^*(\eta_1)\rangle\right\},$$

and if $(p,24)=8$ or $(p,24)=24$, we have

$$\mathcal{E}_\sharp^{\dim+1}(X)\cong\left\{\begin{pmatrix}1 & f_{12}\\ 1 & 1\end{pmatrix}\,\middle|\, f_{12}\in\langle 2\pi_2^*(i_1)\rangle\right\}.\qquad\square$$

References

[Araki and Toda 1965] S. Araki and H. Toda, “Multiplicative structures in mod q cohomology theories, I”, *Osaka J. Math.* **2** (1965), 71–115. MR 32 #449 Zbl 0129.15201

[Arkowitz 1990] M. Arkowitz, “The group of self-homotopy equivalences—a survey”, pp. 170–203 in *Groups of self-equivalences and related topics* (Montreal, 1988), edited by R. A. Piccinini, Lecture Notes in Math. **1425**, Springer, Berlin, 1990. MR 91i:55001 Zbl 0713.55004

[Arkowitz and Maruyama 1998] M. Arkowitz and K.-i. Maruyama, “Self-homotopy equivalences which induce the identity on homology, cohomology or homotopy groups”, *Topology Appl.* **87**:2 (1998), 133–154. MR 99i:55009 Zbl 0929.55008

[Blakers and Massey 1952] A. L. Blakers and W. S. Massey, “The homotopy groups of a triad, II”, *Ann. of Math.* (2) **55** (1952), 192–201. MR 13,485f Zbl 0046.40604

[Gray 1975] B. Gray, *Homotopy theory: An introduction to algebraic topology*, Pure and Applied Mathematics **64**, Academic Press, New York, 1975. MR 53 #6528 Zbl 0322.55001

[Hilton 1965] P. Hilton, *Homotopy theory and duality*, Gordon and Breach, New York, 1965. MR 33 #6624 Zbl 0155.50801

[Jeong 2010] M. H. Jeong, "Certain subgroups of self-homotopy equivalences of the wedge of two Moore spaces", *Commun. Korean Math. Soc.* **25**:1 (2010), 111–117. MR 2011a:55008 Zbl 1213.55005

[Puppe 1958] D. Puppe, "Homotopiemengen und ihre induzierten Abbildungen, I", *Math. Z.* **69** (1958), 299–344. MR 20 #6698 Zbl 0092.39803

[Spanier 1966] E. H. Spanier, *Algebraic topology*, McGraw-Hill, New York, 1966. MR 35 #1007 Zbl 0145.43303

[Toda 1962] H. Toda, *Composition methods in homotopy groups of spheres*, Annals of Mathematics Studies **49**, Princeton University Press, 1962. MR 26 #777 Zbl 0101.40703

Received August 2, 2013. Revised October 17, 2013.

HO WON CHOI
DEPARTMENT OF MATHEMATICS
KOREA UNIVERSITY
SEOUL 702-701
SOUTH KOREA
howon@korea.ac.kr

KEE YOUNG LEE
DEPARTMENT OF MATHEMATICS
KOREA UNIVERSITY
SEJONG 425-791
SOUTH KOREA
keyolee@korea.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.272.59

MODULAR TRANSFORMATIONS INVOLVING THE MORDELL INTEGRAL IN RAMANUJAN'S LOST NOTEBOOK

YOUN-SEO CHOI

For my teacher Bruce C. Berndt on his 75th birthday.

In his "lost notebook" (p. 202 of the 1988 edition), S. Ramanujan recorded modular transformations involving the Mordell integral, q-hypergeometric series, and generalized Lambert series. He gave no proofs; here we prove these formulas and use them to derive modular transformations of third-order mock theta functions. Mordell's formula, the properties of q-hypergeometric series and Appell–Lerch sums play central roles in the proofs.

1. Introduction

For a complex number q with $|q| < 1$, we define the notation

$$(a;q)_\infty := \prod_{m=0}^{\infty}(1-aq^m) \quad \text{and} \quad (a;q)_n := \frac{(a;q)_\infty}{(aq^n;q)_\infty} \quad \text{for any integer } n.$$

L. J. Mordell [1920; 1933] studied the integral

$$\int_{-\infty}^{\infty} \frac{e^{at^2+bt}}{e^{ct}+d}\,dt,$$

where $\Re(a) < 0$. This integral appeared in the work of L. Kronecker [1889a; 1889b] and B. Riemann (as described by C. L. Siegel [1932]). However, Mordell was the first to analyze its behavior relative to modular transformations, so we refer to it as the Mordell integral. In [Mordell 1920] he derived the formula

$$(1)\quad \int_{-\infty}^{\infty} \frac{e^{\pi i\tau t^2-2\pi xt}}{e^{2\pi t}-e^{2\pi i\theta}}\,dt = e^{-\pi i(\theta^2\tau+2\theta x+2\theta)}\,\frac{F[(x+\theta\tau)/\tau,-1/\tau]+i\tau F(x+\theta\tau,\tau)}{\tau\theta_{11}(x+\theta\tau,\tau)}.$$

This work was supported by NRF-2012047647 from the National Research Foundation of Korea.
MSC2010: primary 11B65; secondary 11F27, 33D99.
Keywords: Mordell integral, Appell–Lerch sum, Ramanujan.

where, for $\Im(\tau) > 0$ and setting $q = e^{\pi i \tau}$,

$$iF(x,\tau) := \sum_{m=-\infty}^{\infty} \frac{(-1)^m q^{m^2+m+1/4} e^{(2m+1)\pi i x}}{1+q^{2m+1}},$$

$$i\theta_{11}(x,\tau) := \sum_{m=-\infty}^{\infty} (-1)^m q^{m^2+m+1/4} e^{(2m+1)\pi i x}.$$

To get (1), he mainly used functional equations satisfied by the functions $F(x,\tau)$ and $\theta_{11}(x,\tau)$.

S. Ramanujan studied definite integrals and recorded modular transformations involving the Mordell integral. In his lost notebook [1988, p. 9], he stated two modular transformations involving Mordell integrals and his tenth-order mock theta functions $\phi(q) := \sum_{n=0}^{\infty} q^{n(n+1)/2}/(q;q^2)_{n+1}$ and $\psi(q) := \sum_{n=0}^{\infty} q^{(n+1)(n+2)/2}/(q;q^2)_{n+1}$:

$$\int_0^\infty \frac{e^{-\pi n x^2}}{\cosh \frac{2\pi x}{\sqrt{5}} + \frac{1+\sqrt{5}}{4}}\,dx + \frac{1}{\sqrt{n}} e^{\frac{\pi}{5n}} \psi(-e^{-\frac{\pi}{n}}) = \sqrt{\frac{5+\sqrt{5}}{2}}\, e^{-\frac{\pi n}{5}} \phi(-e^{-\pi n}) - \frac{\sqrt{5}+1}{2\sqrt{n}} e^{-\frac{\pi}{5n}} \phi(-e^{\frac{\pi}{n}}),$$

$$\int_0^\infty \frac{e^{-\pi n x^2}}{\cosh \frac{2\pi x}{\sqrt{5}} + \frac{1-\sqrt{5}}{4}}\,dx + \frac{1}{\sqrt{n}} e^{\frac{\pi}{5n}} \psi(-e^{-\frac{\pi}{n}}) = -\sqrt{\frac{5-\sqrt{5}}{2}}\, e^{\frac{\pi n}{5}} \phi(-e^{-\pi n}) + \frac{\sqrt{5}-1}{2\sqrt{n}} e^{-\frac{\pi}{5n}} \phi(-e^{\frac{\pi}{n}}).$$

In [Choi 2002], we proved these equations. In the lost notebook Ramanujan [1988, p. 202] also wrote (without proofs) two equations involving a Mordell integral, hypergeometric series and generalized Lambert series. Namely, for $q_1 = e^{-\frac{\pi}{3n}}$ and $q = e^{-3\pi n}$,

$$\begin{aligned}
\frac{2}{\sqrt{3}} \int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}} \cos \pi t x}{e^{\frac{2\pi x}{3}} + 1 + e^{-\frac{2\pi x}{3}}}\,dx &= q^{\frac{1}{18}} \sum_{m=1}^{\infty} \frac{q^{\frac{(2m-1)^2}{6}}}{(-e^{\pi t} q^{\frac{1}{3}}; q^{\frac{2}{3}})_m (-e^{-\pi t} q^{\frac{1}{3}}; q^{\frac{2}{3}})_m} \\
&\quad + \frac{e^{-\frac{3\pi t^2}{4n}} q_1^{\frac{1}{2}}}{\sqrt{n}} \sum_{m=1}^{\infty} \frac{q_1^{\frac{3}{2}(2m-1)^2}}{(-e^{\frac{\pi i t}{n}} q_1^3; q_1^6)_m (-e^{-\frac{\pi i t}{n}} q_1^3; q_1^6)_m} \\
= \frac{q^{-\frac{1}{36}}}{(q^{\frac{2}{3}}; q^{\frac{2}{3}})_\infty} \Bigg\{ \sum_{m=1}^{\infty} (-1)^{m+1} q^{\frac{(2m-1)^2}{4}} &\left(\frac{1}{1+e^{\pi t} q^{\frac{2m-1}{3}}} + \frac{1}{1+e^{-\pi t} q^{\frac{2m-1}{3}}} - 1 \right) \\
+ \frac{e^{-\frac{3\pi t^2}{4n}}}{n} \sum_{m=1}^{\infty} (-1)^{m+1} q_1^{\frac{9}{4}(2m-1)^2} &\left(\frac{1}{1+e^{\frac{\pi i t}{n}} q_1^{3(2m-1)}} + \frac{1}{1+e^{-\frac{\pi i t}{n}} q_1^{3(2m-1)}} - 1 \right) \Bigg\}.
\end{aligned}$$

We prove these equations in this paper. Proving these identities is equivalent to proving the following two theorems.

Theorem 1. *For a positive number n, set $q=e^{-3\pi n}$ and $q_1=e^{-\frac{\pi}{3n}}$. For a number t such that $\Re(\frac{t}{n})\pm\frac{2}{3}\notin\mathbb{Z}$ and $\Re(\frac{t}{n})\pm\frac{4}{3}\notin\mathbb{Z}$, we have*

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos \pi t x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{1}{18}}\sum_{m=1}^{\infty}\frac{q^{\frac{(2m-1)^2}{6}}}{(-e^{\pi t}q^{\frac{1}{3}};q^{\frac{2}{3}})_m(-e^{-\pi t}q^{\frac{1}{3}};q^{\frac{2}{3}})_m} \\ + \frac{e^{-\frac{3\pi t^2}{4n}}q_1^{\frac{1}{2}}}{\sqrt{n}}\sum_{m=1}^{\infty}\frac{q_1^{\frac{3}{2}(2m-1)^2}}{(-e^{\frac{\pi i t}{n}}q_1^3;q_1^6)_m(-e^{-\frac{\pi i t}{n}}q_1^3;q_1^6)_m}.$$

Theorem 2. *For a positive number n, set $q=e^{-3\pi n}$ and $q_1=e^{-\frac{\pi}{3n}}$. We have*

$$q^{\frac{1}{18}}\sum_{m=1}^{\infty}\frac{q^{\frac{(2m-1)^2}{6}}}{(-e^{\pi t}q^{\frac{1}{3}};q^{\frac{2}{3}})_m(-e^{-\pi t}q^{\frac{1}{3}};q^{\frac{2}{3}})_m} + \frac{e^{-\frac{3\pi t^2}{4n}}q_1^{\frac{1}{2}}}{\sqrt{n}}\sum_{m=1}^{\infty}\frac{q_1^{\frac{3}{2}(2m-1)^2}}{(-e^{\frac{\pi i t}{n}}q_1^3;q_1^6)_m(-e^{-\frac{\pi i t}{n}}q_1^3;q_1^6)_m}$$
$$=\frac{q^{-\frac{1}{36}}}{(q^{\frac{2}{3}};q^{\frac{2}{3}})_\infty}\Bigg\{\sum_{m=1}^{\infty}(-1)^{m+1}q^{\frac{(2m-1)^2}{4}}\left(\frac{1}{1+e^{\pi t}q^{\frac{2m-1}{3}}}+\frac{1}{1+e^{-\pi t}q^{\frac{2m-1}{3}}}-1\right) \\ +\frac{e^{-\frac{3\pi t^2}{4n}}}{n}\sum_{m=1}^{\infty}(-1)^{m+1}q_1^{\frac{9}{4}(2m-1)^2}\left(\frac{1}{1+e^{\frac{\pi i t}{n}}q_1^{3(2m-1)}}+\frac{1}{1+e^{-\frac{\pi i t}{n}}q_1^{3(2m-1)}}-1\right)\Bigg\}.$$

G. E. Andrews [1981] also studied modular transformations consisting of the Mordell integral and the three functions

$$M_1(q):=\sum_{n=-\infty}^{\infty}\frac{q^{2n^2+n}}{1+q^{2n}},\quad M_2(q):=\sum_{n=-\infty}^{\infty}\frac{q^{2n^2-n}}{1+q^{2n-1}},$$
$$M_3(q):=\sum_{n=-\infty}^{\infty}\frac{q^{2n^2+2n}}{1+q^{2n+1}}.$$

These functions are related to the classical theta functions $\vartheta_2(0,q)$ and $\vartheta_4(0,q)$, and the first two of them appear in Ramanujan's lost notebook.

In [Choi 2011], we made the definition

$$f(\alpha,z;q):=\sum_{m=0}^{\infty}\frac{q^{m^2-3m}\alpha^m z^{2m}}{(-z;q)_m(-\frac{\alpha z}{q};q)_m}.$$

If we let $\alpha=z=q$, we see that $f(q,q;q)$ is one of Ramanujan's famous third-order mock theta functions, $f(q)$, from his letter [Berndt and Rankin 1995]. We can rewrite the right-hand side of the equation in Theorem 1 in terms of $f(\alpha, z; q)$, namely,

$$q^{\frac{2}{9}} f(e^{-2\pi t} q^{\frac{2}{3}}, e^{\pi t} q^{\frac{1}{3}}; q^{\frac{2}{3}}) + \frac{e^{-\frac{3\pi t^2}{4n}} q_1^2}{\sqrt{n}} f(e^{-\frac{2\pi i t}{n}} q_1^6, e^{\frac{\pi i t}{n}} q_1^3; q_1^6).$$

Ramanujan's equations involve the hypergeometric series

$$\sum_{m=1}^{\infty} \frac{q^{\frac{(2m-1)^2}{6}}}{(-e^{\pi t} q^{\frac{1}{3}}; q^{\frac{2}{3}})_m (-e^{-\pi t} q^{\frac{1}{3}}; q^{\frac{2}{3}})_m},$$

$$\sum_{m=1}^{\infty} \frac{q_1^{\frac{3}{2}(2m-1)^2}}{(-e^{\frac{\pi i t}{n}} q_1^3; q_1^6)_m (-e^{-\frac{\pi i t}{n}} q_1^3; q_1^6)_m}.$$

These are special cases of the function

$$g_3(z, q) := \sum_{m=1}^{\infty} \frac{q^{m(m-1)}}{(z; q)_m (z^{-1} q; q)_m}. \tag{2}$$

Andrews and F. G. Garvan [1989] called attention to what they called the "mock theta conjectures", which roughly say that Ramanujan's fifth-order mock theta functions are not, in fact, theta functions. These were proved by D. Hickerson [1988]; though he did not use the function (2) in the proof, he remarked that he could express the conjectures in terms of it. Since then g_3 and a couple of other so-called *universal mock theta functions* have acquired a central role in the study of mock theta functions; see the survey by B. Gordon and R. McIntosh [2012] for discussion.

The function g_3 also satisfies certain modular transformations [Gordon and McIntosh 2012]. For $q = e^{-\alpha}$, $q_1 = e^{-\pi^2/\alpha}$, and

$$h_3(e^{2\pi i r}, q) := \frac{4 \sin^2 \pi r}{(q; q)_\infty} \sum_{m=-\infty}^{\infty} \frac{(-1)^m q^{\frac{m(3m+1)}{2}}}{(1 - e^{2\pi i r} q^m)(1 - e^{-2\pi i r} q^m)},$$

one of the modular transformations satisfied by g_3 is

$$\begin{aligned} q^{\frac{3}{2} r(1-r) - \frac{1}{24}} g_3(q^r, q) = {} & \sqrt{\frac{\pi}{2\alpha}} \csc \pi r \; q_1^{-\frac{1}{6}} h_3(e^{2\pi i r}, q_1^4), \\ & - \sqrt{\frac{3\alpha}{2\pi}} \int_0^\infty e^{-\frac{3}{2}\alpha x^2} \frac{\cosh(3r-1)\alpha x + \cosh(3r-2)\alpha x}{\cosh \frac{3}{2}\alpha x} \, dx. \end{aligned}$$

With the function $g_3(z, q)$, we can rewrite the right-hand side of Ramanujan's first equation (page 60) as

$$q^{\frac{2}{9}} g_3(-e^{\pi t} q^{\frac{1}{3}}, q^{\frac{2}{3}}) + \frac{e^{-\frac{3\pi t^2}{4n}} q_1^2}{\sqrt{n}} g_3(-e^{\frac{\pi i t}{n}} q_1^3; q_1^6).$$

Ramanujan listed four third-order mock theta functions $f(q)$, $\phi(q)$, $\psi(q)$, and $\chi(q)$ in his last letter to G. H. Hardy [Berndt and Rankin 1995]. G. N. Watson [1936] later added three further third-order mock theta functions $\omega(q)$, $\upsilon(q)$ and $\rho(q)$, and derived modular transformations for the seven third-order mock theta functions using Cauchy's theorem. One of the modular transformations is

$$q^{-\frac{1}{24}} f(q) = 2\sqrt{\frac{2\pi}{\alpha}} q_1^{\frac{4}{3}} \omega(q_1^2) + 4\sqrt{\frac{3\alpha}{2\pi}} \int_0^\infty e^{-\frac{3}{2}\alpha x^2} \frac{\sinh \alpha x}{\sinh \frac{3}{2}\alpha x} \, dx$$

where $q = e^{-\alpha}$ and $q_1 = -\pi^2/\alpha$. Gordon and McIntosh [2003; 2012] introduced two more third-order mock theta functions $\xi(q)$ and $\rho(q)$ and their modular transformations.

In his thesis, S. Zwegers [2002] studied the normalized Appell–Lerch sum which is defined by

$$\mu(u, v; \tau) = \frac{1}{f(-e^{2\pi i v}, -e^{2\pi i \tau - 2\pi i v})} \sum_{m=-\infty}^{\infty} \frac{(-1)^m e^{\pi i m(m+1)\tau + 2\pi i m v}}{1 - e^{2\pi i m \tau + 2\pi i u}}$$

where $u,\ v \notin \mathbb{Z}\tau + \mathbb{Z}$ and $\tau \in \mathcal{H}$. He showed the symmetry property, the elliptic transformation properties, and the modular transformation properties satisfied by the normalized Appell–Lerch sum. One of the modular transformation properties contains the Mordell integral, namely,

$$\left(\frac{\tau}{i}\right)^{-\frac{1}{2}} e^{\frac{\pi i (u-v)^2}{\tau}} \mu\left(\frac{u}{\tau}, \frac{v}{\tau}; -\frac{1}{\tau}\right) = -\mu(u, v; \tau) + \frac{1}{2} \int_{-\infty}^{\infty} \frac{e^{\pi i x^2 \tau - 2\pi x(u-v)}}{\cosh \pi x} \, dx.$$

With these properties, Zwegers explained that $\mu(u, v; \tau)$ behaves nearly like a Jacobi form of weight $1/2$ in two variables.

Recently, B. Chern and R. C. Rhoades [2012] proved the modular transformation

$$\begin{aligned}\widetilde{R}(z;\tau) - \frac{e^{\frac{3\pi i z^2}{\tau}}}{\sqrt{i\tau}} \widetilde{R}\left(\frac{z}{\tau}; -\frac{1}{\tau}\right) &= e^{-\frac{\pi i \tau}{3}} i \sin 2\pi z \int_{-\infty}^{\infty} e^{3\pi i \tau x^2 - 6\pi z x} \frac{\cosh 2\pi\tau x}{\cosh \pi x} \, dx \\ &\quad + e^{-\frac{\pi i \tau}{3}} \cos 2\pi z \int_{-\infty}^{\infty} e^{3\pi i \tau x^2 - 6\pi z x} \frac{\sinh 2\pi \tau x}{\cosh \pi x} \, dx\end{aligned}$$

where

$$\widetilde{R}(z;\tau) := \frac{i e^{\frac{\pi i \tau}{12}}}{2 \sin \pi z} \sum_{m=0}^{\infty} \frac{e^{2\pi i \tau m^2}}{(e^{2\pi i (z+\tau)}; e^{2\pi i \tau})_m (e^{-2\pi i (z-\tau)}; e^{2\pi i \tau})_m}.$$

They employed the results in Zwegers' thesis [2002] to prove this equation. By results in [Garvan 1988], we can rewrite $\widetilde{R}$ in terms of g_3:

$$\begin{aligned}
&\widetilde{R}(z;\tau)\\
&= \frac{i e^{\frac{\pi i \tau}{12}}}{2\sin \pi z}(1-e^{2\pi i z})\left(1+e^{2\pi i z}\sum_{m=1}^{\infty}\frac{e^{2\pi i \tau m(m-1)}}{(e^{2\pi i z};e^{2\pi i \tau})_m(e^{-2\pi i(z-\tau)};e^{2\pi i \tau})_m}\right)\\
&= \frac{i e^{\frac{\pi i \tau}{12}}}{2\sin \pi z}(1-e^{2\pi i z})(1+e^{2\pi i z}g_3(e^{2\pi i z},e^{2\pi i \tau})).
\end{aligned}$$

In their paper, Chern and Rhoades [2012] also discussed and proved two more identities involving the Mordell integral and partial theta functions. In this paper, Ramanujan's theta function $f(a,b)$ is used instead of the Jacobi theta functions. The definition of Ramanujan's theta functions is, for $|ab|<1$,

$$f(a,b) := \sum_{m=-\infty}^{\infty} a^{m(m+1)/2} b^{m(m-1)/2}.$$

By the Jacobi triple product identity, this equals $(-a;ab)_\infty(-b;ab)_\infty(ab;ab)_\infty$.

In Section 2, we introduce Lemmas 1 and 2. The identities in these lemmas include generalized Lambert series which are the Appell–Lerch sums. The transformation for the Appell–Lerch sum in [Zwegers 2002] plays a central role in the proofs of Lemmas 1 and 2. In Section 3, we prove Theorem 1 twice with distinct methods. We first prove Theorem 1 by using Lemmas 1 and 2, Mordell's formula, the modular transformation for a theta function θ_{11}, and the evaluations of the contour integrals. Secondly, we prove Theorem 1 by proving

$$\frac{1}{\sqrt{3}}\int_{-\infty}^{\infty}\frac{e^{\frac{\pi i \tau x^2}{3}-2\pi z x}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = g(z;\tau)+\frac{e^{\frac{3\pi i z^2}{\tau}}}{\sqrt{-i\tau}}g\left(-\frac{z}{\tau};-\frac{1}{\tau}\right) \tag{3}$$

where

$$g(z;\tau) := \frac{e^{\frac{2\pi i \tau}{3}}}{(e^{2\pi i \tau};e^{2\pi i \tau})_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m e^{3\pi i \tau m(m+1)}}{1+e^{2\pi i z+2\pi i \tau(m+\frac{1}{2})}}.$$

To prove the equation above, we discuss the elliptic transformation properties of $g(z;\tau)$, evaluate the contour integrals, and employ Liouville's theorem. In Section 4, we prove Theorem 2 by using Equation (6) and some results in the first proof of Theorem 1. In Section 5, with Theorem 1, we derive modular transformations for third-order mock theta functions which are similar to the modular transformations for tenth-order mock theta functions in the lost notebook [1988, p. 9].

2. Lemmas

To prove Theorems 1 and 2, we require the following lemmas.

Lemma 1. *For a complex number q with $|q| < 1$, we have*

$$\begin{aligned}
&\sum_{m=1}^{\infty} \frac{q^{\frac{2m(m-1)}{3}}}{(-tq^{\frac{1}{3}};q^{\frac{2}{3}})_m(-t^{-1}q^{\frac{1}{3}};q^{\frac{2}{3}})_m} \\
&= \frac{tq^{\frac{2}{3}}}{f(-t^3q^{\frac{4}{3}},-t^{-3}q^{\frac{2}{3}})} \sum_{m=-\infty}^{\infty} \frac{(-1)^m q^{m^2+\frac{7}{3}m} t^{3m}}{1+q^{2m+1}} \\
&\quad + \frac{t^{-1}q^{\frac{2}{3}}}{f(-t^{-3}q^{\frac{4}{3}},-t^{3}q^{\frac{2}{3}})} \sum_{m=-\infty}^{\infty} \frac{(-1)^m q^{m^2+\frac{7}{3}m} t^{-3m}}{1+q^{2m+1}} + \frac{(q^2;q^2)_\infty^3}{(q^{\frac{2}{3}};q^{\frac{2}{3}})_\infty f(t^3q,t^{-3}q)} \\
&\quad + \frac{q^{\frac{1}{3}}(q^2;q^2)_\infty^3 f(q^{\frac{1}{3}},q^{\frac{5}{3}})}{(q^{\frac{2}{3}};q^{\frac{2}{3}})_\infty f(q,q) f(t^{-3}q,t^3q)} \left(\frac{t^{-2}f(-t^{-3}q^2,-t^3)}{f(-t^3q^{\frac{4}{3}},-t^{-3}q^{\frac{2}{3}})} + \frac{t^2 f(-t^3q^2,-t^{-3})}{f(-t^{-3}q^{\frac{4}{3}},-t^3q^{\frac{2}{3}})} \right).
\end{aligned}$$

Proof. Garvan [1988] showed that, for $|q| < |z| < |q|^{-1}$ and $z \neq 1$,

$$z^{-1}\left(-1+\sum_{m=0}^{\infty}\frac{q^{m^2}}{(z;q)_{m+1}(q/z;q)_m}\right) = \frac{1}{(q;q)_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{\frac{3m(m+1)}{2}}}{1-q^m z}. \tag{4}$$

Hickerson [1988, p. 649] remarked that

$$z^{-1}\left(-1+\sum_{m=0}^{\infty}\frac{q^{m^2}}{(z;q)_{m+1}(q/z;q)_m}\right) = \sum_{m=1}^{\infty}\frac{q^{m(m-1)}}{(z;q)_m(z^{-1}q;q)_m}. \tag{5}$$

Combining the two results above, we have

$$\sum_{m=1}^{\infty}\frac{q^{m(m-1)}}{(z;q)_m(z^{-1}q;q)_m} = \frac{1}{(q;q)_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{\frac{3m(m+1)}{2}}}{1-zq^m}. \tag{6}$$

We also see this equation in [Gordon and McIntosh 2012, p. 104]. Now, replacing q and z by $q^{2/3}$ and $-tq^{1/3}$, respectively, (6) becomes

$$\sum_{m=1}^{\infty}\frac{q^{\frac{2m(m-1)}{3}}}{(-tq^{\frac{1}{3}};q^{\frac{2}{3}})_m(-t^{-1}q^{\frac{1}{3}};q^{\frac{2}{3}})_m} = \frac{1}{(q^{\frac{2}{3}};q^{\frac{2}{3}})_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m(m+1)}}{1+q^{\frac{2}{3}m+\frac{1}{3}}t}. \tag{7}$$

In [Choi 2004, p. 378], the author showed that

$$\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m(m+1)}}{1-q^{2m}z} = \frac{(q^2;q^2)_\infty^2}{(z;q^2)_\infty(q^2/z;q^2)_\infty}, \tag{8}$$

which was also recorded by Ramanujan [1988, p. 59] in the lost notebook without proofs. Using (8) with z replaced by $-t^3q$, the Jacobi triple product identity and a straightforward calculation show that

$$(9)\quad \sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m(m+1)}}{1+q^{\frac{2}{3}m+\frac{1}{3}}t}=\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m(m+1)}(1-q^{\frac{2}{3}m+\frac{1}{3}}t+q^{\frac{4}{3}m+\frac{2}{3}}t^2)}{1+q^{2m+1}t^3}$$
$$=\frac{(q^2;q^2)_\infty^3}{f(t^3q,t^{-3}q)}+\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m^2+\frac{7}{3}m+\frac{2}{3}}t^{-2}}{1+q^{2m+1}t^{-3}}$$
$$+\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m^2+\frac{7}{3}m+\frac{2}{3}}t^{2}}{1+q^{2m+1}t^{3}}.$$

The two sums on the right side of the equation above are Appell–Lerch sums. In his thesis, Zwegers [2002] showed that the normalized Appell–Lerch sum satisfies

$$(10)\quad \frac{z}{f(-hz,-\frac{q}{hz})}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{\frac{m(m+1)}{2}}(hz)^m}{1-q^m tz}-\frac{1}{f(-h,-\frac{q}{h})}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{\frac{m(m+1)}{2}}}{h^m 1-q^m t}$$
$$=-\frac{(q;q)_\infty^3 f(-htz,-\frac{q}{htz})f(-z,-\frac{q}{z})}{f(-t,-\frac{q}{t})f(-h,-\frac{q}{h})f(-tz,-\frac{q}{tz})f(-hz,-\frac{q}{hz})}$$

where $q=e^{2\pi i\tau}$, $h=e^{2\pi i\nu}$, $t=e^{2\pi iu}$ and $z=e^{2\pi iz'}$, such that $\nu,\ u,\ z'\notin\mathbb{Z}$ and $u,\ \nu,\ u+z',\ \nu+z'\notin\mathbb{Z}\tau+\mathbb{Z}$. Hence, employing the Jacobi triple product identity, using (7) and (9), applying (10) with $q,\ t,\ h$, and z replaced by $q^2,\ -q,\ t^{-3}q^{4/3}$, and t^3, respectively, then again with $q,\ t,\ h$, and z replaced by $q^2,\ -q,\ t^3q^{4/3}$, and t^{-3}, respectively, and employing the fact that $f(q^{7/3},q^{-1/3})=q^{1/3}f(q^{1/3},q^{5/3})$, we obtain Lemma 1 after a slight rearrangement. □

Lemma 2. *Set $\omega=e^{\frac{2\pi i}{3}}$. For a complex number q with $|q|<1$, we have*

$$(11)\quad q^2\sum_{m=1}^{\infty}\frac{q^{6m(m-1)}}{(-tq^3;q^6)_m(-t^{-1}q^3;q^6)_m}$$
$$+\frac{i}{\sqrt{3}}\left\{\frac{1}{f(-\omega^2t^{-1}q^2,-\omega^{-2}t)}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m^2+m}\omega^{2m}t^{-m}}{1+q^{2m+1}}\right.$$
$$\left.+\frac{1}{f(-\omega^2tq^2,-\omega^{-2}t^{-1})}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m^2+m}\omega^{2m}t^{m}}{1+q^{2m+1}}-1\right\}$$
$$=-\frac{(q^2;q^2)_\infty^3}{3(q^6;q^6)_\infty f(tq,t^{-1}q)}$$
$$+\frac{i}{\sqrt{3}}\frac{(q^2;q^2)_\infty^4 f(-t,-t^{-1}q^2)}{(1-\omega^2)f(q,q)f(q,q^5)f(tq,t^{-1}q)}$$
$$\times\left(\frac{1}{f(-\omega^2t^{-1}q^2,-\omega t)}-\frac{t^{-1}}{f(-\omega^2tq^2,-\omega t^{-1})}\right).$$

Proof. We first consider the left side of (11). Employing the Jacobi triple product identity, we easily verify that

$$f(-\omega^2,-\omega q^2)=(1-\omega^2)(q^6;q^6)_\infty,\quad f(\omega^2 q,\omega q)=\frac{(q^2;q^2)(q^6;q^6)_\infty}{f(q,q^5)},$$
$$f(-\omega^2t^{-1},-\omega tq^2)=-\omega^2t^{-1}f(-\omega^2t^{-1}q^2,-\omega t).$$

Then, using (10) with $q,\ h,\ t$, and z replaced by $q^2,\ \omega^2t^{-1},\ -q$, and t, respectively, applying the equations above, and rearranging terms, we obtain

$$(12)\quad \frac{1}{f(-\omega^2t^{-1}q^2,-\omega t)}\sum_{m=-\infty}^{\infty}\frac{(-1)^mq^{m(m+1)}\omega^{2m}t^{-m}}{1+q^{2m+1}}$$
$$=\frac{1}{(1-\omega)(q^6;q^6)_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^mq^{m(m+1)}\omega^{2m}}{1+q^{2m+1}t}$$
$$+\frac{(q^2;q^2)_\infty^4f(-t,-t^{-1}q^2)}{(1-\omega^2)f(q,q)f(q,q^5)f(-\omega^2t^{-1}q^2,-\omega t)f(tq,t^{-1}q)}.$$

A straightforward calculation and the Jacobi triple product identity lead us to

$$(13)\quad \sum_{m=-\infty}^{\infty}\frac{(-1)^mq^{m(m+1)}\omega^{2m}}{1+q^{2m+1}t^{-1}}$$
$$=\sum_{m=-\infty}^{\infty}\frac{(-1)^{m+1}q^{m(m+1)}\omega^{m+1}(q^{2m+1}t+1-1)}{1+q^{2m+1}t}$$
$$=\sum_{m=-\infty}^{\infty}(-1)^{m+1}q^{m(m+1)}\omega^{m+1}+\sum_{m=-\infty}^{\infty}\frac{(-1)^mq^{m(m+1)}\omega^{m+1}}{1+q^{2m+1}t}a$$
$$=(1-\omega)(q^6;q^6)_\infty+\sum_{m=-\infty}^{\infty}\frac{(-1)^mq^{m(m+1)}\omega^{m+1}}{1+q^{2m+1}t}.$$

Using (12), again using (12) with t replaced by t^{-1}, and applying (13), we obtain

$$(14)\quad \frac{1}{f(-\omega^2t^{-1}q^2,-\omega t)}\sum_{m=-\infty}^{\infty}\frac{(-1)^mq^{m(m+1)}\omega^{2m}t^{-m}}{1+q^{2m+1}}$$
$$+\frac{1}{f(-\omega^2tq^2,-\omega t^{-1})}\sum_{m=-\infty}^{\infty}\frac{(-1)^mq^{m(m+1)}\omega^{2m}t^m}{1+q^{2m+1}}-1$$
$$=\frac{1}{(1-\omega)(q^6;q^6)_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^mq^{m(m+1)}(\omega^{2m}+\omega^{m+1})}{1+q^{2m+1}t}$$
$$+\frac{(q^2;q^2)_\infty^4f(-t,-t^{-1}q^2)}{(1-\omega^2)f(q,q)f(q,q^5)f(tq,t^{-1}q)}$$
$$\times\left(\frac{1}{f(-\omega^2t^{-1}q^2,-\omega t)}-\frac{t^{-1}}{f(-\omega^2tq^2,-\omega t^{-1})}\right).$$

We now consider

$$\sum_{m=1}^{\infty}\frac{q^{6m(m-1)+2}}{(-tq^3;q^6)_m(-t^{-1}q^3;q^6)_m} + \frac{i}{\sqrt{3}(1-\omega)(q^6;q^6)_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m(m+1)}(\omega^{2m}+\omega^{m+1})}{1+q^{2m+1}t}. \tag{15}$$

An elementary calculation shows that $(1+\omega)/(1-\omega) = i/\sqrt{3}$ and $\omega^2/(1-\omega) = -i/\sqrt{3}$. Using these, we find that

$$\begin{aligned}\frac{1}{1-\omega}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m(m+1)}(\omega^{2m}+\omega^{m+1})}{1+q^{2m+1}t} &= \frac{i}{\sqrt{3}}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{9m^2+3m}}{1+q^{6m+1}t} + \frac{2i}{\sqrt{3}}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{9m^2+9m+2}}{1+q^{6m+3}t} \\ &\quad + \frac{i}{\sqrt{3}}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{9m^2+15m+6}}{1+q^{6m+5}t}.\end{aligned} \tag{16}$$

Therefore, applying (6) with q and z replaced by q^6 and $-tq^3$, respectively, (16), and (8) with z replaced by $-tq$, we deduce that (15) equals

$$\begin{aligned}&\frac{1}{(q^6;q^6)_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{9m^2+9m+2}}{1+q^{6m+3}t} - \frac{1}{3(q^6;q^6)_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{9m^2+3m}}{1+q^{6m+1}t} \\ &\quad - \frac{2}{3(q^6;q^6)_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{9m^2+9m+2}}{1+q^{6m+3}t} \\ &\quad - \frac{1}{3(q^6;q^6)_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{9m^2+15m+6}}{1+q^{6m+5}t} \\ &= -\frac{1}{3(q^6;q^6)_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m^2+m}}{1+q^{2m+1}t} \\ &= -\frac{1}{3(q^6;q^6)_\infty}\frac{(q^2;q^2)_\infty^3}{f(tq,t^{-1}q)}.\end{aligned} \tag{17}$$

In conclusion, by combining (14) and (17) we have derived Lemma 2. □

3. The proofs of the first identity

First proof of Theorem 1. By a simple calculation and integration by substitution, we obtain

$$\begin{aligned}
(18)\quad &\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}} \cos \pi t x}{e^{\frac{2\pi x}{3}} + 1 + e^{-\frac{2\pi x}{3}}}\, dx \\
&= \frac{\sqrt{3}}{4i} e^{-\frac{3\pi t^2}{4n}} \left\{ e^{\frac{\pi i}{3}} \int_{-\infty - i\frac{t}{2n}}^{\infty - i\frac{t}{2n}} \frac{e^{-3\pi n y^2}}{e^{2\pi(y+i\frac{t}{2n})} + e^{\frac{\pi i}{3}}}\, dy \right. \\
&\quad + e^{\frac{\pi i}{3}} \int_{-\infty + i\frac{t}{2n}}^{\infty + i\frac{t}{2n}} \frac{e^{-3\pi n y^2}}{e^{2\pi(y-i\frac{t}{2n})} + e^{\frac{\pi i}{3}}}\, dy \\
&\quad - e^{-\frac{\pi i}{3}} \int_{-\infty - i\frac{t}{2n}}^{\infty - i\frac{t}{2n}} \frac{e^{-3\pi n y^2}}{e^{2\pi(y+i\frac{t}{2n})} + e^{-\frac{\pi i}{3}}}\, dy \\
&\quad \left. - e^{-\frac{\pi i}{3}} \int_{-\infty + i\frac{t}{2n}}^{\infty + i\frac{t}{2n}} \frac{e^{-3\pi n y^2}}{e^{2\pi(y-i\frac{t}{2n})} + e^{-\frac{\pi i}{3}}}\, dy \right\}.
\end{aligned}$$

For a sufficiently large positive number d, we consider the integral

$$\int_\gamma \frac{e^{-3\pi n y^2}}{e^{2\pi(y+i\frac{t}{2n})} + e^{\frac{\pi i}{3}}}\, dy$$

taken around the rectangle γ whose vertices are at the points $\pm d$ and $\pm d - i\frac{t}{2n}$. We easily verify that

$$\frac{e^{-3\pi n y^2}}{e^{2\pi(y+i\frac{t}{2n})} + e^{\frac{\pi i}{3}}}$$

has simple poles at $i(-\frac{t}{2n} + \frac{2}{3} + k)$ for integers k. Assume that $\Re(\frac{t}{n}) > 0$. Since $\Re(\frac{t}{2n}) \pm \frac{1}{3} \notin \mathbb{Z}$ and $\Re(\frac{t}{2n}) \pm \frac{2}{3} \notin \mathbb{Z}$, after some elementary manipulation and employing Cauchy's residue theorem, we find that

$$\begin{aligned}
&\int_{-d-i\frac{t}{2n}}^{d-i\frac{t}{2n}} \frac{e^{-3\pi n y^2}}{e^{2\pi(y+i\frac{t}{2n})} + e^{\frac{\pi i}{3}}}\, dy \\
&\quad = \sum_{0 \le k < \Re\frac{t}{2n} - \frac{2}{3}} i e^{\frac{2\pi i}{3} + \frac{3\pi t^2}{4n} - 2\pi t - 3\pi t k} q^{-\frac{4}{9} - k^2 - \frac{4}{3}k} \\
&\qquad + \left(\int_{-d}^{d} + \int_{d}^{d-i\frac{t}{2n}} + \int_{-d-i\frac{t}{2n}}^{-d} \right) \frac{e^{-3\pi n y^2}}{e^{2\pi(y+i\frac{t}{2n})} + e^{\frac{\pi i}{3}}}\, dy.
\end{aligned}$$

Since

$$\left| \left(\int_{d}^{d-i\frac{t}{2n}} + \int_{-d-i\frac{t}{2n}}^{-d} \right) \frac{e^{-3\pi n y^2}}{e^{2\pi(y+i\frac{t}{2n})} + e^{\frac{\pi i}{3}}}\, dy \right| \le 2 \frac{e^{-3\pi n(d^2 - (\frac{t}{2n})^2)}}{e^{2\pi d} - 1} \frac{|t|}{n},$$

we find that the sum of these integrals tends to 0 as d tends to ∞. Thus, letting $d \to \infty$ we verify that

$$\int_{-\infty-i\frac{t}{2n}}^{\infty-i\frac{t}{2n}} \frac{e^{-3\pi ny^2}}{e^{2\pi(y+i\frac{t}{2n})}+e^{\frac{\pi i}{3}}}\,dy$$
$$= \sum_{0\le k<\Re\frac{t}{2n}-\frac{2}{3}} ie^{\frac{2\pi i}{3}+\frac{3\pi t^2}{4n}-2\pi t-3\pi tk}q^{-\frac{4}{9}-k^2-\frac{4}{3}k}+\int_{-\infty}^{\infty}\frac{e^{-3\pi ny^2}}{e^{2\pi(y+i\frac{t}{2n})}+e^{\frac{\pi i}{3}}}\,dy.$$

We can also establish similar results for the other three integrals in (18) for $\Re(\frac{t}{n})>0$. We then apply these results to (18) and collect the sums to obtain

$$-\sum_{0\le k<\Re\frac{t}{2n}-\frac{2}{3}} e^{-2\pi t-3\pi tk}q^{-\frac{4}{9}-k^2-\frac{4}{3}k} - \sum_{-\Re\frac{t}{2n}-\frac{2}{3}<k\le -1} e^{2\pi t+3\pi tk}q^{-\frac{4}{9}-k^2-\frac{4}{3}k}$$
$$+\sum_{0\le k<\Re\frac{t}{2n}-\frac{1}{3}} e^{-\pi t-3\pi tk}q^{-\frac{1}{9}-k^2-\frac{2}{3}k} + \sum_{-\Re\frac{t}{2n}-\frac{1}{3}<k\le -1} e^{\pi t+3\pi tk}q^{-\frac{1}{9}-k^2-\frac{2}{3}k}.$$

Replacing k by $-k-1$ in the second and fourth sums above, we find that the four sums above cancel. Thus, for $\Re(\frac{t}{n})>0$,

$$\int_0^\infty \frac{e^{-\frac{\pi nx^2}{3}}\cos\pi tx}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx \tag{19}$$
$$=\frac{\sqrt{3}}{4i}e^{-\frac{3\pi t^2}{4n}}\Biggl\{e^{\frac{\pi i}{3}}\int_{-\infty}^{\infty}\Biggl(\frac{e^{-3\pi ny^2}}{e^{2\pi(y+i\frac{t}{2n})}+e^{\frac{\pi i}{3}}}+\frac{e^{-3\pi ny^2}}{e^{2\pi(y-i\frac{t}{2n})}+e^{\frac{\pi i}{3}}}\Biggr)\,dy$$
$$-e^{-\frac{\pi i}{3}}\int_{-\infty}^{\infty}\Biggl(\frac{e^{-3\pi ny^2}}{e^{2\pi(y+i\frac{t}{2n})}+e^{-\frac{\pi i}{3}}}+\frac{e^{-3\pi ny^2}}{e^{2\pi(y-i\frac{t}{2n})}+e^{-\frac{\pi i}{3}}}\Biggr)\,dy\Biggr\}.$$

For $\Re(\frac{t}{n})<0$, a similar process also brings us to (19). Also, for $\Re(\frac{t}{n})=0$, we directly derive (19) from (18). Therefore, for any positive number n and any number t such that $\Re(\frac{t}{n})\pm\frac{1}{3}\notin\mathbb{Z}$ and $\Re(\frac{t}{n})\pm\frac{2}{3}\notin\mathbb{Z}$, we obtain (19).

Next we must evaluate the integrals on the right side of (19). We need the modular transformation formula for θ_{11}

$$\theta_{11}\Bigl(\frac{x}{\tau},-\frac{1}{\tau}\Bigr)=-i\sqrt{-i\tau}\,e^{\pi ix^2/\tau}\,\theta_{11}(x,\tau). \tag{20}$$

Additionally, $F(x,\tau)$ and θ_{11} satisfy the transformation formulas

$$\theta_{11}(x,\tau)=-\theta_{11}(x+1,\tau)=-\theta_{11}(-x,\tau)=-e^{\pi i(2x+\tau)}\theta_{11}(x+\tau,\tau),$$
$$F(x,\tau)=-F(x+1,\tau)=-F(x+\tau,\tau)+\theta_{11}(x,\tau)=-F(-x+\tau,\tau)$$
$$=F(-x,\tau)+\theta_{11}(x,\tau).$$

We employ these formulas to evaluate the four integrals on the right-hand side of (19). Recall that $q_1=e^{-\frac{\pi}{3n}}$ and $q=e^{-3\pi n}$. We first consider the first integral on the right side of (19). Replacing τ, x, and θ by $3in$, 0, and $\frac{2}{3}-\frac{t}{2n}$, respectively, in Mordell's formula (1), we find that

$$\int_{-\infty}^{\infty}\frac{e^{-3\pi ny^2}}{e^{2\pi(y+i\frac{t}{2n})}-e^{\frac{4\pi i}{3}}}\,dy=e^{-\frac{\pi it}{n}}\int_{-\infty}^{\infty}\frac{e^{-3\pi ny^2}}{e^{2\pi y}-e^{2\pi i(\frac{2}{3}-\frac{t}{2n})}}\,dy$$
$$=e^{\frac{3\pi t^2}{4n}-2\pi t-\frac{4\pi i}{3}}q^{-\frac{4}{9}}\frac{F(\frac{2}{3}-\frac{t}{2n},-\frac{1}{3in})-3nF(2in-i\frac{3t}{2},3in)}{3in\theta_{11}(2in-i\frac{3t}{2},3in)}.$$

We are able to establish a similar result for each of the remaining three integrals. From (20), we deduce that

$$\theta_{11}\Big(2in-\tfrac{3}{2}it,3in\Big)=\frac{i}{\sqrt{3n}}e^{\frac{4}{3}\pi n-2\pi t+\frac{3\pi t^2}{4n}}\theta_{11}\Big(\frac{2}{3}-\frac{t}{2n},-\frac{1}{3in}\Big),$$
$$\theta_{11}\Big(in-\tfrac{3}{2}it,3in\Big)=\frac{i}{\sqrt{3n}}e^{\frac{1}{3}\pi n-\pi t+\frac{3\pi t^2}{4n}}\theta_{11}\Big(\frac{1}{3}-\frac{t}{2n},-\frac{1}{3in}\Big).$$

Using the evaluations of the four integrals, employing the above modular transformations for θ_{11} and the formulas satisfied by θ_{11} and F, simplifying terms, and employing the definitions of θ_{11} and F, we obtain

$$(21)\quad \frac{2}{\sqrt{3}}\int_0^{\infty}\frac{e^{-\frac{\pi nx^2}{3}}\cos\pi tx}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx$$
$$=\frac{e^{\pi t}q^{\frac{8}{9}}}{f(-e^{3\pi t}q^{\frac{4}{3}},-e^{-3\pi t}q^{\frac{2}{3}})}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m^2+\frac{7}{3}m}e^{3m\pi t}}{1+q^{2m+1}}$$
$$+\frac{e^{-\pi t}q^{\frac{8}{9}}}{f(-e^{-3\pi t}q^{\frac{4}{3}},-e^{3\pi t}q^{\frac{2}{3}})}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{m^2+\frac{7}{3}m}e^{-3m\pi t}}{1+q^{2m+1}}$$
$$\times\left\{\frac{1}{f(-\omega^2e^{-\frac{\pi it}{n}}q_1^2,-\omega^{-2}e^{\frac{\pi it}{n}})}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q_1^{m^2+m}\omega^{2m}e^{-\frac{m\pi it}{n}}}{1+q_1^{2m+1}}\right.$$
$$\left.+\frac{1}{f(-\omega^2e^{\frac{\pi it}{n}}q_1^2,-\omega^{-2}e^{-\frac{\pi it}{n}})}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q_1^{m^2+m}\omega^{2m}e^{\frac{m\pi it}{n}}}{1+q_1^{2m+1}}-1\right\}.$$

We are now ready to complete the proof. Use Lemma 1 with t replaced by $e^{\pi t}$ and employ Lemma 2 with q and t replaced by q_1 and $e^{\frac{\pi it}{n}}$, respectively. After some elementary manipulations, we find that the sum of the new left-hand sides of Lemma 1 and (11) equals

(22)

$$
\begin{aligned}
&q^{\frac{2}{9}} \sum_{m=1}^{\infty} \frac{q^{\frac{2m(m-1)}{3}}}{(-e^{\pi t} q^{\frac{1}{3}}; q^{\frac{2}{3}})_m (-e^{-\pi t} q^{\frac{1}{3}}; q^{\frac{2}{3}})_m} \\
&\quad + \frac{e^{-\frac{3\pi t^2}{4n}} q_1^2}{\sqrt{n}} \sum_{m=1}^{\infty} \frac{q_1^{6m(m-1)}}{(-e^{i\frac{\pi t}{n}} q_1^3; q_1^6)_m (-e^{-i\frac{\pi t}{n}} q_1^3; q_1^6)_m} \\
&\quad - \frac{e^{\pi t} q^{\frac{8}{9}}}{f(-e^{3\pi t} q^{\frac{4}{3}}, -e^{-3\pi t} q^{\frac{2}{3}})} \sum_{m=-\infty}^{\infty} \frac{(-1)^m q^{m^2+\frac{7}{3}m} e^{3m\pi t}}{1+q^{2m+1}} \\
&\quad - \frac{e^{-\pi t} q^{\frac{8}{9}}}{f(-e^{-3\pi t} q^{\frac{4}{3}}, -e^{3\pi t} q^{\frac{2}{3}})} \sum_{m=-\infty}^{\infty} \frac{(-1)^m q^{m^2+\frac{7}{3}m} e^{-3m\pi t}}{1+q^{2m+1}} \\
&\quad + i \frac{e^{-\frac{3\pi t^2}{4n}}}{\sqrt{3n}} \left\{ \frac{1}{f(-\omega^2 e^{-\frac{\pi i t}{n}} q_1^2, -\omega^{-2} e^{\frac{\pi i t}{n}})} \sum_{m=-\infty}^{\infty} \frac{(-1)^m q_1^{m^2+m} \omega^{2m} e^{-\frac{m\pi i t}{n}}}{1+q_1^{2m+1}} \right. \\
&\qquad \left. + \frac{1}{f(-\omega^2 e^{\frac{\pi i t}{n}} q_1^2, -\omega^{-2} e^{-\frac{\pi i t}{n}})} \sum_{m=-\infty}^{\infty} \frac{(-1)^m q_1^{m^2+m} \omega^{2m} e^{\frac{m\pi i t}{n}}}{1+q_1^{2m+1}} - 1 \right\}
\end{aligned}
$$

and the sum of the new right-hand sides of Lemma 1 and (11) equals

(23)

$$
\begin{aligned}
&q^{\frac{2}{9}} \left\{ \frac{(q^2;q^2)_\infty^3}{(q^{\frac{2}{3}}; q^{\frac{2}{3}})_\infty f(e^{3\pi t} q, e^{-3\pi t} q)} + \frac{q^{\frac{1}{3}} (q^2;q^2)_\infty^3 f(q^{\frac{1}{3}}, q^{\frac{5}{3}}) f(-e^{-3\pi t} q^2, -e^{3\pi t})}{(q^{\frac{2}{3}}; q^{\frac{2}{3}})_\infty f(q,q) f(e^{-3\pi t} q, e^{3\pi t} q)} \right. \\
&\qquad \left. \times \left(\frac{e^{-2\pi t}}{f(-e^{3\pi t} q^{\frac{4}{3}}, -e^{-3\pi t} q^{\frac{2}{3}})} - \frac{e^{-\pi t}}{f(-e^{-3\pi t} q^{\frac{4}{3}}, -e^{3\pi t} q^{\frac{2}{3}})} \right) \right\} \\
&\quad + \frac{e^{-\frac{3\pi t^2}{4n}}}{\sqrt{n}} \left\{ -\frac{(q_1^2;q_1^2)_\infty^3}{3 (q_1^6;q_1^6)_\infty f(e^{\frac{\pi i t}{n}} q_1, e^{-\frac{\pi i t}{n}} q_1)} \right. \\
&\qquad + \frac{i}{\sqrt{3}} \frac{(q_1^2;q_1^2)_\infty^4 f(-e^{\frac{\pi i t}{n}}, -e^{-\frac{\pi i t}{n}} q_1^2)}{(1-\omega^2) f(q_1,q_1) f(q_1,q_1^5) f(e^{\frac{\pi i t}{n}} q_1, e^{-\frac{\pi i t}{n}} q_1)} \\
&\qquad \left. \times \left(\frac{1}{f(-\omega^2 e^{-\frac{\pi i t}{n}} q_1^2, -\omega e^{\frac{\pi i t}{n}})} - \frac{e^{-\frac{\pi i t}{n}}}{f(-\omega^2 e^{\frac{\pi i t}{n}} q_1^2, -\omega e^{-\frac{\pi i t}{n}})} \right) \right\}.
\end{aligned}
$$

Next we prove that (23) is identically equal to zero. Using the definition of θ_{11}, the Jacobi triple product identity, and the transformation formula (20) for θ_{11}, we deduce the following formula for Ramanujan's theta function f:

$$
f(-e^{2\pi i(x+\tau)}, -e^{-2\pi i x}) = \frac{i}{\sqrt{-i\tau}} e^{-\pi i (x+\frac{\tau}{4}+\frac{x^2-x}{\tau}+\frac{1}{4\tau})} f(-e^{2\pi i \frac{x-1}{\tau}}, -e^{-2\pi i \frac{x}{\tau}}).
$$

Set $\tau = 3in$, and recall that $q_1 = e^{-\frac{\pi}{3n}}$ and $q = e^{-3\pi n}$ to obtain

(24)
$$f(-e^{2\pi ix}q^2, -e^{-2\pi ix}) = \frac{i}{\sqrt{3n}} q^{-\frac{1}{4}} e^{-\pi ix - \frac{\pi}{3n}(x-\frac{1}{2})^2} f(-e^{\frac{2\pi x}{3n}} q_1^2, -e^{-\frac{2\pi x}{3n}}).$$

Since $\lim_{x\to 0}(1-e^{-\frac{2\pi}{3n}x})/(1-e^{-2\pi ix}) = -i/(3n)$, dividing both sides of (24) by $1-e^{-2\pi ix}$ and tending x to 0 leads us to find that

$$(q^2;q^2)_\infty^3 = \frac{1}{3n\sqrt{3n}} q^{-\frac{1}{4}} q_1^{\frac{1}{4}} (q_1^2;q_1^2)_\infty^3. \tag{25}$$

Applying (24) twice with $x = \frac{1}{2} - \frac{3}{2}in - \frac{3}{2}it$ and $x = -in$, respectively, (25), and the fact that $i\sqrt{3}(1-\omega)e^{-\frac{\pi i}{3}} = 3$, and employing the Jacobi triple product identity, we obtain

$$\frac{q^{\frac{2}{9}}(q^2;q^2)_\infty^3}{(q^{\frac{2}{3}};q^{\frac{2}{3}})_\infty f(e^{3\pi t}q, e^{-3\pi t}q)} = \frac{e^{-\frac{3\pi t^2}{4n}}}{\sqrt{n}} \frac{(q_1^2;q_1^2)_\infty^3}{3(q_1^6;q_1^6)_\infty f(e^{\frac{\pi it}{n}}q_1, e^{-\frac{\pi it}{n}}q_1)}. \tag{26}$$

Applying (24) with x replaced by $\frac{1}{2} - \frac{1}{2}in$, $\frac{3it}{2}$, $\frac{1}{2} - \frac{3}{2}in$, $-\frac{3}{2}it - in$, and $\frac{3}{2}it - in$, respectively, and employing the Jacobi triple product,we obtain

(27)
$$f(q^{\frac{5}{3}}, q^{\frac{1}{3}}) = \frac{1}{\sqrt{3n}} q^{-\frac{1}{9}} f(-e^{-\frac{\pi i}{3}}q_1, -e^{\frac{\pi i}{3}}q_1) = \frac{1}{\sqrt{3n}} q^{-\frac{1}{9}} \frac{(q_1^2;q_1^2)_\infty (q_1^6;q_1^6)_\infty}{f(q_1, q_1^5)},$$

$$f(-e^{-3\pi t}q^2, -e^{3\pi t}) = -\frac{iq^{-\frac{1}{4}} q_1^{\frac{1}{4}} e^{\frac{3\pi t}{2} + \frac{3\pi t^2}{4n} - \frac{\pi it}{2n}}}{\sqrt{3n}} f(-e^{\frac{\pi it}{n}}, -e^{-\frac{\pi it}{n}}q_1^2), \tag{28}$$

$$f(q,q) = \frac{1}{\sqrt{3n}} f(q_1, q_1), \tag{29}$$

(30)
$$f(-e^{3\pi t}q^{\frac{4}{3}}, -e^{-3\pi t}q^{\frac{2}{3}}) = \frac{iq^{-\frac{1}{36}} q_1^{\frac{1}{4}} e^{-\frac{\pi i}{3} - \frac{\pi t}{2} + \frac{3\pi t^2}{4n} - \frac{\pi it}{2n}}}{\sqrt{3n}} f(-e^{-\frac{\pi it}{n}}\omega^2 q_1^2, -e^{\frac{\pi it}{n}}\omega),$$

(31)
$$f(-e^{-3\pi t}q^{\frac{4}{3}}, -e^{3\pi t}q^{\frac{2}{3}}) = \frac{iq^{-\frac{1}{36}} q_1^{\frac{1}{4}} e^{-\frac{\pi i}{3} + \frac{\pi t}{2} + \frac{3\pi t^2}{4n} + \frac{\pi it}{2n}}}{\sqrt{3n}} f(-e^{\frac{\pi it}{n}}\omega^2 q_1^2, -e^{-\frac{\pi it}{n}}\omega).$$

Employing (26)–(31) and using the fact that $e^{\frac{\pi i}{3}} = \sqrt{3}i/(1-\omega^2)$, we conclude that

$$(32)\quad \frac{q^{\frac{5}{9}}(q^2;q^2)_\infty^3 f(q^{\frac{1}{3}},q^{\frac{5}{3}})f(-e^{-3\pi t}q^2,-e^{3\pi t})}{(q^{\frac{2}{3}};q^{\frac{2}{3}})_\infty f(q,q)f(e^{-3\pi t}q,e^{3\pi t}q)}$$
$$\times\left(\frac{e^{-2\pi t}}{f(-e^{3\pi t}q^{\frac{4}{3}},-e^{-3\pi t}q^{\frac{2}{3}})}-\frac{e^{-\pi t}}{f(-e^{-3\pi t}q^{\frac{4}{3}},-e^{3\pi t}q^{\frac{2}{3}})}\right)$$
$$=-\frac{e^{-\frac{3\pi t^2}{4n}}}{\sqrt{n}}\frac{i}{\sqrt{3}}\frac{(q_1^2;q_1^2)_\infty^4 f(-e^{\frac{\pi i t}{n}},-e^{-\frac{\pi i t}{n}}q_1^2)}{(1-\omega^2)f(q_1,q_1)f(q_1,q_1^5)f(e^{\frac{\pi i t}{n}}q_1,e^{-\frac{\pi i t}{n}}q_1)}$$
$$\times\left(\frac{1}{f(-\omega^2 e^{-\frac{\pi i t}{n}}q_1^2,-\omega e^{\frac{\pi i t}{n}})}-\frac{e^{-\frac{\pi i t}{n}}}{f(-\omega^2 e^{\frac{\pi i t}{n}}q_1^2,-\omega e^{-\frac{\pi i t}{n}})}\right).$$

As a result, combining (26) and (32), we know that (23) equals 0. Thus, (22) equals 0. Therefore, comparing (21) and (22), we have proved Theorem 1. □

Second proof of Theorem 1. We now consider the equation

$$(33)\quad \frac{1}{\sqrt{3}}\int_{-\infty}^{\infty}\frac{e^{\frac{\pi i\tau x^2}{3}-2\pi z x}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = g(z;\tau)+\frac{e^{\frac{3\pi i z^2}{\tau}}}{\sqrt{-i\tau}}g\Big(-\frac{z}{\tau};-\frac{1}{\tau}\Big)$$

where

$$g(z;\tau):=\frac{e^{\frac{2\pi i\tau}{3}}}{(e^{2\pi i\tau};e^{2\pi i\tau})_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m e^{3\pi i\tau m(m+1)}}{1+e^{2\pi i z+2\pi i\tau(m+\frac{1}{2})}}.$$

Comparing the definitions of g and g_3, we find that

$$g(z;\tau)=e^{\frac{2\pi i\tau}{3}}g_3(e^{2\pi i z+\pi i(\tau+1)},e^{2\pi i\tau}).$$

We now set $\tau = in$, $q=e^{-2\pi n}$, and $q_1=e^{-\frac{2\pi}{n}}$. Using (6) with z replaced by $e^{2\pi i z}$, we get

$$\begin{aligned} g(z;in) &= \frac{q^{\frac{1}{3}}}{(q;q)_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{\frac{3m(m+1)}{2}}}{1+e^{2\pi i z}q^{(m+\frac{1}{2})}} \\ &= q^{\frac{1}{3}}\sum_{m=1}^{\infty}\frac{q^{m(m-1)}}{(-e^{2\pi i z}q^{\frac{1}{2}};q)_m(-e^{-2\pi i z}q^{\frac{1}{2}};q)_m}. \end{aligned}$$

Similarly, we obtain

$$g\Big(\frac{iz}{n};\frac{i}{n}\Big)=q_1^{\frac{1}{3}}\sum_{m=1}^{\infty}\frac{q_1^{m(m-1)}}{(-e^{-\frac{2\pi z}{n}}q_1^{\frac{1}{2}};q_1)_m(-e^{\frac{2\pi z}{n}}q_1^{\frac{1}{2}};q_1)_m}.$$

Applying these results to (33), we easily verify that proving (33) is equivalent to proving the equation in Theorem 1. So, we prove (33) instead of Theorem 1.

We first discuss the right-hand side of (33). From the definition of $g(z;\tau)$, we see that $g(z;\tau)$ is a meromorphic function of z with simple poles in $(\frac{1}{2}+\mathbb{Z})\tau+\frac{1}{2}+\mathbb{Z}$. By a direct calculation, we can determine that its residue at $-\frac{1}{2}\tau-\frac{1}{2}$ is $-q^{1/3}/(2\pi i(q;q)_\infty)$. We will find two functional equations for the function $g(z;\tau)$. By the definition of $g(z;\tau)$, we easily get

$$g(z+1;\tau)=g(z;\tau). \tag{34}$$

Using the definition of $g(z;\tau)$ and the Jacobi triple product identity, we obtain

$$\begin{aligned}
g(z+\tau;\tau)&=\frac{e^{\frac{5\pi i\tau}{3}+2\pi iz}}{(e^{2\pi i\tau};e^{2\pi i\tau})_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m e^{\pi i\tau(3m^2-m)}}{1+e^{2\pi iz+2\pi i\tau(m+\frac{1}{2})}}\\
&=e^{\frac{5\pi i\tau}{3}+2\pi iz}-\frac{e^{\frac{8\pi i\tau}{3}+4\pi iz}}{(e^{2\pi i\tau};e^{2\pi i\tau})_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m e^{\pi i\tau(3m^2+m)}}{1+e^{2\pi iz+2\pi i\tau(m+\frac{1}{2})}}\\
&=e^{\frac{5\pi i\tau}{3}+2\pi iz}-e^{\frac{8\pi i\tau}{3}+4\pi iz}+e^{3\pi i\tau+6\pi iz}g(z;\tau).
\end{aligned}$$

In particular,

$$g(z+\tau;\tau)-e^{3\pi i\tau+6\pi iz}g(z;\tau)=e^{\frac{5\pi i\tau}{3}+2\pi iz}-e^{\frac{8\pi i\tau}{3}+4\pi iz}. \tag{35}$$

Let $G(z;\tau)$ denote the right-hand side of (33). Then, using the functional equations (34) and (35), we get

$$\begin{aligned}
G(z+1;\tau)&=g(z+1;\tau)+\frac{e^{\frac{3\pi i(z+1)^2}{\tau}}}{\sqrt{-i\tau}}g\Big(-\frac{z+1}{\tau};-\frac{1}{\tau}\Big)\\
&=g(z;\tau)+\frac{e^{\frac{3\pi i(z+1)^2}{\tau}}}{\sqrt{-i\tau}}\Big\{e^{-\frac{5\pi i}{3\tau}-\frac{2\pi iz}{\tau}}-e^{-\frac{8\pi i}{3\tau}-\frac{4\pi iz}{\tau}}\\
&\qquad\qquad+e^{-\frac{3\pi i}{\tau}-\frac{6\pi iz}{\tau}}g\Big(-\frac{z}{\tau};-\frac{1}{\tau}\Big)\Big\}.
\end{aligned}$$

Thus,

$$G(z+1;\tau)-G(z;\tau)=\frac{e^{\frac{3\pi i(z+1)^2}{\tau}}}{\sqrt{-i\tau}}\Big(e^{-\frac{5\pi i}{3\tau}-\frac{2\pi iz}{\tau}}-e^{-\frac{8\pi i}{3\tau}-\frac{4\pi iz}{\tau}}\Big). \tag{36}$$

Again, using the functional equations (34) and (35), we obtain

$$\begin{aligned}
G(z+\tau;\tau)&=g(z+\tau;\tau)+\frac{e^{\frac{3\pi i(z+\tau)^2}{\tau}}}{\sqrt{-i\tau}}g\Big(-\frac{z+\tau}{\tau};-\frac{1}{\tau}\Big)\\
&=e^{\frac{5\pi i\tau}{3}+2\pi iz}-e^{\frac{8\pi i\tau}{3}+4\pi iz}+e^{3\pi i\tau+6\pi iz}g(z;\tau)\\
&\qquad+\frac{e^{\frac{3\pi iz^2}{\tau}+6\pi iz+3\pi i\tau}}{\sqrt{-i\tau}}g\Big(-\frac{z}{\tau};-\frac{1}{\tau}\Big).
\end{aligned}$$

So,

$$G(z+\tau;\tau)-e^{3\pi i\tau+6\pi iz}G(z;\tau)=e^{\frac{5\pi i\tau}{3}+2\pi iz}-e^{\frac{8\pi i\tau}{3}+4\pi iz}. \tag{37}$$

Therefore, $G(z;\tau)$ satisfies the functional equations (36) and (37). Recall that the residue of the function $g(z;\tau)$ at $-\frac{1}{2}\tau-\frac{1}{2}$ is $-q^{1/3}/(2\pi i(q;q)_\infty)$. A simple calculation shows that the residue of the function $\left(e^{3\pi iz^2/\tau}/\sqrt{-i\tau}\right)g\left(-\frac{z}{\tau};-\frac{1}{\tau}\right)$ at $-\frac{1}{2}\tau-\frac{1}{2}$ is $q^{1/3}/(2\pi i(q;q)_\infty)$. Using these results and the two functional equations satisfied by $G(z;\tau)$, we easily verify that $G(z;\tau)$ is a holomorphic function of z.

Now we discuss the left-hand side of (33). Let $H(z;\tau)$ denote the left-hand side of (33). Then, by the definition of $H(z;\tau)$, we get

$$\begin{aligned}H(z+1;\tau)-H(z;\tau)&=\frac{1}{\sqrt{3}}\int_{-\infty}^{\infty}\frac{e^{\frac{\pi i\tau x^2}{3}-2\pi zx}(e^{-2\pi x}-1)}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx\\&=\frac{e^{\frac{\pi i(3z+2)^2}{3\tau}}}{\sqrt{3}}\int_{-\infty}^{\infty}e^{\frac{\pi i\tau}{3}\{x+\frac{i}{\tau}(3z+2)\}^2}\,dx\\&\qquad-\frac{e^{\frac{\pi i(3z+1)^2}{3\tau}}}{\sqrt{3}}\int_{-\infty}^{\infty}e^{\frac{\pi i\tau}{3}\{x+\frac{i}{\tau}(3z+1)\}^2}\,dx.\end{aligned}$$

Recall that $\tau=in$. If z is a real number then we easily show that each of two integrals equals $\sqrt{3i/\tau}$. Assume that z is a complex number such that $\Im(z)\neq 0$. We consider the first integral on the right-hand side of the equation above.

$$\int_{-\infty}^{\infty}e^{\frac{\pi i\tau}{3}\{x+\frac{i}{\tau}(3z+2)\}^2}\,dx=\int_{-\infty+\frac{i}{\tau}(3z+2)}^{\infty+\frac{i}{\tau}(3z+2)}e^{\frac{\pi i\tau}{3}x^2}\,dx.$$

For a positive number t, we consider the integral

$$\int_\gamma e^{\frac{\pi i\tau}{3}x^2}\,dx$$

taken around the rectangle γ whose vertices are at the points $\pm t$ and $\pm t+\frac{i}{\tau}(3z+2)$. By Cauchy's residue theorem, we easily get that the integral above equals 0. We first evaluate

$$\int_{-t}^{-t+\frac{i}{\tau}(3z+2)}e^{\frac{\pi i\tau}{3}x^2}\,dx.$$

Let $z=a+bi$ where a and b are real and $b\neq 0$. We only need to consider three cases: $(3a+2)/n=0$, $(3a+2)/n>0$, and $(3a+2)/n<0$. If $(3a+2)/n=0$, then

$$\left|\int_{-t}^{-t+\frac{i}{\tau}(3z+2)}e^{\frac{\pi i\tau}{3}x^2}\,dx\right|\le\frac{3b}{n}e^{-\frac{\pi n}{3}t^2+\frac{3\pi b^2}{n}}.$$

Thus, $\int_{-t}^{-t+\frac{i}{\tau}(3z+2)} e^{\frac{\pi i\tau}{3}x^2}\,dx$ tends to 0 as t tends to ∞. If $(3a+2)/n$ is positive (or negative) then there is a real number c such that $-t < c < -t+(3a+2)/n$ (or $-t+(3a+2)/n < c < -t$) and

$$\left|\int_{-t}^{-t+\frac{i}{\tau}(3z+2)} e^{\frac{\pi i\tau}{3}x^2}\,dx\right| \le \frac{\sqrt{(3a+2)^2+9b^2}}{n} e^{-\frac{\pi n}{3}c^2+\frac{3\pi b^2}{n}}.$$

Thus, the integral $\int_{-t}^{-t+\frac{i}{\tau}(3z+2)} e^{\frac{\pi i\tau}{3}x^2}\,dx$ tends to 0 as t tends to ∞. Similarly, $\int_{t}^{t+\frac{i}{\tau}(3z+2)} e^{\frac{\pi i\tau}{3}x^2}\,dx$ tends to 0 as t tends to ∞. Therefore, we see that

$$\int_{-\infty+\frac{i}{\tau}(3z+2)}^{\infty+\frac{i}{\tau}(3z+2)} e^{\frac{\pi i\tau}{3}x^2}\,dx = \int_{-\infty}^{\infty} e^{\frac{\pi i\tau}{3}x^2}\,dx = \sqrt{\frac{3i}{\tau}}.$$

After a simple calculation, we obtain

$$H(z+1;\tau)-H(z;\tau) = \frac{e^{\frac{3\pi i(z+1)^2}{\tau}}}{\sqrt{-i\tau}}\left(e^{-\frac{5\pi i}{3\tau}-\frac{2\pi iz}{\tau}} - e^{-\frac{8\pi i}{3\tau}-\frac{4\pi iz}{\tau}}\right).$$

Next, we discuss $e^{-3\pi i\tau-6\pi iz}H(z+\tau;\tau)-H(z;\tau)$. After simple calculations and integration by substitution, we get

$$\begin{aligned} &e^{-3\pi i\tau-6\pi iz}H(z+\tau;\tau)-H(z;\tau)\\ &\quad= \frac{1}{\sqrt{3}}\int_{-\infty+3i}^{\infty+3i} \frac{e^{\frac{\pi i\tau x^2}{3}-2\pi zx}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx - \frac{1}{\sqrt{3}}\int_{-\infty}^{\infty} \frac{e^{\frac{\pi i\tau x^2}{3}-2\pi zx}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx. \end{aligned}$$

For a positive number s, we consider the integral

$$\int_{\delta} \frac{e^{\frac{\pi i\tau x^2}{3}-2\pi zx}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx$$

taken around the rectangle δ whose vertices are at the points $\pm s$ and $\pm s+3i$. By Cauchy's residue theorem, after some elementary algebra, we find that the integral above equals $\sqrt{3}e^{-\frac{\pi i\tau}{3}-2\pi iz}(1-e^{-\pi i\tau-2\pi iz})$. We first evaluate

$$\int_{s}^{s+3i} \frac{e^{\frac{\pi i\tau x^2}{3}-2\pi zx}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx.$$

We again recall that $\tau = in$. Since, for any y such that $0 < y < 3$,

$$\left|\frac{e^{\frac{\pi i\tau(s+yi)^2}{3}-2\pi z(s+yi)}}{e^{\frac{2\pi(s+yi)}{3}}+1+e^{-\frac{2\pi(s+yi)}{3}}}\right|$$

tends to 0 as s tends to ∞, we easily find that the integral above tends to 0 as s tends to ∞. Similarly, we deduce that

$$\int_{-s}^{-s+3i} \frac{e^{\frac{\pi i \tau x^2}{3} - 2\pi z x}}{e^{\frac{2\pi x}{3}} + 1 + e^{-\frac{2\pi x}{3}}}\, dx$$

tends to 0 as s tends to ∞. Therefore, we obtain

$$e^{-3\pi i \tau - 6\pi i z} H(z+\tau;\tau) - H(z;\tau) = e^{-\frac{\pi i \tau}{3} - 2\pi i z}(1 - e^{-\pi i \tau - 2\pi i z}).$$

After some elementary algebra and elementary manipulation, we find that $H(z;\tau)$ also satisfies the same functional equations as $G(z;\tau)$ and is a holomorphic function of z.

Let $F(z;\tau) := H(z;\tau) - G(z;\tau)$. Then, by the functional equations satisfied by H and G, we obtain

$$F(z+1;\tau) = F(z;\tau) \quad \text{and} \quad F(z+\tau;\tau) = e^{3\pi i \tau + 6\pi i z} F(z;\tau).$$

Let T be a set of complex numbers such that for any $t \in T$, $0 \le \Re(t) \le 1$ and $0 \le \Im(t) \le n$. Since T is a compact set, $F(z;\tau)$ is bounded on T. For any $t' \in \mathbb{C} \setminus T$, there are two integers k and l and a complex number t such that $t \in T$ and $t' = t + k\tau + l$. Thus, using repeatedly the functional equations satisfied by $F(z;\tau)$,

$$F(t') = F(t + k\tau + l) = F(t + k\tau) = e^{-3\pi n k^2 + 6\pi i t k} F(t).$$

Hence,

$$|F(t')| = e^{-3\pi n k^2 - 6\pi k \Im t} |F(t)| \le e^{-3\pi n\{(|k|-1)^2 - 1\}} |F(t)|. \tag{38}$$

So, we are able to say that F is bounded on $\mathbb{C}$. Therefore, by Liouville's theorem, $F(z;\tau)$ is a constant. In (38), F tends to 0 as k tends to ∞. This implies that $F(z;\tau) = 0$. Finally, we have proved (33). $\square$

4. The proof of the second identity

Proof of Theorem 2. Employing (6) with a moderate modification, we derive

$$\sum_{m=1}^{\infty} \frac{q^{\frac{2m(m-1)}{3}}}{(-e^{\pi t} q^{\frac{1}{3}}; q^{\frac{2}{3}})_m (-e^{-\pi t} q^{\frac{1}{3}}; q^{\frac{2}{3}})_m} = \frac{1}{(q^{\frac{2}{3}}; q^{\frac{2}{3}})_\infty} \sum_{m=-\infty}^{\infty} \frac{(-1)^m q^{m(m+1)}}{1 + q^{\frac{2}{3}m + \frac{1}{3}} e^{\pi t}}, \tag{39}$$

$$\sum_{m=1}^{\infty} \frac{q_1^{6m(m-1)}}{(-e^{\frac{\pi i t}{n}} q_1^3; q_1^6)_m (-e^{-\frac{\pi i t}{n}} q_1^3; q_1^6)_m} = \frac{1}{(q_1^6; q_1^6)_\infty} \sum_{m=-\infty}^{\infty} \frac{(-1)^m q_1^{9m^2+9m}}{1 + q_1^{6m+3} e^{\frac{\pi i t}{n}}}. \tag{40}$$

By a straightforward calculation, we easily find that

$$\begin{aligned}
&\sum_{m=-\infty}^{\infty} \frac{(-1)^m q^{m(m+1)}}{1+q^{\frac{2}{3}m+\frac{1}{3}}e^{\pi t}} \\
&= \sum_{m=1}^{\infty} \frac{(-1)^{m+1} q^{m(m-1)}}{1+q^{\frac{2}{3}m-\frac{1}{3}}e^{\pi t}} + \sum_{m=1}^{\infty} \frac{(-1)^m q^{m(m-1)}}{1+q^{-\frac{2}{3}m+\frac{1}{3}}e^{\pi t}} \\
&= \sum_{m=1}^{\infty} \frac{(-1)^{m+1} q^{m(m-1)}}{1+q^{\frac{2}{3}m-\frac{1}{3}}e^{\pi t}} + \sum_{m=1}^{\infty} \frac{(-1)^m q^{m(m-1)}(q^{\frac{2}{3}m-\frac{1}{3}}e^{-\pi t}+1-1)}{1+q^{\frac{2}{3}m-\frac{1}{3}}e^{-\pi t}} \\
&= \sum_{m=1}^{\infty} \frac{(-1)^{m+1} q^{m(m-1)}}{1+q^{\frac{2}{3}m-\frac{1}{3}}e^{\pi t}} + \sum_{m=1}^{\infty} \frac{(-1)^{m+1} q^{m(m-1)}}{1+q^{\frac{2}{3}m-\frac{1}{3}}e^{-\pi t}} - \sum_{m=1}^{\infty} (-1)^{m+1} q^{m(m-1)}.
\end{aligned} \tag{41}$$

Similarly, we obtain

$$\begin{aligned}
\sum_{m=-\infty}^{\infty} \frac{(-1)^m q_1^{9m^2+9m}}{1+q_1^{6m+3}e^{\frac{\pi i t}{n}}} &= \sum_{m=1}^{\infty} \frac{(-1)^{m+1} q_1^{9m^2-9m}}{1+q_1^{6m-3}e^{\frac{\pi i t}{n}}} \\
&\quad + \sum_{m=1}^{\infty} \frac{(-1)^{m+1} q_1^{9m^2-9m}}{1+q_1^{6m-3}e^{-\frac{\pi i t}{n}}} - \sum_{m=1}^{\infty} (-1)^{m+1} q_1^{9m^2-9m}.
\end{aligned} \tag{42}$$

We previously derived that

$$(q^{\frac{2}{3}};q^{\frac{2}{3}})_\infty = \frac{i(1-\omega)}{\sqrt{3n}} q^{-\frac{1}{36}} q_1^{\frac{1}{4}} e^{-\frac{\pi i}{3}} (q_1^6;q_1^6)_\infty \quad \text{and} \quad i(1-\omega)e^{-\frac{\pi i}{3}} = \sqrt{3}.$$

Thus, using these, we obtain

$$\frac{q^{-\frac{1}{36}}}{(q^{\frac{2}{3}};q^{\frac{2}{3}})_\infty} \frac{q_1^{\frac{9}{4}}}{n} = \frac{q_1^2}{\sqrt{n}\,(q_1^6;q_1^6)_\infty}. \tag{43}$$

Combining equations (39)–(43) completes the proof. □

5. Modular transformations derived from Ramanujan's identity

In this section, we derive modular transformations for third-order mock theta functions from Theorem 1.

Ramanujan's third-order mock theta functions are defined by

$$f(q) = \sum_{m=0}^{\infty} \frac{q^{m^2}}{(-q;q)_m^2}, \quad \phi(q) = \sum_{m=0}^{\infty} \frac{q^{m^2}}{(-q^2;q^2)_m}, \quad \psi(q) = \sum_{m=1}^{\infty} \frac{q^{m^2}}{(q;q^2)_m},$$

$$\chi(q) = \sum_{m=0}^{\infty} \frac{q^{m^2}}{(-\omega q;q)_m(-\omega^2;q)_m}.$$

Watson's third-order mock theta functions are defined by

$$\omega(q)=\sum_{m=1}^{\infty}\frac{q^{2m(m-1)}}{(q;q^2)_m^2},\quad \upsilon(q)=\sum_{m=1}^{\infty}\frac{q^{m(m-1)}}{(-q;q^2)_m},$$
$$\rho(q)=\sum_{m=1}^{\infty}\frac{q^{2m(m-1)}}{(\omega q;q^2)_m(\omega^2 q;q^2)_m}.$$

Gordon and McIntosh's third-order mock theta functions are defined by

$$\xi(q)=1+2\sum_{m=1}^{\infty}\frac{q^{6m(m-1)}}{(q;q^6)_m(q^5;q^6)_m},\quad \sigma(q)=\sum_{m=1}^{\infty}\frac{q^{3m(m-1)}}{(-q;q^3)_m(-q^2;q^3)_m}.$$

To apply Theorem 1 directly to these functions, we first need new representations for Ramanujan's mock theta functions. With his formula for basic hypergeometric series, Watson [1936] gave new representations for $\phi(q)$ and $\psi(q)$, namely,

$$\phi(q)=\frac{1}{(q;q)_\infty}\left(1+2\sum_{m=1}^{\infty}\frac{(-1)^m(1+q^m)q^{m(3m+1)/2}}{1+q^{2m}}\right), \tag{44}$$

$$\psi(q)=\frac{1}{(q^4;q^4)_\infty}\sum_{m=-\infty}^{\infty}\frac{(-1)^m q^{6m(m+1)+1}}{1-q^{4m+1}}. \tag{45}$$

Then, using the definition of $f(q)$ and applying (5) with z replaced by -1, we deduce that

$$f(q)=2-2\sum_{m=1}^{\infty}\frac{q^{m(m-1)}}{(-1;q)_m(-q;q)_m}.$$

Using the definition of $\phi(q)$ and applying (5) with z replaced by i and $-i$, respectively, we obtain

$$\begin{aligned}\phi(q)&=(1-i)\left(1+i\sum_{m=1}^{\infty}\frac{q^{m(m-1)}}{(i;q)_m(-iq;q)_m}\right)\\ &=(1+i)\left(1-i\sum_{m=1}^{\infty}\frac{q^{m(m-1)}}{(-i;q)_m(iq;q)_m}\right).\end{aligned}$$

Using (45) and applying (6) with q and z replaced by q^4 and q, respectively, we deduce that

$$\psi(q)=q\sum_{m=1}^{\infty}\frac{q^{4m(m-1)}}{(q;q^4)_m(q^3;q^4)_m}.$$

Using the definition of $\chi(q)$ and applying (5) with z replaced by $-\omega$ and $-\omega^2$, respectively, we have

$$\begin{aligned}\chi(q) &= (1+\omega)\left(1-\omega\sum_{m=1}^{\infty}\frac{q^{m(m-1)}}{(-\omega;q)_m(-\omega^2q;q)_m}\right)\\ &= (1+\omega^2)\left(1-\omega^2\sum_{m=1}^{\infty}\frac{q^{m(m-1)}}{(-\omega^2;q)_m(-\omega q;q)_m}\right).\end{aligned}$$

We are now ready to derive modular transformations from Theorem 1. We record here the ones which are derived directly from Theorem 1 and expressed in terms of Mordell integrals and third-order mock theta functions. Similar modular transformations can be found in [Gordon and McIntosh 2012].

Using Theorem 1 with t replaced by $n-\frac{i}{2}$ and $n+\frac{i}{2}$, respectively, we obtain

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos\pi(n-\frac{i}{2})x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}\left(\frac{\phi(q^{\frac{2}{3}})}{1+i}+i\right)-\frac{\sqrt{2}q^{\frac{1}{4}}q_1^{-\frac{1}{16}}}{(1+i)\sqrt{n}}\psi(q_1^{\frac{3}{2}}),$$

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos\pi(n+\frac{i}{2})x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}\left(\frac{\phi(q^{\frac{2}{3}})}{1-i}-i\right)-\frac{\sqrt{2}q^{\frac{1}{4}}q_1^{-\frac{1}{16}}}{(1-i)\sqrt{n}}\psi(q_1^{\frac{3}{2}}).$$

Adding the two results above and calculating straightforwardly, we have

$$\frac{4}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos\pi n x\cosh\frac{\pi x}{2}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}\phi(q^{\frac{2}{3}})-\sqrt{\frac{2}{n}}q^{\frac{1}{4}}q_1^{-\frac{1}{16}}\psi(q_1^{\frac{3}{2}}). \tag{46}$$

Using Theorem 1 with t replaced by $\frac{n}{2}-i$ and $-\frac{n}{2}-i$, respectively, we obtain

$$\begin{aligned}&\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos\pi(\frac{n}{2}-i)x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx\\ &\qquad = q^{\frac{1}{18}}\psi(q^{\frac{1}{6}})-\frac{q^{\frac{1}{16}}q_1^{-\frac{1}{4}}}{\sqrt{2n}}\phi(q_1^6)+\frac{q^{\frac{1}{16}}q_1^{-\frac{1}{4}}e^{\pi i/4}}{\sqrt{n}},\end{aligned}$$

$$\begin{aligned}&\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos\pi(\frac{n}{2}+i)x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx\\ &\qquad = q^{\frac{1}{18}}\psi(q^{\frac{1}{6}})-\frac{q^{\frac{1}{16}}q_1^{-\frac{1}{4}}}{\sqrt{2n}}\phi(q_1^6)+\frac{q^{\frac{1}{16}}q_1^{-\frac{1}{4}}e^{-\pi i/4}}{\sqrt{n}}.\end{aligned}$$

Adding the two results above and calculating straightforwardly, we find that

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos\frac{\pi n x}{2}\cosh\pi x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{1}{18}}\psi(q^{\frac{1}{6}})-\frac{q^{\frac{1}{16}}q_1^{-\frac{1}{4}}}{\sqrt{2n}}(\phi(q_1^6)-1). \tag{47}$$

Using Theorem 1 with t replaced by $n+\frac{2}{3}i$ and $-n+\frac{2}{3}i$, respectively, we obtain

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos\pi(n+\frac{2}{3}i)x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}(\omega^2+\chi(q^{\frac{1}{3}})) - \frac{q^{\frac{1}{4}}q_1}{2\sqrt{n}}(\xi(q_1)-1),$$
$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos\pi(n-\frac{2}{3}i)x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}(\omega+\chi(q^{\frac{1}{3}})) - \frac{q^{\frac{1}{4}}q_1}{2\sqrt{n}}(\xi(q_1)-1).$$

Adding the two results above and calculating straightforwardly, we find that

$$\frac{4}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos\pi n x\cosh\frac{2\pi x}{3}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}(2\chi(q^{\frac{1}{3}})-1) - \frac{q^{\frac{1}{4}}q_1}{\sqrt{n}}(\xi(q_1)-1).$$

Using Theorem 1 with t replaced by n, $\frac{n}{2}$, i, 0, $\frac{i}{2}$, $-\frac{i}{3}$, $\frac{2i}{3}$, and $\frac{n}{3}$, respectively, we obtain,

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos\pi n x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}\big(1-\tfrac{1}{2}f(q)\big) + \frac{q^{\frac{1}{4}}q_1^2}{\sqrt{n}}\omega(q_1^3),$$

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos\frac{\pi n x}{2}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = -q^{\frac{1}{18}}\psi(-q^{\frac{1}{6}}) + \frac{q^{\frac{1}{16}}q_1^2}{\sqrt{n}}\upsilon(q_1^6) \tag{48}$$

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cosh\pi x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}\omega(q^{\frac{1}{3}}) + \frac{q_1^{-\frac{1}{4}}}{\sqrt{n}}\big(1-\tfrac{1}{2}f(q_1^6)\big),$$

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}\omega(-q^{\frac{1}{3}}) + \frac{q_1^2}{\sqrt{n}}\omega(-q_1^3), \tag{49}$$

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cosh\frac{\pi x}{2}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}\upsilon(q^{\frac{2}{3}}) - \frac{q_1^{-\frac{1}{16}}}{\sqrt{n}}\psi(-q_1^{\frac{3}{2}}), \tag{50}$$

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cosh\frac{\pi x}{3}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}\rho(q^{\frac{1}{3}}) + \frac{q_1^{\frac{7}{4}}}{\sqrt{n}}\sigma(q_1^2),$$

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cosh\frac{2\pi x}{3}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}\rho(-q^{\frac{1}{3}}) + \frac{q_1}{2\sqrt{n}}(\xi(-q_1)-1),$$

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi n x^2}{3}}\cos\frac{\pi n x}{3}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = q^{\frac{2}{9}}\sigma(q^{\frac{2}{9}}) + \frac{q^{\frac{1}{36}}q_1^2}{\sqrt{n}}\rho(q_1^3).$$

Here, using (46)–(50), we give evaluations for specific Mordell integrals and new representations for Ramanujan's third-order mock theta functions ϕ, ψ, and ω.

Replacing n by $\frac{1}{2}$ in (46), we obtain

$$\frac{4}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi x^2}{6}}\cos\frac{\pi x}{2}\cosh\frac{\pi x}{2}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = e^{-\frac{\pi}{3}}\left\{\phi(e^{-\pi})-2q\psi(e^{-\pi})\right\}. \tag{51}$$

Replacing n by 2 in (47) and multiplying 2 to both sides of (47), we find

$$\frac{4}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{2\pi x^2}{3}}\cos\pi x\cosh\pi x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = e^{-\frac{\pi}{3}}\left\{2q\psi(e^{-\pi})-\phi(e^{-\pi})+1\right\}. \tag{52}$$

Adding (51) and (52), we obtain

$$\frac{4}{\sqrt{3}}\left(\int_0^\infty \frac{e^{-\frac{\pi x^2}{6}}\cos\frac{\pi x}{2}\cosh\frac{\pi x}{2}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx+\int_0^\infty \frac{e^{-\frac{2\pi x^2}{3}}\cos\pi x\cosh\pi x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx\right)=e^{-\frac{\pi}{3}}.$$

Replacing n by 2 in (48) and replacing n by $\frac{1}{2}$ in (50), we obtain

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{2\pi x^2}{3}}\cos\pi x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = -e^{-\frac{\pi}{3}}\psi(-e^{-\pi})+\frac{e^{-\frac{17}{24}\pi}}{\sqrt{2}}\upsilon(e^{-\pi}) \tag{53}$$

$$\frac{2}{\sqrt{3}}\int_0^\infty \frac{e^{-\frac{\pi x^2}{6}}\cosh\frac{\pi x}{2}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = e^{-\frac{\pi}{3}}\upsilon(e^{-\pi})-\sqrt{2}e^{\frac{\pi}{24}}\psi(-e^{-\pi}). \tag{54}$$

Comparing (53) and (54), we have

$$\int_0^\infty \frac{e^{-\frac{2\pi x^2}{3}}\cos\pi x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx = \frac{e^{-\frac{3}{8}\pi}}{\sqrt{2}}\int_0^\infty \frac{e^{-\frac{\pi x^2}{6}}\cosh\frac{\pi x}{2}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx.$$

Ramanujan [1988] recorded

$$\phi(q)+2\psi(q)=\frac{(-q;-q)_\infty}{(q;-q)_\infty^2}, \tag{55}$$

which was proved by Watson [1936]. The right-hand side of (55) can be expressed in terms of theta functions. Using (51), (52), and (55), we obtain new representations for ϕ and ψ which are

$$\phi(e^{-\pi})=\frac{2}{\sqrt{3}}e^{\frac{\pi}{3}}\int_0^\infty \frac{e^{-\frac{\pi x^2}{6}}\cos\frac{\pi x}{2}\cosh\frac{\pi x}{2}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx+\frac{1}{2}\frac{(-e^{-\pi};-e^{-\pi})_\infty}{(e^{-\pi};-e^{-\pi})_\infty^2}$$

$$\psi(e^{-\pi})=-\frac{1}{\sqrt{3}}e^{\frac{\pi}{3}}\int_0^\infty \frac{e^{-\frac{2\pi x^2}{3}}\cos\pi x\cosh\pi x}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx+\frac{1}{4}\frac{(-e^{-\pi};-e^{-\pi})_\infty}{(e^{-\pi};-e^{-\pi})_\infty^2}.$$

Replacing n by 1 in (49), we obtain a new representation for ω, namely,

$$\omega(-e^{-\pi})=\frac{1}{\sqrt{3}}e^{\frac{2\pi}{3}}\int_0^\infty \frac{e^{-\frac{\pi x^2}{3}}}{e^{\frac{2\pi x}{3}}+1+e^{-\frac{2\pi x}{3}}}\,dx.$$

Acknowledgments

The author is grateful to Bruce C. Berndt for introducing him to Ramanujan's modular transformations in Ramanujan's Lost Notebook. The author also thanks the referees for their inspiration and valuable comments.

References

[Andrews 1981] G. E. Andrews, "Mordell integrals and Ramanujan's 'lost' notebook", pp. 10–48 in *Analytic number theory* (Philadelphia, 1980), edited by M. I. Knopp, Lecture Notes in Math. **899**, Springer, Berlin, 1981. MR 83m:33004 Zbl 0482.33002

[Andrews and Garvan 1989] G. E. Andrews and F. G. Garvan, "Ramanujan's "lost" notebook, VI: The mock theta conjectures", *Adv. in Math.* **73**:2 (1989), 242–255. MR 90d:11115

[Berndt and Rankin 1995] B. C. Berndt and R. A. Rankin, *Ramanujan: letters and commentary*, History of Mathematics **9**, Amer. Math. Soc., Providence, RI, 1995. MR 97c:01034 Zbl 0842.01026

[Chern and Rhoades 2012] B. Chern and R. C. Rhoades, "The Mordell integral and quantum modular forms", preprint, 2012.

[Choi 2002] Y.-S. Choi, "Tenth order mock theta functions in Ramanujan's lost notebook, IV", *Trans. Amer. Math. Soc.* **354**:2 (2002), 705–733. MR 2002k:11022 Zbl 1043.33012

[Choi 2004] Y.-S. Choi, "Generalization of two identities in Ramanujan's lost notebook", *Acta Arith.* **114**:4 (2004), 369–389. MR 2005g:11024 Zbl 1099.11006

[Choi 2011] Y.-S. Choi, "The basic bilateral hypergeometric series and the mock theta functions", *Ramanujan J.* **24**:3 (2011), 345–386. MR 2012e:33042 Zbl 1225.33019

[Garvan 1988] F. G. Garvan, "New combinatorial interpretations of Ramanujan's partition congruences mod 5, 7 and 11", *Trans. Amer. Math. Soc.* **305**:1 (1988), 47–77. MR 89b:11081 Zbl 0641.10009

[Gordon and McIntosh 2003] B. Gordon and R. J. McIntosh, "Modular transformations of Ramanujan's fifth and seventh order mock theta functions", *Ramanujan J.* **7**:1-3 (2003), 193–222. MR 2005a:11020 Zbl 1031.11008

[Gordon and McIntosh 2012] B. Gordon and R. J. McIntosh, "A survey of classical mock theta functions", pp. 95–144 in *Partitions, q-series, and modular forms* (Gainesville, FL, 2008), edited by K. Alladi and F. Garvan, Dev. Math. **23**, Springer, New York, 2012. MR 3051186 Zbl 1246.33006

[Hickerson 1988] D. Hickerson, "A proof of the mock theta conjectures", *Invent. Math.* **94**:3 (1988), 639–660. MR 90f:11028a Zbl 0661.10059

[Kronecker 1889a] L. Kronecker, "Summirung der Gaussschen Reihen $\sum_{h=0}^{h=n-1} e^{\frac{2h^2\pi i}{n}}$", *J. Reine Angew. Math.* **105** (1889), 267–268. JFM 21.0251.01

[Kronecker 1889b] L. Kronecker, "Bemerkungen über die Darstellung von Reihen durch Integrale", *J. Reine Angew. Math.* **105** (1889), 345–351. JFM 21.0281.01

[Mordell 1920] L. J. Mordell, "The value of the definite integral $\int_{-\infty}^{\infty} a^{at^2+bt}/(e^{ct}+d)\,dt$", *Quart. J. Pure Appl. Math.* **48** (1920), 329–342. JFM 59.0197.02

[Mordell 1933] L. J. Mordell, "The definite integral $\int_{-\infty}^{\infty} e^{ax^2+bx}/(e^{ax}+d)da$ and the analytic theory of numbers", *Acta Math.* **61**:1 (1933), 323–360. MR 1555379 Zbl 0008.05501

[Ramanujan 1988] S. Ramanujan, *The lost notebook and other unpublished papers*, Narosa, New Delhi, 1988. MR 89j:01078 Zbl 0639.01023

[Siegel 1932] C. L. Siegel, "Über Riemanns Nachlaß zur analytischen Zahlentheorie", *Quell. Stud. Gesch. Math. Astro. Phys. B* **2** (1932), 45–80. Zbl 0004.10501 JFM 58.1037.07

[Watson 1936] G. N. Watson, "The final problem: an account of the mock theta functions", *J. London Math. Soc.* (2) **11**:1 (1936), 55–80. MR 1573993 Zbl 0013.11502 JFM 62.0430.02

[Zwegers 2002] S. Zwegers, *Mock theta functions*, thesis, Utrecht University, 2002, Available at http://dspace.library.uu.nl/handle/1874/878. Zbl 1194.11058

Received April 9, 2013. Revised August 19, 2013.

YOUN-SEO CHOI
THE SCHOOL OF MATHEMATICS
KOREA INSTITUTE FOR ADVANCED STUDY
85 HOEGIRO, DONGDAEMUN-GU
SEOUL 130-722
SOUTH KOREA
y-choi2@kias.re.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.272.87

THE D-TOPOLOGY FOR DIFFEOLOGICAL SPACES

J. DANIEL CHRISTENSEN, GORDON SINNAMON AND ENXIN WU

Diffeological spaces are generalizations of smooth manifolds which include singular spaces and function spaces. For each diffeological space, Iglesias-Zemmour introduced a natural topology called the D-topology. However, the D-topology has not yet been studied seriously in the existing literature. In this paper, we develop the basic theory of the D-topology for diffeological spaces. We explain that the topological spaces that arise as the D-topology of a diffeological space are exactly the Δ-generated spaces and give results and examples which help to determine when a space is Δ-generated. Our most substantial results show how the D-topology on the function space $C^\infty(M, N)$ between smooth manifolds compares to other well-known topologies.

1. Introduction

Smooth manifolds are some of the most important objects in mathematics. They contain a wealth of geometric information, such as tangent spaces, tangent bundles, differential forms, de Rham cohomology, etc., and this information can be put to great use in proving theorems and making calculations. However, the category of smooth manifolds and smooth maps is not closed under many useful constructions, such as subspaces, quotients, function spaces, etc. On the other hand, various convenient categories of topological spaces are closed under these constructions, but the geometric information is missing. Can we have the best of both worlds?

MSC2010: primary 57P99; secondary 58D99, 57R99.
Keywords: diffeological space, D-topology, topologies on function spaces, Δ-generated spaces.

Since the 1970s, the category of smooth manifolds has been enlarged in several different ways to a well-behaved category as described above, and these approaches are nicely summarized and compared in [Stacey 2011]. In this paper, we work with diffeological spaces, which were introduced by J. Souriau [1980; 1984], and in particular we study the natural topology that any diffeological space has.

A diffeological space is a set X along with a specified set of maps $U \to X$ for each open set U in $\mathbb{R}^n$ and each $n \in \mathbb{N}$, satisfying a presheaf condition, a sheaf condition, and a nontriviality condition (see Definition 2.1). Given a diffeological space X, the D-topology on X is the largest topology making all of the specified maps $U \to X$ continuous. In this paper, we make the first detailed study of the D-topology. Our results include theorems giving properties and characterizations of the D-topology as well as many examples which show the behavior that can occur and which rule out some natural conjectures.

Our interest in these topics comes from several directions. First, it is known that the topological spaces which arise as the D-topology of diffeological spaces are precisely the Δ-generated spaces [Shimakawa et al. 2010], which were introduced by Jeff Smith as a possible convenient category for homotopy theory and were studied in [Dugger 2003; Fajstrup and Rosický 2008]. Some of our results help to further understand which spaces are Δ-generated, and we include illustrative examples.

Second, for any diffeological spaces X and Y, the set $C^\infty(X, Y)$ of smooth maps from X to Y is itself a diffeological space in a natural way and thus can be endowed with the D-topology. Since the topology arises completely canonically, it is instructive to compare it with other topologies that arise in geometry and analysis when X and Y are taken to be smooth manifolds. A large part of this paper is devoted to this comparison, and again we give both theorems and illustrative examples.

Finally, this paper arose from work on the homotopy theory of diffeological spaces [Christensen and Wu 2014] and can be viewed as the topological groundwork for this project. It is for this reason that we need to focus on an approach that produces a well-behaved category, rather than working with a theory of infinite-dimensional manifolds, such as the one thoroughly developed in the book [Kriegl and Michor 1997]. We will, however, make use of results from that book, as many of the underlying ideas are related.

Here is an outline of the paper, with a summary of the main results:

In Section 2, we review some basics of diffeological spaces. For example, we recall that the category of diffeological spaces is complete, cocomplete and cartesian closed, and that it contains the category of smooth manifolds as a full subcategory. Moreover, like smooth manifolds, every diffeological space is formed by gluing together open subsets of $\mathbb{R}^n$, with the difference that n can vary and that the gluings are not necessarily via diffeomorphisms.

In Section 3, we study the D-topology of a diffeological space, which was introduced by Iglesias-Zemmour in [1985]. We show that the D-topology is determined by the smooth curves (Theorem 3.7), while diffeologies are not (Example 3.8). We recall a result of [Shimakawa et al. 2010] which says that the topological spaces arising as the D-topology of a diffeological space are exactly the Δ-generated spaces (Proposition 3.10). We give a necessary condition and a sufficient condition for a space to be Δ-generated (Propositions 3.4 and 3.11) and show that neither is necessary and sufficient (Proposition 3.12 and Example 3.14). We can associate two topologies to a subset of a diffeological space. We discuss some conditions under which the two topologies coincide (Lemmas 3.17 and 3.18, Proposition 3.21, and Corollary 4.15).

Section 4 contains our most substantial results. We compare the D-topology on function spaces between smooth manifolds with other well-known topologies. The results are (1) the D-topology is almost always strictly finer than the compact-open topology (Proposition 4.2 and Example 4.5); (2) the D-topology is always finer than the weak topology (Proposition 4.4) and always coarser than the strong topology (Theorem 4.13); (3) we give a full characterization of the D-topology as the smallest Δ-generated topology containing the weak topology (Theorem 4.7); (4) as a consequence, we show that the weak topology is equal to the D-topology if and only if the weak topology is locally path-connected (Corollary 4.9); (5) in particular, when the codomain is $\mathbb{R}^n$ or the domain is compact, the D-topology coincides with the weak topology (Corollary 4.10 and Corollary 4.14), but not always (Example 4.6).

All smooth manifolds in this paper are assumed to be Hausdorff, finite-dimensional, second-countable and without boundary.

2. Background on diffeological spaces

Here is some background on diffeological spaces. While we often cite early sources, almost all of the material in this section is in the book [Iglesias-Zemmour 2013], which we recommend as a good reference.

Definition 2.1 [Souriau 1984]. A *diffeological space* is a set X together with a specified set $\mathcal{D}_X$ of maps $U \to X$ (called *plots*) for each open set U in $\mathbb{R}^n$ and for each $n \in \mathbb{N}$, such that for all open subsets $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$:

(1) (Covering) Every constant map $U \to X$ is a plot.

(2) (Smooth compatibility) If $U \to X$ is a plot and $V \to U$ is smooth, then the composition $V \to U \to X$ is also a plot.

(3) (Sheaf condition) If $U = \bigcup_i U_i$ is an open cover and $U \to X$ is a set map such that each restriction $U_i \to X$ is a plot, then $U \to X$ is a plot.

We usually use the underlying set X to denote the diffeological space $(X, \mathscr{D}_X)$.

Definition 2.2 [Souriau 1984]. Let X and Y be two diffeological spaces, and let $f : X \to Y$ be a set map. We say that f is *smooth* if for every plot $p : U \to X$ of X, the composition $f \circ p$ is a plot of Y.

The collection of all diffeological spaces with smooth maps forms a category, which we denote $\mathfrak{D}$iff. Given two diffeological spaces X and Y, we write $C^\infty(X, Y)$ for the set of all smooth maps from X to Y. An isomorphism in $\mathfrak{D}$iff will be called a *diffeomorphism*.

Every smooth manifold M is canonically a diffeological space with the same underlying set and plots taken to be all smooth maps $U \to M$ in the usual sense. We call this the *standard diffeology* on M. By using charts, it is easy to see that smooth maps in the usual sense between smooth manifolds coincide with smooth maps between them with the standard diffeology. This gives the following standard result, which can be found, for example, in [Iglesias-Zemmour 2013, Section 4.3].

Theorem 2.3. *There is a fully faithful functor from the category of smooth manifolds to* $\mathfrak{D}$iff.

From now on, unless we say otherwise, every smooth manifold considered as a diffeological space is equipped with the standard diffeology.

Proposition 2.4 [Iglesias-Zemmour 1985]. *Given a set X, let $\mathscr{D}$ be the set of all diffeologies on X ordered by inclusion. Then $\mathscr{D}$ is a complete lattice.*

This follows from the fact that $\mathscr{D}$ is closed under arbitrary (small) intersection. The largest element in $\mathscr{D}$ is called the *indiscrete diffeology* on X, which consists of all set maps $U \to X$, and the smallest element in $\mathscr{D}$ is called the *discrete diffeology* on X, which consists of all locally constant maps $U \to X$.

The smallest diffeology on X containing a set of maps $A = \{U_i \to X\}_{i \in I}$ is called the diffeology *generated* by A. It consists of all maps $f : V \to X$ such that there exists an open cover $\{V_j\}$ of V with the property that f restricted to each V_j is either constant or factors through some element $U_i \to X$ in A via a smooth map $V_j \to U_i$. The standard diffeology on a smooth manifold is generated by any smooth atlas on the manifold. For every diffeological space X, $\mathscr{D}_X$ is generated by $\bigcup_{n \in \mathbb{N}} C^\infty(\mathbb{R}^n, X)$.

Generalizing the previous paragraph, let $A = \{f_j : X_j \to X\}_{j \in J}$ be a set of functions from some diffeological spaces to a fixed set X. Then there exists a smallest diffeology on X making all f_j smooth, and we call it the *final diffeology* defined by A. For a diffeological space X with an equivalence relation $\sim$, the final diffeology defined by the quotient map $\{X \twoheadrightarrow X/{\sim}\}$ is called the *quotient diffeology*. Similarly, let $B = \{g_k : Y \to Y_k\}_{k \in K}$ be a set of functions from a fixed set Y to some diffeological spaces. Then there exists a largest diffeology on Y making all g_k smooth, and we call it the *initial diffeology* defined by B. For a diffeological space X and a subset A of X, the initial diffeology defined by the inclusion map $\{A \hookrightarrow X\}$ is called the *subset diffeology*. More generally, we have the following well-known result:

Theorem 2.5. *The category* $\mathfrak{Diff}$ *is both complete and cocomplete.*

This is proved in [Baez and Hoffnung 2011] but can be found implicitly in earlier work. We give a brief sketch here. The forgetful functor $\mathfrak{Diff} \to \mathfrak{Set}$ to the category of sets preserves both limits and colimits since it has both left and right adjoints, given by the discrete and indiscrete diffeologies. The diffeology on the (co)limit is the initial (final) diffeology defined by the natural maps. In more detail, let $F : J \to \mathfrak{Diff}$ be a functor from a small category J and write $\bar{F}$ for the composite $J \to \mathfrak{Diff} \to \mathfrak{Set}$. Then $U \to \lim \bar{F}$ is a plot if and only if the composite $U \to \lim \bar{F} \to \bar{F}(j)$ is a plot of $F(j)$ for each $j \in \mathrm{Obj}(J)$. It is not hard to check directly that $\lim \bar{F}$ with this diffeology is $\lim F$. Similarly, $p : U \to \operatorname{colim} \bar{F}$ is a plot if and only if there is an open cover $\{U_i\}$ of U such that the restriction $p|_{U_i}$ factors as $U_i \to \bar{F}(j) \to \operatorname{colim} \bar{F}$ for some $j \in \mathrm{Obj}(J)$, with the first map a plot of $F(j)$. It is not hard to check directly that $\operatorname{colim} \bar{F}$ with this diffeology is $\operatorname{colim} F$.

The category of diffeological spaces also enjoys another convenient property:

Theorem 2.6 [Iglesias-Zemmour 1985]. *The category* $\mathfrak{Diff}$ *is cartesian closed.*

Given two diffeological spaces X and Y, the set of maps

$$\{U \to C^\infty(X, Y) \mid U \times X \to Y \text{ is smooth}\}$$

forms a diffeology on $C^\infty(X, Y)$. We call it the *functional diffeology* on $C^\infty(X, Y)$, and we always equip hom-sets with the functional diffeology. Furthermore, for each diffeological space Y, $- \times Y : \mathfrak{Diff} \rightleftharpoons \mathfrak{Diff} : C^\infty(Y, -)$ is an adjoint pair.

A smooth manifold of dimension n is formed by gluing together some open subsets of $\mathbb{R}^n$ via diffeomorphisms. A diffeological space is also formed by gluing together open subsets of $\mathbb{R}^n$ (with the standard diffeology) via smooth maps, possibly for all $n \in \mathbb{N}$. To make this precise, we introduce the following concept:

Let $\mathcal{DS}$ be the category with objects all open subsets of $\mathbb{R}^n$ for all $n \in \mathbb{N}$ and morphisms the smooth maps between them. Given a diffeological space X, we define $\mathcal{DS}/X$ to be the category with objects all plots of X and morphisms the commutative triangles

$$\begin{tikzcd} U \arrow[rr, "f"] \arrow[dr, "p"'] & & V \arrow[dl, "q"] \\ & X & \end{tikzcd}$$

with p, q plots of X and f a smooth map. We call $\mathcal{DS}/X$ the *category of plots* of X. It is equipped with a forgetful functor $F : \mathcal{DS}/X \to \mathfrak{Diff}$ sending a plot

$U \to X$ to U regarded as a diffeological space and sending the morphism displayed above to f. We can use F to show that any diffeological space X can be built out of Euclidean spaces.

Proposition 2.7. *The colimit of the functor* $F : \mathcal{DS}/X \to \mathfrak{D}\mathrm{iff}$ *is* X.

Proof. Clearly there is a natural cocone $F \to X$ sending the above commutative triangle to itself. For each diffeological space Y and cocone $g : F \to Y$, we define a set map $h : X \to Y$ by sending $x \in X$ to $g(x)(\mathbb{R}^0)$, where by abuse of notation the second x denotes the plot $\mathbb{R}^0 \to X$ with image $x \in X$. Note that h induces a (unique) cocone map since $h(p(u)) = g(p(u)) = g(p) \circ u$ for each plot $p : U \to X$ and each $u \in U$, which also implies the smoothness of h. □

The result is essentially the same as [Iglesias-Zemmour 2013, Exercise 33].

Given a diffeological space X, the category $\mathcal{DS}/X$ can be used to define geometric structures on X. See [Iglesias-Zemmour 2013; Souriau 1985; Laubinger 2006] for a discussion of differential forms and the de Rham cohomology of a diffeological space, and see [Hector 1995; Laubinger 2006] for tangent spaces and tangent bundles.

3. The D-topology

We can associate to every diffeological space the following interesting topology:

Definition 3.1 [Iglesias-Zemmour 1985; 2013, Chapter 2]. Given a diffeological space X, the final topology induced by its plots, where each domain is equipped with the standard topology, is called the *D-topology* on X.

In more detail, if $(X, \mathcal{D})$ is a diffeological space, then a subset A of X is open in the D-topology of X if and only if $p^{-1}(A)$ is open for each $p \in \mathcal{D}$. We call such subsets *D-open*. If $\mathcal{D}$ is generated by a subset $\mathcal{D}'$, then A is D-open if and only if $p^{-1}(A)$ is open for each $p \in \mathcal{D}'$.

A smooth map $X \to X'$ is continuous when X and X' are equipped with the D-topology, and so this defines a functor $D : \mathfrak{D}\mathrm{iff} \to \mathfrak{T}\mathrm{op}$ to the category of topological spaces.

Example 3.2. (1) The D-topology on a smooth manifold with the standard diffeology coincides with the usual topology on the manifold.

(2) The D-topology on a discrete diffeological space is discrete, and the D-topology on an indiscrete diffeological space is indiscrete.

Every topological space Y has a natural diffeology, called the *continuous diffeology*, whose plots $U \to Y$ are the continuous maps. This was defined in [Donato 1984, Section 2.8]. A continuous map $Y \to Y'$ between topological spaces is smooth when Y and Y' are equipped with the continuous diffeology, and so this defines a functor $C : \mathfrak{T}\mathrm{op} \to \mathfrak{D}\mathrm{iff}$.

Proposition 3.3. *The functors $D : \mathfrak{Diff} \rightleftharpoons \mathfrak{Top} : C$ are adjoint, and we have $C \circ D \circ C = C$ and $D \circ C \circ D = D$.*

Proof. The adjointness is [Shimakawa et al. 2010, Proposition 3.1], and the rest is easy. □

Proposition 3.4 [Hector 1995; Laubinger 2006]. *For each diffeological space, the D-topology is locally path-connected.*

However, not every locally path-connected space comes from a diffeological space; see Example 3.14.

3.1. *The D-topology is determined by smooth curves.*

Definition 3.5. We say that a sequence x_m in $\mathbb{R}^n$ *converges fast* to x in $\mathbb{R}^n$ if for each $k \in \mathbb{N}$ the sequence $m^k(x_m - x)$ is bounded.

Note that every convergent sequence has a subsequence which converges fast.

Lemma 3.6 (Special Curve Lemma [Kriegl and Michor 1997, p. 18]). *Let x_m be a sequence which converges fast to x in $\mathbb{R}^n$. Then there is a smooth curve $c : \mathbb{R} \to \mathbb{R}^n$ such that $c(t) = x$ for $t \leq 0$, $c(t) = x_1$ for $t \geq 1$, $c(1/m) = x_m$ for each $m \in \mathbb{Z}^+$, and c maps $[1/(m+1), 1/m]$ to the line segment joining x_{m+1} and x_m.*

Theorem 3.7. *The D-topology on a diffeological space X is determined by the set $C^\infty(\mathbb{R}, X)$, in the sense that a subset A of X is D-open if and only if $p^{-1}(A)$ is open for every $p \in C^\infty(\mathbb{R}, X)$.*

Proof. ($\Rightarrow$) This follows from the definition of the D-topology.

($\Leftarrow$) Suppose that $p^{-1}(A)$ is open for every $p \in C^\infty(\mathbb{R}, X)$. Consider a plot $q : U \to X$, and let $x \in q^{-1}(A)$. Suppose that $\{x_m\}$ converges fast to x. By the Special Curve Lemma, there is a smooth curve $c : \mathbb{R} \to U$ such that $c(1/m) = x_m$ for each m and $c(0) = x$. Since $c^{-1}(q^{-1}(A))$ is open, x_m is in $q^{-1}(A)$ for m sufficiently large. So $q^{-1}(A)$ is open in U. □

Example 3.8. Let X be $\mathbb{R}^2$ with the standard diffeology, and let Y be the set $\mathbb{R}^2$ with the diffeology generated by $C^\infty(\mathbb{R}, \mathbb{R}^2)$. Then $D(X)$ is homeomorphic to $D(Y)$ since $C^\infty(\mathbb{R}, X) = C^\infty(\mathbb{R}, Y)$, but X and Y are not diffeomorphic since the identity map $\mathbb{R}^2 \to \mathbb{R}^2$ does not locally factor through curves. In other words, the D-topology is determined by smooth curves, but the diffeology is not.

In this example, Y has the smallest diffeology such that $C^\infty(\mathbb{R}, \mathbb{R}^2)$ consists of the usual smooth curves. In contrast, by Boman's theorem [Kriegl and Michor 1997, Corollary 3.14], X has the largest diffeology such that $C^\infty(\mathbb{R}, \mathbb{R}^2)$ consists of the usual smooth curves. That is, $p : U \to X$ is a plot if and only if for every smooth function $c : \mathbb{R} \to U$, the composite $p \circ c$ is in $C^\infty(\mathbb{R}, X)$.

3.2. ***Relationship with Δ-generated topological spaces.*** Write Δ^n for the standard n-simplex in $\mathfrak{Top}$.

Definition 3.9. A topological space X is called Δ-*generated* if the following condition holds: $A \subseteq X$ is open if and only if $f^{-1}(A)$ is open in Δ^n for each continuous map $f : \Delta^n \to X$ and each $n \in \mathbb{N}$.

It is not hard to show that being Δ-generated is the same as being $\mathbb{R}$-generated or $[0, 1]$-generated; that is, one can determine the open sets of a Δ-generated space using just the continuous maps $\mathbb{R} \to X$ or $[0, 1] \to X$. This follows from the existence of a surjective continuous map $\mathbb{R} \to \Delta^n$ that exhibits Δ^n as a quotient of $\mathbb{R}$. Note the similarity to Theorem 3.7. More on Δ-generated topological spaces can be found in [Dugger 2003; Fajstrup and Rosický 2008].

Proposition 3.10 [Shimakawa et al. 2010]. *The spaces in the image of the functor D are exactly the Δ-generated topological spaces.*

Since the argument is easy, we include a proof.

Proof. Let X be a diffeological space, and consider $A \subseteq D(X)$. Suppose $f^{-1}(A)$ is open in $\mathbb{R}$ for all continuous $f : \mathbb{R} \to D(X)$. Then $f^{-1}(A)$ is open in $\mathbb{R}$ for all smooth $f : \mathbb{R} \to X$. Thus A is open in $D(X)$, and so $D(X)$ is Δ-generated.

Suppose that Y is Δ-generated. By adjointness, the identity map $D(C(Y)) \to Y$ is continuous. We claim that it is a homeomorphism, and so Y is in the image of D. Indeed, suppose $A \subseteq D(C(Y))$ is open. That is, $f^{-1}(A)$ is open in $\mathbb{R}$ for all smooth $f : \mathbb{R} \to C(Y)$. That is, $f^{-1}(A)$ is open in $\mathbb{R}$ for all continuous $f : \mathbb{R} \to Y$. Then, since Y is Δ-generated, A is open in Y. □

Because of this, it will be helpful to better understand which topological spaces are Δ-generated.

Proposition 3.11. *Every locally path-connected first-countable topological space is Δ-generated.*

Proof. Let (X, τ) be a locally path-connected first-countable topological space. Then for each $x \in X$, there exists a neighborhood basis $\{A_i\}_{i=1}^{\infty}$ of x such that

(1) each A_i is path-connected; and

(2) $A_{i+1} \subseteq A_i$.

This is because for a neighborhood basis $\{B_i\}_{i=1}^{\infty}$ of x, we can define A_1 to be the path-component of B_1 containing x and A_i to be the path-component of $A_{i-1} \cap B_i$ containing x for $i \geq 2$. Since X is locally path-connected, each A_i is open.

Now let τ' be the final topology on X for all continuous maps $\Delta^n \to (X, \tau)$ for all $n \in \mathbb{N}$. Clearly $\tau \subseteq \tau'$. Suppose A is not in τ. This means that there exists $x \in A$ such that for each $U \in \tau$ which is a neighborhood of x, there exists $x_U \in U \setminus A$. Let $\{A_i\}_{i=1}^{\infty}$ be a neighborhood basis for x with the above two properties, and write $x_n \in A_n \setminus A$ accordingly. Define $f : [0, 1] \to X$ by letting $f|_{[1/(i+1),1/i]}$ be a continuous path connecting x_{i+1} to x_i in A_i, and $f(0) = x$. It is easy to see that f is continuous for (X, τ), but $f^{-1}(A)$ is not open in [0, 1]. So A is not in τ'. □

It follows from Propositions 3.4 and 3.10 that every Δ-generated space is locally path-connected. However, not every Δ-generated space is first-countable.

Proposition 3.12. *Let X be a set with the complement-finite topology. We write* $\mathrm{card}(X)$ *for its cardinality. Then*

(1) *X is Δ-generated if* $\mathrm{card}(X) < \mathrm{card}(\mathbb{N})$ *or* $\mathrm{card}(X) \geq \mathrm{card}(\mathbb{R})$;

(2) *X is not Δ-generated if* $\mathrm{card}(X) = \mathrm{card}(\mathbb{N})$.

Note that X is not first-countable when $\mathrm{card}(X) \geq \mathrm{card}(\mathbb{R})$. This provides a counterexample to the converse of Proposition 3.11.

Proof. (1) If X is a finite set, then the complement-finite topology is the discrete topology. Hence X is Δ-generated.

Assume $\mathrm{card}(X) \geq \mathrm{card}(\mathbb{R})$, and let B be a nonclosed subset of X, that is, $B \neq X$ and $\mathrm{card}(B) \geq \mathrm{card}(\mathbb{N})$. We must construct a continuous map $f : \mathbb{R} \to X$ such that $f^{-1}(B)$ is not closed in $\mathbb{R}$. Note that in this case, every injection $\mathbb{R} \to X$ is continuous.

Take an injection $\tilde{f} : \{1/n\}_{n \in \mathbb{Z}^+} \to B$. We can extend this to an injection $f : \mathbb{R} \to X$ with $f(0) \in X \setminus B$. This map is what we are looking for.

(2) If $\mathrm{card}(X) = \mathrm{card}(\mathbb{N})$, then every continuous map $[0, 1] \to X$ is constant. Otherwise, since every point in X is closed, [0, 1] would be a disjoint union of at least two and at most countably many nonempty closed subsets, which contradicts a theorem of Sierpiński (see, e.g., [van Mill 2001, A.10.6] or the slick argument posted by Gowers [2010]). Since X is not discrete, it is not Δ-generated. □

Remark 3.13. Assume the continuum hypothesis. Then the above proposition says that a set X with the complement-finite topology is Δ-generated if and only if X is not an infinite countable set.

Here is an example showing that not every locally path-connected topological space is the D-topology of a diffeological space:

Example 3.14. As a set, let X be the disjoint union of copies of the closed unit interval indexed by the set J of countable ordinals. We write elements in X as x_a with $x \in [0, 1]$ and $a \in J$. Let Y be the quotient set $X/{\sim}$, where the only nontrivial relations are $1_a \sim 1_b$ for all $a, b \in J$. Since we will only work with Y, we denote the elements of Y in the same way as those of X. The topology on Y is generated by the following basis:

(1) the open interval (x_a, y_a) for each $0 \leq x < y \leq 1$ and $a \in J$;

(2) the set $U_{a,x} := \big(\bigcup_{a \leq b \in J} [0_b, 1_b]\big) \cup \big(\bigcup_{c<a} (x_c, 1_c]\big)$ for each $a \in J$ and $x \in [0, 1)$.

One can show that Y is locally path-connected (but not first-countable). However, Y is not Δ-generated. Indeed, let $A = \bigcup_{a \in J} (0_a, 1_a]$. Then A is not open in Y. For every continuous map $f : \Delta^n \to X$, we claim that $f^{-1}(A)$ is open in Δ^n. Otherwise, there exists $u \in f^{-1}(A)$ such that no open neighborhood of u is contained in $f^{-1}(A)$. Since the intervals (x_a, y_a) are open, we must have $f(u) = 1_a$, the common point. Choose a sequence (u_i) converging to u such that each u_i is not in $f^{-1}(A)$. Then $f(u_i) = 0_{b_i}$ for some countable ordinals b_i. Let b be a countable ordinal larger than each b_i. Then $U_{b,0}$ is an open set containing $f(u)$ but none of the $f(u_i)$, so $f(u_i)$ is not convergent to $f(u) = 1_a$, which contradicts the continuity of f.

3.3. ***Two topologies related to a subset of a diffeological space.*** Let X be a diffeological space, and let Y be a quotient set of X. Then we can give Y two topologies:

(1) the D-topology of the quotient diffeology on Y;

(2) the quotient topology of the D-topology on X.

Since $D : \mathfrak{Diff} \to \mathfrak{Top}$ is a left adjoint, these two topologies are the same.

Similarly, let X be a diffeological space, and let A be a subset of X. Then we can give A two topologies:

(1) $\tau_1(A)$: the D-topology of the subset diffeology on A;

(2) $\tau_2(A)$: the subtopology of the D-topology on X.

However, these two topologies are not always the same. In general, we can only conclude that $\tau_2(A) \subseteq \tau_1(A)$.

Example 3.15. (1) Let A be a subset of $\mathbb{R}$. Then $\tau_1(A)$ is discrete if and only if A is totally disconnected under the subtopology of $\mathbb{R}$. In particular, if $A = \mathbb{Q}$, then $\tau_1(\mathbb{Q})$ is the discrete topology, which is strictly finer than the subtopology $\tau_2(\mathbb{Q})$.

(2) Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous and nowhere differentiable function, and let $A = \{(x, f(x)) \mid x \in \mathbb{R}\}$ be its graph, equipped with the subset diffeology of $\mathbb{R}^2$. Then $\tau_1(A)$ is the discrete topology, which is strictly finer than the subtopology of $\mathbb{R}^2$. Here is the proof. Let $g : \mathbb{R} \to \mathbb{R}^2$ be a smooth map whose image is in A, and define $y, z : \mathbb{R} \to \mathbb{R}$ by $g(t) = (y(t), z(t))$. Assume that $y'(a) \neq 0$ for some $a \in \mathbb{R}$. Then by the inverse function theorem, $y : \mathbb{R} \to \mathbb{R}$ is a local diffeomorphism around a. Since $\operatorname{Im}(g) \subseteq A$, we have $z = f \circ y$, which implies that $f = z \circ y^{-1}$ around $y(a)$, contradicting nowhere-differentiability of f. Therefore, any plot of the form $\mathbb{R} \to A$ is constant. By Theorem 3.7, $\tau_1(A)$ is discrete. On the other hand, the subtopology $\tau_2(A)$ is homeomorphic to the usual topology on $\mathbb{R}$.

Definition 3.16 [Iglesias-Zemmour 2013, 2.14]. When $\tau_1(A) = \tau_2(A)$, we say that A is an *embedded subset* of X.

We are interested in conditions under which this holds.

Lemma 3.17. *Let A be a convex subset of $\mathbb{R}^n$. Then A is an embedded subset of $\mathbb{R}^n$.*

Proof. Following the idea of the proof of [Kriegl and Michor 1997, Lemma 24.6(3)], let $B \subseteq A$ be closed in the $\tau_1(A)$-topology, and let $\overline{B}$ be the closure of B in A for the $\tau_2(A)$-topology. Note that the $\tau_2(A)$-topology is the same as the subtopology of $\mathbb{R}^n$. Hence, for any $b \in \overline{B}$, we can find a sequence b_n in B which converges fast to b. Since A is convex, the Special Curve Lemma (Lemma 3.6) says that there is a smooth curve $c : \mathbb{R} \to A$ such that $c(0) = b$ and $c(1/n) = b_n$ for each $n \in \mathbb{Z}^+$. Therefore, $b \in B$ by the definition of the D-topology. □

Lemma 3.18. *If A is a D-open subset of a diffeological space X, then A is an embedded subset of X.*

Proof. Let B be in $\tau_1(A)$. To show that B is in $\tau_2(A)$, it suffices to show that B is D-open in X. Let $p : U \to X$ be an arbitrary plot of X. Since A is D-open in X, $p^{-1}(A)$ is an open subset of U. Hence, the composition of $p^{-1}(A) \hookrightarrow U \to X$ is also a plot for X, which factors through the inclusion map $A \hookrightarrow X$. Since $B \in \tau_1(A)$, $(p|_{p^{-1}(A)})^{-1}(B)$ is open in $p^{-1}(A)$, which implies that $p^{-1}(B)$ is open in U. Thus B is D-open in X, as required. □

Example 3.19. $\mathrm{GL}(n, \mathbb{R})$ is D-open in $M(n, \mathbb{R}) \cong \mathbb{R}^{n^2}$, so it is an embedded subset.

Also see Corollary 4.15 for another example. Note that Lemma 3.18 is not true if we change D-open to D-closed.

Example 3.20. Let $A = \{1/n\}_{n\in\mathbb{Z}^+} \cup \{0\} \subset \mathbb{R}$. Then A is D-closed in $\mathbb{R}$. It is easy to check that $\tau_1(A)$ is discrete and is strictly finer than $\tau_2(A)$.

Proposition 3.21. *Let X be a diffeological space and let A be a subset of X. If there exists a D-open neighborhood C of A in X together with a smooth retraction $r : C \to A$, then A is embedded in X. (Here both C and A are equipped with the subset diffeologies from X.)*

Proof. Let $B \in \tau_1(A)$. Then $r^{-1}(B) \in \tau_1(C) = \tau_2(C)$ is D-open in X. Therefore, $B = A \cap r^{-1}(B) \in \tau_2(A)$. □

Example 3.22. Given a smooth manifold M of dimension $n > 0$, by the strong Whitney embedding theorem, there is a smooth embedding $M \hookrightarrow \mathbb{R}^{2n}$. If we view M as a subset of $\mathbb{R}^{2n}$, then it is an embedded subset, since there is an open tubular neighborhood U of M in $\mathbb{R}^{2n}$ together with a smooth retraction $U \to M$.

4. The D-topology on function spaces

Let M and N be smooth manifolds. Recall that the set $C^\infty(M, N)$ of smooth maps from M to N has a functional diffeology described just after Theorem 2.6. In this section, we consider the topological space obtained by taking the D-topology associated to this diffeology, and we compare it to other well-known topologies on this set: the compact-open topology, the weak topology, and the strong topology.

Here is a review of these three topologies and their relationship. The books [Hirsch 1976; Kriegl and Michor 1997; Michor 1980] are good references for the weak and strong topologies.

The compact-open topology on $C^\infty(M, N)$ has a subbasis which consists of the sets $A(K, W) = \{f \in C^\infty(M, N) \mid f(K) \subseteq W\}$, where K is a nonempty compact subset of M and W is an open subset of N. (This makes sense for any diffeological spaces M and N, where K is then required to be compact in $D(M)$ and W to be open in $D(N)$.)

We now describe a subbasis for the weak topology on $C^\infty(M, N)$. For $r \in \mathbb{N}$, (U, ϕ) a chart of M, (V, ψ) a chart of N, $K \subseteq U$ compact, $f \in C^\infty(M, N)$ with $f(K) \subseteq V$, and $\epsilon > 0$, we define the set $N^r(f, (U, \phi), (V, \psi), K, \epsilon)$ to be $\{g \in C^\infty(M, N) \mid g(K) \subseteq V$ and $\|D^i(\psi \circ f \circ \phi^{-1})(x) - D^i(\psi \circ g \circ \phi^{-1})(x)\| < \epsilon$ for each $x \in \phi(K)$ and each multi-index i with $|i| \leq r\}$. These sets form a subbasis for the weak topology. Here $i = (i_1, \ldots, i_m)$ is a multi-index in $\mathbb{N}^m$ with $m = \dim(M)$, $|i| = i_1 + \cdots + i_m$, and D^i is the differential operator $\partial^{|i|}/(\partial x_1^{i_1} \cdots \partial x_m^{i_m})$.

A subbasis for the strong topology on $C^\infty(M, N)$ is similar, but it allows constraints using multiple charts. More precisely, if $N^r(f, (U_i, \phi_i), (V_i, \psi_i), K_i, \epsilon_i)$ is a family of subbasic sets for the weak topology such that the collection $\{U_i\}$ is locally finite, then the intersection of this family is a subbasic set for the strong topology. In fact, one can show that these intersections form a base for the strong topology.

Each of these is at least as fine as the previous one, that is,

$$\text{compact-open topology} \subseteq \text{weak topology} \subseteq \text{strong topology}.$$

The first inclusion is proved in Lemma A.2, and the second is clear. The compact-open topology and the weak topology coincide if and only if M or N is zero-dimensional (see Example 4.5). Moreover, the weak topology and the strong topology coincide if the domain M is compact and are different if M is noncompact and N has positive dimension (see [Hirsch 1976, pp. 35–36]).

Now we start our comparison of the D-topology with these topologies. The following lemma is needed for the subsequent proposition.

Lemma 4.1. *Let X and Y be two diffeological spaces such that $D(X)$ is locally compact Hausdorff. Then the natural bijection $D(X \times Y) \to D(X) \times D(Y)$ is a homeomorphism.*

Note that when X is a smooth manifold, $D(X)$ is locally compact Hausdorff.

Proof. First observe that the natural bijection $D(U \times V) \to D(U) \times D(V)$ is a homeomorphism for U and V open subsets of Euclidean spaces, since in this case the D-topology is the usual topology. The functors $D : \mathfrak{Diff} \to \mathfrak{Top}$, $Z \times - : \mathfrak{Diff} \to \mathfrak{Diff}$ for any diffeological space Z and $W \times - : \mathfrak{Top} \to \mathfrak{Top}$ for any locally compact Hausdorff space W all preserve colimits since they are left adjoints. Thus the claim follows from Proposition 2.7, using that $D(X)$ is locally compact Hausdorff, as is each $D(U)$ for U an open subset of some Euclidean space. □

For general X and Y, one can show using a similar argument that the D-topology on $D(X \times Y)$ corresponds under the bijection above to the smallest Δ-generated topology containing the product topology on $D(X) \times D(Y)$.

Proposition 4.2. *For diffeological spaces X and Y, the D-topology on $C^\infty(X, Y)$ contains the compact-open topology.*

This result is a stepping stone to proving the stronger statement that the D-topology contains the weak topology.

Proof. Recall that the compact-open topology has a subbasis which consists of the sets $A(K, W) = \{f \in C^\infty(X, Y) \mid f(K) \subseteq W\}$, where K is a nonempty compact subset of $D(X)$ and W is an open subset of $D(Y)$. We will show that each $A(K, W)$ is D-open. Let $\phi : U \to C^\infty(X, Y)$ be a plot of $C^\infty(X, Y)$. Since the corresponding map $\bar{\phi} : U \times X \to Y$ is smooth, $\bar{\phi}^{-1}(W)$ is open in $D(U \times X)$. So for each $u \in \phi^{-1}(A(K, W))$, $\{u\} \times K$ is in the open set $\bar{\phi}^{-1}(W)$. Note that the natural map $D(U \times X) \to D(U) \times D(X)$ is a homeomorphism by Lemma 4.1. By the compactness of K and the definition of the product topology, $V \times K \subseteq \bar{\phi}^{-1}(W)$ for some open neighborhood V of u in U, which implies that $\phi^{-1}(A(K, W))$ is open in U. Thus $A(K, W)$ is open in the D-topology. □

We will see in Example 4.5 that the D-topology is almost always strictly finer than the compact-open topology.

The next lemma will be used to show that the D-topology contains the weak topology for function spaces between smooth manifolds.

Lemma 4.3. *Let U be an open subset in $\mathbb{R}^n$ and let i be a multi-index in $\mathbb{N}^n$. Then $D^i : C^\infty(U, \mathbb{R}) \to C^\infty(U, \mathbb{R})$ is smooth.*

Proof. Let $\phi : V \to C^\infty(U, \mathbb{R})$ be a plot with $\dim(V) = m$. This means that the associated map $\bar{\phi} : V \times U \to \mathbb{R}$ defined by $\bar{\phi}(v, u) = \phi(v)(u)$ is smooth. Write j for the multi-index $(0_m, i) \in \mathbb{N}^{m+n}$, with 0_m a sequence of m zeros. Then $D^j(\bar{\phi}) : V \times U \to \mathbb{R}$ is smooth. Since $D^j(\bar{\phi})(v, u) = D^i(\phi(v))(u)$, $D^i \circ \phi$ is a plot, which implies the smoothness of D^i. □

Note that the smoothness of D^i does not imply its continuity in general. It is an easy exercise that for $|i| > 0$ and $n > 0$, D^i is not continuous in the compact-open topology but is continuous in both the weak and strong topologies.

Now we can compare the D-topology with the weak topology for function spaces between smooth manifolds.

Proposition 4.4. *Let M and N be smooth manifolds. Then the D-topology on $C^\infty(M, N)$ contains the weak topology.*

Proof. Recall that the weak topology on $C^\infty(M, N)$ has the sets

$$N^r(f, (U, \phi), (V, \psi), K, \epsilon),$$

described at the beginning of Section 4, as a subbasis.

Let $p : W \to C^\infty(M, N)$ be a plot, that is,

$$\bar{p} : W \times M \to N \quad \text{given by } \bar{p}(w, x) = p(w)(x)$$

is smooth. If $w \in p^{-1}(N^r(f, (U, \phi), (V, \psi), K, \epsilon))$, then by Proposition 4.2, Lemma 4.3, and the facts that ϕ and ψ are diffeomorphisms, only finitely many differentials are considered, K is compact and V is open, it is not hard to see that there exists an open neighborhood W' of w in W such that

$$W' \subseteq p^{-1}\big(N^r(f, (U, \phi), (V, \psi), K, \epsilon)\big).$$

Therefore, $N^r(f, (U, \phi), (V, \psi), K, \epsilon)$ is D-open. □

Since the weak topology is almost always strictly finer than the compact-open topology, so is the D-topology.

Example 4.5. The D-topology on $C^\infty(\mathbb{R}, \mathbb{R})$ is strictly finer than the compact-open topology. To prove this, consider $U = N^1(\hat{0}, (\mathbb{R}, \mathrm{id}), (\mathbb{R}, \mathrm{id}), [-1, 1], 1)$, where $\hat{0}$ is the zero function. This is open in the weak topology and thus is open in the D-topology. We claim that no open neighborhood of $\hat{0}$ in the compact-open topology of $C^\infty(\mathbb{R}, \mathbb{R})$ is contained in U. Otherwise, we may assume $\hat{0} \in A(K, (-\epsilon, \epsilon)) \subseteq U$ for some $\epsilon > 0$ and some compact K, since if $\hat{0} \in A(K_1, W_1) \cap \cdots \cap A(K_m, W_m)$, then $0 \in W_i$ for each i and

$$\hat{0} \in A(K_1 \cup \cdots \cup K_m, W_1 \cap \cdots \cap W_m) \subseteq A(K_1, W_1) \cap \cdots \cap A(K_m, W_m).$$

Then clearly $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = (\epsilon/2)\sin(2x/\epsilon)$ is in $A(K, (-\epsilon, \epsilon))$ for any K. But f is not in U since $f'(0) = 1$.

Using a similar argument, with bump functions, one can show that when M and N are smooth manifolds of dimension at least 1, then the weak topology is strictly finer than the compact-open topology. Thus the D-topology is strictly finer than the compact-open topology in this situation.

In general, the weak topology is different from the D-topology on $C^\infty(M, N)$.

Example 4.6. (1) Let $\mathbb{N}$ and $\{0, 1\}$ be equipped with the discrete diffeologies. Let $f : \mathbb{N} \to \{0, 1\}$ be the constant function sending everything to 0, and let $f_n : \mathbb{N} \to \{0, 1\}$ be defined by $f_n^{-1}(0) = \{0, 1, \dots, n\}$. Note that f_n converges to f in the weak topology for the following reason. Since each element in the subbasis of the weak topology depends only on the values of the function and its derivatives on a compact subset of $\mathbb{N}$, any of them containing f must contain all f_n for n large enough.

On the other hand, we claim that for each n there does not exist a continuous path $F : [0, 1] \to C^\infty(\mathbb{N}, \{0, 1\})$ with $F(0) = f_n$ and $F(1) = f$, where the codomain is given the weak topology. Since the weak topology contains the compact-open topology, such an F gives rise to a continuous function $[0, 1] \times \mathbb{N} \to \{0, 1\}$, that is, a homotopy from $D(f_n)$ to $D(f)$. Since these maps are clearly not homotopic, no such F exists.

Thus the weak topology is not locally path-connected. It follows from Proposition 3.4 that the weak topology is different from the D-topology on $C^\infty(\mathbb{N}, \{0, 1\})$.

The above argument in fact shows that every continuous path in $C^\infty(\mathbb{N}, \{0, 1\})$ with respect to a topology containing the compact-open topology is constant. In particular, this holds for the D-topology, and since the D-topology is Δ-generated, it must be discrete.

(2) Let X be a countable disjoint union of copies of S^1; that is, $X = \coprod_{i \in \mathbb{N}} X_i$ with each $X_i = S^1$. Then the weak topology on $C^\infty(X, S^1)$ is not locally path-connected, by a similar argument with $f : X \to S^1$ defined by $f|_{X_i} = \mathrm{id} : X_i \to S^1$ and $f_n : X \to S^1$ defined by

$$f_n|_{X_i} = \begin{cases} \mathrm{id} & \text{if } i = 0, 1, \dots, n, \\ -\mathrm{id} & \text{otherwise.} \end{cases}$$

(3) The weak topology on $C^\infty(\mathbb{R}^2 \setminus (\{0\} \times \mathbb{Z}), S^1)$ is not locally path-connected, by a similar argument with $f : \mathbb{R}^2 \setminus (\{0\} \times \mathbb{Z}) \to S^1$ defined by

$$f(x, y) = \frac{1 - e^{2\pi(x+iy)}}{|1 - e^{2\pi(x+iy)}|},$$

and $f_n : \mathbb{R}^2 \setminus (\{0\} \times \mathbb{Z}) \to S^1$ defined by

$$f_n(x, y) = f(x, \phi_n(y)),$$

where $\phi_n : \mathbb{R} \to \mathbb{R}$ is a strictly increasing smooth function with $\phi_n(t) = t$ for $|t| \leq n$ and $|\phi_n(t)| < n + 1$ for all t.

These examples all show that the weak topology is not locally path-connected, and, in particular, that it is not Δ-generated. The D-topology is a Δ-generated topology containing the weak topology, and the following theorem says that, given this, it is as close to the weak topology as possible.

Theorem 4.7. *For M and N smooth manifolds, the D-topology on $C^\infty(M, N)$ is the smallest Δ-generated topology containing the weak topology.*

Proof. First note that by Proposition 4.4, the D-topology contains the weak topology, and by Proposition 3.10, the D-topology is Δ-generated. So we must prove that the D-topology on $C^\infty(M, N)$ is contained in every Δ-generated topology containing the weak topology.

So let τ be a Δ-generated topology containing the weak topology and assume that $A \subseteq C^\infty(M, N)$ is not open in τ. Since τ is Δ-generated, there is a τ-continuous map $p : \mathbb{R} \to C^\infty(M, N)$ such that $p^{-1}(A)$ is not open in $\mathbb{R}$. Since τ contains the weak topology, p is weakly continuous. By composing with a translation in $\mathbb{R}$, we can assume that 0 is a noninterior point of $p^{-1}(A)$. Thus we can find a sequence t_r of real numbers converging to 0 so that $p(t_r) \notin A$ for each r. By Theorem A.5, there is a smooth curve $q : \mathbb{R} \to C^\infty(M, N)$ such that $q(2^{-j}) = p(t_{r_j}) \notin A$ for each j and $q(0) = p(0)$. This shows that A is not open in the D-topology. □

Since every Δ-generated space is locally path-connected (see Propositions 3.4 and 3.10), the previous result is in fact a special case of the next result.

Theorem 4.8. *Let M and N be smooth manifolds. Then the D-topology on $C^\infty(M, N)$ is the smallest locally path-connected topology containing the weak topology.*

Proof. Suppose τ is a locally path-connected topology that contains the weak topology, A is not τ-open, and $f \in A$ is not τ-interior to A. Since the weak topology on $C^\infty(M, N)$ is first-countable, there is a countable weak neighborhood basis $(W_r)_{r=1}^\infty$ of f. Contained in each W_r there is a path-connected τ-neighborhood T_r of f. For each r, choose an $f_r \in T_r \setminus A$ and a τ-continuous (and therefore weakly continuous) path from f to f_r lying entirely in $T_r \subseteq W_r$. We can concatenate these paths to produce a weakly continuous path p such that $p(0) = f$ and $p(2^{-r}) = f_r$. By Theorem A.5, there is a smooth curve $q : \mathbb{R} \to C^\infty(M, N)$ such that $q(0) = f$ and $q(2^{-j}) = f_{r_j}$. Then $q^{-1}(A)$ contains 0 but not 2^{-j} for any j, so A is not open in the D-topology. □

As a corollary, we have the following necessary and sufficient condition for the weak topology to be equal to the D-topology.

Corollary 4.9. *Let M and N be smooth manifolds. Then the weak topology on $C^\infty(M, N)$ coincides with the D-topology if and only if the weak topology is locally path-connected.*

Proof. This follows from Theorem 4.8 (or from Theorem 4.7, using that the weak topology is second-countable [Hirsch 1976, pp. 35–36]). □

This allows us to give a situation in which the D-topology and the weak topology coincide. (See also Corollary 4.14.)

Corollary 4.10. *For M a smooth manifold, the weak topology on $C^\infty(M, \mathbb{R}^n)$ coincides with the D-topology.*

Proof. By Lemma A.3, the weak topology on $C^\infty(M, \mathbb{R}^n)$ has a basis of convex sets. A linear path is smooth and hence weakly continuous, so it follows that this topology is locally path-connected. □

Our next goal is to show that the D-topology is contained in the strong topology. We first need some preliminary results.

Lemma 4.11. *Let M be a smooth manifold and let N be an open subset of $\mathbb{R}^d$. Then the D-topology on $C^\infty(M, N)$ is contained in any topology that contains the weak topology and has a basis of convex sets.*

Here we say that a subset of $C^\infty(M, N)$ is *convex* if it is convex when regarded as a subset of the real vector space $C^\infty(M, \mathbb{R}^d)$.

Proof. A convex set isn't necessarily path-connected, since linear paths may not be continuous. Thus Theorem 4.8 doesn't apply directly. However, in the proof of Theorem 4.8, all that is used is that the subsets T_r are path-connected in the weak topology. Since linear paths are smooth, they are weakly continuous, and so the proof goes through. □

Lemma 4.12. *Let M be a smooth manifold and let N be an open subset of $\mathbb{R}^d$. Then $C^\infty(M, N)$ is an open subspace of $C^\infty(M, \mathbb{R}^d)$ when both are equipped with the strong topology.*

Proof. We first prove that the strong topology on $C^\infty(M, N)$ is the subspace topology of the strong topology on $C^\infty(M, \mathbb{R}^d)$. Since the inclusion map $N \to \mathbb{R}^d$ induces a continuous map in the strong topologies (see [Hirsch 1976, Exercise 10(b), p. 65]), the intersection of a strong open set in $C^\infty(M, \mathbb{R}^d)$ with $C^\infty(M, N)$ is open in $C^\infty(M, N)$. On the other hand, the data for each weak subbasic set A in $C^\infty(M, N)$ defines a weak subbasic set in $C^\infty(M, \mathbb{R}^d)$ whose intersection with $C^\infty(M, N)$ is A. Since the strong subbasic sets are certain intersections of the weak subbasic sets, our claim follows.

Now we show that $C^\infty(M, N)$ is an open subset of $C^\infty(M, \mathbb{R}^d)$, following the argument in Lemma A.2. For $f \in C^\infty(M, N)$, choose charts for M and N and compact sets $K_i \subseteq M$ as described in Lemma A.1(b). Then

$$f \in \bigcap_{i=1}^{\infty} N^0(f, (U_i, \phi_i), (N, \mathrm{id}), K_i, 1) \subseteq C^\infty(M, N),$$

where each $N^0(f, (U_i, \phi_i), (N, \mathrm{id}), K_i, 1)$ is understood to be a subbasic set for $C^\infty(M, \mathbb{R}^d)$. So $C^\infty(M, N)$ is open in the strong topology. □

Theorem 4.13. *Let M and N be smooth manifolds. Then the D-topology on $C^\infty(M, N)$ is contained in the strong topology.*

Proof. Choose an embedding $N \hookrightarrow \mathbb{R}^d$, and let U be an open tubular neighborhood of N in $\mathbb{R}^d$, so that the inclusion $i : N \to U$ has a smooth retract $r : U \to N$. Since i and r induce continuous maps in both the strong topology (see [Hirsch 1976, Exercise 10, p. 65]) and the D-topology (an easy argument), $C^\infty(M, N)$ is a subspace of $C^\infty(M, U)$ when both are equipped with either of these topologies. So if these topologies agree on $C^\infty(M, U)$, then they agree on $C^\infty(M, N)$. Thus it suffices to prove the result when N is open in $\mathbb{R}^d$. Assume that this is the case.

We first prove that the strong topology on $C^\infty(M, \mathbb{R}^d)$ has a basis of convex sets. If $A := \bigcap_i N^r(f, (U_i, \phi_i), (V_i, \psi_i), K_i, \epsilon_i)$ is a basic open set of the strong topology, as described at the beginning of Section 4, and if $g \in A$, then by the proof of Lemma A.3,

$$g \in \bigcap_i N^r(g, (U_i, \phi_i), (\mathbb{R}^d, \mathrm{id}), K_i, \epsilon_i''') \subseteq A,$$

which shows that A is covered by convex strong open sets.

By Lemma 4.12, $C^\infty(M, N)$ is open in $C^\infty(M, \mathbb{R}^d)$, so it too has a basis of convex sets. Thus, by Lemma 4.11, the D-topology on $C^\infty(M, N)$ is contained in the strong topology. □

Corollary 4.14. *Let M and N be smooth manifolds with M compact. Then the D-topology on $C^\infty(M, N)$ coincides with the weak topology.*

Proof. The D-topology is trapped between the weak topology (Proposition 4.4) and the strong topology (Theorem 4.13), and these coincide when M is compact. □

Here is one application of our results:

Corollary 4.15. *Let M be a smooth compact manifold, and let $\mathfrak{Diff}(M)$ be the set of all diffeomorphisms from M to itself with the subset diffeology of $C^\infty(M, M)$. Then $\mathfrak{Diff}(M)$ is D-open in $C^\infty(M, M)$. Hence, $\mathfrak{Diff}(M)$ is an embedded subset of $C^\infty(M, M)$ (see Definition 3.16).*

Proof. As mentioned in Corollary 4.14, when M is compact, the weak, strong and D-topologies on $C^\infty(M, M)$ all coincide. The first claim is then the restatement of [Hirsch 1976, Theorem 2.1.7], and the second part follows from Lemma 3.18. □

Similarly, many results in [Hirsch 1976, Chapter 2] can be translated into results for the D-topology.

When M is noncompact and N has positive dimension, the weak topology is different from the strong topology [Hirsch 1976, pp. 35–36]. Since the weak topology and the D-topology coincide for $C^\infty(M, \mathbb{R}^n)$, it follows that the D-topology and the strong topology are different for $C^\infty(M, \mathbb{R}^n)$ when M is noncompact. We can make this explicit in the next example.

Example 4.16. It is not hard to show that the strong topology on $C^\infty(\mathbb{R}, \mathbb{R})$ has a basis $\{B^k_\delta(f) \mid k \in \mathbb{N},\ \delta : \mathbb{R} \to \mathbb{R}^+ \text{ continuous},\ f \in C^\infty(\mathbb{R}, \mathbb{R})\}$, where

$$B^k_\delta(f) = \left\{ g \in C^\infty(\mathbb{R}, \mathbb{R}) \;\middle|\; \sum_{i=0}^{k} (f^{(i)}(x) - g^{(i)}(x))^2 < \delta(x) \text{ for each } x \in \mathbb{R} \right\}.$$

On the other hand, the D-topology agrees with the weak topology on $C^\infty(\mathbb{R}, \mathbb{R})$, so it has a basis $\{\tilde{B}^k_\epsilon(f) \mid k \in \mathbb{N}, \epsilon \in \mathbb{R}^+, f \in C^\infty(\mathbb{R}, \mathbb{R})\}$, where

$$\tilde{B}^k_\epsilon(f) = \left\{ g \in C^\infty(\mathbb{R}, \mathbb{R}) \;\middle|\; \sum_{i=0}^{k} (f^{(i)}(x) - g^{(i)}(x))^2 < \epsilon \text{ for each } x \text{ in } [-k, k] \right\}.$$

It follows that the strong topology is strictly finer than the D-topology on $C^\infty(\mathbb{R}, \mathbb{R})$.

On the other hand, it can be the case that the D-topology is different from the weak topology but agrees with the strong topology. For example, this happens in case (1) of Example 4.6, where it is easy to see that the strong topology is also discrete.

Remark 4.17. The book [Kriegl and Michor 1997] also studies function spaces between smooth manifolds, but uses a different smooth structure on the function space to ensure that the resulting object has the desired local models. By Lemma 42.5 of that book, their smooth structure has fewer smooth curves than the diffeology studied here, and as a result the natural topology discussed in their Remark 42.2 is larger than the D-topology. In fact, according to that remark, it is larger than the strong topology (which they call the WO^∞-topology).

Appendix: The weak topology on function spaces

In this appendix, our goal is to prove a theorem about the weak topology on function spaces which is analogous to the Special Curve Lemma (Lemma 3.6). This is Theorem A.5. Before proving the theorem, we collect together and prove some basic results about the weak topology on function spaces and state the following lemma.

Lemma A.1. *Let M and N be smooth manifolds.*

(a) *There exist a locally finite countable atlas $\{(U_i, \phi_i)\}_{i\in\mathbb{N}}$ of M and a compact set $K_i \subseteq U_i$, for each i, such that $M = \bigcup_i \mathring{K}_i$, where $\mathring{K}_i$ denotes the interior of K_i.*

(b) *For any smooth map $f : M \to N$, there exist $\{(U_i, \phi_i, K_i)\}_{i\in\mathbb{N}}$ as in (a) and a countable atlas $\{(V_i, \psi_i)\}_{i\in\mathbb{N}}$ of N such that $f(K_i) \subseteq V_i$ for each i.*

Recall that for M and N smooth manifolds, the weak topology on $C^\infty(M, N)$ has as subbasic neighborhoods the sets $N^r(f, (U, \phi), (V, \psi), K, \epsilon)$ described at the beginning of Section 4.

Lemma A.2. *Let M and N be smooth manifolds. Then the weak topology on $C^\infty(M, N)$ contains the compact-open topology.*

Proof. Consider $A(K, W) = \{g \in C^\infty(M, N) \mid g(K) \subseteq W\}$, where $K \subseteq M$ is compact and $W \subseteq N$ is open. Let $f \in A(K, W)$. Choose charts for M and N and compact sets K_i as described in Lemma A.1(b). Choose j so that $K \subseteq \bigcup_{i=1}^{j} K_i$. Then

$$f \in \bigcap_{i=1}^{j} N^0(f, (U_i, \phi_i), (V_i \cap V, \psi_i), K_i \cap K, 1) \subseteq A(K, W),$$

so $A(K, W)$ is open in the weak topology. □

Lemma A.3. *Let M be a smooth manifold. The sets $N^r(f, (U, \phi), (\mathbb{R}^d, \mathrm{id}), K, \epsilon)$, where $r \in \mathbb{N}$, $f \in C^\infty(M, \mathbb{R}^d)$, (U, ϕ) is a chart of M, $K \subseteq U$ is compact and $\epsilon > 0$, form a subbasis for the weak topology on $C^\infty(M, \mathbb{R}^d)$. In particular, the weak topology on $C^\infty(M, \mathbb{R}^d)$ has a basis of convex sets.*

Proof. Consider a subbasic set $A := N^r(f, (U, \phi), (V, \psi), K, \epsilon)$ containing a function g. First observe that $g \in A' := N^r(g, (U, \phi), (V, \psi), K, \epsilon') \subseteq A$ for some ϵ', since these sets are determined by comparing finitely many norms on a compact set. One can then show that $A'' := N^r(g, (U, \phi), (V, \mathrm{id}), K, \epsilon'') \subseteq A'$ for some ϵ'', using bounds on the derivatives of ψ on $g(K)$. Finally, we claim that $A''' := N^r(g, (U, \phi), (\mathbb{R}^d, \mathrm{id}), K, \epsilon''') \subseteq A''$ for some ϵ'''. To see this, cover $g(K)$ by finitely many open balls $B_1, \dots, B_n$ such that $2B_\ell \subseteq V$ for each ℓ, and let ϵ''' be the minimum of the radii and ϵ''. Then if $h \in A'''$ and $x \in K$, we have $g(x) \in B_\ell$ for some ℓ and $|g(x) - h(x)| < \epsilon'''$, so $h(x) \in 2B_\ell \subseteq V$. □

For N open in $\mathbb{R}^d$, we will implicitly use that the inclusion map induces a continuous map $C^\infty(M, N) \subseteq C^\infty(M, \mathbb{R}^d)$ in the weak topologies, which follows from the fact that the weak topology is functorial in the second variable (see [Hirsch 1976, Exercise 10(a), p. 64]). (In fact, the weak topology and the subspace topology on $C^\infty(M, N)$ agree, but we won't need this.) Although $C^\infty(M, N)$ need not be an open subset of $C^\infty(M, \mathbb{R}^d)$, it has the following weaker property.

Lemma A.4. *Let M be a smooth manifold and let N be an open subset of $\mathbb{R}^d$. If f is in $C^\infty(M, N)$ and K is a compact subset of M, then there is a convex basic weak $C^\infty(M, \mathbb{R}^d)$-neighborhood of f whose elements map K into N.*

Proof. The set $\{g \in C^\infty(M, \mathbb{R}^d) \mid g(K) \subseteq N\}$ is open in the compact-open topology on $C^\infty(M, \mathbb{R}^d)$ and so is open in the weak topology by Lemma A.2. By Lemma A.3, the weak topology on $C^\infty(M, \mathbb{R}^d)$ has a basis of convex sets. Thus any $f : M \to N$ has such a convex basic set as a weak neighborhood. $\square$

Theorem A.5. *Let M and N be smooth manifolds. Suppose $p : \mathbb{R} \to C^\infty(M, N)$ is weakly continuous and t_r is a sequence of real numbers converging to zero. Then there is a subsequence t_{r_j} and a smooth curve $q : \mathbb{R} \to C^\infty(M, N)$ such that $q(2^{-j}) = p(t_{r_j})$ for each j and $q(0) = p(0)$.*

Proof. We first reduce to the case where N is open in $\mathbb{R}^d$. As in Theorem 4.13, choose an embedding $N \hookrightarrow \mathbb{R}^d$, and let U be an open tubular neighborhood of N in $\mathbb{R}^d$, so that the inclusion $i : N \to U$ has a smooth retract $r : U \to N$. By [Hirsch 1976, Exercise 10(a), p. 64], the map $\mathbb{R} \to C^\infty(M, U)$ sending t to $i \circ p(t)$ is weakly continuous, so if the theorem holds for $C^\infty(M, U)$, then there is a smooth curve $q : \mathbb{R} \to C^\infty(M, U)$ such that $q(2^{-j}) = i \circ p(t_{r_j})$ for each j and $q(0) = i \circ p(0)$. Then the map sending t to $r \circ q(t)$ is smooth, $r \circ q(2^{-j}) = p(t_{r_j})$ for each j, and $r \circ q(0) = p(0)$, so we are done. Thus we may assume that N is open in $\mathbb{R}^d$.

If t_r is eventually constant, we may take q to be a constant function, so suppose it is not. Choose charts $(U_k, \phi_k)_{k=1}^\infty$ for M and compact sets $K_k \subseteq U_k$ as described in Lemma A.1(a). Let $f = p(0)$. For $j = 1, 2, \ldots$, the sets,

$$A_j = \bigcap_{k=1}^{j} N^j\big(f, (U_k, \phi_k), (\mathbb{R}^d, \mathrm{id}), K_k, 2^{-(j+1)^2}\big)$$

are weak $C^\infty(M, \mathbb{R}^d)$-neighborhoods of f, so we may choose a strictly monotone subsequence t_{r_j} such that $p(t_{r_j}) \in A_j$ for each j. Set $f_j = p(t_{r_j})$. Now compose p with a continuous function taking 2^{-j} to t_{r_j} for each j to obtain a weakly continuous function p_0 that satisfies $p_0(2^{-j}) = f_j$ for $j = 1, 2, \ldots$ and $p_0(0) = f$.

Fix k. By Lemma A.4, for each $t \in [0, 1]$, there is a convex neighborhood of $p_0(t)$ whose elements map K_k into N. By compactness, there is a $\delta_k > 0$ such that any subinterval of $[0, 1]$ of length at most $2\delta_k$ is mapped by p_0 into one of these neighborhoods. Thus, for each t, any convex combination of elements in $p([t - \delta_k, t + \delta_k] \cap [0, 1])$ maps K_k into N. Let $\tau_0, \tau_1, \ldots$ be the strictly decreasing sequence obtained by ordering the set $\{1, 1/2, 1/4, \ldots\} \cup \{\delta_k, 2\delta_k, \ldots, \lfloor 1/\delta_k \rfloor \delta_k\}$. Note that $\tau_0 = 1$ and $\tau_{j-1} - \tau_j \le \delta_k$ for $j = 1, 2, \ldots$.

Fix a nondecreasing $\mu \in C^\infty(\mathbb{R}, [0, 1])$ such that $\mu = 0$ in a neighborhood of $(-\infty, 0]$ and $\mu = 1$ in a neighborhood of $[1, \infty)$. Let

$$\mathcal{M}_\ell = 1 + 2 \max_{\ell' \le \ell} \max_{t \in [0,1]} |\mu^{(\ell')}(t)|.$$

Define $q_k : \mathbb{R} \to C^\infty(M, \mathbb{R}^d)$ by $q_k(t) = p_0(0)$ for $t \le 0$, $q_k(t) = p_0(1)$ for $t \ge 1$,

and

$$q_k(t) = p_0(\tau_j) + \mu\left(\frac{t-\tau_j}{\tau_{j-1}-\tau_j}\right)(p_0(\tau_{j-1}) - p_0(\tau_j))$$

for $\tau_j \le t \le \tau_{j-1}$, $j = 1, 2, \ldots$. Note that for each $t \in (0, 1]$, $q_k(t)$ is a convex combination of elements of $p_0([t-\delta_k, t+\delta_k] \cap [0, 1])$. Clearly, q_k is constant on $(-\infty, 0]$ and constant on $[1, \infty)$. The choice of μ ensures that it is constant in a neighborhood of τ_{j-1} for each j and smooth on (τ_j, τ_{j-1}). Thus, q_k is smooth on $(\mathbb{R} \setminus \{0\}) \times M$. To see that it is also smooth at $t = 0$, fix a positive integer κ, set $F = f \circ \phi_\kappa^{-1}$, $F_j = f_j \circ \phi_\kappa^{-1}$ for each j, and $Q(t, s) = q_k(t)(\phi_\kappa^{-1}(s)) - F(s)$. It will suffice to show that all partial derivatives of Q exist and equal zero on $S := \{0\} \times \phi_\kappa(\mathring{K}_\kappa)$. Certainly $Q = 0$ there, and if $\mathcal{D}$ is any composition of partial differentiation operators such that $\mathcal{D}Q$ vanishes on S, then the partial derivative of $\mathcal{D}Q$ with respect to any of $s_1, \ldots, s_m$ also vanishes there. To complete the induction, it is enough to show that the partial derivative of $\mathcal{D}Q$ with respect to t also vanishes on S.

Where Q is C^∞, the order of mixed partials is unimportant, so $\mathcal{D}Q = D_t^\ell D_s^i Q$ off S for some $\ell \ge 0$ and some multi-index i. Choose J so that $2^{-J} < \delta_k$. Then $2^{-J}, 2^{-J-1}, 2^{-J-2}, \ldots$ is a tail of the sequence $\tau_0, \tau_1, \ldots$. So if $j > J$ and $2^{-j} \le t \le 2^{1-j}$, then

$$q_k(t) = f_j + \mu(2^j t - 1)(f_{j-1} - f_j),$$

and, for $s \in \phi_\kappa(U_\kappa)$,

$$(D_t^\ell D_s^i Q)(t, s) = \begin{cases} (D_s^i(F_j - F))(s) + \mu(2^j t - 1)(D_s^i(F_{j-1} - F_j))(s) & \text{if } \ell = 0, \\ \mu^{(\ell)}(2^j t - 1)\, 2^{\ell j} (D_s^i(F_{j-1} - F_j))(s) & \text{if } \ell \ge 1. \end{cases}$$

If $j > \max(J, \kappa, |i|, \ell + 2)$, then

$$f_j \in A_j \subseteq N^j\big(f, (U_\kappa, \phi_\kappa), (\mathbb{R}^d, \mathrm{id}), K_\kappa, 2^{-(j+1)^2}\big),$$

and

$$f_{j-1} \in A_{j-1} \subseteq N^{j-1}\big(f, (U_\kappa, \phi_\kappa), (\mathbb{R}^d, \mathrm{id}), K_\kappa, 2^{-j^2}\big),$$

so

$$|D_s^i(F_j - F)| \le 2^{-(j+1)^2} \le 2^{-j^2} \quad \text{and} \quad |D_s^i(F_{j-1} - F)| \le 2^{-j^2}$$

on $\phi_\kappa(K_\kappa)$. Thus, for any $s \in \phi_k(\mathring{K}_\kappa)$,

$$|(\mathcal{D}Q)(t, s) - (\mathcal{D}Q)(0, s)| = |(D_t^\ell D_s^i Q)(t, s)| \le \mathcal{M}_\ell 2^{\ell j} 2^{-j^2} \le \mathcal{M}_\ell t^2,$$

where we have used that $\ell < j - 2$ in the last inequality. Since j can be arbitrarily large, this inequality holds for all sufficiently small t, so the partial derivative of $\mathcal{D}Q$ with respect to t (from the right) exists and equals zero. The partial derivative from the left is trivially zero. This completes the induction and the proof that q_k is smooth.

Before allowing k to vary, observe that $q_k(\tau_j) = p_0(\tau_j)$ for each j, and in particular, $q_k(2^{-j}) = p_0(2^{-j}) = f_j$ for each j.

Let $(\nu_k)_{k=1}^{\infty}$ be a smooth partition of unity on M with ν_k supported in $\mathring{K}_k$ and define q by $q(t)(x) = \sum_{k=1}^{\infty} \nu_k(x) q_k(t)(x)$. Then $q : \mathbb{R} \to C^\infty(M, \mathbb{R}^d)$ is a smooth curve such that $q(2^{-j}) = f_j = p(t_{r_j})$ for each j, and of course $q(0) = p(0)$. It remains to show that $q(t)$ takes values in N for each $t \in \mathbb{R}$. Let $x \in M$. There are finitely many k such that $\nu_k(x) \neq 0$; among them, choose k' so that $\delta_{k'}$ is as large as possible. Then, for any t and any k such that $\nu_k(x) \neq 0$, $q_k(t)$ is a convex combination of elements of $p_0([t - \delta_{k'}, t + \delta_{k'}] \cap [0, 1])$. Thus, $\sum_{k=1}^{\infty} \nu_k(x) q_k(t)$ is also a convex combination of elements of $p_0([t - \delta_{k'}, t + \delta_{k'}] \cap [0, 1])$, and therefore maps $K_{k'}$ to N. But $\nu_{k'}(x) \neq 0$, so $x \in K_{k'}$. Hence, $\sum_{k=1}^{\infty} \nu_k(x) q_k(t)(x) \in N$, that is, $q(t)(x) \in N$. We conclude that $q : \mathbb{R} \to C^\infty(M, N)$. This completes the proof. $\square$

Acknowledgements

We would like to thank Andrew Stacey and Chris Schommer-Pries for very helpful conversations, and Jeremy Brazas for the idea behind Example 3.14.

References

[Baez and Hoffnung 2011] J. C. Baez and A. E. Hoffnung, "Convenient categories of smooth spaces", *Trans. Amer. Math. Soc.* **363**:11 (2011), 5789–5825. MR 2012h:18016 Zbl 1237.58006

[Christensen and Wu 2014] J. D. Christensen and E. Wu, "The homotopy theory of diffeological spaces", preprint, 2014. arXiv 1311.6394

[Donato 1984] P. Donato, *Revêtement et groupe fondamental des espaces différentiels homogènes*, thesis, Université de Provence, Marseille, 1984.

[Dugger 2003] D. Dugger, "Notes on delta-generated spaces", preprint, 2003, available at http://www.uoregon.edu/~ddugger/delta.html.

[Fajstrup and Rosický 2008] L. Fajstrup and J. Rosický, "A convenient category for directed homotopy", *Theory Appl. Categ.* **21**:1 (2008), 7–20. MR 2009g:18007 Zbl 1157.18003 arXiv 0708.3937

[Gowers 2010] T. Gowers, "Answer to 'Why are the integers with the cofinite topology not path-connected?"', MathOverflow, 2010, available at http://mathoverflow.net/questions/48970/#48977.

[Hector 1995] G. Hector, "Géométrie et topologie des espaces difféologiques", pp. 55–80 in *Analysis and geometry in foliated manifolds* (Santiago de Compostela, 1994), edited by X. Masa et al., World Scientific, River Edge, NJ, 1995. MR 98f:58008 Zbl 0993.58500

[Hirsch 1976] M. W. Hirsch, *Differential topology*, Graduate Texts in Mathematics **33**, Springer, New York, 1976. MR 56 #6669 Zbl 0356.57001

[Iglesias-Zemmour 1985] P. J. Iglesias-Zemmour, *Fibrations difféologiques et homotopie*, thesis, Université de Provence, 1985.

[Iglesias-Zemmour 2013] P. J. Iglesias-Zemmour, *Diffeology*, Mathematical Surveys and Monographs **185**, American Mathematical Society, Providence, RI, 2013. MR 3025051 Zbl 1269.53003

[Kriegl and Michor 1997] A. Kriegl and P. W. Michor, *The convenient setting of global analysis*, Mathematical Surveys and Monographs **53**, American Mathematical Society, Providence, RI, 1997. MR 98i:58015 Zbl 0889.58001

[Laubinger 2006] M. Laubinger, "Diffeological spaces", *Proyecciones* (*Antofagasta*) **25**:2 (2006), 151–178. MR 2007f:57059 Zbl 1169.57308

[Michor 1980] P. W. Michor, *Manifolds of differentiable mappings*, Shiva Mathematics Series **3**, Shiva, Nantwich, 1980. MR 83g:58009 Zbl 0433.58001

[van Mill 2001] J. van Mill, *The infinite-dimensional topology of function spaces*, North-Holland Mathematical Library **64**, North-Holland, Amsterdam, 2001. MR 2002h:57031 Zbl 0969.54003

[Shimakawa et al. 2010] K. Shimakawa, K. Yoshida, and T. Haraguchi, "Homology and cohomology via enriched bifunctors", preprint, 2010. arXiv 1010.3336

[Souriau 1980] J.-M. Souriau, "Groupes différentiels", pp. 91–128 in *Differential geometrical methods in mathematical physics* (Aix-en-Provence/Salamanca, 1979), edited by P. L. Garcia et al., Lecture Notes in Math. **836**, Springer, Berlin, 1980. MR 84b:22038 Zbl 0501.58010

[Souriau 1984] J.-M. Souriau, "Groupes différentiels de physique mathématique", pp. 73–119 in *South Rhone seminar on geometry, II* (Lyon, 1983), edited by P. Dazord and N. Desolneux-Moulis, Hermann, Paris, 1984. MR 86a:58001 Zbl 0541.58002

[Souriau 1985] J.-M. Souriau, "Un algorithme générateur de structures quantiques", pp. 341–399 in *Élie Cartan et les mathématiques d'aujourd'hui* (Lyon, 1984), Astérisque, Numéro hors série, Société Mathématique de France, Paris, 1985. MR 87g:58045 Zbl 0608.58028

[Stacey 2011] A. Stacey, "Comparative smootheology", *Theory Appl. Categ.* **25**:4 (2011), 64–117. MR 2012g:18012 Zbl 1220.18013

Received August 9, 2013. Revised March 10, 2014.

J. DANIEL CHRISTENSEN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WESTERN ONTARIO
LONDON, ON N6A 5B7
CANADA
jdc@uwo.ca

GORDON SINNAMON
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WESTERN ONTARIO
LONDON, ON N6A 5B7
CANADA
sinnamon@uwo.ca

ENXIN WU
FACULTY OF MATHEMATICS
UNIVERSITY OF VIENNA
OSKAR-MORGENSTERN-PLATZ 1
1090 VIENNA
AUSTRIA
enxin.wu@univie.ac.at

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.272.111

ON THE ATKIN POLYNOMIALS

AHMAD EL-GUINDY AND MOURAD E. H. ISMAIL

We identify the Atkin polynomials in terms of associated Jacobi polynomials. Our identification then takes advantage of the theory of orthogonal polynomials and their asymptotics to establish many new properties of the Atkin polynomials. This shows that corecursive polynomials may lead to interesting sets of orthogonal polynomials.

1. Introduction

In unpublished work, Oliver Atkin introduced a family of orthogonal polynomials with fascinating number-theoretic properties: They are the unique family of monic orthogonal polynomials corresponding to a unique scalar product on the space of polynomials in the modular j-invariant for which all Hecke operators are self-adjoint. Furthermore, their reductions modulo a prime $p \geq 5$ are also very significant in the theory of elliptic curves, as they match the supersingular polynomial at p whenever the degrees agree. For all the number-theoretic definitions, as well as beautiful proofs of these and other facts about the Atkin polynomials, we refer the reader to the excellent [Kaneko and Zagier 1998], where Atkin's results were popularized, simplified, and expanded upon.

The Atkin polynomials are generated by the recurrence relation

$$A_{n+1}(x) = \left[x - 24\frac{144n^2-29}{(2n+1)(2n-1)}\right]A_n(x) - 36\frac{(12n-13)(12n-7)(12n-5)(12n+1)}{n(n-1)(2n-1)^2}A_{n-1}(x), \quad n > 1, \tag{1-1}$$

through the initial conditions

$$A_0(x) = 1, \quad A_1(x) = x - 720, \quad A_2(x) = x^2 - 1640x + 269280. \tag{1-2}$$

Research supported by the DSFP at King Saud University through grant DSFP/MATH 01 and by the NPST program of King Saud University, project number 10-MAT1293-02.
MSC2010: primary 33C47, 11F03; secondary 33C05.
Keywords: associated Jacobi polynomials, asymptotics, weight functions, scaled invariant, modular function, birth and death process polynomials, supersingular polynomials, Hecke operators.

The polynomials $\{A_n(x)\}$ are orthogonal with respect to an absolutely continuous measure supported on [0, 1728] (see Section 7).

In this paper we show that the Atkin polynomials are related to the associated Jacobi polynomials of Wimp [1987] and of Ismail and Masson [1991]. This identification leads to many new properties of the polynomials $\{A_n(x)\}$.

It is worth pointing out that the way the Atkin polynomials are defined, that is, defining $P_0(x)$, $P_1(x)$ and $P_2(x)$, then using a recurrence relation to generate the rest, is not unusual in the literature on orthogonal polynomials. The idea is to start with two monic polynomials, $P_k(x)$ and $P_{k+1}(x)$, of degrees k and $k+1$, respectively, with real, simple and interlacing zeros. Then use the division algorithm to generate the monic polynomials $P_n(x)$, $0 \le n < k$; we are guaranteed to have a sequence of monic orthogonal polynomials $\{P_j(x) : 0 \le j \le k+1\}$. Now use any three-term recurrence relation of the form

$$P_{n+1}(x) = (x-\alpha_n)P_n(x) + \beta_n P_{n-1}(x), \tag{1-3}$$

where $\alpha_n \in \mathbb{R}$ and $\beta_n > 0$ for $n > k$, to generate the polynomials $\{P_n(x)\}$ for $n > k+1$. The construction above is referred to as "Wendroff's Theorem" in the orthogonal polynomial literature. The interested reader may consult [Ismail 2009] or [Chihara 1978] for the precise statement and the detailed proof of Wendroff's theorem. This is also related to the concept of corecursive polynomials [Chihara 1978].

In Section 2, we recall some preliminary facts about associated Jacobi polynomials and orthogonal polynomials in general. In Section 3, we obtain a representation of (a scaled version of) the Atkin polynomials as a linear combination of the associated Jacobi polynomials of Wimp [1987] and of Ismail and Masson [1991]. Building on that, we provide an explicit representation of the coefficients of the Atkin polynomials in Section 4, a representation in terms of certain hypergeometric functions and an asymptotic expansion in Section 5, and a generating function identity in Section 6. Lastly, in Section 7 we give an explicit description of the weight function in terms of certain ${}_2F_1$ functions.

We shall follow the standard notation for hypergeometric functions and orthogonal polynomials as in [Andrews et al. 1999; Ismail 2009; Luke 1969; Rainville 1960; Szegő 1975]. In particular we use $F\left(\begin{smallmatrix} a,b \\ c \end{smallmatrix} \,\middle|\, z\right)$ to mean to mean ${}_2F_1\left(\begin{smallmatrix} a,b \\ c \end{smallmatrix} \,\middle|\, z\right)$.

2. Preliminaries

Let $\{\lambda_n\}$ and $\{\mu_n\}$ be the birth and death rates of a birth and death process; that is, $\lambda_n > 0$ and $\mu_{n+1} > 0$ for all $n \ge 0$, with $\mu_0 \ge 0$. Such a process generates a sequence of orthogonal polynomials through a three-term recurrence relation

$$-x\,Q_n(x) = \lambda_n Q_{n+1}(x) - (\lambda_n+\mu_n)Q_n(x) + \mu_n Q_{n-1}(x), \quad n > 0, \tag{2-1}$$

with the initial conditions

$$Q_0(x)=1, \qquad Q_1(x)=(\lambda_0+\mu_0-x)/\lambda_0. \tag{2-2}$$

The corresponding monic polynomials satisfy

$$x\,\widetilde{Q}_n(x)=\widetilde{Q}_{n+1}+(\lambda_n+\mu_n)\widetilde{Q}_n(x)-\lambda_{n-1}\mu_n\widetilde{Q}_{n-1}(x), \tag{2-3}$$

with $\widetilde{Q}_0(x)=1$, $\widetilde{Q}_1(x)=x-\lambda_0-\mu_0$. When $\mu_0\neq 0$ there is a second natural birth and death process with birth rates $\{\lambda_n\}$ and death rates $\{\tilde{\mu}_n\}$ with $\tilde{\mu}_n=\mu_n$ for $n>0$ but $\tilde{\mu}_0=0$ [Ismail et al. 1988]. The latter birth and death generate a second family of orthogonal polynomials satisfying (2-1) but with initial conditions $Q_0(x)=1$, $Q_1(x)=(\lambda_0-x)/\lambda_0$. This observation is due to Ismail, Letessier and Valent [Ismail et al. 1988].

The associated polynomials of $\{Q_n(x)\}$ correspond to the birth and death rates $\{\lambda_{n+c}\}$ and death rates $\{\mu_{n+c}\}$, when such rates are well defined. Since we consider $c\geq 0$, usually $\mu_c>0$. Thus we usually have two families of associated polynomials. One is defined when μ_c is defined from the pattern of μ_n. When $\mu_c\neq 0$, a second family arises if μ_{n+c}, when $n=0$ is interpreted as zero.

Recall that the Jacobi polynomials $\{P_n^{(\alpha,\beta)}(x)\}$ can be defined by the three-term recurrence relation

$$\begin{aligned} 2(n+1)(n+\alpha+\beta+1)(\alpha+\beta+2n)P_{n+1}^{(\alpha,\beta)}(x) \\ =(\alpha+\beta+2n+1)\big[(\alpha^2-\beta^2)+x(\alpha+\beta+2n+2)(\alpha+\beta+2n)\big]P_n^{(\alpha,\beta)}(x) \\ -2(\alpha+n)(\beta+n)(\alpha+\beta+2n+2)P_{n-1}^{(\alpha,\beta)}(x), \end{aligned} \tag{2-4}$$

for $n\geq 0$, with $P_{-1}^{(\alpha,\beta)}(x)=0$, $P_0^{(\alpha,\beta)}(x)=1$. We now set

$$\begin{aligned} V_n^{(\alpha,\beta)}(x) &= \frac{n!\,(\alpha+\beta+1)_n}{(\alpha+\beta+1)_{2n}}\,P_n^{(\alpha,\beta)}(2x-1) \\ &= \frac{n!}{(n+\alpha+\beta+1)_n}\,P_n^{(\alpha,\beta)}(2x-1). \end{aligned} \tag{2-5}$$

One can easily verify that the polynomials $\{V_n^{(\alpha,\beta)}(x)\}$ are monic birth and death process polynomials $\widetilde{Q}_n$, with rates

$$\begin{aligned} \lambda_n &= \frac{(n+\beta+1)(n+\alpha+\beta+1)}{(2n+\alpha+\beta+1)(2n+\alpha+\beta+2)}, \\ \mu_n &= \frac{n(n+\alpha)}{(2n+\alpha+\beta)(2n+\alpha+\beta+1)}. \end{aligned} \tag{2-6}$$

Wimp [1987] considered the recurrence relation obtained by formally replacing n by $n+c$ in (2-4), and he showed that the new relation has two linearly independent solutions $P_n^{(\alpha,\beta)}(x;c)$ and $P_{n-1}^{(\alpha,\beta)}(x;c+1)$. Ismail and Masson [1991] identified the birth and death rates corresponding to that three-term recurrence relation and provided two linearly independent solutions $P_n^{(\alpha,\beta)}(x;c)$ and $\mathcal{P}_n^{(\alpha,\beta)}(x;c)$. They then used the notation

$$R_n^{(\alpha,\beta)}(x;c) = P_n^{(\alpha,\beta)}(2x-1;c), \quad \mathcal{R}_n^{(\alpha,\beta)}(x;c) = \mathcal{P}_n^{(\alpha,\beta)}(2x-1;c). \tag{2-7}$$

We shall use the notation

$$\begin{aligned} V_n^{(\alpha,\beta)}(x;c) &= \frac{(c+1)_n(\alpha+\beta+c+1)_n}{(\alpha+\beta+2c+1)_{2n}} R_n^{(\alpha.\beta)}(x;c), \\ \mathcal{V}_n^{(\alpha,\beta)}(x;c) &= \frac{(c+1)_n(\alpha+\beta+c+1)_n}{(\alpha+\beta+2c+1)_{2n}} \mathcal{R}_n^{(\alpha.\beta)}(x;c). \end{aligned} \tag{2-8}$$

To lighten our notation, we shall occasionally omit the parameters when the context is clear. We consider the birth and death rates

$$\begin{aligned} \lambda_n &= \frac{(n+c+\beta+1)(n+c+\alpha+\beta+1)}{(2n+2c+\alpha+\beta+1)(2n+2c+\alpha+\beta+2)}, \quad n \geq 0, \\ \mu_n &= \frac{(n+c)(n+c+\alpha)}{(2n+2c+\alpha+\beta)(2n+2c+\alpha+\beta+1)}, \quad n > 0, \end{aligned} \tag{2-9}$$

with

$$\mu_0 := \begin{cases} \dfrac{c(c+\alpha)}{(2c+\alpha+\beta)(2c+\alpha+\beta+1)} & \text{for } V, \\ 0 & \text{for } \mathcal{V}. \end{cases} \tag{2-10}$$

3. The Atkin polynomials

In order to compare the Atkin polynomials with other results in the literature we need to renormalize them. Let

$$A_n(1728y) = (1728)^n \mathcal{A}_n(y). \tag{3-1}$$

The polynomials $\mathcal{A}_n$ are now generated by the recurrence

$$\begin{aligned} &\mathcal{A}_{n+1}(x) \\ &= \left[x - \frac{2\left(n^2-\frac{29}{144}\right)}{4n^2-1}\right]\mathcal{A}_n(x) - \frac{\left(n-\frac{13}{12}\right)\left(n-\frac{7}{12}\right)\left(n-\frac{5}{12}\right)\left(n+\frac{1}{12}\right)}{2n(2n-1)^2(2n-2)}\mathcal{A}_{n-1}(x) \end{aligned} \tag{3-2}$$

for $n > 1$. The initial conditions are

$$\mathcal{A}_0(x) = 1, \quad \mathcal{A}_1(x) = x - \tfrac{5}{12}, \quad \mathcal{A}_2(x) = x^2 - \tfrac{205}{216}x + \tfrac{935}{10368}. \tag{3-3}$$

Kaneko and Zagier [1998] wrote the recurrence relation (1-1) in the monic form (2-3). Indeed their (19), when written in terms of the $\mathcal{A}_n$, corresponds to (2-3) with

$$\lambda_n = \frac{\left(n-\frac{1}{12}\right)\left(n+\frac{5}{12}\right)}{2n(2n+1)}, \quad \mu_n = \frac{\left(n-\frac{5}{12}\right)\left(n+\frac{1}{12}\right)}{2n(2n-1)}. \tag{3-4}$$

From (2-8), we see that $V_n^{(\alpha,\beta)}(x;c)$ and $\mathcal{V}_n^{(\alpha,\beta)}(x;c)$ satisfy the second-order difference equation

$$T_{n+1}(x)=\left(x+\frac{\alpha^2-\beta^2-(2n+2c+\alpha+\beta)(2n+2c+\alpha+\beta+2)}{2(2n+2c+\alpha+\beta)(2n+2c+\alpha+\beta+2)}\right)T_n(x) - \frac{(n+c)(n+c+\alpha)(n+c+\beta)(n+c+\alpha+\beta)}{(2n+2c+\alpha+\beta-1)(2n+2c+\alpha+\beta)^2(2n+2c+\alpha+\beta+1)}T_{n-1}(x) \tag{3-5}$$

for $n\geq 1$, with the initial conditions $V_0=\mathcal{V}_0=1$ and

$$\begin{aligned} V_1^{(\alpha,\beta)}(x;c)&=x-(\lambda_0+\mu_0),\\ \mathcal{V}_1^{(\alpha,\beta)}(x;c)&=x-\lambda_0, \end{aligned} \tag{3-6}$$

where λ_n and μ_n are defined as in (2-9)–(2-10). On the other hand, we see that the sequence $\{\mathcal{A}_{n+1}(x)\}_{n=-1}^{\infty}$ is a solution of the second-order difference equation

$$T_{n+1}(x)=\left(x-\frac{7+36(2n+1)(2n+3)}{72(2n+1)(2n+3)}\right)T_n(x) - \frac{\left(n-\frac{1}{12}\right)\left(n+\frac{5}{12}\right)\left(n+\frac{7}{12}\right)\left(n+\frac{13}{12}\right)}{(2n)(2n+1)^2(2n+2)}T_{n-1}(x) \tag{3-7}$$

for $n\geq 1$. It is not hard to check that (3-7) is identical to (3-5) in exactly four cases, namely,

$$(\alpha,\beta,c)\in S:=\left\{\left(-\tfrac{1}{2},-\tfrac{2}{3},\tfrac{13}{12}\right),\left(\tfrac{1}{2},-\tfrac{2}{3},\tfrac{7}{12}\right),\left(-\tfrac{1}{2},\tfrac{2}{3},\tfrac{5}{12}\right),\left(\tfrac{1}{2},\tfrac{2}{3},-\tfrac{1}{12}\right)\right\}.$$

Theorem 3.1. *For $n\geq 0$ and $(\alpha,\beta,c)\in S$, we have the following representations for $\mathcal{A}_{n+1}(x)$:*

$$\mathcal{A}_{n+1}(x)=\left(x-\tfrac{5}{12}\right)V_n^{(\alpha,\beta)}(x;c)-\tfrac{91}{384}V_{n-1}^{(\alpha,\beta)}(x;c+1), \tag{3-8}$$

$$\mathcal{A}_{n+1}(x)=(x-8)V_n^{(-1/2,2/3)}\left(x;\tfrac{5}{12}\right)+\tfrac{91}{12}\mathcal{V}_n^{(-1/2,2/3)}\left(x;\tfrac{5}{12}\right), \tag{3-9}$$

$$\mathcal{A}_{n+1}(x)=x\,V_n^{(1/2,-2/3)}\left(x;\tfrac{7}{12}\right)-\tfrac{5}{12}\mathcal{V}_n^{(1/2,-2/3)}\left(x;\tfrac{7}{12}\right). \tag{3-10}$$

Proof. It is straightforward to check that for any $(\alpha,\beta,c)\in S$,

$$\{V_n^{(\alpha,\beta)}(x;c),\mathcal{V}_n^{(\alpha,\beta)}(x;c)\}$$

is a basis of solutions of (3-7), and the same is true for

$$\{V_n^{(\alpha,\beta)}(x;c),V_{n-1}^{(\alpha,\beta)}(x;c+1)\}.$$

The results follow by simple linear algebra on the equations corresponding to $n=0$ and $n=1$. □

We note that $V_n^{(\alpha,\beta)}(x;c)$ is the same for the four triples in S, whereas we have two possibilities for $\mathcal{V}_n^{(\alpha,\beta)}(x;c)$, depending on whether $\beta=\frac{2}{3}$ or $\beta=-\frac{2}{3}$. For convenience we explicitly write down the first few of these polynomials:

$$\begin{aligned}&V_0^{(\alpha,\beta)}(x;c)=1,\\ &V_1^{(\alpha,\beta)}(x;c)=x-\tfrac{115}{216},\\ &V_2^{(\alpha,\beta)}(x;c)=x^2-\tfrac{187}{180}x+\tfrac{11621}{55296};\end{aligned}\tag{3-11}$$

$$\begin{aligned}&V_{-1}^{(\alpha,\beta)}(x;c+1)=0,\\ &V_0^{(\alpha,\beta)}(x;c+1)=1,\\ &V_1^{(\alpha,\beta)}(x;c+1)=x-\tfrac{547}{1080};\end{aligned}\tag{3-12}$$

$$\begin{aligned}&\mathcal{V}_0^{(1/2,-2/3)}\big(x;\tfrac{7}{12}\big)=\mathcal{V}_0^{(-1/2,-2/3)}\big(x;\tfrac{13}{12}\big)=1,\\ &\mathcal{V}_1^{(1/2,-2/3)}\big(x;\tfrac{7}{12}\big)=\mathcal{V}_1^{(-1/2,-2/3)}\big(x;\tfrac{13}{12}\big)=x-\tfrac{187}{864},\\ &\mathcal{V}_2^{(1/2,-2/3)}\big(x;\tfrac{7}{12}\big)=\mathcal{V}_2^{(-1/2,-2/3)}\big(x;\tfrac{13}{12}\big)=x^2-\tfrac{347}{480}x+\tfrac{124729}{2488320};\end{aligned}\tag{3-13}$$

$$\begin{aligned}&\mathcal{V}_0^{(1/2,2/3)}\big(x;\tfrac{-1}{12}\big)=\mathcal{V}_0^{(-1/2,2/3)}\big(x;\tfrac{5}{12}\big)=1,\\ &\mathcal{V}_1^{(1/2,2/3)}\big(x;\tfrac{-1}{12}\big)=\mathcal{V}_1^{(-1/2,2/3)}\big(x;\tfrac{5}{12}\big)=x-\tfrac{475}{864},\\ &\mathcal{V}_2^{(1/2,2/3)}\big(x;\tfrac{-1}{12}\big)=\mathcal{V}_2^{(-1/2,2/3)}\big(x;\tfrac{5}{12}\big)=x^2-\tfrac{169}{160}x+\tfrac{108965}{497664}.\end{aligned}\tag{3-14}$$

One can check the first few cases of Theorem 3.1 using the equalities

$$\begin{aligned}&\mathcal{A}_1(x)=x-\tfrac{5}{12},\\ &\mathcal{A}_2(x)=x^2-\tfrac{205}{216}x+\tfrac{935}{10368},\\ &\mathcal{A}_3(x)=x^3-\tfrac{131}{90}x^2+\tfrac{28277}{55296}x-\tfrac{124729}{5971968}.\end{aligned}\tag{3-15}$$

4. Explicit representations

Wimp [1987, p. 987] gave an explicit formula for $R_n^{(\alpha,\beta)}(x;c)$. When translated in terms of the V_n polynomials it becomes

$$\begin{aligned}V_n^{(\alpha,\beta)}(x;c)&=(-1)^n\frac{(c+1)_n(\beta+c+1)_n}{(\alpha+\beta+2c+n+1)_n\,n!}\\ &\quad\times\sum_{k=0}^{n}\frac{(-n)_k(n+2c+\alpha+\beta+1)_k}{(c+1)_k(c+\beta+1)_k}x^k\\ &\quad\times{}_4F_3\left(\begin{matrix}k-n,\,n+k+\alpha+\beta+2c+1,\,c+\beta,\,c\\ k+\beta+c+1,\,k+c+1,\,\alpha+\beta+2c\end{matrix}\,\middle|\,1\right).\end{aligned}\tag{4-1}$$

On the other hand, Ismail and Masson [1991, Theorem 3.3] gave a similar formula for $\mathcal{R}_n^{(\alpha,\beta)}(x;c)$, which leads to

$$\begin{aligned}\text{(4-2)}\quad & \mathcal{V}_n^{(\alpha,\beta)}(x;c)\\ &= (-1)^n \frac{(c+1)_n(\beta+c+1)_n}{(\alpha+\beta+2c+n+1)_n\, n!}\\ &\quad\times \sum_{k=0}^{n} \frac{(-n)_k(n+2c+\alpha+\beta+1)_k}{(c+1)_k(c+\beta+1)_k} x^k\\ &\quad\quad \times {}_4F_3\left(\begin{matrix} k-n,\, n+k+\alpha+\beta+2c+1,\, c+\beta+1,\, c \\ k+\beta+c+1,\, k+c+1,\, \alpha+\beta+2c+1\end{matrix}\,\middle|\, 1\right).\end{aligned}$$

The following theorem establishes an analogous representation of $\mathcal{A}_{n+1}(x)$:

Theorem 4.1. *For $n \geq 0$, we have*

$$\begin{aligned}\text{(4-3)}\quad & \mathcal{A}_{n+1}(x) =\\ & \frac{\left(\frac{19}{12}\right)_n\left(\frac{11}{12}\right)_n}{(n+2)_n(-n)_n}\\ &\quad \times \Bigg[{}_3F_2\left(\begin{matrix} -n,\, n+2,\, \frac{7}{12} \\ \frac{19}{12},\, 2\end{matrix}\,\middle|\, 1\right)\\ &\quad\quad + \sum_{k=0}^{n} \frac{(-n)_k(n+2)_k}{\left(\frac{19}{12}\right)_k\left(\frac{11}{12}\right)_k} x^{k+1} \Bigg\{ \frac{6}{5}\, {}_4F_3\left(\begin{matrix} k-n,\, n+k+2,\, \frac{11}{12},\, \frac{-5}{12} \\ k+\frac{11}{12},\, k+\frac{19}{12},\, 1\end{matrix}\,\middle|\, 1\right)\\ &\quad\quad\quad - \frac{1}{5}\, {}_4F_3\left(\begin{matrix} k-n,\, n+k+2,\, \frac{-1}{12},\, \frac{-5}{12} \\ k+\frac{11}{12},\, k+\frac{19}{12},\, 1\end{matrix}\,\middle|\, 1\right)\Bigg\}\Bigg].\end{aligned}$$

Proof. From (3-10) we have

$$\mathcal{A}_{n+1}(x) = x V_n^{(1/2,-2/3)}\left(x;\, \tfrac{7}{12}\right) - \tfrac{5}{12}\mathcal{V}_n^{(1/2,-2/3)}\left(x;\, \tfrac{7}{12}\right), \quad n \geq 0;$$

we see that the coefficient of x^{k+1} in $\mathcal{A}_{n+1}(x)$ is given by

$$\begin{aligned}\text{(4-4)}\quad & (-1)^n \frac{\left(\frac{19}{12}\right)_n\left(\frac{11}{12}\right)_n}{(n+2)_n\, n!} \frac{(-n)_k(n+2)_k}{\left(\frac{19}{12}\right)_k\left(\frac{11}{12}\right)_k}\\ &\quad \times \Bigg[{}_4F_3\left(\begin{matrix} k-n,\, n+k+2,\, \frac{-1}{12},\, \frac{7}{12} \\ k+\frac{11}{12},\, k+\frac{19}{12},\, 1\end{matrix}\,\middle|\, 1\right)\\ &\quad\quad - \frac{5}{12} \frac{(k-n)(n+k+2)}{\left(k+\frac{19}{12}\right)\left(k+\frac{11}{12}\right)}\, {}_4F_3\left(\begin{matrix} k+1-n,\, n+k+3,\, \frac{11}{12},\, \frac{7}{12} \\ k+1+\frac{11}{12},\, k+1+\frac{19}{12},\, 2\end{matrix}\,\middle|\, 1\right)\Bigg].\end{aligned}$$

The coefficient of y^m in

$$
{}_4F_3\left(\begin{matrix} k-n,\, n+k+2,\, \frac{-1}{12},\, \frac{7}{12} \\ k+\frac{11}{12},\, k+\frac{19}{12},\, 1 \end{matrix}\,\middle|\, y\right)
- \frac{5y}{12}\frac{(k-n)(n+k+2)}{\left(k+\frac{19}{12}\right)\left(k+\frac{11}{12}\right)}\,{}_4F_3\left(\begin{matrix} k+1-n,\, n+k+3,\, \frac{11}{12},\, \frac{7}{12} \\ k+1+\frac{11}{12},\, k+1+\frac{19}{12},\, 2 \end{matrix}\,\middle|\, y\right)
$$

is

$$
\frac{(k-n)_m(n+k+2)_m\left(\frac{-1}{12}\right)_m\left(\frac{7}{12}\right)_m}{\left(k+\frac{11}{12}\right)_m\left(k+\frac{19}{12}\right)_m(m!)^2}
- \frac{5m}{12}\frac{(k-n)(n+k+2)}{\left(k+\frac{19}{12}\right)\left(k+\frac{11}{12}\right)}\frac{(k-n+1)_{m-1}(n+k+3)_{m-1}\left(\frac{11}{12}\right)_{m-1}\left(\frac{7}{12}\right)_{m-1}}{\left(k+1+\frac{11}{12}\right)_{m-1}\left(k+1+\frac{19}{12}\right)_{m-1}(2)_{m-1}(m)!}.
$$

Using the identity $(z)_m = z(z+1)_{m-1}$, we get that this coefficient is

$$
\begin{aligned}
&\frac{(k-n)_m(n+k+2)_m\left(\frac{-1}{12}\right)_m\left(\frac{7}{12}\right)_m}{\left(k+\frac{11}{12}\right)_m\left(k+\frac{19}{12}\right)_m(m!)^2}\left(1+\frac{\left(\frac{-5}{12}\right)(-12m)}{\left(\frac{7}{12}+m-1\right)}\right) \\
&\qquad = \frac{(k-n)_m(n+k+2)_m\left(\frac{-1}{12}\right)_m\left(\frac{-5}{12}\right)_m}{\left(k+\frac{11}{12}\right)_m\left(k+\frac{19}{12}\right)_m(m!)^2}\left(1-\tfrac{72}{5}m\right) \\
&\qquad = \frac{1}{5}\frac{(k-n)_m(n+k+2)_m\left(\frac{-5}{12}\right)_m}{\left(k+\frac{11}{12}\right)_m\left(k+\frac{19}{12}\right)_m(m!)^2}\left(6\left(\tfrac{11}{12}\right)_m-\left(\tfrac{-1}{12}\right)_m\right).
\end{aligned}
\tag{4-5}
$$

In the last equality, we used

$$
\left(\tfrac{-1}{12}\right)_m\left(m-\tfrac{5}{72}\right) = \left(\tfrac{-1}{12}\right)_m\left[\left(m-\tfrac{1}{12}\right)+\tfrac{1}{72}\right] = -\tfrac{1}{12}\left(\tfrac{11}{12}\right)_m + \tfrac{1}{72}\left(\tfrac{-1}{12}\right)_m.
$$

It now follows that

$$
\begin{aligned}
&\left[{}_4F_3\left(\begin{matrix} k-n,\, n+k+2,\, \frac{-1}{12},\, \frac{7}{12} \\ k+\frac{11}{12},\, k+\frac{19}{12},\, 1 \end{matrix}\,\middle|\, y\right)\right. \\
&\qquad \left. - \frac{5y}{12}\frac{(k-n)(n+k+2)}{\left(k+\frac{19}{12}\right)\left(k+\frac{11}{12}\right)}\,{}_4F_3\left(\begin{matrix} k+1-n,\, n+k+3,\, \frac{11}{12},\, \frac{7}{12} \\ k+1+\frac{11}{12},\, k+1+\frac{19}{12},\, 2 \end{matrix}\,\middle|\, y\right)\right] \\
&\qquad = \tfrac{6}{5}\,{}_4F_3\left(\begin{matrix} k-n,\, n+k+2,\, \frac{11}{12},\, \frac{-5}{12} \\ k+\frac{11}{12},\, k+\frac{19}{12},\, 1 \end{matrix}\,\middle|\, y\right) \\
&\qquad\qquad - \tfrac{1}{5}\,{}_4F_3\left(\begin{matrix} k-n,\, n+k+2,\, \frac{-1}{12},\, \frac{-5}{12} \\ k+\frac{11}{12},\, k+\frac{19}{12},\, 1 \end{matrix}\,\middle|\, y\right).
\end{aligned}
\tag{4-6}
$$

The result now follows by substituting (4-6) with $y = 1$ into (4-4) . $\square$

Remark 4.2. There is another explicit representation of a somewhat different form than (4-3) for the Atkin polynomials. Indeed, it follows from Theorem 4(ii) in [Kaneko and Zagier 1998] that

$$\mathscr{A}_n(x)=\sum_{i=0}^{n}\sum_{m=0}^{i}(-1)^m\binom{\frac{-1}{12}}{i-m}\binom{\frac{-5}{12}}{i-m}\binom{n+\frac{1}{12}}{m}\binom{n-\frac{7}{12}}{m}\binom{2n-1}{m}^{-1}x^{n-i}. \tag{4-7}$$

5. Asymptotics

Wimp [1987, Proof of Theorem 1] showed that the functions u_n and y_n (u_n and v_n in his notation) defined by

$$\begin{aligned} u_n^{(\alpha,\beta)}(x;c)&=(-1)^n\frac{\Gamma(n+\beta+c+1)}{\Gamma(n+c+1)}F\left(\begin{matrix}-n-c,\,n+\alpha+\beta+c+1\\ 1+\beta\end{matrix}\,\middle|\,x\right),\\ y_n^{(\alpha,\beta)}(x;c)&=(-1)^n\frac{\Gamma(n+\alpha+c+1)}{\Gamma(n+\alpha+\beta+c+1)}F\left(\begin{matrix}-n-\beta-c,\,n+\alpha+c+1\\ 1-\beta\end{matrix}\,\middle|\,x\right),\end{aligned} \tag{5-1}$$

satisfy the same recurrence relation satisfied by R_n and $\mathscr{R}_n$, and thus the latter can be represented as linear combinations of the former. We shall slightly modify these functions so as to replace the gamma factors by rising factorials (thus getting rational rather than transcendental coefficients when the parameters are rational) as follows. Set

$$\begin{aligned} U_n^{(\alpha,\beta)}(x;c)&=\frac{\Gamma(c+1)}{\Gamma(\beta+c+1)}u_n^{(\alpha,\beta)}(x;c),\\ Y_n^{(\alpha,\beta)}(x;c)&=\frac{\Gamma(\alpha+\beta+c+1)}{\Gamma(\alpha+c+1)}y_n^{(\alpha,\beta)}(x;c).\end{aligned} \tag{5-2}$$

Thus we have

$$\begin{aligned} U_n^{(\alpha,\beta)}(x;c)&=(-1)^n\frac{(\beta+c+1)_n}{(c+1)_n}F\left(\begin{matrix}-n-c,\,n+\alpha+\beta+c+1\\ 1+\beta\end{matrix}\,\middle|\,x\right),\\ Y_n^{(\alpha,\beta)}(x;c)&=(-1)^n\frac{(\alpha+c+1)_n}{(\alpha+\beta+c+1)_n}F\left(\begin{matrix}-n-\beta-c,\,n+\alpha+c+1\\ 1-\beta\end{matrix}\,\middle|\,x\right).\end{aligned} \tag{5-3}$$

Note that since the factors multiplied by u_n and y_n in (5-2) are independent of n, U_n and Y_n satisfy the same recurrence as R_n and $\mathscr{R}_n$. Indeed, after a simple Kummer transformation, Formula (28) on p. 988 of [Wimp 1987] can be written as

$$\begin{aligned} R_n=\frac{(\beta+c)(\alpha+\beta+c)}{\beta(\alpha+\beta+2c)}&F\left(\begin{matrix}c,\,1-(\alpha+\beta+c)\\ 1-\beta\end{matrix}\,\middle|\,x\right)U_n\\ &-\frac{c(\alpha+c)}{\beta(\alpha+\beta+2c)}F\left(\begin{matrix}\beta+c,\,1-(\alpha+c)\\ 1+\beta\end{matrix}\,\middle|\,x\right)Y_n.\end{aligned} \tag{5-4}$$

Similarly, Theorem 3.10 of [Ismail and Masson 1991] leads to

$$\mathcal{R}_n = F\left(\begin{matrix} c, -(\alpha+\beta+c) \\ -\beta \end{matrix} \,\middle|\, x\right) U_n - \frac{c(\alpha+c)}{\beta(\beta+1)} x F\left(\begin{matrix} 1+\beta+c, 1-(\alpha+c) \\ 2+\beta \end{matrix} \,\middle|\, x\right) Y_n. \tag{5-5}$$

The following theorem provides the analogous representation for the Atkin polynomials:

Theorem 5.1. *Let U_n and Y_n be as in* (5-3), *and set*

$$\begin{aligned} \widetilde{U}_n^{(\alpha,\beta)}(x;c) &= \frac{(c+1)_n(\alpha+\beta+c+1)_n}{(\alpha+\beta+2c+1)_{2n}} U_n^{(\alpha,\beta)}(x;c), \\ \widetilde{Y}_n^{(\alpha,\beta)}(x;c) &= \frac{(c+1)_n(\alpha+\beta+c+1)_n}{(\alpha+\beta+2c+1)_{2n}} Y_n^{(\alpha,\beta)}(x;c). \end{aligned} \tag{5-6}$$

Then we have

$$\mathcal{A}_{n+1}(x) = C(x)\widetilde{U}_n^{(1/2,-2/3)}\big(x;\tfrac{7}{12}\big) + D(x)\widetilde{Y}_n^{(1/2,-2/3)}\big(x;\tfrac{7}{12}\big), \quad n \ge 0, \tag{5-7}$$

with $C(x)$ and $D(x)$ given by

$$\begin{aligned} C(x) &:= \frac{-1}{60}\left(24F\left(\begin{matrix} \frac{-5}{12}, \frac{-5}{12} \\ \frac{-1}{3} \end{matrix} \,\middle|\, x\right) + F\left(\begin{matrix} \frac{-5}{12}, \frac{-5}{12} \\ \frac{2}{3} \end{matrix} \,\middle|\, x\right)\right), \\ D(x) &:= \frac{91}{384} x\left(4F\left(\begin{matrix} \frac{-1}{12}, \frac{-1}{12} \\ \frac{1}{3} \end{matrix} \,\middle|\, x\right) - 5F\left(\begin{matrix} \frac{11}{12}, \frac{-1}{12} \\ \frac{4}{3} \end{matrix} \,\middle|\, x\right)\right). \end{aligned} \tag{5-8}$$

Proof. From (5-4) and (5-5), we see that

$$\begin{aligned} x R_n^{(1/2,-2/3)}\big(x;\tfrac{7}{12}\big) &- \tfrac{5}{12}\mathcal{R}_n^{(1/2,-2/3)}\big(x;\tfrac{7}{12}\big) \\ = \frac{5}{12} U_n^{(1/2,-2/3)}\big(x;\tfrac{7}{12}\big)&\left(\frac{\left(\frac{-1}{12}\right)}{\left(-\frac{2}{3}\right)} x F\left(\begin{matrix} \frac{7}{12}, \frac{7}{12} \\ \frac{5}{3} \end{matrix} \,\middle|\, x\right) - F\left(\begin{matrix} \frac{7}{12}, \frac{-5}{12} \\ \frac{2}{3} \end{matrix} \,\middle|\, x\right)\right) \\ - x Y_n^{(1/2,-2/3)}\big(x;\tfrac{7}{12}\big)&\left(\frac{\left(\frac{7}{12}\right)\left(\frac{13}{12}\right)}{\left(-\frac{2}{3}\right)} F\left(\begin{matrix} \frac{-1}{12}, \frac{-1}{12} \\ \frac{1}{3} \end{matrix} \,\middle|\, x\right) - \frac{5}{12}\frac{\left(\frac{7}{12}\right)\left(\frac{13}{12}\right)}{\left(-\frac{2}{3}\right)\left(\frac{1}{3}\right)} F\left(\begin{matrix} \frac{11}{12}, \frac{-1}{12} \\ \frac{4}{3} \end{matrix} \,\middle|\, x\right)\right). \end{aligned} \tag{5-9}$$

Expanding the hypergeometric series in powers of x, we get, after some computation,

(5-10) $$x R_n^{(1/2,-2/3)}\big(x;\tfrac{7}{12}\big)-\tfrac{5}{12}\mathcal{R}_n^{(1/2,-2/3)}\big(x;\tfrac{7}{12}\big)$$
$$=\frac{-1}{60}\left(24F\left(\begin{matrix}\frac{-5}{12},\frac{-5}{12}\\ \frac{-1}{3}\end{matrix}\,\middle|\,x\right)+F\left(\begin{matrix}\frac{-5}{12},\frac{-5}{12}\\ \frac{2}{3}\end{matrix}\,\middle|\,x\right)\right)U_n^{(1/2,-2/3)}\left(x;\frac{7}{12}\right)$$
$$+\frac{91}{384}x\left(4F\left(\begin{matrix}\frac{-1}{12},\frac{-1}{12}\\ \frac{1}{3}\end{matrix}\,\middle|\,x\right)-5F\left(\begin{matrix}\frac{11}{12},\frac{-1}{12}\\ \frac{4}{3}\end{matrix}\,\middle|\,x\right)\right)Y_n^{(1/2,-2/3)}\left(x;\frac{7}{12}\right),$$

and the result follows from (3-10). □

Theorem 5.1 enables us to obtain an asymptotic formula for the Atkin polynomials:

Theorem 5.2. *Let $C(x)$ and $D(x)$ be as in* (5-8). *For fixed $\theta\in(0,\pi/2)$, the following asymptotic formula holds as $n\to\infty$:*

(5-11) $$\mathcal{A}_{n+1}(\sin^2\theta)$$
$$\sim\frac{(-1)^n}{2^{2n+1}(\cos\theta)(\sin\theta)^{\frac{7}{6}}}C(\sin^2\theta)\frac{\Gamma\big(\frac13\big)(\sin\theta)^{\frac23}}{\Gamma\big(\frac{11}{12}\big)\Gamma\big(\frac{17}{12}\big)}\cos\left[2(n-1)\theta+\frac{\pi}{12}\right]$$
$$+D(\sin^2\theta)\frac{\Gamma\big(\frac53\big)}{\Gamma\big(\frac{13}{12}\big)\Gamma\big(\frac{19}{12}\big)}\cos\left[2(n-1)\theta-\frac{7\pi}{12}\right]$$

Proof. We start by recalling the following asymptotic formula, due to Watson, [Luke 1969, (8), p. 237] (all of our asymptotic formulas will be as $n\to\infty$).

(5-12) $$F\left(\begin{matrix}b-n,n+a\\ d\end{matrix}\,\middle|\,\sin^2\theta\right)$$
$$\sim\frac{\Gamma(d)n^{-d+\frac12}}{\sqrt{\pi}}\;\frac{(\cos\theta)^{d-a-b-\frac12}}{(\sin\theta)^{d-\frac12}}\cos\left[2n\theta+(a-b)\theta-\frac{\pi}{2}\left(d-\frac12\right)\right]$$

for fixed $\theta\in(0,\pi)$. Note that Stirling's formula can be written as

(5-13) $$\Gamma(n+a)\sim\sqrt{2\pi}\,n^{n+a-\frac12}e^{-n}\quad\text{as } n\to\infty,$$

from which we deduce
$$\frac{\Gamma(n+a)}{\Gamma(n+b)}\sim n^{a-b}.$$

Hence

$$u_n(\sin^2\theta)\sim\frac{(-1)^n\Gamma(1+\beta)(\cos\theta)^{-\alpha-\frac12}}{\sqrt{\pi n}(\sin\theta)^{\beta+\frac12}}\cos\left[2n\theta-(\alpha+\beta+2c+1)\theta-\frac{\pi}{2}\left(\beta+\frac12\right)\right],$$
$$y_n(\sin^2\theta)\sim\frac{(-1)^n\Gamma(1-\beta)(\cos\theta)^{-\alpha-\frac12}}{\sqrt{\pi n}(\sin\theta)^{-\beta+\frac12}}\cos\left[2n\theta-(\alpha+\beta+2c+1)\theta+\frac{\pi}{2}\left(\beta-\frac12\right)\right].$$

Also, from (5-13) we get

$$
\begin{aligned}
\text{(5-14)}\quad \frac{(c+1)_n(\alpha+\beta+c+1)_n}{(\alpha+\beta+2c+1)_{2n}} &= \frac{\Gamma(\alpha+\beta+2c+1)}{\Gamma(c+1)\Gamma(\alpha+\beta+c+1)} \frac{\Gamma(n+c+1)\Gamma(n+\alpha+\beta+c+1)}{\Gamma(2n+\alpha+\beta+2c+1)} \\
&\sim \frac{\Gamma(\alpha+\beta+2c+1)}{\Gamma(c+1)\Gamma(\alpha+\beta+c+1)} \sqrt{\pi n}\left(\tfrac{1}{2}\right)^{2n+\alpha+\beta+2c}.
\end{aligned}
$$

Substituting $(\alpha, \beta, c) = \left(\frac{1}{2}, \frac{-2}{3}, \frac{7}{12}\right)$, we see that

$$
\text{(5-15)}\quad
\begin{aligned}
\widetilde{U}_n(\sin^2\theta) &\sim \frac{(-1)^n\Gamma\left(\frac{1}{3}\right)(\sin\theta)^{1/6}}{2^{2n+1}\cos\theta\,\Gamma\left(\frac{11}{12}\right)\Gamma\left(\frac{17}{12}\right)} \cos\left[2(n-1)\theta+\frac{\pi}{12}\right], \\
\widetilde{Y}_n(\sin^2\theta) &\sim \frac{(-1)^n\Gamma\left(\frac{5}{3}\right)(\sin\theta)^{-7/6}}{2^{2n+1}\cos\theta\,\Gamma\left(\frac{13}{12}\right)\Gamma\left(\frac{19}{12}\right)} \cos\left[2(n-1)\theta-\frac{7\pi}{12}\right],
\end{aligned}
$$

and the result follows from (5-7) and (5-8). □

6. Generating functions

We start by recalling a remarkable identity of Flensted-Jensen and Koornwinder [1973]. The interested reader could also consult [Wimp 1987] for more details on various other authors who presented variants of this identity as well as other proofs.

Lemma 6.1. *Let t, x, a, b, d be complex numbers with $x \notin [1, \infty)$ and*

$$
\text{(6-1)}\quad |t| < \frac{1}{|\sqrt{x}+\sqrt{x-1}|^2}.
$$

Then

$$
\begin{aligned}
\text{(6-2)}\quad \sum_{n=0}^{\infty} \frac{(d+a)_n(b)_n}{(a+b+1)_n} F\left(\begin{matrix} -n-a,\, n+b \\ d \end{matrix} \,\middle|\, x\right)\frac{(-t)^n}{n!} \\
= \left(\frac{z_2-t}{z_2+t}\right)^{a+d}\left(\frac{2}{z_2-t}\right)^{b} F\left(\begin{matrix} -a,\, b \\ d \end{matrix} \,\middle|\, \frac{t+z_1}{2t}\right) F\left(\begin{matrix} a+d,\, a+1 \\ a+b+1 \end{matrix} \,\middle|\, \frac{2t}{t+z_2}\right),
\end{aligned}
$$

where $z_1 = 1-\sqrt{(1+t)^2-4xt}$ and $z_2 = 1+\sqrt{(1+t)^2-4xt}$.

To simplify notation we shall write, for $t \neq 0$,

$$
\text{(6-3)}\quad
\begin{aligned}
\delta &= \frac{t+z_1}{2t} = \frac{(1+t)-\sqrt{(1+t)^2-4xt}}{2t}, \\
\epsilon &= \frac{t+z_2}{2t} = \frac{(1+t)+\sqrt{(1+t)^2-4xt}}{2t}.
\end{aligned}
$$

We clearly have

$$t(y-\delta)(y-\epsilon) = ty^2-(1+t)y+x. \tag{6-4}$$

Obviously $z_2+t = 2t\epsilon$, and we also have $z_2-t = 2t(\epsilon-1)$. Furthermore we have $\delta\epsilon = x/t$. Thus we can rewrite (6-2) for $x \neq 0$ as

$$\sum_{n=0}^{\infty} \frac{(d+a)_n(b)_n}{(a+b+1)_n} F\left(\begin{matrix} -n-a,\, n+b \\ d \end{matrix} \,\middle|\, x\right) \frac{(-t)^n}{n!} = \frac{(x-t\delta)^{a+d-b}\delta^b}{x^{a+d}} F\left(\begin{matrix} -a,\, b \\ d \end{matrix} \,\middle|\, \delta\right) F\left(\begin{matrix} a+d,\, a+1 \\ a+b+1 \end{matrix} \,\middle|\, \frac{t}{x}\delta\right). \tag{6-5}$$

The following proposition provides a generating function for U_n and Y_n:

Proposition 6.2. *Let U_n and Y_n be as in* (5-2). *Let t and x be such that $x \notin [1, \infty)$ is nonzero and $|t(\sqrt{x}+\sqrt{x-1})^2| < 1$, and set δ as in* (6-3). *Then*:

$$\sum_{n=0}^{\infty} \frac{(\alpha+\beta+c+1)_n(c+1)_n}{(\alpha+\beta+2c+2)_n} U_n(x) \frac{t^n}{n!} = \frac{\delta^{\alpha+\beta+c+1}}{x^{\beta+c+1}(x-t\delta)^{\alpha}} F\left(\begin{matrix} -c,\, \alpha+\beta+c+1 \\ 1+\beta \end{matrix} \,\middle|\, \delta\right) F\left(\begin{matrix} \beta+c+1,\, c+1 \\ \alpha+\beta+2c+2 \end{matrix} \,\middle|\, \frac{t\delta}{x}\right), \tag{6-6}$$

$$\sum_{n=0}^{\infty} \frac{(\alpha+\beta+c+1)_n(c+1)_n}{(\alpha+\beta+2c+2)_n} Y_n(x) \frac{t^n}{n!} = \frac{\delta^{\alpha+c+1}}{x^{c+1}(x-t\delta)^{\alpha}} F\left(\begin{matrix} -\beta-c,\, \alpha+c+1 \\ 1-\beta \end{matrix} \,\middle|\, \delta\right) F\left(\begin{matrix} c+1,\, \beta+c+1 \\ \alpha+\beta+2c+2 \end{matrix} \,\middle|\, \frac{t\delta}{x}\right). \tag{6-7}$$

Proof. From (5-1), we see that

$$\frac{\Gamma(c+1)}{\Gamma(\beta+c+1)} \frac{(\alpha+\beta+c+1)_n(c+1)_n}{(\alpha+\beta+2c+2)_n} u_n = \frac{(c+\beta+1)_n(\alpha+\beta+c+1)_n}{(\alpha+\beta+2c+2)_n} F\left(\begin{matrix} -n-c,\, n+\alpha+\beta+c+1 \\ 1+\beta \end{matrix} \,\middle|\, x\right) \tag{6-8}$$

and

$$\frac{\Gamma(\alpha+\beta+c+1)}{\Gamma(\alpha+c+1)} \frac{(\alpha+\beta+c+1)_n(c+1)_n}{(\alpha+\beta+2c+2)_n} y_n = \frac{(c+1)_n(\alpha+c+1)_n}{(\alpha+\beta+2c+2)_n} F\left(\begin{matrix} -n-\beta-c,\, n+\alpha+c+1 \\ 1-\beta \end{matrix} \,\middle|\, x\right). \tag{6-9}$$

The identities (6-6) and (6-7) follow from applying (6-5) with the choices $(a, b, d) = (c, \alpha+\beta+c+1, \beta+1)$ and $(a, b, d) = (\beta+c, \alpha+c+1, 1-\beta)$, respectively. □

Remark 6.3. The result in Proposition 6.2 is essentially due to Wimp. However, we take this opportunity to correct a misprint in the statement of Theorem 5 in [Wimp 1987]: In the first line of page 999, the parameter "$\gamma+c+\beta$" should be replaced by "$\gamma+c-\beta$" (in our notation, the later is $\alpha+c+1$ while the former would be $\alpha+c+1+2\beta$, which indeed doesn't ever seem to figure in the theory).

We next obtain a generating function identity for the Atkin polynomials scaled by a rather unexpected appearance of the Catalan numbers. The right-hand side of the generating series has four summands; each is up to a relatively simple multiple a product of three hypergeometric functions in the variables x, δ and $1/\epsilon = t\delta/x$.

Theorem 6.4. *Let $C(x)$ and $D(x)$ be as in* (5-8), *and δ as in* (6-3). *Furthermore, let $\{C_n = 1/(n+1)\binom{2n}{n}\}_n$ denote the sequence of Catalan numbers.*

(1) *For $0 < x < 1$ and $|t| < 1$, we have*

$$\sum_{n=0}^{\infty} C_{n+1}\mathcal{A}_{n+1}(x)t^n = \frac{\delta^{17/12}}{x^{11/12}\sqrt{x-t\delta}} F\left(\begin{matrix}\frac{11}{12},\ \frac{19}{12}\\ 3\end{matrix}\,\middle|\,\frac{t\delta}{x}\right) \times\left[C(x)F\left(\begin{matrix}\frac{-7}{12},\ \frac{17}{12}\\ \frac{1}{3}\end{matrix}\,\middle|\,\delta\right)+D(x)\left(\frac{x}{\delta}\right)^{2/3}F\left(\begin{matrix}\frac{1}{12},\ \frac{25}{12}\\ \frac{5}{3}\end{matrix}\,\middle|\,\delta\right)\right]. \tag{6-10}$$

(2) *For $|t| < 1$, we have*

$$\sum_{n=0}^{\infty} C_{n+1}\mathcal{A}_{n+1}(0)(-t)^n = \tfrac{-5}{12} F\left(\begin{matrix}\frac{11}{12},\ \frac{17}{12}\\ 3\end{matrix}\,\middle|\,t\right), \tag{6-11}$$

and consequently for $n \geq 0$, we have

$$\mathcal{A}_{n+1}(0) = (-1)^n\left(\tfrac{-5}{12}\right)\frac{\left(\frac{11}{12}\right)_n\left(\frac{17}{12}\right)_n}{(2n+1)!}. \tag{6-12}$$

Proof. Note that for $0 \leq x < 1$, we have

$$|\sqrt{x}+\sqrt{x-1}|^2 = |\sqrt{x}+i\sqrt{1-x}|^2 = 1,$$

so (6-1) indeed translates into $|t| < 1$. Now, using (5-6), we see that

$$\begin{aligned}\frac{(\alpha+\beta+c+1)_n(c+1)_n}{(\alpha+\beta+2c+2)_n}U_n &= \frac{(\alpha+\beta+2c+1)_{2n}}{(\alpha+\beta+2c+2)_n}\widetilde{U}_n \\ &= (\alpha+\beta+2c+1)\frac{(\alpha+\beta+2c+1+n)_n}{(\alpha+\beta+2c+1+n)}\widetilde{U}_n,\end{aligned} \tag{6-13}$$

with a similar identity for Y_n, and (6-10) now follows from (5-7) and Proposition 6.2 by substituting $(\alpha, \beta, c) = \left(\frac{1}{2}, \frac{-2}{3}, \frac{7}{12}\right)$.

When $x = 0$ and $|t| < 1$, then, in the notation of (6-2), we have $t+z_1 = 0$ and $z_2+t = 2(1+t)$, and hence $z_2-t = 2$. Furthermore, we have $C(0) = \frac{-5}{12}$ and

$D(0) = 0$. It thus follows from (6-2) and (5-7) that

$$2\sum_{n=0}^{\infty}\binom{2n+1}{n}\mathscr{A}_{n+1}(0)\frac{t^n}{2+n} = \frac{-5}{12(1+t)^{11/12}}F\left(\begin{matrix}\frac{11}{12},\frac{19}{12}\\3\end{matrix}\,\middle|\,\frac{t}{1+t}\right). \tag{6-14}$$

Replacing t with $(-t)$ and applying the Pfaff–Kummer transformation [Erdélyi et al. 1953, Formula (2) on p. 105], we obtain (6-11), from which (6-12) follows by comparing coefficients and simplifying. □

Remark 6.5. Formula (6-12) can also be obtained directly from the defining recursion of the Atkin polynomials, as in Proposition 6 of [Kaneko and Zagier 1998]. In that same proposition, and again using only the defining recurrence (1-1), Kaneko and Zagier also obtain a formula equivalent to

$$\mathscr{A}_{n+1}(1) = \frac{7}{12}\frac{\left(\frac{11}{12}\right)_n\left(\frac{19}{12}\right)_n}{(2n+1)!}. \tag{6-15}$$

Taking a hint from (6-11), it is straightforward to prove directly from (6-15) that for $|t| < 1$, we have

$$\sum_{n=0}^{\infty}C_{n+1}\mathscr{A}_{n+1}(1)t^n = \frac{7}{12}F\left(\begin{matrix}\frac{11}{12},\frac{19}{12}\\3\end{matrix}\,\middle|\,t\right). \tag{6-16}$$

Alternatively, one can prove (6-16) in a manner similar to (6-11), bearing in mind that we have $C(1) = D(1) = 0$, whereas $\widetilde{U}_n^{(1/2,-2/3)}\left(x;\,\frac{7}{12}\right)$ and $\widetilde{V}_n^{(1/2,-2/3)}\left(x;\,\frac{7}{12}\right)$ have simple poles at $x = 1$, and thus their product is to be interpreted in the limit $x \to 1^-$ as the derivative of the former multiplied by the residue of the latter.

7. The weight function for the Atkin polynomials

Kaneko and Zagier [1998] gave the weight function for the Atkin polynomials $A_n(j)$ on [0, 1728] as

$$w(j) = \frac{6}{\pi}\theta'(j), \tag{7-1}$$

where $\theta : [0, 1728] \to [\pi/3, \pi/2]$ is the inverse of the monotone increasing function $\theta \mapsto j(e^{i\theta})$, where $j(\tau)$ is the usual modular j-invariant from the theory of modular forms. In this section we derive an explicit description of the weight function in terms of hypergeometric series. Formula (25) on p. 20 of [Erdélyi et al. 1953] states that an inverse for the scaled j-invariant given by

$$J(z) = \frac{j(z)}{1728}$$

is obtainable by the formula

$$z = e^{2\pi i/3} \frac{F - \lambda e^{i\pi/3} J^{1/3} F^*}{F - \lambda e^{-i\pi/3} J^{1/3} F^*}, \tag{7-2}$$

where

$$\begin{aligned} F(J) &= {}_2F_1\left(\begin{matrix} \frac{1}{12}, \frac{1}{12} \\ \frac{2}{3} \end{matrix} \,\middle|\, J \right), \\ F^*(J) &= {}_2F_1\left(\begin{matrix} \frac{5}{12}, \frac{5}{12} \\ \frac{4}{3} \end{matrix} \,\middle|\, J \right), \\ \lambda &= \frac{\Gamma\left(\frac{2}{3}\right)\Gamma\left(\frac{5}{12}\right)\Gamma\left(\frac{11}{12}\right)}{\Gamma\left(\frac{4}{3}\right)\Gamma\left(\frac{1}{12}\right)\Gamma\left(\frac{7}{12}\right)} = (2-\sqrt{3}) \frac{\Gamma\left(\frac{2}{3}\right)\Gamma^2\left(\frac{11}{12}\right)}{\Gamma\left(\frac{4}{3}\right)\Gamma^2\left(\frac{7}{12}\right)}. \end{aligned} \tag{7-3}$$

We must note that this is one inverse of many as J is invariant under modular transformations. This particular formula gives, easily, that $z(0) = e^{2\pi i/3}$. In order to use the same intervals as in [Kaneko and Zagier 1998], we consider another inverse, corresponding to applying $z \mapsto -1/z$, thus obtaining

$$z(J) = e^{\pi i/3} \frac{F - \lambda e^{-i\pi/3} J^{1/3} F^*}{F - \lambda e^{i\pi/3} J^{1/3} F^*}. \tag{7-4}$$

It is straightforward to verify that using (7-4), we get $z(0) = e^{\pi i/3}$ and $z(1) = e^{\pi i/2}$. For $0 \le J \le 1$, F and F^* are computed in terms of the converging hypergeometric series and hence are real. Thus in the ratio

$$\frac{F(J) - \lambda e^{-i\pi/3} J^{1/3} F^*(J)}{F(J) - \lambda e^{i\pi/3} J^{1/3} F^*(J)}$$

the denominator is the complex conjugate of the numerator. Hence the ratio has absolute value equal to 1, and is of the form $e^{i\rho}$. We will show below that $0 \le \rho \le \pi/6$. Thus an explicit description of the function $\theta(j) : [0, 1728] \to [\pi/3, \pi/2]$ is given by $\theta(j) = \phi(j/1728)$, where $\phi(J) : [0, 1] \to [\pi/3, \pi/2]$ is defined by

$$\phi(J) = \frac{\pi}{3} - i \log\left(\frac{F(J) - \lambda e^{-i\pi/3} J^{1/3} F^*(J)}{F(J) - \lambda e^{i\pi/3} J^{1/3} F^*(J)} \right) = \frac{\pi}{3} + \rho(J),$$

and we have

$$\begin{aligned} \phi'(J) &= -i \frac{F(J) - \lambda e^{i\pi/3} J^{1/3} F^*(J)}{F(J) - \lambda e^{-i\pi/3} J^{1/3} F^*(J)} \left(\frac{F(J) - \lambda e^{-i\pi/3} J^{1/3} F^*(J)}{F(J) - \lambda e^{i\pi/3} J^{1/3} F^*(J)} \right)' \\ &= -i \frac{W(J)}{|F(J) - \lambda e^{-i\pi/3} J^{1/3} F^*(J)|^2}, \end{aligned} \tag{7-5}$$

where $W(J)$ is given explicitly by

(7-6) $W(J)$

$$= \big(F(J) - \lambda e^{i\pi/3} J^{1/3} F^*(J)\big)\big(F'(J) - \lambda e^{-i\pi/3} J^{1/3} (F^*)'(J) - \tfrac{\lambda}{3} e^{-i\pi/3} J^{-2/3} F^*(J)\big)$$
$$- \big(F(J) - \lambda e^{-i\pi/3} J^{1/3} F^*(J)\big)\big(F'(J) - \lambda e^{i\pi/3} J^{1/3} (F^*)'(J) - \tfrac{\lambda}{3} e^{i\pi/3} J^{-2/3} F^*(J)\big)$$
$$= \tfrac{\lambda}{3} J^{-2/3} i\sqrt{3}\big(F(J)F^*(J) + 3JF(J)(F^*)'(J) - 3JF'(J)F^*(J)\big).$$

We also note that W is the Wronskian of two linearly independent solutions for the equation

$$z(1-z)\frac{d^2u}{dz^2} + (c - (1+a+b)z)\frac{du}{dz} - abu = 0,$$

where here $a = b = \frac{1}{12}$ and $c = \frac{2}{3}$. It follows that W itself satisfies the equation

$$z(1-z)\frac{dW}{dz} = ((a+b+1)z - c)W. \tag{7-7}$$

On the open interval (0, 1), (7-7) has solution

$$W(J) = BJ^{-2/3}(1-J)^{-1/2}. \tag{7-8}$$

To determine the constant B we compare the coefficient of $J^{-2/3}$ in (7-8) and (7-6) to get

$$B = \frac{i\lambda}{\sqrt{3}}.$$

Hence

$$\phi'(J) = \frac{\lambda}{\sqrt{3}} \frac{J^{-2/3}(1-J)^{-1/2}}{|F(J) - \lambda e^{-i\pi/3} J^{1/3} F^*(J)|^2}. \tag{7-9}$$

The fact that the derivative is positive for $0 \le J \le 1$ implies that $\phi(J)$ is monotone increasing, and hence that it is bounded between $\phi(0)$ and $\phi(1)$, as we claimed above.

Note that

$$w(j) = \frac{6}{\pi}\theta'(j) = \frac{6}{1728\pi}\phi'\left(\frac{j}{1728}\right) \tag{7-10}$$
$$= \frac{6\lambda}{1728\pi\sqrt{3}} \frac{12(12^2 j^{-2/3})((1728-j)^{-1/2} 12^{3/2})}{\big|12F\big(\frac{j}{1728}\big) - \lambda e^{-i\pi/3} j^{1/3} F^*\big(\frac{j}{1728}\big)\big|^2}.$$

We have thus proved the following theorem:

Theorem 7.1. *Let* λ *be as in* (7-3). *Then the normalized weight function for the Atkin polynomials* $A_n(j)$ *on the interval* [0, 1728] *is given by*

$$w(j) = \frac{144\lambda}{\pi} \frac{j^{-2/3}(1728-j)^{-1/2}}{\left|12F\left(\begin{matrix}\frac{1}{12}, \frac{1}{12} \\ \frac{2}{3}\end{matrix} \,\middle|\, \frac{j}{1728}\right) - \lambda e^{-i\pi/3} j^{1/3} F\left(\begin{matrix}\frac{5}{12}, \frac{5}{12} \\ \frac{4}{3}\end{matrix} \,\middle|\, \frac{j}{1728}\right)\right|^2}. \tag{7-11}$$

Acknowledgements

Mourad Ismail wishes to acknowledge the hospitality of Texas A&M University at Qatar during his visit in September 2012, when this work started. Both authors thank Richard Askey, Jet Foncannon (Wimp), Masanobu Kaneko, and Don Zagier for their interest in this work and for very helpful hints. They are also grateful for the referee for constructive remarks and for suggesting a number of additional references related to the paper.

References

[Andrews et al. 1999] G. E. Andrews, R. Askey, and R. Roy, *Special functions*, Encyclopedia of Mathematics and its Applications **71**, Cambridge University Press, 1999. MR 2000g:33001 Zbl 0920.33001

[Chihara 1978] T. S. Chihara, *An introduction to orthogonal polynomials*, Mathematics and its Applications **13**, Gordon and Breach, New York, 1978. MR 58 #1979 Zbl 0389.33008

[Erdélyi et al. 1953] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher transcendental functions, I*, McGraw-Hill, New York, 1953. MR 15,419i Zbl 0051.30303

[Flensted-Jensen and Koornwinder 1973] M. Flensted-Jensen and T. Koornwinder, "The convolution structure for Jacobi function expansions", *Ark. Mat.* **11** (1973), 245–262. MR 49 #5688 Zbl 0267.42009

[Ismail 2009] M. E. H. Ismail, *Classical and quantum orthogonal polynomials in one variable*, Encyclopedia of Mathematics and its Applications **98**, Cambridge University Press, 2009. MR 2010i:33001 Zbl 1172.42008

[Ismail and Masson 1991] M. E. H. Ismail and D. R. Masson, "Two families of orthogonal polynomials related to Jacobi polynomials", pp. 359–375 in *Proceedings of the U.S.-Western Europe Regional Conference on Padé Approximants and Related Topics* (Boulder, CO, 1988), vol. 21, 1991. MR 93a:33018 Zbl 0744.33004

[Ismail et al. 1988] M. E. H. Ismail, J. Letessier, and G. Valent, "Linear birth and death models and associated Laguerre and Meixner polynomials", *J. Approx. Theory* **55**:3 (1988), 337–348. MR 89m:60204 Zbl 0656.60092

[Kaneko and Zagier 1998] M. Kaneko and D. Zagier, "Supersingular j-invariants, hypergeometric series, and Atkin's orthogonal polynomials", pp. 97–126 in *Computational perspectives on number theory* (Chicago, IL, 1995), edited by D. A. Buell and J. T. Teitelbaum, AMS/IP Stud. Adv. Math. **7**, Amer. Math. Soc., Providence, RI, 1998. MR 99b:11064 Zbl 0955.11018

[Luke 1969] Y. L. Luke, *The special functions and their approximations, I*, Mathematics in Science and Engineering **53**, Academic Press, New York, 1969. MR 39 #3039 Zbl 0193.01701

[Rainville 1960] E. D. Rainville, *Special functions*, Macmillan, New York, 1960. MR 21 #6447 Zbl 0092.06503

[Szegő 1975] G. Szegő, *Orthogonal polynomials*, 4th ed., Colloquium Publications **23**, Amer. Math. Soc., Providence, R.I., 1975. MR 51 #8724 Zbl 0305.42011

[Wimp 1987] J. Wimp, "Explicit formulas for the associated Jacobi polynomials and some applications", *Canad. J. Math.* **39**:4 (1987), 983–1000. MR 88k:33023 Zbl 0643.33009

Received April 14, 2013. Revised September 9, 2013.

AHMAD EL-GUINDY
DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
CAIRO UNIVERSITY
GIZA 12613
EGYPT
Current address:
SCIENCE PROGRAM
TEXAS A&M UNIVERSITY AT QATAR
DOHA, 23874
QATAR
a.elguindy@gmail.com

MOURAD E. H. ISMAIL
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CENTRAL FLORIDA
ORLANDO, FL 32828
UNITED STATES
and
DEPARTMENT OF MATHEMATICS
KING SAUD UNIVERSITY
RIYADH
SAUDI ARABIA
mourad.eh.ismail@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.272.131

EVOLVING CONVEX CURVES TO CONSTANT-WIDTH ONES BY A PERIMETER-PRESERVING FLOW

LAIYUAN GAO AND SHENGLIANG PAN

This paper deals with a curve evolution problem which, if the curvature of the initial convex curve satisfies a certain pinching condition, keeps the convexity and preserves the perimeter, while increasing the enclosed area of the evolving curve, and which leads to a limiting curve of constant width. In particular, under this flow the limiting curve is a circle if and only if the initial convex curve is centrosymmetric.

1. Introduction

Denote by S^1 the unit circle centered at the origin of the Euclidean plane $\mathbb{R}^2$. Let $X_0(\varphi)$, $\varphi \in S^1$, be a closed C^∞ curve in the plane. A curve evolution problem is usually defined as

$$\begin{cases} (\partial X/\partial t)(\varphi,t) = \beta(\varphi,t)N(\varphi,t), & (\varphi,t) \in S^1 \times (0,T), \\ X(\varphi,0) = X_0(\varphi), & \varphi \in S^1, \end{cases}$$

where $X(\varphi,t) = (x(\varphi,t), y(\varphi,t))$ is the position vector of the evolving curve, $N(\varphi,t)$ its unit normal vector field and $\beta(\varphi,t)$ some geometric quantity depending on the evolving curve. Such problems arise in many fields, such as image processing [Cao 2003], phase transitions [Gurtin 1993], etc. In fact, the above evolution problem has been studied extensively, for example, for the popular curve-shortening flow [Gage 1984; Gage and Hamilton 1986; Grayson 1987], the area-preserving flows [Gage 1986; Mao et al. 2013; Ma and Cheng 2014], the perimeter-preserving flows [Pan and Yang 2008; Ma and Zhu 2012] and in other related research [Angenent 1991; Chow and Tsai 1996; Andrews 1998; Urbas 1999; Chao et al. 2013]. One can find more background material in the book [Chou and Zhu 2001].

Let θ be the tangential angle, i.e., the oriented angle from the positive x-axis to the unit tangential vector of the curve. If the initial curve X_0 is strictly convex then it can be parameterized by θ. In this paper, we will focus on the following curve

---

This work is supported by the National Science Foundation of China (grant 11171254).
MSC2010: 35K15, 35K55, 53A04.
Keywords: convex curves, curves of constant width, perimeter-preserving curve flow.

evolution problem in the plane $\mathbb{R}^2$:

$$\begin{cases} (\partial X/\partial t)(\theta,t) = (w(\theta,t)-\eta(\theta,t))N(\theta,t), & (\theta,t)\in S^1\times(0,T), \\ X(\theta,0) = X_0(\theta), & \theta\in S^1, \end{cases} \tag{1-1}$$

where $w(\theta,t)$ is the width of the evolving curve and $\eta(\theta,t)$ is $\rho(\theta,t)+\rho(\theta+\pi,t)$; here $\rho(\theta,t)$ is the radius of curvature of the curve. Using the Minkowski support function of a convex curve (see [Hsiung 1981; Schneider 1993; Groemer 1996]), one can easily see that $w(\theta,t)-\eta(\theta,t) = -(\partial^2 w/\partial\theta^2)(\theta,t)$, and thus it is obvious that constant-width curves are invariant (i.e., stable) under this flow.

The following theorem is the main result of our paper.

Main Theorem. *Let $X_0(\theta)$ be a strictly convex smooth curve in the plane $\mathbb{R}^2$ which evolves according to the flow* (1-1). *Denote by $\rho_0(\theta)$ the radius of curvature of $X_0(\theta)$ and set*

$$M = \max\{\rho_0(\theta) \mid \theta\in S^1\}, \quad m = \min\{\rho_0(\theta)\mid \theta\in S^1\}.$$

If the pinching condition

$$M < 3m \tag{1-2}$$

holds for $X_0(\theta)$, i.e., $\frac{1}{3} < m/M \le 1$, then (1-1) *has a global solution $X(\theta,t)$ for $(\theta,t)\in S^1\times[0,\infty)$. As time passes, the flow keeps the convexity, preserves the perimeter while increasing the enclosed area of the evolving curve, and makes the curve more and more circular. As the time t goes to infinity, the curve $X(\cdot\,,t)$ evolves smoothly to a curve of constant width L_0/π, where L_0 is the perimeter of the initial convex curve $X_0(\theta)$. In particular, the limiting curve is a circle if and only if the initial curve is centrosymmetric.*

If a smooth simple closed curve evolves under the curve shortening flow then it converges to a round point (see [Gage 1984; Gage and Hamilton 1986; Grayson 1987]). In the cases of nonlocal flows for convex curves, the limiting curves are finite circles (see [Gage 1986; Jiang and Pan 2008; Pan and Zhang 2010; Ma and Cheng 2014]). Forming a striking contrast to these researches, although in the present case the evolving curve keeps its convexity and becomes more and more circular, the limiting curve of the flow is only of constant width rather than being a circle.

This paper is organized as follows. In Section 2, we will compute the evolution equations of the commonly used geometric quantities, and reduce the nonlinear problem (1-1) to the Cauchy problem

$$\begin{cases} (\partial\rho/\partial t)(\theta,t) = (\partial^2\eta/\partial\theta^2)(\theta,t), \\ (\partial\eta/\partial t)(\theta,t) = 2(\partial^2\eta/\partial\theta^2)(\theta,t), \\ \rho(\theta,0) = \rho_0(\theta), \\ \eta(\theta,0) = \eta_0(\theta), \end{cases} \tag{1-3}$$

where θ is the tangential angle and (θ, t) is in $S^1 \times [0, T)$. In Section 3, we will show that the Cauchy problem (1-3) has a bounded positive solution in $S^1 \times [0, +\infty)$, provided that condition (1-2) holds. We will prove that the evolving curve maintains its convexity and is of the same length as the initial convex curve. As time tends to infinity, the asymptotic behavior of the evolving curve will be considered. In Section 4 we will give several examples.

2. Some preparations

In this section, we will first calculate the evolution equations of the commonly used geometric quantities, and then give the equivalence between the curve evolution problem (1-1) and the Cauchy problem (1-3). Now, we suppose that there exists a family of convex curves $X(\varphi, t)$ evolving according to (1-1).

To make the tangential angle θ a variable independent of time t, let us consider the following flow instead of (1-1):

$$\begin{cases} (\partial \widetilde{X}/\partial t)(\theta, t) = \alpha(\theta, t) T(\theta, t) + (w(\theta, t) - \eta(\theta, t)) N(\theta, t), \\ \widetilde{X}(\theta, 0) = X_0(\theta), \end{cases} \tag{2-1}$$

where $\alpha = \alpha(\theta, t)$ is to be determined. Set $\beta(\theta, t) = w(\theta, t) - \eta(\theta, t)$. By [Chou and Zhu 2001, Proposition 1.1, p. 6], the solution of (2-1), $\widetilde{X}(\cdot, t)$, differs from the solution of (1-1), $X(\cdot, t)$, only by altering the parametrization. Therefore, we just need to calculate the evolution equations of ρ and η under the flow (2-1).

Let s be the arc length of the curve $\widetilde{X}(\cdot, t)$. The metric of the curve is given by $g(\varphi, t) = \|\partial \widetilde{X}/\partial \varphi\|$. From the Frenet formulae it follows that

$$\begin{aligned} \frac{\partial g}{\partial t} &= \frac{1}{g} \left\langle \frac{\partial}{\partial t} \frac{\partial \widetilde{X}}{\partial \varphi}, \frac{\partial \widetilde{X}}{\partial \varphi} \right\rangle = \frac{1}{g} \left\langle \frac{\partial}{\partial \varphi} (\alpha T + \beta N), \frac{\partial \widetilde{X}}{\partial \varphi} \right\rangle \\ &= \frac{1}{g} \left\langle g \frac{\partial}{\partial s} (\alpha T + \beta N), g T \right\rangle = \left(\frac{\partial \alpha}{\partial s} - \beta \kappa \right) g. \end{aligned}$$

Therefore, one gets

$$\begin{aligned} \frac{\partial}{\partial t} \frac{\partial}{\partial s} &= \frac{\partial}{\partial t} \left(\frac{1}{g} \frac{\partial}{\partial \varphi} \right) = -\frac{1}{g^2} \frac{\partial g}{\partial t} \frac{\partial}{\partial \varphi} + \frac{1}{g} \frac{\partial}{\partial t} \frac{\partial}{\partial \varphi} \\ &= -\frac{1}{g} \frac{\partial g}{\partial t} \frac{\partial}{\partial s} + \frac{1}{g} \frac{\partial}{\partial \varphi} \frac{\partial}{\partial t} = \left(-\frac{\partial \alpha}{\partial s} + \beta \kappa \right) \frac{\partial}{\partial s} + \frac{\partial}{\partial s} \frac{\partial}{\partial t} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial T}{\partial t} &= \frac{\partial}{\partial t} \frac{\partial X}{\partial s} = \frac{\partial}{\partial s} \frac{\partial X}{\partial t} + \left(-\frac{\partial \alpha}{\partial s} + \beta \kappa \right) \frac{\partial X}{\partial s} \\ &= \frac{\partial}{\partial s} (\alpha T + \beta N) + \left(-\frac{\partial \alpha}{\partial s} + \beta \kappa \right) T = \left(\alpha \kappa + \frac{\partial \beta}{\partial s} \right) N, \end{aligned}$$

$$\frac{\partial N}{\partial t}=\left\langle\frac{\partial N}{\partial t},T\right\rangle T+\left\langle\frac{\partial N}{\partial t},N\right\rangle N=-\left\langle N,\frac{\partial T}{\partial t}\right\rangle T+0=-\left(\alpha\kappa+\frac{\partial\beta}{\partial s}\right)T.$$

Since the Frenet frame can be expressed via the tangential θ as $T=(\cos\theta,\sin\theta)$, $N=(-\sin\theta,\cos\theta)$, one can obtain the Frenet formulae

$$\frac{\partial T}{\partial\theta}=N,\qquad \frac{\partial N}{\partial\theta}=-T. \tag{2-2}$$

The definition of curvature κ implies that $\partial\theta/\partial s=\kappa$ or $\partial s/\partial\theta=\rho$. Noticing that

$$\left(\alpha\kappa+\frac{\partial\beta}{\partial s}\right)N=\frac{\partial T}{\partial t}=\frac{\partial}{\partial t}(\cos\theta,\sin\theta)=(-\sin\theta,\cos\theta)\frac{\partial\theta}{\partial t}=\frac{\partial\theta}{\partial t}N,$$

one obtains that

$$\frac{\partial\theta}{\partial t}=\alpha\kappa+\frac{\partial\beta}{\partial s}. \tag{2-3}$$

From (2-3), if we set $\alpha=-(1/\kappa)(\partial\beta/\partial s)=-\partial\beta/\partial\theta$, then the tangential angle θ is independent of t and so are T and N. The evolution equation of the Minkowski support function p of the evolving curve is given by

$$\frac{\partial p}{\partial t}=-\frac{\partial}{\partial t}\langle X,N\rangle=-\left\langle\frac{\partial X}{\partial t},N\right\rangle+0=-\beta=\eta-w.$$

Since

$$\begin{aligned}\frac{\partial p}{\partial\theta}&=-\left\langle\frac{\partial X}{\partial\theta},N\right\rangle-\left\langle X,\frac{\partial N}{\partial\theta}\right\rangle=\langle X,T\rangle,\\ \frac{\partial^2 p}{\partial\theta^2}&=\frac{\partial}{\partial\theta}\langle X,T\rangle=\left\langle\frac{\partial X}{\partial s}\frac{\partial s}{\partial\theta},T\right\rangle+\langle X,N\rangle=\rho-p,\end{aligned} \tag{2-4}$$

one gets

$$\rho=\frac{\partial^2 p}{\partial\theta^2}+p$$

and

$$\begin{aligned}\eta(\theta,t)&=\rho(\theta,t)+\rho(\theta+\pi,t)\\ &=p(\theta,t)+\frac{\partial^2 p}{\partial\theta^2}(\theta,t)+p(\theta+\pi,t)+\frac{\partial^2 p}{\partial\theta^2}(\theta+\pi,t)\\ &=\frac{\partial^2 w}{\partial\theta^2}(\theta,t)+w(\theta,t).\end{aligned}$$

From the evolution equation of the support function p and from the definition of width, $w(\theta,t):=p(\theta,t)+p(\theta+\pi,t)$, we have

$$\frac{\partial w}{\partial t}=2\eta-2w=2\frac{\partial^2 w}{\partial\theta^2}$$

and

$$
\begin{aligned}
\frac{\partial \rho}{\partial t} &= \frac{\partial}{\partial t}\left(\frac{\partial^2 p}{\partial \theta^2} + p\right) = \frac{\partial^2}{\partial \theta^2}\frac{\partial p}{\partial t} + \frac{\partial p}{\partial t} = \frac{\partial^2}{\partial \theta^2}(\eta - w) + (\eta - w) \\
&= \frac{\partial^2 \eta}{\partial \theta^2} - \frac{\partial^2 w}{\partial \theta^2} + \eta - w = \frac{\partial^2 \eta}{\partial \theta^2}, \\
\frac{\partial \eta}{\partial t} &= \frac{\partial}{\partial t}\big(\rho(\theta,t) + \rho(\theta+\pi,t)\big) = 2\frac{\partial^2 \eta}{\partial \theta^2}.
\end{aligned}
\tag{2-5}
$$

Now, we can conclude that if there is a family of convex curves $X(\cdot\,,t)$ evolving according to the flow (2-1), then the Cauchy problem (1-3) is solvable for some $T > 0$. The following theorem can tell us that the contrary also holds.

Theorem 2.1. *The curve evolution problem* (1-1) *is equivalent to the Cauchy problem* (1-3) *for some $T > 0$, if the initial curve $X(\varphi, 0) = X_0(\varphi)$ is smooth and strictly convex.*

Proof. We just need to prove that if the Cauchy problem (1-3) has a solution $\eta(\theta,t)$ for some $T > 0$ then the evolution problem (2-1) is solvable for $(\theta,t) \in S^1 \times [0,T)$. Define a family of curves $\widetilde{X}(\theta,t) = (\tilde{x}(\theta,t), \tilde{y}(\theta,t)) + (C_1(t), C_2(t))$ by setting

$$
\begin{aligned}
\tilde{x}(\theta,t) &= \int_0^\theta \rho(\phi,t)\cos\phi\,d\phi, & \tilde{y}(\theta,t) &= \int_0^\theta \rho(\phi,t)\sin\phi\,d\phi. \\
C_1(t) &= \int_0^t \Big(\frac{\partial \eta}{\partial \theta}(0,\tau) - \frac{\partial w}{\partial \theta}(0,\tau)\Big)\,d\tau, & C_2(t) &= \int_0^t (w(0,\tau) - \eta(0,\tau))\,d\tau.
\end{aligned}
$$

Direct computation gives us

$$
\begin{aligned}
\frac{\partial \tilde{x}}{\partial t} &= \int_0^\theta \frac{\partial \rho}{\partial t}(\phi,t)\cos\phi\,d\phi = \int_0^\theta \frac{\partial^2 \eta}{\partial \phi^2}(\phi,t)\cos\phi\,d\phi \\
&= \frac{\partial \eta}{\partial \theta}\cos\theta - \frac{\partial \eta}{\partial \theta}(0,t) + \eta\sin\theta - \int_0^\theta \eta\cos\phi\,d\phi \\
&= \frac{\partial \eta}{\partial \theta}\cos\theta - \frac{\partial \eta}{\partial \theta}(0,t) + \eta\sin\theta - \int_0^\theta \left(\frac{\partial^2 w}{\partial \phi^2} + w\right)\cos\phi\,d\phi \\
&= \left(\frac{\partial \eta}{\partial \theta} - \frac{\partial w}{\partial \theta}\right)\cos\theta - \left(\frac{\partial \eta}{\partial \theta}(0,t) - \frac{\partial w}{\partial \theta}(0,t)\right) + (\eta - w)\sin\theta.
\end{aligned}
$$

And similarly, one can get

$$
\frac{\partial \tilde{y}}{\partial t} = \left(\frac{\partial \eta}{\partial \theta} - \frac{\partial w}{\partial \theta}\right)\sin\theta + (\eta(0,t) - w(0,t)) + (w - \eta)\cos\theta.
$$

So the curve $\widetilde{X}(\theta,t)$ satisfies

$$\frac{\partial \widetilde{X}}{\partial t} = -\frac{\partial}{\partial \theta}(w-\eta)T + (w-\eta)N,$$

where $\{T, N\}$ is the Frenet frame of the curve $\widetilde{X}$, which implies that the flow (2-1) has a solution since we have chosen $\alpha = -(1/\kappa)(\partial\beta/\partial s) = -\partial\beta/\partial\theta$. Therefore the original problem (1-2) is also solvable on $S^1 \times [0, T)$. □

3. Global existence and convergence of the flow

Because we can reduce the curve evolution problem (1-1) to a Cauchy problem (1-3) for small t, the local existence of the flow (1-1) is a direct corollary of the classical theory of heat equations. In this section, we will first prove that problem (1-1) has a unique, convex and smooth solution curve $X(\varphi,t)$ on $S^1 \times [0,+\infty)$; i.e., the Cauchy problem (1-3) has a positive and smooth solution $(\rho(\cdot,t), \eta(\cdot,t))$ for $t \geq 0$, provided that the pinching condition (1-2) holds for the initial curve. Then we will show that the curve $X(\cdot,t)$ evolves to a constant-width curve smoothly.

Lemma 3.1. *The Cauchy problem* (1-3) *has a global solution $\rho(\cdot,t)$ for $t \geq 0$. If the pinching condition* (1-2) *holds*, *then there exist two positive constants C_1 and C_2 such that*

$$C_1 \leq \rho(\theta,t) \leq C_2 \tag{3-1}$$

for $(\theta,t) \in S^1 \times [0,+\infty)$.

Proof. The local and global existence of solutions for the Cauchy problem (1-3) is a direct corollary of the classical theory for heat equations. Suppose (1-3) has a positive solution $\rho(\theta,t)$ on $S^1 \times [0,T)$ for some $T > 0$. Since $\partial\eta/\partial t = 2\partial^2\eta/\partial\theta^2$ and $\eta(\theta,0) = \eta_0(\theta)$ is a positive smooth function, we know that $\eta(\theta,t)$ is defined on $S^1 \times [0,+\infty)$ and is smooth. Furthermore, by the maximum principle, $n \leq \eta(\theta,t) \leq N$, where $n := \min\{\eta_0(\theta) \mid \theta \in S^1\}$, $N := \max\{\eta_0(\theta) \mid \theta \in S^1\}$. By the evolution equation of η and Writinger's inequality, we have

$$\frac{d}{dt}\int_0^{2\pi}\left(\frac{\partial^k \eta}{\partial\theta^k}\right)^2 d\theta = -4\int_0^{2\pi}\left(\frac{\partial^{k+1}\eta}{\partial\theta^{k+1}}\right)^2 d\theta \leq -16\int_0^{2\pi}\left(\frac{\partial^k \eta}{\partial\theta^k}\right)^2 d\theta,$$

where k is a positive integer. And thus

$$\int_0^{2\pi}\left(\frac{\partial^k \eta}{\partial\theta^k}\right)^2 d\theta \leq \left[\int_0^{2\pi}\left(\frac{\partial^k \eta_0}{\partial\theta^k}\right)^2 d\theta\right] e^{-16t}.$$

Hence, by Sobolev's inequality, one gets

$$\max_{\theta\in[0,2\pi]}\left|\frac{\partial^k \eta}{\partial\theta^k}\right| \leq C(k)e^{-8t}, \tag{3-2}$$

where

$$C(k)=\frac{1}{\sqrt{2\pi}}\left(\int_0^{2\pi}\left(\frac{\partial^k\eta_0}{\partial\theta^k}\right)^2 d\theta\right)^{\frac{1}{2}}+\sqrt{2\pi}\left(\int_0^{2\pi}\left(\frac{\partial^{k+1}\eta_0}{\partial\theta^{k+1}}\right)^2 d\theta\right)^{\frac{1}{2}}.$$

Notice that $\partial\rho/\partial t=\partial^2\eta/\partial\theta^2\le C(2)e^{-8t}$; i.e.,

$$\rho(\theta,t)\le\rho_0(\theta)+C(2)\tfrac{1}{8}(1-e^{-8t})\le M+\tfrac{1}{8}C(2),$$

where $M=\max\{\rho_0(\theta)\mid\theta\in S^1\}$. Letting $C_2=M+\frac{1}{8}C(2)$ gives us

$$\rho(\theta,t)\le C_2. \tag{3-3}$$

From (1-3), we get $\partial\rho/\partial t=\partial^2\eta/\partial\theta^2=\frac{1}{2}\partial\eta/\partial t$, which yields

$$\rho=\rho_0+\tfrac{1}{2}(\eta-\eta_0). \tag{3-4}$$

By the maximum principle, we know that $\eta(\theta,t)\ge\min\{\eta_0(\theta)\mid\theta\in S^1\}>0$. If $\eta_0(\theta)$ attains its minimum n at θ_n, then one has

$$\begin{aligned}\rho(\theta,t)&=\rho_0(\theta)+\tfrac{1}{2}(\eta(\theta,t)-\eta_0(\theta))\\&\ge\rho_0(\theta)+\tfrac{1}{2}(n-\rho_0(\theta)-\rho_0(\theta+\pi))\\&=\tfrac{1}{2}\big(\rho_0(\theta)+\rho_0(\theta_n)+\rho_0(\theta_n+\pi)-\rho_0(\theta+\pi)\big)\\&\ge\tfrac{1}{2}(3m-M)>0.\end{aligned}$$

Namely, there exists a positive constant $C_1=\frac{1}{2}(3m-M)$ such that

$$\rho(\theta,t)\ge C_1. \tag{3-5}$$

Combining (3-3) and (3-5), we complete the proof of (3-1). □

Corollary 3.2. *If the pinching condition* (1-2) *holds for the strictly convex initial curve $X_0(\varphi)$, then the problem* (1-1) *has a unique global solution $X(\varphi,t)$ on $S^1\times[0,+\infty)$ and $X(\cdot,t)$ is a strictly convex curve for all $t>0$.*

Lemma 3.3. *Under the condition of Corollary 3.2, the convex evolving curve converges to a constant-width curve smoothly.*

Proof. By the evolution equation of w and the closing condition of the evolving curve, we have almost the same estimate for w as that for η:

$$\max_{\theta\in[0,2\pi]}\left|\frac{\partial^k w}{\partial\theta^k}\right|\le C(X_0,k)e^{-8t}, \tag{3-6}$$

where $C(X_0,k)$ is a positive constant depending only on the initial data X_0 and k. By the Arzelà–Ascoli theorem, there exists a subsequence $\{w(\theta,t_i)\}$ convergent as t_i goes to infinity. Since $\lim_{t\to\infty}|\partial w/\partial\theta|=0$, $\lim_{t_i\to\infty}w(\theta,t_i)$ equals some constant.

Noticing that $\int_0^{2\pi} w(\theta,t)\,d\theta = 2L_0$, we obtain that $\lim_{t_i\to\infty} w(\theta,t_i) = L_0/\pi$. Since this equality holds for any convergent subsequence of $\{w(\theta,t)\}$, we can claim that $\{w(\theta,t)\}$ is convergent:

$$\lim_{t\to\infty} w(\theta,t) = \frac{L_0}{\pi}. \tag{3-7}$$

Similarly, we also have

$$\lim_{t\to\infty} \eta(\theta,t) = \frac{L_0}{\pi}. \tag{3-8}$$

From (3-4) it follows that

$$\lim_{t\to\infty} \rho(\theta,t) = \frac{L_0}{2\pi} + \frac{1}{2}(\rho_0(\theta) - \rho_0(\theta+\pi)). \tag{3-9}$$

Since $M < 3m$ (condition (1-2)), one gets

$$\lim_{t\to\infty} \rho(\theta,t) \geq m + \tfrac{1}{2}(m-M) > 0. \tag{3-10}$$

By (3-6) and (3-10), the limit of the evolving curve is convex and is of constant width. By (3-4) and (3-9), we have

$$\rho(\theta,t) - \lim_{t\to\infty} \rho(\theta,t) = \frac{1}{2}\left(\eta(\theta,t) - \frac{L_0}{\pi}\right).$$

Thus that the evolving curve converges smoothly is a corollary of (3-2). □

Lemma 3.4. *Under the condition of Corollary 3.2, the flow* (1-1) *keeps the perimeter of the evolving curve X and increases the enclosed area.*

Proof. Let $L(t)$ be the perimeter of the evolving curve $X(\cdot\,,t)$ and $A(t)$ the enclosed area. The variational formulae of $L(t)$ and $A(t)$ in [Gage 1986] give us

$$\frac{dL}{dt} = -\int_0^L \beta\kappa\,ds, \quad \frac{dA}{dt} = -\int_0^L \beta\,ds. \tag{3-11}$$

Under the flow (1-1), the perimeter evolves according to

$$\frac{dL}{dt} = -\int_0^{2\pi} (w(\theta,t) - \eta(\theta,t))\,d\theta = \int_0^{2\pi} \frac{\partial^2 w}{\partial\theta^2}(\theta,t)\,d\theta = 0, \tag{3-12}$$

which implies that the flow (1-1) keeps the perimeter of the evolving curve. By Gage's variational formulae (3-11), the enclosed area evolves according to

$$\begin{aligned}\frac{dA}{dt} &= -\int_0^{2\pi} (w(\theta,t)-\eta(\theta,t))\rho(\theta,t)\,d\theta \;=\; \int_0^{2\pi} \frac{\partial^2 w}{\partial\theta^2}(\theta,t)\rho(\theta,t)\,d\theta \\ &= \int_0^{\pi} \frac{\partial^2 w}{\partial\theta^2}(\theta,t)\rho(\theta,t)\,d\theta + \int_\pi^{2\pi} \frac{\partial^2 w}{\partial\theta^2}(\theta,t)\rho(\theta,t)\,d\theta \\ &= \int_0^{\pi} \frac{\partial^2 w}{\partial\theta^2}(\theta,t)\rho(\theta,t)\,d\theta + \int_0^{\pi} \frac{\partial^2 w}{\partial\xi^2}(\xi+\pi,t)\rho(\xi+\pi,t)\,d\xi \\ &= \int_0^{\pi} \frac{\partial^2 w}{\partial\theta^2}(\theta,t)(\rho(\theta,t)+\rho(\theta+\pi,t))\,d\theta \\ &= \int_0^{\pi} \frac{\partial^2 w}{\partial\theta^2}(\theta,t)\left(\frac{\partial^2 w}{\partial\theta^2}(\theta,t)+w(\theta,t)\right)d\theta \\ &= \int_0^{\pi} \left(\frac{\partial^2 w}{\partial\theta^2}(\theta,t)\right)^2 d\theta - \int_0^{\pi} \left(\frac{\partial w}{\partial\theta}(\theta,t)\right)^2 d\theta,\end{aligned}$$

where the fact that $w(\theta,t)$ is a periodic function with period π with respect to θ is used. Now, the Wirtinger inequality implies $dA/dt \geq 0$. Namely, flow (1-1) increases the area enclosed by the evolving curve. □

From the previous lemma, it follows that

$$\frac{d}{dt}(L^2 - 4\pi A) \leq 0,$$

which tells us that the isoperimetric deficit of the evolving curve is decreasing and thus the curve becomes more and more circular during the evolution process.

Generally speaking, the pinching condition (1-2) can not be omitted, because we have a lot of convex curves such that the right-hand side of (3-9) is negative for some θ. However, an example in the next section shows that the pinching inequality (1-2) is just a sufficient condition to guarantee the global existence of convex curve $X(\cdot\,,t)$. We do not know how to weaken this condition.

Next, we will follow the idea from [Lin and Tsai 2009] to study the geometric behavior of the flow (1-1) (using Fourier series). Now suppose that (1-1) has a global solution on $S^1 \times [0,+\infty)$ and that each evolving curve is strictly convex.

The Fourier expansion of the support function $p(\theta,t)$ of the evolving curve can be written as

$$p(\theta,t) = \frac{L_0}{2\pi} + \sum_{k=1}^{\infty}[a_k(t)\cos(k\theta) + b_k(t)\sin(k\theta)],$$

where θ is the tangential angle. By the definitions of width and radius of curvature, we have

$$\rho(\theta,t)=\frac{\partial^2 p}{\partial\theta^2}(\theta,t)+p(\theta,t)=\frac{L_0}{2\pi}+\sum_{k=1}^{\infty}[a_k(t)\cos(k\theta)+b_k(t)\sin(k\theta)](1-k^2),$$

$$w(\theta,t)=p(\theta,t)+p(\theta+\pi,t)=\frac{L_0}{\pi}+2\sum_{k=1}^{\infty}[a_{2k}(t)\cos(2k\theta)+b_{2k}(t)\sin(2k\theta)].$$

Since $\partial w/\partial t=2(\partial^2 w/\partial\theta^2)$, we have, by comparing the coefficients of both sides,

$$w(\theta,t)=\frac{L_0}{\pi}+2\sum_{k=1}^{\infty}[a_{2k}(0)\cos(2k\theta)+b_{2k}(0)\sin(2k\theta)]e^{-8k^2t}. \tag{3-13}$$

Therefore,

$$\begin{aligned}\eta(\theta,t)&=\frac{\partial^2 w}{\partial\theta^2}(\theta,t)+w(\theta,t)\\&=\frac{L_0}{\pi}+2\sum_{k=1}^{\infty}[a_{2k}(0)\cos(2k\theta)+b_{2k}(0)\sin(2k\theta)](1-4k^2)e^{-8k^2t},\end{aligned}$$

and thus

$$\begin{aligned}\rho(\theta,t)=\frac{L_0}{2\pi}+\sum_{k=1}^{\infty}&[a_{2k-1}(0)\cos((2k-1)\theta)+b_{2k-1}(0)\sin((2k-1)\theta)](4k-4k^2)\\&+\sum_{k=1}^{\infty}[a_{2k}(0)\cos(2k\theta)+b_{2k}(0)\sin(2k\theta)](1-4k^2)e^{-8k^2t}.\end{aligned}$$

As we know, $\partial p/\partial t=\eta-w=\partial^2 w/\partial\theta^2=\frac{1}{2}(\partial w/\partial t)$. Integrating this yields

$$\begin{aligned}p(\theta,t)=\frac{L_0}{2\pi}+\sum_{k=1}^{\infty}&[a_{2k-1}(0)\cos((2k-1)\theta)+b_{2k-1}(0)\sin((2k-1)\theta)]\\&+\sum_{k=1}^{\infty}[a_{2k}(0)\cos(2k\theta)+b_{2k}(0)\sin(2k\theta)]e^{-8k^2t}.\end{aligned} \tag{3-14}$$

The formula above is useful because we can use (2-4) and the definition of the support function to draw the graph of the evolving curve $X=(x,y)$ according to the following parametrization of convex curves (see [Green and Osher 1999]):

$$\begin{aligned}X&=\langle X,T\rangle T+\langle X,N\rangle N=\frac{\partial p}{\partial\theta}T-pN\\&=\left(p\sin\theta+\frac{\partial p}{\partial\theta}\cos\theta,\ -p\cos\theta+\frac{\partial p}{\partial\theta}\sin\theta\right).\end{aligned}$$

At the end of this section, we prove the last part of the Main Theorem.

Lemma 3.5. *If the initial curve X_0 is centrosymmetric, then the flow* (1-1) *has a global solution on $S^1 \times [0,\infty)$ and the limiting curve is a circle, and vice versa.*

Proof. If the initial curve X_0 is centrosymmetric and the symmetric center is the origin of the plane, then the support function and the radius of curvature of X_0 satisfy

$$p_0(\theta) = p_0(\theta+\pi), \quad \rho_0(\theta) = \rho_0(\theta+\pi).$$

By the evolution equations of ρ and η (see (2-5)), $\partial/\partial t\big(\rho(\theta,t) - \frac{1}{2}\eta(\theta,t)\big) = 0$. Thus we get

$$\rho(\theta,t) - \tfrac{1}{2}\eta(\theta,t) = \rho_0(\theta) - \tfrac{1}{2}\eta_0(\theta) = \rho_0(\theta) - \tfrac{1}{2}\big(\rho_0(\theta) + \rho_0(\theta+\pi)\big) = 0.$$

The maximum principle tells us that $0 < \frac{1}{2}n \le \rho(\theta,t) \le \frac{1}{2}N$ (n, N are defined in the proof of Lemma 3.1). Since $\frac{1}{2}\eta(\theta,t)$ converges to $L_0/2\pi$, $\rho(\theta,t)$ also tends to $L_0/2\pi$ as $t \to \infty$. Therefore, the limiting curve is a circle.

If the flow (1-1) has a global solution on $S^1 \times [0,\infty)$ and the limiting curve is a circle, then (3-14) implies that $a_{2k-1}(0) = b_{2k-1}(0) = 0$ for $k = 1, 2, \dots$. Therefore

$$p_0(\theta) = p_0(\theta+\pi).$$

Namely, X_0 is centrosymmetric with respect to the origin. □

Now, combining Corollary 3.2 and Lemmas 3.3–3.5, we complete the proof of the Main Theorem.

4. Examples

In this section, we will illustrate several examples. We have said that the pinching condition (1-2) cannot be omitted in the Main Theorem. In the following, a convex curve is given to show that (3-9) is negative for some θ. Define a function on S^1 by

$$p_0(\theta) = 10 - \cos(2\theta) + \cos(3\theta) + \tfrac{1}{8}\cos(5\theta) \quad \text{for } \theta \in [0, 2\pi].$$

We can construct a closed curve $X_0(\theta) = (x(\theta), y(\theta))$ by setting

$$x = p_0 \sin\theta + \frac{dp_0}{d\theta}\cos\theta, \quad y = -p_0\cos\theta + \frac{dp_0}{d\theta}\sin\theta.$$

The support function of $X_0(\theta)$ is $p_0(\theta)$, and we claim that X_0 is convex, since we can find that the minimum of the radius of curvature

$$\rho_0(\theta) = \frac{d^2 p_0}{d\theta^2} + p_0 = 10 + 3\cos(2\theta) - 8\cos(3\theta) - 3\cos(5\theta)$$

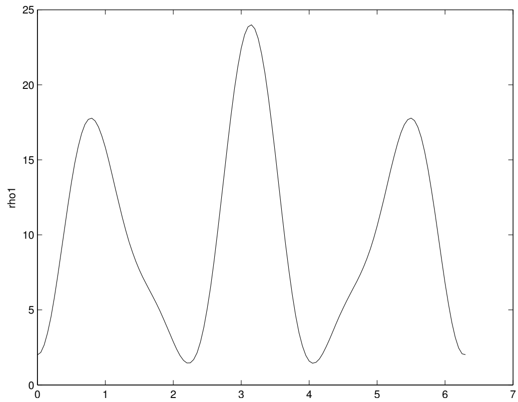

Figure 1. $\rho_0(\theta)$.

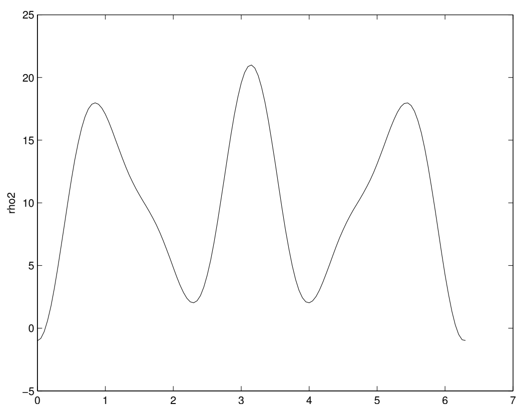

Figure 2. $\rho_\infty(\theta)$.

is $1.4409\ldots$, with the help of Matlab 7.8.0. However, the minimum of

$$\rho_\infty(\theta) := \frac{L_0}{2\pi} + \frac{1}{2}(\rho_0(\theta) - \rho_0(\theta+\pi)) = 10 - 8\cos(3\theta) - 3\cos(5\theta)$$

is -1. Figures 1 and 2 are the images of functions $\rho_0(\theta)$ and $\rho_\infty(\theta)$, respectively. A part of the "limiting curve" is given in Figure 3, in which singularities and self-intersections may occur near $x = 0$.

If we set the support function of a convex curve X_0 to be

$$p_0(\theta) = 19 + 2\cos(2\theta) + \cos(3\theta), \quad \text{for } \theta \in [0, 2\pi],$$

then the radius of curvature is

$$\rho_0(\theta) = 19 - 6\cos(2\theta) - 8\cos(3\theta),$$

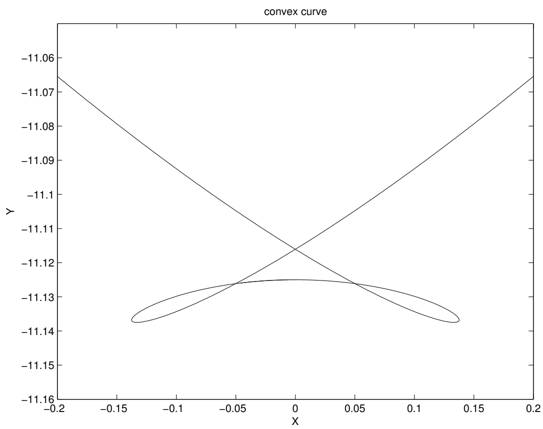

Figure 3. A portion of the limiting curve, with singularities and self-intersection near $x = 0$.

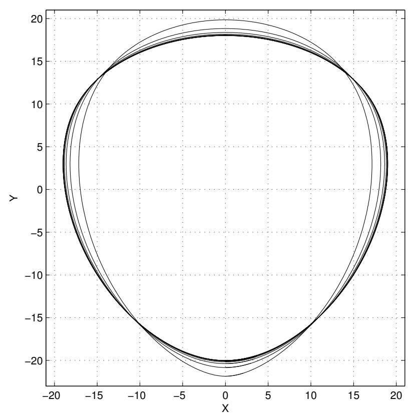

Figure 4. Convergence to a curve of constant width 38.

with minimum equal to 5 and maximum equal to $30.6366\ldots$, again using Matlab. Although this convex curve does not satisfy the pinching condition of the Main Theorem, numerical experiment shows that, under the flow (1-1), it keeps its convexity ($\rho_{\min}(t) \geq 5$, for every $t \in [0, \infty)$) and converges to a curve of constant width 38. Figure 4 describes the evolution process.

Our last example is a centrosymmetric convex curve X_0 with support function

$$p_0(\theta) = 15 + 3\cos(2\theta), \quad \text{for } \theta \in [0, 2\pi].$$

If X_0 evolves according to the flow (1-1) then the family of evolving curves converges to a circle. The evolution process is demonstrated in Figure 5.

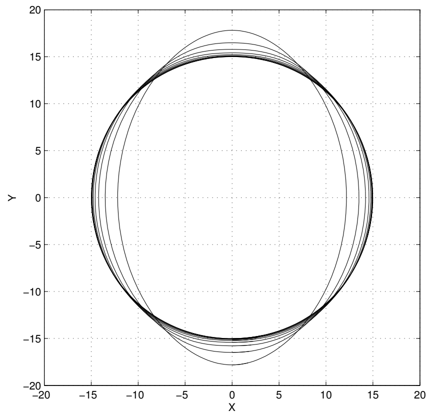

Figure 5. Convergence to a circle.

Acknowledgments

The authors would like to thank the referee for their careful reading of the manuscript and for giving us some corrections and useful suggestions.

References

[Andrews 1998] B. Andrews, "Evolving convex curves", *Calc. Var. Partial Differential Equations* **7**:4 (1998), 315–371. MR 99k:58038 Zbl 0931.53030

[Angenent 1991] S. Angenent, "On the formation of singularities in the curve shortening flow", *J. Differential Geom.* **33**:3 (1991), 601–633. MR 92c:58016 Zbl 0731.53002

[Cao 2003] F. Cao, *Geometric curve evolution and image processing*, Lecture Notes in Math. **1805**, Springer, Berlin, 2003. MR 2004b:53113 Zbl 01896083

[Chao et al. 2013] X.-L. Chao, X.-R. Ling, and X.-L. Wang, "On a planar area-preserving curvature flow", *Proc. Amer. Math. Soc.* **141**:5 (2013), 1783–1789. MR 3020863 Zbl 06146523

[Chou and Zhu 2001] K.-S. Chou and X.-P. Zhu, *The curve shortening problem*, CRC Press, Boca Raton, FL, 2001. MR 2003e:53088 Zbl 1061.53045

[Chow and Tsai 1996] B. Chow and D.-H. Tsai, "Geometric expansion of convex plane curves", *J. Differential Geom.* **44**:2 (1996), 312–330. MR 97m:58041 Zbl 0872.58052

[Gage 1984] M. E. Gage, "Curve shortening makes convex curves circular", *Invent. Math.* **76**:2 (1984), 357–364. MR 85i:52004 Zbl 0542.53004

[Gage 1986] M. E. Gage, "On an area-preserving evolution equation for plane curves", pp. 51–62 in *Nonlinear problems in geometry* (Mobile, AL, 1985), edited by D. M. DeTurck, Contemp. Math. **51**, Amer. Math. Soc., Providence, RI, 1986. MR 87g:53003 Zbl 0608.53002

[Gage and Hamilton 1986] M. E. Gage and R. S. Hamilton, "The heat equation shrinking convex plane curves", *J. Differential Geom.* **23**:1 (1986), 69–96. MR 87m:53003 Zbl 0621.53001

[Grayson 1987] M. A. Grayson, "The heat equation shrinks embedded plane curves to round points", *J. Differential Geom.* **26**:2 (1987), 285–314. MR 89b:53005 Zbl 0667.53001

[Green and Osher 1999] M. Green and S. Osher, "Steiner polynomials, Wulff flows, and some new isoperimetric inequalities for convex plane curves", *Asian J. Math.* **3**:3 (1999), 659–676. MR 2001j:53089 Zbl 0969.53040

[Groemer 1996] H. Groemer, *Geometric applications of Fourier series and spherical harmonics*, Encyclopedia Math. Appl. **61**, Cambridge University Press, 1996. MR 97j:52001 Zbl 0877.52002

[Gurtin 1993] M. E. Gurtin, *Thermomechanics of evolving phase boundaries in the plane*, Clarendon, New York, 1993. MR 97k:73001 Zbl 0787.73004

[Hsiung 1981] C. C. Hsiung, *A first course in differential geometry*, Wiley, New York, 1981. MR 83c:53001 Zbl 0458.53001

[Jiang and Pan 2008] L. Jiang and S. Pan, "On a non-local curve evolution problem in the plane", *Comm. Anal. Geom.* **16**:1 (2008), 1–26. MR 2009e:53083 Zbl 1151.35392

[Lin and Tsai 2009] Y.-C. Lin and D.-H. Tsai, "Evolving a convex closed curve to another one via a length-preserving linear flow", *J. Differential Equations* **247**:9 (2009), 2620–2636. MR 2010j:53135 Zbl 1187.35119

[Ma and Cheng 2014] L. Ma and L. Cheng, "A non-local area preserving curve flow", *Geom. Dedicata* **171** (2014), 231–247. arXiv 0907.1430

[Ma and Zhu 2012] L. Ma and A. Zhu, "On a length preserving curve flow", *Monatsh. Math.* **165**:1 (2012), 57–78. MR 2886123 Zbl 1235.35175

[Mao et al. 2013] Y. Mao, S. Pan, and Y. Wang, "An area-preserving flow for closed convex plane curves", *Internat. J. Math.* **24**:4 (2013), Art. ID #1350029. MR 3062969 Zbl 06180188

[Pan and Yang 2008] S. Pan and J. Yang, "On a non-local perimeter-preserving curve evolution problem for convex plane curves", *Manuscripta Math.* **127**:4 (2008), 469–484. MR 2010h:53099 Zbl 1169.35033

[Pan and Zhang 2010] S. Pan and H. Zhang, "On a curve expanding flow with a non-local term", *Comm. Contemp. Math.* **12**:5 (2010), 815–829. MR 2012c:53102 Zbl 1206.35136

[Schneider 1993] R. Schneider, *Convex bodies: the Brunn–Minkowski theory*, Encyclopedia Math. Appl. **44**, Cambridge University Press, 1993. MR 94d:52007 Zbl 0798.52001

[Urbas 1999] J. Urbas, "Convex curves moving homothetically by negative powers of their curvature", *Asian J. Math.* **3**:3 (1999), 635–658. MR 2001m:53119 Zbl 0970.53039

Received September 1, 2013. Revised December 22, 2013.

LAIYUAN GAO
DEPARTMENT OF MATHEMATICS
TONGJI UNIVERSITY
SHANGHAI, 200092
CHINA
5lygao@mail.tongji.edu.cn

SHENGLIANG PAN
DEPARTMENT OF MATHEMATICS
TONGJI UNIVERSITY
SHANGHAI, 200092
CHINA
slpan@tongji.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.272.147

HILBERT SERIES OF CERTAIN JET SCHEMES OF DETERMINANTAL VARIETIES

SUDHIR R. GHORPADE, BOYAN JONOV AND B. A. SETHURAMAN

We consider the affine variety $\mathcal{Z}_{2,2}^{m,n}$ of first-order jets over $\mathcal{Z}_2^{m,n}$, where $\mathcal{Z}_2^{m,n}$ is the classical determinantal variety given by the vanishing of all 2×2 minors of a generic $m \times n$ matrix. When $2 < m \leq n$, this jet scheme $\mathcal{Z}_{2,2}^{m,n}$ has two irreducible components: a trivial component, isomorphic to an affine space, and a nontrivial component that is the closure of the jets supported over the smooth locus of $\mathcal{Z}_2^{m,n}$. This second component is referred to as the *principal component* of $\mathcal{Z}_{2,2}^{m,n}$; it is, in fact, a cone and can also be regarded as a projective subvariety of $\mathbb{P}^{2mn-1}$. We prove that the degree of the principal component of $\mathcal{Z}_{2,2}^{m,n}$ is the square of the degree of $\mathcal{Z}_2^{m,n}$ and, more generally, the Hilbert series of the principal component of $\mathcal{Z}_{2,2}^{m,n}$ is the square of the Hilbert series of $\mathcal{Z}_2^{m,n}$. As an application, we compute the a-invariant of the principal component of $\mathcal{Z}_{2,2}^{m,n}$ and show that the principal component of $\mathcal{Z}_{2,2}^{m,n}$ is Gorenstein if and only if $m = n$.

1. Introduction

Let $\mathbb{F}$ be an algebraically closed field and m, n, r be integers with $1 \leq r \leq m \leq n$. Let $\mathcal{Z}_r^{m,n}$ denote the affine variety in $\mathbb{A}_{\mathbb{F}}^{mn}$ defined by the vanishing of all $r \times r$ minors of an $m \times n$ matrix whose entries are independent indeterminates over $\mathbb{F}$. Equivalently $\mathcal{Z}_r^{m,n}$ is the locus of $m \times n$ matrices over $\mathbb{F}$ of rank $< r$. This is a classical and well-studied object and a number of its properties are known. For example, we know that $\mathcal{Z}_r^{m,n}$ is irreducible, rational, arithmetically Cohen–Macaulay and projectively normal. Moreover the multiplicity of $\mathcal{Z}_r^{m,n}$ (at its vertex, since $\mathcal{Z}_r^{m,n}$ is evidently a cone) or, equivalently, the degree of the corresponding projective subvariety of $\mathbb{P}_{\mathbb{F}}^{mn-1}$ is given by the following elegant formula (see [Abhyankar 1988, Remarks 20.18 and 20.19] or [Ghorpade 1994, Corollary 6.2]; see also [Herzog and Trung 1992] for an alternative proof and [Arbarello et al. 1985, Chapter 2, §4] or [Ghorpade and Krattenthaler 2004, p. 352] for an alternative approach and a different formula):

MSC2010: primary 14M12; secondary 05E40.
Keywords: jet schemes, Hilbert series, determinantal varieties.

$$e(\mathscr{X}_r^{m,n}) = \det_{1\le i,j\le r-1}\left(\binom{m+n-i-j}{m-i}\right). \tag{1}$$

More generally, the Hilbert series of $\mathscr{X}_r^{m,n}$ (or, more precisely, of the corresponding projective subvariety of $\mathbb{P}_{\mathbb{F}}^{mn-1}$) is also known and is explicitly given by

$$\frac{\sum_{k\ge 0} h_k t^k}{(1-t)^d}, \tag{2}$$

where $d = (r-1)(m+n-r+1)$ is the dimension of $\mathscr{X}_r^{m,n}$ (as an affine variety), and the coefficients h_k are given by sums of binomial determinants as follows:

$$h_k = \sum_{k_1+\cdots+k_{r-1}=k} \det_{1\le i,j\le r-1}\left(\binom{m-i}{k_i}\binom{n-j}{k_i+i-j}\right).$$

For a proof of this formula, we refer to [Ghorpade 1996] (see also [Galligo 1985] and [Conca and Herzog 1994]). Using this, or otherwise (see [Svanes 1974]), it can be shown that $\mathscr{X}_r^{m,n}$ is Gorenstein if and only if $m = n$. Moreover one can also show that the a-invariant of the (homogeneous) coordinate ring of $\mathscr{X}_r^{m,n}$ (which, by definition, is the least degree of a generator of its graded canonical module) is $n(1-r)$; see, e.g., [Gräbe 1988] or [Ghorpade 1996, Theorem 4].

We now turn to jet schemes, which have been of much recent interest due in large part to Nash's suggestion [1995] that jet schemes should give information about singularities of the base; see, e.g., [Mustaţă 2001; 2002; Ein and Mustaţă 2009]. If $\mathscr{X}$ is a scheme of finite type over $\mathbb{F}$ and k a positive integer, then a $(k-1)$-jet on $\mathscr{X}$ is a morphism $\operatorname{Spec}\mathbb{F}[t]/(t^k) \to \mathscr{X}$. The set of $(k-1)$-jets on $\mathscr{X}$ forms a scheme of finite type over $\mathbb{F}$, denoted $\mathscr{J}_{k-1}(\mathscr{X})$ and called the $(k-1)$-th jet scheme of $\mathscr{X}$. A little more concretely, suppose $\mathscr{X}$ is the affine scheme $\operatorname{Spec} S/I$ defined by the ideal $I = \langle f_1, \dots, f_s\rangle$ in the polynomial ring $S = \mathbb{F}[X_1, \dots, X_N]$. Consider independent indeterminates t and $X_i^{(\ell)}$ ($i = 1, \dots, N$ and $\ell = 0, \dots, k-1$) over $\mathbb{F}$ and the corresponding polynomial ring $S^{(k)}$ in the Nk variables $X_i^{(\ell)}$. For each $j = 1, \dots, s$, the polynomial

$$f_j\big(X_1^{(0)} + tX_1^{(1)} + \cdots + t^{k-1}X_1^{(k-1)}, \dots, X_N^{(0)} + tX_N^{(1)} + \cdots + t^{k-1}X_N^{(k-1)}\big)$$

is of the form

$$f_j^{(0)} + tf_j^{(1)} + \cdots + t^{k-1}f_j^{(k-1)} \quad \text{modulo } \langle t^k\rangle$$

for unique $f_j^{(\ell)} \in S^{(k)}$ $(0 \le \ell < k)$. Then $\mathscr{J}_{k-1}(\mathscr{X})$ is the affine scheme $\operatorname{Spec} S^{(k)}/I'$, where I' is the ideal generated by all $f_j^{(\ell)}$, $1 \le j \le s$, $0 \le \ell < k$, (Often in the literature, authors conflate the algebraic set in $\mathbb{A}^{Nk}$ consisting of the zeros of the polynomials $f_j^{(\ell)}$ with $\mathscr{J}_{k-1}(\mathscr{X})$ itself. This is generally harmless, especially when considering topological properties such as components, since the points of this algebraic set correspond bijectively with the set of closed points of $\mathcal{J}_{k-1}(\mathcal{X})$ as $\mathbb{F}$ is algebraically closed, and the set of closed points of an affine scheme is dense in the scheme. See [Liu 2002, Chapter 2, Remark 3.49], for instance.)

When $\mathcal{X}$ is smooth of dimension d, the jet scheme $\mathcal{J}_{k-1}(\mathcal{X})$ is known to be smooth of dimension kd. In general, $\mathcal{J}_{k-1}(\mathcal{X})$ can have multiple irreducible components, and these include a principal component that corresponds to the closure of the set of jets supported over the smooth points of the base scheme $\mathcal{X}$. These components are usually quite complicated and interesting. In fact, very little seems to be known about the structure of these components and their numerical invariants such as multiplicities. For example, even when $\mathcal{X}$ is a monomial scheme such as the one given by $X_1X_2\cdots X_e = 0$, where $e \le N$, determining the irreducible components and the multiplicity of $\mathcal{J}_{k-1}(\mathcal{X})$ appears to require some effort; see, e.g., [Goward and Smith 2006] and [Yuen 2007b]. Irreducible components of jet schemes of toric surfaces are discussed in [Mourtada 2011], while the irreducibility of jet schemes of the commuting matrix pairs scheme is discussed in [Sethuraman and Šivic 2009]. In a more recent work [Bruschek et al. 2011], the Hilbert series of arc spaces (that are, in a sense, limits of k-th jet schemes as $k \to \infty$) of seemingly simple objects such as the double line $y^2 = 0$ are shown to have connections with the Rogers–Ramanujan identities.

Now determinantal varieties such as $\mathcal{X}_r^{m,n}$ above are natural examples of singular algebraic varieties, and it is not surprising that the study of their jet schemes has been of considerable interest. This was done first by Košir and Sethuraman [2005a; 2005b] (see also [Yuen 2007a]). To describe the related results, henceforth we fix positive integers r, k, m, n with $r \le m \le n$, and let $\mathcal{X}_{r,k}^{m,n}$ denote the $(k-1)$-th jet scheme on $\mathcal{X}_r^{m,n}$. It was shown in [Košir and Sethuraman 2005a] that $\mathcal{X}_{r,k}^{m,n}$ is irreducible of codimension $k(n-m+1)$ when $r = m$, and if $r < m$, then it can have $\ge 1 + \lfloor k/2 \rfloor$ irreducible components with equality when $r = 2$ or $k = 2$. A more unified result was obtained in [Docampo 2013], showing that $\mathcal{X}_{r,k}^{m,n}$ has exactly $k + 1 - \lceil k/r \rceil$ irreducible components. At any rate, the best understood case with multiple components is $\mathcal{X}_{2,2}^{m,n}$, where $2 < m \le n$. In this case $\mathcal{X}_{2,2}^{m,n} = Z_0 \cup Z_1$, where Z_1 is isomorphic to $\mathbb{A}^{mn}$ while Z_0 is the principal component which is the closure of the jets supported over the smooth points of the base variety $\mathcal{X}_2^{m,n}$. Here it will be convenient to consider $2mn$ indeterminates, denoted $x_{i,j}, y_{i,j}$ for $1 \le i \le m$, $1 \le j \le n$, and the corresponding polynomial ring $R = \mathbb{F}[x_{i,j}, y_{i,j} : 1 \le i \le m,\ 1 \le j \le n]$. Also let $\mathcal{I} = \mathcal{I}_{2,2}^{m,n}$ and $\mathcal{I}_0$ denote, respectively, the ideals of R corresponding to the jet scheme $\mathcal{X}_{2,2}^{m,n}$ and its principal component Z_0. In [Košir and Sethuraman 2005b], it was shown that both $\mathcal{I}$ and $\mathcal{I}_0$ are homogeneous radical ideals of R (so that $\mathcal{I}_0$ is prime), and moreover their Gröbner bases were explicitly determined. The leading term ideal $\mathrm{LT}(\mathcal{I}_0)$ of $\mathcal{I}_0$ with respect to this Gröbner basis is generated by squarefree monomials and hence $R/\mathrm{LT}(\mathcal{I}_0)$ is the Stanley–Reisner ring of a simplicial complex Δ_0. Jonov [2011] subsequently studied this simplicial complex. He showed that Δ_0 is shellable and thus deduced that $R/\mathcal{I}_0$ is Cohen–Macaulay. (This last result was independently obtained in [Smith and Weyman 2007] as well, using a geometric technique for computing syzygies.) Jonov also found a formula for the multiplicity of $R/\mathcal{I}_0$, namely,

$$e(R/\mathcal{I}_0) = \sum_{\substack{i=1 \\ (i,j)\neq(m,n)}}^{m} \sum_{j=1}^{n} \binom{m+n-i-j}{m-i} \det \begin{pmatrix} \binom{i+n-2}{i-1} & \binom{m+j-2}{m-1} \\ \binom{i+n-3}{i-2} & \binom{m+j-3}{m-2} \end{pmatrix}. \tag{3}$$

Equation (3) above is the starting point of the present paper. We first show that the right side of this equation simplifies remarkably to yield the pretty result

$$e(R/\mathcal{I}_0) = \binom{m+n-2}{m-1}^2 = e(\mathcal{Z}_2^{m,n})^2.$$

(this was already alluded to in [Jonov 2011, Remark 2.8]). Next we proceed to determine the Hilbert series of $R/\mathcal{I}_0$ or of the principal component Z_0. We use the well-known connections between the Hilbert series of $R/\mathcal{I}_0$, that of $R/\mathrm{LT}(\mathcal{I}_0)$ and the shelling of the facets of the simplicial complex Δ_0 obtained in [Jonov 2011]. With some effort we are led to an initial formula for the Hilbert series of $R/\mathcal{I}_0$, which is enormously complicated and involves multiple sums of products of binomials in the same vein as the right side of (3). But we persist with the combinatorics and are eventually rewarded with the main result of this paper. Namely, just like the multiplicity, the Hilbert series of $R/\mathcal{I}_0$ is precisely the square of the Hilbert series of the base determinantal variety $\mathcal{Z}_2^{m,n}$. As a corollary of this, we are able to determine the a-invariant of $R/\mathcal{I}_0$ and the Hilbert series of its graded canonical module. Moreover we show that, as in the case of classical determinantal varieties, Z_0 is Gorenstein if and only if $m = n$.

The proofs given here are completely elementary but highly combinatorial and rather intricate. Heuristically it appears to us that up to some flat deformation (such as the Gröbner deformation of $\mathcal{I}_0$ to $\mathrm{LT}(\mathcal{I}_0)$, which preserves the Hilbert series), the coordinate ring of the principal component (suitably deformed) should look like the tensor product of the coordinate ring of the base (similarly deformed) with itself. (This would reflect the fact that, at the smooth points, the base variety locally looks like its tangent space.) It would follow then that the Hilbert series of the principal component is the square of that of $\mathcal{Z}_2^{m,n}$. We emphasize that this is only heuristics (with all of its ever-present dangers); nevertheless we suspect that analogous results relating the Hilbert series of the principal component to that of the base scheme should hold more generally for all $\mathcal{Z}_{r,k}^{m,n}$, and possibly also for jet schemes over a wider class of affine base schemes. We do not know how to prove this, and leave it open for investigation.

2. Binomials and lattice paths

In this section we collect some preliminaries concerning binomial coefficients, alterations of summations, and lattice paths. These will be useful in the sequel.

2.1. ***Binomials.*** To begin with, let us recall that the binomial coefficient $\binom{s}{a}$ is defined for any integer parameters s, a (and with the standard convention that the empty product is taken as 1) as follows:

$$\binom{s}{a} = \begin{cases} \dfrac{s(s-1)\cdots(s-a+1)}{a!} & \text{if } a \geq 0, \\ 0 & \text{if } a < 0. \end{cases}$$

In fact, this definition makes sense not only for any $s \in \mathbb{Z}$ but also for s in any overring of $\mathbb{Z}$ and in particular, s can be an indeterminate over $\mathbb{Q}$ in which case $\binom{s}{a}$ is a polynomial in s of degree a, provided $a \geq 0$. Now let $s, a \in \mathbb{Z}$. Note that

$$\binom{s}{a} = 0 \iff \text{either } a < 0 \text{ or } a > s \geq 0. \tag{4}$$

One has to be careful with the validity of some of the familiar identities; for example,

$$\binom{s}{a} = \binom{s}{s-a} \iff \text{either } s \geq 0 \text{ or } s < a < 0, \tag{5}$$

whereas some standard identities such as the Pascal triangle identity or its alternative equivalent version below are valid for arbitrary integer parameters:

$$\binom{s}{a-1} + \binom{s}{a} = \binom{s+1}{a} \quad \text{and} \quad \binom{s+a}{a} + \binom{s+a}{a+1} = \binom{s+a+1}{a+1}. \tag{6}$$

The equivalence of the two identities above follows from the simple fact below, which is also valid for arbitrary integer parameters:

$$\binom{s+a}{a} = (-1)^a \binom{-s-1}{a}, \quad \text{that is,} \quad \binom{s}{a} = (-1)^a \binom{a-s-1}{a}. \tag{7}$$

We now record some basic facts, which are often used in later sections. Proofs are easy and are briefly outlined for the sake of completeness.

Lemma 1. *For any $e, s, t \in \mathbb{Z}$ with $s \leq t$, we have*

$$\sum_{s<d\leq t} \binom{d}{e} = \binom{t+1}{e+1} - \binom{s+1}{e+1}.$$

Proof. Induct on $t - s$, using the first identity in (6) to rewrite $\binom{t+1}{e+1}$. □

The following result is a version of the so-called Chu–Vandermonde identity.

Lemma 2. *For any $s, t, \alpha, \beta \in \mathbb{Z}$, we have*

$$\sum_{j\in\mathbb{Z}} \binom{s}{\alpha+j}\binom{t}{\beta-j} = \binom{s+t}{\alpha+\beta} \tag{8}$$

and

$$\sum_{j\in\mathbb{Z}} \binom{s+\alpha+j}{\alpha+j}\binom{t+\beta-j}{\beta-j} = \binom{s+t+\alpha+\beta+1}{\alpha+\beta}, \tag{9}$$

where, *in view of* (4), *the summation on the left in* (8) *as well as in* (9) *is essentially finite in the sense that all except finitely many summands are zero.*

Proof. Let X be an indeterminate over $\mathbb{Q}$. Use the binomial theorem, namely,

$$(1+X)^d = \sum_{i=0}^{\infty} \binom{d}{i} X^i,$$

which is valid in the formal power series ring $\mathbb{Q}[[X]]$ for any $d \in \mathbb{Z}$, and compare the coefficients of $X^{\alpha+\beta}$ on the two sides of the identity $(1+X)^s(1+X)^t = (1+X)^{s+t}$ to obtain (8). Now (8) and (7) imply (9). $\square$

2.2. ***Alterations of summations.*** As in (8) and (9) above, we will often deal with summations that are *essentially finite*, by which we mean that the parameters in the sum range over an infinite set, but the summand is zero for all except finitely many values of parameters, and so the summation is, in fact, finite. It is, however, very useful that the parameters range freely over a seemingly infinite set so that useful alterations such as the ones listed below can be readily made. These are too obvious to be stated as lemmas and proved formally. But for ease of reference, we record below some elementary transformations of essentially finite summations. In what follows, $f : \mathbb{Z}^2 \to \mathbb{Q}$ will denote a rational-valued function of two integer parameters with the property that the *support* of f, namely, the set $\{(s_1, s_2) \in \mathbb{Z}^2 : f(s_1, s_2) \neq 0\}$ is finite or more generally, it is *diagonally finite*, that is, for each $k \in \mathbb{Z}$, the set $\{(s_1, s_2) \in \mathbb{Z}^2 : s_1 + s_2 = k$ and $f(s_1, s_2) \neq 0\}$ is finite. In this case, for any $\nu \in \mathbb{Z}$ and any $\alpha, \beta \in \mathbb{Z}$ such that $\alpha + \beta = \nu$, we have

$$\sum_{s_1+s_2=k-\nu} f(s_1, s_2) = \sum_{t_1+t_2=k} f(t_1-\alpha, t_2-\beta), \tag{10}$$

where writing $s_1 + s_2 = k - \nu$ below the first summation indicates that the sum is over all $(s_1, s_2) \in \mathbb{Z}^2$ satisfying $s_1 + s_2 = k - \nu$. A similar meaning applies for the second summation and in fact, for all such summations appearing in the sequel. Since the "diagonal condition" $t_1 + t_2 = k$ is symmetric, we also have

$$\sum_{t_1+t_2=k} f(t_1, t_2) = \sum_{t_1+t_2=k} f(t_2, t_1). \tag{11}$$

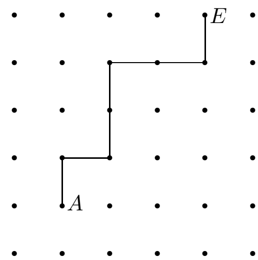

Figure 1. A lattice path from $A = (1, 1)$ to $E = (4, 5)$.

Thus, for example, using (10) and (11), we find

$$\sum_{t_1+t_2=k} f(t_1, t_2) = \sum_{t_1+t_2=k} f(t_2+1, t_1-1) = \sum_{t_1+t_2=k} f(t_1+1, t_2-1).$$

2.3. ***Lattice paths.*** Let $A = (a, a')$ and $E = (e, e')$ be points in the integer lattice $\mathbb{Z}^2$. By a *lattice path* from A to E we mean a finite sequence $L = (P_0, P_1, \dots, P_t)$ of points in $\mathbb{Z}^2$ with $P_0 = A$, $P_t = E$ and

$$P_i - P_{i-1} = (1, 0) \text{ or } (0, 1) \quad \text{for } i = 1, \dots, t.$$

The lattice path L can and will be identified with its point set $\{P_j : 0 \le j \le t\}$; indeed L is obtained by simply arranging the elements of this set in a lexicographic order. The point $A = P_0$ is called the *initial point* of L while $E = P_t$ is called the *end point* of L. We say that a point P_j is a *NE-turn* of the lattice path L if $0 < j < t$ and $P_j - P_{j-1} = (0, 1)$ while $P_{j+1} - P_j = (1, 0)$. Note that a lattice path is also determined by its NE turns.

In more intuitive terms, a lattice path consists of vertical or horizontal steps of length 1, and a NE-turn is simply a northeast turn. For example, a lattice path from $A = (1, 1)$ to $E = (4, 5)$ may be depicted as in Figure 1, and it may be noted that the points $(1, 2)$ and $(2, 4)$ are its NE turns.

If we let $\mathcal{P}(A \to E)$ denote the set of lattice paths from $A = (a, a')$ to $E = (e, e')$ and, for any $k \in \mathbb{Z}$, let $\mathcal{P}_k(A \to E)$ denote the subset of $\mathcal{P}(A \to E)$ consisting of lattice paths with exactly k NE turns, then it is easily seen that

$$\begin{aligned} |\mathcal{P}(A \to E)| &= \binom{e-a+e'-a'}{e-a}, \\ |\mathcal{P}_k(A \to E)| &= \binom{e-a}{k}\binom{e'-a'}{k}, \end{aligned} \tag{12}$$

where as usual, for a finite set $\mathcal{P}$, we denote by $|\mathcal{P}|$ the cardinality of $\mathcal{P}$. Given any two d-tuples $\mathcal{A} = (A_1, \dots, A_d)$ and $\mathcal{E} = (E_1, \dots, E_d)$ of points in $\mathbb{Z}^2$, by a *lattice path* from $\mathcal{A}$ to $\mathcal{E}$ we mean a d-tuple $\mathcal{L} = (L_1, \dots, L_d)$, where L_r is a lattice path from A_r to E_r, for $1 \le r \le d$. We call $\mathcal{L}$ to be *nonintersecting* if no two of the paths $L_1, \dots, L_d$ have a point in common. We say that $\mathscr{L}$ has k NE turns if the total number of NE turns in the d paths $L_1, \dots, L_d$ is k. The set of nonintersecting lattice paths from $\mathscr{A} = (A_1, \dots, A_d)$ to $\mathscr{E} = (E_1, \dots, E_d)$ will be denoted by $\mathscr{P}(A_1 \to E_1, \dots, A_d \to E_d)$ or simply by $\mathscr{P}(\mathscr{A} \to \mathscr{E})$, and its subset consisting of nonintersecting lattice paths with exactly k NE turns will be denoted by $\mathscr{P}_k(A_1 \to E_1, \dots, A_d \to E_d)$ or simply by $\mathscr{P}_k(\mathscr{A} \to \mathscr{E})$.

Proposition 3. *Let d be a positive integer and let $A_r = (a_r, a'_r)$ and $E_r = (e_r, e'_r)$, $r = 1, \dots, d$, be points in $\mathbb{Z}^2$. Also let $\mathscr{A} = (A_1, \dots, A_d)$ and $\mathscr{E} = (E_1, \dots, E_d)$.*

(i) *Suppose*

$$a_1 \le \dots \le a_d, \quad e_1 \le \dots \le e_d \quad \text{and} \quad a'_1 \ge \dots \ge a'_d, \quad e'_1 \ge \dots \ge e'_d.$$

Then the number of nonintersecting lattice paths from $\mathscr{A}$ to $\mathscr{E}$ is equal to

$$\det\left(\binom{e_j - a_i + e'_j - a'_i}{e_j - a_i}_{1 \le i, j \le d}\right) \tag{13}$$

(ii) *Let $k \in \mathbb{Z}$ and suppose*

$$a_1 \le \dots \le a_d, \quad e_1 < \dots < e_d \quad \text{and} \quad a'_1 > \dots > a'_d, \quad e'_1 \ge \dots \ge e'_d.$$

Then the number of nonintersecting lattice paths from $\mathscr{A}$ to $\mathscr{E}$ with exactly k NE turns is equal to

$$\sum_{k_1 + \dots + k_d = k} \det\left(\binom{e_j - a_i + i - j}{k_i + i - j}\binom{e'_j - a'_i - i + j}{k_i}_{1 \le i, j \le d}\right) \tag{14}$$

Part (i) of the above proposition is due to Gessel and Viennot [1985, Theorem 1], although some of the ideas can be traced back to Chaundy [1932], Karlin and McGregor [1959], and Lindström [1973]. The statement here is a little more general than that of [Gessel and Viennot 1985], and a proof can be found, for example, in [Ghorpade 2001, §3] or [Krattenthaler 1995b, §2.2]. Part (ii) was proved independently by Modak [1992], Krattenthaler [1995a] and Kulkarni [1996] (see also [Ghorpade 1996]), although the hypothesis in [Modak 1992] and [Kulkarni 1996] on the coordinates of the initial and the end points is slightly more restrictive than in (ii) above where we follow [Krattenthaler 1995a, Theorem 1]. The following consequence is frequently used in Section 4.

Corollary 4. *For any $a, b, c, d, s \in \mathbb{Z}$ with $a < c$ and $b \ge d$, the cardinality of $\mathscr{P}_s((1, 2) \to (a, b), (1, 1) \to (c, d))$ is given by*

$$\sum_{s_1 + s_2 = s} \binom{a-1}{s_1}\binom{b-2}{s_1}\binom{c-1}{s_2}\binom{d-1}{s_2} - \binom{a}{s_2+1}\binom{b-2}{s_2}\binom{c-2}{s_1-1}\binom{d-1}{s_1}.$$

Proof. This is just a special case of part (ii) of Proposition 3. $\square$

3. Multiplicity

As in the Introduction, we fix in the remainder of this paper an algebraically closed field $\mathbb{F}$ and integers m, n with $2 < m \leq n$. Also let $x_{i,j}$, $y_{i,j}$, $1 \leq i \leq m$, $1 \leq j \leq n$, be independent indeterminates over $\mathbb{F}$. Denote by V_x the set

$$\{x_{i,j} : 1 \leq i \leq m \text{ and } 1 \leq j \leq n\}$$

of the "x-variables", and by V_y a similar set of the "y-variables". Let $V = V_x \cup V_y$ and let $R = \mathbb{F}[V]$ be the corresponding polynomial ring in $2mn$ variables; also let $R_x = \mathbb{F}[V_x]$ and $R_y = \mathbb{F}[V_y]$ be the corresponding polynomial rings in mn variables. By the *support* of a monomial F in R, denoted $\operatorname{supp}(F)$, we mean the subset of V consisting of the variables appearing in F. Clearly a monomial F in R can be uniquely written as

$$F = F_x F_y, \quad \text{where } F_x, F_y \text{ are monomials with } F_x \in R_x \text{ and } F_y \in R_y, \tag{15}$$

and moreover F is squarefree if and only if both F_x and F_y are squarefree. Note that squarefree monomials can be identified with their supports, and in particular, faces of a simplicial complex Δ with vertex set V can be viewed as squarefree monomials in R. With this in view, we may not distinguish between a squarefree monomial and its support, and we may sometimes write $x_{i,j} \in G$ rather than $x_{i,j} \mid G$ when G is a squarefree monomial in R and $x_{i,j}$ is a variable appearing in it. A monomial G in R_x will be called a *lattice path monomial* in R_x if there is a positive integer t and variables $x_{i_1,j_1}, \ldots, x_{i_t,j_t}$ in V_x such that

$$G = \prod_{s=1}^{t} x_{i_s,j_s} \quad \text{with } (i_s - i_{s-1},\, j_s - j_{s-1}) = (1,0) \text{ or } (0,1) \text{ for } 1 < s \leq t. \tag{16}$$

In this case G is called a lattice path monomial from x_{i_1,j_1} to x_{i_t,j_t}, and we will refer to x_{i_1,j_1} as the *leader* of G and denote it by $\mu(G)$. Note that $\mu(G) = x_{i_1,j_1}$ depends only on G (and not on the given ordering of the variables appearing in it) since (i_1, j_1) is lexicographically the least among the pairs (i, j) for which $x_{i,j} \in \operatorname{supp}(G)$. A variable x_{i_s,j_s} in $\operatorname{supp}(G)$ will be called an *ES-turn* of G if $1 < s < t$, $i_s = i_{s-1}$, and $j_s = j_{s+1}$. Analogously a variable x_{i_s,j_s} in $\operatorname{supp}(G)$ will be called a *SE-turn* of G if $1 < s < t$, $j_s = j_{s-1}$, and $i_s = i_{s+1}$. Moreover we will call a variable x_{i_s,j_s} in $\operatorname{supp}(G)$ the *midpoint of a segment* in G if $1 < s < t$ and either $i_{s-1} = i_s = i_{s+1}$ (horizontal segment) or $j_{s-1} = j_s = j_{s+1}$ (vertical segment). It may be noted that a variable x_{i_s,j_s} with $1 < s < t$ is either an ES-turn or a SE-turn or the midpoint of a segment in G.

Evidently lattice path monomials in R_x correspond to lattice paths in the sense of Section 2.3 if we turn the $m \times n$ rectangular matrix $(x_{i,j})$ left by 90° and identify the variable $x_{i,j}$ with the lattice point (i, j). In this way leaders correspond to initial

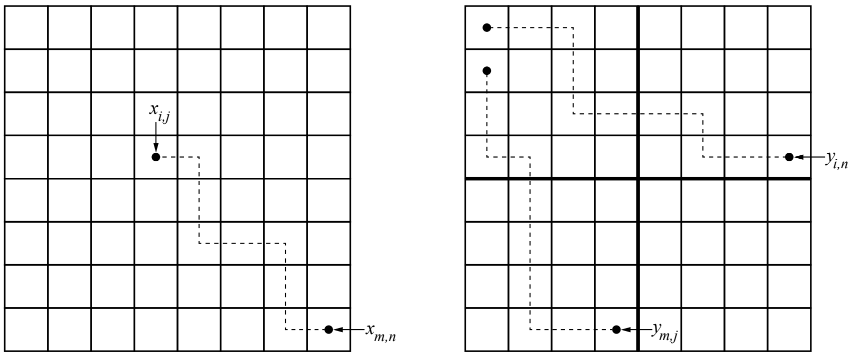

Figure 2. Lattice path monomials F_x and $F_y = F_y^{\mathsf{U}} F_y^{\mathsf{L}}$ in Proposition 5.

points while ES turns correspond to NE turns. Lattice path monomials in R_y together with their leaders, ES turns, SE turns, and midpoints of segments are similarly defined (and similarly identified with lattice paths in the sense of Section 2.3).

We have noted in the introduction that a Gröbner basis (with respect to reverse lexicographic order on monomials with the $2mn$ variables arranged suitably) of the ideal $\mathscr{I}$ of the variety $\mathscr{Z}_{2,2}^{m,n}$ of first-order jets over $\mathscr{Z}_2^{m,n}$, as well as of the ideal $\mathscr{I}_0$ of the principal component Z_0 of $\mathscr{Z}_{2,2}^{m,n}$, was determined in [Košir and Sethuraman 2005b]. As a consequence, one can write down the generators of the leading term ideal of $\mathscr{I}_0$ (see [Jonov 2011, Proposition 1.1]), say $\mathrm{LT}(\mathscr{I}_0)$, and deduce that $R/\mathrm{LT}(\mathscr{I}_0)$ is the Stanley–Reisner ring of a simplicial complex Δ_0 with V as its set of vertices. A precise description of the facets of Δ_0 was given by Jonov [2011, §2], and we recall it below.

Proposition 5. *A squarefree monomial F, decomposed as in* (15) *above, is a facet of Δ_0 if and only if there is a unique $(i, j) \in \mathbb{Z}^2$, with $1 \le i \le m$, $1 \le j \le n$, such that $(i, j) \ne (m, n)$ and F_x is a lattice path monomial from $x_{i,j}$ to $x_{m,n}$, whereas $F_y = F_y^{\mathsf{U}} F_y^{\mathsf{L}}$, where F_y^{U} is a lattice path monomial from $y_{1,1}$ to $y_{i,n}$, F_y^{L} is a lattice path monomial from $y_{2,1}$ to $y_{m,j}$, and the supports of F_y^{U} and F_y^{L} are disjoint.*

The lattice path monomials F_x and $F_y = F_y^{\mathsf{U}} F_y^{\mathsf{L}}$ are illustrated in Figure 2 by the corresponding "paths" in rectangular matrices.

Using Proposition 5 together with the first identity in (12) and part (i) of Proposition 3, Jonov showed that the simplicial complex Δ_0 is pure (i.e., all its facets have the same dimension) and deduced the dimension and the formula stated in the introduction for the multiplicity of the coordinate ring $R/\mathscr{I}_0$ of Z_0.

Corollary 6. *The* (*Krull*) *dimension of $R/\mathscr{I}_0$ is* $2(m+n-1)$ *and the multiplicity of $R/\mathscr{I}_0$ is given by* (3).

Now here is the pretty result about the multiplicity that was alluded to in the introduction, namely, that the formula (3) admits a remarkable simplification.

Theorem 7. *The multiplicity of* $R/\mathcal{I}_0$ *is given by*

$$e(R/\mathcal{I}_0) = \binom{m+n-2}{m-1}^2. \tag{17}$$

Proof. For $1 \le i \le m$ and $1 \le j \le n$, let $\Delta_{i,j}$ denote the 2×2 determinant in (3). Observe that if $(i, j) = (m, n)$, then $\Delta_{i,j} = 0$. Thus, by expanding this determinant and rearranging the summands, we can write

$$e(R/\mathcal{I}_0) = \sum_{i=1}^{m} \binom{i+n-2}{i-1} \sum_{j=1}^{n} S_{i,j} - \sum_{i=1}^{m} \binom{i+n-3}{i-2} \sum_{j=1}^{n} T_{i,j},$$

where, for $1 \le i \le m$ and $1 \le j \le n$, we have put

$$S_{i,j} = \binom{m+n-i-j}{m-i}\binom{m+j-3}{m-2} \quad \text{and} \quad T_{i,j} = \binom{m+n-i-j}{m-i}\binom{m+j-2}{m-1}.$$

Rewriting $S_{i,j}$ using (5) and then noting that the resulting product is zero if $j < 1$ or $j > n$, thanks to (4), we see from Equation (9) in Lemma 2 that

$$\sum_{j=1}^{n} S_{i,j} = \sum_{j} \binom{m+n-i-j}{n-j}\binom{m+j-3}{j-1} = \binom{2m+n-i-2}{n-1},$$

for each $i = 1, \dots, m$. In a similar manner,

$$\sum_{j=1}^{n} T_{i,j} = \sum_{j} \binom{m+n-i-j}{n-j}\binom{m+j-2}{j-1} = \binom{2m+n-i-1}{n-1},$$

for each $i = 1, \dots, m$. It follows that $e(R/\mathcal{I}_0)$ is given by the telescoping sum

$$e(R/\mathcal{I}_0) = \sum_{i=1}^{m} (a_i - a_{i-1}), \quad \text{where } a_i := \binom{i+n-2}{i-1}\binom{2m+n-i-2}{n-1},$$

for $0 \le i \le m$. Since $a_0 = 0$ and $a_m = \binom{m+n-2}{m-1}^2$, we obtain the desired result. $\square$

It may be noted that in view of (1) and (17), the multiplicity of the principal component Z_0 is precisely the square of the multiplicity of the base variety $\mathcal{Z}_2^{m,n}$.

4. Hilbert series

Let us begin by recalling that a *shelling* of a pure simplicial complex Δ is a linear ordering $F_1, \dots, F_e$ of its facets such that for all positive integers i, j, with $j < i \le e$, there exist some $v \in F_i \setminus F_j$ and some positive integer $k < i$ such that $F_i \setminus F_k = \{v\}$. Given such a shelling and any $t \in \{1, \dots, e\}$, we let

$$c(F_t) = \{v \in F_t : \text{there exists } s < t \text{ such that } F_t \setminus F_s = \{v\}\}.$$

Elements of $c(F_t)$ will be referred to as the *corners* of F_t. It may be noted that $c(F_t)$ is nonempty if and only if $t > 1$. Recall also that a simplicial complex Δ is said to be *shellable* if it is pure and it has a shelling. The following result is well known (see [Bruns and Conca 2003, Theorem 6.3]).

Proposition 8. *Let Δ be a shellable simplicial complex and let R_Δ denote its Stanley–Reisner ring. Then*:

(i) *R_Δ is Cohen–Macaulay and its* (*Krull*) *dimension* $\dim R_\Delta$ *is* $1+\dim\Delta$.

(ii) *Suppose $d = \dim R_\Delta$ and $F_1, \dots, F_e$ is a shelling of Δ. Then the Hilbert series of R_Δ is given by*

$$\frac{\sum_{j\geq 0} h_j z^j}{(1-z)^d}, \quad \textit{where } h_j = \bigl|\{t \in \{1,\dots,e\} : |c(F_t)| = j\}\bigr| \textit{ for } j \geq 0.$$

Jonov [2011] showed that the simplicial complex Δ_0 mentioned in the previous section is shellable and concluded using part (i) of Proposition 8 that the coordinate ring of $R/\mathcal{I}_0$ of the principal component Z_0 of $\mathfrak{X}_{2,2}^{m,n}$ is Cohen–Macaulay. We shall now proceed to use part (ii) of Proposition 8 to determine the Hilbert series of $R/\mathcal{I}_0$. We will use the notation and terminology introduced at the beginning of Section 3. Further we introduce the following "antilexicographic" linear order on V_x, that is, on the x-variables. For any $x_{a,b}, x_{c,d} \in V_x$, define

$$x_{a,b} \prec x_{c,d} \iff \text{either } a > c \quad \text{or} \quad a = c \text{ and } b > d.$$

Given a lattice path monomial G as in (16), the *spread* of G, denoted $\mathrm{sp}(G)$, is the set of variables that are on or below the corresponding lattice path; more precisely,

$$\mathrm{sp}(G) = \{x_{a,b} : i_s \leq a \leq m \text{ and } 1 \leq b \leq j_s \text{ for some } s = 1, \dots, t\}.$$

The notion of spread is defined for lattice path monomials in R_y in exactly the same manner. It may be observed that if G, H are lattice path monomials (both in R_x or both in R_y), then the condition $\mathrm{sp}(G) \subseteq \mathrm{sp}(H)$ means, roughly speaking, that H is to the right of G; moreover, if $\mu(G) = \mu(H)$ and $\mathrm{sp}(G) = \mathrm{sp}(H)$, then we must have $G = H$.

Notice that the lattice path monomials $F_y^{\sf U}$ and $F_y^{\sf L}$ of Proposition 5 have the property that $\mathrm{sp}(F_y^{\sf L}) \subseteq \mathrm{sp}(F_y^{\sf U})$.

Following [Jonov 2011], we now define a partial order on the facets of Δ_0.

Definition 9. For any facets P, Q of Δ_0 with decompositions $P = P_x P_y^{\sf U} P_y^{\sf L}$ and $Q = Q_x Q_y^{\sf U} Q_y^{\sf L}$ as in Proposition 5, define $P < Q$ if one of the following four conditions hold: (i) $\mu(P_x) \prec \mu(Q_x)$, (ii) $\mu(P_x) = \mu(Q_x)$ and $\mathrm{sp}(P_x) \subsetneq \mathrm{sp}(Q_x)$, (iii) $P_x = Q_x$ and $\mathrm{sp}(P_y^{\sf U}) \subsetneq \mathrm{sp}(Q_y^{\sf U})$, (iv) $P_x = Q_x$, $P_y^{\sf U} = Q_y^{\sf U}$ and $\mathrm{sp}(P_y^{\sf L}) \subsetneq \mathrm{sp}(Q_y^{\sf L})$.

The next result is a consequence of [Jonov 2011, Theorem 3.2] and its proof.

Proposition 10. *The relation $<$ in Definition 9 defines a partial order and any extension of it to a total order on the facets of Δ_0 gives a shelling of Δ_0.*

The terminology of ES turns can be extended from lattice path monomials to facets of Δ_0 as follows. For any facet F of Δ_0 having a decomposition $F = F_x F_y^{\mathsf{U}} F_y^{\mathsf{L}}$ as in Proposition 5, by an *ES-turn* of F we shall mean an ES-turn of either F_x or F_y^{L} or F_y^{U}. It turns out that the corners of a facet of Δ_0 are essentially its ES turns or the leader of its x-component. There are, however, some subtleties involved and a precise relation is given below.

Lemma 11. *Let F be a facet of Δ_0 and $F = F_x F_y^{\mathsf{U}} F_y^{\mathsf{L}}$ be its decomposition as in Proposition 5. Also let $v \in V$ be a vertex of Δ_0. Then*:

(i) *If $v \in c(F)$, then either $v = \mu(F_x)$ or v is an ES-turn of F. In particular, $x_{m,n} \notin c(F)$ and $y_{m,n} \notin c(F)$.*

(ii) *If $\mu(F_x) = x_{i,j}$, with $(i, j) \neq (m, n-1)$, then $\mu(F_x) \in c(F)$. Moreover $x_{m,n-1} \notin c(F)$.*

(iii) *If v is an ES-turn of F_x, then $v \in c(F)$.*

(iv) *If v is an ES-turn of F_y^{U} or of F_y^{L}, then $v \in c(F)$, except when v is an ES-turn of F_y^{U} such that $v = y_{1,2}$ or when v is an ES-turn of F_y^{U} such that $v = y_{m-1,j+1}$ and $\mu(F_x) = x_{m,j}$ for some $j < n$.*

Proof. (i) Let $P = P_x P_y^{\mathsf{U}} P_y^{\mathsf{L}}$ be a facet of Δ_0 such that $F \setminus P = \{v\}$ and $F > P$. The latter implies that one of the four possibilities in Definition 9 must arise. First suppose $\mu(P_x) \prec \mu(F_x)$. Then $\mu(F_x)$ is a vertex of F that is smaller than $\mu(P_x)$ in the standard lexicographic order, and hence $\mu(F_x) \notin P_x$; consequently $v = \mu(F_x)$, and we are done. Now suppose $\mu(P_x) = \mu(F_x)$ and $\mathrm{sp}(P_x) \subsetneq \mathrm{sp}(F_x)$. Then $P_x \neq F_x$ and hence $F_x \setminus P_x = \{v\}$. Note that since $\mu(F_x)$ and $x_{m,n}$ are in P_x, the vertex v is an ES-turn, SE-turn, or the midpoint of a segment of F_x. In case it is the midpoint of a segment of F_x, the other two vertices in that segment must be in P_x, and since P_x is a lattice path monomial, we see that $v \in P_x$, which is a contradiction. Also if $v = x_{k,l}$ (say) is a SE-turn of F_x, then $x_{k-1,l}$ and $x_{k,l+1}$ must be in F_x and hence in P_x. But then P_x must contain $x_{k-1,l+1}$, which is a contradiction since $x_{k-1,l+1} \notin \mathrm{sp}(F_x)$. It follows that v is an ES-turn of F_x. Next suppose $P_x = F_x$ and $\mathrm{sp}(P_y^{\mathsf{U}}) \subsetneq \mathrm{sp}(F_y^{\mathsf{U}})$. Then $F_y^{\mathsf{U}} \setminus P_y^{\mathsf{U}} = \{v\}$. Since $\mu(P_x) = \mu(F_x)$, in view of Proposition 5, we see that the initial and the terminal variables of P_y^{U} and F_y^{U} coincide, and so v is neither of these. Arguing as in the preceding case, we can rule out the possibilities that v is a SE-turn or the midpoint of a segment of F_y^{U}. Hence v is an ES-turn of F_y^{U}. In a similar manner, we see that if $P_x = F_x$, $P_y^{\mathsf{U}} = F_y^{\mathsf{U}}$ and $\mathrm{sp}(P_y^{\mathsf{L}}) \subsetneq \mathrm{sp}(F_y^{\mathsf{L}})$, then v is a ES-turn of F_y^{L}. Thus (i) is proved.

(ii) Let $\mu(F_x) = x_{i,j}$ with $(i, j) \neq (m, n-1)$. Then either $x_{i,j+1} \in F_x$ or $x_{i+1,j} \in F_x$. First suppose $x_{i,j+1} \in F_x$. We define a new facet P as follows. Let $P_x = F_x \setminus \{x_{i,j}\}$

and $P_y^{\mathsf{L}} = F_y^{\mathsf{L}} \cup \{y_{m,j+1}\}$. To define P_y^{U}, we take $P_y^{\mathsf{U}} = F_y^{\mathsf{U}}$ in the case $y_{m,j+1} \notin F_y^{\mathsf{U}}$. If $y_{m,j+1} \in F_y^{\mathsf{U}}$, then this must mean that $i = m$, and hence $j < n-1$. We therefore define $P_y^{\mathsf{U}} = (F_y^{\mathsf{U}} \setminus \{y_{m,j+1}\}) \cup \{y_{m-1,j+2}\}$. Observe that $P = P_x P_y^{\mathsf{U}} P_y^{\mathsf{L}}$ is a facet of Δ_0 and since $\mu(P_x) \prec \mu(F_x)$, we have $P < F$. It follows that $\mu(F_x) \in c(F)$. Next suppose $x_{i+1,j} \in F_x$. We first assume that $(i, j) \neq (m-1, n)$. Now define a new facet P as follows. First we let $P_x = F_x \setminus \{x_{i,j}\}$. If $y_{i+1,n} \notin F_y^{\mathsf{L}}$, then we let $P_y^{\mathsf{U}} = F_y^{\mathsf{U}} \cup \{y_{i+1,n}\}$ and $P_y^{\mathsf{L}} = F_y^{\mathsf{L}}$. If $y_{i+1,n} \in F_y^{\mathsf{L}}$, then j must equal n. If now $i \leq m-2$, then we let $P_y^{\mathsf{L}} = (F_y^{\mathsf{L}} \setminus \{y_{i+1,n}\}) \cup \{y_{i+2,n-1}\}$. We are left with the special case $i = m-1$, $j = n$. Here we let $P_x = \{x_{m,n-1},\ x_{m,n}\}$, $P_y^{\mathsf{U}} = F_y^{\mathsf{U}} \cup \{y_{m,n}\}$, and $P_y^{\mathsf{L}} = F_y^{\mathsf{L}} \setminus \{y_{m,n}\}$. In all three cases, it is easy to verify that $P = P_x P_y^{\mathsf{U}} P_y^{\mathsf{L}}$ is a facet of Δ_0 such that $F \setminus P = \{x_{i,j}\}$ and $P < F$. Consequently $\mu(F_x) \in c(F)$. Finally we show that $x_{m,n-1} \notin c(F)$. Assume, on the contrary, that there is a facet P of Δ_0 such that $F \setminus P = \{x_{m,n-1}\}$. By (i) above, $\mu(F) = x_{m,n-1}$ because there can be no ES-turn at $x_{m,n-1}$. In view of Proposition 5, P must contain at least one variable other than $x_{m,n}$, and since $x_{m,n-1} \notin P$, it follows that $x_{m-1,n} \in P$. This forces $\mu(F_x) \prec \mu(P_x)$, which violates the fact that $P < F$. Thus (ii) is proved.

(iii) Let $v = x_{k,l}$ be an ES-turn of F_x. Define $P_x = F_x \setminus \{x_{k,l}\} \cup \{x_{k+1,l-1}\}$ and $P_y = F_y$. It is clear that $P = P_x P_y$ is a facet of Δ_0 such that $P < F$ and $F \setminus P = \{v\}$. This proves (iii).

(iv) First suppose $v = y_{k,l}$ is an ES-turn of F_y^{L}. Then $k < m$ and $l > 1$. Define $P_x = F_x$, $P_y^{\mathsf{U}} = F_y^{\mathsf{U}}$, and $P_y^{\mathsf{L}} = F_y^{\mathsf{L}} \setminus \{y_{k,l}\} \cup \{y_{k+1,l-1}\}$. It is easy to see that $P = P_x P_y^{\mathsf{U}} P_y^{\mathsf{L}}$ is facet of Δ_0 such that $P < F$ and $F \setminus P = \{v\}$. Next suppose $v = y_{k,l}$ is an ES-turn of F_y^{U}. Then once again $k < m$ and $l > 1$. In case $y_{k+1,l-1}$ is not in F_y^{L}, we define $P_x = F_x$, $P_y^{\mathsf{L}} = F_y^{\mathsf{L}}$, and $P_y^{\mathsf{U}} = F_y^{\mathsf{U}} \setminus \{y_{k,l}\} \cup \{y_{k+1,l-1}\}$, whereas in case $y_{k+1,l-1}$ is in F_y^{L} and also $k < m-1$ and $l > 2$, we define $P_x = F_x$, $P_y^{\mathsf{U}} = F_y^{\mathsf{U}} \setminus \{y_{k,l}\} \cup \{y_{k+1,l-1}\}$, and $P_y^{\mathsf{L}} = F_y^{\mathsf{L}} \setminus \{y_{k+1,l-1}\} \cup \{y_{k+2,l-2}\}$. We verify that in both the cases, $P = P_x P_y^{\mathsf{U}} P_y^{\mathsf{L}}$ is a facet of Δ_0 such that $P < F$ and $F \setminus P = \{v\}$.

When $l = 2$, it is easy to see that $v = y_{k,2}$ can be an ES-turn of F_y^{U} only when $k = 1$ lest F_y^{U} and F_y^{L} intersect at $y_{k,1}$. We now show that $y_{1,2}$ is not a corner of F. Suppose that $P = P_x P_y^{\mathsf{U}} P_y^{\mathsf{L}}$ is a facet of Δ_0 such that $F \setminus P = \{v\}$, $v = y_{1,2}$ and $F > P$. By Proposition 5, P_y^{U} must start at $y_{1,1}$ and P_y^{L} must start at $y_{2,1}$. For P_y^{U} to avoid $v = y_{1,2}$, it must be the case that P_y^{U} contains $y_{2,1}$. But this contradicts the fact that P_y^{U} and P_y^{L} do not intersect.

We are left with the situation where $k = m-1$ and $v = y_{k,l}$ is an ES-turn of F_y^{U} and moreover $y_{m,l-1} \in F_y^{\mathsf{L}}$. Now since F_y^{U} has an ES-turn at $y_{m-1,l}$, we see that $l > 1$ and both $y_{m-1,l-1}$ and $y_{m,l}$ are in F_y^{U}. In particular, $y_{m,l} \notin F_y^{\mathsf{L}}$ and since $y_{m,l-1} \in F_y^{\mathsf{L}}$, in view of Proposition 5, it follows that F_y^{L} ends at $y_{m,l-1}$, while F_y^{U} ends at $y_{m,n}$ and also that $\mu(F_x) = x_{m,l-1}$. Now if there were a facet $P = P_x P_y^{\mathsf{U}} P_y^{\mathsf{L}}$ of Δ_0 such that $F \setminus P = \{v\}$ and $F > P$, then $P_x = F_x$ and $P_y^{\mathsf{L}} = F_y^{\mathsf{L}}$, whereas

$F_y^{\mathsf{U}} \setminus P_y^{\mathsf{U}} = \{y_{m-1,l}\}$. But then P_y^{U} is a lattice path monomial that contains both $y_{m-1,l-1}$ and $y_{m,l}$ and does not contain $y_{m-1,l}$; so it must contain $y_{m,l-1}$. This is a contradiction since $y_{m,l-1} \in F_y^{\mathsf{L}} = P_y^{\mathsf{L}}$ and the monomials P_y^{U} and P_y^{L} have no variable in common. This completes the proof. □

For any integers i, j, k with $k \geq 0$, $1 \leq i \leq m$ and $1 \leq j \leq n$, we define $C_{i,j}^k$ to be the number of facets $F = F_x F_y$ of Δ_0 such that $\mu(F_x) = x_{i,j}$ and F has exactly k ES turns that are in $c(F)$. We state a useful consequence of Lemma 11:

Corollary 12. *The Hilbert series of the coordinate ring $R/\mathcal{I}_0$ of the principal component Z_0 of $\mathcal{L}_{2,2}^{m,n}$ is given by*

$$\frac{\sum_{k\geq 0} h_k z^k}{(1-z)^{2(m+n-1)}}, \tag{18}$$

where $h_0 = 1$, and for $k \geq 1$,

$$h_k = C_{m,n-1}^k + \sum_{\substack{(i,j)\neq(m,n-1)\\(i,j)\neq(m,n)}} C_{i,j}^{k-1}, \tag{19}$$

where the last sum is over all pairs (i, j) of integers satisfying $1 \leq i \leq m$ and $1 \leq j \leq n$, with $(i, j) \neq (m, n-1)$ and $(i, j) \neq (m, n)$.

Proof. It is well-known that the (Krull) dimension as well as the Hilbert series of $R/\mathcal{I}_0$ coincides with that of $R/\mathrm{LT}(\mathcal{I}_0)$ (see, e.g., [Bruns and Conca 2003, §3]), where $\mathrm{LT}(\mathcal{I}_0)$ denotes the leading term ideal of $\mathcal{I}_0$ as in [Košir and Sethuraman 2005b] and [Jonov 2011, Proposition 1.1]. Now Δ_0 is precisely the simplicial complex such that $R/\mathrm{LT}(\mathcal{I}_0)$ is the Stanley–Reisner ring of Δ_0. Thus it follows from Corollary 6 and part (ii) of Proposition 8 that the Hilbert series of $R/\mathcal{I}_0$ is given by (18), where $h_0 = 1$, and for $k \geq 1$,

$$h_k = \big|\{F : F \text{ a facet of } \Delta_0 \text{ with } |c(F)| = k\}\big|.$$

Partitioning the facets $F = F_x F_y$ in the above set in accordance with the values of $\mu(F_x)$ and noting from Proposition 5 that $\mu(F_x) \neq (m, n)$, and then applying Lemma 11, we obtain the desired result. □

We have seen in Section 3 that lattice path monomials can be related to lattice paths in the sense of Section 2.3 if we rotate to the left by 90° and identify the variable $x_{i,j}$ with the point (i, j) of $\mathbb{Z}^2$. Also recall that for any $(a, a'), (e, e') \in \mathbb{Z}^2$ and $s \in \mathbb{Z}$, we denote by $\mathcal{P}_s((a, a') \to (e, e'))$ the set of lattice paths from (a, a') to (e, e') with s NE turns. Likewise if $(a_i, a_i'), (e_i, e_i') \in \mathbb{Z}^2$ for $i = 1, 2$ and $s \in \mathbb{Z}$, then by $\mathcal{P}_s((a_1, a_1') \to (e_1, e_1'), (a_2, a_2') \to (e_2, e_2'))$ we denote the set of pairs (L_1, L_2) of nonintersecting lattice paths such that L_i is from (a_i, a_i') to (e_i, e_i') for $i = 1, 2$,

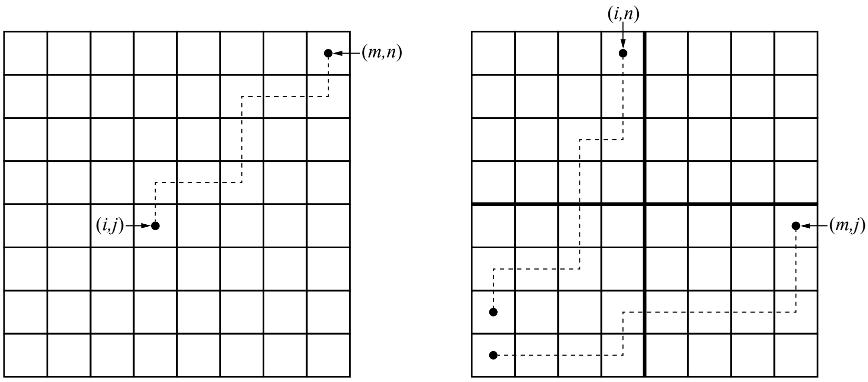

Figure 3. Lattice paths L and (L_1, L_2) corresponding to F_x and $(F_y^{\mathsf{U}}, F_y^{\mathsf{L}})$.

and the paths L_1 and L_2 together have exactly s NE turns. Evidently these sets are empty (and hence of cardinality 0) when $s < 0$.

Lemma 13. *Let $s, i, j \in \mathbb{Z}$ with $s \geq 0$, $1 \leq i \leq m$ and $1 \leq j \leq n$.*

(i) *If $i \neq m$, then*

$$C_{i,j}^s = \sum_{s_1+s_2=s} \left|\mathcal{P}_{s_1}((i, j) \to (m, n))\right| \left|\mathcal{P}_{s_2}((1, 2) \to (i, n), (1, 1) \to (m, j))\right|,$$

where the sum is over pairs (s_1, s_2) of nonnegative integers with $s_1 + s_2 = s$.

(ii) *If $1 < j < n - 1$, then*

$$\begin{aligned} C_{m,j}^s = &\sum_{p=1}^{m-1} \sum_{q=j+1}^{n-1} \left|\mathcal{P}_{s-1}((1, 2) \to (p, q), (1, 1) \to (m, j))\right| \\ &+ \sum_{p=1}^{m-2} \left|\mathcal{P}_{s-1}((1, 2) \to (p, j), (1, 1) \to (m, j))\right| \\ &+ \left|\mathcal{P}_s((1, 2) \to (m-1, j), (1, 1) \to (m, j))\right|. \end{aligned}$$

(iii) $C_{m,1}^s = \binom{n-2}{s}\binom{m-1}{s}$ *and*

$$\begin{aligned} C_{m,n-1}^s = &\sum_{p=1}^{m-2} \left|\mathcal{P}_{s-1}((1, 2) \to (p, n-1), (1, 1) \to (m, n-1))\right| \\ &+ \left|\mathcal{P}_s((1, 2) \to (m-1, n-1), (1, 1) \to (m, n-1))\right|. \end{aligned}$$

Proof. Let $i, j \in \mathbb{Z}$ with $1 \leq i \leq m$, $1 \leq j \leq n$, and $(i, j) \neq (m, n)$. By a 90° rotation to the left, we see from Proposition 5 that the facets $F = F_x F_y$ of Δ_0 with $\mu(F_x) = x_{i,j}$ are in one-to-one correspondence with the triples (L, L_1^*, L_2^*)

of lattice paths, where L is from (i, j) to (m, n), while L_1^* is from $(1, 1)$ to (i, n) and L_2^* is from $(2, 1)$ to (m, j), and moreover L_1^*, L_2^* are nonintersecting. We will now modify L_1^*, L_2^* slightly keeping in mind the hypothesis in Corollary 4. To this end, first note that $(1, 2) \in L_1^*$ since $2 < m \leq n$. Thus if we let $L_1 := L_1^* \setminus \{(1, 1)\}$ and $L_2 := L_2^* \cup \{(1, 1)\}$, then (L_1^*, L_2^*) and (L_1, L_2) are pairs of nonintersecting lattice paths that determine each other and have exactly the same NE turns, except that if L_1^* had a NE turn at $(1, 2)$, then L_1 will not have a NE turn at $(1, 2)$. Note though that, by Lemma 11 (iv), $y_{1,2}$ is not a corner of any facet, and this switch will therefore not affect the count of corners. Consequently the facets $F = F_x F_y$ of Δ_0 with $\mu(F_x) = x_{i,j}$ are in one-to-one correspondence with

$$\mathcal{P}((i, j) \to (m, n)) \times \mathcal{P}((1, 2) \to (i, n), (1, 1) \to (m, j)).$$

The lattice paths L and (L_1, L_2) corresponding to the components F_x and $(F_y^{\mathsf{U}}, F_y^{\mathsf{L}})$ of the facet $F = F_x F_y$ are illustrated in Figure 3; these may be compared with Figure 2 that depicts the lattice path monomials F_x and $F_y = F_y^{\mathsf{U}} F_y^{\mathsf{L}}$.

(i) Suppose $i \neq m$. Then, from Lemma 11, we see that, for every facet $F = F_x F_y$ of Δ_0 with $\mu(F_x) = x_{i,j}$, all the ES turns of F_x, F_y^{U} or F_y^{L} that are in $c(F)$ correspond to the NE turns of the corresponding lattice paths L, L_1 or L_2. From this, we readily obtain the formula in (i).

(ii) Suppose $i = m$ and $1 < j < n - 1$. Then for a facet $F = F_x F_y$ of Δ_0 with $\mu(F_x) = x_{m,j}$, the lattice path L corresponding to F_x is from (m, j) to (m, n) and evidently this has no NE turns. Consider in $\mathcal{P}((1, 2) \to (i, n), (1, 1) \to (m, j))$ the pair (L_1, L_2) corresponding to $(F_y^{\mathsf{U}}, F_y^{\mathsf{L}})$. Suppose the last NE-turn of L_1 is at $(p, q + 1)$. Note that if $q < j$, then we must have $(m, j) \in L_1$, which contradicts the fact that L_1, L_2 are nonintersecting. Thus $1 \leq p \leq m - 1$ and $j \leq q < n$. Moreover if $q = j$, then by part (iv) of Lemma 11, we see that either $p \leq m - 2$ or the NE-turn $(p, q + 1)$ is not in $c(F)$. It follows that L_1 can be replaced by its truncation $\tilde{L}_1$, which is a lattice path from $(1, 2)$ to (p, q) such that $\tilde{L}_1$ and L_2 are nonintersecting. Moreover the number of NE turns of $\tilde{L}_1$ in $c(F)$ are exactly one less than the number of NE turns of L_1 in $c(F)$, except when $(p, q) = (m - 1, j)$ in which case they are the same. Thus by varying (p, q) over an appropriate range, we obtain the formula in (ii).

(iii) If $(i, j) = (m, 1)$ and $F = F_x F_y$ is a facet of Δ_0 with $\mu(F_x) = x_{m,1}$, then the path L corresponding to F_x as well as the path L_2 corresponding to F_y^{L} have no NE turns. Moreover every NE-turn of the path $L_1 \in \mathcal{P}((1, 2) \to (m, n))$ corresponding to F_y^{U} is necessarily in $c(F)$, thanks to Lemma 11. Thus, in view of (12), we see that $C_{m,1}^s = \binom{n-2}{s}\binom{m-1}{s}$. Finally if $(i, j) = (m, n - 1)$, then arguing as in (ii) above, we see that for a facet $F = F_x F_y$ of Δ_0 with $\mu(F_x) = x_{m,n-1}$, the lattice path L corresponding to F_x has no NE turns and the last NE-turn of the lattice path L_1 corresponding to F_y^{U} must be (p, n) for some $p = 1, \ldots, m - 1$. Moreover by Lemma 11, this turn is counted as a corner (i.e., $x_{p,n} \in c(F)$) if and only if $p < m-1$. Thus upon replacing L_1 by its truncation up to $(p, n-1)$, we obtain the desired formula for $C^s_{m,n-1}$ in (iii). $\square$

We can already use the results obtained thus far to write down an explicit formula for the Hilbert series of the graded ring $R/\mathcal{I}_0$ corresponding to Z_0. Indeed it suffices to combine Corollary 12, Lemma 13, and Corollary 4. However the resulting formula is much too complicated and we will instead use results in Section 2 for simplifying various terms in (19) so as to eventually arrive at an elegant formula for (18).

Lemma 14. *Let k be a positive integer. Then $C^k_{m,n-1}$ is equal to*

$$\sum_{t_1+t_2=k} \binom{m-2}{t_1}\binom{n-2}{t_1}\binom{m-1}{t_2}\binom{n-2}{t_2} - \binom{m-1}{t_2+1}\binom{n-2}{t_2}\binom{m-2}{t_1-1}\binom{n-2}{t_1}.$$

Proof. For $s \in \mathbb{Z}$, let $f(s) := \binom{m-1}{s}\binom{n-2}{s}$ and $g(s) := \binom{m-2}{s-1}\binom{n-2}{s}$. By Corollary 4,

$$\sum_{p=1}^{m-2} \left|\mathcal{P}_{k-1}((1,2) \to (p, n-1), (1,1) \to (m, n-1))\right| \tag{20}$$

$$= \sum_{p=1}^{m-2} \sum_{s_1+s_2=k-1} \binom{p-1}{s_1}\binom{n-3}{s_1} f(s_2) - \binom{p}{s_2+1}\binom{n-3}{s_2} g(s_1)$$

$$= \sum_{s_1+s_2=k-1} \left(\sum_{p'=0}^{m-3} \binom{p'}{s_1}\right)\binom{n-3}{s_1} f(s_2) - \left(\sum_{p=1}^{m-2} \binom{p}{s_2+1}\right)\binom{n-3}{s_2} g(s_1)$$

$$= \sum_{s_1+s_2=k-1} \binom{m-2}{s_1+1}\binom{n-3}{s_1} f(s_2) - \binom{m-1}{s_2+2}\binom{n-3}{s_2} g(s_1)$$

$$= \sum_{t_1+t_2=k} \binom{m-2}{t_1}\binom{n-3}{t_1-1} f(t_2) - \binom{m-1}{t_2+1}\binom{n-3}{t_2-1} g(t_1),$$

where the penultimate equality follows from Lemma 1 since $\binom{0}{s_1+1} = 0 = \binom{1}{s_2+2}$ for $s_1, s_2 \geq 0$, and also since $\binom{n-3}{s_1} f(s_2) = 0 = \binom{n-3}{s_2} g(s_1)$ if $s_1 < 0$ or $s_2 < 0$, while the last equality follows by altering the summations (twice!) as in (10). On the other hand, by Corollary 4, $\left|\mathcal{P}_k((1,2) \to (m-1, n-1), (1,1) \to (m, n-1))\right|$ is equal to

$$\sum_{t_1+t_2=k} \binom{m-2}{t_1}\binom{n-3}{t_1} f(t_2) - \binom{m-1}{t_2+1}\binom{n-3}{t_2} g(t_1). \tag{21}$$

Now combining (20) and (21), using (6), and then using part (iii) of Lemma 13, we obtain the desired result. $\square$

Lemma 15. *Let k be a positive integer. Then* $\sum_{i=1}^{m-1}\sum_{j=1}^{n} C_{i,j}^{k-1}$ *is equal to*

$$\sum_{t_1+t_2=k}\binom{m}{t_2}\binom{n}{t_1+1}\binom{m-1}{t_1}\binom{n-2}{t_2-1}-\binom{m-1}{t_1}\binom{n}{t_2}\binom{m-1}{t_2-1}\binom{n-2}{t_1}.$$

Proof. Using (12) and part (i) of Lemma 13, we see that $\sum_{i=1}^{m-1}\sum_{j=1}^{n} C_{i,j}^{k-1}$ equals

$$\sum_{i=1}^{m-1}\sum_{j=1}^{n}\sum_{k_1+k_2=k-1}\binom{m-i}{k_1}\binom{n-j}{k_1}\left|\mathcal{P}_{k_2}((1,2)\to(i,n),(1,1)\to(m,j))\right|.$$

Applying Corollary 4 and then suitably interchanging summations and noting that the summands below are zero if $k_1<0$ or $s_1<0$ or $s_2<0$, this can be written as

$$\sum_{\substack{k_1+s_1+s_2=k-1\\ k_1,s_1,s_2\geq 0}} M_1N_1\binom{m-1}{s_2}\binom{n-2}{s_1}-M_2N_2\binom{m-2}{s_1-1}\binom{n-2}{s_2},\tag{22}$$

where, for any given $k_1, s_1, s_2\geq 0$, we have temporarily put

$$M_1=\sum_{i=1}^{m-1}\binom{m-i}{k_1}\binom{i-1}{s_1},\quad N_1=\sum_{j=1}^{n}\binom{n-j}{k_1}\binom{j-1}{s_2}=\binom{n}{k_1+s_2+1},$$

$$M_2=\sum_{i=1}^{m-1}\binom{m-i}{k_1}\binom{i}{s_2+1},\quad N_2=\sum_{j=1}^{n}\binom{n-j}{k_1}\binom{j-1}{s_1}=\binom{n}{k_1+s_1+1},$$

and where the simplified expressions for N_1, N_2 follow by rewriting each of the summands in N_1 and N_2 using (5), invoking (4) (noting that $k_1, s_1, s_2\geq 0$), and then applying (9) for suitable values of s, t, α and β. A similar simplification is possible in M_1 and M_2 if we add and subtract the term corresponding to $i=m$, and in view of (4), this is only necessary if $k_1=0$. Thus

$$M_1=\binom{m}{k_1+s_1+1}-\delta_{0,k_1}\binom{m-1}{s_1}\quad\text{and}\quad M_2=\binom{m+1}{k_1+s_2+2}-\delta_{0,k_1}\binom{m}{s_2+1},$$

where δ is the Kronecker delta. Substituting the simplified values of M_1, N_1, M_2, N_2 in (22), and letting

$$A(s_1,s_2):=\binom{m-1}{s_2}\binom{n-2}{s_1},\quad B(s_1,s_2):=\binom{m-2}{s_1-1}\binom{n-2}{s_2}$$

for $s_1, s_2\in\mathbb{Z}$, we see that (22) is of the form E_3+S_3, where

$$E_3=\sum_{\substack{k_1+s_1+s_2=k-1\\ k_1,s_1,s_2\geq 0}}\binom{m}{k-s_2}\binom{n}{k-s_1}A(s_1,s_2)-\binom{m+1}{k-s_1+1}\binom{n}{k-s_2}B(s_1,s_2),$$

and S_3 is the part where the Kronecker delta is nonzero:

$$S_3 = \sum_{s_1+s_2=k-1} \binom{m}{s_2+1}\binom{n}{s_1+1} B(s_1, s_2) - \binom{m-1}{s_1}\binom{n}{s_2+1} A(s_1, s_2).$$

Altering the summation as in (10), we see that S_3 can be written as

$$\sum_{t_1+t_2=k} \binom{m}{t_2}\binom{n}{t_1+1}\binom{m-2}{t_1-1}\binom{n-2}{t_2-1} - \binom{m-1}{t_1}\binom{n}{t_2}\binom{m-1}{t_2-1}\binom{n-2}{t_1}. \tag{23}$$

On the other hand, in view of (4) and (11), we can write

$$E_3 = \sum_{\ell=0}^{k-1} \sum_{s_1+s_2=\ell} \binom{m}{k-s_1}\binom{n}{k-s_2} A(s_2, s_1) - \binom{m+1}{k-s_1+1}\binom{n}{k-s_2} B(s_1, s_2).$$

By (6), we have

$$\binom{m+1}{k-s_1+1} = \binom{m}{k-s_1} + \binom{m}{k-(s_1-1)}.$$

Using this to split the second summand in E_3 into two parts and combining one of the parts with the first summand in E_3 and then applying (6) once again, we see that

$$E_3 = \sum_{\ell=0}^{k-1} \sum_{s_1+s_2=\ell} f(s_1, s_2) - f(s_1-1, s_2),$$

where

$$f(s_1, s_2) := \binom{m}{k-s_1}\binom{n}{k-s_2}\binom{m-2}{s_1}\binom{n-2}{s_2}$$

for $s_1, s_2 \in \mathbb{Z}$. Now in view of (10), we find that E_3 is given by the telescoping sum

$$E_3 = \sum_{\ell=0}^{k-1} F_\ell - F_{\ell-1}, \quad \text{where } F_\ell := \sum_{s_1+s_2=\ell} f(s_1, s_2) \text{ for } \ell \in \mathbb{Z}.$$

From the definition of f, we see that $F_{-1} = 0$, and thus $E_3 = F_{k-1}$, that is,

$$E_3 = \sum_{s_1+s_2=k-1} \binom{m}{k-s_1}\binom{n}{k-s_2}\binom{m-2}{s_1}\binom{n-2}{s_2}.$$

Now we can replace $k - s_1, k - s_2$ by $s_2 + 1, s_1 + 1$, respectively, in the above summand, and then alter the summation using (10) to obtain

$$E_3 = \sum_{t_1+t_2=k} \binom{m}{t_2}\binom{n}{t_1+1}\binom{m-2}{t_1}\binom{n-2}{t_2-1}. \tag{24}$$

Finally, by adding (24) and (23) termwise and using (6), we obtain the desired formula for $E_3 + S_3$, i.e., for $\sum_{i=1}^{m-1} \sum_{j=1}^{n} C_{i,j}^{k-1}$. □

Lemma 16. *Let k be a positive integer. Then* $\sum_{j=1}^{n-2} C_{m,j}^{k-1}$ *is equal to*

$$\sum_{t_1+t_2=k} \binom{m-1}{t_1}\binom{n-2}{t_1}\binom{m-1}{t_2-1}\binom{n-2}{t_2} - \binom{m}{t_2+1}\binom{n-2}{t_2}\binom{m-2}{t_1-2}\binom{n-2}{t_1}.$$

Proof. The desired result is easily verified when $n \le 3$ and so we assume that $n > 3$. For $j, s \in \mathbb{Z}$, let

$$f_j(s) := \binom{m-1}{s}\binom{j-1}{s}, \quad g_j(s) := \binom{m-2}{s-1}\binom{j-1}{s}.$$

In view of parts (iii) and (ii) of Lemma 13 together with (4) and Corollary 4, we see that

$$C_{m,1}^{k-1} = \binom{n-2}{k-1}\binom{m-1}{k-1} \quad \text{and} \quad \sum_{j=2}^{n-2} C_{m,j}^{k-1} = S_4 + S_5 + S_6, \tag{25}$$

where

$$S_4 = \sum_{j=2}^{n-2}\sum_{p=1}^{m-1}\sum_{q=j+1}^{n-1}\sum_{\substack{s_1+s_2=k-2\\ s_1,s_2\ge 0}} \binom{p-1}{s_1}\binom{q-2}{s_1} f_j(s_2) - \binom{p}{s_2+1}\binom{q-2}{s_2} g_j(s_1),$$

$$S_5 = \sum_{j=2}^{n-2}\sum_{p=1}^{m-2}\sum_{\substack{s_1+s_2=k-2\\ s_1,s_2\ge 0}} \binom{p-1}{s_1}\binom{j-2}{s_1} f_j(s_2) - \binom{p}{s_2+1}\binom{j-2}{s_2} g_j(s_1),$$

$$S_6 = \sum_{j=2}^{n-2}\sum_{s_1+s_2=k-1} \binom{m-2}{s_1}\binom{j-2}{s_1} f_j(s_2) - \binom{m-1}{s_2+1}\binom{j-2}{s_2} g_j(s_1).$$

Interchanging s_1 and s_2 in the second summand for S_6 as in (11), we can write

$$S_6 = \sum_{s_1+s_2=k-1} \lambda(s_1, s_2)\left(\binom{m-2}{s_1}\binom{m-1}{s_2} - \binom{m-1}{s_1+1}\binom{m-2}{s_2-1}\right), \tag{26}$$

where, for $s_1, s_2 \in \mathbb{Z}$, we let

$$\lambda(s_1, s_2) := \sum_{j=2}^{n-2} \binom{j-2}{s_1}\binom{j-1}{s_2}.$$

Next, by Lemma 1,

$$\sum_{p=1}^{m-2} \binom{p-1}{s_1} = \binom{m-2}{s_1+1} \quad \text{and} \quad \sum_{p=1}^{m-2} \binom{p}{s_2+1} = \binom{m-1}{s_2+2} \quad \text{for } s_1, s_2 \ge 0.$$

Consequently, by interchanging summations and rearranging terms, we find

$$\text{(27)}\quad S_5 = \sum_{j=2}^{n-2} \sum_{\substack{s_1+s_2=k-2\\ s_1,s_2\geq 0}} \binom{m-2}{s_1+1}\binom{j-2}{s_1} f_j(s_2) - \binom{m-1}{s_2+2}\binom{j-2}{s_2} g_j(s_1)$$

$$= \sum_{s_1+s_2=k-2} \lambda(s_1,s_2)\left(\binom{m-2}{s_1+1}\binom{m-1}{s_2} - \binom{m-1}{s_1+2}\binom{m-2}{s_2-1}\right)$$

$$= \sum_{s_1+s_2=k-1} \lambda(s_1-1,s_2)\left(\binom{m-2}{s_1}\binom{m-1}{s_2} - \binom{m-1}{s_1+1}\binom{m-2}{s_2-1}\right),$$

where the penultimate equality follows from (4) and (11) by interchanging s_1 and s_2 in the second summand of the preceding formula, while the last equality follows from (10). Now, using (6), we easily see that

$$\lambda(s_1-1,s_2)+\lambda(s_1,s_2) = \nu(s_1,s_2) \quad \text{for any } s_1,s_2\in\mathbb{Z},$$

where

$$\nu(s_1,s_2) := \sum_{j=2}^{n-2} \binom{j-1}{s_1}\binom{j-1}{s_2}.$$

Hence we can combine (27) and (26) to obtain

$$\text{(28)}\qquad S_5+S_6 = \sum_{s_1+s_2=k-1} \nu(s_1,s_2)\left(\binom{m-2}{s_1}\binom{m-1}{s_2} - \binom{m-1}{s_1+1}\binom{m-2}{s_2-1}\right).$$

It remains to consider S_4 or rather $C_{m,1}^{k-1}+S_4$. This is a little more complicated, but it can be handled using arguments similar to those in the proof of Lemma 15 as follows. First, by interchanging summations and using Lemma 1, we find

$$S_4 = \sum_{j=2}^{n-2} \sum_{\substack{s_1+s_2=k-2\\ s_1,s_2\geq 0}} \binom{m-1}{s_1+1}\theta(s_1) f_j(s_2) - \binom{m}{s_2+2}\theta(s_2) g_j(s_1),$$

where, for $s\in\mathbb{Z}$, we have let

$$\theta(s) := \binom{n-2}{s+1} - \binom{j-1}{s+1}.$$

Now observe that if $s_1<0$ or $s_2<0$, then $\theta(s_1)f_j(s_2) = 0 = \theta(s_2)g_j(s_1)$. Thus we may drop the condition $s_1,s_2\geq 0$ in the above expression for S_4, and then alter each of the two summations over (s_1,s_2) using (10) to write

$$S_4 = \sum_{j=2}^{n-2} \sum_{s_1+s_2=k-1} \binom{m-1}{s_1}\theta(s_1-1) f_j(s_2) - \binom{m}{s_2+1}\theta(s_2-1) g_j(s_1).$$

Next we collate the terms involving j and bring the summation over j inside, and note that, by Lemma 1, $\sum_{j=2}^{n-2}\binom{j-1}{s}=\binom{n-2}{s+1}-\delta_{0,s}$ for any $s\geq 0$. This yields

$$\begin{aligned}S_4=\sum_{s_1+s_2=k-1}&\binom{m-1}{s_1}\binom{n-2}{s_1}\binom{m-1}{s_2}\left(\binom{n-2}{s_2+1}-\delta_{0,s_2}\right)\\&-\binom{m}{s_2+1}\binom{n-2}{s_2}\binom{m-2}{s_1-1}\left(\binom{n-2}{s_1+1}-\delta_{0,s_1}\right)\\&-\binom{m-1}{s_1}\binom{m-1}{s_2}\nu(s_1,s_2)+\binom{m}{s_2+1}\binom{m-2}{s_1-1}\nu(s_1,s_2).\end{aligned}$$

Since $\binom{m-2}{s_1-1}=0$ when $s_1=0$, the only contribution of the terms involving Kronecker delta is when $s_2=0$, and it is $-\binom{m-1}{k-1}\binom{n-2}{k-1}$, that is, precisely $-C^{k-1}_{m,1}$. It follows that $C^{k-1}_{m,1}+S_4=S_4^*+E_4$, where

$$S_4^*=\sum_{s_1+s_2=k-1}\binom{m-1}{s_1}\binom{n-2}{s_1}\binom{m-1}{s_2}\binom{n-2}{s_2+1}-\binom{m}{s_2+1}\binom{n-2}{s_2}\binom{m-2}{s_1-1}\binom{n-2}{s_1+1}$$

and

$$\begin{aligned}E_4&=\sum_{s_1+s_2=k-1}\nu(s_1,s_2)\left(\binom{m}{s_2+1}\binom{m-2}{s_1-1}-\binom{m-1}{s_1}\binom{m-1}{s_2}\right)\\&=\sum_{s_1+s_2=k-1}\nu(s_1,s_2)\left(\binom{m}{s_1+1}\binom{m-2}{s_2-1}-\binom{m-1}{s_1}\binom{m-1}{s_2}\right),\end{aligned}\tag{29}$$

where the last equality follows by interchanging s_1 and s_2, while noting that ν is symmetric in s_1, s_2.

Now combining (28) and (29), and then, making an easy calculation using (6), we see that

$$E_4+S_5+S_6=\sum_{s_1+s_2=k-1}\nu(s_1,s_2)\left(\binom{m-1}{s_1}\binom{m-2}{s_2-1}-\binom{m-2}{s_1-1}\binom{m-1}{s_2}\right)=0,$$

where the last equality follows by interchanging s_1 and s_2 in one of the summations above. Thus $\sum_{j=1}^{n-2}C^{k-1}_{m,j}=S_4^*$. Finally, using (10), we readily see that S_4^* is precisely the desired formula in the statement of the lemma. □

Corollary 17. *Let k be a positive integer. Then $C^k_{m,n-1}+\sum_{j=1}^{n-2}C^{k-1}_{m,j}$ is equal to*

$$\sum_{t_1+t_2=k}\binom{m-1}{t_1}\binom{n-2}{t_1}\binom{m-1}{t_2}\binom{n-2}{t_2}-\binom{m-1}{t_2+1}\binom{n-2}{t_2}\binom{m-1}{t_1-1}\binom{n-2}{t_1}.$$

Proof. Consider the formula for $\sum_{j=1}^{n-2}C^{k-1}_{m,j}$ given by Lemma 16. This is a difference of two summations over $(t_1, t_2) \in \mathbb{Z}^2$ with $t_1 + t_2 = k$. Alter the first of these summations by interchanging t_1 and t_2, while putting $\binom{m}{t_2+1} = \binom{m-1}{t_2} + \binom{m-1}{t_2+1}$ in the second summation to split it into two summations. Then, using (6), we readily see that the formula for $\sum_{j=1}^{n-2} C_{m,j}^{k-1}$ becomes

$$\sum_{t_1+t_2=k} \binom{m-2}{t_1-1}\binom{n-2}{t_1}\binom{m-1}{t_2}\binom{n-2}{t_2} - \binom{m-1}{t_2+1}\binom{n-2}{t_2}\binom{m-2}{t_1-2}\binom{n-2}{t_1}.$$

This can be added termwise, using (6) once again, with the formula for $C_{m,n-1}^{k}$ given by Lemma 14, to obtain the desired result. □

We are now ready for our main theorem.

Theorem 18. *The Hilbert series of $R/\mathcal{I}_0$ is given by*

$$\left(\frac{\sum_{e=0}^{m-1}\binom{m-1}{e}\binom{n-1}{e}z^e}{(1-z)^{m+n-1}}\right)^2. \tag{30}$$

Proof. First note that (30) is of the form $(1-z)^{-2(m+n-1)}\sum_{k=0}^{2m-2} h_k^* z^k$, where

$$h_k^* = \sum_{t_1+t_2=k} \binom{m-1}{t_1}\binom{n-1}{t_1}\binom{m-1}{t_2}\binom{n-1}{t_2} \quad \text{for } k \in \mathbb{Z}. \tag{31}$$

On the other hand, by Corollary 12, we see that the Hilbert series of $R/\mathcal{I}_0$ is given by $(1-z)^{-2(m+n-1)}\sum_{k\geq 0} h_k z^k$, where $h_0 = 1$, and

$$h_k = \left(C_{m,n-1}^{k} + \sum_{j=1}^{n-2} C_{m,j}^{k-1}\right) + \sum_{i=1}^{m-1}\sum_{j=1}^{n} C_{i,j}^{k-1} \quad \text{for } k \geq 1. \tag{32}$$

It is clear that $h_0^* = 1 = h_0$ and so it suffices to show that $h_k^* = h_k$ for all $k \geq 1$. In view of Corollary 17 and Lemma 15, this is equivalent to showing that

$$\sum_{t_1+t_2=k} P_1(t_1,t_2) - P_2(t_1,t_2) + P_3(t_1,t_2) - P_4(t_1,t_2) - P(t_1,t_2) = 0 \quad \text{for } k \geq 1,$$

where $P_i(t_1, t_2)$ for $i = 1, \ldots, 4$, and $P(t_1, t_2)$ are the relevant summands, namely,

$$P_1(t_1,t_2) := \binom{m-1}{t_1}\binom{n-2}{t_1}\binom{m-1}{t_2}\binom{n-2}{t_2},$$

$$P_2(t_1,t_2) := \binom{m-1}{t_2+1}\binom{n-2}{t_2}\binom{m-1}{t_1-1}\binom{n-2}{t_1},$$

$$P_3(t_1,t_2) := \binom{m}{t_2}\binom{n}{t_1+1}\binom{m-1}{t_1}\binom{n-2}{t_2-1},$$

$$P_4(t_1,t_2) := \binom{m-1}{t_1}\binom{n}{t_2}\binom{m-1}{t_2-1}\binom{n-2}{t_1},$$

and

$$P(t_1,t_2) := \binom{m-1}{t_1}\binom{n-1}{t_1}\binom{m-1}{t_2}\binom{n-1}{t_2}$$

for $t_1, t_2 \in \mathbb{Z}$. To this end, we will make an extensive use of alterations as in (10) and (11); more specifically, the fact that

$$\sum_{t_1+t_2=k} f(t_1,t_2) = \sum_{t_1+t_2=k} f(t_2,t_1) = \sum_{t_1+t_2=k} f(t_1+1,t_2-1) = \sum_{t_1+t_2=k} f(t_2+1,t_1-1)$$

for any $f : \mathbb{Z}^2 \to \mathbb{Q}$ with finite support and any $k \in \mathbb{Z}$. Now fix any positive integer k and any $(t_1,t_2) \in \mathbb{Z}^2$ with $t_1+t_2 = k$. Observe that

$$P_3(t_1-1,t_2+1) - P_4(t_2,t_1) = \binom{m-1}{t_2+1}\binom{n}{t_1}\binom{m-1}{t_1-1}\binom{n-2}{t_2}.$$

Using (6) twice, we may substitute $\binom{n-2}{t_1} + \binom{n-2}{t_1-1} + \binom{n-1}{t_1-1}$ for $\binom{n}{t_1}$ in the right-hand side of the above identity to obtain

$$-P_2(t_1,t_2) + P_3(t_1-1,t_2+1) - P_4(t_2,t_1) = Q_1(t_1,t_2) + Q_2(t_1,t_2),$$

where

$$Q_1(t_1,t_2) := \binom{m-1}{t_2+1}\binom{n-2}{t_1-1}\binom{m-1}{t_1-1}\binom{n-2}{t_2},$$

$$Q_2(t_1,t_2) := \binom{m-1}{t_2+1}\binom{n-1}{t_1-1}\binom{m-1}{t_1-1}\binom{n-2}{t_2}.$$

Finally observe that $P_1(t_1,t_2) + Q_1(t_1+1,t_2-1) + Q_2(t_2+1,t_1-1) = P(t_1,t_2)$. This yields the desired result. □

It may be noted that in view of (2) and (30), the Hilbert series of the principal component Z_0 is precisely the square of the Hilbert series of the base variety $\mathscr{Z}_2^{m,n}$, and, as such, Theorem 7 could be deduced as a consequence of Theorem 18.

As an application of Theorem 18, we will now compute the a-invariant of the coordinate ring $R/\mathscr{I}_0$ of the principal component Z_0 of $\mathscr{Z}_{2,2}^{m,n}$ and determine when Z_0 is Gorenstein. Recall that if A is a finitely generated, positively graded Cohen–Macaulay algebra over a field, then A admits a graded canonical module ω_A and the a-invariant of A is defined as the negative of the least degree of a generator of ω_A. If the Hilbert series of A is given by $H_A(z) = h(z)/(1-z)^d$, where $d = \dim A$ and $h(z) \in \mathbb{Q}[z]$ with $h(1) \neq 0$, then the a-invariant of A is the order of the pole of $H_A(z)$ at infinity, which is $-(d - \deg h(z))$. Moreover the Hilbert series of ω_A is given by $H_{\omega_A}(z) = (-1)^d H_A(z^{-1})$. As a general reference for these notions and results, one may consult [Bruns and Herzog 1993], especially Sections 3.6 and 4.4. The following result is an analogue of a theorem of Gräbe [1988] (see also

[Ghorpade 1996, Theorem 4]) for classical determinantal varieties which says that if $1 \le r \le m \le n$, then the a-invariant of (the coordinate ring of) $\mathscr{Z}_r^{m,n}$ is $-(r-1)n$.

Corollary 19. *The a-invariant of $R/\mathscr{I}_0$ is equal to $-2n$ and the Hilbert series of the graded canonical module of $R/\mathscr{I}_0$ is given by*

$$\left(\frac{\sum_{e=0}^{m-1} \binom{m-1}{e}\binom{n-1}{e} z^{m+n-e-1}}{(1-z)^{m+n-1}} \right)^2. \tag{33}$$

Proof. We know from [Jonov 2011, Theorem 1.2] that $A = R/\mathscr{I}_0$ is Cohen–Macaulay and it is obviously a finitely generated, positively graded $\mathbb{F}$-algebra. Moreover, by Theorem 18, the Hilbert series of A is given by $h_0(z)/(1-z)^{2(m+n-1)}$, where

$$h_0(z) = \left(\sum_{e=0}^{m-1} \binom{m-1}{e}\binom{n-1}{e} z^e \right)^2.$$

Since $2 \le m \le n$, we see that $h_0(z)$ is a polynomial in z of degree $2(m-1)$, with leading coefficient $\binom{n-1}{m-1}^2$, and all other coefficients nonnegative integers; in particular, $h_0(1) \ne 0$. Hence the a-invariant of $A = R/\mathscr{I}_0$ is

$$2(m-1) - 2(m+n-1) = -2n,$$

and also that the Hilbert series of ω_A is given by (33). $\square$

The following result is an analogue of a theorem of Svanes [1974] (see also [Conca and Herzog 1994]) for classical determinantal varieties which says that for any $r \ge 1$, (the coordinate ring of) $\mathscr{Z}_r^{m,n}$ is Gorenstein if and only if $m = n$.

Corollary 20. *The coordinate ring $R/\mathscr{I}_0$ of Z_0 is Gorenstein if and only if $m = n$.*

Proof. By [Jonov 2011, Theorem 1.2] and [Košir and Sethuraman 2005b, Proposition 3.3], $A = R/\mathscr{I}_0$ is a Cohen–Macaulay domain. Hence from a well-known result of Stanley [1978, Theorem 4.4] (see also [Bruns and Herzog 1993, Corollary 4.4.6]), we see that A is Gorenstein if and only if $H_A(z) = (-1)^d z^a H_A(z^{-1})$ for some $a \in \mathbb{Z}$. Moreover, in this case, the integer a is necessarily the a-invariant of A. Thus, from Corollary 19, we see that $R/\mathscr{I}_0$ is Gorenstein if and only if

$$\left(\sum_{e=0}^{m-1} \binom{m-1}{e}\binom{n-1}{e} z^e \right)^2 = \left(\sum_{e=0}^{m-1} \binom{m-1}{e}\binom{n-1}{e} z^{m-1-e} \right)^2.$$

Since both the polynomials inside the square brackets on the two sides of the above equality have positive leading coefficients, it follows that $R/\mathscr{I}_0$ is Gorenstein if and only if $\binom{n-1}{e} = \binom{n-1}{m-1-e}$ for all $e = 0, 1, \dots, m-1$. Since $1 < m-1 \le n-1$, the latter clearly holds if and only if $m = n$. $\square$

Acknowledgments

Ghorpade is partially supported by Indo-Russian project INT/RFBR/P-114 from the Department of Science and Technology, Government of India and IRCC Award grant 12IRAWD009 from Indian Institute of Technology Bombay. Sethuraman was supported in part by NSF grants DMS-0700904 and CCF-1318260 during the preparation of the paper. He would like to thank the mathematics department at Indian Institute of Technology Bombay for its warm hospitality throughout the year-long visit during some of which part of this work was done. Ghorpade would like to thank the mathematics department at California State University Northridge for the summer visit during which this collaboration took root. The work presented in this paper originated from the Master's thesis of Jonov at California State University Northridge and is a continuation of his earlier paper [Jonov 2011].

References

[Abhyankar 1988] S. S. Abhyankar, *Enumerative combinatorics of Young tableaux*, Monographs and Textbooks in Pure and Applied Mathematics **115**, Marcel Dekker, New York, 1988. MR 89e:05011 Zbl 0643.05001

[Arbarello et al. 1985] E. Arbarello, M. Cornalba, P. A. Griffiths, and J. Harris, *Geometry of algebraic curves, I*, Grundlehren der Mathematischen Wissenschaften **267**, Springer, New York, 1985. MR 86h:14019 Zbl 0559.14017

[Bruns and Conca 2003] W. Bruns and A. Conca, "Gröbner bases and determinantal ideals", pp. 9–66 in *Commutative algebra, singularities and computer algebra* (Sinaia, 2002), edited by J. Herzog and V. Vuletescu, NATO Sci. Ser. II Math. Phys. Chem. **115**, Kluwer, Dordrecht, 2003. MR 2005a:13025 Zbl 1052.13010

[Bruns and Herzog 1993] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, Cambridge Studies in Advanced Mathematics **39**, Cambridge University Press, 1993. MR 95h:13020 Zbl 0788.13005

[Bruschek et al. 2011] C. Bruschek, H. Mourtada, and J. Schepers, "Arc spaces and Rogers–Ramanujan identities", pp. 211–220 in *23rd International Conference on Formal Power Series and Algebraic Combinatorics* (Reykjavik, 2011), Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2011. MR 2012m:14025

[Chaundy 1932] T. Chaundy, "The unrestricted plane partition", *Quart. J. Math., Oxf. Ser.* **3** (1932), 76–80. Zbl 0004.09704

[Conca and Herzog 1994] A. Conca and J. Herzog, "On the Hilbert function of determinantal rings and their canonical module", *Proc. Amer. Math. Soc.* **122**:3 (1994), 677–681. MR 95a:13016 Zbl 0823.13008

[Docampo 2013] R. Docampo, "Arcs on determinantal varieties", *Trans. Amer. Math. Soc.* **365**:5 (2013), 2241–2269. MR 3020097 Zbl 1272.14014

[Ein and Mustaţă 2009] L. Ein and M. Mustaţă, "Jet schemes and singularities", pp. 505–546 in *Algebraic geometry—Seattle 2005, II* (Seattle, WA, 2005), edited by D. Abramovich et al., Proc. Sympos. Pure Math. **80**, Amer. Math. Soc., Providence, RI, 2009. MR 2010h:14004 Zbl 1181.14019

[Galligo 1985] A. Galligo, "Computations of some Hilbert functions related with Schubert calculus", pp. 79–97 in *Algebraic geometry* (Sitges, Barcelona, 1983), edited by E. Casas-Alvero et al., Lecture Notes in Math. **1124**, Springer, Berlin, 1985. MR 87j:14083 Zbl 0589.14039

[Gessel and Viennot 1985] I. Gessel and G. Viennot, "Binomial determinants, paths, and hook length formulae", *Adv. in Math.* **58**:3 (1985), 300–321. MR 87e:05008 Zbl 0579.05004

[Ghorpade 1994] S. R. Ghorpade, "Abhyankar's work on Young tableaux and some recent developments", pp. 215–249 in *Algebraic geometry and its applications* (West Lafayette, IN, 1990), edited by C. Bajaj, Springer, New York, 1994. MR 95h:05148 Zbl 0814.14046

[Ghorpade 1996] S. R. Ghorpade, "Young bitableaux, lattice paths and Hilbert functions", *J. Statist. Plann. Inference* **54**:1 (1996), 55–66. MR 98c:05165 Zbl 0864.05088

[Ghorpade 2001] S. R. Ghorpade, "A note on Hodge's postulation formula for Schubert varieties", pp. 211–219 in *Geometric and combinatorial aspects of commutative algebra* (Messina, 1999), edited by J. Herzog and G. Restuccia, Lecture Notes in Pure and Appl. Math. **217**, Dekker, New York, 2001. MR 2002d:14079 Zbl 0986.14028

[Ghorpade and Krattenthaler 2004] S. R. Ghorpade and C. Krattenthaler, "The Hilbert series of Pfaffian rings", pp. 337–356 in *Algebra, arithmetic and geometry with applications* (West Lafayette, IN, 2000), edited by C. Christensen et al., Springer, Berlin, 2004. MR 2005c:13010 Zbl 1083.13504

[Goward and Smith 2006] R. A. Goward, Jr. and K. E. Smith, "The jet scheme of a monomial scheme", *Comm. Algebra* **34**:5 (2006), 1591–1598. MR 2007e:13037 Zbl 1120.14055

[Gräbe 1988] H.-G. Gräbe, *Streckungsringe*, Dissertation B, Erfurt/Mühlhausen, 1988.

[Herzog and Trung 1992] J. Herzog and N. V. Trung, "Gröbner bases and multiplicity of determinantal and Pfaffian ideals", *Adv. Math.* **96**:1 (1992), 1–37. MR 94a:13012 Zbl 0778.13022

[Jonov 2011] B. Jonov, "Initial complex associated to a jet scheme of a determinantal variety", *J. Pure Appl. Algebra* **215**:5 (2011), 806–811. MR 2011j:14103 Zbl 1219.13013

[Karlin and McGregor 1959] S. Karlin and J. McGregor, "Coincidence probabilities", *Pacific J. Math.* **9** (1959), 1141–1164. MR 22 #5072 Zbl 0092.34503

[Košir and Sethuraman 2005a] T. Košir and B. A. Sethuraman, "Determinantal varieties over truncated polynomial rings", *J. Pure Appl. Algebra* **195**:1 (2005), 75–95. MR 2005h:13020 Zbl 1085.14043

[Košir and Sethuraman 2005b] T. Košir and B. A. Sethuraman, "A Groebner basis for the 2×2 determinantal ideal mod t^2", *J. Algebra* **292**:1 (2005), 138–153. MR 2006f:13011 Zbl 1105.13009

[Krattenthaler 1995a] C. Krattenthaler, "Counting nonintersecting lattice paths with turns", *Sém. Lothar. Combin.* **34** (1995), Art. B34i, 17 pp. MR 97e:05018 Zbl 0855.05004

[Krattenthaler 1995b] C. Krattenthaler, *The major counting of nonintersecting lattice paths and generating functions for tableaux*, Mem. Amer. Math. Soc. **552**, 1995. MR 95i:05109 Zbl 0830.05003

[Kulkarni 1996] D. M. Kulkarni, "Counting of paths and coefficients of the Hilbert polynomial of a determinantal ideal", *Discrete Math.* **154**:1-3 (1996), 141–151. MR 97f:13018 Zbl 0853.05005

[Lindström 1973] B. Lindström, "On the vector representations of induced matroids", *Bull. London Math. Soc.* **5** (1973), 85–90. MR 49 #95 Zbl 0262.05018

[Liu 2002] Q. Liu, *Algebraic geometry and arithmetic curves*, Oxford Graduate Texts in Mathematics **6**, Oxford University Press, Oxford, 2002. MR 2003g:14001 Zbl 0996.14005

[Modak 1992] M. R. Modak, "Combinatorial meaning of the coefficients of a Hilbert polynomial", *Proc. Indian Acad. Sci. Math. Sci.* **102**:2 (1992), 93–123. MR 94e:13022 Zbl 0761.13006

[Mourtada 2011] H. Mourtada, "Jet schemes of toric surfaces", *C. R. Math. Acad. Sci. Paris* **349**:9-10 (2011), 563–566. MR 2012e:14029 Zbl 1222.14031

[Mustaţă 2001] M. Mustaţă, "Jet schemes of locally complete intersection canonical singularities", *Invent. Math.* **145**:3 (2001), 397–424. MR 2002f:14005 Zbl 1091.14004

[Mustaţă 2002] M. Mustaţă, "Singularities of pairs via jet schemes", *J. Amer. Math. Soc.* **15**:3 (2002), 599–615. MR 2003b:14005 Zbl 0998.14009

[Nash 1995] J. F. Nash, Jr., "Arc structure of singularities", *Duke Math. J.* **81**:1 (1995), 31–38. MR 98f:14011 Zbl 0880.14010

[Sethuraman and Šivic 2009] B. A. Sethuraman and K. Šivic, "Jet schemes of the commuting matrix pairs scheme", *Proc. Amer. Math. Soc.* **137**:12 (2009), 3953–3967. MR 2011b:14117 Zbl 1185.14046

[Smith and Weyman 2007] G. Smith and J. Weyman, "Resolutions of tangent schemes of determinantal ideals", private communication, 2007.

[Stanley 1978] R. P. Stanley, "Hilbert functions of graded algebras", *Advances in Math.* **28**:1 (1978), 57–83. MR 58 #5637 Zbl 0384.13012

[Svanes 1974] T. Svanes, "Coherent cohomology on Schubert subschemes of flag schemes and applications", *Advances in Math.* **14** (1974), 369–453. MR 54 #7490 Zbl 0308.14008

[Yuen 2007a] C. Yuen, "Jet schemes of determinantal varieties", pp. 261–270 in *Algebra, geometry and their interactions*, edited by A. Corso et al., Contemp. Math. **448**, Amer. Math. Soc., Providence, RI, 2007. MR 2009b:14093 Zbl 1142.14032

[Yuen 2007b] C. Yuen, "The multiplicity of jet schemes of a simple normal crossing divisor", *Comm. Algebra* **35**:12 (2007), 3909–3911. MR 2008i:13051 Zbl 1127.14052

Received July 5, 2013.

SUDHIR R. GHORPADE
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY BOMBAY
POWAI, MUMBAI 400076
INDIA
srg@math.iitb.ac.in

BOYAN JONOV
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA SANTA BARBARA
SANTA BARBARA, CA 93106
UNITED STATES
boyan@math.ucsb.edu

B. A. SETHURAMAN
DEPARTMENT OF MATHEMATICS
CALIFORNIA STATE UNIVERSITY NORTHRIDGE
NORTHRIDGE, CA 91330
UNITED STATES
al.sethuraman@csun.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.272.177

ON A LIU–YAU TYPE INEQUALITY FOR SURFACES

OUSSAMA HIJAZI, SEBASTIÁN MONTIEL AND SIMON RAULOT

Let Ω be a compact mean-convex domain with smooth boundary $\Sigma := \partial\Omega$, in an initial data set (M^3, g, K), which has no apparent horizon in its interior. If Σ is spacelike in a spacetime $(\mathscr{E}^4, g_{\mathscr{E}})$ with spacelike mean curvature vector $\mathscr{H}$ such that Σ admits an isometric and isospin immersion into $\mathbb{R}^3$ with mean curvature H_0, then

$$\int_\Sigma |\mathscr{H}|\, d\Sigma \le \int_\Sigma \frac{H_0^2}{|\mathscr{H}|}\, d\Sigma.$$

If equality occurs, we prove that there exists a local isometric immersion of Ω in $\mathbb{R}^{3,1}$ (the Minkowski spacetime) with second fundamental form given by K. We also examine, under weaker conditions, the case where the spacetime is the $(n+2)$-dimensional Minkowski space $\mathbb{R}^{n+1,1}$ and establish a stronger rigidity result.

1. Introduction

Let $(\mathscr{E}^4, g_{\mathscr{E}})$ be a spacetime satisfying the Einstein field equations; that is, $(\mathscr{E}^4, g_{\mathscr{E}})$ is a 4-dimensional time-oriented Lorentzian manifold such that

$$\mathrm{Ric}_{\mathscr{E}} - \tfrac{1}{2} R_{\mathscr{E}} g_{\mathscr{E}} = \mathscr{T},$$

where $R_{\mathscr{E}}$ (respectively, $\mathrm{Ric}_{\mathscr{E}}$) denotes the scalar curvature (respectively, the Ricci curvature) of $(\mathscr{E}, g_{\mathscr{E}})$, and $\mathscr{T}$ is the energy-momentum tensor which describes the matter content of the ambient spacetime. We also assume that $(\mathscr{E}^4, g_{\mathscr{E}})$ satisfies the dominant energy condition; that is, its energy-momentum tensor $\mathscr{T}$ has the property that, for every future-directed causal vector $\eta \in \Gamma(T\mathscr{E})$, the vector field dual to the one-form $-\mathscr{T}(\eta, \cdot)$ is a future-directed causal vector of $T\mathscr{E}$.

Let M^3 be an immersed spacelike hypersurface of $(\mathscr{E}^4, g_{\mathscr{E}})$ with induced Riemannian metric g. Assume that T is the future-directed timelike normal vector to M and denote by K the associated second fundamental form defined by $K(X, Y) = g_{\mathscr{E}}(\nabla^{\mathscr{E}}_X T, Y)$ for all $X, Y \in \Gamma(TM)$. Here $\nabla^{\mathscr{E}}$ denotes the Levi-Civita

S. Montiel was partially supported by the Spanish MEC-FEDER grant no. MTM2011-22547.
MSC2010: 53C27, 53C40, 53C80.
Keywords: manifolds with boundary, Dirac operator, Einstein equations, initial data set, mean curvature, holographic principle.

connection of the spacetime. Then the Gauss, Codazzi and Einstein equations provide *constraint equations* on M, given by

$$\begin{cases} \mu = \frac{1}{2}\big(R - |K|_M^2 + (\mathrm{tr}_M(K))^2\big), \\ J = -\delta(K - \mathrm{tr}_M(K)g), \end{cases}$$

where R is the scalar curvature of (M^3, g), $|K|^2$ and $\mathrm{tr}(K)$ denote the squared norm and the trace of K with respect to g, and δ is the divergence on M. Here μ and J are the energy and momentum density of the matter fields, and are given by

$$\mu = \mathcal{T}(T, T) \quad \text{and} \quad J_i = \mathcal{T}(e_i, T)$$

for $1 \leq i \leq 3$, where $\{e_1, e_2, e_3\}$ is a local basis of the spatial tangent space of M. The dominant energy condition for the spacetime implies that $\mu \geq |J|$ as functions on M. A triplet (M^3, g, K) which satisfies the dominant energy condition is called an *initial data set*.

Now we consider a codimension-two spacelike orientable surface Σ^2 in the spacetime $\mathscr{E}^4$. We will represent by $\mathscr{H}$ the mean curvature vector field on Σ^2, defined as

$$\mathscr{H} = \mathrm{tr}\, II,$$

where II is the second fundamental form of this immersion. Since the normal space at each point of Σ^2 is a Lorentzian plane, it can be spanned by two future-directed null normal vector fields $\mathcal{N}_+$ and $\mathcal{N}_-$, normalized in such a way that $\langle \mathcal{N}_+, \mathcal{N}_- \rangle = -\frac{1}{2}$. We denote by θ_+ and θ_- the components of $\mathscr{H}$ with respect to $\mathcal{N}_+$ and $\mathcal{N}_-$. They are the so-called future-directed null expansions of $\mathscr{H}$, and measure the area growth when Σ^2 varies in the corresponding directions. It is clear that

$$|\mathscr{H}|^2 = -\theta_+\theta_-.$$

If θ_+ and θ_- are both negative, the surface will be called a *trapped* surface. A surface with $\theta_+ = 0$ or $\theta_- = 0$ is called an *apparent horizon* (or a *marginally trapped* surface). Note that if Σ^2 is trapped or marginally trapped, then the mean curvature vector $\mathscr{H}$ is a causal vector at each point. This is why the mean curvature field $\mathscr{H}$ being spacelike everywhere is equivalent to Σ being an *untrapped* surface.

In the case that Σ^2 spans a spacelike hypersurface in the spacetime, that is, when there exists a spacelike hypersurface Ω^3 immersed in $\mathscr{E}^4$ such that $\partial\Omega^3 = \Sigma^2$, the normal null vector fields $\mathcal{N}_+$ and $\mathcal{N}_-$ may be ordered in such a way that they project onto directions tangent to Ω^3 which are, respectively, *outer* and *inner* normal at each point of Σ^2. In other words, if N is the inner normal unit vector field on Σ^2 tangent to Ω^3 and T is the future-directed timelike normal to Ω^3 in $\mathscr{E}^4$, we put

$$\mathcal{N}_+ = \tfrac{1}{2}(T - N) \quad \text{and} \quad \mathcal{N}_- = \tfrac{1}{2}(T + N).$$

The second fundamental form of Σ^2 in $\mathscr{E}^4$ is given in terms of the Lorentzian basis of the normal bundle the hypersurface Ω by

$$II(X,Y) = g(AX,Y)N + g(BX,Y)T$$

for all $X, Y \in \Gamma(T\Sigma)$, where $AX := -\nabla_X N$ denotes the shape operator of Σ^2 in Ω^3 and ∇ is the Levi-Civita connection of the Riemannian metric g on M. The mean curvature vector field $\mathscr{H}$ of Σ in $\mathscr{E}$ can be reexpressed by

$$\mathscr{H} = \theta_+ \mathscr{N}_- + \theta_- \mathscr{N}_+ = HN + \mathrm{tr}_\Sigma\,(K)\,T,$$

where $H = \mathrm{tr}\,A$ is the mean curvature of Σ^2 in Ω^3 and $\mathrm{tr}_\Sigma\,(K)$ is the trace on Σ^2 of the shape operator K of Ω^3 in $\mathscr{E}^4$. The norm of $\mathscr{H}$ can be also reexpressed as

$$|\mathscr{H}|^2 = H^2 - \mathrm{tr}_\Sigma\,(K)^2 = -\theta_+\theta_-, \tag{1}$$

where $\theta_\pm = \mathrm{tr}_\Sigma\,(K) \pm H$ are the future-directed null expansions of $\mathscr{H}$. The spacelike surfaces with $\theta_+ < 0$ (respectively, $\theta_- < 0$) are referred to as *outer* (respectively, *inner*) *trapped* surfaces. It is easy to see that untrapped submanifolds, that is, codimension-two spacelike submanifolds of a spacetime with spacelike mean curvature vector field, naturally divide into two disjoint classes:

Lemma 1. *Let Σ^2 be a compact spacelike codimension-two submanifold embedded in a spacetime $\mathscr{E}^4$. Suppose that its mean curvature vector field $\mathscr{H}$ is spacelike and that Σ^2 is the boundary of a spacelike hypersurface Ω^3 in $\mathscr{E}^4$. Then Ω^3 is either mean-convex or mean-concave.*

Proof. It suffices to take into account that if (θ_+, θ_-) are the future-directed null expansions of the mean curvature vector field $\mathscr{H}$ associated to the embedding of Σ^2 in the domain Ω^3, we have, from (1),

$$0 < |\mathscr{H}|^2 = -\theta_+\theta_- \quad \text{and} \quad \theta_+ - \theta_- = 2H,$$

where H is the inner mean curvature function of Σ^2 in Ω^3. The first of these two equalities implies that θ_+ and θ_- have opposite signs everywhere on Σ^2. Then, from the second one, we have that either $H > 0$ or $H < 0$ on the whole of Σ^2. $\square$

Note that this fact obviously holds for higher-dimensional initial data sets. In the following, an untrapped surface (respectively, a codimension-two untrapped submanifold) which bounds a compact, connected and mean-convex spacelike hypersurface will be referred to as an *outer untrapped surface* (respectively, an *outer untrapped submanifold*). It is worth noting that round spheres in Euclidean slices are untrapped surfaces. The same occurs in general for large radial spheres in asymptotically flat spacelike hypersurfaces.

We now give the precise statement of our main result:

Theorem 2. *Let Ω be a compact domain with an outer untrapped boundary surface $\Sigma := \partial\Omega$ in an initial data set (M^3, g, K). If Ω has no apparent horizon in its interior, then for all $\varphi \in \Gamma(\mathbb{S}\Sigma)$,*

$$\int_\Sigma \left(\frac{1}{|\mathcal{H}|} |\not{D}\varphi|^2 - \frac{|\mathcal{H}|}{4} |\varphi|^2 \right) d\Sigma \geq 0, \tag{2}$$

where $\mathbb{S}\Sigma$ is the extrinsic spinor bundle on Σ and $\not{D}$ is the extrinsic Dirac operator (see Section 2). *Moreover, if equality occurs, then there exists a local isometric immersion of Ω in $\mathbb{R}^{3,1}$ with K as second fundamental form.*

As a direct application, we prove the following result:

Theorem 3. *Under the conditions of Theorem 2, assume furthermore that Σ admits an isometric and isospin immersion into $\mathbb{R}^3$ with mean curvature H_0. Then*

$$\int_\Sigma |\mathcal{H}|\, d\Sigma \leq \int_\Sigma \frac{H_0^2}{|\mathcal{H}|}\, d\Sigma. \tag{3}$$

Moreover, if equality occurs, then Σ is connected and there exists a local isometric immersion of Ω in $\mathbb{R}^{3,1}$ with second fundamental form given by K and mean curvature vector of Σ satisfying $|\mathcal{H}| = H_0$.

If we consider the case of codimension-two outer untrapped submanifolds in the $(n+2)$-dimensional Minkowski spacetime $\mathbb{R}^{n+1,1}$, we prove that we can remove the assumption on the nonexistence of apparent horizons (see Theorem 14). Moreover, in this situation, we completely characterize the equality case. Namely:

Theorem 4. *Let Σ be a codimension-two outer untrapped submanifold in $\mathbb{R}^{n+1,1}$. If Σ admits an isometric and isospin immersion into $\mathbb{R}^{n+1}$ with mean curvature H_0, then inequality* (3) *holds and equality is achieved if and only if Σ lies in a hyperplane in $\mathbb{R}^{n+1,1}$ and Σ is connected.*

Remark 5. In Theorems 3 and 4, we assumed that the boundary hypersurface of a compact domain in a certain spin manifold admits an *isospin immersion* into a Euclidean space. In general, an $(n + 1)$-dimensional spin manifold induces a spin structure on each of its orientable immersed hypersurfaces through their corresponding immersions (see Section 2.2 below). Two distinct immersions of an orientable manifold Σ^n into two (possibly different) $(n + 1)$-dimensional spin manifolds are said to be *isospin* when the spin structures induced on Σ^n from the corresponding ambient manifolds coincide (up to an equivalence). Recall that spin structures on Σ^n are parametrized by the cohomology group $H^1(\Sigma^n, \mathbb{Z}_2)$. Thus, for example, if Σ^n is a simply connected manifold, any two immersions of Σ^n in two arbitrary $(n + 1)$-dimensional spin manifolds must be isospin. Consequently if the surface Σ in Theorem 3 has genus zero or the hypersurface Σ in Theorem 4 is simply connected, we only need to suppose that they are mean-convex in their initial data sets and that they can be immersed as hypersurfaces in a Euclidean space.

Also it is clear that when the two immersions defined on Σ^n lie in the same ambient space and are *regularly homotopic*, the associated induced spin structures are equivalent. In fact, two immersions are said to be regularly homotopic (*isotopic*, according to Pinkall [1985] and others) if we may pass continuously from one to the other through a family of immersions. Consequently they determine the same class in $H^1(\Sigma^n, \mathbb{Z}_2)$. Indeed in the case $n = 2$, two spin structures induced from the spin structure of the 3-dimensional spin ambient space through two different embeddings are equivalent if and only if they are regularly homotopic (besides the previous reference, see [Hass and Hughes 1985, pp. 104–105] and [Benedetti and Silhol 1995, p. 656]).

Then take any compact mean-convex surface Σ embedded in $\mathbb{R}^3$. This surface bounds a compact domain in three-dimensional Euclidean space which is a totally geodesic initial data set in the Minkowski space $\mathbb{R}^{3,1}$. If we slightly deform this surface, the positivity of the mean curvature is preserved by continuity, and, from the arguments above, the same holds for the induced spin structure. So there are examples of mean-convex boundaries in initial data sets of spacetimes admitting isospin immersions in Euclidean spaces. Many of them are nonconvex. In fact, take Σ to be, for instance, a right cylinder with two half-spheres closing its extremes (after smoothing) or a torus of revolution thin enough (if we want to have some point with negative Gauss curvature).

Note that if Σ is not convex, we cannot use the Weyl theorem and so we do not know whether it is possible to immerse Σ isometrically in Euclidean space $\mathbb{R}^3$. This is why in this case, Theorems 3 and 4 should be viewed as comparison theorems for the mean curvatures of two immersions in the spirit of a classical result by Herglotz. Indeed, Herglotz [1943] gave a succinct proof of Cohn-Vossen's rigidity result for convex surfaces based on an integral inequality involving the second fundamental forms of two embeddings (see, e.g., [Montiel and Ros 1997, Section 7.4]). Our Theorem 3 provides an inequality of this type which could be a first step in enlarging the Cohn-Vossen theorem to include Euclidean mean-convex compact surfaces.

In this direction, one can easily see that Theorem 4 implies that the integral of the mean curvature is preserved through *bendings* of compact mean-convex hypersurfaces embedded in a Euclidean space. This was first proved by Almgren and Rivin [1998] (see also [Rivin and Schlenker 1999]).

Recall that Liu and Yau [2006] (see also [Liu and Yau 2003]) proved the following positivity result: Let (Ω^3, g, K) be an initial data set for the Einstein equation. Suppose that the boundary $\partial\Omega$ has finitely many components Σ_i, $1 \le i \le l$, each of which has positive Gauss curvature and spacelike mean curvature vector in the spacetime. Then for all i,

$$\int_{\Sigma_i} |\mathscr{H}| \, d\Sigma \leq \int_{\Sigma_i} H_0 \, d\Sigma. \tag{4}$$

Moreover, if equality occurs for some $i \in \{1, \dots, l\}$, then $\partial\Omega$ is connected and the spacetime is flat along Ω.

The proof of this result relies on a generalized version of the positive mass theorem and on the resolution of the Jang equation. One of the key ingredients in the proof is provided by the Weyl embedding theorem [1916], which asserts that the condition that Σ embeds isometrically as a strictly convex hypersurface in $\mathbb{R}^3$ is equivalent to Σ having positive Gauss curvature. Note that by the Cauchy–Schwarz inequality, inequality (4) implies (3).

More recently, Eichmair, Miao and Wang [Eichmair et al. 2012] generalized inequality (4) for time-symmetric initial data under weaker convexity assumptions for the embedding of Σ in $\mathbb{R}^3$. We point out that, in contrast to Liu and Yau's result, we do not assume that the immersion is a *strictly convex embedding*. In particular, the mean curvature H_0 is not assumed to be positive.

2. The Riemannian setting

2.1. *Preliminaries on spin manifolds.* Let (M, g) be an $(n+1)$-dimensional Riemannian spin manifold, which we will suppose from now on to be connected, and denote by ∇ the Levi-Civita connection on its tangent bundle TM. We choose a spin structure on M and consider the corresponding spinor bundle $\mathbb{S}M$, a rank-$2^{(n+1)/2}$ complex vector bundle. Denote by γ the Clifford multiplication

$$\gamma : \mathbb{C}\ell(M) \longrightarrow \mathrm{End}(\mathbb{S}M), \tag{5}$$

which is a fiber-preserving algebra morphism. Then $\mathbb{S}M$ becomes a bundle of complex left modules over the Clifford bundle $\mathbb{C}\ell(M)$ over the manifold M. When $(n+1)$ is even, the spinor bundle splits into the direct sum of the *positive* and *negative* chiral subbundles:

$$\mathbb{S}M = \mathbb{S}M^+ \oplus \mathbb{S}M^-, \tag{6}$$

where $\mathbb{S}M^\pm$ are defined to be the ± 1-eigenspaces of the endomorphism $\gamma(\omega_{n+1})$, with $\omega_{n+1} = i^{(n+2)/2} e_1 e_2 \cdots e_{n+1}$ the complex volume form.

On the spinor bundle $\mathbb{S}M$, one has (see [Lawson and Michelsohn 1989]) a natural Hermitian metric, denoted by $\langle \cdot\,, \cdot \rangle$, and the spinorial Levi-Civita connection ∇ acting on spinor fields. It is well-known that the Hermitian scalar product, the Levi-Civita connection ∇ and the Clifford multiplication (5) satisfy, for any spinor fields $\psi, \varphi \in \Gamma(\mathbb{S}M)$ and any tangent vector fields $X, Y \in \Gamma(TM)$, the compatibility conditions

(7) $$\langle\gamma(X)\psi,\gamma(X)\varphi\rangle = |X|^2\langle\psi,\varphi\rangle,$$

(8) $$X\langle\psi,\varphi\rangle = \langle\nabla_X\psi,\varphi\rangle + \langle\psi,\nabla_X\varphi\rangle,$$

(9) $$\nabla_X\big(\gamma(Y)\psi\big) = \gamma(\nabla_X Y)\psi + \gamma(Y)\nabla_X\psi.$$

Since $\nabla\omega_{n+1} = 0$, for $n+1$ even, the decomposition (6) is orthogonal and ∇ preserves this decomposition.

The Dirac operator D on $\mathbb{S}M$ is the first-order elliptic differential operator locally given by

$$D = \sum_{i=1}^{n+1}\gamma(e_i)\nabla_{e_i},$$

where $\{e_1,\dots,e_{n+1}\}$ is a local orthonormal frame of TM. When $(n+1)$ is even, the Dirac operator interchanges positive and negative spinor fields; that is,

$$D : \Gamma(\mathbb{S}M^{\pm}) \longmapsto \Gamma(\mathbb{S}M^{\mp}).$$

2.2. *Hypersurfaces and induced structures.* In this section, we compare the restrictions $\not{\mathbb{S}}\Sigma$ of the spinor bundle $\mathbb{S}M$ of a spin manifold M to an orientable hypersurface Σ immersed into M, and its Dirac-type operator $\not{D}$ to the intrinsic spinor bundle $\mathbb{S}\Sigma$ of the induced spin structure on Σ and its fundamental Dirac operator D_Σ. A fundamental case will be when the hypersurface Σ is just the boundary ∂M of a manifold M. These facts are in general well-known (see, for example, [Bureš 1993; Trautman 1995; Bär 1998; Baum et al. 1990; Hijazi et al. 2001a; 2001b; 2002; Hijazi and Montiel 2014]). For completeness, we introduce the notation and key facts.

Denote by $\not{\nabla}$ the Levi-Civita connection associated with the induced Riemannian metric on Σ. The Gauss formula says that

(10) $$\not{\nabla}_X Y = \nabla_X Y - g(AX,Y)N,$$

where X, Y are vector fields tangent to the hypersurface Σ, the vector field N is a global unit field normal to Σ, and A stands for the shape operator corresponding to N; that is,

(11) $$\nabla_X N = -AX \quad \text{for all } X \in \Gamma(T\Sigma).$$

We have that the restriction

$$\not{\mathbb{S}}\Sigma := \mathbb{S}M|_\Sigma$$

is a left module over $\mathbb{C}\ell(\Sigma)$ for the induced Clifford multiplication

$$\not{\gamma} : \mathbb{C}\ell(\Sigma) \longrightarrow \mathrm{End}(\not{\mathbb{S}}\Sigma)$$

given by

$$\slashed{\gamma}(X)\psi = \gamma(X)\gamma(N)\psi \tag{12}$$

for every $\psi \in \Gamma(\slashed{\mathbb{S}}\Sigma)$ and $X \in \Gamma(T\Sigma)$. (Note that a spinor field on the ambient manifold M and its restriction to the hypersurface Σ will be denoted by the same symbol.) Consider the Hermitian metric $\langle\,\cdot\,,\cdot\,\rangle$ on $\slashed{\mathbb{S}}\Sigma$ induced from that of $\mathbb{S}M$. This metric immediately satisfies the compatibility condition (7) if one considers the Riemannian metric on Σ induced from M and the Clifford multiplication $\slashed{\gamma}$ defined in (12). Now the Gauss formula (10) implies that the spin connection $\slashed{\nabla}$ on $\slashed{\mathbb{S}}\Sigma$ is given by the spinorial Gauss formula

$$\slashed{\nabla}_X\psi = \nabla_X\psi - \tfrac{1}{2}\slashed{\gamma}(AX)\psi = \nabla_X\psi - \tfrac{1}{2}\gamma(AX)\gamma(N)\psi \tag{13}$$

for every $\psi \in \Gamma(\slashed{\mathbb{S}}\Sigma)$ and $X \in \Gamma(T\Sigma)$. Note that the compatibility conditions (7), (8) and (9) are satisfied by $(\slashed{\mathbb{S}}\Sigma, \slashed{\gamma}, \langle\,\cdot\,,\cdot\,\rangle, \slashed{\nabla})$.

Denote by $\slashed{D} : \Gamma(\slashed{\mathbb{S}}\Sigma) \to \Gamma(\slashed{\mathbb{S}}\Sigma)$ the Dirac operator associated with the Dirac bundle $\slashed{\mathbb{S}}\Sigma$ over the hypersurface. It is a well-known fact that $\slashed{D}$ is a first-order elliptic differential operator which is formally L^2-selfadjoint. By (13), for any spinor field $\psi \in \Gamma(\mathbb{S}M)$,

$$\slashed{D}\psi = \sum_{j=1}^{n} \slashed{\gamma}(e_j)\slashed{\nabla}_{e_j}\psi = \tfrac{1}{2}H\psi - \gamma(N)\sum_{j=1}^{n}\gamma(e_j)\nabla_{e_j}\psi,$$

where $\{e_1, \dots, e_n\}$ is a local orthonormal frame of $T\Sigma$ and $H = \operatorname{tr} A$ is the mean curvature of Σ corresponding to the orientation N. Using (13) and (11), it is straightforward to see that the skew-commutativity rule

$$\slashed{D}(\gamma(N)\psi) = -\gamma(N)\slashed{D}\psi \tag{14}$$

holds for any spinor field $\psi \in \Gamma(\slashed{\mathbb{S}}\Sigma)$. It is important to point out that, from this fact, *the spectrum of $\slashed{D}$ is always symmetric with respect to zero*, while this is the case for the Dirac operator D_Σ of the intrinsic spinor bundle *only when n is even*. Indeed, in this case, we have an isomorphism of Dirac bundles

$$(\slashed{\mathbb{S}}\Sigma, \slashed{\gamma}, \slashed{D}) \equiv (\mathbb{S}\Sigma, \gamma_\Sigma, D_\Sigma),$$

and the decomposition $\slashed{\mathbb{S}}\Sigma = \slashed{\mathbb{S}}\Sigma^+ \oplus \slashed{\mathbb{S}}\Sigma^-$, given by

$$\slashed{\mathbb{S}}\Sigma^\pm := \{\psi \in \slashed{\mathbb{S}}\Sigma \mid i\gamma(N)\psi = \pm\psi\},$$

corresponds to the chiral decomposition of the spinor bundle $\mathbb{S}\Sigma$. Hence $\slashed{D}$ interchanges $\slashed{\mathbb{S}}\Sigma^+$ and $\slashed{\mathbb{S}}\Sigma^-$.

When n is odd the spectrum of D_Σ is not necessarily symmetric. In fact, in this case, the spectrum of $\slashed{D}$ is just the symmetrization of the spectrum of D_Σ.

This is why the decomposition of $\mathbb{S}M$ into positive and negative chiral spinors induces an orthogonal and $\gamma\!\!\!/$, $D\!\!\!\!/$-invariant decomposition $\$\Sigma = \$\Sigma_+ \oplus \$\Sigma_-$, with $\$\Sigma_\pm := (\mathbb{S}M^\pm)_{|\Sigma}$, in such a way that

$$(\$\Sigma_\pm, \gamma\!\!\!/, D\!\!\!\!/_{|\$\Sigma_\pm}) \equiv (\mathbb{S}\Sigma, \pm\gamma_\Sigma, \pm D_\Sigma).$$

Also, $\gamma(N)$ interchanges the decomposition, and both maps $\gamma(N) : \$\Sigma_\pm \to \$\Sigma_\mp$ are isomorphisms.

Consequently, studying the spectrum of the induced operator $D\!\!\!\!/$ is equivalent to studying the spectrum of the Dirac operator D_Σ of the Riemannian spin structure induced on the hypersurface Σ.

2.3. ***A spinorial Reilly-type inequality for manifolds with boundary.*** Here, we prove a spinorial Reilly-type inequality (see [Liu and Yau 2003] and [Raulot 2013]).

Recall that on a compact $(n+1)$-dimensional Riemannian spin manifold M with boundary $\Sigma = \partial M$, for any spinor field $\psi \in \Gamma(\mathbb{S}M)$, the fundamental Schrödinger–Lichnerowicz formula is given by:

$$\int_\Sigma \left(\langle D\!\!\!\!/\psi, \psi\rangle - \frac{H}{2}|\psi|^2 \right) d\Sigma = \int_M \left(\tfrac{1}{4} R|\psi|^2 + |\nabla\psi|^2 - |D\psi|^2 \right) dM, \tag{15}$$

where R is the scalar curvature of M. Note that the assumption $R \geq 0$ is quite natural and has been used intensively to get, in particular, lower bounds on both D and $D\!\!\!\!/$. However, in our situation (see Section 3.1), we have a weaker assumption on the scalar curvature. More precisely, we assume that there exists a smooth vector field $X \in \Gamma(TM)$ such that

$$R \geq 2|X|^2 + 2\delta(X), \tag{16}$$

where $|X|^2 = g(X, X)$ and δ is the divergence of $X = \sum_{j=1}^{n} X^j e_j \in \Gamma(TM)$, locally given by

$$\delta(X) = -\sum_{i=1}^{n+1} e_i(X^i).$$

Then we prove an adapted Reilly-type inequality. Namely:

Proposition 6. *Let M a compact Riemannian spin manifold with boundary Σ such that there exists a smooth vector field $X \in \Gamma(TM)$ satisfying* (16). *Then*

$$\int_\Sigma \left\langle D\!\!\!\!/\psi - \tfrac{1}{2}(H + g(X, N))\psi, \psi \right\rangle d\Sigma \geq \int_M \left(\tfrac{1}{2}|\nabla\psi|^2 - |D\psi|^2 \right) dM. \tag{17}$$

Moreover, equality occurs if and only if the spinor field ψ satisfies

$$\nabla_Y \psi = -g(X, Y)\psi \tag{18}$$

for all $Y \in \Gamma(TM)$.

Proof. First note that, since

$$\delta(|\psi|^2 X) = -X(|\psi|^2) + |\psi|^2\delta(X),$$

the Stokes formula gives

$$\begin{aligned}\int_M \frac{R}{4}|\psi|^2\,dM &= \int_M \Big(\frac{R}{4} - \frac{1}{2}\delta(X)\Big)|\psi|^2\,dM + \frac{1}{2}\int_M \delta(X)|\psi|^2\,dM \\ &= \frac{1}{4}\int_M (R - 2\delta(X))|\psi|^2\,dM + \frac{1}{2}\int_M X(|\psi|^2)\,dM \\ &\qquad + \frac{1}{2}\int_\Sigma g(X,N)|\psi|^2\,d\Sigma.\end{aligned}$$

Inserting this identity in (15) leads to

$$\begin{aligned}&\int_\Sigma \big\langle D\!\!\!/\psi - \tfrac{1}{2}(H + g(X,N)), \psi\big\rangle\,d\Sigma \\ &\quad = \int_M \big(\tfrac{1}{4}(R - 2\delta(X))|\psi|^2 + \tfrac{1}{2}X(|\psi|^2)\big)\,dM + \int_M (|\nabla\psi|^2 - |D\psi|^2)\,dM\end{aligned}$$

and, using (16), we conclude that

$$\begin{aligned}&\int_\Sigma \big\langle D\!\!\!/\psi - \tfrac{1}{2}(H + g(X,N)), \psi\big\rangle\,d\Sigma \\ &\quad \geq \int_M \big(\tfrac{1}{2}|X|^2|\psi|^2 + \tfrac{1}{2}X(|\psi|^2)\big)\,dM + \int_M (|\nabla\psi|^2 - |D\psi|^2)\,dM.\end{aligned}\tag{19}$$

If we let $\widetilde{\nabla}_Y\psi := \nabla_Y\psi + g(X,Y)\psi$, it is straightforward to compute

$$|\widetilde{\nabla}\psi|^2 = |\nabla\psi|^2 + |X|^2|\psi|^2 + 2\,\mathrm{Re}\langle\nabla_X\psi, \psi\rangle,$$

and since $2\,\mathrm{Re}\langle\nabla_X\psi,\psi\rangle = X(|\psi|^2)$, we get

$$\tfrac{1}{2}X(|\psi|^2) \geq -\tfrac{1}{2}|\nabla\psi|^2 - \tfrac{1}{2}|X|^2|\psi|^2,$$

with equality if and only if $\widetilde{\nabla}\psi = 0$. Combining this last inequality with (19) finishes the proof. □

2.4. ***A local boundary elliptic condition for the Dirac operator.*** As before, Σ is the boundary of an $(n+1)$-dimensional Riemannian spin compact manifold M. We define two pointwise projections

$$P_\pm : \mathbb{S}\Sigma \longrightarrow \mathbb{S}\Sigma$$

on the induced Dirac bundle over the hypersurface by

$$P_\pm = \tfrac{1}{2}(\mathrm{Id}_{\mathbb{S}\Sigma} \pm i\gamma(N)).\tag{20}$$

It is a well-known fact that these two orthogonal projections $P_\pm$ acting on the spin bundle $\mathbb{S}\Sigma$ provide local elliptic boundary conditions for the Dirac operator D of M. The ellipticity of these boundary conditions and that of the Dirac operator D allow us to solve boundary value problems for D on M by prescribing, on the boundary Σ, the corresponding $P_\pm$-projections of the solutions. Namely, we have:

Proposition 7 [Hijazi and Montiel 2014]. *Let M be a compact Riemannian spin manifold with boundary a hypersurface Σ. If $\varphi \in \Gamma(\mathbb{S}\Sigma)$ is a smooth spinor field of the induced Dirac bundle, then the boundary value problem*

$$\begin{cases} D\psi = 0 & \text{on } M, \\ P_\pm(\psi_{|\Sigma}) = P_\pm\varphi & \text{on } \Sigma \end{cases}$$

for the Dirac operator has a unique smooth solution $\psi \in \Gamma(\mathbb{S}M)$.

For a more general discussion on boundary conditions for the Dirac operator, we refer to [Booß-Bavnbek and Wojciechowski 1993], [Ballmann and Bär 2012] or [Bartnik and Chruściel 2005].

2.5. ***A holographic principle for the existence of parallel spinors.*** It is by now standard (see [Hijazi et al. 2001b; 2002]) to make use of (15) for a compact Riemannian spin manifold M with nonnegative scalar curvature R, together with the solution of an appropriate boundary value problem for the Dirac operator D of M, in order to establish a certain integral inequality for the induced Dirac operator $\not{D}$ of the boundary hypersurface $\partial M = \Sigma$. Raulot [2013] uses such arguments for compact manifolds whose scalar curvature satisfies (16). In this section, we generalize the holographic principle for the existence of parallel spinors proved in [Hijazi and Montiel 2014] in the context studied in [Raulot 2013].

First we need to recall the following fact:

Lemma 8 [Hijazi et al. 2002]. *For any smooth spinor field $\psi \in \Gamma(\mathbb{S}\Sigma)$,*

$$\int_\Sigma \langle \not{D}\psi, \psi\rangle \, d\Sigma = 2\int_\Sigma \langle \not{D} P_+\psi, P_-\psi\rangle \, d\Sigma.$$

The proof simply relies on the self-adjointness of the Dirac operator $\not{D}$ and on the identities

$$\not{D} P_\pm = P_\mp \not{D}, \tag{21}$$

which are obtained using (14) and (20).

Proposition 9. *Let M be a compact Riemannian spin manifold with scalar curvature satisfying* (16) *such that*

$$F := H + g(X, N) > 0.$$

For any $\varphi \in \Gamma(\mathbb{S}\Sigma)$, *one has*

$$0 \leq \int_\Sigma \left(\frac{1}{F} |\not{D} P_+ \varphi|^2 - \frac{F}{4} |P_+ \varphi|^2 \right) d\Sigma. \tag{22}$$

Moreover equality holds if and only if there exists a parallel spinor field $\psi \in \Gamma(\mathbb{S}M)$ *such that* $P_+\psi = P_+\varphi$ *along the boundary hypersurface* Σ *and the vector field* X *vanishes identically on* M.

Proof. Take any spinor field $\varphi \in \Gamma(\mathbb{S}\Sigma)$ of the induced spinor bundle on the hypersurface and consider the boundary value problem

$$\begin{cases} D\psi = 0 & \text{on } M, \\ P_+\psi = P_+\varphi & \text{on } \Sigma \end{cases}$$

for the Dirac operator D and the boundary condition P_+. The existence and uniqueness of a smooth solution $\psi \in \Gamma(\mathbb{S}M)$ for this boundary problem is ensured by Proposition 7. This solution ψ, inserted in inequality (17), translates to

$$0 \leq \frac{1}{2} \int_M |\nabla \psi|^2 \, dM \leq \int_\Sigma \left(\langle \not{D}\psi, \psi \rangle - \frac{F}{2} |\psi|^2 \right) d\Sigma. \tag{23}$$

Note that if equality is achieved, then ψ is a parallel spinor field satisfying (18). Since such a spinor field has no zeros, the vector field X vanishes identically on the whole of M. Inequality (23) combined with Lemma 8, together with the fact that the decomposition

$$\psi = P_+\psi + P_-\psi$$

is pointwise orthogonal, imply

$$0 \leq \int_\Sigma \left(2\langle \not{D} P_+\psi, P_-\psi \rangle - \frac{F}{2} |P_+\psi|^2 - \frac{F}{2} |P_-\psi|^2 \right) d\Sigma. \tag{24}$$

Since the function F is assumed to be positive on Σ, it follows that

$$0 \leq \left| \sqrt{\frac{2}{F}} \not{D} P_+\psi - \sqrt{\frac{F}{2}} P_-\psi \right|^2 = \frac{2}{F} |\not{D} P_+\psi|^2 + \frac{F}{2} |P_-\psi|^2 - 2\langle \not{D} P_+\psi, P_-\psi \rangle.$$

In other words,

$$2\langle \not{D} P_+\psi, P_-\psi \rangle - \frac{F}{2} |P_-\psi|^2 \leq \frac{2}{F} |\not{D} P_+\psi|^2,$$

which, when combined with inequality (24), implies inequality (22). Now, if equality holds, we already noticed that the spinor field ψ must be parallel with $P_+\psi = P_+\varphi$ and $X \equiv 0$.

Conversely, if we assume that there is a parallel spinor field ψ on M and $X \equiv 0$, then we are in the situation covered in [Hijazi and Montiel 2014]. □

With this, we are ready to state the main result of this section:

Theorem 10. *Let M be a compact Riemannian spin $(n+1)$-dimensional manifold, and $X \in \Gamma(TM)$ such that*

$$R \geq 2|X|^2 + 2\delta(X) \quad \text{and} \quad F := H + g(X, N) > 0.$$

Then, for any spinor field $\varphi \in \Gamma(\mathbb{S}\Sigma)$, one has

$$0 \leq \int_\Sigma \left(\frac{1}{F} |D\!\!\!/ \varphi|^2 - \frac{F}{4} |\varphi|^2 \right) d\Sigma. \tag{25}$$

Equality holds if and only if there exist two parallel spinor fields $\Psi^+, \Psi^- \in \Gamma(\mathbb{S}M)$ such that $P_+\Psi^+ = P_+\varphi$ and $P_-\Psi^- = P_-\varphi$ on the boundary and $X \equiv 0$.

Proof. From the symmetry between the two boundary conditions P_+ and P_- for the Dirac operator on M (see Proposition 7 and Lemma 8), one can repeat the proof of Proposition 9 to get the inequality corresponding to (22) where the *positive* projection P_+ is replaced by the *negative* one P_-. Hence, for any spinor field $\varphi \in \Gamma(\mathbb{S}\Sigma)$, we also have

$$0 \leq \int_\Sigma \left(\frac{1}{F} |D\!\!\!/ P_-\varphi|^2 - \frac{F}{4} |P_-\varphi|^2 \right) d\Sigma. \tag{26}$$

Taking into account the relation (21) and the pointwise orthogonality of the projections $P_\pm$, the sum of the two inequalities (22) and (26) yields (25). The equality case is a consequence of Proposition 9. ☐

Remark 11. Note that, as observed in [Hijazi and Montiel 2014], equality in (25) does not imply that the two parallel spinors in Theorem 10 coincide.

We should also mention that inequality (25) has a nice interpretation in terms of the first eigenvalue of the boundary Dirac operator $D\!\!\!/_F$ associated with the conformal metric $g_F = F^2 g$. More precisely:

Corollary 12. *Let (M^{n+1}, g) be an $(n+1)$-dimensional compact connected Riemannian spin manifold satisfying the assumptions of Theorem 10. Then the first nonnegative eigenvalue $\lambda_1(D\!\!\!/_F)$ of the Dirac operator corresponding to the conformal metric $g_F = F^2 g$ satisfies*

$$\lambda_1(D\!\!\!/_F) \geq \tfrac{1}{2},$$

and equality holds if and only if M admits a nontrivial parallel spinor (and $X \equiv 0$). In this case, the eigenspace corresponding to $\lambda_1(D\!\!\!/_F) = \frac{1}{2}$ consists of restrictions to Σ of parallel spinor fields on M multiplied by the function $F^{-(n-1)/2}$. Furthermore the boundary hypersurface Σ has to be connected.

The proof is omitted since it is similar to [Hijazi and Montiel 2014, Theorem 1].

2.6. ***A discussion on quasilocal masses.*** In this section, we consider a 3-dimensional compact connected Riemannian manifold (M^3, g) with nonnegative scalar curvature, whose boundary Σ^2 has positive mean curvature H. Note that since M is a 3-dimensional manifold, it is necessarily spin. Moreover we also assume that there exists an immersion ι_0 of the surface Σ in $\mathbb{R}^3$ with mean curvature H_0.

One of the fundamental results in classical general relativity is certainly the proof of the positivity of the total energy by Schoen and Yau [1981] and Witten [1981]. This led to the more ambitious claim of associating energy to extended, but finite, spacetime domains, that is, at the quasilocal level. Obviously the quasilocal data could provide a more detailed characterization of the states of the gravitational field than the global ones, so they are interesting in their own right. For a complete review of these topics, we refer to [Szabados 2004]. It is currently required that a quasilocal mass satisfies natural properties, among which are:

(I) *Nonnegativity*: $\mathcal{M}(\Sigma) \geq 0$.

(II) *Rigidity*: $\mathcal{M}(\Sigma) = 0$ if and only if Σ is in the Minkowski spacetime.

(III) *Monotonicity*: If $\Sigma_1 = \partial M_1$ and $\Sigma_2 = \partial M_2$ such that $M_1 \subset M_2$, then $\mathcal{M}(\Sigma_1) \leq \mathcal{M}(\Sigma_2)$.

(IV) *ADM limit*: If (Σ_k) is a sequence of surfaces that exhaust an asymptotically flat manifold (N^3, g), then

$$\lim_{k\to\infty} \mathcal{M}(\Sigma_k) = m_{\mathrm{ADM}}(g),$$

where $m_{\mathrm{ADM}}(g)$ is the ADM mass of (N, g).

(V) *Black hole limit*: If Σ is a horizon in an asymptotically flat manifold (N^3, g), then

$$\mathcal{M}(\Sigma) = \sqrt{\frac{A}{16\pi}},$$

where A is the area of Σ.

Brown and York [1993] proposed the following definition for the quasilocal mass of a surface Σ (now called the Brown–York mass):

$$m_{BY}(\Sigma) := \frac{1}{8\pi} \int_\Sigma (H_0 - H)\, d\Sigma.$$

The nonnegativity of $m_{BY}(\Sigma)$ is proved in [Shi and Tam 2002] under additional assumptions. Indeed they impose that ι_0 is a *strictly convex* isometric embedding, which by the Weyl embedding theorem [1916] is equivalent to the fact that Σ has positive Gauss curvature. Moreover, in this situation, the embedding ι_0 is unique up to an isometry of $\mathbb{R}^3$.

Recently Lam [2011] proposed in his thesis the definition

$$m_L(\Sigma) := \frac{1}{16\pi}\int_\Sigma \frac{1}{H_0}(H_0^2 - H^2)\, d\Sigma.$$

He proves that $m_L(\Sigma)$ has several interesting properties for certain surfaces in complete asymptotically flat Riemannian manifolds that are the graphs of smooth functions over $\mathbb{R}^3$ (see the same work for a precise description). More precisely, it satisfies Properties (I), (III), (IV) and (V). Moreover, using the Cauchy–Schwarz inequality, it is straightforward to check that $m_{BY}(\Sigma) \geq m_L(\Sigma)$.

From [Hijazi and Montiel 2014], we can define a quasilocal mass similar to the Brown–York and Lam masses, and prove its nonnegativity in the more general context described in the beginning of this section. Indeed, if we let

$$m(\Sigma) := \frac{1}{16\pi}\int_\Sigma \frac{1}{H}(H_0^2 - H^2)\, d\Sigma,$$

then, from the immersion ι_0, there exists a spinor field $\Psi_0 \in \Gamma(\$\Sigma)$ satisfying the Dirac equation

$$\not{D}\Psi_0 = \frac{H_0}{2}\Psi_0 \quad \text{and} \quad |\Psi_0| = 1.$$

It is obtained by taking the restriction to Σ of a parallel spinor field on $\mathbb{R}^3$. Now taking Ψ_0 in inequality (25) with $X \equiv 0$ and $F = H$ gives $m(\Sigma) \geq 0$. Moreover, from the same reference, $m(\Sigma) = 0$ if and only if M is a Euclidean domain and the embedding of Σ in M and its immersion in $\mathbb{R}^3$ are congruent. In other words, properties (I) and (II) are satisfied.

Note that if we assume that Σ has positive Gauss curvature (which is a stronger assumption) then using the Cauchy–Schwarz inequality implies that $m(\Sigma) \geq m_{BY}(\Sigma)$, and the nonnegativity of $m(\Sigma)$ follows from the nonnegativity of the Brown–York mass. On the other hand, it is also proved in [Hijazi and Montiel 2014, Proof of Corollary 10] that (IV) holds. However it is clear from the definition that the mass $m(\Sigma)$ is not defined for minimal surfaces (and so for apparent horizons). Moreover the monotonicity property (III) is not satisfied in general. Take for example the 3-dimensional Schwarzschild manifold $(N^3, g) = (\mathbb{R}^3 \setminus \{0\}, u^4 g_{\text{eucl}})$, where $u := 1 + M/2r$, $M > 0$, and g_{eucl} is the Euclidean metric. For a sphere $\mathbb{S}^2_r$ in N^3, its isometric image in $\mathbb{R}^3$ is $\mathbb{S}^2_{ru^2}$. Thus $H_0 = 2/ru^2$ and since the Schwarzschild metric is conformal to the Euclidean metric,

$$H = u^{-2}\left(\frac{2}{r} + \frac{4}{u}\frac{\partial u}{\partial r}\right).$$

A direct computation gives

$$m(\mathbb{S}^2_r) = M\frac{r + M/2}{r - M/2},$$

and so $m(\mathbb{S}^2_r)$ is monotonically *decreasing* to the ADM mass M as r goes to infinity.

3. Spacelike surfaces in initial data sets

3.1. ***The Jang equation.*** In this section, we recall some well-known facts about the Jang equation (for more details, we refer to [Schoen and Yau 1981], [Yau 2001] or [Andersson et al. 2011]). This equation first was used by Jang [1978] in his attempt to prove the positive mass theorem using the inverse mean curvature flow. However, as shown by Schoen and Yau [1981], this equation can be used to reduce the proof of the general positive mass theorem to the case of time-symmetric initial data sets (that is, $K_{ij} = 0$) previously obtained by the same authors [1979]. More recently, Liu and Yau [2003; 2006] defined a quasilocal mass, generalizing the Brown–York quasilocal mass, and proved its positivity using the Jang equation. Other similar applications of the Jang equation can be found in, for example, [Wang and Yau 2007; 2009].

The problem can be stated as follows: Let (M^3, g, K) be an initial data set for the Einstein equation and consider the four-dimensional manifold $M \times \mathbb{R}$ equipped with the *Riemannian* metric $\langle\cdot\,,\cdot\rangle := g \oplus dt^2$. The problem is to find a smooth function $u : M \to \mathbb{R}$ such that the hypersurface $\widehat{M}$ of $M \times \mathbb{R}$ obtained by taking the graph of u over M satisfies the equation

$$H_{\widehat{M}} = \mathrm{tr}_{\widehat{M}}(K),$$

where $H_{\widehat{M}}$ denotes the mean curvature of $\widehat{M}$ in $(M \times \mathbb{R}, \langle\cdot\,,\cdot\rangle)$ and $\mathrm{tr}_{\widehat{M}}(\cdot)$ is the trace on $\widehat{M}$ with respect to the induced metric. This geometric problem is equivalent to solving the nonlinear second-order elliptic equation

$$\sum_{i,j=1}^{3}\Big(g^{ij} - \frac{u^i u^j}{1+|\nabla u|^2}\Big)\Big(\frac{(\nabla^2 u)_{ij}}{\sqrt{1+|\nabla u|^2}} - K_{ij}\Big) = 0, \tag{27}$$

where ∇ (respectively, ∇^2) denotes the Levi-Civita connection (respectively, the Hessian) of the metric g, $u^i = g^{ij}u_j$ and $u_j = e_j(u)$. Note that the metric induced by $\langle\cdot\,,\cdot\rangle$ on $\widehat{M}$ is

$$\hat{g}_{ij} = g_{ij} + u_i u_j$$

and can be viewed as a deformation of the metric g on M. In the following, we adopt the convention that M and $\widehat{M}$ denote, respectively, the Riemannian manifolds (M, g) and $(M, \hat{g})$. Analogously, if ∇ denotes the Levi-Civita connection for M, then $\widehat{\nabla}$ denotes that on $\widehat{M}$ and so on. Since we assume that the initial data set (M^3, g, K) comes from a spacetime satisfying the dominant energy condition, we have that the relation

$$0 \le 2(\mu - |J|) \le \widehat{R} - 2|X|^2_{\hat{g}} - 2\hat{\delta}(X) \tag{28}$$

holds on $\widehat{M}$, where

$$X = \omega - \widehat{\nabla} \log(f), \tag{29}$$

ω is the tangent part of the vector field dual to $-K(\cdot, \hat{\nu})$, $f = -\langle \partial_t, \hat{\nu}\rangle$ and $\hat{\nu}$ denotes the unit normal vector field to $\widehat{M}$ in $M \times \mathbb{R}$. All the quantities K_{ij}, μ and J are defined on $M \times \mathbb{R}$ by parallel transport along the $\mathbb{R}$-factor. Moreover equality occurs in (28) if and only if $\mu = |J|$ and the second fundamental form of $\widehat{M}$ in $M \times \mathbb{R}$ is K.

It is important to note here that in Theorem 2 we assume that there is no apparent horizon in the interior of Ω so that there exists a global solution of the Jang equation which does not blow up.

3.2. *Proof of Theorem 2.* From [Yau 2001], and since we assumed that Ω has no apparent horizon in its interior, there exists a smooth solution u on Ω of the Jang equation (27), defined with the Dirichlet boundary condition

$$u_{|\Sigma} \equiv 0.$$

This boundary condition ensures that the metrics $\hat{g}$ and g coincide on the boundary Σ so that the Dirac operators $\not{D}$ acting on $\$\Sigma$ and $\widehat{\not{D}}$ on $\widehat{\$}\Sigma$ also coincide. Moreover, from a calculation in the same work,

$$\widehat{H} - \hat{g}(X, \widehat{N}) = f^{-1} H - \sigma |\nabla u| \operatorname{tr}_\Sigma(K),$$

where $\widehat{N}$ denotes the unit outward normal vector field of Σ in $\widehat{\Omega}$ and $\sigma \in \{\pm 1\}$. From this equality and since $f = -\langle \partial_t, \hat{\nu}\rangle = 1/\sqrt{1 + |\nabla u|^2}$, we easily see that

$$F := \widehat{H} - \hat{g}(X, \widehat{N}) \ge |\mathcal{H}| = \sqrt{H^2 - \operatorname{tr}_\Sigma(K)^2}. \tag{30}$$

Since we assume that Σ has a spacelike mean curvature vector $\mathcal{H}$, this implies that the function F is positive on Σ. From the discussion of Section 3.1, we also have that the resulting Riemannian manifold $\widehat{\Omega}$ satisfies the condition (16) because of (28), the vector field X being defined here by (29). Clearly all the assumptions of Theorem 10 are fulfilled and we deduce that for all $\varphi \in \Gamma(\$\Sigma)$,

$$0 \le \int_\Sigma \left(\frac{1}{F} |\not{D}\varphi|^2 - \frac{F}{4} |\varphi|^2 \right) d\Sigma,$$

which by inequality (30) implies inequality (2).

Now assume that equality is achieved. Once again we apply Theorem 10, and then $\widehat{\Omega}$ has at least a parallel spinor field Φ. In particular, $\widehat{\Omega}$ is Ricci-flat, and since it is a 3-dimensional domain, it is flat. Moreover, if we have equality in (28), then the second fundamental form of $\widehat{\Omega}$ in $M \times \mathbb{R}$ is K_{ij}. So we can choose a coordinate system $\widehat{x} = (\widehat{x}_1, \widehat{x}_2, \widehat{x}_3)$ in a neighborhood $\mathcal{U}$ of a point $p \in \Omega$ such that $\hat{g}_{ij} = \delta_{ij}$. In this chart,

$$g_{ij} = \delta_{ij} - \frac{\partial u}{\partial \widehat{x}_i} \frac{\partial u}{\partial \widehat{x}_j},$$

and this shows that if $(\widehat{x}_1, \widehat{x}_2, \widehat{x}_3, t)$ denotes coordinates in the Minkowski spacetime, the graph of u over $\mathcal{U}$ isometrically embeds in $\mathbb{R}^{3,1}$ with second fundamental form given by K_{ij}. Then it is clear that Ω locally embeds in the Minkowski spacetime with K as second fundamental form as asserted. □

As a first consequence, we have the estimate proved by Raulot [2013] for the first eigenvalue of the Dirac operator on Σ.

Corollary 13. *Under the same conditions of Theorem 2, the first eigenvalue $\lambda_1(D_\Sigma)$ of the Dirac operator satisfies*

$$\lambda_1(D_\Sigma)^2 \geq \tfrac{1}{4} \inf_\Sigma |\mathcal{H}|^2.$$

Moreover, if equality occurs, then Σ is connected and there exists a local isometric embedding of Ω as a spacelike hypersurface in $\mathbb{R}^{3,1}$ with K as second fundamental form.

Proof. The inequality on $\lambda_1(D_\Sigma)$ follows directly by taking $\varphi = \Phi \in \Gamma(\$\Sigma)$ in (2), where Φ is an eigenspinor for the Dirac operator $\not{D}$ associated with the eigenvalue $\lambda_1(\not{D})$ (which equals $\lambda_1(D_\Sigma)$). On the other hand, the second part of the equality case follows directly from Theorem 2. For the connectedness of Σ, it is enough to remark that, from [Hijazi et al. 2001a], the eigenspace associated to $\lambda_1(\not{D})$ corresponds to the restriction to Σ of the space of parallel spinor fields on the domain $\widehat{\Omega}$ obtained by solving the Jang equation. Then, assuming that Σ has several connected components, we fix one of them, say Σ_0, and define a spinor field on Σ by

$$\widetilde{\Phi} = \begin{cases} \Phi_0 & \text{on } \Sigma_0, \\ 0 & \text{on } \Sigma - \Sigma_0, \end{cases}$$

where Φ_0 is an eigenspinor for the extrinsic Dirac operator $\not{D}$ associated to the eigenvalue $\lambda_1(\not{D})$. It is then straightforward to check that $\widetilde{\Phi}$ is also an eigenspinor associated to $\lambda_1(\not{D})$ so that it comes from the restriction of a parallel spinor on $\widehat{\Omega}$. However, since such a spinor field has constant norm, it is impossible unless Σ is connected. □

Proof of Theorem 3. In order to establish inequality (3) it is sufficient to apply inequality (2) to the restriction to Σ of a parallel spinor field on $\mathbb{R}^3$. From the equality case of Theorem 2, we deduce that Ω locally embeds in the Minkowski spacetime with K as a second fundamental form. On the other hand, we have equality in (30) so that $\widehat{H} = |\mathcal{H}|$, and then equality in (3) now reads

$$\int_\Sigma \left(\widehat{H} - \frac{H_0^2}{\widehat{H}} \right) d\Sigma = 0.$$

We conclude by applying the rigidity part of [Hijazi and Montiel 2014, Theorem 3] to the compact Ricci-flat manifold $\widehat{\Omega}$ to deduce that Σ is connected and $|\mathcal{H}| = H_0$. □

3.3. ***Codimension-two outer untrapped submanifolds in the Minkowski spacetime.*** In this section, we prove that inequality (2) holds in the case of codimension-two outer untrapped submanifolds of the Minkowski spacetime without any assumption on the existence of apparent horizon. More precisely, we prove:

Theorem 14. *Let Σ^n be a codimension-two outer untrapped submanifold of the $(n+2)$-dimensional Minkowski spacetime $(\mathbb{R}^{n+1,1}, \langle \cdot , \cdot \rangle)$. Then inequality* (2) *holds. Moreover equality holds if and only if Σ lies in a hyperplane of $\mathbb{R}^{n+1,1}$.*

Proof. First we note that by assumption Σ factorizes through a compact and connected spacelike hypersurface Ω of $\mathbb{R}^{n+1,1}$. This factorization provides us a Lorentzian orthonormal reference $\{T, N\}$ for the normal plane of Σ in $\mathbb{R}^{n+1,1}$, and, since Σ is the boundary of a mean-convex domain Ω and has spacelike mean curvature vector, we deduce that the corresponding future-directed null expansions satisfy $\theta_+ > 0$ and $\theta_- < 0$. On the other hand, from the work of Bartnik and Simon [1982] and a straightforward generalization in [Miao et al. 2010, Lemma 4.1], the submanifold Σ spans a compact, smoothly immersed, maximal hypersurface Ω' in $\mathbb{R}^{n+1,1}$. This means that Σ factorizes through another spacelike hypersurface Ω' of $\mathbb{R}^{n+1,1}$. The new factorization provides us a different Lorentzian orthonormal reference $\{T', N'\}$ for the normal plane of Σ in $\mathbb{R}^{n+1,1}$. In fact, it is obvious that there must be a function $f \in C^\infty(\Sigma)$ such that

$$T' = (\cosh f)T - (\sinh f)N \quad \text{and} \quad N' = -(\sinh f)T + (\cosh f)N.$$

It is clear that this new reference determines a new pair of null vectors $T' \pm N'$ and a new future-directed null expansion of $\mathcal{H}$

$$\theta'_+ = e^f \theta_+ \quad \text{and} \quad \theta'_- = e^{-f} \theta_-, \tag{31}$$

which satisfies $\theta'_+ > 0$ and $\theta'_- < 0$. In particular, we get that $2H' = \theta'_+ - \theta'_- > 0$. Moreover, since Ω' is maximal, we have $\operatorname{tr}(K') = 0$, and the Gauss formula gives $R' = |K'|^2 \geq 0$. Here R' is the scalar curvature of Ω' equipped with the metric induced by the Minkowski spacetime, and K' is the associated second fundamental form. On the other hand, since Σ has a spacelike mean curvature vector, we deduce

$$0 < |\mathcal{H}| = \sqrt{-\theta'_+\theta'_-} = \sqrt{H'^2 - \mathrm{tr}_\Sigma(K')^2} \le H', \tag{32}$$

so we conclude that Ω' is such that $R' \ge 0$ and $H' > 0$. Now we can apply Theorem 10 to Ω' with $X \equiv 0$, and then for all $\varphi \in \Gamma(\$\Sigma)$,

$$0 \le \int_\Sigma \left(\frac{1}{H'} |\not{D}\varphi|^2 - \frac{1}{4} H' |\varphi|^2 \right) d\Sigma. \tag{33}$$

Inequality (2) follows using inequality (32). Assume now that equality is achieved. From the equality case of (33), we deduce that Ω' has at least a parallel spinor so that Ω' is Ricci-flat. In particular, it has zero scalar curvature, and since $R' = |K'|^2 = 0$, Ω' has to be totally geodesic in $\mathbb{R}^{n+1,1}$, hence Σ lies in a hyperplane of $\mathbb{R}^{n+1,1}$. Conversely, if Σ is a codimension-two submanifold with spacelike mean curvature vector which lies in a hyperplane $\mathbb{R}^{n+1,1}$, then its second fundamental form K is zero since a hyperplane P^{n+1} is totally geodesic. In particular, the squared norm of the mean curvature vector of Σ satisfies

$$|\mathcal{H}|^2 = H^2 - \mathrm{tr}_\Sigma(K)^2 = H^2, \tag{34}$$

where H is the mean curvature of Σ in the hyperplane P. Note that $|\mathcal{H}| > 0$ since $H > 0$. Consider now a parallel spinor field Φ_0 on $\mathbb{R}^{n+1,1}$. The spinorial Gauss formula from the totally geodesic immersion of the hyperplane P^{n+1} in $\mathbb{R}^{n+1,1}$ and then the one from Σ^n into P^{n+1} tell us that Φ_0 satisfies

$$\not{\nabla}_Y \Phi_0 = -\tfrac{1}{2} \gamma\!\!/(AY)\Phi_0$$

for all $Y \in \Gamma(T\Sigma)$, where A is the Weingarten map of Σ^n in P^{n+1}. Taking the trace of this identity gives

$$\not{D}\Phi_0 = \tfrac{1}{2} H \Phi_0 = \tfrac{1}{2} |\mathcal{H}| \Phi_0,$$

where the last equality comes from (34). It is now straightforward to check that equality holds in (2) for $\varphi = \Phi_0$. □

Note that Theorem 4 is obtained as a direct application of the previous result. As an application we obtain the n-dimensional counterpart of Corollary 13 in the Minkowski spacetime with an optimal rigidity statement:

Corollary 15. *Let Σ^n be a codimension-two outer untrapped submanifold in $\mathbb{R}^{n+1,1}$. Then*

$$|\lambda_1(D_\Sigma)| \ge \tfrac{1}{2} \inf_\Sigma |\mathcal{H}|.$$

Moreover equality occurs if and only if Σ is a totally umbilical round sphere in a spacelike hyperplane of $\mathbb{R}^{n+1,1}$.

Proof. It is enough to apply the previous theorem to an eigenspinor for $\not{D}$ associated with the eigenvalue $\lambda_1(\not{D})$, and we directly have the result. From Theorem 14, Σ lies in a totally geodesic spacelike hyperplane P^{n+1} with constant positive mean curvature H. Then the Alexandrov theorem allows to conclude that Σ is a totally umbilical sphere in P^{n+1}. The converse is clear by taking the restriction of a parallel spinor of the Minkowski space to Σ via the totally geodesic immersion of $\mathbb{R}^{n+1}$ in $\mathbb{R}^{n+1,1}$. □

References

[Almgren and Rivin 1998] F. J. Almgren, Jr. and I. Rivin, "The mean curvature integral is invariant under bending", pp. 1–21 in *The Epstein birthday schrift*, edited by I. Rivin et al., Geom. Topol. Monogr. **1**, University of Warwick, Institute of Mathematics, Warwick, 1998. MR 2000a:53133 Zbl 0914.53007 arXiv math/9810183

[Andersson et al. 2011] L. Andersson, M. Eichmair, and J. Metzger, "Jang's equation and its applications to marginally trapped surfaces", pp. 13–45 in *Complex analysis and dynamical systems IV, Part 2: General relativity, geometry, and PDE* (Nahariya, 2009), edited by M. Agranovsky et al., Contemp. Math. **554**, Amer. Math. Soc., Providence, RI, 2011. MR 2884392 Zbl 1235.53019

[Ballmann and Bär 2012] W. Ballmann and C. Bär, "Boundary value problems for elliptic differential operators of first order", pp. 1–78 in *Algebra and geometry: in memory of C. C. Hsiung* (Bethlehem, PA, 2010), edited by H.-D. Cao and S.-T. Yau, Surv. Differ. Geom. **17**, International Press, Boston, 2012. MR 3076058

[Bär 1998] C. Bär, "Extrinsic bounds for eigenvalues of the Dirac operator", *Ann. Global Anal. Geom.* **16**:6 (1998), 573–596. MR 99k:58183 Zbl 0921.58065

[Bartnik and Chruściel 2005] R. A. Bartnik and P. T. Chruściel, "Boundary value problems for Dirac-type equations", *J. Reine Angew. Math.* **579** (2005), 13–73. MR 2005k:58040 Zbl 1174.58305

[Bartnik and Simon 1982] R. A. Bartnik and L. Simon, "Spacelike hypersurfaces with prescribed boundary values and mean curvature", *Comm. Math. Phys.* **87**:1 (1982), 131–152. MR 84j:58126 Zbl 0512.53055

[Baum et al. 1990] H. Baum, T. Friedrich, R. Grunewald, and I. Kath, *Twistor and Killing spinors on Riemannian manifolds*, Seminarberichte **108**, Humboldt Universität, Sektion Mathematik, Berlin, 1990. MR 92h:53055 Zbl 0705.53004

[Benedetti and Silhol 1995] R. Benedetti and R. Silhol, "Spin and Pin$^-$ structures, immersed and embedded surfaces and a result of Segre on real cubic surfaces", *Topology* **34**:3 (1995), 651–678. MR 96f:57033 Zbl 0996.57519

[Booß-Bavnbek and Wojciechowski 1993] B. Booß-Bavnbek and K. P. Wojciechowski, *Elliptic boundary problems for Dirac operators*, Birkhäuser, Boston, 1993. MR 94h:58168 Zbl 0797.58004

[Brown and York 1993] J. D. Brown and J. W. York, Jr., "Quasilocal energy and conserved charges derived from the gravitational action", *Phys. Rev. D* (3) **47**:4 (1993), 1407–1419. MR 94h:83046

[Bureš 1993] J. Bureš, "Dirac operators on hypersurfaces", *Comment. Math. Univ. Carolin.* **34**:2 (1993), 313–322. MR 95g:58215 Zbl 0781.53031

[Eichmair et al. 2012] M. Eichmair, P. Miao, and X. Wang, "Extension of a theorem of Shi and Tam", *Calc. Var. Partial Differential Equations* **43**:1-2 (2012), 45–56. MR 2012m:53075 Zbl 1238.53025

[Hass and Hughes 1985] J. Hass and J. Hughes, "Immersions of surfaces in 3-manifolds", *Topology* **24**:1 (1985), 97–112. MR 87a:57037 Zbl 0527.57020

[Herglotz 1943] G. Herglotz, "Über die Starrheit der Eiflächen", *Abh. Math. Sem. Hansischen Univ.* **15** (1943), 127–129. MR 7,322a Zbl 0028.09401

[Hijazi and Montiel 2014] O. Hijazi and S. Montiel, "A holographic principle for the existence of parallel spinor fields and an inequality of Shi–Tam type", *Asian J. Math* **18**:3 (2014), 489–506.

[Hijazi et al. 2001a] O. Hijazi, S. Montiel, and X. Zhang, "Dirac operator on embedded hypersurfaces", *Math. Res. Lett.* **8**:1-2 (2001), 195–208. MR 2002j:53053 Zbl 0988.53019

[Hijazi et al. 2001b] O. Hijazi, S. Montiel, and X. Zhang, "Eigenvalues of the Dirac operator on manifolds with boundary", *Comm. Math. Phys.* **221**:2 (2001), 255–265. MR 2002g:58051 Zbl 0997.58015

[Hijazi et al. 2002] O. Hijazi, S. Montiel, and X. Zhang, "Conformal lower bounds for the Dirac operator of embedded hypersurfaces", *Asian J. Math.* **6**:1 (2002), 23–36. MR 2003i:58058 Zbl 1016.58013

[Jang 1978] P. S. Jang, "On the positivity of energy in general relativity", *J. Math. Phys.* **19**:5 (1978), 1152–1155. MR 80b:83012a

[Lam 2011] G. Lam, *The graph cases of the Riemannian positive mass and Penrose inequalities in all dimensions*, thesis, Duke University, Durham, NC, 2011, Available at http://www.math.duke.edu/~bray/PhD%20Theses/Lam.pdf.

[Lawson and Michelsohn 1989] H. B. Lawson, Jr. and M.-L. Michelsohn, *Spin geometry*, Princeton Mathematical Series **38**, Princeton University Press, 1989. MR 91g:53001 Zbl 0688.57001

[Liu and Yau 2003] C.-C. M. Liu and S.-T. Yau, "Positivity of quasilocal mass", *Phys. Rev. Lett.* **90**:23 (2003), Article ID #231102. MR 2004g:83031 Zbl 1267.83028

[Liu and Yau 2006] C.-C. M. Liu and S.-T. Yau, "Positivity of quasi-local mass, II", *J. Amer. Math. Soc.* **19**:1 (2006), 181–204. MR 2007f:83032 Zbl 1081.83008

[Miao et al. 2010] P. Miao, Y. Shi, and L.-F. Tam, "On geometric problems related to Brown–York and Liu–Yau quasilocal mass", *Comm. Math. Phys.* **298**:2 (2010), 437–459. MR 2011h:58022 Zbl 1200.58018

[Montiel and Ros 1997] S. Montiel and A. Ros, *Curvas y superficies*, Proyecto Sur, Armilla, 1997. Translated as *Curves and surfaces*, 2nd ed., Graduate Studies in Mathematics **69**, Amer. Math. Soc., Providence, RI, 2009. MR 2010f:53007 Zbl 1075.53001

[Pinkall 1985] U. Pinkall, "Regular homotopy classes of immersed surfaces", *Topology* **24**:4 (1985), 421–434. MR 87e:57028 Zbl 0583.57020

[Raulot 2013] S. Raulot, "The Dirac operator on untrapped surfaces", *Comm. Math. Phys.* **318**:2 (2013), 411–427. MR 3020163 Zbl 06145993

[Rivin and Schlenker 1999] I. Rivin and J.-M. Schlenker, "The Schläfli formula in Einstein manifolds with boundary", *Electron. Res. Announc. Amer. Math. Soc.* **5** (1999), 18–23. MR 2000a:53076 Zbl 0910.53036

[Schoen and Yau 1979] R. Schoen and S. T. Yau, "On the proof of the positive mass conjecture in general relativity", *Comm. Math. Phys.* **65**:1 (1979), 45–76. MR 80j:83024 Zbl 0405.53045

[Schoen and Yau 1981] R. Schoen and S. T. Yau, "Proof of the positive mass theorem, II", *Comm. Math. Phys.* **79**:2 (1981), 231–260. MR 83i:83045 Zbl 0494.53028

[Shi and Tam 2002] Y. Shi and L.-F. Tam, "Positive mass theorem and the boundary behaviors of compact manifolds with nonnegative scalar curvature", *J. Differential Geom.* **62**:1 (2002), 79–125. MR 2005b:53046 Zbl 1071.53018

[Szabados 2004] L. B. Szabados, “Quasi-local energy-momentum and angular momentum in general relativity”, *Living Rev. Relativity* **7**:4 (2004).

[Trautman 1995] A. Trautman, “The Dirac operator on hypersurfaces”, *Acta Phys. Polon. B* **26**:7 (1995), 1283–1310. MR 96m:58236 Zbl 0966.58504

[Wang and Yau 2007] M.-T. Wang and S.-T. Yau, “A generalization of Liu–Yau’s quasi-local mass”, *Comm. Anal. Geom.* **15**:2 (2007), 249–282. MR 2008h:53046 Zbl 1171.53336

[Wang and Yau 2009] M.-T. Wang and S.-T. Yau, “Isometric embeddings into the Minkowski space and new quasi-local mass”, *Comm. Math. Phys.* **288**:3 (2009), 919–942. MR 2010d:53077 Zbl 1195.53039

[Weyl 1916] H. Weyl, “Über die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement”, *Vierteljahrsschr. Naturf. Ges.* (*Zürich*) **61** (1916), 40–72. JFM 46.1115.03

[Witten 1981] E. Witten, “A new proof of the positive energy theorem”, *Comm. Math. Phys.* **80**:3 (1981), 381–402. MR 83e:83035 Zbl 1051.83532

[Yau 2001] S. T. Yau, “Geometry of three manifolds and existence of black hole due to boundary effect”, *Adv. Theor. Math. Phys.* **5**:4 (2001), 755–767. MR 2003j:53052 Zbl 1019.53016

Received June 30, 2013. Revised September 17, 2013.

OUSSAMA HIJAZI
INSTITUT ÉLIE CARTAN DE LORRAINE
UNIVERSITÉ DE LORRAINE
NANCY I, BOÎTE POSTALE 239
54506 VANDOEUVRE-LÈS-NANCY CEDEX
FRANCE
oussama.hijazi@univ-lorraine.fr

SEBASTIÁN MONTIEL
DEPARTAMENTO DE GEOMETRÍA Y TOPOLOGÍA
UNIVERSIDAD DE GRANADA
18071 GRANADA
SPAIN
smontiel@ugr.es

SIMON RAULOT
LABORATOIRE DE MATHÉMATIQUES RAPHAËL SALEM UMR 6085 CNRS
UNIVERSITÉ DE ROUEN AVENUE DE L’UNIVERSITÉ
BOÎTE POSTALE 12 TECHNOPÔLE DU MADRILLET
76801 SAINT-ÉTIENNE-DU-ROUVRAY
FRANCE
simon.raulot@univ-rouen.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.272.201

NONLINEAR EULER SUMS

ISTVÁN MEZŐ

We work out some formulas for nonlinear Euler sums involving multiple zeta values. As applications of these formulas, we give new closed form sums of several nonlinear Euler series, we present sums for powers of the digamma function and deduce the Landen identities for the polylogarithms by finite combinatorial identities.

1. Introduction

In a letter from Goldbach to Euler, Goldbach proposed to investigate infinite series of the form

$$\sum_{n=1}^{\infty}\frac{1}{n^a}\sum_{k=1}^{n}\frac{1}{k^b}.$$

See for the historical details. In 1742 and 1743 Euler presented a number of closed form expression for such sums and their variations. The most fundamental one is the following [Borwein and Bradley 2006]:

$$\sum_{n=1}^{\infty}\frac{1}{n^2}\sum_{k=1}^{n-1}\frac{1}{k}=\sum_{n=1}^{\infty}\frac{1}{n^3}=8\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\sum_{k=1}^{n-1}\frac{1}{k}. \tag{1}$$

The research of István Mező was supported by the Scientific Research Foundation of Nanjing University of Information Science and Technology.
MSC2010: 33B15, 11M32.
Keywords: Euler sums, nonlinear Euler sums, polygamma functions, polylogarithm functions, Landen's identities, multiple zeta values.

The sum in the middle is a zeta function value and in the present day we consider the value of this series as a "fundamental constant", which cannot be traced back to "more fundamental" ones.

In the past two hundred years it had been apparent that the above sums and their generalizations — nowadays they are called *Euler sums* — often can be traced back to zeta function values. To treat these sums, we adopt the modern notations and notions. The *multiple polylogarithm* is defined by

$$\zeta(s_1, s_2, \dots, s_m; z) = \sum_{0<n_m<\cdots<n_1} \frac{z^{n_1}}{n_1^{s_1} n_2^{s_2} \cdots n_m^{s_m}}, \tag{2}$$

with the appropriate restriction on the powers to get a convergent series. In particular,

$$\zeta(s; 1) = \zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \quad (\Re(s) > 1)$$

is the classical *Riemann zeta function* [Andrews et al. 1999]. Typically z is set to 1 or -1, in which cases we are dealing with a *multiple zeta function* or *alternating multiple zeta function*, respectively. We remark that in the literature there exists a more general version of the above multiple zeta function, called the *colored multiple zeta function* [Bigotte et al. 2002]. It is defined as

$$\zeta(s_1, s_2, \dots, s_m; \sigma_1, \sigma_2, \dots, \sigma_m) = \sum_{0<n_m<\cdots<n_1} \frac{\sigma_1^{n_1} \sigma_2^{n_2} \cdots \sigma_m^{n_m}}{n_1^{s_1} n_2^{s_2} \cdots n_m^{s_m}}.$$

The sum

$$\sum_{i=1}^{m} s_i$$

is the *weight* of the zeta function, while m is the *depth*. A brief survey on multiple polylogarithms can be found in [Bowman and Bradley 2001].

The finite sums inside the sums are called *generalized harmonic numbers* and are denoted by $H_{n,r}$ (or $H_n^{(r)}$, but we use the former, because our expressions will involve powers):

$$H_{n,r} = \sum_{k=1}^{n} \frac{1}{k^r} \quad (n \ge 1,\ r \ge 1),$$

with the convention $H_{0,r} = 0$ for all $r = 1, 2, \dots$. The numbers $H_{n,1} =: H_n$ are called *harmonic numbers*.

With these, the above relations under (1) can be written in the short form

$$\zeta(2, 1) := \sum_{n=1}^{\infty} \frac{H_{n-1}}{n^2} = \zeta(3) = 8\zeta(2, 1; -1).$$

For Euler's original proof, see [Euler 1776]. These relations were rediscovered many times, as the references [Briggs et al. 1955; Bruckman 1982; Farnum and Tissier 1999; Klamkin and Steinberg 1952] show.

We mention that the general expression in terms of zeta values of the sum

$$\sum_{n=1}^{\infty} \frac{H_n}{n^a}$$

was known already to Euler, who found that

$$\sum_{n=1}^{\infty} \frac{H_n}{n^a} = \Bigl(1+\frac{a}{2}\Bigr)\zeta(a+1) - \frac{1}{2}\sum_{k=1}^{a-2} \zeta(k+1)\zeta(a-k) \quad (a \geq 2). \tag{3}$$

Naturally, then, researchers after Euler have turned to generalizations and alterations of these sums. In the next section we present some existing directions, then we show in which direction we proceed.

2. Existing results and research directions

2.1. ***Alternating Euler sums.*** In the past and present, the alternating Euler sums and their modifications and generalizations have attracted the attention of a large number of mathematicians. For example, the alternating Euler sums, like

$$\sum_{n=1}^{\infty} \frac{1}{n^a} \sum_{k=1}^{n} \frac{(-1)^k}{k^r},$$

are investigated in [Bailey et al. 1994; de Doelder 1991; Li 2011; Sitaramachandra Rao 1987], to name a few. We mention one sum from [Li 2011, Proposition 3.2]:

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} = \frac{3}{2}\zeta(2)\log 2 - \zeta(3).$$

In [Sitaramachandra Rao 1987] one can find an exhaustive bibliography on alternating Euler sums. It turns out that these sums are reducible to zeta values in many cases, see [Flajolet and Salvy 1998, Theorem 7.1].

2.2. ***Analytic extension of Euler sums.*** T. Apostol and T. H. Vu [1984] started to investigate Euler sums as functions of the power of n:

$$h(s) = \sum_{n=1}^{\infty} \frac{H_n}{n^s}.$$

They showed that this function can be continued to the whole s-plane as a meromorphic function with a second-order pole at $s = 1$, and simple poles at $s = 0$ and at the negative odd integers.

In the same paper this result was extended to the function

$$h(s,z)=\sum_{n=1}^{\infty}\frac{H_{n,z}}{n^s}.$$

These results were further specified and extended by Boyadzhiev [2008; 2009] to

$$\begin{aligned}
u(s)&=\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^s}H_n,\\
v(s)&=\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^s}\left(1-\frac{1}{2}+\frac{1}{3}+\cdots+\frac{(-1)^{n+1}}{n}\right),\\
w(s)&=\sum_{n=1}^{\infty}\frac{1}{n^s}\left(1-\frac{1}{2}+\frac{1}{3}+\cdots+\frac{(-1)^{n+1}}{n}\right).
\end{aligned}$$

2.3. *Nonlinear Euler sums up to now.* Another direction of investigation is the nonlinear case. In this case one considers sums like

$$\sum_{n=1}^{\infty}\frac{H_{n,r_1}H_{n,r_2}\cdots H_{n,r_p}}{n^a}. \tag{4}$$

These are called *nonlinear Euler sums*. In this case just sporadic results are known; one can find some of them in the references [Borwein and Borwein 1995; Chu 1997; de Doelder 1991; Shen 1995; Sofo and Hassani 2012]. Moreover, V. Adamchik [1997] investigated the relation between such nonlinear Euler sums and several sums on the Stirling numbers of the fist kind. D. F. Connon [2008a; 2008b; 2008c; 2008d; 2008e; 2008f; 2008g; 2008h] has found a large number of connections between specific nonlinear Euler sums and the Riemann and Hurwitz zeta functions. To mention two beautiful results, we cite an expression for $\zeta(4)$ and $\zeta(5)$ [Connon 2008c, formulas (4.3.45f) and (4.3.57b)]:

$$\zeta(4)=\frac{1}{6}\sum_{n=1}^{\infty}\frac{H_n^2+H_{n,2}}{n^2},\qquad \zeta(5)=\frac{1}{24}\sum_{n=1}^{\infty}\frac{2H_{n,3}+3H_nH_{n,2}+H_n^3}{n^2}.$$

To present some additional examples from the literature, we cite two sums from [Borwein and Borwein 1995]:

$$\sum_{n=1}^{\infty}\frac{H_n^2}{n^2}=\frac{17}{4}\zeta(4),\qquad \sum_{n=1}^{\infty}\frac{H_nH_{n+1}}{(n+1)^2}=3\zeta(4)=\frac{\pi^4}{30},$$

and from [Flajolet and Salvy 1998]:

$$\sum_{n=1}^{\infty} \frac{H_n^3}{n^4} = \frac{231}{16}\zeta(7) - \frac{51}{4}\zeta(3)\zeta(4) + 2\zeta(2)\zeta(5),$$

$$\sum_{n=1}^{\infty} \frac{H_n^4}{(n+1)^3} = \frac{185}{8}\zeta(7) - \frac{43}{2}\zeta(3)\zeta(4) + 5\zeta(2)\zeta(5).$$

That all of these sums can be expressed as multiple zeta values is not known. The only available result concerns quadratic sums:

Theorem 1 [Flajolet and Salvy 1998, p. 25, Theorem 4.2]. *If $p_1 + p_2 + q$ is even, and $p_1 > 1$, $p_2 > 1$, $q > 1$, the quadratic sums*

$$\sum_{n=1}^{\infty} \frac{H_{n,p_1} H_{n,p_2}}{n^q}$$

are reducible to linear sums.

The theorem exactly gives the reduction, but the formulas are rather complicated to cite.

Finally, we cite a nice example of a nonlinear alternating sum from [Borwein and Borwein 1995]:

$$\sum_{n=1}^{\infty} \frac{1}{n^2}\left(1 - \frac{1}{2} + \cdots + \frac{(-1)^{n+1}}{n}\right)^2 = -\frac{13}{8}\zeta(4) + \frac{5}{2}\zeta(2)\log^2 2 + \frac{1}{12}\log^4 2 + 2\,\mathrm{Li}_4\left(\tfrac{1}{2}\right).$$

Here

$$\mathrm{Li}_k(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^k} = \zeta(k; z) \tag{5}$$

is a special multiple zeta function, called the *polylogarithm*. (The special value $\mathrm{Li}_4\left(\frac{1}{2}\right)$, like $\zeta(3)$, does not seem to be evaluable in terms of more fundamental constants.)

No general reduction formula is known for nonlinear alternating sums, but computer-based calculations are available in several cases; see [Bailey et al. 1994].

An exhaustive and up-to-date bibliography on Euler sums and their generalizations can be found at http://www.usna.edu/Users/math/meh/biblio.html.

3. New nonlinear Euler sum formulas

Now we turn to our own results.

In this section we demonstrate how we can trace back some specific quadratic Euler sums to linear ones. To express these sums in a convenient form, we use the concept of E. A. Ulanskii, who defined the *nonstrict multiple polylogarithm* [2003] as follows:

$$\mathrm{Le}_{(s_1,\ldots,s_m)}(z) = \sum_{1\le n_m\le\cdots\le n_1} \frac{z^{n_1}}{n_1^{s_1} n_2^{s_2}\cdots n_m^{s_m}}.$$

As it can be seen, this definition differs from the multiple polylogarithm function in the nonstrictness of the relations. We shall always set the parameter z to 1, and we refer to the function

$$\mathrm{Le}_{(s_1,\ldots,s_m)}(1) =: \mathrm{Le}(s_1,\ldots,s_m)$$

as the *nonstrict multiple zeta function*. Ulanskii did not deal with the specific values of these sums but with the functional relations among them. For example, he proved the following theorem, which will be extremely useful for us.

Theorem 2. *The nonstrict multiple polylogarithm function can be written as a sum of multiple polylogarithms as*

$$\mathrm{Le}_{(s_1,\ldots,s_m)}(z) = \sum_{\rho} \zeta(\rho; z),$$

where ρ runs through all sets of the form $(s_1 * \cdots * s_m)$, the symbol $*$ standing either for $+$ or the comma; the total number of such sets is 2^{m-1}. Moreover, ζ is defined under (2).

For further reference we specify this theorem with $m = 2$ and $m = 3$, with $z = 1$ (in which form these relations has appeared in [Hoffman 1992]):

$$\mathrm{Le}(s_1,s_2) = \zeta(s_1,s_2) + \zeta(s_1+s_2), \tag{6}$$

$$\mathrm{Le}(s_1,s_2,s_3) = \zeta(s_1,s_2,s_3) + \zeta(s_1+s_2+s_3) + \zeta(s_1,s_2+s_3) + \zeta(s_1+s_2,s_3). \tag{7}$$

Finally, we introduce the notations

$$H(a,b) = \sum_{n=1}^{\infty} \frac{H_{n,b}}{n^a},$$

$$H(a,b,c) = \sum_{n=1}^{\infty} \frac{H_{n,b}H_{n,c}}{n^a},$$

$$H(a,b,c,d) = \sum_{n=1}^{\infty} \frac{H_{n,b}H_{n,c}H_{n,d}}{n^a}.$$

We can call these sums ordinary (or first-order), quadratic (or second-order) and cubic Euler sums, respectively.

3.1. ***Quadratic sums: the homogeneous case.*** We now show how we can trace back $H(a,b,b)$ to nonstrict multiple zeta function — and so, by Theorem 2, to multiple zeta values — and $H(a,b)$. Since there are extensive tables for multiple zeta values and many results for $H(a,b)$, we can calculate sums like

$$H(a,b,b)=\sum_{n=1}^{\infty}\frac{H_{n,b}^2}{n^a}$$

with a relatively small effort.

Theorem 3. *For homogeneous quadratic Euler sums we have the reduction*

$$H(a,b,b)=2\,\mathrm{Le}(a,b,b)-H(a,2b),$$

or, if we write out the definitions,

$$\sum_{n=1}^{\infty}\frac{H_{n,b}^2}{n^a}=2\,\mathrm{Le}(a,b,b)-\sum_{n=1}^{\infty}\frac{H_{n,2b}}{n^a}.$$

Proof. Let us write out the sums:

$$H(a,b,b)=\sum_{n=1}^{\infty}\frac{H_{n,b}^2}{n^a}=\sum_{n=1}^{\infty}\frac{1}{n^a}\sum_{m=1}^{n}\frac{1}{m^b}\sum_{k=1}^{n}\frac{1}{k^b}.$$

On the other hand,

$$\mathrm{Le}(a,b,b)=\sum_{n=1}^{\infty}\frac{1}{n^a}\sum_{m=1}^{n}\frac{1}{m^b}\sum_{k=1}^{m}\frac{1}{k^b}.$$

Geometrically, the sum $H(a,b,b)$ runs through a two dimensional square with integer coordinates. On the other hand, $\mathrm{Le}(a,b,b)$ runs through the lower triangle of this square, including the diagonal. By symmetry of the terms of the sums, the lattice points of this square are equal if we mirror them with respect to the main diagonal of the square. Therefore the sum $H(a,b,b)$ equals twice $\mathrm{Le}(a,b,b)$ minus the diagonal, which is counted twice. At the diagonal the inner sums equal

$$\sum_{m=1}^{n}\frac{1}{m^{2b}}.$$

Summing on the index n, we have our relation. $\square$

Employing the above theorem, in the next subsection we provide a concrete example. This example is chosen to be very typical. It uses almost all the usual tricks which lead to the zeta expression of an Euler sum.

3.2. ***The homogeneous quadratic sums $H(2,2,2)$, $H(2,1,1)$ and $H(3,1,1)$.*** According to the theorem of Flajolet and Salvy (in this paper Theorem 1), the sum

$$\sum_{n=1}^{\infty} \frac{H_{n,2}^2}{n^2}$$

reduces to zeta values. Without using the evaluations of Flajolet and Salvy, we employ our above theorem. This implies the next representation.

Theorem 4.
$$\sum_{n=1}^{\infty} \frac{H_{n,2}^2}{n^2} = \frac{19}{24}\zeta(6) + \zeta^2(3).$$

Proof. The sum in the left-hand side equals

$$H(2,2,2) = 2\,\mathrm{Le}(2,2,2) - H(2,4).$$

Our goal is to reduce the expression on the right to Riemann zeta values. By (7),

$$\mathrm{Le}(2,2,2) = \zeta(6) + \zeta(4,2) + \zeta(2,4) + \zeta(2,2,2).$$

All the values $\zeta(4,2)$, $\zeta(2,4)$ and $\zeta(2,2,2)$ can be found in [Li 2011]:

$$\zeta(4,2) = \zeta^2(3) - \frac{4}{3}\zeta(6), \quad \zeta(2,4) = -\zeta^2(3) + \frac{25}{12}\zeta(6), \quad \zeta(2,2,2) = \frac{3}{16}\zeta(6).$$

Altogether,

$$\mathrm{Le}(2,2,2) = \frac{31}{16}\zeta(6) = \frac{31}{15120}\pi^6.$$

Now we deal with the sum $H(2,4)$. We could not find in the literature directly this sum, but in [Flajolet and Salvy 1998, p. 16, formula (b)] we can find that

$$H(4,2) = \zeta^2(3) - \frac{1}{3}\zeta(6).$$

We now apply the reflection formula [Boyadzhiev 2002; Flajolet and Salvy 1998]

$$H(a,b) + H(b,a) = \zeta(a)\zeta(b) + \zeta(a+b)$$

to obtain

$$H(2,4) = \zeta(2)\zeta(4) + \zeta(6) - H(b,a) = \frac{37}{12}\zeta(6) - \zeta^2(3).$$

(Here we used the fact that $\zeta(2)\zeta(4) = \frac{7}{4}\zeta(6)$.) Hence

$$H(2,2,2) = 2 \cdot \frac{31}{16}\zeta(6) - \Big(\frac{37}{12}\zeta(6) - \zeta^2(3)\Big) = \frac{19}{24}\zeta(6) + \zeta^2(3).$$

This is what we wanted to prove. $\square$

Connon [2007] gave an "elementary" evaluation for the sum $H(2, 1, 1)$, which was evaluated earlier by de Doelder [1991]. Their result is

$$H(2, 1, 1) = \sum_{n=1}^{\infty} \frac{H_n^2}{n^2} = \frac{17}{4}\zeta(4).$$

Once we have Theorem 3, the evaluation of this sum reduces to looking for the appropriate values of the multiple zeta. These values can be found in [Borwein and Girgensohn 1996], hence our method gives a third (and the easiest) proof.

In the same paper Connon notes that he could not evaluate the sum $H(3, 1, 1)$. Our method and the multiple zeta values from the paper [Borwein and Girgensohn 1996] give immediately that

$$H(3, 1, 1) = \sum_{n=1}^{\infty} \frac{H_n^2}{n^3} = \frac{7}{2}\zeta(5) - \zeta(2)\zeta(3).$$

However, Mathematica can evaluate this sum automatically.

Connon gave an integral representation for $H(q, 1, 1)$ for integer $q > 1$:

$$\sum_{n=1}^{\infty} \frac{H_n^2}{n^q} = \int_0^1 \int_0^1 \frac{\mathrm{Li}_{q-2}\big((1-t)(1-u)\big) \log t \log u}{(1-t)(1-u)} \, du \, dt.$$

By the results above for $q = 3$ and knowing that $\mathrm{Li}_1(x) = -\log(1-x)$, we get the closed form of the following integral:

$$\int_0^1 \int_0^1 \frac{\log\big(1-(1-t)(1-u)\big) \log t \log u}{(1-t)(1-u)} \, du \, dt = \zeta(2)\zeta(3) - \frac{7}{2}\zeta(5). \tag{8}$$

3.3. ***Quadratic sums: the inhomogeneous case.*** Another kind of approach helps us to evaluate inhomogeneous quadratic sums, i.e., sums of the form

$$\sum_{n=1}^{\infty} \frac{H_{n,b} H_{n,c}}{n^a}.$$

Namely, the next theorem is true.

Theorem 5. *Inhomogeneous quadratic Euler sums can be expressed by the nonstrict multiple zeta function as*

$$H(a, b, c) = \mathrm{Le}(a, b, c) + \mathrm{Le}(a, c, b) - \mathrm{Le}(a, b+c).$$

If $c = b$ this formula reduces to the formula presented in Theorem 3.

Proof. $$H(a,b,c)=\sum_{n=1}^{\infty}\frac{H_{n,b}H_{n,c}}{n^a}=\sum_{n=1}^{\infty}\frac{1}{n^a}\left(\sum_{m=1}^{n}\frac{1}{m^b}\right)\left(\sum_{k=1}^{n}\frac{1}{k^c}\right)$$
$$=\sum_{n=1}^{\infty}\frac{1}{n^a}\left(\sum_{m=1}^{n}\frac{1}{m^b}\right)\left(\sum_{k=1}^{m}\frac{1}{k^c}\right)+\sum_{n=1}^{\infty}\frac{1}{n^a}\left(\sum_{m=1}^{n}\frac{1}{m^b}\right)\left(\sum_{k=m+1}^{n}\frac{1}{k^c}\right).$$

The first sum is nothing but $\mathrm{Le}(a,b,c)$, while the second can be rearranged as

$$\sum_{n=1}^{\infty}\frac{1}{n^a}\left(\sum_{m=1}^{n}\frac{1}{m^b}\right)\left(\sum_{k=m+1}^{n}\frac{1}{k^c}\right)=\sum_{n=1}^{\infty}\frac{1}{n^a}\left(\sum_{k=2}^{n}\frac{1}{k^c}\right)\left(\sum_{m=1}^{k-1}\frac{1}{m^b}\right).$$

Since the sum over m is empty if $k=1$, we can start the sum over k from 1. Hence, at this point, we have that

$$H(a,b,c)=\mathrm{Le}(a,b,c)+\sum_{n=1}^{\infty}\frac{1}{n^a}\left(\sum_{k=1}^{n}\frac{1}{k^c}\right)\left(\sum_{m=1}^{k-1}\frac{1}{m^b}\right).$$

The latter sum almost equals $\mathrm{Le}(a,c,b)$, but here the sum on m runs up to $k-1$ instead of k. We can resolve this as follows:

$$\sum_{n=1}^{\infty}\frac{1}{n^a}\left(\sum_{k=1}^{n}\frac{1}{k^c}\right)\left(\sum_{m=1}^{k-1}\frac{1}{m^b}\right)=\sum_{n=1}^{\infty}\frac{1}{n^a}\left(\sum_{k=1}^{n}\frac{1}{k^c}\right)\left(\sum_{m=1}^{k}\frac{1}{m^b}\right)-\sum_{n=1}^{\infty}\frac{1}{n^a}\left(\sum_{k=1}^{n}\frac{1}{k^{c+b}}\right).$$

The right-hand side equals $\mathrm{Le}(a,c,b)-\mathrm{Le}(a,c+b)$. Substituting this into the ultimate expression of $H(a,b,c)$, we are done. □

3.4. *Two nonhomogeneous quadratic sums: $H(2,1,2)$ and $H(2,2,3)$.* By using the theorem of the last subsection, we evaluate the next inhomogeneous quadratic sums.

Theorem 6. *We have*

$$\sum_{n=1}^{\infty}\frac{H_nH_{n,2}}{n^2}=\zeta(2)\zeta(3)+\zeta(5),$$
$$\sum_{n=1}^{\infty}\frac{H_{n,2}H_{n,3}}{n^2}=\frac{131}{16}\zeta(7)-\frac{5}{2}\zeta(2)\zeta(5)-\frac{3}{2}\zeta(3)\zeta(4).$$

Note that the theorem of Flajolet and Salvy does not apply to these sums.

Proof. These proofs are again instructive. First,

$$\sum_{n=1}^{\infty}\frac{H_nH_{n,2}}{n^2}=H(2,1,2)=\mathrm{Le}(2,1,2)+\mathrm{Le}(2,2,1)-\mathrm{Le}(2,3). \tag{9}$$

Moreover, according to (7),

$$\operatorname{Le}(2,1,2)=\zeta(5)+\zeta(3,2)+\zeta(2,3)+\zeta(2,1,2).$$

In [Borwein and Girgensohn 1996] we find that

$$\zeta(2,1,2)=\frac{9}{2}\zeta(5)-2\zeta(2)\zeta(3),\quad \zeta(3,2)=-\frac{11}{2}\zeta(5)+3\zeta(2)\zeta(3).$$

We could not find the value of $\zeta(2,3)$ directly, but it can be easily deduced by a result of Boyadzhiev [2002]. Namely,

$$\zeta(2,3)=\sum_{n=1}^{\infty}\frac{H_{n-1,3}}{n^2}=\sum_{n=1}^{\infty}\frac{H_{n,3}}{n^2}-\zeta(5).$$

The zeta expression of the harmonic sum on the right — and even a more general form — is worked out by Boyadzhiev in the same paper:

$$\sum_{n=1}^{\infty}\frac{H_{n,3}}{n^2}=\frac{11}{2}\zeta(5)-2\zeta(2)\zeta(3).$$

(In general, he found that

$$\sum_{n=1}^{\infty}\frac{H_{n,3}}{n^p}=\zeta(3)\zeta(p)+\Big(1+\frac{p^3+5p}{12}\Big)\zeta(p+3)-\frac{1}{4}\sum_{k=1}^{p-1}k(k+1)\zeta(k+2)\zeta(p-k+1)$$
$$-\frac{1}{4}p(p+1)H(p+2,1)-\frac{1}{2}[H(p+1,2)+\zeta(p+1)\zeta(2)],$$

if p is even.)

So

$$\zeta(2,3)=\frac{9}{2}\zeta(5)-2\zeta(2)\zeta(3). \tag{10}$$

Altogether, we get that

$$\operatorname{Le}(2,1,2)=\frac{9}{2}\zeta(5)-\zeta(2)\zeta(3).$$

We also need the value of $\operatorname{Le}(2,2,1)$. Again, employing (7),

$$\operatorname{Le}(2,2,1)=\zeta(5)+\zeta(4,1)+\zeta(2,3)+\zeta(2,2,1).$$

Using the tables in [Borwein and Girgensohn 1996], we find that

$$\zeta(4,1)=2\zeta(5)-\zeta(2)\zeta(3),\quad \zeta(2,2,1)=-\frac{11}{2}\zeta(5)+3\zeta(2)\zeta(3).$$

Applying to (10) the value of $\zeta(2,3)$ calculated above, we get the simple zeta value of $\operatorname{Le}(2,2,1)$:

$$\operatorname{Le}(2,2,1)=2\zeta(5).$$

The last undetermined zeta value in (9) does not cause any problem:

$$\mathrm{Le}(2,3)=\zeta(5)+\zeta(2,3)=\zeta(5)+\frac{9}{2}\zeta(5)-2\zeta(2)\zeta(3)=\frac{11}{2}\zeta(5)-2\zeta(2)\zeta(3).$$

Collecting the nonstrict zeta values in (9) for $H(2,1,2)$, we find that they really sum to $\zeta(5)+\zeta(2)\zeta(3)$.

Now we turn to the second nonlinear Euler sum $H(2,2,3)$. It equals

$$\mathrm{Le}(2,2,3)+\mathrm{Le}(2,3,2)-\mathrm{Le}(2,5). \tag{11}$$

We follow the same pattern as above:

$$\mathrm{Le}(2,2,3)=\zeta(7)+\zeta(4,3)+\zeta(2,5)+\zeta(2,2,3).$$

The zeta expressions of $\zeta(4,3)$ and $\zeta(2,5)$ can be calculated from the results in [Bailey et al. 1994], under the notations $\sigma_h(3,4)$ and $\sigma_h(5,2)$, respectively. They equal

$$\zeta(4,3)=17\zeta(7)-10\zeta(2)\zeta(5), \tag{12}$$

$$\zeta(2,5)=10\zeta(7)-2\zeta(3)\zeta(4)-4\zeta(2)\zeta(5). \tag{13}$$

(In fact, the next expressions are deduced in [Bailey et al. 1994] and by the same authors in [Borwein et al. 1995]:

$$\begin{aligned}\zeta(m,n)=\frac{1}{2}\left(\binom{m+n}{m}-1\right)\zeta(m+n)+\zeta(m)\zeta(n)\\ -\sum_{j=1}^{m+n}\left(\binom{2j-2}{m-1}+\binom{2j-2}{n-1}\right)\zeta(2j-1)\zeta(m+n-2j+1)\end{aligned} \tag{14}$$

if m is odd and n is even, while

$$\begin{aligned}\zeta(m,n)=-\frac{1}{2}\left(\binom{m+n}{m}+1\right)\zeta(m+n)\\ +\sum_{j=1}^{m+n}\left(\binom{2j-2}{m-1}+\binom{2j-2}{n-1}\right)\zeta(2j-1)\zeta(m+n-2j+1)\end{aligned}$$

if m is even and n is odd.)

The value of $\zeta(2,2,3)$ is listed in [Borwein and Girgensohn 1996]:

$$\zeta(2,2,3)=-\frac{291}{16}\zeta(7)-\frac{3}{2}\zeta(3)\zeta(4)+12\zeta(2)\zeta(5).$$

Hence

$$\mathrm{Le}(2,2,3)=\frac{157}{16}\zeta(7)-2\zeta(2)\zeta(5)-\frac{7}{2}\zeta(3)\zeta(4). \tag{15}$$

The calculation of Le(2, 3, 2) needs a bit less work, because

$$\mathrm{Le}(2, 3, 2) = \zeta(7) + \zeta(5, 2) + \zeta(2, 5) + \zeta(2, 3, 2),$$

and we can apply the reflection formula [Wan 2012]

$$\zeta(a, b) + \zeta(b, a) = \zeta(a)\zeta(b) - \zeta(a + b).$$

We have that

$$\zeta(5, 2) + \zeta(2, 5) = \zeta(2)\zeta(5) - \zeta(7),$$

and so

$$\mathrm{Le}(2, 3, 2) = \zeta(2)\zeta(5) + \zeta(2, 3, 2).$$

The value

$$\zeta(2, 3, 2) = \frac{75}{8}\zeta(7) - \frac{11}{2}\zeta(2)\zeta(5)$$

can be found in [Borwein and Girgensohn 1996]. Hence

$$\mathrm{Le}(2, 3, 2) = \frac{75}{8}\zeta(7) - \frac{9}{2}\zeta(2)\zeta(5). \tag{16}$$

Only Le(2, 5) is missing in (11).

$$\mathrm{Le}(2, 5) = \zeta(7) + \zeta(2, 5) = 11\zeta(7) - 4\zeta(2)\zeta(5) - 2\zeta(4)\zeta(3), \tag{17}$$

as we can see from (13).

Substituting (15), (16), and (17) into (11), we have the second sum in Theorem 6. □

3.5. ***Homogeneous cubic sums.*** The geometric approach we applied in Section 3.1 to homogeneous quadratic sums can be generalized to homogeneous cubic sums as well.

Theorem 7. *The homogeneous cubic Euler sums can be reduced to multiple zeta values of depth* 3 *and* 4, *and to Euler sums of order one and two. Namely*,

$$H(a, b, b, b) = 6\zeta(a, b, b, b) + 6\zeta(a + b, b, b) + 3H(a, b, 2b) - 2H(a, 3b).$$

Proof. The sum

$$H(a, b, b, b) = \sum_{n=1}^{\infty} \frac{1}{n^a} \sum_{m=1}^{n} \frac{1}{m^b} \sum_{k=1}^{n} \frac{1}{k^b} \sum_{l=1}^{n} \frac{1}{l^b}$$

can be considered as a sum on the infinite cubic lattice with positive integer coordinates. We subtract from this the second-order sums on the principal planes $m = k$, $k = l$ and $m = l$. Since we have subtracted the main diagonal $m = k = l$ three times, we can add it two times. Then, by symmetry, we have six times the sum in the "lower" part of the cube, with (integer) coordinates $m = 1, 2, \ldots, n$,

$k = 1, \ldots, m-1$ and $l = 1, \ldots, k-1$. Again, by symmetry, the second-order sums on the principal planes $m = k$, $k = l$ and $m = l$ are identical and equal $H(a, b, 2b)$; and the main diagonal $m = k = l$ corresponds to the sum $H(a, 3b)$. Hence, we have the relation

$$\frac{H(a, b, b, b) - 3H(a, b, 2b) + 2H(a, 3b)}{6} = \sum_{n=1}^{\infty} \frac{1}{n^a} \sum_{m=1}^{n} \frac{1}{m^b} \sum_{k=1}^{m-1} \frac{1}{k^b} \sum_{l=1}^{k-1} \frac{1}{l^b}.$$

The sum on the right-hand side can be easily rewritten as a multiple zeta expression, if we separate the terms $m = 1, 2, \ldots, n-1$ and $m = n$:

$$\sum_{n=1}^{\infty} \frac{1}{n^a} \sum_{m=1}^{n} \frac{1}{m^b} \sum_{k=1}^{m-1} \frac{1}{k^b} \sum_{l=1}^{k-1} \frac{1}{l^b} = \zeta(a, b, b, b) + \zeta(a+b, b, b).$$

Substituting this into the above relation and rearranging we have our theorem. □

3.6. *The inhomogeneous quadratic sum $H(4, 1, 2)$.* We apply the theorem of the above section to prove the next identity.

Theorem 8. $$\sum_{n=1}^{\infty} \frac{H_n H_{n,2}}{n^4} = \frac{3}{4}\zeta(3)\zeta(4) + 2\zeta(2)\zeta(5) - \frac{51}{16}\zeta(7).$$

Note that here the reduction theorem of Flajolet and Salvy does not apply.

Proof. We specialize Theorem 7 to $a = 4$ and $b = 1$. Then

$$H(4, 1, 1, 1) = 6\zeta(4, 1, 1, 1) + 6\zeta(5, 1, 1) + 3H(4, 1, 2) - 2H(4, 3). \tag{18}$$

The sum on the left-hand side equals

$$H(4, 1, 1, 1) = \sum_{n=1}^{\infty} \frac{H_n^3}{n^4} = \frac{231}{16}\zeta(7) - \frac{51}{4}\zeta(3)\zeta(4) + 2\zeta(2)\zeta(5), \tag{19}$$

as one can find in [Flajolet and Salvy 1998, p. 16]. Moreover, an important simplification can be done on the right-hand side, since

$$\zeta(4, 1, 1, 1) = \zeta(5, 1, 1). \tag{20}$$

This is an observation of J. Borwein, D. Bradley and D. Broadhurst, see the paragraph after formula (30) in [1997]. The general version that they proved is the following:

$$\zeta(m+2, \{1\}_n) = \zeta(n+2, \{1\}_m),$$

where $\{1\}_n$ means that we repeat the argument n times. Identity (20) comes if we substitute $m = 2$ and $n = 3$. Other examples are

$$\zeta(2, \{1\}_n) = \zeta(n+2), \quad \zeta(3, \{1\}_n) = \zeta(n+2, 1),$$

and so on.

The value of $\zeta(5,1,1)$ can be found in [Borwein and Girgensohn 1996]:

$$\zeta(5,1,1) = -\frac{5}{4}\zeta(3)\zeta(4) + 5\zeta(7) - 2\zeta(5)\zeta(2).$$

Thus, with respect to (18) and (19) we have the temporary result

$$\begin{aligned}\frac{231}{16}\zeta(7) - \frac{51}{4}\zeta(3)\zeta(4) + 2\zeta(2)\zeta(5)&\\ = -15\zeta(3)\zeta(4) + 60\zeta(7) - 24\zeta(5)\zeta(2) + 3H(4,1,2) - 2H(4,3),&\end{aligned}$$

which can be rearranged:

$$3H(4,1,2) = -\frac{729}{16}\zeta(7) + \frac{9}{4}\zeta(3)\zeta(4) + 26\zeta(2)\zeta(5) + 2H(4,3). \tag{21}$$

The sum $H(4,3)$ can be deduced from the formula of Bailey, Borwein, and Girgensohn (14):

$$H(4,3) = \sum_{n=1}^{\infty}\frac{H_{n,3}}{n^4} = \sum_{n=0}^{\infty}\frac{H_{n+1,3}}{(n+1)^4} = \sum_{n=0}^{\infty}\frac{H_{n,3} + 1/(n+1)^3}{(n+1)^4} = \zeta(4,3) + \zeta(7).$$

By using (12),

$$H(4,3) = 18\zeta(7) - 10\zeta(2)\zeta(5).$$

Substituting this into (21), we are done. □

4. Generating functions of nonlinear Euler sums

Up to this point, we were interested in closed form expression for quadratic and cubic Euler sums. In several cases, using polylogarithms and several tricks, we can involve a free parameter z in these sums and express them with known special functions. To be more concrete, we can find the generating functions for H_n^2 and H_n^3 as well. We shall deduce the formulas in the next theorem.

Theorem 9. *For any $|z| < 1$ the ordinary generating functions of H_n^2 and H_n^3 are*

$$\sum_{n=1}^{\infty} H_n^2 z^n = \frac{1}{1-z}\big(\mathrm{Li}_2(z) + \log^2(1-z)\big), \tag{22}$$

$$\begin{aligned}\sum_{n=1}^{\infty} H_n^3 z^n = \frac{1}{1-z}\Big(-\frac{\pi^2}{2}\log(1-x) - \log^3(1-z) + \frac{3}{2}\log^2(1-z)\log z&\\ + 3\,\mathrm{Li}_3(1-z) + \mathrm{Li}_3(z) - 3\zeta(3)\Big).&\end{aligned} \tag{23}$$

The first relation is easy to prove and is not new; one can find it, for example, in [Mező 2013]. For the sake of completeness, we give its proof. To our knowledge, the second formula is new.

Proof. Note that $H_{n-1}^2 = \left(H_n - \frac{1}{n}\right)^2 = H_n^2 + \frac{1}{n^2} - 2\frac{H_n}{n}$, whence

$$\sum_{n=1}^{\infty} H_{n-1}^2 z^n = \sum_{n=1}^{\infty} H_n^2 z^n + \sum_{n=1}^{\infty} \frac{z^n}{n^2} - 2\sum_{n=1}^{\infty} \frac{H_n}{n} z^n. \tag{24}$$

The second sum is $\mathrm{Li}_2(z)$ (see definition (5)), while the last sum equals

$$\sum_{n=1}^{\infty} \frac{H_n}{n} z^n = \mathrm{Li}_2(z) + \tfrac{1}{2}\log^2(1-z), \tag{25}$$

by [Borwein and Borwein 1995]. If we temporarily introduce the function

$$f(z) = \sum_{n=1}^{\infty} H_n^2 z^n,$$

then (24) and (25) imply that

$$zf(z) = f(z) + \mathrm{Li}_2(z) - 2\big(\mathrm{Li}_2(z) + \tfrac{1}{2}\log^2(1-z)\big);$$

hence

$$f(z) = \sum_{n=1}^{\infty} H_n^2 z^n = \frac{\mathrm{Li}_2(z) + \log^2(1-z)}{1-z}.$$

Let us prove the second formula. Our initial point is almost the same as above:

$$H_n^3 = \left(H_{n-1} + \frac{1}{n}\right)^3 = H_{n-1}^3 + 3H_{n-1}^2\frac{1}{n} + 3H_{n-1}\frac{1}{n^2} + \frac{1}{n^3}.$$

This time we set

$$f(z) = \sum_{n=1}^{\infty} H_n^3 z^n.$$

Then

$$f(z) = zf(z) + 3\sum_{n=0}^{\infty} \frac{H_n^2}{n+1} z^{n+1} + 3\sum_{n=0}^{\infty} \frac{H_n}{(n+1)^2} z^{n+1} + \mathrm{Li}_3(z). \tag{26}$$

To calculate the first sum, we utilize the first formula of the theorem:

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{H_n^2}{n+1} z^{n+1} &= \int_0^z \left(\frac{\mathrm{Li}_2(x) + \log^2(1-x)}{1-x}\right) dx \\ &= -\frac{\pi^2}{3}\log(1-z) - \frac{1}{3}\log^3(1-z) + \log^2(1-z)\log z \\ &\qquad + \log(1-z)\,\mathrm{Li}_2(z) + 2\,\mathrm{Li}_3(1-z) - 2\zeta(3).\end{aligned} \tag{27}$$

This can be seen directly by differentiation. The integration constant $2\zeta(3)$ comes if we substitute $z = 0$.

Now we deal with the second sum on the right-hand side of (26).

$$
\begin{aligned}
(28)\quad \sum_{n=1}^{\infty} \frac{H_n}{(n+1)^2} z^{n+1}
&= \int_0^z \sum_{n=1}^{\infty} \frac{H_n}{n+1} x^n \, dx = \int_0^z \frac{\log^2(1-x)}{2x} \, dx \\
&= \frac{1}{2} \log^2(1-z) \log z + \log(1-z) \operatorname{Li}_2(1-z) - \operatorname{Li}_3(1-z) + \zeta(3).
\end{aligned}
$$

Taking the derivative of the right-hand side, this can be justified. The integration constant comes if we substitute $z = 0$.

Substituting (27) and (28) into (26) and utilizing [Lewin 1991, formula (1.5)]

$$\operatorname{Li}_2(z) + \operatorname{Li}_2(1-z) = \zeta(2) - \log z \log(1-z)$$

to simplify, we have proven Theorem 9. □

It is interesting that S. Ramanujan dealt with a similar function as in (28), but in the denominator there is $(n+1)^3$ in place of $(n+1)^2$ in his function:

$$h(z) = \sum_{n=1}^{\infty} \frac{H_n}{(n+1)^3} z^{n+1}.$$

He could not provide a closed form for this function but he showed that it can be analytically continued to the whole complex plane in z and proved some functional equations for h. Details can be found in [Berndt 1985, p. 253]. Such generating functions also appear in a beautiful paper of Guillera and Sondow [2008].

4.1. ***Some series as consequences of Theorem 9.*** We note an interesting alternating nonlinear sum as a corollary of formula (27) in the proof of Theorem 9:

$$\sum_{n=1}^{\infty} (-1)^{n+1} \frac{H_n^2}{n+1} = \frac{\pi^2}{12} \log 2 - \frac{1}{3} \log^3 2 - \frac{1}{4} \zeta(3).$$

The proof can be done by substituting $z = -1$ into (27) and handling the occurring imaginary values. One of them is $\log(-1)$, the other one is $\operatorname{Li}_3(2)$. By a formula of Lewin's book [1981, (6.7), p. 154],

$$\operatorname{Li}_3(2) = \operatorname{Li}_3\left(\tfrac{1}{2}\right) + \frac{\pi^2}{3} \log 2 - \frac{1}{6} \log^3 2 - \frac{1}{2} i\pi \log^2 2.$$

Since $\log(-1) = i\pi$ in the principal branch, the imaginary parts cancel — as they must — and then we can finish the proof using the special values

$$\operatorname{Li}_2(-1) = -\frac{\pi^2}{12}, \quad \operatorname{Li}_3\left(\tfrac{1}{2}\right) = \frac{1}{24}\left(4 \log^3(2) + 21\zeta(3) - 2\pi^2 \log 2\right).$$

Another consequence of the calculations in (28) is the classic result of Euler, which is nothing but (1):

$$\sum_{n=1}^{\infty} \frac{H_n}{(n+1)^2} = \zeta(3).$$

To prove this we let z tend to 1 from the left. (Taking the limit is not straightforward, we have to check the Taylor series around 1 to see that we have the right to do this. Finally we see that all the terms cancel, and just the constant term $\zeta(3)$ remains.)

Nice sums of infinite series involving the square and third power of the *digamma function* are consequences of Theorem 9. This function is the logarithmic derivative of the Euler Γ function and can be defined by the sum [Gradshteyn and Ryzhik 2007]

$$\psi(x) = -\gamma + \sum_{n=0}^{\infty}\Big(\frac{1}{n+1} - \frac{1}{n+x}\Big) \quad (x \in \mathbb{R} \setminus \{0, -1, -2, \ldots\}).$$

Here $\gamma = -\lim\limits_{n\to\infty}\Big(\log n - \sum\limits_{k=1}^{n} \frac{1}{k}\Big) \approx 0.577215664901533$ is the Euler–Mascheroni constant.

The derivatives of the digamma function ψ', ψ'', ... are called *trigamma, tetragamma functions*, etc. In general, these derivatives are called *polygamma functions* and denoted by ψ_n ($\psi_0 = \psi$, $\psi_1 = \psi'$, ...). Since

$$\psi_k(n) = (-1)^{k+1} k!\,(\zeta(k+1) - H_{n-1,k+1}), \tag{29}$$

it is straightforward to see that the polygamma functions have the generating functions

$$\sum_{n=1}^{\infty} \psi_k(n) z^n = \frac{z}{1-z}(-1)^k k!\,(\mathrm{Li}_{k+1}(z) - \zeta(k+1)) \quad (|z| < 1,\ k = 1, 2, \ldots).$$

If $k = 0$, we have that

$$\sum_{n=1}^{\infty} \psi(n) z^n = \frac{z}{z-1}(\gamma + \log(1-z)).$$

From the general representation (29) it follows that at a positive integer n the digamma function equals

$$\psi(n) = H_{n-1} - \gamma. \tag{30}$$

We have infinite series for the second and third power of the digamma function:

$$\sum_{n=1}^{\infty} \frac{\psi^2(n+1)}{2^n} = \gamma^2 - 4\gamma \log 2 + \log^2 2 + \zeta(2),$$

$$\sum_{n=1}^{\infty} \frac{\psi^3(n+1)}{2^n} = \frac{\pi^2}{3}\log 2 + \frac{1}{3}\log^3 2 + \zeta(3) - \frac{\pi^2}{2}\gamma - 3\gamma \log^2 2 + 6\gamma^2 \log 2 - \gamma^3.$$

It is interesting that the second formula includes all the most frequently appearing constants: π, γ, e, $\zeta(3)$, and $\log 2$.

We shall prove just the second identity, because the first one is similar but simpler. Using Theorem 9, we can see that

$$\sum_{n=1}^{\infty} \frac{H_n^3}{2^n} = \frac{\pi^2}{3}\log 2 + \frac{1}{3}\log^3 2 + \zeta(3)$$

and

$$\sum_{n=1}^{\infty} \frac{H_n^2}{2^n} = \zeta(2) + \log^2 2 = \frac{\pi^2}{6} + \log^2 2.$$

From the generating function

$$\sum_{n=1}^{\infty} H_n z^n = -\frac{\log(1-z)}{1-z},$$

it is obvious that

$$\sum_{n=1}^{\infty} \frac{H_n}{2^n} = \log 4.$$

Since

$$\psi^3(n+1) = H_n^3 - 3\gamma H_n^2 + 3\gamma^2 H_n - \gamma^3,$$

the result follows after dividing by 2^n and summing over n.

5. The Landen functional equations of the dilogarithm and trilogarithm functions

As an application of generating functions of the above nonlinear Euler sums we give proofs for the functional equations of the dilogarithm and trilogarithm functions. The proof relies on finite identities and on a result of Euler with respect to binomial transforms.

More concretely, we shall reprove the functional equation of the dilogarithm function:

$$\mathrm{Li}_2\Big(\frac{x}{1+x}\Big) = -\frac{1}{2}\log^2(1+x) - \mathrm{Li}_2(-x). \tag{31}$$

This is called *Landen's equation* [Lewin 1981, (1.12), p. 5].

We also show a new proof of the Landen functional equation for the trilogarithm:

$$\mathrm{Li}_3\Big(\frac{x}{1+x}\Big) = \zeta(3) + \zeta(2)\log(1+x) - \tfrac{1}{2}\log^2(1+x)\log(-x) + \tfrac{1}{6}\log^3(1+x) - \mathrm{Li}_3(-x) + \mathrm{Li}_3(1+x). \tag{32}$$

This is proved in [Lewin 1981, p. 155].

We remark that these equations are also presented in [Lewin 1991] on p. 2, but there is a typo there: in equation (1.13) in place of the coefficient $\frac{1}{6}$ there is $\frac{1}{2}$, which is incorrect.

The known proofs are analytic. We present proofs which are based on finite combinatorial identities. Moreover, we show a reason why there probably does not exist a functional equation of Landen type for higher-order polylogarithms.

Closed-form expressions for $\mathrm{Li}_2\big(\frac{1}{2}\big)$ and $\mathrm{Li}_3\big(\frac{1}{2}\big)$ are also known [Lewin 1991, pages 1 and 2]:

$$\mathrm{Li}_2\big(\tfrac{1}{2}\big) = \frac{\pi^2}{12} - \frac{1}{2}\log^2 2, \quad \mathrm{Li}_3\big(\tfrac{1}{2}\big) = \frac{7}{8}\zeta(3) - \frac{\pi^2}{12}\log 2 - \frac{1}{6}\log^3 2.$$

But there is no such formula for $\mathrm{Li}_4\big(\frac{1}{2}\big)$; see the remark after equation (7.92) in [Lewin 1981, p. 211].

We try to get closer to the constant $\mathrm{Li}_4\big(\frac{1}{2}\big)$ and we show that

$$\mathrm{Li}_4\big(\tfrac{1}{2}\big) = \frac{\pi^4}{180} + \frac{\pi^2}{48}\log^2 2 - \frac{1}{24}\log^4 2 - \frac{7}{16}\log(2)\zeta(3) + \frac{1}{2}\sum_{n=1}^{\infty}(-1)^{n+1}\frac{H_n H_{n,2}}{n}. \tag{33}$$

The last sum on the right does not seem to be reducible to known constants. When we tried to reduce it, we found that in its expression $\mathrm{Li}_4\big(\frac{1}{2}\big)$ appears, so we would get a $0 = 0$-type identity upon substituting this into (33).

The new proofs of the Landen identities are based on the representations of the generalized harmonic numbers:

$$H_{n,2} = \sum_{k=1}^{n}\binom{n}{k}(-1)^{k+1}\frac{H_k}{k}, \tag{34}$$

$$H_{n,3} = \frac{1}{2}\sum_{k=1}^{n}\binom{n}{k}\frac{(-1)^{k+1}}{k}(H_k^2 + H_{k,2}), \tag{35}$$

$$H_{n,4} = \frac{1}{6}\sum_{k=1}^{n}\binom{n}{k}\frac{(-1)^{k+1}}{k}(H_k^3 + 3H_n H_{n,2} + 2H_{n,3}), \tag{36}$$

for all $n \geq 1$. (It is interesting that in the last sum, the term $H_k^3 + 3H_n H_{n,2} + 2H_{n,3}$ appears in [Adamchik 1997; Connon 2008a]. To see how to derive identities like this, we refer to [Connon 2008c].)

To prove (31) and (32) we need an identity, due to Euler, giving the generating function of a sequence's binomial transform. Recall that for an arbitrary real sequence a_n, the *binomial transform* of a_n is the sequence b_n defined by

$$b_n = \sum_{k=0}^{n} \binom{n}{k} a_k, \quad \text{or, equivalently,} \quad a_n = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} b_k.$$

If a_n has the generating function $a(x)$, that is, $\sum_{n=0}^{\infty} a_n x^n = a(x)$, then b_n has the generating function

$$\sum_{n=0}^{\infty} b_n x^n = \frac{1}{1-x}\, a\Big(\frac{x}{1-x}\Big).$$

For more information on binomial transforms and Euler's result, see [Dumont 1981; Mező and Dil 2009; Seidel 1877].

5.1. ***The Landen equation for the dilogarithm.*** It is straightforward to see that

$$\sum_{n=1}^{\infty} H_{n,k} x^n = \frac{\mathrm{Li}_k(x)}{1-x},$$

and from (34) we also know that $H_{n,2}$ is the inverse binomial transform of $-H_n/n$. Hence

$$\frac{\mathrm{Li}_2(x)}{1-x} = \frac{1}{1-x}\, a\Big(\frac{x}{1-x}\Big),$$

where $a(x)$ is the generating function of H_n/n. The denominator $1+x$ cancels, and we apply the substitution $x \to x/(1+x)$ to get

$$\mathrm{Li}_2\Big(\frac{x}{1+x}\Big) = a(x).$$

Finally, to prove (31) we realize that

$$\begin{aligned} a(x) = -\sum_{n=1}^{\infty} \frac{(-1)^n}{n} H_n x^n &= -\sum_{n=1}^{\infty} \frac{(-1)^n}{n} \Big(H_{n-1} + \frac{1}{n}\Big) x^n \\ &= -\sum_{n=1}^{\infty} \frac{(-1)^n}{n} H_{n-1} x^n - \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} x^n. \end{aligned}$$

The last two sums equal respectively $\frac{1}{2}\log^2(1+x)$ and $\mathrm{Li}_2(-x)$ (in the latter case by definition). These prove (31).

5.2. ***The Landen equation for the trilogarithm.*** Identity (35) shows that

$$\mathrm{Li}_3\Big(\frac{x}{1+x}\Big) = -\frac{1}{2} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} H_n^2 x^n - \frac{1}{2} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} H_{n,2} x^n. \tag{37}$$

Let us deal with the first sum. We prove that

$$\sum_{n=1}^{\infty}\frac{H_n^2}{n}x^n=\mathrm{Li}_3(x)+2\mathrm{Li}_2(1-x)\log(1-x)+\mathrm{Li}_2(x)\log(1-x)\\ -\tfrac{1}{3}\log^3(1-x)+2\log x\log^2(1-x)-\tfrac{1}{3}\pi^2\log(1-x)\tag{38}$$

for all $|x|<1$.

Applying (22), we have that

$$\sum_{n=1}^{\infty}\frac{H_n^2}{n}x^n=\int_0^x\frac{\mathrm{Li}_2(y)}{y(1-y)}\,dy+\int_0^x\frac{\log^2(y)}{y(1-y)}\,dy.$$

These integrands have primitive functions:

$$\int_0^x\frac{\mathrm{Li}_2(y)}{y(1-y)}\,dy=2\mathrm{Li}_3(1-x)+\mathrm{Li}_3(x)+\mathrm{Li}_2(x)\log(1-x)\\ +\log x\log^2(1-x)-\tfrac{1}{3}\pi^2\log(1-x)-2\zeta(3),\tag{39}$$

and

$$\int_0^x\frac{\log^2(y)}{y(1-y)}\,dy=-2\mathrm{Li}_3(1-x)+2\mathrm{Li}_2(1-x)\log(1-x)\\ -\tfrac{1}{3}\log^3(1-x)+\log x\log^2(1-x)+2\zeta(3),\tag{40}$$

as can be seen by differentiation. (The integration constants come if we substitute $x=0$.) These two integrals together give (38).

Similarly,

$$\sum_{n=1}^{\infty}\frac{H_{n,2}}{n}x^n=\int_0^x\frac{\mathrm{Li}_2(y)}{y(1-y)}\,dy.\tag{41}$$

This integral is the same as (39).

Collecting the results under (38) and (41) (considering (39)) and putting them into (37), we get the Landen equation for the trilogarithm, after a simplification.

5.3. ***The Landen equation for the tetralogarithm and higher-order polylogs.*** Let us go to the tetralogarithm $\mathrm{Li}_4(x)$. Identity (36) immediately gives

$$\mathrm{Li}_4\Big(\frac{x}{1+x}\Big)=\frac{1}{6}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\big(H_n^3+3H_nH_{n,2}+2H_{n,3}\big)x^n.$$

This shows why finding a functional equation of Landen type for $\mathrm{Li}_4(x)$ is not hopeful: the product $H_nH_{n,2}$ does not seem to have a generating function expressible by standard functions for all $|x|<1$. This is probably true for higher-order polylogarithms as well, because those harmonic number expressions probably contain H_n^4 and other powers and products of generalized harmonic numbers.

We note that Theorem 1 of [Ulanskii 2003] does provide a general Landen functional equation for these polylogarithms. However, that equation uses multiple zeta functions, which probably cannot be reduced to polylogarithms and ordinary logarithms.

6. Collected sums

We close the paper collecting the calculated sums.

$$\sum_{n=1}^{\infty}\frac{H_n^2}{n^2}=\frac{17}{4}\zeta(4),\qquad \sum_{n=1}^{\infty}\frac{H_n^2}{n^3}=\frac{7}{2}\zeta(5)-\zeta(2)\zeta(3),$$
$$\sum_{n=1}^{\infty}\frac{H_{n,2}^2}{n^2}=\frac{19}{24}\zeta(6)+\zeta^2(3),\qquad \sum_{n=1}^{\infty}\frac{H_nH_{n,2}}{n^2}=\zeta(2)\zeta(3)+\zeta(5),$$
$$\sum_{n=1}^{\infty}\frac{H_{n,2}H_{n,3}}{n^2}=\frac{131}{16}\zeta(7)-\frac{5}{2}\zeta(2)\zeta(5)-\frac{3}{2}\zeta(3)\zeta(4),$$
$$\sum_{n=1}^{\infty}\frac{H_nH_{n,2}}{n^4}=\frac{3}{4}\zeta(3)\zeta(4)+2\zeta(2)\zeta(5)-\frac{51}{16}\zeta(7),$$
$$\sum_{n=1}^{\infty}(-1)^{n+1}\frac{H_n^2}{n+1}=\frac{\pi^2}{12}\log 2-\frac{1}{3}\log^3 2-\frac{1}{4}\zeta(3),$$
$$\sum_{n=1}^{\infty}\frac{H_n^3}{2^n}=\frac{\pi^2}{3}\log 2+\frac{1}{3}\log^3(2)+\zeta(3),\qquad \sum_{n=1}^{\infty}\frac{H_n^2}{2^n}=\zeta(2)+\log^2 2.$$

We also present some other sums without proof. The methods of Sections 4 and 5 can help get these as well.

$$\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}H_n^3=\frac{1}{144}\big(\pi^4+18\pi^2\log^2 2-36\log^4 2+162\log(2)\zeta(3)\big),$$
$$\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}H_{n,3}=\frac{19}{1440}\pi^4-\frac{3}{4}\log(2)\zeta(3),$$
$$\sum_{n=1}^{\infty}\frac{(-1)^{n}}{n}H_{n-1}H_{n-1,2}=\frac{7}{8}\log(2)\zeta(3)-\frac{1}{4}\log^2(2)\zeta(2)-\frac{1}{8}\zeta^2(2),$$
$$\sum_{n=1}^{\infty}\frac{(-1)^{n}}{n}H_nH_{n,2}=\zeta(4)-\frac{1}{12}\log^4 2-2\,\mathrm{Li}_4\big(\tfrac{1}{2}\big)+\frac{\pi^2}{24}\log^2 2-\frac{7}{8}\log(2)\zeta(3).$$

To conclude, we record the amusing identity of sums

$$\sum_{n=1}^{\infty} H_{n,2}\frac{(-1)^n}{n!} = \frac{\pi^2}{6e} - \sum_{n=0}^{\infty}\frac{1}{(n+1)^2}\frac{!n}{n!}.$$

Here $!n$ is the subfactorial of n (the number of permutations on n elements that don't fix any of them) and $e = \exp(1)$. The reader can look for a proof as a challenge.

Acknowledgement

The author is grateful to the referee, who recommended some additional literature. Thanks to this, the integral (8) now has a closed form evaluation.

References

[Adamchik 1997] V. Adamchik, "On Stirling numbers and Euler sums", *J. Comput. Appl. Math.* **79**:1 (1997), 119–130. MR 97m:11025 Zbl 0877.39001

[Andrews et al. 1999] G. E. Andrews, R. Askey, and R. Roy, *Special functions*, Encyclopedia of Mathematics and its Applications **71**, Cambridge University Press, 1999. MR 2000g:33001 Zbl 0920.33001

[Apostol and Vu 1984] T. M. Apostol and T. H. Vu, "Dirichlet series related to the Riemann zeta function", *J. Number Theory* **19**:1 (1984), 85–102. MR 85j:11106 Zbl 0539.10032

[Bailey et al. 1994] D. H. Bailey, J. M. Borwein, and R. Girgensohn, "Experimental evaluation of Euler sums", *Experiment. Math.* **3**:1 (1994), 17–30. MR 96e:11168 Zbl 0810.11076

[Berndt 1985] B. C. Berndt, *Ramanujan's notebooks*, vol.1, Springer, New York, 1985. MR 86c:01062 Zbl 0555.10001

[Bigotte et al. 2002] M. Bigotte, G. Jacob, N. E. Oussous, and M. Petitot, "Lyndon words and shuffle algebras for generating the coloured multiple zeta values relations tables", *Theoret. Comput. Sci.* **273**:1-2 (2002), 271–282. MR 2003j:11097 Zbl 1014.68126

[Borwein and Borwein 1995] D. Borwein and J. M. Borwein, "On an intriguing integral and some series related to $\zeta(4)$", *Proc. Amer. Math. Soc.* **123**:4 (1995), 1191–1198. MR 95e:11137 Zbl 0840.11036

[Borwein and Bradley 2006] J. M. Borwein and D. M. Bradley, "Thirty-two Goldbach variations", *Int. J. Number Theory* **2**:1 (2006), 65–103. MR 2007e:11109 Zbl 1094.11031

[Borwein and Girgensohn 1996] J. M. Borwein and R. Girgensohn, "Evaluation of triple Euler sums", *Electron. J. Combin.* **3**:1 (1996), R23, 1–27. MR 97d:11137 Zbl 0884.40005

[Borwein et al. 1995] D. Borwein, J. M. Borwein, and R. Girgensohn, "Explicit evaluation of Euler sums", *Proc. Edinburgh Math. Soc.* (2) **38**:2 (1995), 277–294. MR 96f:11106 Zbl 0819.40003

[Borwein et al. 1997] J. M. Borwein, D. M. Bradley, and D. J. Broadhurst, "Evaluations of k-fold Euler/Zagier sums: A compendium of results for arbitrary k", *Electron. J. Combin.* **4**:2 (1997), R5, 1–21. MR 98b:11091 Zbl 0884.40004

[Bowman and Bradley 2001] D. Bowman and D. M. Bradley, "Multiple polylogarithms: A brief survey", pp. 71–92 in *q-series with applications to combinatorics, number theory, and physics* (Urbana, IL, 2000), edited by B. C. Berndt and K. Ono, Contemp. Math. **291**, Amer. Math. Soc., Providence, RI, 2001. Also available on arXiv: math/0310062. MR 2003c:33021 Zbl 0998.33013

[Boyadzhiev 2002] K. N. Boyadzhiev, "Consecutive evaluation of Euler sums", *Int. J. Math. Math. Sci.* **29**:9 (2002), 555–561. MR 2003d:11124 Zbl 0994.40003

[Boyadzhiev et al. 2008] K. N. Boyadzhiev, H. G. Gadiyar, and R. Padma, "The values of an Euler sum at the negative integers and a relation to a certain convolution of Bernoulli numbers", *Bull. Korean Math. Soc.* **45**:2 (2008), 277–283. MR 2009e:11174 Zbl 1148.11046

[Boyadzhiev et al. 2009] K. N. Boyadzhiev, H. Gopalkrishna Gadiyar, and R. Padma, "Alternating Euler sums at the negative integers", *Hardy-Ramanujan J.* **32** (2009), 24–37. MR 2010f:11144 Zbl 1194.11090

[Briggs et al. 1955] W. E. Briggs, S. Chowla, A. J. Kempner, and W. E. Mientka, "On some infinite series", *Scripta Math.* **21** (1955), 28–30. MR 16,1014d Zbl 0065.30401

[Bruckman 1982] P. S. Bruckman, "Problem H-320", *Fibonacci Quart.* **20**:2 (1982), 186–187.

[Chu 1997] W. Chu, "Hypergeometric series and the Riemann zeta function", *Acta Arith.* **82**:2 (1997), 103–118. MR 98m:11089 Zbl 0881.11066

[Connon 2007] D. F. Connon, "Elementary evaluations of some Euler sums", preprint, 2007. arXiv 0710.5360

[Connon 2008a] D. F. Connon, "Euler–Hurwitz series and non-linear Euler sums", preprint, 2008. arXiv 0803.1304

[Connon 2008b] D. F. Connon, "Some series and integrals involving the Riemann zeta function, binomial coefficients and the harmonic numbers, I", preprint, 2008. arXiv 0710.4022

[Connon 2008c] D. F. Connon, "Some series and integrals involving the Riemann zeta function, binomial coefficients and the harmonic numbers, IIa", preprint, 2008. arXiv 0710.4023

[Connon 2008d] D. F. Connon, "Some series and integrals involving the Riemann zeta function, binomial coefficients and the harmonic numbers, IIb", preprint, 2008. arXiv 0710.4024

[Connon 2008e] D. F. Connon, "Some series and integrals involving the Riemann zeta function, binomial coefficients and the harmonic numbers, III", preprint, 2008. arXiv 0710.4025

[Connon 2008f] D. F. Connon, "Some series and integrals involving the Riemann zeta function, binomial coefficients and the harmonic numbers, IV", preprint, 2008. arXiv 0710.4028

[Connon 2008g] D. F. Connon, "Some series and integrals involving the Riemann zeta function, binomial coefficients and the harmonic numbers, V", preprint, 2008. arXiv 0710.4047

[Connon 2008h] D. F. Connon, "Some series and integrals involving the Riemann zeta function, binomial coefficients and the harmonic numbers, VI", preprint, 2008. arXiv 0710.4032

[de Doelder 1991] P. J. de Doelder, "On some series containing $\psi(x) - \psi(y)$ and $(\psi(x) - \psi(y))^2$ for certain values of x and y", *J. Comput. Appl. Math.* **37**:1-3 (1991), 125–141. MR 92m:40002 Zbl 0782.33001

[Dumont 1981] D. Dumont, "Matrices d'Euler–Seidel", *Sémin. Lothar. Comb.* **5**:B05c (1981), 1–25. Zbl 0925.05025

[Euler 1776] L. Eulerus, "Meditationes circa singulare serierum genus", *Novi Comm. Acad. Sci. Petropolitanae* **20** (1776), 140–186. Reprinted on pp. 217–265 in his *Opera Omnia*, series prima, vol. 15 (*Commentationes analyticae* 2), Teubner, Berlin, 1927; the relevant passage starts on p. 228.

[Farnum and Tissier 1999] N. R. Farnum, "Problems and solutions", solution by A. Tissier of problem 10635, "Apery's constant", *Amer. Math. Monthly* **106**:10 (1999), 965–966. MR 1543578

[Flajolet and Salvy 1998] P. Flajolet and B. Salvy, "Euler sums and contour integral representations", *Experiment. Math.* **7**:1 (1998), 15–35. MR 99c:11110 Zbl 0920.11061

[Gradshteyn and Ryzhik 2007] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed., Academic Press, Amsterdam, 2007. MR 2008g:00005 Zbl 1208.65001

[Guillera and Sondow 2008] J. Guillera and J. Sondow, "Double integrals and infinite products for some classical constants via analytic continuations of Lerch's transcendent", *Ramanujan J.* **16**:3 (2008), 247–270. MR 2009e:11239 Zbl 1216.11075

[Hoffman 1992] M. E. Hoffman, "Multiple harmonic series", *Pacific J. Math.* **152**:2 (1992), 275–290. MR 92i:11089 Zbl 0763.11037

[Klamkin and Steinberg 1952] M. S. Klamkin and R. Steinberg, "Advanced Problems and Solutions: Solutions: 4431", *Amer. Math. Monthly* **59**:7 (1952), 471–472. MR 1528216

[Lewin 1981] L. Lewin, *Polylogarithms and associated functions*, North-Holland, New York, 1981. MR 83b:33019 Zbl 0465.33001

[Lewin 1991] L. Lewin (editor), *Structural properties of polylogarithms*, Mathematical Surveys and Monographs **37**, Amer. Math. Soc., Providence, RI, 1991. MR 93b:11158 Zbl 0745.33009

[Li 2011] Z. Li, "On harmonic sums and alternating Euler sums", preprint, 2011. arXiv 1012.5192v3

[Mező 2013] I. Mező, "A q-Raabe formula and an integral of the fourth Jacobi theta function", *J. Number Theory* **133**:2 (2013), 692–704. MR 2994381 Zbl 06124189

[Mező and Dil 2009] I. Mező and A. Dil, "Euler–Seidel method for certain combinatorial numbers and a new characterization of Fibonacci sequence", *Cent. Eur. J. Math.* **7**:2 (2009), 310–321. MR 2010c:11036 Zbl 1229.11043

[Seidel 1877] L. Seidel, "Über eine einfache Enstehung weise der Bernoullischen Zahlen und einiger verwandten Reihen", *Sitzungsberichte der Münch. Akad. Math. Phys. Classe* **4** (1877), 157–187.

[Shen 1995] L.-C. Shen, "Remarks on some integrals and series involving the Stirling numbers and $\zeta(n)$", *Trans. Amer. Math. Soc.* **347**:4 (1995), 1391–1399. MR 95g:11014 Zbl 0828.11044

[Sitaramachandra Rao 1987] R. Sitaramachandra Rao, "A formula of S. Ramanujan", *J. Number Theory* **25**:1 (1987), 1–19. MR 88c:11048 Zbl 0606.10032

[Sofo and Hassani 2012] A. Sofo and M. Hassani, "Quadratic harmonic number sums", *Appl. Math. E-Notes* **12** (2012), 110–117. MR 2992944

[Ulanskii 2003] E. A. Ulanskii, "Identities for generalized polylogarithms", *Mat. Zametki* **73**:4 (2003), 613–624. In Russian; translated in *Math. Notes* **73**:4 (2003), 571–581. MR 2004f:11065 Zbl 1093.11049

[Wan 2012] J. Wan, "Some notes on weighted sum formulae for doubling zeta values", preprint, 2012. arXiv 1206.2424v1

Received August 23, 2013. Revised October 4, 2013.

ISTVÁN MEZŐ
DEPARTMENT OF MATHEMATICS
NANJING UNIVERSITY OF INFORMATION SCIENCE AND TECHNOLOGY
NANJING, 210044
CHINA

istvanmezo81@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.272.227

BOUNDARY LIMITS FOR FRACTIONAL POISSON a-EXTENSIONS OF L^p BOUNDARY FUNCTIONS IN A CONE

LEI QIAO AND TAO ZHAO

If one replaces the Poisson kernel of a cone by the Poisson a-kernel, then normalized Poisson integrals with respect to the stationary Schrödinger operator converge along approach regions wider than the ordinary nontangential cones. In this paper we present new and simplified proofs of these results. We also generalize the result by Mizuta and Shimomura to the smooth cones.

1. Introduction and main results

Let $\mathbb{R}$ and $\mathbb{R}_+$ be the set of all real numbers and the set of all positive real numbers, respectively. We denote by $\mathbb{R}^n$ $(n \geq 2)$ the n-dimensional Euclidean space. A point in $\mathbb{R}^n$ is denoted by $P = (X, x_n)$, $X = (x_1, x_2, \ldots, x_{n-1})$. The Euclidean distance of two points P and Q in $\mathbb{R}^n$ is denoted by $|P - Q|$. Also $|P - O|$, with O the origin of $\mathbb{R}^n$, is simply denoted by $|P|$. The boundary, the closure and the complement of a set S in $\mathbb{R}^n$ are denoted by ∂S, $\overline{S}$ and S^c, respectively.

We introduce a system of spherical coordinates (r, Θ), $\Theta = (\theta_1, \theta_2, \ldots, \theta_{n-1})$, in $\mathbb{R}^n$ which are related to cartesian coordinates $(x_1, x_2, \ldots, x_{n-1}, x_n)$ by

$$x_1 = r\left(\prod_{j=1}^{n-1} \sin\theta_j\right), \qquad x_n = r\cos\theta_1,$$

for $n \geq 2$, and for $n \geq 3$,

$$x_{n-m+1} = r\left(\prod_{j=1}^{m-1} \sin\theta_j\right)\cos\theta_m \quad (2 \leq m \leq n-1),$$

where $0 \leq r < +\infty$, $-\pi/2 \leq \theta_{n-1} < 3\pi/2$, and if $n \geq 3$, then $0 \leq \theta_j \leq \pi$ $(1 \leq j \leq n-2)$.

This work was supported by the National Natural Science Foundation of China (grants U1304102 and 11301140.)
MSC2010: 31B05, 31B10.
Keywords: boundary limit, Poisson a-integral, stationary Schrödinger operator, cone.

The unit sphere and the upper unit half-sphere are denoted by S^{n-1} and S_+^{n-1}, respectively. For simplicity, a point $(1, \Theta)$ on S^{n-1} and the set $\{\Theta; (1, \Theta) \in \Omega\}$, for a set $\Omega \subset S^{n-1}$, are often identified with Θ and Ω, respectively. For two sets $\Xi \subset \mathbb{R}_+$ and $\Omega \subset S^{n-1}$, the set $\{(r, \Theta) \in \mathbb{R}^n; r \in \Xi, (1, \Theta) \in \Omega\}$ in $\mathbb{R}^n$ is simply denoted by $\Xi \times \Omega$. In particular, the half-space $\mathbb{R}_+ \times S_+^{n-1} = \{(X, x_n) \in \mathbb{R}^n; x_n > 0\}$ will be denoted by T_n.

By $C_n(\Omega)$, we denote the set $\mathbb{R}_+ \times \Omega$ in $\mathbb{R}^n$ with the domain Ω on S^{n-1}. We call it a cone. Then T_n is a special cone obtained by putting $\Omega = S_+^{n-1}$. We denote the sets $I \times \Omega$ and $I \times \partial\Omega$, with I an interval on $\mathbb{R}$, by $C_n(\Omega; I)$ and $S_n(\Omega; I)$. By $S_n(\Omega)$, we denote $S_n(\Omega; (0, +\infty))$, which is $\partial C_n(\Omega) - \{O\}$.

For positive functions h_1 and h_2, we say that $h_1 \lesssim h_2$ if $h_1 \le M h_2$ for some constant $M > 0$. If $h_1 \lesssim h_2$ and $h_2 \lesssim h_1$, we say that $h_1 \approx h_2$.

This article is devoted to the stationary Schrödinger operator

$$\mathrm{SSE}_a = -\Delta + a(P)I,$$

where Δ is the Laplace operator and I is the identity operator. We assume hereafter that the potential $a(P)$ is a nonnegative, locally integrable function in $C_n(\Omega)$, namely, $0 \le a \in L^b_{\mathrm{loc}}(C_n(\Omega))$, with $b > n/2$ if $n \ge 4$, and with $b = 2$ if $n = 2$ or 3. We denote this class of potentials by $\mathscr{A}$.

If $a \in \mathscr{A}$, then the operator SSE_a can be extended in the usual way from the space $C_0^\infty(C_n(\Omega))$ to an essentially self-adjoint operator on $L^2(C_n(\Omega))$ (see [Reed and Simon 1979, Chapter 13]). We shall denote the extended operator by SSE_a as well. The latter has Green function $G^a_\Omega(P, Q)$ vanishing almost everywhere at the boundary and possessing all the analytic properties. For $|P - Q| \to 0$, we normalize it such that $c_n G^a_\Omega(P, Q) \approx -\log|P - Q|$ when $n = 2$, or $c_n G^a_\Omega(P, Q) \approx |P - Q|^{2-n}$ when $n \ge 3$. Here $c_2 = 2\pi$, $c_n = (n-2)s_n$ when $n \ge 3$, and s_n is the surface area $2\pi^{n/2}(\Gamma(n/2))^{-1}$ of S^{n-1}. The Green function $G^a_\Omega(P, Q)$ is positive on $C_n(\Omega)$ and its inner normal derivative $\partial G^a_\Omega(P, Q)/\partial n_Q \ge 0$. We denote this derivative by $\mathrm{PI}^a_\Omega(P, Q)$, which is called the Poisson a-kernel with respect to $C_n(\Omega)$. Then the Poisson a-integral $\mathrm{PI}^a_\Omega f(P)$ $(P \in C_n(\Omega))$ is defined by

$$\mathrm{PI}^a_\Omega f(P) = \int_{S_n(\Omega)} \mathrm{PI}^a_\Omega(P, Q) f(Q)\, d\sigma_Q,$$

where

$$\mathrm{PI}^a_\Omega(P, Q) = \frac{\partial}{\partial n_Q} G^a_\Omega(P, Q),$$

$f \in L^p(\partial C_n(\Omega))$ $(1 \le p < \infty)$ and $d\sigma_Q$ is the surface area element on $S_n(\Omega)$.

Remark 1 [Yoshida 1991]. Let $\Omega = S_+^{n-1}$ and $a = 0$. Then

$$G^0_{S_+^{n-1}}(P, Q) = \begin{cases} \log|P - Q^*| - \log|P - Q| & n = 2, \\ |P - Q|^{2-n} - |P - Q^*|^{2-n} & n \ge 3, \end{cases}$$

where $Q^* = (Y, -y_n)$; that is, Q^* is the mirror image of $Q = (Y, y_n)$ with respect to ∂T_n. Hence, for the two points $P = (X, x_n) \in T_n$ and $Q = (Y, y_n) \in \partial T_n$, we have

$$\mathrm{PI}^0_{S^{n-1}_+}(P, Q) = \begin{cases} 2|P - Q|^{-2} x_n & n = 2, \\ 2(n-2)|P - Q|^{-n} x_n & n \geq 3. \end{cases}$$

Let Ω be a domain on S^{n-1} with smooth boundary. Consider the Dirichlet problem

$$\begin{aligned} (\Lambda_n + \lambda)\varphi = 0 \quad & \text{on } \Omega, \\ \varphi = 0 \quad & \text{on } \partial\Omega, \end{aligned}$$

where Λ_n is the spherical part of the Laplace operator

$$\Delta_n = \frac{n-1}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial r^2} + \frac{\Lambda_n}{r^2}.$$

We denote the least positive eigenvalue of this boundary value problem by λ and the normalized positive eigenfunction corresponding to λ by $\varphi(\Theta)$; $\int_\Omega \varphi^2(\Theta)\, d\sigma_\Theta = 1$, where $d\sigma_\Theta$ is the surface area on S^{n-1}.

To simplify our consideration in the following, we shall assume that if $n \geq 3$, then Ω is a $C^{2,\alpha}$-domain ($0 < \alpha < 1$) on S^{n-1} surrounded by a finite number of mutually disjoint closed hypersurfaces (e.g., see [Gilbarg and Trudinger 1977, pp. 88–89] for the definition of $C^{2,\alpha}$-domain). Then by modifying Miranda's method [1970, pp. 7–8], we can prove the inequality (see [Yoshida 1991, p. 373])

$$\varphi(\Theta) \approx \operatorname{dist}(\Theta, \partial\Omega) \quad (\Theta \in \Omega). \tag{1-1}$$

For any $(1, \Theta) \in \Omega$, we have (see [Courant and Hilbert 1953])

$$\varphi(\Theta) \approx \operatorname{dist}\big((1, \Theta), \partial C_n(\Omega)\big),$$

which yields that

$$\delta(P) \approx r\varphi(\Theta), \tag{1-2}$$

where $\delta(P) = \operatorname{dist}(P, \partial C_n(\Omega))$ and $P = (r, \Theta) \in C_n(\Omega)$.

Solutions of the ordinary differential equation

$$-Q''(r) - \frac{n-1}{r}Q'(r) + \left(\frac{\lambda}{r^2} + a(r)\right)Q(r) = 0, \quad 0 < r < \infty, \tag{1-3}$$

with a parameter λ play an essential role in these questions. It is known (see, for example, [Verzhbinskii and Maz'ya 1971]) that if the potential a belongs to $\mathscr{A}$, then (1-3) has a fundamental system of positive solutions $\{V, W\}$ such that V is nondecreasing with

$$0 \leq V(0+) \leq V(r) \nearrow \infty \quad \text{as } r \to +\infty,$$

and W is monotonically decreasing with

$$+\infty = W(0+) > W(r) \searrow 0 \quad \text{as } r \to +\infty.$$

We will also consider the class $\mathcal{B}$, consisting of the potentials $a \in \mathcal{A}$ such that there exists a finite limit $\lim_{r\to\infty} r^2 a(r) = k \in [0, \infty)$ and moreover $r^{-1}|r^2 a(r) - k| \in L(1, \infty)$. If $a \in \mathcal{B}$, then the (sub)superfunctions are continuous (see [Simon 1982]).

In the rest of paper, we assume that $a \in \mathcal{B}$ and we shall suppress this assumption for simplicity.

Denote

$$\iota_k^{\pm} = \frac{2 - n \pm \sqrt{(n-2)^2 + 4(k+\lambda)}}{2};$$

then the solutions to (1-3) have the asymptotics (see [Hartman 1964])

$$V(r) \approx r^{\iota_k^+}, \quad W(r) \approx r^{\iota_k^-}, \quad \text{as } r \to \infty. \tag{1-4}$$

Let $u(r, \Theta)$ be a function on $C_n(\Omega)$. For any given $r \in \mathbb{R}_+$, the integral

$$\int_\Omega u(r, \Theta)\varphi(\Theta)\, dS_1,$$

is denoted by $N_u(r)$, when it exists. The finite or infinite limit

$$\lim_{r\to\infty} V^{-1}(r) N_u(r)$$

is denoted by $\mathcal{U}_u$, when it exists.

We fix an open, nonempty and bounded set $G \subset \partial C_n(\Omega)$. In $C_n(\Omega)$, we normalize the extension, with respect to G, by

$$\mathcal{P}^a_\Omega f(P) = \frac{\mathrm{PI}^a_\Omega f(P)}{\mathrm{PI}^a_\Omega \chi_G(P)}.$$

Let

$$\Gamma(\zeta) = \{P = (r, \Theta) \in C_n(\Omega) : |(r, \Theta) - \zeta| \lesssim \delta(P)\}$$

be a nontangential cone in $C_n(\Omega)$ with vertex $\zeta \in \partial C_n(\Omega)$.

We define

$$\aleph_p(f, l, P) = \left(\frac{1}{l^{n-1}} \int_{B(P,l)} |f(Q)|^p \, d\sigma_Q\right)^{1/p}$$

and

$$\mathbb{E}^p_f(G) = \{P \in G : \aleph_p(f - f(P), l, P) \to 0 \text{ as } l \to 0\}.$$

Note that if $f \in L^p(\partial C_n(\Omega))$, then $|G \backslash \mathbb{E}^p_f(G)| = 0$ (almost every point is a Lebesgue point).

In the proof we need inequalities between the Green function $G^a_\Omega(P, Q)$ and that of the Laplacian, hereafter denoted by $G_\Omega(P, Q)$. It is well known that, for any potential $a(P) \geq 0$,

$$G^a_\Omega(P, Q) \leq G_\Omega(P, Q). \tag{1-5}$$

The inverse inequality is much more elaborate. Cranston, Fabes and Zhao (see [Cranston et al. 1988]; the case $n = 2$ is implicitly contained in [Cranston 1989]) have proved

$$G^a_\Omega(P, Q) \geq M(\Omega)G_\Omega(P, Q), \tag{1-6}$$

where $M(\Omega) = M(\Omega, a(P))$ is a positive constant and does not depend on points P and Q in $C_n(\Omega)$. If $a = 0$, then obviously $M(\Omega) \equiv 1$.

So we have

$$G^a_\Omega(P, Q) \approx G_\Omega(P, Q),$$

from (1-5) and (1-6), which yields that

$$\mathrm{PI}^a_\Omega(P, Q) \approx \mathrm{PI}_\Omega(P, Q). \tag{1-7}$$

Now we state our results, which are due to Qiao [2012] in the case $a = 0$ by the remark. For related results in the half-space and the unit disc, we refer readers to [Mizuta and Shimomura 2003, Theorem 3; Sjögren 1984; 1997; Rönning 1997; Brundin 1999].

Theorem 2. *Let* $1 \leq p < \infty$ *and* $f \in L^p(\partial C_n(\Omega))$. *Then, for any* $\zeta \in \mathbb{E}^p_f(G)$ (*in particular, for a.e.* $\zeta \in G$), *one has that* $\mathcal{P}^a_\Omega f(P) \to f(\zeta)$ *as* $P \to \zeta$ *along* $\Gamma(\zeta)$.

2. Some lemmas

Lemma 1. *For any* $P = (r, \Theta) \in C_n(\Omega)$ *and any* $Q = (t, \Phi) \in S_n(\Omega)$ *satisfying* $0 < t/r \leq \frac{4}{5}$ (*resp.* $0 < r/t \leq \frac{4}{5}$),

$$\mathrm{PI}^a_\Omega(P, Q) \approx t^{-1}V(t)W(r)\varphi(\Theta) \tag{2-1}$$

$$(\textit{resp.}\ \ \mathrm{PI}^a_\Omega(P, Q) \approx V(r)t^{-1}W(t)\varphi(\Theta)). \tag{2-2}$$

For any $P = (r, \Theta) \in C_n(\Omega)$ *and any* $Q = (t, \Phi) \in S_n(\Omega; (4r/5, 5r/4))$,

$$\mathrm{PI}^a_\Omega(P, Q) \approx \frac{r\varphi(\Theta)}{|P - Q|^n}, \tag{2-3}$$

Proof. These immediately follow from [A. Escassut and Yang 2008, Chapter 11], [Essén and Lewis 1973, Lemma 2], [Azarin 1969, Lemma 4 and Remark] and (1-7). □

Lemma 2. $\mathrm{PI}^a_\Omega 1(P) = O(1)$ *as* $P \to \zeta \in G$.

Proof. Write

$$\mathrm{PI}^a_\Omega 1(P) = \int_{E_1} + \int_{E_2} + \int_{E_3} = U_1(P) + U_2(P) + U_3(P),$$

where

$$E_1 = S_n\big(\Omega; \big(0, \tfrac{4}{5}r\big]\big), \quad E_2 = S_n\big(\Omega; \big[\tfrac{5}{4}r, \infty\big)\big), \quad E_3 = S_n\big(\Omega; \big(\tfrac{4}{5}r, \tfrac{5}{4}r\big)\big).$$

By (1-4), (2-1) and (2-2), we have the estimates

$$U_1(P) \approx W(r)\varphi(\Theta) \int_{E_1} t^{\iota_k^+ - 1}\, d\sigma_Q \approx -\frac{s_n}{\iota_k^-} W\Big(\frac{5}{4}\Big)\varphi(\Theta), \tag{2-4}$$

$$U_2(P) \approx \frac{s_n}{\iota_k^+} V\Big(\frac{4}{5}\Big)\varphi(\Theta). \tag{2-5}$$

Next we shall estimate $U_3(P)$. Take a sufficiently small positive number k such that

$$S_n\big(\Omega; \big(\tfrac{4}{5}r, \tfrac{5}{4}r\big)\big) \subset \bigcup_{P=(r,\Theta)\in\Lambda(k)} B\big(P, \tfrac{1}{2}r\big),$$

where

$$\Lambda(k) = \Big\{P = (r, \Theta) \in C_n(\Omega) : \inf_{z\in\partial\Omega} |(1, \Theta) - (1, z)| < k, 0 < r < \infty\Big\}.$$

Since $P \to \zeta \in G$, we only consider the case $P \in \Lambda(k)$. Now, put

$$H_i(P) = \big\{Q \in E_3 : 2^{i-1}\delta(P) \le |P - Q| < 2^i\delta(P)\big\}.$$

Since $S_n(\Omega) \cap \{Q \in \mathbb{R}^n : |P - Q| < \delta(P)\} = \varnothing$, we have by (1-5) and (2-3) that

$$U_3(P) \approx \sum_{i=1}^{i(P)} \int_{H_i(P)} \frac{r\varphi(\Theta)}{|P - Q|^n}\, d\sigma_Q,$$

where $i(P)$ is a positive integer satisfying $2^{i(P)-1}\delta(P) \le r/2 < 2^{i(P)}\delta(P)$.

By (1-2) we have

$$\int_{H_i(P)} \frac{r\varphi(\Theta)}{|P - Q|^n}\, d\sigma_Q \approx r\varphi(\Theta) \int_{H_i(P)} \frac{1}{\delta(P)}\, d\sigma_Q = \frac{r\varphi(\Theta)}{\delta(P)} \frac{s_n}{2^{i(P)}} \approx \frac{s_n}{2^{i(P)}},$$

for $i = 1, 2, \ldots, i(P)$.

So

$$U_3(P) \approx O(1). \tag{2-6}$$

Combining (2-4)–(2-6), Lemma 2 is proved. □

Lemma 3. $\mathrm{PI}_\Omega^a \chi_G(P) = \mathrm{PI}_\Omega^a 1(P) + O(1)$ *as* $P \to \zeta \in G$.

Proof. In fact, we only need to prove

$$U_4(P) = \int_{S_n(\Omega)-G} \mathrm{PI}_\Omega^a(P, Q)\, d\sigma_Q \lesssim O(1). \tag{2-7}$$

Write

$$\begin{aligned} U_4(P) &= \int_{(S_n(\Omega)-G)\cap E_1} + \int_{(S_n(\Omega)-G)\cap E_2} + \int_{(S_n(\Omega)-G)\cap E_3} \\ &= U_5(P) + U_6(P) + U_7(P). \end{aligned}$$

Obviously

$$U_5(P) \lesssim U_1(P) \approx O(1), \tag{2-8}$$

$$U_6(P) \lesssim U_2(P) \approx O(1). \tag{2-9}$$

Further, we have by (2-3) that

$$\begin{aligned} U_7(P) &\approx r\varphi(\Theta) \int_{(S_n(\Omega)-G)\cap E_3} \frac{1}{|P-Q|^n}\, d\sigma_Q \\ &\lesssim \frac{s_n}{d} |\zeta| \varphi(\Theta) \quad (P \to \zeta \in G), \end{aligned} \tag{2-10}$$

where $d = \inf_{Q \in \partial C_n(\Omega)-G} |Q-\zeta|$.

Combining (2-8)–(2-10), (2-7) holds, which gives the conclusion. □

3. Proof of the theorem

As $P \to \zeta \in G$,

$$\mathrm{PI}_\Omega^a \chi_G(P) = O(1) \neq 0,$$

from Lemmas 2 and 3.

Now let $f \in L^p(\partial C_n(\Omega))$ and $\zeta \in \mathbb{E}_f^p(G)$ be given. We may, without loss of generality, assume that $f(\zeta) = 0$. Furthermore we assume that $P = (r, \Theta) \in \Gamma(\zeta)$. Let $s = |(r, \Theta) - \zeta|$. We write

$$\begin{aligned} \mathrm{PI}_\Omega^a f(P) &= \int_{E_1} + \int_{E_2} + \int_{E_3 \cap B(\zeta, 2s)} + \int_{E_3 \cap B^c(\zeta, 2s)} \\ &= V_1 f(P) + V_2 f(P) + V_3 f(P) + V_4 f(P). \end{aligned}$$

By using Hölder's inequality, (1-4), (2-1) and (2-2), we have the estimates

$$|V_1 f(P)| \lesssim W(r)\varphi(\Theta) \int_{E_1} t^{\alpha-1} f(Q)\, d\sigma_Q \lesssim r^{(1-n)/p} \|f\|_p,$$

$$|V_2 f(P)| \lesssim r^{(1-n)/p} \|f\|_p.$$

Similar to the estimate of $U_3(P)$ in Lemma 2, we only consider the following inequality by (1-2):

$$\int_{H_i(P)} \frac{r\varphi(\Theta)}{|P-Q|^n}\,d\sigma_Q \approx r\varphi(\Theta)\int_{H_i(P)} \frac{1}{(2^{i-1}\delta(P))^n}\,d\sigma_Q$$
$$\lesssim r^{\iota_0^+}\varphi(\Theta)\int_{E_2} t^{\iota_0^- -1}|f(Q)|\,d\sigma_Q \lesssim r^{(1-n)/p}\|f\|_p,$$

for $i=0,1,2,\dots,i(P)$, which is similar to the estimate of $V_2 f(P)$.

So

$$|V_3 f(P)| \lesssim r^{(1-n)/p}\|f\|_p.$$

Notice that $|P-Q| > \frac{1}{2}|\zeta-Q|$ in the case $Q\in E_3\cap B^c(\zeta,2s)$. By (1-2) and (2-3), we have

$$\begin{aligned}
|V_4 f(P)| &\lesssim \delta(P)\int_{E_3\cap B^c(\zeta,2s)} \frac{|f(Q)|}{|P-Q|^n}\,d\sigma_Q\\
&\lesssim \delta(P)\sum_{i=1}^{\infty}\int_{E_3\cap(B(\zeta,2^{i+1}s)\setminus B(\zeta,2^i s))} \frac{|f(Q)|}{|\zeta-Q|^n}\,d\sigma_Q\\
&\lesssim \delta(P)\sum_{i=1}^{\infty}\Big(\frac{1}{2^i s}\Big)^n\int_{E_3\cap B(\zeta,2^{i+1}s)} |f(Q)|\,d\sigma_Q\\
&\lesssim \delta(P)\sum_{i=1}^{\infty}\aleph_1(f,2^{i+1}s,\zeta) \lesssim \delta(P)\sum_{i=1}^{\infty}\int_{2^{i+1}s}^{2^{i+2}s}\frac{\aleph_1(f,l,\zeta)}{l}\,dl\\
&\lesssim \delta(P)\int_s^{\infty}\frac{\aleph_1(f,l,\zeta)}{l}\,dl \lesssim \delta(P)\int_{\delta(P)}^{\infty}\frac{\aleph_1(f,l,\zeta)}{l}\,dl.
\end{aligned}$$

Thus it follows that

$$\begin{aligned}
|\mathcal{P}_\Omega f(P)| &\lesssim \frac{1}{O(1)}\big[|V_1 f(P)|+|V_2 f(P)|+|V_3 f(P)|+|V_4 f(P)|\big]\\
&\lesssim r^{(1-n)/p}\|f\|_p + \delta(P)\int_{\delta(P)}^{\infty}\frac{\aleph_1(f,l,\zeta)}{l}\,dl.
\end{aligned}$$

Using the fact that $s\lesssim\delta(P)\lesssim r\varphi(\Theta)$, we get

$$|\mathcal{P}_\Omega f(P)| \lesssim \aleph_1(f,2s,\zeta) + \delta(P)\int_{\delta(P)}^{\infty}\frac{\aleph_1(f,l,\zeta)}{l}\,dl.$$

It is clear that

$$\int_{\delta(P)}^{\infty}\frac{\aleph_1(f,l,\zeta)}{l}\,dl$$

is a convergent integral, since

$$\frac{\aleph_1(f,l,\zeta)}{l} \lesssim s^{-1-n} s^{n/q} \|f\|_p \lesssim s^{-1-n/p} \|f\|_p,$$

from Hölder's inequality.

Now, as $\delta(P) \to 0$, we also have $s \to 0$. Since $f(\zeta) = 0$ and since we have assumed that $\zeta \in \mathbb{E}_f^p(G)$ (and thus that $\zeta \in \mathbb{E}_f^1(G)$), it follows that $\mathcal{P}_\Omega^a f(P) \to 0 = f(\zeta)$ as $P = (r, \Theta) \to \zeta$ along $\Gamma(\zeta)$. This concludes the proof.

References

[A. Escassut and Yang 2008] W. T. A. Escassut and C. C. Yang, *Some topics on value distribution and differentiability in complex and p-adic analysis*, Mathematics Monograph Series **11**, Science Press, Beijing, 2008. MR 2011j:30002 Zbl 1233.30005

[Azarin 1969] V. S. Azarin, "Generalization of a theorem of Hayman on subharmonic functions in an m-dimensional cone", *Amer. Math. Soc. Transl.* (2) **80** (1969), 119–138. Zbl 0188.41903

[Brundin 1999] M. Brundin, *Approach regions for the square root of the Poisson kernel and weak L^p boundary functions*, Thesis for the degree of Licentiate of Philosophy, Göteborg University and Chalmers University of Technology, 1999, available at http://www.math.chalmers.se/Math/Research/Preprints/1999/56.ps.gz.

[Courant and Hilbert 1953] R. Courant and D. Hilbert, *Methods of mathematical physics, I*, Interscience, New York, 1953. MR 16,426a Zbl 0053.02805

[Cranston 1989] M. Cranston, "Conditional Brownian motion, Whitney squares and the conditional gauge theorem", pp. 109–119 in *Seminar on Stochastic Processes* (Gainesville, FL, 1988), edited by E. Çinlar et al., Progr. Probab. **17**, Birkhäuser, Boston, 1989. MR 90j:60078 Zbl 0689.60077

[Cranston et al. 1988] M. Cranston, E. Fabes, and Z. Zhao, "Conditional gauge and potential theory for the Schrödinger operator", *Trans. Amer. Math. Soc.* **307**:1 (1988), 171–194. MR 90a:60135 Zbl 0652.60076

[Essén and Lewis 1973] M. Essén and J. L. Lewis, "The generalized Ahlfors–Heins theorem in certain d-dimensional cones", *Math. Scand.* **33** (1973), 113–129. MR 50 #629 Zbl 0273.31004

[Gilbarg and Trudinger 1977] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Grundlehren der Mathematischen Wissenschaften **224**, Springer, New York, 1977. MR 57 #13109 Zbl 0361.35003

[Hartman 1964] P. Hartman, *Ordinary differential equations*, Wiley, New York, 1964. MR 30 #1270 Zbl 0125.32102

[Miranda 1970] C. Miranda, *Partial differential equations of elliptic type*, 2nd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete **2**, Springer, Berlin, 1970. MR 44 #1924 Zbl 0198.14101

[Mizuta and Shimomura 2003] Y. Mizuta and T. Shimomura, "Growth properties for modified Poisson integrals in a half space", *Pacific J. Math.* **212**:2 (2003), 333–346. MR 2004m:31002 Zbl 1056.31004

[Qiao 2012] L. Qiao, "Nontangential limits for modified Poisson integrals of boundary functions in a cone", *J. Funct. Spaces Appl.* (2012), Art. ID 825240. MR 2957158 Zbl 1248.31002

[Reed and Simon 1979] M. Reed and B. Simon, *Methods of modern mathematical physics, III: Scattering theory*, Academic Press, New York, 1979. MR 80m:81085 Zbl 0405.47007

[Rönning 1997] J.-O. Rönning, "Convergence results for the square root of the Poisson kernel", *Math. Scand.* **81**:2 (1997), 219–235. MR 99e:31004 Zbl 0902.42009

[Simon 1982] B. Simon, "Schrödinger semigroups", *Bull. Amer. Math. Soc.* (*N.S.*) **7**:3 (1982), 447–526. MR 86b:81001a Zbl 0524.35002

[Sjögren 1984] P. Sjögren, "Une remarque sur la convergence des fonctions propres du laplacien à valeur propre critique", pp. 544–548 in *Théorie du potentiel* (Orsay, 1983), edited by G. Mokobodzki and D. Pinchon, Lecture Notes in Math. **1096**, Springer, Berlin, 1984. MR 88f:31001 Zbl 0563.31003

[Sjögren 1997] P. Sjögren, "Approach regions for the square root of the Poisson kernel and bounded functions", *Bull. Austral. Math. Soc.* **55**:3 (1997), 521–527. MR 98h:31002 Zbl 0903.31002

[Verzhbinskii and Maz'ya 1971] G. M. Verzhbinskii and V. G. Maz'ya, "Asymptotic behavior of solutions of elliptic equations of the second order close to a boundary, I", *Sib. Mat. Zh.* **12**:6 (1971), 1217–1249. In Russian; translated in *Sib. Mat. J.* **12**:6 (1971), 874–899. Zbl 0229.35008

[Yoshida 1991] H. Yoshida, "Harmonic majorization of a subharmonic function on a cone or on a cylinder", *Pacific J. Math.* **148**:2 (1991), 369–395. MR 92c:31007 Zbl 0738.31005

Received August 22, 2013. Revised December 29, 2013.

LEI QIAO
HENAN UNIVERSITY OF ECONOMICS AND LAW
SCHOOL OF MATHEMATICS AND INFORMATION SCIENCE
ZHENGZHOU, 450046
CHINA
qiaocqu@163.com

TAO ZHAO
FACULTY OF SCIENCE
OKAYAMA UNIVERSITY
OKAYAMA, 700-8350
JAPAN
t_t_zhao@yeah.net

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.272.237

JACOBI–TRUDI DETERMINANTS AND CHARACTERS OF MINIMAL AFFINIZATIONS

STEVEN V SAM

In their study of characters of minimal affinizations of representations of orthogonal and symplectic Lie algebras, Chari and Greenstein conjectured that certain Jacobi–Trudi determinants satisfy an alternating sum formula. In this note, we prove their conjecture and slightly more. The proof relies on some symmetries of the ring of symmetric functions discovered by Koike and Terada. Using results of Hernandez, Mukhin and Young, and Naoi, this implies that the characters of minimal affinizations in types B, C, and D are given by a Jacobi–Trudi determinant.

Introduction

In [Chari and Greenstein 2011] (henceforth abbreviated [CG]), the authors study a class of modules over the current algebra $\mathfrak{g} \otimes \mathbb{C}[t]$, where $\mathfrak{g}$ is either a special orthogonal or symplectic Lie algebra (over the complex numbers). These modules are related to the minimal affinizations, a class of irreducible representations for the quantum loop algebra $\mathrm{U}_q(\mathfrak{g} \otimes \mathbb{C}[t, t^{-1}])$. We refer the reader to [CG, §3] for background and references. A character formula, which is similar to a Jacobi–Trudi determinant, for these modules is conjectured in [CG, Conjecture 1.13]. This is inspired by [Nakai and Nakanishi 2006], which conjectures that the characters of minimal affinizations are given by such determinants (see also [Nakai and Nakanishi 2007a; 2007b] for related work).

The aim of this note is to prove [CG, Conjecture 1.13] (see Theorem 1.1). We will give a uniform proof for all types. The conjecture reduces to a combinatorial statement about characters of $\mathfrak{g}$, so we will not need to discuss current or loop algebras any further. In fact, we will prove an extension of the combinatorial statement which removes a restriction on the highest weights considered. Furthermore, using results of Hernandez, Mukhin and Young, and Naoi, this gives a character formula for minimal affinizations of representations of $\mathfrak{g}$ in types B, C, and D (see Remark 1.3).

The method of proof involves passing to a suitable limit (with respect to the rank of the Lie algebra) to take advantage of additional symmetries. This suggests that

The author was supported by a Miller research fellowship.
MSC2010: primary 05E05; secondary 17B10.
Keywords: minimal affinizations, classical Lie algebras, symmetric functions.

there should be a connection to the categories Rep(O) and Rep(Sp) studied in [Sam and Snowden 2013, §4] and a suitable categorification of the involutions i_{O} and i_{Sp} used in Section 3 (which were introduced in [Koike and Terada 1987]), but we have been unable to find one so far.

1. Notation

We need some basic terminology of partitions [Macdonald 1995, §I.1]. A partition λ is a sequence of integers $(\lambda_1, \dots, \lambda_r)$ with $\lambda_1 \geq \cdots \geq \lambda_r \geq 0$. We set $|\lambda| = \sum_i \lambda_i$ and $\ell(\lambda) = \max\{i \mid \lambda_i \neq 0\}$. We write $\mu \subseteq \lambda$ if $\mu_i \leq \lambda_i$ for all i and also say that λ contains μ. The notation a^b means the sequence $(a, a, \dots, a)$ where a appears b times. We use $\lambda^\dagger$ to denote the transpose partition of λ, i.e., $\lambda_i^\dagger = \#\{j \mid \lambda_j \geq i\}$ (in terms of Young diagrams, we are flipping across the diagonal). Let $\boldsymbol{S}_\lambda$ denote the corresponding Schur functor [Fulton and Harris 1991, §6.1]; for the purposes of this note, $\boldsymbol{S}_\lambda$ is a functor from the category of complex vector spaces to itself. Special cases are symmetric powers $\boldsymbol{S}_k = \mathrm{Sym}^k$ and exterior powers $\boldsymbol{S}_{(1^k)} = \bigwedge^k$. We use s_λ to denote the Schur function indexed by λ [Macdonald 1995, §I.3] (it is the character of $\boldsymbol{S}_\lambda$). The product of two Schur functions is a linear combination of Schur functions:

$$s_\mu s_\nu = \sum_\lambda c^\lambda_{\mu,\nu} s_\lambda.$$

The $c^\lambda_{\mu,\nu}$ are the Littlewood–Richardson coefficients [Macdonald 1995, §I.9]. If $c^\lambda_{\mu,\nu} \neq 0$, then $|\lambda| = |\mu| + |\nu|$ and also $\mu \subseteq \lambda$ and $\nu \subseteq \lambda$.

Let G be a complex classical group of type B_n, C_n, or D_{n+1}, i.e., G is either $\mathrm{O}_{2n+1}(\mathbb{C})$, $\mathrm{Sp}_{2n}(\mathbb{C})$, or $\mathrm{O}_{2n+2}(\mathbb{C})$, respectively. Let $\mathfrak{g}$ be the Lie algebra of G. Let $\mathrm{rank}(\mathfrak{g})$ be the rank of $\mathfrak{g}$; i.e., it is n in the cases of type B and C, and it is $n+1$ in the case of type D. We use these groups rather than their Lie algebras to avoid having to make technical remarks later. For the representations considered in [CG], this choice will not be important. We number the nodes of the Dynkin diagram according to Bourbaki notation:

$$\mathrm{B}_n: \quad \begin{array}{ccccccccccc} \circ & - & \circ & - & \cdots & - & \circ & - & \circ & \Rightarrow & \circ \\ 1 & & 2 & & & & n-2 & & n-1 & & n \end{array}$$

$$\mathrm{C}_n: \quad \begin{array}{ccccccccccc} \circ & - & \circ & - & \cdots & - & \circ & - & \circ & \Leftarrow & \circ \\ 1 & & 2 & & & & n-2 & & n-1 & & n \end{array}$$

$$\mathrm{D}_{n+1}: \quad \begin{array}{ccccccccccc} & & & & & & & & \circ & & \\ & & & & & & & & | \; n+1 & & \\ \circ & - & \circ & - & \cdots & - & \circ & - & \circ & - & \circ \\ 1 & & 2 & & & & n-2 & & n-1 & & n \end{array}$$

Let ω_i be the fundamental weights, and let λ be a dominant integral weight which is a linear combination of $\omega_1, \dots, \omega_{n-1}$ (so in particular, we avoid the spin representations in the orthogonal case). We will use a basis $e_1, \dots, e_{\mathrm{rank}(\mathfrak{g})}$ for the weight lattice of G (see [Fulton and Harris 1991, §16.1, §18.1] for details; there the basis is denoted $L_1, \dots, L_{\mathrm{rank}(\mathfrak{g})}$). Given $\lambda = a_1\omega_1 + \cdots + a_{n-1}\omega_{n-1}$, we associate to it the partition

$$(a_1 + \cdots + a_{n-1}, a_2 + \cdots + a_{n-1}, \dots, a_{n-1}).$$

So in particular, the notation $\lambda_i = a_i + \cdots + a_{n-1}$ is defined. Then we have $\lambda = \lambda_1 e_1 + \cdots + \lambda_{n-1} e_{n-1}$. Let V_λ be the corresponding highest weight representation of G. We will denote $V = V_1$, the vector representation. We sometimes use the notation V_λ^{O} or V_λ^{Sp} to emphasize that we are dealing with the orthogonal or symplectic case, respectively.

In general, all finite-dimensional irreducible representations V_λ of G can be indexed by partitions λ (see [Fulton and Harris 1991, §17.3, §19.5] or [Sam and Snowden 2013, §4.1]). We may assume that $\ell(\lambda) \le \mathrm{rank}(\mathfrak{g})$ as long as we are ambivalent about the presence of the sign representation in the orthogonal group case. (The reason we do not use the special orthogonal group is because some irreducible representations of the even orthogonal group are not irreducible when restricted to the special orthogonal group, and so the latter group does not behave as well from the perspective of stability.)

Now we rephrase the definitions in [CG, §1.13] in this notation. First, we have $i_\lambda = \ell(\lambda)$. In the orthogonal case, $\Psi_\lambda = \{e_i + e_j \mid 1 \le i < j \le \ell(\lambda)\}$, and in the symplectic case, $\Psi_\lambda = \{e_i + e_j \mid 1 \le i \le j \le \ell(\lambda)\}$. Define the set

$$\Gamma(\lambda, \Psi_\lambda) = \left\{ (\mu, s) \;\middle|\; \lambda = \mu + \sum_{\beta \in \Psi_\lambda} n_\beta \beta,\ n_\beta \in \mathbb{Z}_{\ge 0},\ \sum_{\beta \in \Psi_\lambda} n_\beta = s \right\}.$$

By the definitions of Ψ_λ, we see that $(\mu, s) \in \Gamma(\lambda, \Psi_\lambda)$ implies that $s = (|\lambda| - |\mu|)/2$.

Define $\boldsymbol{h}_k = \mathrm{char}(V_k^{\mathrm{O}})$ in the orthogonal case and $\boldsymbol{h}_k = \sum_{0 \le r \le k/2} \mathrm{char}(V_{k-2r}^{\mathrm{Sp}})$ in the symplectic case. In both cases, define the Jacobi–Trudi determinant

$$\boldsymbol{H}_\lambda = \det(\boldsymbol{h}_{\lambda_i - i + j}).$$

For $(\nu, s) \in \Gamma(\lambda, \Psi_\lambda)$, define

$$C^\lambda_{\nu,s} = \dim \hom_G(V_\nu, \textstyle\bigwedge^s(\mathfrak{g}) \otimes V_\lambda)$$

(see [CG, §2.7], but there it is c instead of C; we use c for Littlewood–Richardson coefficients).

All of the above definitions make sense for any partition λ with $\ell(\lambda) \le \mathrm{rank}(\mathfrak{g})$. To make this clear, we spell out the conversion between partitions and weights now. Let $r = \mathrm{rank}(\mathfrak{g})$ and let $\lambda = (\lambda_1, \dots, \lambda_r)$ be a partition.

- If $G = \mathrm{Sp}_{2r}(\mathbb{C})$, then V_λ is irreducible with highest weight

$$\sum_{i=1}^{r-1}(\lambda_i - \lambda_{i+1})\omega_i + \lambda_r\omega_r.$$

- If $G = \mathrm{O}_{2r+1}(\mathbb{C})$, then V_λ is irreducible with highest weight

$$\sum_{i=1}^{r-1}(\lambda_i - \lambda_{i+1})\omega_i + 2\lambda_r\omega_r.$$

- If $G = \mathrm{O}_{2r}(\mathbb{C})$, then there are two cases. In both cases, V_λ is an irreducible representation of $\mathrm{O}_{2r}(\mathbb{C})$, but we distinguish between what happens when we pass to the Lie algebra $\mathfrak{so}_{2r}(\mathbb{C})$.
  - If $\lambda_r = 0$, then V_λ is an irreducible representation of $\mathfrak{so}_{2r}(\mathbb{C})$ with highest weight

$$\sum_{i=1}^{r-2}(\lambda_i - \lambda_{i+1})\omega_i + \lambda_{r-1}(\omega_{r-1} + \omega_r).$$

  - If $\lambda_r > 0$, then as a representation of $\mathfrak{so}_{2r}(\mathbb{C})$, V_λ is the direct sum of irreducible representations with highest weights

$$\sum_{i=1}^{r-2}(\lambda_i - \lambda_{i+1})\omega_i + (\lambda_{r-1} - \lambda_r)\omega_{r-1} + (\lambda_{r-1} + \lambda_r)\omega_r$$

    and

$$\sum_{i=1}^{r-2}(\lambda_i - \lambda_{i+1})\omega_i + (\lambda_{r-1} + \lambda_r)\omega_{r-1} + (\lambda_{r-1} - \lambda_r)\omega_r.$$

In the orthogonal case, let d_ν^λ be the multiplicity of V_ν^{Sp} in $\mathbf{S}_\lambda(V^{\mathrm{Sp}})$: here V^{Sp} is the vector representation for $\mathrm{Sp}(2n)$ with $n \ge \ell(\lambda)$ and $\mathbf{S}_\lambda(V^{\mathrm{Sp}})$ is considered as a representation of $\mathrm{Sp}(2n)$. By [Koike and Terada 1987, Proposition 1.5.3], this multiplicity is independent of n as long as $n \ge \ell(\lambda)$, and we have

$$d_\nu^\lambda = \sum_\eta c^\lambda_{\nu,(2\eta)^\dagger}.$$

Similarly, in the symplectic case, let d_ν^λ be the multiplicity of V_ν^{O} in $\mathbf{S}_\lambda(V^{\mathrm{O}})$ (note that we are using branching rules for the *other* group in both cases). Then we have

$$d_\nu^\lambda = \sum_\eta c^\lambda_{\nu,2\eta}.$$

When $\ell(\lambda) \leq n-1$, the following main result proves [CG, Conjecture 1.13].

Theorem 1.1. *Let λ be a partition with $\ell(\lambda) \leq \operatorname{rank}(\mathfrak{g})$. Then*

$$\sum_{(\nu,s)\in\Gamma(\lambda,\Psi_\lambda)} (-1)^s C^\lambda_{\nu,s} \boldsymbol{H}_\nu = \operatorname{char}(V_\lambda). \tag{1.2}$$

Also $\boldsymbol{H}_\lambda = \sum_\nu d^\lambda_\nu \operatorname{char}(V_\nu)$.

Remark 1.3. Under the restriction $\ell(\lambda) \leq n-1$, Chari and Greenstein constructed the module $P(\lambda, 0)^{\Gamma(\lambda,\Psi_\lambda)}$ in [CG], and Theorem 1.1 together with [CG, Theorem 2] shows that its character is $\boldsymbol{H}_\lambda$. In types B and C, Naoi [2013, Remark 4.7] shows that these modules are the "graded limits" of the minimal affinizations of the corresponding simple modules V_λ of $\mathfrak{g}$. A similar result is obtained for a special class of highest weights in type D in [Naoi 2014]. In particular, the characters (considered as representations of $\mathfrak{g}$) of both modules are the same. So the character of the minimal affinization is also $\boldsymbol{H}_\lambda$. In type B, this follows from [Hernandez 2007] (see [Naoi 2013, Remark 4.7]) or from [Mukhin and Young 2012, Corollary 7.6]. □

2. Some identities

Let Q_{-1} be the set of partitions with the following inductive definition. The empty partition belongs to Q_{-1}. A nonempty partition μ belongs to Q_{-1} if and only if the number of rows in μ is one more than the number of columns, i.e., $\ell(\mu) = \mu_1 + 1$, and the partition obtained by deleting the first row and column of μ, i.e., $(\mu_2 - 1, \ldots, \mu_{\ell(\mu)} - 1)$, belongs to Q_{-1}. The first few partitions in Q_{-1} are 0, (1, 1), (2, 1, 1), (2, 2, 2). Define $Q_1 = \{\lambda \mid \lambda^\dagger \in Q_{-1}\}$. We record this definition as the following formula:

$$Q_1^\dagger = Q_{-1}. \tag{2.1}$$

The significance of these sets are the following decompositions (see [Macdonald 1995, I.A.7, Examples 4, 5]):

$$\textstyle\bigwedge^i(\operatorname{Sym}^2(E)) = \bigoplus_{\substack{\mu\in Q_1\\ |\mu|=2i}} \boldsymbol{S}_\mu(E), \tag{2.2}$$

$$\textstyle\bigwedge^i\big(\bigwedge^2(E)\big) = \bigoplus_{\substack{\mu\in Q_{-1}\\ |\mu|=2i}} \boldsymbol{S}_\mu(E). \tag{2.3}$$

We need two of Littlewood's identities [Koike and Terada 1987, Proposition 1.5.3]:

$$\operatorname{char}(V^{\mathrm{O}}_\lambda) = \sum_{\mu\in Q_1} (-1)^{|\mu|/2} \sum_\nu c^\lambda_{\mu,\nu} s_\nu, \tag{2.4}$$

$$\operatorname{char}(V^{\mathrm{Sp}}_\lambda) = \sum_{\mu\in Q_{-1}} (-1)^{|\mu|/2} \sum_\nu c^\lambda_{\mu,\nu} s_\nu. \tag{2.5}$$

Lemma 2.6. *Fix* $(\nu, s) \in \Gamma(\lambda, \Psi_\lambda)$, *where* $\ell(\lambda) \leq \operatorname{rank}(\mathfrak{g})$ *and* $s = (|\lambda| - |\nu|)/2$. *Then* $C^\lambda_{\nu,s} = \sum_{\mu \in Q_{-1}} c^\lambda_{\mu,\nu}$ *in the orthogonal case (for the symplectic case, use* Q_1 *instead of* Q_{-1}*).*

Conversely, if this sum is nonzero, then $(\nu, s) \in \Gamma(\lambda, \Psi_\lambda)$ *for* $s = (|\lambda| - |\nu|)/2$.

Proof. In the orthogonal case, we have $\mathfrak{g} = V_{1,1} = \bigwedge^2(V)$. So we need to calculate the multiplicity of V_ν in $\bigwedge^s\big(\bigwedge^2(V)\big) \otimes V_\lambda$, where $s = (|\lambda| - |\nu|)/2$. By (2.3), we get

$$\bigwedge^s\big(\bigwedge^2(V)\big) = \bigoplus_{\substack{\mu \in Q_{-1} \\ |\mu| = 2s}} \boldsymbol{S}_\mu(V).$$

(In the symplectic case we instead have $\mathfrak{g} = V_2 = \operatorname{Sym}^2(V)$, so all of the following statements will hold if we replace Q_{-1} with Q_1.) We claim that the multiplicity of V_ν in $\boldsymbol{S}_\mu(V) \otimes V_\lambda$ is the Littlewood–Richardson coefficient $c^\lambda_{\mu,\nu}$.

If $\ell(\mu) \leq \operatorname{rank}(\mathfrak{g})$, then as a representation of the orthogonal group (also in the symplectic case), $\boldsymbol{S}_\mu(V)$ is the sum of V_μ and other V_α, where $|\alpha| < |\mu|$ up to twisting V_α with a sign character (this follows from the explicit formula in [Koike and Terada 1987, Proposition 2.5.1]). Also, if V_ν appears in $V_\alpha \otimes V_\lambda$, then we must have $|\nu| \geq |\lambda| - |\alpha|$ by a basic argument with weights. This implies that the multiplicity of V_ν in $\boldsymbol{S}_\mu(V) \otimes V_\lambda$ is the same as the multiplicity of V_ν in $V_\mu \otimes V_\lambda$ under our hypothesis that $|\nu| + |\mu| = |\lambda|$. Furthermore, the multiplicity in this case is the Littlewood–Richardson coefficient $c^\lambda_{\mu,\nu}$ [ibid., Proposition 2.5.2].

If $\ell(\mu) > \operatorname{rank}(\mathfrak{g})$, then the multiplicity of V_ν in $\boldsymbol{S}_\mu(V) \otimes V_\lambda$ is 0 since all V_α in $\boldsymbol{S}_\mu(V)$ satisfy $|\alpha| < |\mu|$. Also, $c^\lambda_{\mu,\nu} = 0$ since $\mu \not\subseteq \lambda$. This proves the claim and the second sentence of the lemma.

Now we handle the last sentence of the lemma. So suppose that $c^\lambda_{\mu,\nu} \neq 0$ for some $\mu \in Q_{-1}$. Set $s = (|\lambda| - |\nu|)/2 = |\mu|/2$. The weights of $\boldsymbol{S}_\mu(V) \subset \bigwedge^s(\mathfrak{g})$ are linear combinations of s roots of $\mathfrak{g}$. In particular, λ is the sum of ν and s roots $\alpha_1, \dots, \alpha_s$ of $\mathfrak{g}$. The possible roots of $\mathfrak{g}$ are $e_i \pm e_j$ and $\pm e_i$. Since $|\nu + e_i - e_j| = |\nu|$ and $|\nu \pm e_i| = |\nu| \pm 1$, the s roots $\alpha_1, \dots, \alpha_s$ must all be of the form $e_i + e_j$, so $(\nu, s) \in \Gamma(\lambda, \Psi_\lambda)$. □

3. Proof of main theorem

Lemma 3.1. *Pick* $\mathrm{X} \in \{\mathrm{B}, \mathrm{C}, \mathrm{D}\}$. *Fix a partition* λ *with* $\ell(\lambda) \leq n$. *Then* (1.2) *is true for the representation* V_λ *for* X_n *if and only if it is true for the representation* V_λ *for* X_m *for any* $m \geq n$.

Proof. By [Koike and Terada 1987, Corollary 2.5.3], the tensor product decomposition $V_\lambda \otimes V_\mu$ is independent of m if $m \geq \ell(\lambda) + \ell(\mu)$, and in this case, the tensor product decomposes as a sum of V_α with $\ell(\alpha) \leq \ell(\lambda) + \ell(\mu)$. The definition of $\boldsymbol{H}_\lambda$ involves multiplying at most $\ell(\lambda) \leq m$ characters, all indexed by one-row partitions, so its definition is independent of m. Certainly the set $\Gamma(\lambda, \Psi_\lambda)$ does not depend on m if $m \geq \ell(\lambda)$. So it remains to show that the coefficients $C^\lambda_{\nu,s}$ are independent of m, but this follows from Lemma 2.6. $\square$

In particular, we may assume that $n = \infty$. In this limit, we can use some additional symmetries of the character ring Λ of $\mathfrak{g}$. Then Λ is the ring of symmetric functions, but is equipped with a new basis which was studied in [ibid.]. Write $s_{[\lambda]} = \operatorname{char}(V_\lambda)$. We use $s^{\mathrm{Sp}}_{[\lambda]}$ or $s^{\mathrm{O}}_{[\lambda]}$ if we need to emphasize the group. Then the $s_{[\lambda]}$, as λ ranges over all partitions, forms a basis for this character ring. The idea is to use (2.4) or (2.5) to exhibit the change of basis between $s_{[\lambda]}$ and the usual Schur functions $s_\mu = \operatorname{char}(\boldsymbol{S}_\mu(V))$. There is an involution (which is an algebra automorphism), denoted i_{O} in the orthogonal case and i_{Sp} in the symplectic case, that sends $s_{[\lambda]}$ to $s_{[\lambda^\dagger]}$ [ibid., Theorem 2.3.4]. Also, we recall that the linear map $\omega\colon s_\lambda \mapsto s_{\lambda^\dagger}$ is an algebra automorphism [Macdonald 1995, §I.3]. We need the following identity [Koike and Terada 1987, Theorem 2.3.2]:

$$\omega(s^{\mathrm{Sp}}_{[\lambda]}) = s^{\mathrm{O}}_{[\lambda^\dagger]}. \tag{3.2}$$

Lemma 3.3. *The involution i_{O} or i_{Sp} sends $\boldsymbol{H}_\nu$ to the Schur function $s_{\nu^\dagger}$.*

Proof. In the orthogonal case, $i_{\mathrm{O}}(\boldsymbol{h}_k) = s_{[1^k]} = \operatorname{char}\big(\bigwedge^k V\big) = s_{1^k}$, and in the symplectic case,

$$i_{\mathrm{Sp}}(\boldsymbol{h}_k) = \sum_{0 \leq r \leq k/2} s_{[1^{k-2r}]} = \operatorname{char}\big(\textstyle\bigwedge^k V\big) = s_{1^k}$$

by basic properties of the decomposition of exterior powers under the action of the symplectic group. Since i_{O} and i_{Sp} are algebra homomorphisms, we see that $\boldsymbol{H}_\nu = \det(\boldsymbol{h}_{\nu_i - i + j})$ gets sent to $\det(s_{1^{\nu_i - i + j}})$, which is the Schur function $s_{\nu^\dagger}$ by the Jacobi–Trudi formula [Macdonald 1995, §I.3, equation (3.5)]. $\square$

Now we focus on the orthogonal case (the symplectic case is almost identical). By (2.4),

$$s_{[\lambda]} = \sum_{\mu \in Q_1} (-1)^{|\mu|/2} \sum_\nu c^\lambda_{\mu,\nu} s_\nu.$$

Since $c^\lambda_{\mu,\nu} = c^{\lambda^\dagger}_{\mu^\dagger,\nu^\dagger}$ (use that $s_\mu s_\nu = \sum_\lambda c^\lambda_{\mu,\nu} s_\lambda$ [Macdonald 1995, §I.9] and the involution ω defined above), and $Q_1^\dagger = Q_{-1}$ (2.1), we can rewrite this as

$$s_{[\lambda^\dagger]} = \sum_{\mu \in Q_{-1}} (-1)^{|\mu|/2} \sum_\nu c^\lambda_{\mu,\nu} s_{\nu^\dagger}.$$

In particular, the coefficient of $s_{\nu^\dagger}$ is $\sum_{\mu \in Q_{-1}} (-1)^{(|\lambda| - |\nu|)/2} c^\lambda_{\mu,\nu}$. By Lemma 2.6, we get

$$s_{[\lambda^\dagger]} = \sum_{(\nu,s) \in \Gamma(\lambda,\Psi_\lambda)} (-1)^s C^\lambda_{\nu,s} s_{\nu^\dagger}.$$

Finally, apply the involution i_{O} to this equation and use Lemma 3.3 to get (1.2). The last part of the theorem follows directly from Lemma 3.3 and (3.2).

References

[Chari and Greenstein 2011] V. Chari and J. Greenstein, "Minimal affinizations as projective objects", *J. Geom. Phys.* **61**:3 (2011), 594–609. MR 2012a:17019 Zbl 1207.81040

[Fulton and Harris 1991] W. Fulton and J. Harris, *Representation theory: A first course*, Graduate Texts in Mathematics **129**, Springer, New York, 1991. MR 93a:20069 Zbl 0744.22001

[Hernandez 2007] D. Hernandez, "On minimal affinizations of representations of quantum groups", *Comm. Math. Phys.* **276**:1 (2007), 221–259. MR 2008h:17014 Zbl 1141.17011

[Koike and Terada 1987] K. Koike and I. Terada, "Young-diagrammatic methods for the representation theory of the classical groups of type B_n, C_n, D_n", *J. Algebra* **107**:2 (1987), 466–511. MR 88i:22035 Zbl 0622.20033

[Macdonald 1995] I. G. Macdonald, *Symmetric functions and Hall polynomials*, 2nd ed., Clarendon Press, New York, 1995. MR 96h:05207 Zbl 0824.05059

[Mukhin and Young 2012] E. Mukhin and C. A. S. Young, "Path description of type B q-characters", *Adv. Math.* **231**:2 (2012), 1119–1150. MR 2955205 Zbl 0608.4095

[Nakai and Nakanishi 2006] W. Nakai and T. Nakanishi, "Paths, tableaux and q-characters of quantum affine algebras: The C_n case", *J. Phys. A* **39**:9 (2006), 2083–2115. MR 2006k:17028 Zbl 1085.17011

[Nakai and Nakanishi 2007a] W. Nakai and T. Nakanishi, "Paths and tableaux descriptions of Jacobi–Trudi determinant associated with quantum affine algebra of type C_n", *SIGMA Symmetry Integrability Geom. Methods Appl.* **3** (2007), Paper 078, 20. MR 2008f:17030 Zbl 1142.17009

[Nakai and Nakanishi 2007b] W. Nakai and T. Nakanishi, "Paths and tableaux descriptions of Jacobi–Trudi determinant associated with quantum affine algebra of type D_n", *J. Algebraic Combin.* **26**:2 (2007), 253–290. MR 2008e:17013 Zbl 1171.17004

[Naoi 2013] K. Naoi, "Demazure modules and graded limits of minimal affinizations", *Represent. Theory* **17** (2013), 524–556. MR 3120578 Zbl 0629.0062

[Naoi 2014] K. Naoi, "Graded limits of minimal affinizations in type D", *SIGMA Symmetry Integrability Geom. Methods Appl.* **10** (2014), Paper 047, 20. MR 3210588

[Sam and Snowden 2013] S. V. Sam and A. Snowden, "Stability patterns in representation theory", preprint, 2013. arXiv 1302.5859v1

Received August 16, 2013. Revised October 15, 2013.

STEVEN V SAM
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, BERKELEY
933 EVANS HALL
BERKELEY, CA 94720
UNITED STATES
svs@math.berkeley.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 272, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.272.245

NORMAL FAMILIES OF HOLOMORPHIC MAPPINGS INTO COMPLEX PROJECTIVE SPACE CONCERNING SHARED HYPERPLANES

LIU YANG, CAIYUN FANG AND XUECHENG PANG

We prove new criteria for normality for holomorphic mappings into the complex projective space using the generalized Zalcman lemma. This improves previous results in one complex variable. An example is included to complement our theory.

1. Introduction

Recall that a family $\mathcal{F}$ of meromorphic functions on a plane domain $D \subset \mathbb{C}$ is *normal* on D if every sequence in $\mathcal{F}$ contains a subsequence that converges uniformly on D (with respect to the spherical metric) to a meromorphic function or to ∞.

The following Picard-type theorem is a consequence of the second main theorem of value distribution theory.

Theorem A [Bergweiler 2006, pp. 78–80] . *Let f be a meromorphic function on the complex plane $\mathbb{C}$. If there exist three mutually distinct points a_1, a_2 and a_3 on the Riemann sphere such that $f(z) - a_j$ (for $j = 1, 2, 3$) has no zero on the complex plane then $f(z)$ is a constant.*

A heuristic principle, bearing Bloch's name and playing an important role in the theory of normal families, says that if the only meromorphic function with a certain property are constant, then a family of meromorphic functions in a plane domain possessing this property is likely to be normal [Bergweiler 2006, pp. 78–80]. For example, the Montel-type theorem associated with Theorem A is true:

Theorem B [Bergweiler 2006, pp. 78–80]. *Let $\mathcal{F}$ be a family of meromorphic functions on a plane domain D. Suppose that there exist three mutually distinct points a_1, a_2 and a_3 on the Riemann sphere such that $f(z) - a_j$ (for $j = 1, 2, 3$) has no zero on D for each $f \in \mathcal{F}$. Then $\mathcal{F}$ is a normal family on D.*

We say that two meromorphic functions f and g on a domain D *share the value* a ($a = \infty$ is allowed) if $f^{-1}(a) = g^{-1}(a)$ as sets (ignoring multiplicities). There

Supported by National Natural Science Foundation of China (Grant No. 11261029 and 11371139).
MSC2010: primary 32A17, 32H25; secondary 30D35, 30D45.
Keywords: normal family, holomorphic mapping, complex projective space.

are many results concerning this notion in value distribution theory, such as R. Nevanlinna's famous theorem [1926] that two meromorphic functions on the complex plane sharing five distinct values coincide identically. (The number 5 cannot be reduced, as the pair e^z, e^{-z}, with shared values $0, 1, -1, \infty$, demonstrates; but Nevanlinna [1926] also showed that if four values are shared and the multiplicities with which these each of these values is taken are the same for the two functions, the two functions differ only by a Möbius transformation. The condition that the multiplicities are the same cannot be relaxed; see [Gundersen 1979].)

More generally, the maximum modulus principle and Montel's theorem yield this extension of Theorem B:

Theorem C. *Let $\mathcal{F}$ be a family of meromorphic functions on a plane domain D. Suppose that there exist three mutually distinct points a_1, a_2 and a_3 on the Riemann sphere such that for each $f, g \in \mathcal{F}$, f and g share a_j (for $j = 1, 2, 3$) on D. Then $\mathcal{F}$ is normal on D.*

The following question arises naturally from Theorem C. *Suppose two families of meromorphic functions share some values a_j. If one is normal, is the other normal?* Recently the problem was solved by Pang and Liu, who showed that if two families of meromorphic functions share four values, the normality of one family implies the normality of the other. They also gave a counterexample to show that the number 4 is sharp.

Theorem D [Liu et al. 2013]. *Let $\mathcal{F}$ and $\mathcal{G}$ be two families of meromorphic functions on a plane domain D. Suppose that there exist four mutually distinct points a_1, a_2, a_3 and a_4 on the Riemann sphere such that for each $f \in \mathcal{F}$, there exists $g \in \mathcal{G}$ such that f and g share a_j for $j = 1, \dots, 4$ on D. If $\mathcal{G}$ is normal on D, then $\mathcal{F}$ is also normal on D.*

The classical Zalcman lemma plays a central role in normal family theory of one complex variable. On the other hand, the study of normal families for holomorphic mappings was initiated by H. Wu in his well-known paper in Acta Math [1967]. Much attention has been given to find the correct generalization of Zalcman's result to several complex variables. In this paper we prove some new normality criteria for holomorphic mappings from plane domains into $\mathbb{P}^s(\mathbb{C})$ using the generalized Zalcman lemma. An example will be included to complement our theory.

2. Basic notions and main results

Basic notions. We start with relevant definitions. For details see [Mai et al. 2005; Shabat 1985, pp. 99–106; Ru 2001, pp. 99–102].

Let $\mathbb{P}^s(\mathbb{C})$ be a complex s-dimensional projective space and $\rho : \mathbb{C}^{s+1} \setminus \{0\} \to \mathbb{P}^s(\mathbb{C})$ be the standard projective mapping. A subset H of $\mathbb{P}^s(\mathbb{C})$ is called a hyperplane if there is a s-dimensional linear subspace $\widetilde{H}$ of $\mathbb{C}^{s+1}$ such that

$$\rho(\widetilde{H} - \{0\}) = H.$$

For a fixed system of homogeneous coordinates $Z = [Z_0 : Z_1 : \cdots : Z_s]$, a hyperplane H of $\mathbb{P}^s(\mathbb{C})$ can be written as

$$H = \big\{[Z_0 : Z_1 : \cdots : Z_s] \in \mathbb{P}^s(\mathbb{C}) \mid \langle Z, \alpha\rangle = 0\big\},$$

where

$$\langle Z, \alpha\rangle := a_0 Z_0 + \cdots + a_s Z_s$$

and $\alpha = (a_0, \dots, a_s) \in \mathbb{C}^{s+1}$ is a nonzero vector. We write it as

$$H = \{\langle Z, \alpha\rangle = 0\}$$

for convenience. In particular, we can take $\alpha \in B$, where B is the set of Euclidean unit vectors in $\mathbb{C}^{s+1}$.

Let $H_1, \dots, H_{s+1}$ be hyperplanes in $\mathbb{P}^s(\mathbb{C})$. Let $\alpha_j = (a_{j0}, \dots, a_{js}) \in B$ be such that

$$H_j = \{\langle Z, \alpha_j\rangle = 0\}$$

for $j = 1, \dots, s+1$. Define

$$D(H_1, \dots, H_{s+1}) := |\det(\alpha_1^T, \dots, \alpha_{s+1}^T)|$$

which only depends on H_j but does not depend on the choice of $\alpha_j \in B$.

Definition 2.1. Let $H_1, \dots, H_q$, with $q \geq s+1$, be hyperplanes in $\mathbb{P}^s(\mathbb{C})$. Define

$$D(H_1, \dots, H_q) := \prod_{1 \leq j_1 < \cdots < j_{s+1} \leq q} |\det(\alpha_{j_1}^T, \dots, \alpha_{j_{s+1}}^T)|.$$

We say the hyperplane family $H_1, \dots, H_q$, $q \geq s+1$, in $\mathbb{P}^s(\mathbb{C})$ is in general position if $D(H_1, \dots, H_q) > 0$.

Let M and N be connected Hermitian manifolds of dimension m and s with Hermitian metrics h_M and h_N, respectively. The space $\mathscr{C}(M; N)$ of continuous mappings between M and N endowed with the compact-open topology is second countable so that a metric can be furnished in $\mathscr{C}(M; N)$ which induces the compact-open topology.

Remark 2.2. A sequence $\{f_n\}$ in $\mathscr{C}(M; N)$ converges to f in $\mathscr{C}(M; N)$ in this topology if and only if $\{f_n\}$ converges to f uniformly on compact subset of M.

The space $\mathscr{H}(M; N)$ of holomorphic mappings from M into N is a closed subspace of $\mathscr{C}(M; N)$.

Definition 2.3. A family $\mathcal{F} \subset \mathcal{H}(M; N)$ is called normal on M if any sequence in $\mathcal{F}$ contains a subsequence which is relatively compact in $\mathcal{H}(M; N)$, that is, if any sequence $\{f_n\} \subset \mathcal{F}$ contains a subsequence which converges to $f \in \mathcal{H}(M; N)$ uniformly on every compact subset of M.

Throughout this paper, we consider the special case where M is a plane domain and N is a complex projective space.

Let $f : D \to \mathbb{P}^s(\mathbb{C})$ be a holomorphic map and U be an open set in D. Any holomorphic mapping $\tilde{f} : U \to \mathbb{C}^{s+1}$ such that $\rho \circ \tilde{f}(z) \equiv f(z)$ in U is called a representation of f on U. For a fixed system of homogeneous coordinates $[Z_0 : Z_1 : \cdots : Z_s]$ we set

$$V_i = \big\{[Z_0 : Z_1 : \cdots : Z_s] \mid Z_i \neq 0\big\}, \quad \text{for } i = 0, \dots, s.$$

Then every $a \in D$ has a neighborhood U of a such that $f(U) \subset V_i$ for some i, and f has a representation

$$\tilde{f} = (f_0, \dots, f_{i-1}, 1, f_{i+1}, \dots, f_s)$$

on U with holomorphic functions $f_0, \dots, f_{i-1}, f_{i+1}, \dots, f_s$.

Definition 2.4. For an open subset U of D we call a representation $\tilde{f} = (f_0, \dots, f_s)$ a reduced representation of f on U if $f_0, \dots, f_s$ are holomorphic functions on U and have no common zero.

Remark 2.5. Every holomorphic map of D into $\mathbb{P}^s(\mathbb{C})$ has a reduced representation on some neighborhood of each point in D. Moreover, let $\tilde{f} = (f_0, \dots, f_s)$ be a reduced representation of f. For an arbitrary nowhere zero holomorphic function h, $(f_0 h, \dots, f_s h)$ is also a reduced representation of f. Conversely, for every reduced representation $(g_0, \dots, g_s)$ of f, each g_i can be written as $g_i = h f_i$ for some nowhere zero holomorphic function h.

Remark 2.6. Every $f \in \mathcal{H}(D; \mathbb{P}^s(\mathbb{C}))$ has a reduced representation on the totality of D [Fujimoto 1974].

We now give the definition of sharing hyperplanes, which extends the definition of sharing values. Take $f \in \mathcal{H}(D; \mathbb{P}^s(\mathbb{C}))$. Let $H = \{\langle Z, \alpha \rangle = 0\}$ be a hyperplane in $\mathbb{P}^s(\mathbb{C})$, where $\alpha = (a_0, \dots, a_s) \in \mathbb{C}^{s+1} - \{0\}$. Let $\tilde{f} = (f_0, \dots, f_s)$ be a reduced representation of f. We consider the holomorphic function on D

$$\langle \tilde{f}(z), H \rangle := a_0 f_0 + \cdots + a_s f_s.$$

Definition 2.7. Suppose f, g are in $\mathcal{H}(D; \mathbb{P}^s(\mathbb{C}))$ and H is a hyperplane in $\mathbb{P}^s(\mathbb{C})$. If there exist some (thus all) reduced representations of f and g respectively such that $\langle \tilde{f}(z), H \rangle$ and $\langle \tilde{g}(z), H \rangle$ share 0 on D, we say that f and g share H on D.

By Remark 2.5, $\langle \tilde{f}(z), H\rangle = 0$ is indeed independent of the choice of the reduced representation of f. Therefore sharing hyperplanes is well defined.

We will use the notation $\langle f(z), H\rangle$ when some properties are independent of the choice of the reduced representation of f. For example, we can say that $\langle f(z), H\rangle$ has finite zeros on D.

H. Fujimoto [1974] gave the relation between m-convergence and quasiregularity. In the case of holomorphic maps, we have the following properties. Suppose $\{f_n\} \subset \mathcal{H}(D; \mathbb{P}^s(\mathbb{C}))$; then $\{f_n\}$ converges uniformly on compact subsets of D to a holomorphic mapping f of D into $\mathbb{P}^s(\mathbb{C})$ if and only if, for any $a \in D$, each f_n has a reduced representation

$$\tilde{f}_n = (f_{n0}, f_{n1}, \dots, f_{ns})$$

on some fixed neighborhood U of a in D such that $\{f_{ni}\}$ converges uniformly on compact subsets of U to a holomorphic function f_i on U, $i = 0, 1, \dots, s$, with the property that

$$\tilde{f} = (f_0, f_1, \dots, f_s)$$

is a reduced representation of f on U.

Main results. Here we shall improve both Theorem C and Theorem D and obtain the following results.

Theorem 2.8. *Suppose $\mathcal{F} \subset \mathcal{H}(D; \mathbb{P}^s(\mathbb{C}))$. Let $H_1, \dots, H_q$, with $q \geq 2s+1$, be hyperplanes in $\mathbb{P}^s(\mathbb{C})$ located in general position. Suppose that for each $f, g \in \mathcal{F}$, f and g share H_j on D, for $j = 1, \dots, q$. Then $\mathcal{F}$ is normal on D.*

Corollary 2.9. *Suppose $\mathcal{F} \subset \mathcal{H}(D; \mathbb{P}^s(\mathbb{C}))$. Let $H_1, \dots, H_q$, with $q \geq 2s+1$, be hyperplanes in $\mathbb{P}^s(\mathbb{C})$ located in general position. Suppose that for each $f \in \mathcal{F}$, f omits H_j on D, for $j = 1, \dots, q$. Then $\mathcal{F}$ is normal on D.*

Proof. Each H_j $(j = 1, \dots, q)$ is a shared value of all $f \in \mathcal{F}$, since $f^{-1}(H_j) = \varnothing$. Thus, the family $\mathcal{F}$ satisfies the assumptions of Theorem 2.8. □

Theorem 2.10. *Suppose $\mathcal{F}, \mathcal{G} \subset \mathcal{H}(D; \mathbb{P}^s(\mathbb{C}))$. Let $q \geq 3s+1$ be a integer, and suppose the following three conditions are satisfied:*

(i) *For each $f \in \mathcal{F}$, there exist $g \in \mathcal{G}$ and q hyperplanes $H_{1,f}, \dots, H_{q,f}$ (which may depend on f) such that f and g share $H_{j,f}$ on D, for $j = 1, \dots, q$.*

(ii) $\inf\{D(H_{1,f}, \dots, H_{q,f}) : f \in \mathcal{F}\} > 0$.

(iii) *$\mathcal{G}$ is normal on D.*

Then $\mathcal{F}$ is a normal family on D.

By Theorem 2.10 we immediately have the following corollary.

Corollary 2.11. *Suppose $\mathcal{F}, \mathcal{G} \subset \mathcal{H}(D; \mathbb{P}^s(\mathbb{C}))$. Let $H_1, \dots, H_q$, with $q \geq 3s+1$, be hyperplanes in $\mathbb{P}^s(\mathbb{C})$ located in general position. Suppose that for each $f \in \mathcal{F}$ there exists $g \in \mathcal{G}$ such that f and g share H_j on D, $j = 1, \dots, q$. If $\mathcal{G}$ is normal on D, then $\mathcal{F}$ is also normal on D.*

The following example shows that the number $3s+1$ in Theorem 2.10 is sharp when $s = 2$.

Example 1. Let Δ be the unit disk. Let $\mathcal{F} = \{f_n(z)\}$, where

$$f_n(z) = \left[\sqrt{-1}\cos nz : \sin nz : \sin nz\right].$$

We denote by $z_{n,1}, z_{n,2}, \dots, z_{n,k_n}$ the zeros of $\sin nz$ in Δ. Let $\mathcal{G} = \{g_n(z)\}$, where

$$g_n(z) = \left[1 : \prod_{1 \leq i \leq k_n} \frac{z - z_{n,i}}{1 - \bar{z}_{n,i} z} : \prod_{1 \leq i \leq k_n} \frac{z - z_{n,i}}{1 - \bar{z}_{n,i} z}\right].$$

Let

$$\begin{aligned}
H_1 &= \{[Z_0 : Z_1 : Z_2] \mid 3Z_0 + Z_1 + 2Z_2 = 0\},\\
H_2 &= \{[Z_0 : Z_1 : Z_2] \mid -5Z_0 + Z_1 + 4Z_2 = 0\},\\
H_3 &= \{[Z_0 : Z_1 : Z_2] \mid 7Z_0 + Z_1 + 6Z_2 = 0\},\\
H_4 &= \{[Z_0 : Z_1 : Z_2] \mid -9Z_0 + Z_1 + 8Z_2 = 0\},\\
H_5 &= \{[Z_0 : Z_1 : Z_2] \mid Z_2 = 0\},\\
H_6 &= \{[Z_0 : Z_1 : Z_2] \mid Z_1 = 0\}.
\end{aligned}$$

Then these hyperplanes are in general position.

One can verify that f_n and g_n share H_j on Δ for $j = 1, \dots, 6$. Clearly, $\mathcal{G}$ is normal on Δ. However, $\mathcal{F}$ fails to be normal on any neighborhood of 0 by Lemma 3.2 in next section.

3. Some lemmas

The following is the general version of the Zalcman lemma.

Lemma 3.1 [Thai et al. 2003]. *Let $\mathcal{F}$ be a family of holomorphic mappings of a domain Ω in $\mathbb{C}^m$ into $\mathbb{P}^s(\mathbb{C})$. The family $\mathcal{F}$ is not normal on Ω if and only if there exist sequences $\{f_n\} \subset \mathcal{F}$, $\{z_n\} \subset \Omega$ with $z_n \to z_0 \in \Omega$, and $\{\rho_n\}$ with $\rho_n > 0$ and $\rho_n \to 0$ such that*

$$h_n(\xi) := f_n(z_n + \rho_n \xi)$$

converges uniformly on compact subsets of $\mathbb{C}$ to a nonconstant holomorphic mapping h of $\mathbb{C}$ into $\mathbb{P}^s(\mathbb{C})$.

Lemma 3.2 [Osserman and Ru 1997]. *Let M be a Riemann surface, and $f_n : M \to \mathbb{P}^s(\mathbb{C})$ be a sequence of holomorphic maps converging uniformly on every compact subset of M to a holomorphic map $f : M \to \mathbb{P}^s(\mathbb{C})$. Given $a, b \in \mathbb{P}^s(\mathbb{C}^*)$, let $f_{a,b}$ be the meromorphic function defined by*

$$f_{a,b} = \frac{\langle \tilde{f}, \alpha \rangle}{\langle \tilde{f}, \beta \rangle},$$

where $\tilde{f}$ is a reduced representation of f on U, and $\alpha, \beta \in (\mathbb{C}^{s+1})^*$ are such that $a = \rho(\alpha)$, $b = \rho(\beta)$. Assume that $\beta(\tilde{f}) \not\equiv 0$ on some U. Let $p \in M$ be such that $\beta(\tilde{f})(p) \neq 0$, and U_p be a neighborhood of p such that $\beta(\tilde{f})(z) \neq 0$ for $z \in U_p$. Then $\{f_{n_{a,b}}\}$ converges uniformly on U_p to the meromorphic function $f_{a,b}$.

Let $\mu > 0$ be an integer. The holomorphic map $f \in \mathcal{H}(\mathbb{C}; \mathbb{P}^s(\mathbb{C}))$ is said to be ramified over a hyperplane $H = \{\langle Z, \alpha \rangle = 0\}$ with multiplicity at least μ if all zeros of $\langle f(z), \alpha \rangle = 0$ have orders at least μ, where $\tilde{f}$ is a local reduced representation of f (it is easy to check that this definition is independent of the choice of reduced representation). If either the image of f completely omits H or $f(\mathbb{C}) \subseteq H$, we shall say that f is ramified over H with multiplicity ∞.

Nochka [1983] improved the result of Green [1977] and proved H. Cartan's conjecture.

Lemma 3.3 [Nochka 1983]. *Suppose that $q(\geq 2s+1)$ hyperplanes $H_1, \dots, H_q$ are given in general position in $\mathbb{P}^s(\mathbb{C})$, along with q positive integers $m_1, \dots, m_q$ (some of them may be ∞). If*

$$\sum_{j=1}^{q} \Bigl(1 - \frac{s}{m_j}\Bigr) > s + 1,$$

then there does not exist a nonconstant holomorphic mapping $f : \mathbb{C} \to \mathbb{P}^s(\mathbb{C})$ such that f intersects H_j with multiplicity at least m_j, $j = 1, \dots, q$.

Lemma 3.4 (first main theorem [Fujimoto 1993, Corollary 3.1.16]). *Let $f : \mathbb{C} \to \mathbb{P}^s(\mathbb{C})$ be a holomorphic map. Let H be a hyperplane in $\mathbb{P}^s(\mathbb{C})$. If $f(\mathbb{C}) \nsubseteq H$, then*

$$T_f(r) = m_f(r, H) + N_f(r, H) + O(1).$$

The second main theorem about linearly degenerated case is also required.

Lemma 3.5 (degenerate second main theorem [Ru 2001, Theorem A3.4.4]). *Let $f = [f_0 : \cdots : f_s] : \mathbb{C} \to \mathbb{P}^s(\mathbb{C})$ be a holomorphic map whose image is contained in some k-dimensional subspace but not in any subspace of dimension lower than k. Let $H_1, \dots, H_q$ be hyperplanes in general position. Assume that $f(\mathbb{C}) \nsubseteq H_j$, for*

$j = 1, \dots, q$. Then the inequality

$$\sum_{j=1}^{q} m_f(r, H_j) + \frac{n+1}{k+1} N(R_f, r) \le (2n-k+1)T_f(r) + O(\log T_f(r))$$

holds for all r outside a set E with finite Lebesgue measure. Here $N(R_f, r)$ is the ramification term.

Lemma 3.6 [Fujimoto 1974]. *Let $f \in \mathcal{H}(\mathbb{C}; \mathbb{P}^s(\mathbb{C}))$. The map f is rational, namely, f is representable as $f = [f_0 : \cdots : f_s]$ with polynomial f_i, $i = 0, \dots, s$, if and only if*

$$\lim_{r\to\infty} \frac{T_f(r)}{\log r} < \infty.$$

Lemma 3.7. *Let $f \in \mathcal{H}(\mathbb{C}; \mathbb{P}^s(\mathbb{C}))$, and $H_1, \dots, H_{2s+1}$ be hyperplanes in $\mathbb{P}^s(\mathbb{C})$ located in general position. If for each hyperplane H_j, $j = 1, \dots, 2s+1$, either $f(\mathbb{C}) \subset H_j$ or $\langle f(z), H_j \rangle$ has finite zeros in $\mathbb{C}$ (no zero point is allowed), then the map f is rational.*

Proof. Let $\tilde{f} = (f_0, \dots, f_s)$ be a reduced representation of f on $\mathbb{C}$. We set the rank of the vector group $\{f_0, \dots, f_s\}$ to be $k+1$, with $0 \le k \le s$. Thus, $f(\mathbb{C})$ is contained in some k-dimensional subspace of $\mathbb{P}^s(\mathbb{C})$ but not in any subspace of dimension lower than k.

Let I be a subset of $\{1, \dots, 2s+1\}$ such that i is in I if and only if $f(\mathbb{C}) \subset H_i$, and let

$$X_I = \bigcap_{i\in I} H_i.$$

We can identify X_I with a projective space of dimension $s - k_1$, where $k_1 = \#I$. So $0 \le k_1 \le s - k$. According to the definition, the restrictions of

$$H_j^* := H_j \bigcap X_I, \quad j \notin I$$

are hyperplanes which are still in general position in $X_I = \mathbb{P}^{s-k_1}(\mathbb{C})$.

Applying Lemma 3.5 to $f = [f_0 : \cdots : f_s] : \mathbb{C} \to \mathbb{P}^{s-k_1}(\mathbb{C})$ and the hyperplanes H_j^*, $j \notin I$, and using the first main theorem about holomorphic curves, it follows that the inequality

$$(2s - k_1 + 1)T_f(r) \le \sum_{j\notin I} N_f(r, H_j^*) + (2(s-k_1) - k + 1)T_f(r) + O(\log T_f(r))$$

holds for all r outside a set with finite Lebesgue measure. Since $\langle f(z), H_j^* \rangle$ has finite zeros in $\mathbb{C}$, this yields the inequality

$$(k_1 + k)T_f(r) \le O(\log T_f(r)) + O(\log r).$$

If $k = k_1 = 0$, the rank of the vector group $\{f_0, \dots, f_s\}$ is 1, which means that f is a constant map.

If $k_1 + k > 0$. Together with Lemma 3.6, the above inequality implies that f is rational. Hence, the lemma is proved. □

4. Proofs of the theorems

Proof of Theorem 2.8. Fix $g \in \mathcal{F}$. Suppose that $\mathcal{F}$ is not normal on some point $z_0 \in D$. Suppose there are k hyperplanes H_{j_l}, $l = 1, \dots, k$, such that

$$g(z_0) \in \bigcap_{l=1}^{k} H_{j_l}.$$

Then $k \leq s$. For otherwise $k \geq s+1$, and because $H_1, \dots, H_q$, $q \geq 2s+1$, are hyperplanes in $\mathbb{P}^s(\mathbb{C})$ located in general position, it follows that $g = [0 : 0 : \cdots : 0]$. This is a contradiction. Therefore, $k \leq s$. Without loss of generality, we assume that there exists a neighborhood $U(z_0) \subset D$ such that for $l = 1, \dots, k_1$,

$$g(U(z_0)) \subset H_l,$$

for $\mu = k_1 + 1, \dots, k$,

$$g(U(z_0)) \cap H_\mu = \{g(z_0)\},$$

and for $\nu = k+1, \dots, 2s+1$,

$$g(D(z_0)) \cap H_\nu = \phi.$$

In other words, these hyperplanes are divided into three groups.

Observing that normality is a local property, we may suppose that $U(z_0)$ is the unit disk Δ, and $z_0 = 0$. Then by Lemma 3.1 there exist points z_n with $z_n \to z_0 \in D$, positive numbers ρ_n with $\rho_n \to 0$, and functions $f_n \in \mathcal{F}$ such that

$$h_n(\xi) := f_n(z_n + \rho_n \xi)$$

converges uniformly on compact subsets of $\mathbb{C}$ to a nonconstant holomorphic mapping h of $\mathbb{C}$ into $\mathbb{P}^s(\mathbb{C})$. Here $\xi \in \mathbb{C}$ satisfies $z_n + \rho_n \xi \in \Delta$.

We consider two cases.

If $z_n/\rho_n \to \infty$, then for each $\xi \in \mathbb{C}$, $z_n + \rho_n \xi \neq z_0$ when n is large enough. It follows that for $i = k_1 + 1, \dots, 2s+1$,

$$\langle f_n(z_n + \rho_n \xi), H_i \rangle \neq 0.$$

The Hurwitz theorem implies that for $i = k_1 + 1, \dots, 2s+1$, $\langle h(\xi), H_i \rangle \neq 0$ or $\langle h(\xi), H_i \rangle \equiv 0$. Thus, $\langle h(\xi), H_j \rangle \neq 0$ or $\langle h(\xi), H_j \rangle \equiv 0$ for $j = 1, \dots, 2s+1$. By Lemma 3.3, h is a constant holomorphic mapping. This contradicts the claim that

h is a nonconstant holomorphic mapping.

If $z_n/\rho_n \not\to \infty$, taking a subsequence and renumbering, we may assume that $z_n/\rho_n \to c$, $c \in \mathbb{C}$. Then

$$f_n(\rho_n \xi) = h_n\Big(\xi - \frac{z_n}{\rho_n}\Big)$$

converges uniformly on compact subsets of $\mathbb{C}$ to a nonconstant holomorphic mapping $h(\xi - c)$. Since for each hyperplane H_j, $j = 1, \dots, 2s+1$, either $h(\mathbb{C}) \subset H_j$ or $\langle h(\xi - c)(z), H_j\rangle$ has finite zeros in $\mathbb{C}$, $h(\xi - c)$ is rational by Lemma 3.7. Since $h(\xi - c)$ is a holomorphic mapping, there exist some constants c_ν, with $\nu = k+1, \dots, 2s+1$, such that

$$\langle h(\xi - c)(z), H_\nu\rangle \equiv c_\nu.$$

Note that $2s - k + 1 \geq s + 1$, and $\{H_j\}$ are in general position. Hence we see that $h(\xi - c)$ is a constant map. Again, this a contradiction. And hence the family $\mathscr{F}$ is normal on D. □

Proof of Theorem 2.10. If $\mathscr{F}$ is not normal on D, then by Lemma 3.1, there exist points $z_n \to z_0 \in D$, positive numbers $\rho_n \to 0$ and functions $f_n \in \mathscr{F}$, such that

$$h_n(\xi) := f_n(z_n + \rho_n \xi),$$

where $\xi \in \mathbb{C}$ satisfies $z_n + \rho_n \xi \in D$, converges uniformly on compact subsets of $\mathbb{C}$ to a nonconstant holomorphic mapping h of $\mathbb{C}$ into $\mathbb{P}^s(\mathbb{C})$.

By condition (i), there exist q hyperplane sequences $\{H_{j,f_n}\}_{n=1}^{\infty}$ and $\{g_n\} \subset \mathscr{G}$ such that for $z \in D$, $j = 1, \dots, q$,

$$\langle g_n(z), H_{j,f_n}\rangle = 0$$

whenever

$$\langle f_n(z), H_{j,f_n}\rangle = 0, \quad z \in D.$$

For $j = 1, \dots, q$, take $\{\alpha_{jn}\}_{n=1}^{\infty} \subset B$ satisfying

$$H_{j,f_n} = \{\langle Z, \alpha_{jn}\rangle = 0\}.$$

Since B is a compact subset of $\mathbb{C}^{s+1}$, there exist $\alpha_j = (a_{j0}, \dots, a_{js}) \in B$ for $j = 1, \dots, q$, and subsequences which (to avoid complication in notation) we again call $\{\alpha_{jn}\}$ satisfying that $\alpha_{jn} \to \alpha_j$ as $n \to \infty$. Let

$$H_j = \{\langle Z, \alpha_j\rangle = 0\}$$

be hyperplanes of $\mathbb{P}^s(\mathbb{C})$, $j = 1, \dots, q$. From condition (i), it follows that

$$D(H_1, \dots, H_q) \geq \liminf_{n\to\infty} D(H_{1,f_n}, \dots, H_{q,f_n}) > 0.$$

Thus, the hyperplanes H_j, $j = 1, \dots, q$, are located in general position.

Claim. There exist at most $2s$ *hyperplanes such that for each hyperplane* H*, either the image of* h *completely omits* H *or* $h(\mathbb{C}) \subset H$. If not, Lemma 3.3 shows that h is a constant holomorphic mapping, which is a contradiction. So there exist at least $s+1$ hyperplanes of H_j, $j = 1, \ldots, q$, such that for $i = 1, \ldots, s+1$, $h(\mathbb{C}) \cap H_i \neq \phi$ and $h(\mathbb{C}) \not\subseteq H_i$.

For a fixed $i \in \{1, \ldots, s+1\}$, suppose that $\xi_i \in h(\mathbb{C}) \cap H_i$. Choose a small neighborhood $U(\xi_i)$ of ξ_i such that $h(\mathbb{C}) \cap H_i = \{\xi_i\}$. Hence $\langle \tilde{h}(\xi_i), H \rangle = 0$ and $\langle \tilde{h}(\xi_i), H \rangle \not\equiv 0$, where $\tilde{h}$ is a local reduced representation. Since h_n converges uniformly to h on $U(\xi_i)$, h_n has a local reduced representation $\tilde{h}_n = (h_{n0}, \ldots, h_{ns})$ such that $\tilde{h}_n$ uniformly converges to a reduced representation $\tilde{h} = (h_0, \ldots, h_s)$ of h on $U(\xi_i)$. Obviously, h_{nk} converges uniformly to h_k on $U(\xi_i)$ for each $k = 0, \ldots, s$. Therefore $\langle \tilde{h}_n(\xi), \alpha_{in} \rangle$ converges uniformly to $\langle \tilde{h}(\xi), \alpha_i \rangle$ on $U(\xi_i)$. By the Hurwitz theorem, there exist $\xi_{in} \to \xi_i$ such that $\langle \tilde{h}_n(\xi_{in}), \alpha_{in} \rangle = 0$, that is, $\langle \tilde{f}_n(z_n + \rho_n \xi_{in}), \alpha_{in} \rangle = 0$.

On the other hand, applying condition (iii), we can find subsequences of $\{g_n\}$ (again denoted by themselves) such that g_n converges uniformly to g on D, where g is a holomorphic mapping of D into $\mathbb{P}^s(\mathbb{C})$. As we noted earlier, g_n has a local reduced representation $\tilde{g}_n = (g_{n0}, \ldots, g_{ns})$ such that $\tilde{g}_n$ uniformly converges to a reduced representation $\tilde{g} = (g_0, \ldots, g_s)$ of g on $U(z_0)$. It follows that

$$\langle \tilde{g}_n(z_n + \rho_n \xi_{in}), \alpha_{in} \rangle = 0.$$

As $n \to \infty$, we have

$$\langle \tilde{g}(z_0), \alpha_i \rangle = 0.$$

So there exist $s+1$ hyperplanes H_i, $i = 1, \ldots, s+1$, which intersect at one point $\rho(\tilde{g}(z_0))$. This contradicts the claim that the hyperplanes H_j, $j = 1, \ldots, q$, are located in general position. This finishes the proof. $\square$

Acknowledgements

The authors thank the referees for valuable suggestions.

References

[Bergweiler 2006] W. Bergweiler, "Bloch's principle", *Comput. Methods Funct. Theory* **6**:1 (2006), 77–108. MR 2007f:30045 Zbl 1101.30034

[Fujimoto 1974] H. Fujimoto, "On families of meromorphic maps into the complex projective space", *Nagoya Math. J.* **54** (1974), 21–51. MR 51 #3543 Zbl 0267.32005

[Fujimoto 1993] H. Fujimoto, *Value distribution theory of the Gauss map of minimal surfaces in* $\mathbb{R}^m$, Aspects of Mathematics **E21**, Vieweg, Braunschweig, 1993. MR 95d:32029 Zbl 1107.32004

[Green 1977] M. L. Green, "The hyperbolicity of the complement of $2n+1$ hyperplanes in general position in $\mathbb{P}_n$ and related results", *Proc. Amer. Math. Soc.* **66**:1 (1977), 109–113. MR 56 #15994 Zbl 0366.32013

[Gundersen 1979] G. G. Gundersen, "Meromorphic functions that share three or four values", *J. London Math. Soc.* (2) **20**:3 (1979), 457–466. MR 80m:30030 Zbl 0413.30025

[Liu et al. 2013] X. J. Liu, S. H. Li, and X. Pang, "A normal criterion about two families of meromorphic functions concerning shared values", *Acta Math. Sin.* (*Engl. Ser.*) **29**:1 (2013), 151–158. MR 3001017 Zbl 1267.30086

[Mai et al. 2005] P. N. Mai, D. D. Thai, and P. N. T. Trang, "Normal families of meromorphic mappings of several complex variables into $\mathbb{P}^N(\mathbb{C})$", *Nagoya Math. J.* **180** (2005), 91–110. MR 2006i:32003 Zbl 1162.32001

[Nevanlinna 1926] R. H. Nevanlinna, "Einige Eindueutigkeitssätze in der Theorie der meromorphen Funktionen", *Acta Math.* **48** (1926), 367–391.

[Nochka 1983] E. I. Nochka, "On the theory of meromorphic functions", *Dokl. Akad. Nauk SSSR* **269**:3 (1983), 547–552. In Russian; translated in *Sov. Math. Dokl.* **27** (1983), 377–381. MR 85i:32038 Zbl 0552.32024

[Osserman and Ru 1997] R. Osserman and M. Ru, "An estimate for the Gauss curvature of minimal surfaces in $\mathbb{R}^m$ whose Gauss map omits a set of hyperplanes", *J. Differential Geom.* **46**:3 (1997), 578–593. MR 98k:53014 Zbl 0918.53003

[Ru 2001] M. Ru, *Nevanlinna theory and its relation to Diophantine approximation*, World Scientific, River Edge, NJ, 2001. MR 2002g:11106 Zbl 0998.30030

[Shabat 1985] B. V. Shabat, *Distribution of values of holomorphic mappings*, Translations of Mathematical Monographs **61**, American Mathematical Society, Providence, RI, 1985. MR 86k:32023 Zbl 0564.32016

[Thai et al. 2003] D. D. Thai, P. N. T. Trang, and P. D. Huong, "Families of normal maps in several complex variables and hyperbolicity of complex spaces", *Complex Var. Theory Appl.* **48**:6 (2003), 469–482. MR 2004c:32003 Zbl 1036.32001

[Wu 1967] H. Wu, "Normal families of holomorphic mappings", *Acta Math.* **119** (1967), 193–233. MR 37 #468 Zbl 0158.33301

Received August 3, 2013. Revised September 22, 2013.

LIU YANG
DEPARTMENT OF MATHEMATICS
EAST CHINA NORMAL UNIVERSITY
SHANGHAI, 200062
CHINA
yangliu20062006@126.com

CAIYUN FANG
SCHOOL OF MATHEMATICS SCIENCES
NANJING NORMAL UNIVERSITY
NANJING, 210023
CHINA
05325@njnu.edu.cn

XUECHENG PANG
DEPARTMENT OF MATHEMATICS
EAST CHINA NORMAL UNIVERSITY
SHANGHAI, 200062
CHINA
xcpang@math.ecnu.edu.cn

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 272 No. 1 November 2014

0030-8730(201411)272:1;1-5