Pacific Journal of Mathematics

Volume 275 No. 2 June 2015

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

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

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 275, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.275.257

A COMBINATORIAL CHARACTERIZATION OF TIGHT FUSION FRAMES

MARCIN BOWNIK, KURT LUOTO AND EDWARD RICHMOND

In this paper we give a combinatorial characterization of tight fusion frame (TFF) sequences using Littlewood–Richardson skew tableaux. The equal rank case has been solved recently by Casazza, Fickus, Mixon, Wang, and Zhou. Our characterization does not have this limitation. We also develop some methods for generating TFF sequences. The basic technique is a majorization principle for TFF sequences combined with spatial and Naimark dualities. We use these methods and our characterization to give necessary and sufficient conditions which are satisfied by the first three highest ranks. We also give a combinatorial interpretation of spatial and Naimark dualities in terms of Littlewood–Richardson coefficients. We exhibit four classes of TFF sequences which have unique maximal elements with respect to majorization partial order. Finally, we give several examples illustrating our techniques including an example of tight fusion frame which can not be constructed by the existing spectral tetris techniques. We end the paper by giving a complete list of maximal TFF sequences in dimensions ≤ 9.

1. Introduction

Fusion frames were introduced in [Casazza and Kutyniok 2004] (under the name *frames of subspaces*) and in [Casazza et al. 2008]. A fusion frame for $\mathbb{R}^N$ is a finite collection of subspaces $\{W_i\}_{i=1}^K$ in $\mathbb{R}^N$ such that there exist constants $0 < \alpha \le \alpha' < \infty$ satisfying

$$\alpha \|x\|^2 \le \sum_{i=1}^{K} \|P_i x\|^2 \le \alpha' \|x\|^2 \qquad \text{for all } x \in \mathbb{R}^N,$$

where P_i is the orthogonal projection onto W_i. Equivalently, $\{W_i\}_{i=1}^K$ is a fusion frame if and only if

$$\alpha \mathbf{I} \le \sum_{i=1}^{K} P_i \le \alpha' \mathbf{I},$$

MSC2010: primary 05E05, 15A57, 42C15; secondary 14N15, 14M15.
Keywords: tight fusion frame, majorization, orthogonal projection, partition, Schur function, Littlewood–Richardson coefficient, Schubert calculus, symmetric functions.

where $\mathbf{I}$ is the identity on $\mathbb{R}^N$. The constants α and α' are called fusion frame bounds. An important class of fusion frames are *tight fusion frames* (TFF), for which $\alpha = \alpha'$ and hence $\sum_{i=1}^K P_i = \alpha\mathbf{I}$. We note that the definition of fusion frames given in [Casazza and Kutyniok 2004; Casazza et al. 2008] applies to closed subspaces in any Hilbert space together with a collection of weights associated to each subspace W_i. Since the scope of this paper is limited to nonweighted finite dimensional TFF, the definition of a fusion frame is only presented for this case.

Fusion frames have been a very active area of research in the frame theory [Casazza and Kutyniok 2013]. A lot of effort was devoted into developing the basic properties and constructing fusion frames with desired properties. In particular, the construction and existence of sparse tight fusion frames was studied in [Calderbank et al. 2011]. Fusion frame potentials have been studied in [Casazza and Fickus 2009] and [Massey et al. 2010]. Applications of fusion frames include sensor networks [Casazza et al. 2008], coding theory [Bodmann 2007; Kutyniok et al. 2009], compressed sensing [Boufounos et al. 2011], and filter banks [Chebira et al. 2011]. In this paper we consider a problem of classifying TFF sequences.

Problem 1.1. Given $N \in \mathbb{N}$, characterize sequences $(L_1, \dots, L_K)$ for which there exists a tight fusion frame $\{W_i\}_{i=1}^K$ with $\dim W_i = L_i$ in N dimensional space. Equivalently, given $\alpha > 1$ such that $\alpha N \in \mathbb{N}$, characterize sequences $(L_1, \dots, L_K)$ such that $\alpha\mathbf{I}$ can be decomposed as a sum of projections $P_1 + \cdots + P_K$ with $\operatorname{rank} P_i = L_i$, $i = 1, \dots, K$.

Casazza, Fickus, Mixon, Wang, and Zhou [Casazza et al. 2011] have recently achieved significant progress in this direction by solving the equal rank case. That is, the authors have classified all triples (K, L, N) such that there exists a tight fusion frame consisting of K subspaces $\{W_i\}_{i=1}^K$ with the same dimension $\dim W_i = L$ in $\mathbb{R}^N$. The answer is highly nontrivial in the most interesting case when L does not divide N and $2L < N$. The authors show that a necessary condition for such sequences (K, L, N) is that $K \geq \lceil N/L \rceil + 1$, whereas a sufficient condition is $K \geq \lceil N/L \rceil + 2$. In a gray area, where $K = \lceil N/L \rceil + 1$, the authors have devised a reduction procedure which replaces the original sequence by another one with the equivalent TFF property (existence or nonexistence). Then, it is shown that after a finite number of steps the original sequence (K, L, N) is reduced to one for which either the necessary condition fails or the sufficient condition holds. However, the results [Casazza et al. 2011] do not say much about a more general problem of classifying TFF sequences with unequal ranks. In this paper we answer Problem 1.1 by giving a combinatorial characterization of TFF sequences using Littlewood–Richardson skew tableaux.

While the concept of fusion frames is relatively new, the problem of representing an operator as a sum of orthogonal projections has been studied for a long time in the operator theory. The first fundamental result of this kind belongs to Fillmore [1969] who characterized finite rank operators which are finite sums of projections; see Theorem 3.1. Fong and Murphy [1985] characterized operators which are positive combinations of projections. Analogous results were recently investigated for C-$*$ algebras and von Neumann algebras; see [Halpern et al. 2013; Kaftal et al. 2011]. However, the most relevant results for us are due to Kruglyak, Rabanovich, and Samoĭlenko [Kruglyak et al. 2002; 2003] who characterized the set of all (α, N) such that $\alpha\mathbf{I}$ is the sum of K orthogonal projections. In other words, their main result (Theorem 7 of the latter reference) gives a minimal length K of a TFF sequence in $\mathbb{R}^N$ with the frame bound α. However, their results do not say anything about the ranks of projections which is a focus of this paper.

In the finite dimensional setting the existence of TFF sequences is intimately related to Horn's problem [Horn 1962] which has been solved by Klyachko [1998], and Knutson and Tao [Knutson and Tao 1999; Knutson et al. 2004], for a survey see [Fulton 2000; Knutson and Tao 2001]. Problem 1.1 can be thought of as a very special kind of Horn's problem where hermitian matrices have only two eigenvalues: 0 and 1, and their sum has only one eigenvalue α. Using Klyachko's result [1998] we show that the existence of TFF sequence $(L_1, \dots, L_K)$ is equivalent to the nonvanishing of a certain Littlewood–Richardson coefficient; see Theorem 4.3. In turn, the latter condition is equivalent to the existence of a matrix satisfying some computationally explicit properties such as: constant row and column sums, and row and column sum dominance; see Corollary 4.4. Our combinatorial characterization enables us to deduce several properties that TFF sequences must satisfy. In addition, it enables us to give an explicit construction procedure of a tight fusion frame corresponding to a given TFF sequence; see Example 7.2.

A fundamental technique of our paper is a majorization principle involving the majorization partial order $\preccurlyeq$ as in the Schur–Horn theorem [Antezana et al. 2007; Kaftal and Weiss 2010], which is also known as the dominance order in algebraic combinatorics [Fulton 1997]. In Section 2 we show that a sequence majorized by a TFF sequence is also a TFF sequence. We also establish the spatial and Naimark dualities for general TFF sequences extending the equal rank results in [Casazza et al. 2011]. In Section 3 we find necessary and sufficient conditions on the first three largest ranks of projections using Filmore's theorem [1969] and a description of possible spectra of a sum of two projections; see Lemma 3.2. The latter result might be of independent interest since its proof uses honeycomb models developed by Knutson and Tao [1999; 2001]. In the same section we also exhibit classes of TFF sequences which have only one maximal element. These include not only the expected case of integer α, but also half-integer scenario, and the corresponding conjugate α's via the Naimark duality. In Section 4 we prove our main characterization result of TFF sequences using Littlewood–Richardson skew tableaux. In addition to illustrating it on specific examples, in Section 5 we give a complete proof of Theorem 3.3 using the combinatorics of the Schur functions. This leads to a partial characterization of TFF sequences which are of the hook type, i.e., sequences ending in repeated 1's. In Section 6 we show that the spatial and Naimark dualities manifest themselves as identities for the corresponding Littlewood–Richardson coefficients. Finally, in Section 7 we give several examples of existence of tight fusion frames using skew Littlewood–Richardson tableaux. In particular, we give an explicit construction of TFF corresponding to the sequence (4, 2, 2, 2, 1) in dimension $N = 6$. This example is remarkable for two reasons. It is the first TFF sequence which is missed by brute force generation involving recursive spatial and Naimark dualities. Furthermore, this example can not be constructed by the existing spectral tetris construction [Calderbank et al. 2011; Casazza et al. 2012], which is an algorithmic method of constructing sparse fusion frames utilized in the equal rank characterization [Casazza et al. 2011]. We end the paper by giving a complete list of maximal TFF sequences for $\alpha \le 2$ in dimensions $N \le 9$.

2. Basic majorization and duality results

Definition 2.1. Fix a positive integer N. Let $L_1 \ge L_2 \ge \cdots \ge L_K > 0$ be a weakly decreasing sequence of positive integers. Such sequence is also known as a *partition* in number theory [Andrews 1976] and algebraic combinatorics [Fulton 1997]. We say that $(L_1, L_2, \dots, L_K)$ is a tight fusion frame (TFF) sequence if there exists orthogonal projections $P_1, \dots, P_K$ such that

$$\alpha \mathbf{I} = \sum_{i=1}^{K} P_i, \quad \text{and} \quad \text{rank}_i = L_i, \tag{2-1}$$

where $\alpha \in \mathbb{R}$ and $\mathbf{I}$ is the identity on $\mathbb{R}^N$. A trace argument shows that $\alpha = \sum_{i=1}^{K} L_i/N \ge 1$. Given $\alpha \ge 1$ such that $\alpha N \in \mathbb{N}$, we define $\text{TFF}(\alpha, N)$ to be the set of all TFF sequences in $\mathbb{R}^N$ with the frame bound α.

Majorization. The following definition comes from the majorization theory of the Schur–Horn theorem; see [Kaftal and Weiss 2010]. In algebraic combinatorics the majorization partial order on partitions is known as the *dominance order*; see [Fulton 1997].

Definition 2.2. Suppose that $\mathbf{L} = (L_1, L_2, \dots, L_K)$ and $\mathbf{L}' = (L'_1, L'_2, \dots, L'_{K'})$ be two weakly decreasing sequences of nonnegative integers. We say that $\mathbf{L}'$ majorizes $\mathbf{L}$, and write $\mathbf{L} \preccurlyeq \mathbf{L}'$ if for all $k \le \min(K, K')$,

$$\sum_{i=1}^{K} L_i = \sum_{i=1}^{K'} L'_i \quad \text{and} \quad \sum_{i=1}^{k} L_i \le \sum_{i=1}^{k} L'_i.$$

Observe that appending zeros at the tails of sequences $\mathbf{L}$, $\mathbf{L}'$ does not affect majorization relation. Moreover, for sequences with only positive terms, the majorization $\mathbf{L} \preccurlyeq \mathbf{L}'$ forces $K \geq K'$.

The majorization principle for TFF sequences takes the following form.

Theorem 2.3. *Let $\mathbf{L}$ and $\mathbf{L}'$ be two weakly decreasing sequences of positive integers such that $\mathbf{L} \preccurlyeq \mathbf{L}'$. Then, $\mathbf{L}' \in \mathrm{TFF}(\alpha, N)$ implies that $\mathbf{L} \in \mathrm{TFF}(\alpha, N)$.*

In the proof of Theorem 2.3 we use the following elementary result on a sum of two projections.

Lemma 2.4. *Fix positive integers $p > q \geq 0$. Let P and Q be two orthogonal projection of ranks p and q, respectively. Then, there exist orthogonal projections P' and Q' of ranks $p-1$ and $q+1$, respectively, such that $P + Q = P' + Q'$.*

Proof. Assume we have two projections P and Q with ranks $p > q$ that act on an N dimensional vector space V. Then, we can decompose V into the eigenspaces of P and Q such that

$$V = V_P \oplus V_P^{\perp}, \qquad V = V_Q \oplus V_Q^{\perp},$$

where V_P and $V_P^{\perp}$ denote the 1-eigenspace and 0-eigenspace, respectively. Since $p > q$, we have that $p + (N - q) > N$ and hence $\dim(V_P \cap V_Q^{\perp}) > 0$. Choose a nonzero vector in $V_P \cap V_Q^{\perp}$ and let R denote the corresponding rank 1 projection. Then, we can decompose $P = \bar{P} + R$, where $\bar{P}$ is a rank $p-1$ projection. Moreover, $Q + R$ is a projection of rank $q+1$. Thus, $P + Q = \bar{P} + (Q + R)$, which completes the proof of the lemma. □

Proof of Theorem 2.3. Since $\mathbf{L} \preccurlyeq \mathbf{L}'$ we can find a sequence of partitions $\mathbf{L} = \mathbf{L}^0 \preccurlyeq \mathbf{L}^1 \preccurlyeq \cdots \preccurlyeq \mathbf{L}^n = \mathbf{L}'$ such that any two consecutive partitions $\mathbf{L}^{j-1}$ and $\mathbf{L}^j$, $j = 1, \ldots, n$, differ at exactly two positions by ± 1. That is, for each $j = 1, \ldots, n$, there exist two positions $m < m' \in \mathbb{N}$ such that

$$\begin{aligned} \mathbf{L}^{j-1} &= (*, \ldots, *, \tilde{L}_m \qquad, *, \ldots, *, \tilde{L}_{m'} \qquad, *, \ldots, *), \\ \mathbf{L}^{j} &= (*, \ldots, *, \tilde{L}_m + 1, *, \ldots, *, \tilde{L}_{m'} - 1, *, \ldots, *), \end{aligned} \tag{2-2}$$

where the remaining values, denoted by $*$, are the same. Such partitions $\mathbf{L}^j$ can be easily constructed by the following recursive procedure.

Given the initial partitions $\mathbf{L}$ and $\mathbf{L}'$ we append extra zeros to $\mathbf{L}'$ so that $\mathbf{L}$ and $\mathbf{L}'$ have the same length. Define m to be the first position such that initial subsequences $(L_1, \ldots, L_m)$ and $(L'_1, \ldots, L'_m)$ are not the same. Likewise, m' is the last position such that the ending subsequences $(L_{m'}, \ldots)$ and $(L'_{m'}, \ldots)$ are not the same. Define $\mathbf{L}^1$ from $\mathbf{L}$ by replacing $L_m \to L_m + 1$ and $L_{m'} \to L_{m'} - 1$. It is not difficult to see that $\mathbf{L}^1$ forms a weakly decreasing sequence and $\mathbf{L} = \mathbf{L}_0 \preccurlyeq \mathbf{L}^1 \preccurlyeq \mathbf{L}'$.

Repeating this procedure recursively we define a sequence $\mathbf{L}^1 \preccurlyeq \mathbf{L}^2 \preccurlyeq \cdots \preccurlyeq \mathbf{L}'$. After a finite number of steps we must arrive at $\mathbf{L}^n = \mathbf{L}'$.

Observe that the ranks in (2-2) satisfy $\tilde{L}_m \geq \tilde{L}_{m'}$. By Lemma 2.4 applied to two projections with ranks $p = \tilde{L}_m + 1 > q = \tilde{L}_{m'} - 1 \geq 0$, if $\mathbf{L}^j \in \mathrm{TFF}(\alpha, N)$, then $\mathbf{L}^{j-1} \in \mathrm{TFF}(\alpha, N)$. Therefore, repeated application of Lemma 2.4 proves Theorem 2.3. □

We remark that the above proof does not use the tightness assumption in any way. Consequently, Theorem 2.3 holds for general (not necessarily tight) fusion frames with a prescribed frame operator.

Dualities. In this subsection we shall establish two dualities for TFF sequences. The first duality involves taking orthogonal projections of the same ambient space and is a straightforward generalization of [Casazza et al. 2011, Theorem 6].

Theorem 2.5. *Suppose that* $(L_1, L_2, \ldots, L_K) \in \mathrm{TFF}(\alpha, N)$. *Then*, $(N - L_K, N - L_{K-1}, \ldots, N - L_1) \in \mathrm{TFF}(K - \alpha, N)$.

Proof. Let $P_1, \ldots, P_K$ be the orthogonal projections with rank $P_i = L_i$ such that $\sum_{i=1}^K P_i = \alpha\mathbf{I}$. Clearly, $\sum_{i=1}^K (\mathbf{I} - P_i) = (K - \alpha)\mathbf{I}$ and $\mathrm{rank}(\mathbf{I} - P_i) = N - L_i$. □

The second result relies on taking more subtle orthogonal complements based on a dilation theorem for tight frames with bound 1, also known as *Parseval frames.* It is known that every Parseval frame can be obtained as a projection of an orthogonal basis of some higher dimensional space. The complementary projection gives rise to another Parseval frame, which is often called the *Naimark complement* of the original frame. This leads to the following result

Theorem 2.6. *Suppose that* $(L_1, L_2, \ldots, L_K) \in \mathrm{TFF}(\alpha, N)$. *Then, the same sequence* $(L_1, L_2, \ldots, L_K) \in \mathrm{TFF}(\tilde{\alpha}, \widetilde{N})$, *where the dimension* $\widetilde{N} = \left(\sum_{i=1}^K L_i - N\right)$ *and the frame bound* $\tilde{\alpha} = \alpha/(\alpha - 1) = \alpha N/\widetilde{N}$.

Proof. For each $k = 0, \ldots, K$, define $\sigma_k = \sum_{i=1}^k L_i$ with the convention that $\sigma_0 = 0$. Our assumption implies that there exists a tight frame $\{v_j\}_{j=1}^{\sigma_K}$ in $\mathbb{R}^N$ such that for each $k = 1, \ldots, K$, the subcollection $\{v_j\}_{j=1+\sigma_{k-1}}^{\sigma_k}$ is an orthonormal sequence which spans the L_k dimensional space W_k from the definition of a TFF. Treating $v_1, \ldots, v_{\sigma_K}$ as column vectors we obtain an $N \times \sigma_K$ matrix U with orthogonal rows each of norm $\sqrt{\alpha} = \sqrt{\sigma_K/N}$. This is due to the fact that $\{v_j\}_{j=1}^{\sigma_K}$ is a tight frame with constant α.

Let $\widetilde{U}$ be an extension of U to a $\sigma_K \times \sigma_K$ matrix with all orthogonal rows of norm $\sqrt{\alpha}$. In other words, $(1/\sqrt{\alpha})\widetilde{U}$ is a unitary extension of $(1/\sqrt{\alpha})U$ which has orthonormal rows. Let $\{w_j\}_{j=1}^{\sigma_K}$ be the column vectors constituting the $(\sigma_K - N) \times \sigma_K$ submatrix of the bottom rows of $\widetilde{U}$. Since $(1/\sqrt{\alpha})\widetilde{U}$ is an orthogonal matrix we have

$$\langle v_j, v_{j'} \rangle + \langle w_j, w_{j'} \rangle = \alpha \delta_{j,j'} \qquad \text{for all } j, j' = 1, \ldots, \sigma_K.$$

By the block orthogonality of vectors v_j we have that for each block $k = 1, \dots, K$,

$$\langle w_j, w_{j'} \rangle = (\alpha - 1)\delta_{j,j'} \qquad \text{for all } j, j' = 1 + \sigma_{k-1}, \dots, \sigma_k.$$

This means that the vectors $\{w_j\}_{j=1+\sigma_{k-1}}^{\sigma_k}$ form an orthogonal set which span some L_k dimensional space $\widetilde{W}_k$. Moreover, $\{w_j\}_{j=1}^{\sigma_K}$ is a tight frame with a constant α for $(\sigma_K - N)$ dimensional space. Consequently, unit norm vectors $\{(1/\sqrt{\alpha-1})w_j\}_{j=1}^{\sigma_K}$, which are block orthonormal, form a tight frame with a constant $\alpha/(\alpha-1)$. This leads to the decomposition $\widetilde{P}_1 + \cdots + \widetilde{P}_K = \alpha/(\alpha-1)\mathbf{I}$, where $\widetilde{P}_k$ is an orthogonal projection onto $\widetilde{W}_k$. □

As an immediate corollary of Theorem 2.6, we can reduce the study of TFF sequences to the case when $1 < \alpha < 2$; the case $\alpha = 2$ does not cause any difficulties as we will see later.

Corollary 2.7. *If $\alpha > 1$ is such that $\alpha N \in \mathbb{N}$, then* $\mathrm{TFF}(\alpha, N) = \mathrm{TFF}(\tilde{\alpha}, \widetilde{N})$, *where* $1/\alpha + 1/\tilde{\alpha} = 1$ *and* $\widetilde{N} = N(\alpha - 1)$.

Observe that if there exists a TFF sequence with parameters (α, N), then by computing traces we necessarily have that $\alpha N \in \mathbb{N}$. Hence, without loss of generality we shall always make this assumption.

3. Estimates on first 3 ranks

In this section we find necessary and sufficient conditions on the first three largest ranks of TFF projections. Our analysis is based on two fundamental results. Theorem 3.1 is due to Fillmore. Lemma 3.2 describes the spectral properties of the sum of two projections, and it can be thought of as a generalization of Lemma 2.4.

Theorem 3.1 [Fillmore 1969, Theorem 1]. *A nonnegative definite hermitian matrix S is a sum of projections if and only if*

$$\text{(3-1)} \qquad \operatorname{trace}(S) \in \mathbb{N}_0 \qquad \textit{and} \qquad \operatorname{trace}(S) \ge \operatorname{rank}(S).$$

Lemma 3.2. *Let P, Q be two orthogonal projections on an N dimensional vector space V with ranks p, q, respectively. For any $\lambda \in \mathbb{R}$, let $m(\lambda)$ be the multiplicity of λ as an eigenvalue of $P + Q$. Then, the following are true*:

(i) $m(\lambda) > 0 \Longrightarrow \lambda \in [0, 2]$,

(ii) $\sum_{\lambda \in [0,2]} m(\lambda) = N$,

(iii) $m(1) \ge |p - q|$,

(iv) $\lambda \in (0, 2) \Longrightarrow m(\lambda) = m(2 - \lambda)$,

(v) $m(0) - m(2) = N - p - q$.

Conversely, if $0 \leq p, q \leq N$ *and* $m : \mathbb{R} \to \mathbb{N}_0$ *satisfies* (*i*)–(*v*), *then there exists orthogonal projections* P, Q *of ranks* p, q, *such that* m *is a multiplicity function of* $P + Q$.

Proof. Since P, Q are hermitian, we can decompose V as a direct sum of eigenspaces

$$V = V_P \oplus V_P^{\perp} = V_Q \oplus V_Q^{\perp}$$

where V_P denotes the 1-eigenspace and $V_P^{\perp}$ the 0 eigenspace of P. Thus, $p = \dim(V_P)$ and $q = \dim(V_Q)$. Parts (i)–(iii) follow by basic linear algebra.

To prove part (iv) we define $f_\lambda : V \to V$ by

$$f_\lambda(v) := v_P + \Big(\frac{\lambda}{\lambda-2}\Big)v'_P,$$

where $v = v_P + v'_P$ is induced by the orthogonal decomposition $V = V_P \oplus V_P^{\perp}$ and $\lambda \in (0, 2)$. Since f_λ is an invertible and linear map, it suffices to show that if $(P + Q)v = \lambda v$, then $(P + Q) f_\lambda(v) = (2 - \lambda) f_\lambda(v)$. Write

$$v_P = x_Q + x'_Q \quad \text{and} \quad v'_P = y_Q + y'_Q$$

according to the decomposition $V = V_Q \oplus V_Q^{\perp}$. Then,

$$(P + Q)v = v_P + x_Q + y_Q = 2x_Q + y_Q + x'_Q = \lambda(x_Q + x'_Q + y_Q + y'_Q)$$

and hence

$$(2 - \lambda)x_Q + (1 - \lambda)y_Q = (\lambda - 1)x'_Q + \lambda y'_Q.$$

This implies that

$$(2 - \lambda)x_Q = (\lambda - 1)y_Q \quad \text{and} \quad (1 - \lambda)x'_Q = \lambda y'_Q \tag{3-2}$$

since $V_Q \cap V_Q^{\perp} = \{0\}$.

By (3-2), we have

$$\begin{aligned}
(P + Q) f_\lambda(v) &= 2x_Q + x'_Q + \Big(\frac{\lambda}{\lambda-2}\Big)y_Q \\
&= (2 - \lambda)v_P + \lambda x_Q + (\lambda - 1)x'_Q + \Big(\frac{\lambda}{\lambda-2}\Big)y_Q \\
&= (2 - \lambda)v_P + \Big(\frac{\lambda(1-\lambda)}{\lambda-2}\Big)y_Q - \lambda y'_Q + \Big(\frac{\lambda}{\lambda-2}\Big)y_Q \\
&= (2 - \lambda)v_P - \lambda y_Q - \lambda y'_Q \\
&= (2 - \lambda)\Big(v_P + \Big(\frac{\lambda}{\lambda-2}\Big)v'_P\Big) = (2 - \lambda) f_\lambda(v).
\end{aligned}$$

This proves part (iv). To prove part (v), we consider the projection map

$$g : V \to V_P + V_Q$$

where $V_P + V_Q$ denotes the span of vectors in V_P, V_Q. We have

$$\dim(V_P + V_Q) = \dim(V_P) + \dim(V_Q) - m(2) = p + q - m(2).$$

But,

$$\dim(V_P + V_Q) = N - \dim(\ker g) = N - m(0).$$

This shows that the properties (i)–(v) are necessary.

A quick way to see the converse direction is to utilize the honeycomb model of Knutson and Tao [1999; 2001]. The honeycombs corresponding to triples $(P, Q, -(P+Q))$, where $p > q$ can be represented by one of the following diagrams. In the case $p = q$ the line corresponding the eigenvalue -1 of $-(P+Q)$ might not be present. We leave the details to the reader. This involves finding multiplicities of unlabeled line segments to satisfy the "zero-tension" property. □

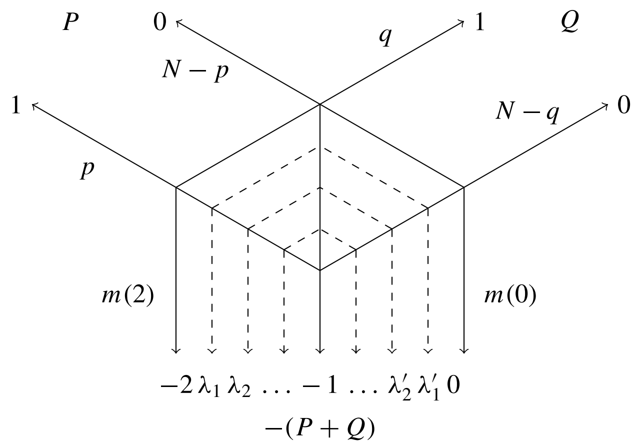


Figure 1. Honeycomb with $m(2) > 0$, $m(0) > 0$ and $\lambda_i' := -2 - \lambda_i$.

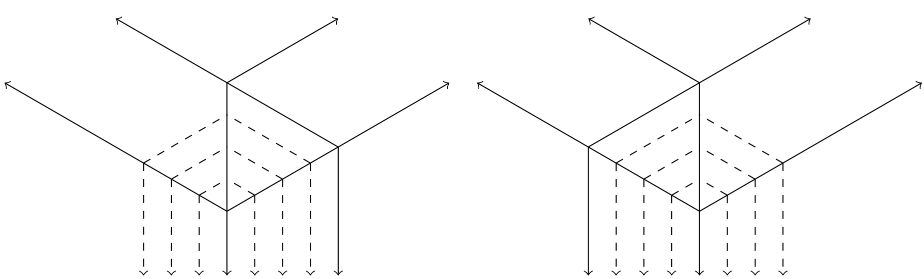


Figure 2. Honeycombs with $m(2) = 0$ and $m(0) = 0$, respectively.

Using Theorem 3.1 and Lemma 3.2, our goal is to find necessary and sufficient conditions on the first three largest ranks of projections in a TFF.

Theorem 3.3. *Suppose that* $1 < \alpha < 2$ *and* $(L_1 \geq L_2 \geq \cdots \geq L_K) \in \mathrm{TFF}(\alpha, N)$. *Then, we have the following necessary conditions*:

$$L_1 \leq (\alpha - 1)N, \tag{3-3}$$

$$L_1 + L_2 \leq N, \tag{3-4}$$

$$L_1 + L_2 + L_3 \leq \begin{cases} N, & \alpha < 3/2, \\ 2(\alpha - 1)N, & \alpha > 3/2. \end{cases} \tag{3-5}$$

Conversely, if $L_1 \geq L_2 \geq L_3$ *satisfy* (3-3), (3-4), *and* (3-5), *then there exists* $\mathbf{L} \in \mathrm{TFF}(\alpha, N)$ *which starts with the sequence* (L_1, L_2, L_3).

Proof. Suppose $\alpha\mathbf{I}$ is written as in (2-1). Then, $S = \alpha\mathbf{I} - P_1$ is an operator with 2 eigenvalues: α with multiplicity $N - L_1$ and $(\alpha - 1)$ with multiplicity L_1. By Theorem 3.1 we must have

$$\alpha N - L_1 \geq N.$$

Solving this for L_1 yields (3-3).

By Lemma 3.2 the sum $P_1 + P_2$ has eigenvalue 1 with multiplicity at least $L_1 - L_2$. Moreover, all other positive eigenvalues of this sum must come in pairs $(2 - \lambda, \lambda)$, where $1 \leq \lambda \leq \alpha < 2$. Thus, by Lemma 3.2(v), $L_1 + L_2 \leq N$. Let $S = \alpha\mathbf{I} - P_1 - P_2$. By Theorem 3.1, S must satisfy (3-1). Note that the trace of S remains constant regardless of choices of P_1 and P_2:

$$\mathrm{trace}(S) = \alpha N - L_1 - L_2.$$

Thus, the rank of S must be minimized to guarantee that it can be written as a sum of projections. The minimal rank of S occurs if $P_1 + P_2$ has eigenvalue α with multiplicity L_2, and thus eigenvalue $2 - \alpha$ with the same multiplicity. Then, the rank of the corresponding S is $N - L_2$. Thus, we have

$$\alpha N - L_1 - L_2 \geq N - L_2.$$

This leads again to (3-3). Therefore, Fillmore's theorem does not introduce new constraints in this case. In other words, (3-3) and (3-4) are both necessary and sufficient conditions for the existence of an element of $\mathrm{TFF}(\alpha, N)$ starting with (L_1, L_2).

Suppose next that $1 < \alpha < 3/2$. Repeating the above arguments, by Lemma 3.2, $P_1 + P_2$ must have all of its $L_1 + L_2$ nonzero eigenvalues (counted with multiplicities) in the interval $[2 - \alpha, \alpha]$. Thus, if $L_1 + L_2 + L_3 > N$, then at least one eigenvalue of $P_1 + P_2 + P_3$ would be at least $(2 - \alpha) + 1 > 3/2 > \alpha$, which is impossible. Thus, (3-5) is necessary.

To prove the converse, assume that $L_1 + L_2 + L_3 \leq N$. Using honeycomb models as in the proof of Lemma 3.2, one can show that there exist projections P_i such that their sum $P_1 + P_2 + P_3$ has the eigenvalue α with multiplicity $L_2 + L_3$, and no eigenvalues bigger than α. This is shown in a two step process. First, we construct P_1 and P_2 such that their sum has eigenvalues: α and $2-\alpha$ both with multiplicities L_2 and 1 with multiplicity $L_1 - L_2$. Then, using a honeycomb model we can add on another projection P_3, such that $P_1 + P_2 + P_3$ has eigenvalue α with multiplicity $L_2 + L_3$. See Figure 3 for an illustration of this honeycomb construction. We now have an operator $S = \alpha\mathbf{I} - (P_1 + P_2 + P_3)$ with the rank $N - L_2 - L_3$. The trace of S remains constant regardless of the choice of such projections:

$$\operatorname{trace}(S) = \alpha N - L_1 - L_2 - L_3.$$

Since $L_1 \leq (\alpha - 1)N$, Fillmore's Theorem 3.1 can be applied to represent S as a sum of projections. This proves that (3-3)–(3-5) are both necessary and sufficient conditions for the first 3 ranks of a TFF sequence in the case $1 < \alpha < 3/2$.

In the special case $\alpha = 3/2$, it is easy to see that $(N/2, N/2, N/2)$ is the unique maximal element in $\mathrm{TFF}(\alpha, N)$; see Theorem 3.4. Unfortunately, the case $3/2 < \alpha < 2$ does not seem to be easily approachable with the techniques of this section. Instead, in Section 5 we shall give another combinatorial proof of Theorem 3.3 which works in the entire range $1 < \alpha < 2$. □

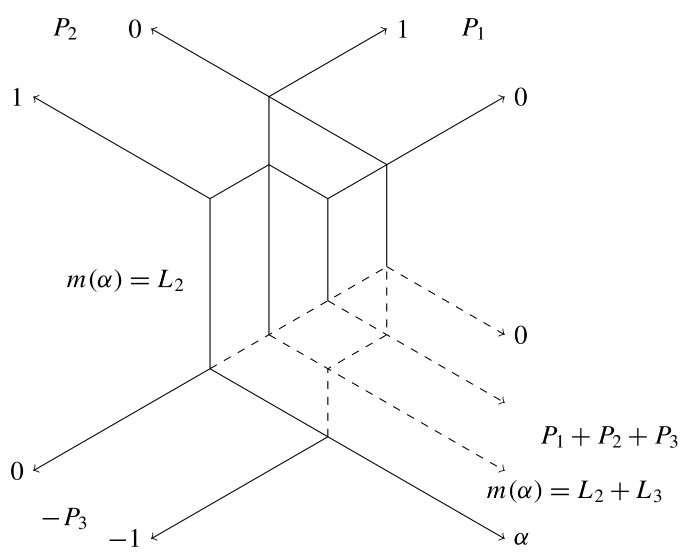


Figure 3. Honeycomb of $P_1 + P_2 + P_3$ with maximum eigenvalue α of multiplicity $L_2 + L_3$.

We end this section by an explicit characterization of TFF sequences for some special values α.

Theorem 3.4. *The set* TFF(α, N) *has exactly one maximal element* $\mathbf{L}$ *with respect to majorization relation* $\preccurlyeq$ *in the following four cases indexed by* $n \in \mathbb{N}$:

$$\alpha = n, \qquad \mathbf{L} = (\underbrace{N, N, \dots, N}_{n}), \tag{3-6}$$

$$\alpha = 1 + \frac{1}{n}, \quad n|N, \qquad \mathbf{L} = \Bigl(\underbrace{\frac{N}{n}, \frac{N}{n}, \dots, \frac{N}{n}}_{n+1}\Bigr), \tag{3-7}$$

$$\alpha = n + \frac{1}{2}, \quad 2|N, \qquad \mathbf{L} = \Bigl(\underbrace{N, \dots, N}_{n-1}, \frac{N}{2}, \frac{N}{2}, \frac{N}{2}\Bigr), \tag{3-8}$$

$$\alpha = 1 + \frac{2}{2n-1}, \quad (2n-1)|N, \tag{3-9}$$

$$\mathbf{L} = \Bigl(\underbrace{\frac{2N}{2n-1}, \dots, \frac{2N}{2n-1}}_{n-1}, \frac{N}{2n-1}, \frac{N}{2n-1}, \frac{N}{2n-1}\Bigr).$$

Proof. The case (3-6) is the easiest and it follows immediately from Theorem 2.3. The case (3-7) is obtained by the duality argument. Indeed, note that if $\alpha = 1 + 1/n$, then n must divide N. Then, by Corollary 2.7, TFF$(\alpha, N) =$ TFF$(\tilde{\alpha}, \widetilde{N})$, where $\tilde{\alpha} = \alpha/(\alpha - 1) = n + 1$ and $\widetilde{N} = (\alpha - 1)N = N/n$.

In particular, we have that TFF$(3/2, N) =$ TFF$(3, N/2)$ has a unique maximal element $(N/2, N/2, N/2)$. By appending $(n-1)$ N's in the front of this sequence we obtain a maximal element of TFF$(n + 1/2, N)$. It remains to show that this is the only maximal element.

Suppose that we have another element $(L_1, \dots, L_K) \in$ TFF$(n+1/2, N)$. Let P_i's be the corresponding projections. Given two hermitian matrices S and T we write $S \le T$ if $\langle Sx, x\rangle \le \langle Tx, x\rangle$ for all $x \in \mathbb{R}^N$. Since $\sum_{i=1}^{n} P_i \le n\mathbf{I}$, $S = \sum_{i=n+1}^{K} P_i$ must have full rank N. By Fillmore's Theorem 3.1, this implies that

$$\operatorname{trace}(S) = \sum_{i=n+1}^{K} L_i \ge N.$$

Thus, $L_1 + \dots + L_n \le (n - 1/2)N$.

Suppose on the contrary that $L_1 + \dots + L_{n+1} > nN$. For each i, let W_i be the corresponding subspace with $\dim W_i = L_i$. By basic linear algebra, the intersection satisfies

$$\dim\Bigl(\bigcap_{i=1}^{n+1} W_i\Bigr) \ge L_1 + \dots + L_{n+1} - nN > 0.$$

This implies that $P_1 + \cdots + P_{n+1}$ has eigenvalue $n+1$ exceeding $\alpha = n+1/2$, which is a contradiction. Thus, we have necessarily $L_1+\cdots+L_{n+1} \le nN$. Clearly,

$$L_1+\cdots+L_{n+2} \le L_1+\cdots+L_K = (n+1/2)N.$$

Consequently, $(L_1, \ldots, L_K) \preccurlyeq \mathbf{L}$, proving (3-8).

Finally, case (3-9) is shown by the duality argument. If $\alpha = 1+2/(2n-1)$, then $2n-1$ must divide N. Then, by Corollary 2.7, $\mathrm{TFF}(\alpha, N) = \mathrm{TFF}(\tilde{\alpha}, \widetilde{N})$, where $\tilde{\alpha} = \alpha/(\alpha-1) = n+1/2$ and $\widetilde{N} = (\alpha-1)N = 2N/(2n-1)$. □

Section 7 provides the list of all maximal elements in $\mathrm{TFF}(\alpha, N)$ for all $\alpha \le 2$ and dimensions $N \le 9$. It is easy to observe that all unique maximal elements in our tables are covered by Theorem 3.4. Hence, it is very tempting to conjecture that for general α and N, if $\mathrm{TFF}(\alpha, N)$ has only one maximal element, then α must necessarily come from one of the four cases of Theorem 3.4.

4. A combinatorial characterization of tight fusion frames

In this section we give a combinatorial characterization of tight fusion frames in the context of Schur functions. The main result of this section, Theorem 4.3, is a direct consequence of Horn's recursion for the hermitian eigenvalue problem (for a survey of this problem see [Fulton 2000]). For completeness, we state the main results of this body of work. For any partition

$$\lambda = (\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_d > 0),$$

let

$$|\lambda| = \sum_{i=1}^{d} \lambda_i$$

denote the size of λ and let d denote the length. We say λ is a rectangular partition if

$$\lambda = (a^b) := (\underbrace{a, \ldots, a}_{b})$$

for some positive integers a, b. For any partition λ, let s_λ denote the corresponding Schur polynomial. The polynomial s_λ is a homogeneous polynomial of degree $|\lambda|$. It is well known that the Schur polynomials form a linear basis of the algebra of symmetric polynomials with integer coefficients. Hence for any collection of partitions $\lambda^1, \ldots, \lambda^K$ we can define the corresponding *Littlewood–Richardson coefficients* $c(\lambda^1, \ldots, \lambda^K; \mu)$ as the product structure constants of

$$\prod_{i=1}^{K} s_{\lambda^i} = \sum_{\mu} c(\lambda^1, \ldots, \lambda^K; \mu)\, s_\mu.$$

The Littlewood–Richardson coefficients defined above play an important role in the hermitian eigenvalue problem. To state these results, we first need some notation. There is a standard identification between sets of positive integers of size r and partitions of length at most r. For any set $I = \{i_1 < i_2 < \cdots < i_r\}$, define the partition

$$\lambda(I) := (i_r - r, i_{r-1} - r + 1, \ldots, i_1 - 1).$$

Let $(\beta^1, \ldots, \beta^{K+1}) \in (\mathbb{R}^N)^{K+1}$ denote a collection of sequences where each $\beta^i := (\beta^i_1 \geq \cdots \geq \beta^i_N)$. The goal of the hermitian eigenvalue problem is to determine for which sequences $(\beta^1, \ldots, \beta^{K+1})$ do there exist $N \times N$ hermitian matrices $H_1, \ldots, H_{K+1}$ such that the eigenvalues of H_i are given by the sequence β^i and

$$\sum_{i=1}^{K} H_i = H_{K+1}.$$

The following theorem, due to Klyachko, gives a remarkable characterization in terms of a collection of inequalities parametrized by nonzero Littlewood–Richardson coefficients.

Theorem 4.1 [Klyachko 1998]. *Let $(\beta^1, \ldots, \beta^{K+1}) \in (\mathbb{R}^N)^{K+1}$ be a collection of sequences of nonincreasing real numbers such that*

$$\sum_{i=1}^{K} \sum_{j=1}^{N} \beta^i_j = \sum_{j'=1}^{N} \beta^{K+1}_{j'}.$$

Then the following are equivalent:

(1) *There exist $N \times N$ hermitian matrices $H_1, \ldots, H_{K+1}$ with spectra $(\beta^1, \ldots, \beta^{K+1})$ such that*

$$\sum_{i=1}^{K} H_i = H_{K+1}.$$

(2) *For every $r < N$, the sequence $(\beta^1, \ldots, \beta^{K+1})$ satisfies the inequality*

$$\sum_{i=1}^{K} \sum_{j \in I^j} \beta^i_j \geq \sum_{j' \in I^{K+1}} \beta^{K+1}_{j'} \tag{4-1}$$

for every collection of subsets $I^1, \ldots, I^{K+1}$ of size r of integers $\{1, 2, \ldots, N\}$ where the Littlewood–Richardson coefficient

$$c\big(\lambda(I^1), \ldots, \lambda(I^K); \lambda(I^{K+1})\big) \neq 0.$$

The inequalities given in (4-1) are called Horn's inequalities and were initially defined in a very different way by Horn [1962]. While Horn's list of inequalities are, a priori, different than Klyachko's list (4-1), they were shown to be equivalent as a consequence of the saturation theorem of Knutson and Tao [1999]. What is amazing about this equivalence is that Horn's initial definition of the inequalities (4-1) uses a recursion unrelated to Littlewood–Richardson coefficients. Horn's recursion, in light of Theorem 4.1, can be stated as follows:

Theorem 4.2. *Let $I^1, \dots, I^{K+1}$ be subsets of size r of integers $\{1, 2, \dots, N\}$ such that*

$$\sum_{i=1}^{K} \sum_{j=1}^{r} \lambda(I^i)_j = \sum_{j'=1}^{r} \lambda(I^{K+1})_{j'}. \tag{4-2}$$

The following are equivalent:

(1) *The Littlewood–Richardson coefficient*

$$c(\lambda(I^1), \dots, \lambda(I^K); \lambda(I^{K+1})) \neq 0.$$

(2) *There exist $r \times r$ hermitian matrices $H_1, \dots, H_{K+1}$ that have spectra*

$$(\lambda(I^1), \dots, \lambda(I^{K+1}))$$

such that

$$\sum_{i=1}^{K} H_i = H_{K+1}. \tag{4-3}$$

The recursion says that a collection of subsets $I^1, \dots, I^{K+1}$ corresponds to a Horn inequality if and only if the corresponding collection of partitions are eigenvalues of some $r \times r$ hermitian matrices which satisfy (4-3). Hence Horn's inequalities can be defined recursively by induction on N. We also remark that (4-2) is a necessary condition for the corresponding Littlewood–Richardson coefficient to be nonzero since it is equivalent to the grading condition when multiplying Schur functions in the graded algebra of symmetric functions.

We now apply Theorem 4.2 to the case of tight fusion frames. Suppose that $(L_1 \geq L_2 \geq \cdots \geq L_K) \in \mathrm{TFF}(\alpha, N)$ and that $M := \sum_{i=1}^{K} L_i$. Then there exist orthogonal projections $P_1, \dots, P_K$ such that

$$\sum_{i=1}^{K} N P_i = M\mathbf{I}. \tag{4-4}$$

Since P_i is an orthogonal projection, the spectrum of the hermitian matrix NP_i is given by

$$(\underbrace{N, \dots, N}_{L_i}, \underbrace{0, \dots, 0}_{N-L_i}).$$

Let (N^{L_i}) denote the corresponding rectangular partition to the spectra above. The following is a direct corollary of Theorem 4.2.

Theorem 4.3. *Fix an integer N and let $(L_1 \geq L_2 \cdots \geq L_K)$ be a sequence of nonnegative integers such that $L_1 \leq N$. Let $M := \sum_{i=1}^{K} L_i$ and $\alpha = M/N$. The following are equivalent*:

(1) *The sequence $(L_1 \geq L_2 \geq \cdots \geq L_K) \in \mathrm{TFF}(\alpha, N)$.*

(2) *The Littlewood–Richardson coefficient*

$$c\big((N^{L_1}), \dots, (N^{L_K}); (M^N)\big) \neq 0.$$

Proof. Assume part (1). Then there exist orthogonal projections $P_1, \dots, P_K$ with ranks $(L_1, \dots, L_K)$ such that

$$\sum_{i=1}^{K} P_i = \alpha \mathbf{I}. \tag{4-5}$$

Multiplying both sides of (4-5) by N gives (4-4). Applying Theorem 4.2 gives part (2).

Conversely, if we assume part (2) then by Theorem 4.2, there exists a collection of $N \times N$ matrices which satisfy (4-4) and have spectra $(N^{L_1}), \dots, (N^{L_K})$. Scaling by $1/N$ yields the desired tight fusion frame. □

The condition that $c\big((N^{L_1}), \dots, (N^{L_K}); (M^N)\big) \neq 0$ can be made computationally explicit by the following existence condition. With the notation of Theorem 4.3 we consider the following properties for an $N \times M$ matrix $A = A[i, j]$.

(i) integral nonnegativity: $A[i, j] \in \mathbb{Z}_{\geq 0}$

(ii) row sum: $\sum_{j=1}^{M} A[i, j] = M$ for all i

(iii) column sum: $\sum_{i=1}^{N} A[i, j] = N$ for all j

(iv) row sum dominance: $\sum_{j=1}^{l} (A[i, j] - A[i+1, j]) \geq A[i+1, l+1]$ for all i, l

(v) column sum dominance: $\sum_{i=1}^{l} (A[i, j] - A[i, j+1]) \geq A[l+1, j+1]$ for all j, l

Note that we can take l to be zero in conditions (iv) and (v) since in this case the sums are by definition equal to zero. Observe that properties (iv) and (v) require dominance with one additional summand in the later row or column. Also note that (ii) and (iii) are the only properties dependent on the size of the matrix A. Let A be an $N \times M$ matrix and consider the sequence $(L_1, \dots L_K)$. We can partition A into a sequence of column block matrices

$$A = [A_1 \mid A_2 \mid \cdots \mid A_K]$$

where each A_i is the corresponding $N \times L_i$ submatrix of A. We now have the following addition to Theorem 4.3.

Corollary 4.4. *Conditions* (1) *and* (2) *in Theorem 4.3 are equivalent to*:

(3) *There exists an $N \times M$ matrix A which satisfies properties* (i)–(iv) *and whose column block submatrices $A_1, \dots, A_K$ each satisfy property* (v).

Moreover, the coefficient $c\big((N^{L_1}), \dots, (N^{L_K}); (M^N)\big)$ equals the number of $N \times M$ matrices A which satisfy (3).

Proof. We refer to [Fulton 1997] for definitions and details of Littlewood–Richardson skew tableaux. Consider the Littlewood–Richardson coefficients $c^{\nu}_{\lambda,\mu}$ corresponding to the product of two Schur functions

$$s_\lambda s_\mu = \sum_{\nu} c^{\nu}_{\lambda,\,\mu}\, s_\nu.$$

It is well known that the number $c^{\nu}_{\lambda,\,\mu}$ is precisely equal to the number of Littlewood–Richardson skew tableaux ν/λ of content μ. Now suppose there exists an $N \times M$ matrix A which satisfies the conditions of Corollary 4.4 with respect to a sequence $\mathbf{L} = (L_1, \dots, L_K)$. For any $k \le K$ let

$$A(k) := [A_1 \mid \cdots \mid A_k]$$

denote the submatrix of A consisting of the matrices $A_1, \dots, A_k$. By properties (i) and (iv), the row sums of $A(k)$ yield a partition

$$\mu^k := \Big(\sum_j A(k)[i, j]\Big)_{i=1}^{N} \tag{4-6}$$

given in the standard weakly decreasing form. It is easy to see that μ^k/μ^{k-1} is a well defined skew partition. Consider the Young diagram corresponding to μ^k/μ^{k-1}. We can fill the boxes of the j-th row of this diagram with $A_k[j, 1]$ 1s, $A_k[j, 2]$ 2s, $A_k[j, 3]$ 3's and so forth in weakly increasing order. Property (iv) implies that the shape of μ^k/μ^{k-1} is a valid skew partition and that the entries in each column are strictly decreasing. Property (v) implies that the row reading word is reverse lattice. Together this implies the resulting skew tableau is a Littlewood–Richardson skew tableau. Property (iii) implies that content of the tableau is that of the rectangular partition (N^{L_k}). Hence the existence of the matrix $A(k)$ implies that the Littlewood–Richardson coefficient

$$c^{\mu^k}_{\mu^{k-1},\,(N^{L_k})} \neq 0. \tag{4-7}$$

Properties (ii) and (iii) imply that $\mu^K = (M^N)$. By induction on k, multiplying the Schur functions $s_{(N^{L_1})}, \dots s_{(N^{L_K})}$ gives

$$c\big((N^{L_1}), \dots, (N^{L_K}); (M^N)\big) \neq 0.$$

Conversely, if the second part of Theorem 4.3 holds, then there exists a sequence of partitions $\mu^1, \mu^2, \dots, \mu^K$ such that (4-7) holds with $\mu^1 = (N^{L_1})$ and $\mu^K = (M^N)$. In particular, there exists a Littlewood–Richardson skew tableau of shape μ^k/μ^{k-1} with content (N^{L_k}). We can construct a matrix A_k which satisfies property (v) using the entries of this Littlewood–Richardson skew tableau by reversing the argument above. Moreover, taking $A = [A_1 \mid \cdots \mid A_K]$, we have that A satisfies all the conditions part (3) of Corollary 4.4.

Finally, the Littlewood–Richardson rule states that

$$c^{\mu^k}_{\mu^{k-1},\,(N^{L_k})}$$

is precisely the number of Littlewood–Richardson skew tableaux of shape μ^k/μ^{k-1} with content (N^{L_k}). Hence, the second part of Corollary 4.4 follows from the bijection given by (4-6). This completes the proof. □

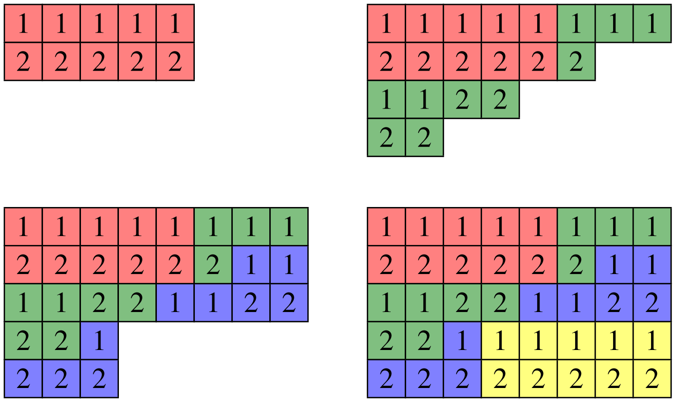
Example 4.5. We consider two examples where tight fusion frames exist for $N = 5$ and $M = 8$.

First, consider the sequence $\mathbf{L} = (2, 2, 2, 2)$. The matrix

$$A = \begin{pmatrix} 5 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 2 & 2 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 5 \end{pmatrix}$$

satisfies the conditions in Corollary 4.4. We write out the corresponding Young tableaux to the partitions μ^1, μ^2, μ^3 and μ^4 with content given by the submatrices

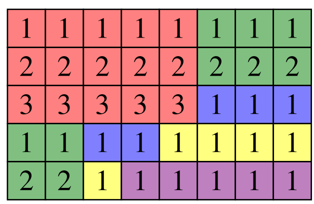
$A(1)$, $A(2)$, $A(3)$, $A(4)$:



Note that the all the data can be encoded in the final partition μ^4 as a union of skew Littlewood–Richardson tableaux.

For the second example, we consider $\mathbf{L} = (3, 2, 1, 1, 1)$, the matrix A and the corresponding union of Littlewood–Richardson tableaux

$$A = \left(\begin{array}{ccc|cc|c|c|c} 5 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 2 & 4 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 1 & 5 \end{array}\right)$$



.

We end this section with an important observation concerning the partial sum of orthogonal projections $P_1 + \cdots + P_k$. Let A be a matrix as in Corollary 4.4 for some $\mathbf{L} \in \mathrm{TFF}(\alpha, N)$ and consider the sequence of partitions $\mu^1, \ldots, \mu^K$ corresponding to A. For each $k \leq K$, we have that the coefficient

$$c\big((N^{L_1}), \ldots, (N^{L_k}); \mu^k\big) \neq 0.$$

This fact together with Theorem 4.2 yields the following corollary

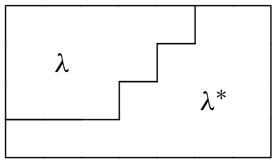
Corollary 4.6. *Let A be a matrix as in Corollary 4.4 for some $\mathbf{L} \in \mathrm{TFF}(\alpha, N)$ and consider the sequence of partitions $\mu^1, \ldots, \mu^K$ corresponding to A. Then there exist orthogonal projections $P_1, \ldots, P_K$ such that (4-4) holds and each partial sum $\sum_{i=1}^{k} P_i$ has spectrum $(\mu_1^k/N, \ldots, \mu_N^k/N)$.*

We remark that Corollary 4.6 is useful for the construction of explicit tight fusion frames. We illustrate this process later in Example 7.2.

5. Combinatorial majorization and hook type sequences

In this section we give alternate proofs of Theorem 2.3 on majorization as well as Theorem 3.3 on estimates using the combinatorics of Schur functions and Theorem 4.3. We begin with some fundamental definitions and lemmas on Schur functions. Let $\lambda \subseteq (M^N)$. We define the dual partition of λ in (M^N) to be

$$\lambda^* := (M - \lambda_N \geq M - \lambda_{N-1} \geq \cdots \geq M - \lambda_1).$$



Lemma 5.1. *Let $\lambda \subseteq (M^N)$ and let $p(\lambda)$ denote the number of parts of λ equal to M. Assume that for some positive integer k we have*

$$|\lambda| = N(M - k).$$

Then

$$c\big(\lambda, \underbrace{(N), \ldots, (N)}_{k}; (M^N)\big) \neq 0$$

if and only if $k \geq N - p(\lambda)$.

Proof. The lemma follows from two elementary facts about Schur functions. Consider the product

$$(s_{(N)})^k = \sum_{\mu} c((N), \ldots, (N); \mu)\, s_\mu$$

By the Pieri rule, we have that $c((N), \ldots, (N); \mu) \neq 0$ if and only if μ has length less than or equal to k and $|\mu| = Nk$. Furthermore, if $\lambda, \mu \subseteq (M^N)$, then $c_{\lambda,\mu}^{(M^N)} \neq 0$ if and only if $\mu = \lambda^*$. It is easy to check that λ^* appears as a summand in the product $(s_{(N)})^k$ precisely when $k \geq N - p(\lambda)$. □

The following theorem on the product of Schur functions corresponding to rectangular partitions is proved by Okada.

Theorem 5.2 [Okada 1998, Theorem 2.4]. *Fix integers a, b, N_1, N_2 with $a \geq b$. The product of Schur functions*

$$s_{(N_1^a)} s_{(N_2^b)} = \sum_{\lambda} s_\lambda, \tag{5-1}$$

where the sum is over all partitions λ with length $\leq a + b$ such that

- $\lambda_{b+1} = \lambda_{b+2} = \cdots = \lambda_a = N_1$,
- $\lambda_b \geq \max\{N_1, N_2\}$,
- $\lambda_i + \lambda_{a+b+1-i} = N_1 + N_2$ *for all $i \in \{1, \ldots, b\}$.*

We now give an alternate proof of Theorem 2.3 using Theorem 5.2 in the case when $N_1 = N_2$.

Lemma 5.3. *Fix a positive integer N and let $0 < a < b$. Then the Littlewood–Richardson coefficients*

$$c^{\lambda}_{(N^b),\,(N^a)} \leq c^{\lambda}_{(N^{b-1}),\,(N^{a+1})}.$$

In particular, Theorem 2.3 on majorization of tight fusion frames follows.

Proof. It is easy to check the λ that appear in the summation (5-1) for the pair $((N^b), (N^a))$ are contained in the λ that appear in the summation (5-1) for the pair $((N^{b-1}), (N^{a+1}))$. This proves the inequality. The application to tight fusion frames follows from Theorem 4.3. □

It is easy to see that by majorization, the following theorem is equivalent to Theorem 3.3 on estimates.

Theorem 5.4. *Assume the conditions in Theorem 4.3. Further assume that $\alpha = M/N < 2$. If $(L_1 \geq L_2 \geq \cdots \geq L_K) \in \mathrm{TFF}(\alpha, N)$, then we have the following necessary conditions:*

(1) $L_1 \leq M - N$.

(2) $L_1 + L_2 \leq N$.

(3) *If $\alpha > 3/2$, then $L_1 + L_2 + L_3 \leq 2(M - N)$.*

(4) *If $\alpha < 3/2$, then $L_1 + L_2 + L_3 \leq N$.*

Conversely, suppose L_1, L_2, L_3 satisfy conditions (1)–(4) *and $L_4 = \cdots = L_K = 1$. Then $(L_1 \geq L_2 \geq \cdots \geq L_K) \in \mathrm{TFF}(\alpha, N)$.*

Proof. Recall that for any partition $\lambda \subseteq (M^N)$, we let $p(\lambda)$ denote the number of parts of λ equal to M. First we prove part (1). By majorization, it suffices to assume that $L_2 = 1$. Part (1) now follows from Lemma 5.1 by setting $\lambda = (N^{L_1})$ and observing that $p((N^{L_1})) = 0$.

We now prove part (2). By majorization, it suffices to assume that $L_3 = 1$. Consider the product

$$s_{(N^{L_1})} s_{(N^{L_2})} = \sum_{\lambda} s_{\lambda}. \tag{5-2}$$

By Theorem 5.2, we have that $\lambda_1 + \lambda_{L_1+L_2} = 2N$ for every λ in the sum (5-2). If $\lambda \subseteq (M^N)$, then $\lambda_1 \leq M$. Note that such a partition λ exists since $(L_1 \geq L_2 \geq \cdots \geq L_K) \in \mathrm{TFF}(\alpha, N)$. Hence

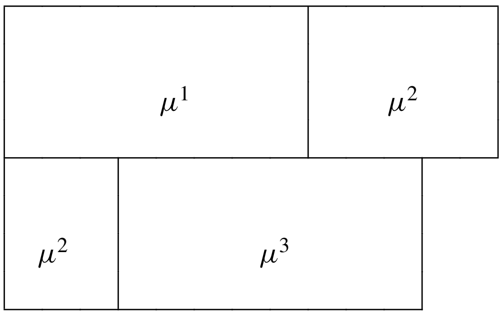
$$\lambda_{L_1+L_2} = 2N - \lambda_1 \geq 2N - M > 0$$

since $\alpha < 2$. This implies that $L_1 + L_2 \leq N$ since (M^N) has only N parts.

For part (3), we assume that $L_4 = 1$. First, if $L_2 + L_3 \leq L_1$, then by part (1), $L_1 + L_2 + L_3 \leq 2(M - N)$. Next, we assume $L_1 \leq L_2 + L_3$. Consider the product

$$s_{(N^{L_1})} s_{(N^{L_2})} s_{(N^{L_3})} = \sum_{\lambda} c\big((N^{L_1}), (N^{L_2}), (N^{L_3}); \lambda\big)\, s_\lambda. \tag{5-3}$$

Since $\alpha > 3/2$, for any $\lambda \subseteq (M^N)$ such that $c\big((N^{L_1}), (N^{L_2}), (N^{L_3}); \lambda\big) \neq 0$, we have $p(\lambda) \leq L_1$. This can be seen by considering L_2 and L_3 as large as possible, hence $L_1 = L_2 = L_3$. One can show using the Littlewood–Richardson rule that since $3N < 2M$, three layered bricks of width N cannot span M more than once; see diagram below.



By Lemma 5.1,

$$M - L_1 - L_2 - L_3 \geq N - p(\lambda) \geq N - L_1.$$

Hence $L_2 + L_3 \leq M - N$. This proves part (3).

For part (4), fix any λ in the summand found in (5-2) such that $\lambda \subseteq (M^N)$. Since $\alpha < 3/2$, we have

$$\lambda_{L_1+L_2} = 2N - \lambda_1 \geq 2N - M > M - N.$$

Hence the rectangular partition $((M - N + 1)^{L_1+L_2}) \subseteq \lambda$. Comparing the two products

$$s_\lambda s_{(N^{L_3})} = \sum_{\mu'} c^{\mu'}_{\lambda,(N^{L_3})} s_{\mu'} \tag{5-4}$$

and

$$s_{((M-N+1)^{L_1+L_2})} s_{(N^{L_3})} = \sum_{\mu} s_\mu \tag{5-5}$$

we have that any partition μ' from (5-4) such that $c^{\mu'}_{\lambda,(N^{L_3})} \neq 0$ contains some μ from (5-5). Therefore it is enough to consider the partitions μ from (5-5). By Theorem 5.2, we get

$$\mu_1 + \mu_{L_1+L_2+L_3} = M - N + 1 + N = M + 1$$

for every μ in the sum (5-5). Hence if $\mu \subseteq (M^N)$, then $\mu_{L_1+L_2+L_3} > 0$ since $\mu_1 \le M$. Note that such a partition μ exists since $(L_1 \ge L_2 \ge \cdots \ge L_K) \in \mathrm{TFF}(\alpha, N)$. Thus $L_1 + L_2 + L_3 \le N$. This proves part (4).

To prove sufficiency, we construct λ in the sum (5-3) such that $\lambda \subseteq (M^N)$ and $c\big((N^{L_1}), (N^{L_2}), (N^{L_3}); \lambda\big) \neq 0$. One can show using the Littlewood–Richardson rule that parts (1)–(4) imply that such a λ exists. Furthermore, we can construct λ such that $p(\lambda)$ satisfies the following:

- If $L_1 \ge L_2 + L_3$, then $p(\lambda) = L_2 + L_3$.
- If $L_1 < L_2 + L_3$ and $\alpha < 3/2$, then $p(\lambda) = L_2 + L_3$.
- If $L_1 < L_2 + L_3$ and $\alpha > 3/2$, then $p(\lambda) = \lfloor \frac{1}{2}(L_1 + L_2 + L_3) \rfloor$.

See Figures 4 and 5 for a diagram of λ in the cases above. If $p(\lambda) = L_2 + L_3$, then Lemma 5.1 implies we need to check that $L_1 \le M - N$, which is a necessary condition. The last case is the most delicate. We construct λ with a partition μ_2 of the shape as in Figure 5 with a notch of height x. If $L_1 + L_2 + L_3$ is even, then we choose x such that $p(\lambda) = x + L_3 = L_1 + L_2 - x$, and hence $p(\lambda) = \frac{1}{2}(L_1 + L_2 + L_3)$. If $L_1 + L_2 + L_3$ is odd, then $p(\lambda) = \lfloor \frac{1}{2}(L_1 + L_2 + L_3) \rfloor$. Either way, Lemma 5.1 implies we need to check that $L_1 + L_2 + L_3 \le 2(M - N)$, which again is a necessary condition. This completes the proof of the theorem. □

Remark 5.5. Parts (2) and (4) of Theorem 5.4 can be generalized to the following. Claim: Let $2 \le k \le K$. If $\alpha < k/(k-1)$, then $L_1 + \cdots + L_k \le N$. The proof follows the same argument as the proof of Theorem 5.4 part (4).

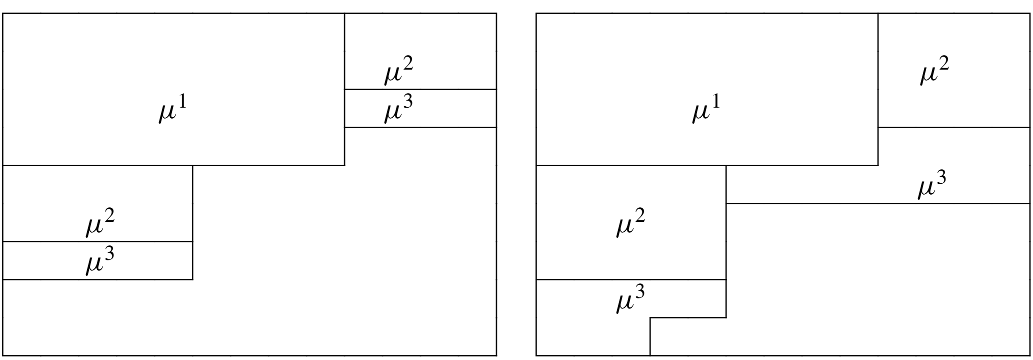


Figure 4. Construction of λ for $\alpha < 3/2$ as a union of Littlewood–Richardson skew tableaux μ^1, μ^2, μ^3 when $L_2 + L_3 \le L_1$ and $L_1 \le L_2 + L_3$, respectively. This construction is possible since $L_1 + L_2 + L_3 \le N$ and $2M < 3N$.

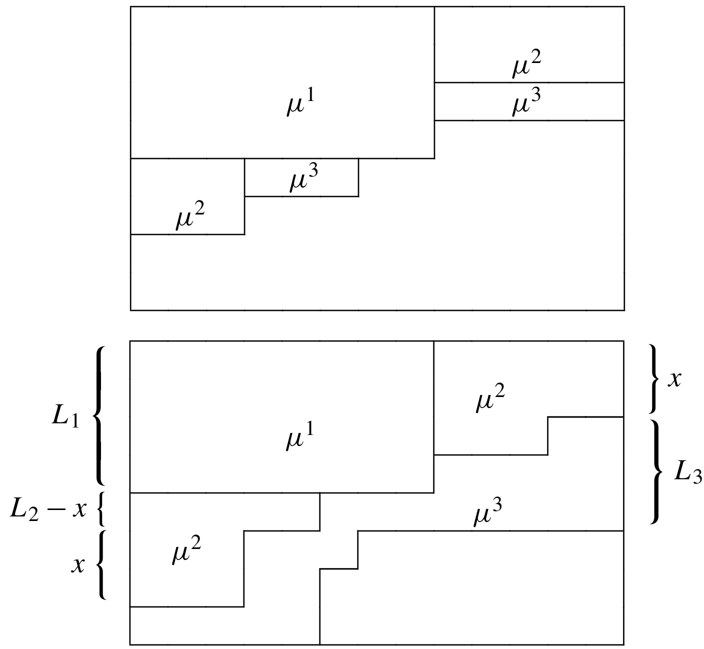


Figure 5. Construction of λ for $\alpha > 3/2$ as a union of Littlewood–Richardson skew tableaux μ^1, μ^2, μ^3 when $L_2 + L_3 \le L_1$ and $L_1 \le L_2 + L_3$, resp.

6. Combinatorial spatial and Naimark duality

Theorems 2.5 and 2.6 establish spatial and Naimark dualities for tight fusion frames. By Theorem 4.3, we have the analogous results for Littlewood–Richardson coefficients.

Corollary 6.1. *Assume we have a sequence of integers* $(L_1 \ge \cdots \ge L_K)$ *as in Theorem 4.3. Then*

$$c\big((N^{L_1}), \dots, (N^{L_K}); (M^N)\big) \neq 0 \iff c\big((N^{N-L_1}), \dots, (N^{N-L_K}); ((KN-M)^N)\big) \neq 0$$

and

$$c\big((N^{L_1}), \dots, (N^{L_K}); (M^N)\big) \neq 0 \iff c\big(((M-N)^{L_1}), \dots, ((M-N)^{L_K}); (M^{(M-N)})\big) \neq 0.$$

In this section we prove a much stronger version of the corollary above. In particular, we prove that these Littlewood–Richardson coefficients are equal. We will frequently reference properties (i)–(v) for matrices defined in the paragraph preceding Corollary 4.4 using lower case roman numerals. We first consider spatial duality.

Theorem 6.2. *The Littlewood–Richardson coefficients*

$$c\big((N^{L_1}),\dots,(N^{L_K});(M^N)\big) = c\big((N^{N-L_1}),\dots,(N^{N-L_K});((KN-M)^N)\big). \tag{6-1}$$

The coefficient $c\big((N^{L_1}),\dots,(N^{L_K});(M^N)\big)$ is precisely the number of $N\times M$ matrices A which satisfy the conditions given in the Corollary 4.4. We will call such a collection of matrices the set of configuration matrices corresponding to $(L_1,\dots,L_K;N)$. We prove Theorem 6.2 by providing a bijection between the configuration matrices corresponding to the coefficients in (6-1).

Suppose that $c\big((N^{L_1}),\dots,(N^{L_K});(M^N)\big)\neq 0$ and fix a configuration matrix $A=[A_1\mid A_2\mid\cdots\mid A_K]$. For each A_i, we construct an $N\times(N-L_i)$ matrix B_i as follows. Decompose

$$A_i=\sum_{j=1}^{N} C_j$$

as a sum of binary matrices which for all integers y, j, satisfy

(1) $\sum_{x=1}^{N} C_j[x,y]=1.$

(2) $\sum_{x=1}^{N'} (C_j[x,y]-C_j[x,y+1])\geq 0$ for all $N'<N$.

(3) $\sum_{x=1}^{N'} (C_j[x,y]-C_{j+1}[x,y])\geq 0$ for all $N'<N$.

Consider A_2 from Example 4.5. We have

$$\begin{pmatrix}3&0\\0&1\\2&2\\0&2\\0&0\end{pmatrix}=\begin{pmatrix}1&0\\0&1\\0&0\\0&0\\0&0\end{pmatrix}+\begin{pmatrix}1&0\\0&0\\0&1\\0&0\\0&0\end{pmatrix}+\begin{pmatrix}1&0\\0&0\\0&1\\0&0\\0&0\end{pmatrix}+\begin{pmatrix}0&0\\0&0\\1&0\\0&1\\0&0\end{pmatrix}+\begin{pmatrix}0&0\\0&0\\1&0\\0&1\\0&0\end{pmatrix}.$$

It is easy to see that this decomposition of A_i is unique since A_i satisfies properties (i), (iii) and (v). For each C_j, define the $N\times(N-L_i)$ matrix C'_j to be the unique binary matrix which satisfies conditions (1), (2) and that $[C_j\mid C'_j]$ is invertible. For example, if $N = 5$ then

$$C_j = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \rightsquigarrow \quad C'_j = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Define

$$B_i := \sum_{i=1}^{N} C'_j$$

and consider the $N \times (KN - M)$ matrix

$$B := [B_K \mid B_{K-1} \mid \cdots \mid B_1].$$

Note that the binary decomposition of B_i into C'_j also satisfies conditions (1)–(3) if we order the C'_j in reverse. Moreover, if we apply this algorithm to the matrix B, we will recover the matrix A. We now record some important observations on the submatrices A_i and B_i. First, if $x < y$, then

$$A_i[x, y] = B_i[x, y] = 0. \tag{6-2}$$

Second,

$$A_i[x, y] + B_i[x, x - y] = A_i[x + 1, y + 1] + B_i[x + 1, x - y + 1]. \tag{6-3}$$

In the equations above we take $A_i[x, y] = 0$ (respectively, $B_i[x, y] = 0$) if x, y lie outside the boundaries of A_i (respectively, B_i). In the case when $x = y$, we get

$$A_i[x, x] = A_i[x + 1, x + 1] + B_i[x + 1, 1]. \tag{6-4}$$

Theorem 6.2 follows from the proceeding proposition.

Proposition 6.3. *The $N \times (KN - M)$ matrix B is a configuration matrix for the sequence $(N - L_K, \ldots, N - L_1; KN - M)$.*

Proof. The most challenging part of this proof is to show that the matrix B satisfies property (iv). Hence the majority of this argument is dedicated to the proof this property. We first consider the other properties. Properties (i)–(iii) are immediate by construction of B. Property (v) follows from the fact that each B_i is a sum of binary matrices which satisfy conditions (1)–(3). We now prove that B satisfies property (iv) by contradiction. Suppose there exist integers i, l such that

$$\sum_{j=1}^{l} (B[i, j] - B[i + 1, j]) < B[i + 1, l + 1]. \tag{6-5}$$

We define the integers k, l' as follows. Let k denote largest integer for which the partial sum

$$l' := \sum_{j=1}^{k} (N - L_{K-j+1}) \le l.$$

Hence l' is the number of columns of the submatrix $[B_K \mid \cdots \mid B_{K-k+1}]$ of B.

Observe that each row sum of the matrix $[A_j \mid B_j]$ is equal to N. Combining this observation with (6-5) gives

$$\begin{aligned}
&\sum_{j=1}^{l} (B[i, j] - B[i+1, j]) \\
&\qquad = \sum_{j=1}^{l'} (B[i, j] - B[i+1, j]) + \sum_{j=l'+1}^{l} (B[i, j] - B[i+1, j]) \\
&\qquad = \sum_{j=M-(kN-l'-1)}^{M} (A[i+1, j] - A[i, j]) + \sum_{j=l'+1}^{l} (B[i, j] - B[i+1, j]) \\
&\qquad < B[i+1, l+1].
\end{aligned}$$

Rewriting this inequality yields

$$\begin{aligned}
&\sum_{j=M-kN+l'+1}^{M} (A[i+1, j] - A[i, j]) \\
&\qquad\qquad < B[i+1, l+1] - \sum_{j=l'+1}^{l} (B[i, j] - B[i+1, j]) \\
&\qquad\qquad = B[i+1, l'+1] + \sum_{j=l'+1}^{l} (B[i+1, j+1] - B[i, j]).
\end{aligned}$$

The matrix entries of B appearing on the right hand side of the above equation are all contained in the submatrix B_{K-k}. Applying (6-3) and (6-4), we get

(6-6)
$$\sum_{j=M-kN+l'+1}^{M} (A[i+1, j] - A[i, j]) < \sum_{j=0}^{l-l'} (A_{K-k}[i, i-j] - A_{K-k}[i+1, i-j+1]).$$

By (6-2), $A_{K-k}[x, y] = 0$ if $y > x$. Hence we can extend the right hand side of (6-6) to

$$\sum_{j=M-kN+l'+1}^{M} (A[i+1, j] - A[i, j])$$
$$< A_{K-k}[i, i-l+l'] + \sum_{j=0}^{L_{K-k}-(i+1)+(l-l')} (A_{K-k}[i, L_{K-k}-j] - A_{K-k}[i+1, L_{K-k}-j]).$$

Now the fact that A satisfies properties (ii), contradicts the fact that it also satisfies property (iv). This completes the proof. ☐

Example 6.4. Let $N = 4$ and consider the sequence $\mathbf{L} = (2, 2, 2, 1)$. Using Corollary 4.4, the matrix A below implies that $\mathbf{L} \in \mathrm{TFF}(7/4, 4)$.

$$A = \left(\begin{array}{cc|cc|cc|c} 4 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 2 & 2 & 0 \\ 0 & 0 & 0 & 1 & 0 & 2 & 4 \end{array}\right)$$

We get

$$B = \left(\begin{array}{ccc|cc|cc|cc} 4 & 0 & 0 & 4 & 0 & 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 2 & 2 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 1 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 3 & 0 & 4 \end{array}\right)$$

and hence $(3, 2, 2, 2) \in \mathrm{TFF}(9/4, 4)$.

We now give the analogous theorem on combinatorial Naimark duality.

Theorem 6.5. *The Littlewood–Richardson coefficients*

$$c\big((N^{L_1}), \dots, (N^{L_K}); (M^N)\big) = c\big(((M-N)^{L_1}), \dots, ((M-N)^{L_K}); (M^{(M-N)})\big). \tag{6-7}$$

As with Theorem 6.2, we define a bijection between configuration matrices corresponding to the Littlewood–Richardson coefficients in (6-7). Fix a configuration matrix A corresponding to the sequence $(L_1, \dots, L_K; N)$ and consider the Littlewood–Richardson skew tableaux μ^k/μ^{k-1} where μ^k is defined in (4-6). To each μ^k/μ^{k-1} we define the $L_k \times M$ binary matrix T_k by

$$T_k[x, y] := \begin{cases} 1, & \text{if } x \text{ appears in column } y \text{ of } \mu^k/\mu^{k-1} \\ 0, & \text{otherwise.} \end{cases}$$

The partition shape of μ^k can be recovered from the matrices $T_1, \dots T_K$ as follows. Define the matrix $T(k)$ by "stacking" the matrices $T_1, \dots, T_k$ (see Example 6.7

below). In other words,

$$T(k) := \begin{pmatrix} T_1 \\ \hline \vdots \\ \hline T_k \end{pmatrix}.$$

Since A satisfies property (iv), the partition μ^k can be recovered by upward justifying the nonzero entries of $T(k)$. In particular, the entire collection $T_1, \ldots, T_K$ uniquely determines the matrix A.

We now define the "complementary" $L_k \times M$ matrix S_k by

$$S_k[x, y] := 1 - T_k[x, M - y + 1]$$

and $S(k)$ as the corresponding column matrix with block entries $S_1, \ldots, S_k$. It is easy that if the nonzero entries of $S(k)$ are justified upwards, we get the dual partition $(\mu^k)^*$ in rectangle (M^{M_k}) where $M_k := \sum_{i=1}^{k} L_i$. Hence $S_1, \ldots, S_K$ determines some matrix B in the same way that $T_1, \ldots, T_K$ determines A. Also note that we can recover T_k from S_k by applying the complementary operation to S_k. Theorem 6.5 is a consequence of the following proposition.

Proposition 6.6. *The collection $S_1, \ldots, S_K$ determines a configuration matrix for the sequence $(L_1, \ldots, L_K; M - N)$.*

Proof. Let $B = [B_1 \mid \cdots \mid B_K]$ denote the matrix corresponding to the collection $S_1, \ldots, S_K$. We will show that B is a configuration matrix for the sequence $(L_1, \ldots, L_K; M - N)$. In this case, property (v) is the most challenging to prove. Hence most the argument to dedicated to this part of the proof.

First, note that B trivially satisfies property (i). Next, we observe that A satisfies properties (ii) and (iii) if and only if the matrix $T(K)$ has M columns where each column sum is equal to N. Since $S(K)$ has the same number of columns as $T(K)$ with column sums of $M - N$, we get that B also satisfies properties (ii) and (iii). Property (iv) follows from the fact that if we upwards justify the entries of $S(k)$ we get the shape of the dual partition $(\mu^k)^*$.

We now prove that B satisfies property (v) by contradiction. Suppose there exists B_k and integers j, l such that

$$\sum_{i=1}^{l} (B_k[i, j] - B_k[i, j+1]) < B_k[l+1, j+1].$$

This implies there exists an integer l' such that

$$\sum_{i=l'+1}^{M} (S_k[j, i] - S_k[j+1, i]) < 0 \tag{6-8}$$

with

$$S_k[j, l'+1] = 0 \quad \text{and} \quad S_k[j+1, l'+1] = 1. \tag{6-9}$$

Conversely, assume there exists an integer l' such that (6-8) and (6-9) are true. By (6-9), there exists an integer l'' such that

$$\sum_{i=l'+1}^{M} S_k[j,i] = \sum_{i=1}^{l''} B_k[i,j] \quad \text{and} \quad \sum_{i=l'+1}^{M} S_k[j+1,i] \le \sum_{i=1}^{l''+1} B_k[i,j+1].$$

Hence by (6-8),

$$-B_k[l''+1, j+1] + \sum_{i=1}^{l''}(B_k[i,j] - B_k[i,j+1]) \le \sum_{i=l'+1}^{M}(S_k[j,i] - S_k[j+1,i]) < 0.$$

Observe that if (6-8) is true for l, then there is always some integer $l' \le l$ for which both (6-8) and (6-9) are true. Thus the failure of property (v) is equivalent to (6-8). By definition of S_k and (6-8), we have

$$\sum_{i=1}^{M-l'}(T_k[j+1,i] - T_k[j,i]) < 0.$$

Since the row sums of T_k equal N,

$$\sum_{i=M-l'+1}^{M}(T_k[j,i] - T_k[j+1,i]) < 0.$$

Therefore the matrix A also fails to satisfy property (v) which is a contradiction. This completes the proof. □

Example 6.7. Consider $N = 4$ and $\mathbf{L} = (2, 2, 2, 1)$ as in Example 6.4. Then μ^4, as a union of Littlewood–Richardson skew tableaux, is equal to

1	1	1	1	1	1	1
2	2	2	2	2	1	1
1	2	2	1	1	2	2
2	2	2	1	1	1	1

We have

$$T(4)=\begin{pmatrix}1&1&1&1&0&0&0\\1&1&1&1&0&0&0\\\hline 1&0&0&0&1&1&1\\1&1&1&0&1&0&0\\\hline 0&0&0&1&1&1&1\\0&1&1&0&0&1&1\\\hline 0&0&0&1&1&1&1\end{pmatrix}\quad\rightsquigarrow\quad S(4)=\begin{pmatrix}1&1&1&0&0&0&0\\1&1&1&0&0&0&0\\\hline 0&0&0&1&1&1&0\\1&1&0&1&0&0&0\\\hline 0&0&0&0&1&1&1\\0&0&1&1&0&0&1\\\hline 0&0&0&0&1&1&1\end{pmatrix}$$

Upward justifying the nonzero entries of $S(4)$ gives the union of Littlewood–Richardson skew tableaux and corresponding configuration matrix

$$\begin{array}{|c|c|c|c|c|c|c|}\hline 1&1&1&1&1&1&1\\\hline 2&2&2&2&1&1&2\\\hline 2&2&2&2&1&1&1\\\hline\end{array}\qquad B=\left(\begin{array}{cc|cc|cc|c}3&0&3&0&1&0&0\\0&3&0&1&2&1&0\\0&0&0&2&0&2&3\end{array}\right),$$

and hence $(2,2,2,1)\in\mathrm{TFF}(7/3,3)$.

7. Examples and tables of TFF sequences

This section is divided into two parts. In the first part we give several examples of existence of tight fusion frames using skew Littlewood–Richardson tableaux as in Example 4.5. In the second part, we give a complete list of tight fusion frame sequences for $N\le 9$ and $\alpha\le 2$ by listing all maximal elements in the partial order induced by majorization.

Examples of skew Littlewood–Richardson tableaux. The following are some examples of Littlewood–Richardson tableaux in the cases of $N=3,5,7$. Readers who are interested in combinatorial spatial and Naimark duality as discussed in Section 6 are encouraged to apply the bijective constructions to these examples.

$N=3$ and $\mathbf{L}=(3,2,1)$, $\mathbf{L}=(2,1,1,1)$, and $\mathbf{L}=(1,1,1,1)$:

$$\begin{array}{|c|c|c|c|c|c|}\hline 1&1&1&1&1&1\\\hline 2&2&2&2&2&2\\\hline 3&3&3&1&1&1\\\hline\end{array}\qquad\begin{array}{|c|c|c|c|c|}\hline 1&1&1&1&1\\\hline 2&2&2&1&1\\\hline 1&1&1&1&1\\\hline\end{array}\qquad\begin{array}{|c|c|c|c|}\hline 1&1&1&1\\\hline 1&1&1&1\\\hline 1&1&1&1\\\hline\end{array}$$

$N=5$ and $\mathbf{L}=(2,2,2,2)$ and $\mathbf{L}=(3,3,3,3)$:

$$\begin{array}{|c|c|c|c|c|c|c|c|}\hline 1&1&1&1&1&1&1&1\\\hline 2&2&2&2&2&2&1&1\\\hline 1&1&2&2&1&1&2&2\\\hline 2&2&1&1&1&1&1&1\\\hline 2&2&2&2&2&2&2&2\\\hline\end{array}\qquad\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|}\hline 1&1&1&1&1&1&1&1&1&1&1&1\\\hline 2&2&2&2&2&2&2&2&1&1&2&2\\\hline 3&3&3&3&3&3&1&1&1&1&1&1\\\hline 2&2&3&3&2&2&2&2&2&2&2&2\\\hline 3&3&3&3&3&3&3&3&3&3&3&3\\\hline\end{array}$$

$N = 7$ and $\mathbf{L} = (4, 3, 3, 1, 1)$ and $\mathbf{L} = (3, 2, 2, 2, 1)$:

1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	1	1	1
4	4	4	4	4	4	4	1	1	2	2	2
1	1	3	3	3	1	1	2	2	3	3	3
2	2	2	2	3	3	1	1	1	1	1	1
3	3	3	3	1	1	1	1	1	1	1	1

1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	1	1	1
1	1	1	1	1	1	1	2	2	2
2	2	2	2	2	1	1	1	1	1
1	2	2	1	1	2	2	2	2	2
2	2	2	1	1	1	1	1	1	1

Tables of maximal tight fusion frames. At the end of this section we give a complete list of tight fusion frames for $N \leq 9$ and $\alpha \leq 2$ by listing all maximal elements in the partial order induced by majorization. These lists are generated by applying the techniques developed in this paper. In particular, we use the following methods

- Constructing Littlewood–Richardson tableaux as in Corollary 4.4.
- Recursive construction using spatial and Naimark duality.
- Recursive construction using Lemma 7.1.
- Applying inequalities of Theorem 3.3 or 5.4.

The following lemma follows from the Naimark duality.

Lemma 7.1. *Assume that* $L_1 = N(\alpha - 1)$. *Then,* $\mathbf{L} \in \mathrm{TFF}(\alpha, N)$ *if and only if* $\mathbf{L}' \in \mathrm{TFF}(\tilde{\alpha} - 1, N(\alpha - 1))$ *where* $\mathbf{L}' = (L_2 \geq \cdots \geq L_k)$ *and* $1/\alpha + 1/\tilde{\alpha} = 1$.

It is easy to see that maximality under the majorization partial order is preserved under these dualities and the lemma above. Unfortunately, there are several TFF sequences missed by majorization and the recursive generation techniques mentioned above. These sequences were only found by brute force construction of Littlewood–Richardson tableaux. The first maximal tight fusion frame sequence missed by recursion is $(4, 2, 2, 2, 1)$ where $N = 6$. Hence, it might be of interest to illustrate how to construct a tight fusion frame for this sequence.

Example 7.2. Let $N = 6$ and $\mathbf{L} = (4, 2, 2, 2, 1)$. The first step in our construction is identifying a skew Littlewood–Richardson tableaux corresponding to our TFF sequence.

1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	1	1
3	3	3	3	3	3	1	1	1	2	2
4	4	4	4	4	4	1	1	1	1	1
1	2	2	1	2	1	2	2	2	2	2
2	2	2	2	2	1	1	1	1	1	1

The above tableaux shows the existence of projections $P_1, \dots, P_5$ in $\mathbb{R}^6$ with

$$\sum_{i=1}^{5} P_i = \frac{11}{6}\mathbf{I}, \quad \operatorname{rank} P_1 = 4,\ \operatorname{rank} P_2 = \operatorname{rank} P_3 = \operatorname{rank} P_4 = 2,\ \operatorname{rank} P_5 = 1. \tag{7-1}$$

By Corollary 4.6, the tableaux also contains information on the eigenvalues of the intermediate partial sums of projections in (7-1).

sum of projections	eigenvalue list
P_1	$(1, 1, 1, 1, 0, 0)$
$P_1 + P_2$	$\left(\frac{11}{6}, \frac{9}{6}, 1, 1, \frac{3}{6}, \frac{1}{6}\right)$
$P_1 + P_2 + P_3$	$\left(\frac{11}{6}, \frac{11}{6}, \frac{11}{6}, 1, \frac{5}{6}, \frac{4}{6}\right)$
$P_1 + \dots + P_4$	$\left(\frac{11}{6}, \frac{11}{6}, \frac{11}{6}, \frac{11}{6}, \frac{11}{6}, \frac{5}{6}\right)$
$P_1 + \dots + P_5$	$\left(\frac{11}{6}, \frac{11}{6}, \frac{11}{6}, \frac{11}{6}, \frac{11}{6}, \frac{11}{6}\right)$

Equipped with this information and a symbolic computation program such as Mathematica we can construct an explicit tight fusion frame in $\mathbb{R}^6$ associated with the sequence $(4, 2, 2, 2, 1)$. The matrix below shows an orthonormal basis (column) vectors for the corresponding ranges of projections P_i, $i = 1, \dots, 5$.

$$\left(\begin{array}{cccc|cc|cc|cc|c}
1 & 0 & 0 & 0 & \frac{5}{6} & 0 & -\sqrt{\frac{5}{72}} & 0 & \sqrt{\frac{5}{72}} & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2\sqrt{2}} & -\frac{1}{3} & -\frac{1}{2\sqrt{2}} & \frac{1}{3} & \frac{1}{3} \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & \frac{\sqrt{5}}{3} & 0 & \frac{\sqrt{5}}{6} & \frac{\sqrt{5}}{6} \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \sqrt{\frac{5}{12}} & -\sqrt{\frac{5}{12}} \\
0 & 0 & 0 & 0 & 0 & -\frac{\sqrt{3}}{2} & \frac{1}{2\sqrt{6}} & -\frac{1}{\sqrt{3}} & \frac{1}{2\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\
0 & 0 & 0 & 0 & -\frac{\sqrt{11}}{6} & 0 & -\sqrt{\frac{55}{72}} & 0 & \sqrt{\frac{55}{72}} & 0 & 0
\end{array}\right)$$

A direct calculation shows that: (i) columns are orthonormal to each other in every block, and (ii) rows are orthogonal with norms $\sqrt{11/6}$. This proves the existence of a TFF (7-1).

It is worth noting that the Example 7.2 can not be obtained using the spectral tetris construction (STC). The STC has been recently introduced by Casazza et al. [2012] who gave an algorithmic way of constructing sparse fusion frames. Among other things, the authors have shown that the ranks $\mathbf{L}$ of spectral tetris fusion frames must necessarily satisfy $\mathbf{L} \preccurlyeq \mathbf{L}'$, where $\mathbf{L}'$ is a sequence of ranks of the reference fusion frame.

Moreover, in the tight case this condition is also sufficient, and hence [Casazza et al. 2012, Theorem 3.3] characterizes possible ranks obtained by the STC in the case when the frame bound $\alpha \geq 2$. Combining this with Naimark complements, see Theorem 2.6, this yields TFFs also in the case $1 < \alpha < 2$. In particular, we have $\mathrm{TFF}(11/6, 6) = \mathrm{TFF}(11/5, 5)$. A direct calculation of the reference fusion frame corresponding to eigenvalues (11/5, 11/5, 11/5, 11/5, 11/5) yields a TFF sequence $(3, 3, 3, 2)$. This happens to be another maximal element of TFF(11/6, 6) which is not comparable with (4, 2, 2, 2, 1) with respect to the majorization relation $\preccurlyeq$. Hence, the above example can not be obtained by the STC even when paired with Naimark duality.

List of maximal TFF sequences for $N \leq 9$ and $\alpha \leq 2$.

$N=3$	
α	max elements
1	(3)
4/3	(1, 1, 1, 1)
5/3	(2, 1, 1, 1)
2	(3, 3)

$N=4$	
α	max elements
1	(4)
5/4	(1, 1, 1, 1, 1)
6/4	(2, 2, 2)
7/4	(3, 1, 1, 1, 1), (2, 2, 2, 1)
2	(4, 4)

$N=5$	
α	max elements
1	(5)
6/5	(1, 1, 1, 1, 1, 1)
7/5	(2, 2, 1, 1, 1)
8/5	(3, 2, 1, 1, 1), (2, 2, 2, 2)
9/5	(4, 1, 1, 1, 1, 1), (3, 2, 2, 2)
2	(5, 5)

$N=6$	
α	max elements
1	(6)
7/6	(1, 1, 1, 1, 1, 1, 1)
8/6	(2, 2, 2, 2)
9/6	(3, 3, 3)
10/6	(4, 2, 2, 2)
11/6	(5, 1, 1, 1, 1, 1, 1), (4, 2, 2, 2, 1), (3, 3, 3, 2)
2	(6, 6)

$N = 7$	
α	max elements
1	(7)
8/7	(1, 1, 1, 1, 1, 1, 1, 1)
9/7	(2, 2, 2, 1, 1, 1)
10/7	(3, 3, 1, 1, 1, 1), (3, 2, 2, 2, 1)
11/7	(4, 3, 1, 1, 1, 1), (4, 2, 2, 2, 1)
12/7	(5, 2, 2, 1, 1, 1), (4, 3, 3, 1, 1), (3, 3, 3, 3)
13/7	(6, 1, 1, 1, 1, 1, 1, 1), (5, 2, 2, 2, 2), (4, 3, 3, 3)
2	(7, 7)

$N = 8$	
α	max elements
1	(8)
9/8	(1, 1, 1, 1, 1, 1, 1, 1, 1)
10/8	(2, 2, 2, 2, 2)
11/8	(3, 2, 2, 2, 2), (3, 3, 2, 1, 1, 1)
12/8	(4, 4, 4)
13/8	(5, 3, 2, 1, 1, 1), (5, 2, 2, 2, 2), (4, 4, 2, 2, 1)
14/8	(6, 2, 2, 2, 2), (5, 3, 3, 2, 1), (4, 4, 4, 2)
15/8	(7, 1, 1, 1, 1, 1, 1, 1), (6, 2, 2, 2, 2, 1), (5, 3, 3, 2, 2), (4, 4, 4, 3)
2	(8, 8)

$N = 9$	
α	max elements
1	(9)
10/9	(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
11/9	(2, 2, 2, 2, 1, 1, 1)
12/9	(3, 3, 3, 3)
13/9	(4, 4, 1, 1, 1, 1, 1), (4, 3, 2, 2, 2), (3, 3, 3, 3, 1)
14/9	(5, 4, 1, 1, 1, 1, 1), (5, 3, 2, 2, 2), (4, 3, 3, 3, 1)
15/9	(6, 3, 3, 3)
16/9	(7, 2, 2, 2, 1, 1, 1), (6, 3, 3, 3, 1), (5, 4, 4, 2, 1), (4, 4, 4, 4)
17/9	(8, 1, 1, 1, 1, 1, 1, 1, 1, 1), (7, 2, 2, 2, 2, 2), (6, 3, 3, 3, 2), (5, 4, 4, 4)
2	(9, 9)

The following sequences are missed by the recursive construction of spatial and Naimark duality and Lemma 7.1. These sequences were obtained applying Theorem 3.3 and constructing Littlewood–Richardson tableaux.

N	TFF sequences
6	(4, 2, 2, 2, 1)
7	(4, 3, 3, 1, 1), (5, 2, 2, 2, 2)
8	(5, 3, 3, 2, 1), (6, 2, 2, 2, 2, 1), (5, 3, 3, 2, 2)
9	(6, 3, 3, 3, 1), (5, 4, 4, 2, 1), (7, 2, 2, 2, 2, 2), (6, 3, 3, 3, 2), (5, 4, 4, 4)

We also remark that the sequences (4, 4, 2, 2, 1) for $N = 8$ and (4, 3, 3, 3, 1) for $N = 9$ can be recursively found from the "missed" sequences (4, 3, 3, 1, 1) for $N = 7$ and (4, 2, 2, 2, 1) for $N = 6$, respectively. Finally, among the eleven maximal TFF sequences listed above only the sequence (5, 4, 4, 4) for $N = 9$ can be obtained using the spectral tetris construction (STC) paired with the Naimark duality. A construction of explicit tight fusion frames corresponding to the remaining ten maximal TFF sequences requires a procedure similar to Example 7.2.

Acknowledgments

Authors would like to thank Stephanie van Willigenburg for useful discussions. Bownik was partially supported by NSF grant DMS-1265711. Luoto and Richmond were partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

[Andrews 1976] G. E. Andrews, *The theory of partitions*, Encyclopedia of Mathematics and its Applications **2**, Addison-Wesley, Reading, MA, 1976. MR 58 #27738 Zbl 0371.10001

[Antezana et al. 2007] J. Antezana, P. G. Massey, M. A. Ruiz, and D. Stojanoff, "The Schur–Horn theorem for operators and frames with prescribed norms and frame operator", *Illinois J. Math.* **51**:2 (2007), 537–560. MR 2009g:42049 Zbl 1137.42008

[Bodmann 2007] B. G. Bodmann, "Optimal linear transmission by loss-insensitive packet encoding", *Appl. Comput. Harmon. Anal.* **22**:3 (2007), 274–285. MR 2008c:42031 Zbl 1193.42113

[Boufounos et al. 2011] P. Boufounos, G. Kutyniok, and H. Rauhut, "Sparse recovery from combined fusion frame measurements", *IEEE Trans. Inform. Theory* **57**:6 (2011), 3864–3876. MR 2012e:94032

[Calderbank et al. 2011] R. Calderbank, P. G. Casazza, A. Heinecke, G. Kutyniok, and A. Pezeshki, "Sparse fusion frames: existence and construction", *Adv. Comput. Math.* **35**:1 (2011), 1–31. MR 2012d:42067 Zbl 1264.94042

[Casazza and Fickus 2009] P. G. Casazza and M. Fickus, "Minimizing fusion frame potential", *Acta Appl. Math.* **107**:1-3 (2009), 7–24. MR 2010e:42042 Zbl 1175.42010

[Casazza and Kutyniok 2004] P. G. Casazza and G. Kutyniok, "Frames of subspaces", pp. 87–113 in *Wavelets, frames and operator theory* (College Park, MD, 2003), edited by C. Heil et al., Contemporary Mathematics **345**, American Mathematical Society, Providence, RI, 2004. MR 2005e:42090 Zbl 1058.42019

[Casazza and Kutyniok 2013] P. G. Casazza and G. Kutyniok, "Fusion frames", pp. 437–477 in *Finite frames: theory and applications*, edited by P. G. Casazza and G. Kutyniok, Birkhäuser, New York, 2013. MR 2964018 Zbl 1261.42048

[Casazza et al. 2008] P. G. Casazza, G. Kutyniok, and S. Li, "Fusion frames and distributed processing", *Appl. Comput. Harmon. Anal.* **25**:1 (2008), 114–132. MR 2009d:42094 Zbl 1258.42029

[Casazza et al. 2011] P. G. Casazza, M. Fickus, D. G. Mixon, Y. Wang, and Z. Zhou, "Constructing tight fusion frames", *Appl. Comput. Harmon. Anal.* **30**:2 (2011), 175–187. MR 2012c:42069 Zbl 1221.42052

[Casazza et al. 2012] P. G. Casazza, M. Fickus, A. Heinecke, Y. Wang, and Z. Zhou, "Spectral tetris fusion frame constructions", *J. Fourier Anal. Appl.* **18**:4 (2012), 828–851. MR 2984371 Zbl 1257.42042

[Chebira et al. 2011] A. Chebira, M. Fickus, and D. G. Mixon, "Filter bank fusion frames", *IEEE Trans. Signal Process.* **59**:3 (2011), 953–963. MR 2012d:94015

[Fillmore 1969] P. A. Fillmore, "On sums of projections", *J. Funct. Anal.* **4**:1 (1969), 146–152. MR 39 #7455 Zbl 0176.43404

[Fong and Murphy 1985] C. K. Fong and G. J. Murphy, "Averages of projections", *J. Operator Theory* **13**:2 (1985), 219–225. MR 86h:47071 Zbl 0614.47011

[Fulton 1997] W. Fulton, *Young tableaux: with applications to representation theory and geometry*, London Mathematical Society Student Texts **35**, Cambridge University Press, 1997. MR 99f:05119 Zbl 0878.14034

[Fulton 2000] W. Fulton, "Eigenvalues, invariant factors, highest weights, and Schubert calculus", *Bull. Amer. Math. Soc.* (*N.S.*) **37**:3 (2000), 209–249. MR 2001g:15023 Zbl 0994.15021

[Halpern et al. 2013] H. Halpern, V. Kaftal, P. W. Ng, and S. Zhang, "Finite sums of projections in von Neumann algebras", *Trans. Amer. Math. Soc.* **365**:5 (2013), 2409–2445. MR 3020103 Zbl 1275.47088

[Horn 1962] A. Horn, "Eigenvalues of sums of Hermitian matrices", *Pacific J. Math.* **12** (1962), 225–241. MR 25 #3941 Zbl 0112.01501

[Kaftal and Weiss 2010] V. Kaftal and G. Weiss, "An infinite dimensional Schur–Horn theorem and majorization theory", *J. Funct. Anal.* **259**:12 (2010), 3115–3162. MR 2011k:47030 Zbl 1202.15035

[Kaftal et al. 2011] V. Kaftal, P. W. Ng, and S. Zhang, "Positive combinations and sums of projections in purely infinite simple $\mathbb{C}^*$-algebras and their multiplier algebras", *Proc. Amer. Math. Soc.* **139**:8 (2011), 2735–2746. MR 2012f:46100 Zbl 1230.46049

[Klyachko 1998] A. A. Klyachko, "Stable bundles, representation theory and Hermitian operators", *Selecta Math.* (*N.S.*) **4**:3 (1998), 419–445. MR 2000b:14054 Zbl 0915.14010

[Knutson and Tao 1999] A. Knutson and T. Tao, "The honeycomb model of $\mathrm{GL}_n(\mathbb{C})$ tensor products, I: Proof of the saturation conjecture", *J. Amer. Math. Soc.* **12**:4 (1999), 1055–1090. MR 2000c:20066 Zbl 0944.05097

[Knutson and Tao 2001] A. Knutson and T. Tao, "Honeycombs and sums of Hermitian matrices", *Notices Amer. Math. Soc.* **48**:2 (2001), 175–186. MR 2002g:15020 Zbl 1047.15006

[Knutson et al. 2004] A. Knutson, T. Tao, and C. Woodward, "The honeycomb model of $\mathrm{GL}_n(\mathbb{C})$ tensor products, II: Puzzles determine facets of the Littlewood–Richardson cone", *J. Amer. Math. Soc.* **17**:1 (2004), 19–48. MR 2005f:14105 Zbl 1043.05111

[Kruglyak et al. 2002] S. A. Kruglyak, V. I. Rabanovich, and Y. S. Samoĭlenko, "О суммах проекторов", *Funkts. Anal. Prilozh.* **36**:3 (2002), 20–35. Translated as "On sums of projections" in *Funct. Anal. Appl.* **36**:3 (2002), 182–195. MR 2004e:47021 Zbl 1038.47001

[Kruglyak et al. 2003] S. A. Kruglyak, V. I. Rabanovich, and Y. S. Samoĭlenko, "Decomposition of a scalar matrix into a sum of orthogonal projections", *Linear Algebra Appl.* **370** (2003), 217–225. MR 2004f:15045 Zbl 1028.15010

[Kutyniok et al. 2009] G. Kutyniok, A. Pezeshki, R. Calderbank, and T. Liu, "Robust dimension reduction, fusion frames, and Grassmannian packings", *Appl. Comput. Harmon. Anal.* **26**:1 (2009), 64–76. MR 2010a:94023 Zbl 1283.42046

[Massey et al. 2010] P. G. Massey, M. A. Ruiz, and D. Stojanoff, "The structure of minimizers of the frame potential on fusion frames", *J. Fourier Anal. Appl.* **16**:4 (2010), 514–543. MR 2011f:42038 Zbl 1194.42039

[Okada 1998] S. Okada, "Applications of minor summation formulas to rectangular-shaped representations of classical groups", *J. Algebra* **205**:2 (1998), 337–367. MR 99g:20081 Zbl 0915.20023

Received October 7, 2013. Revised September 27, 2014.

MARCIN BOWNIK
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OREGON
EUGENE, OR 97403
UNITED STATES
mbownik@uoregon.edu

KURT LUOTO
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER BC V6T 1Z2
CANADA
kwluoto@math.ubc.ca

EDWARD RICHMOND
DEPARTMENT OF MATHEMATICS
OKLAHOMA STATE UNIVERSITY
STILLWATER, OK 74078
UNITED STATES
edward.richmond@okstate.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 275, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.275.295

COMBINATORICS OF FINITE ABELIAN GROUPS AND WEIL REPRESENTATIONS

KUNAL DUTTA AND AMRITANSHU PRASAD

The Weil representation of the symplectic group associated to a finite abelian group of odd order is shown to have a multiplicity-free decomposition. When the abelian group is p-primary, the irreducible representations occurring in the Weil representation are parametrized by a partially ordered set which is independent of p. As p varies, the dimension of the irreducible representation corresponding to each parameter is shown to be a polynomial in p which is calculated explicitly. The commuting algebra of the Weil representation has a basis indexed by another partially ordered set which is independent of p. The expansions of the projection operators onto the irreducible invariant subspaces in terms of this basis are calculated. The coefficients are again polynomials in p. These results remain valid in the more general setting of finitely generated torsion modules over a Dedekind domain.

1. Introduction

1A. ***Overview.*** Heisenberg groups were introduced by Weyl [1949, Chapter 4] in his mathematical formulation of quantum kinematics. Best known among them are the Lie groups whose Lie algebras are spanned by position and momentum operators which satisfy Heisenberg's commutation relations. Weyl also considered Heisenberg groups which are finite modulo their centers, such as the Pauli group (generated by the Pauli matrices), which he used to characterize the kinematics of electron spin.

A fundamental property of Heisenberg groups, predicted by Weyl and proved by Stone [1930] and von Neumann [1931] for real Heisenberg groups is known as the Stone–von Neumann theorem. Mackey [1949] extended this theorem to locally compact Heisenberg groups (see Section 1C for the case that is pertinent to this paper, and [Prasad 2011] for a more detailed and general exposition). By considering Heisenberg groups associated to finite fields, local fields and adèles, Weil [1964] demonstrated the importance of Heisenberg groups in number theory.

Dutta was a student at the Institute of Mathematical Sciences, Chennai when this paper was written. He thanks the Institute for its hospitality and support.

MSC2010: primary 05E10, 11F27; secondary 81R05.

Keywords: Weil representation, Heisenberg group, Clifford group, finite abelian group.

Weil exploited the Stone–von Neumann–Mackey theorem to construct a projective representation of a group of automorphisms of the Heisenberg group, now commonly known as the Weil representation. Along with parabolic induction and the technique of Deligne and Lusztig [1976] using l-adic cohomology, the Weil representation is one of the most important techniques for constructing representations of reductive groups over finite fields (see [Gérardin 1977; Srinivasan 1979]) or local fields (see [Gérardin 1975] and Mœglin, Vignéras and Waldspurger [Mœglin et al. 1987]).

Tanaka [1967a; 1967b] showed how the Weil representation can be used to construct all the irreducible representations of $\mathrm{SL}_2(\mathbb{Z}/p^k\mathbb{Z})$ for odd p by looking at Weil representations associated to the abelian groups $\mathbb{Z}/p^k\mathbb{Z} \oplus \mathbb{Z}/p^l\mathbb{Z}$ for $l \leq k$. However, most of the literature on Weil representations associated to finite abelian groups has focused on vector spaces over finite fields and on constructing representations of classical groups over finite fields.

The representation theory of groups over finite principal ideal local rings was initiated by Kloosterman [1946], who studied $\mathrm{SL}_2(\mathbb{Z}/p^k\mathbb{Z})$. In contrast to general linear groups over finite fields, whose character theory was worked out by Green [1955], the representation theory for general linear groups over these rings is quite hard. It has been shown (by Aubert, Onn, Prasad and Stasinski [Aubert et al. 2010] and Singla [2010]) that this problem is intricately related to the problem of understanding the representations of automorphism groups of finitely generated torsion modules over discrete valuation rings. However, explicit constructions have been available either for a very small class of representations [Hill 1995a; 1995b] or for a very small class of groups [Onn 2008; Stasinski 2009; Singla 2010].

This article concerns the decomposition of the Weil representation of the full symplectic group associated to a finite abelian group of odd order (and more generally, a finite module of odd order over a Dedekind domain) into irreducible representations. When the module in question is elementary (e.g., $(\mathbb{Z}/p\mathbb{Z})^n$ for some odd prime p), it is well-known that the Weil representation, which may be realized on the space of functions on the abelian group, breaks up into two irreducible subspaces consisting of even and odd functions. Besides this, only the case where all the invariant factors are equal (e.g., $(\mathbb{Z}/p^k\mathbb{Z})^n$) has been understood completely (see [Prasad 1998, Theorem 2] for the case where k is even, and Cliff, McNeilly and Szechtman [Cliff et al. 2000] for the general case). A small part of the decomposition has been explained in the general case by Cliff, McNeilly and Szechtman [Cliff et al. 2003]. Maktouf and Torasso [2014; 2012] have shown that the restriction of the Weil representation of a symplectic group over a p-adic field to a maximal compact subgroup or to a maximal elliptic torus is multiplicity-free and have given an explicit description of the irreducible subrepresentations.

In this paper, we describe all the invariant subspaces for the Weil representation for all finite modules of odd order over a Dedekind domain. To be specific, it is shown that the Weil representation has a multiplicity-free decomposition (Theorem 5.5). When the underlying finitely generated torsion module is primary of type λ for some partition λ (see (4.1) and (10.1)), the irreducible components are parametrized by elements of a partially ordered set which depends only on λ, and not on the underlying ring. As the local ring varies, for a fixed element of this partially ordered set, the dimension of the corresponding representation is shown to be a polynomial in the order of its residue field whose coefficients do not depend on the ring (Theorem 9.3). These polynomials are computed explicitly (Theorem 9.12). The centralizer algebra of the Weil representation also has a combinatorial basis indexed by a partially ordered set which depends only on λ and not on the underlying ring. The projection operators onto the irreducible invariant subspaces, when expressed in terms of this basis, are also shown to have coefficients which are polynomials in the order of the residue field whose coefficients also do not depend on the ring (Theorem 9.17), and these polynomials are computed explicitly (Theorems 9.18 and 9.19). Thus the decomposition of the Weil representation into irreducible invariant subspaces is, despite its apparent complexity, combinatorial in nature.

The results in this paper could serve as a starting point from which more subtle constructions involving the Weil representation (such as Howe duality) which have worked so well in the case of classical groups over finite fields can be extended to groups of automorphisms of finitely generated torsion modules over a discrete valuation ring.

It is worth noting that every Heisenberg group that is finite modulo its center is isomorphic to one of the groups considered here (for the precise statement, see Prasad, Shapiro and Vemuri [Prasad et al. 2010], particularly Section 3 and Corollary 5.7). For example, the seemingly different Heisenberg groups used by Tanaka [1967a] to construct representations in the principal series and cuspidal series of finite SL_2 are isomorphic. The difference lies in the realization of the special linear group as a group of automorphisms. The decomposition of any Weil representation associated to a finite abelian group will therefore always be a refinement of one of the decompositions described in this paper.

In order to concentrate on the important ideas without being distracted by technicalities, the main body of this paper uses the setting of finite abelian groups. Section 10 explains how to carry over the results to finitely generated torsion modules over discrete valuation rings and even more generally, finite modules over Dedekind domains.

To obtain our results, we use the combinatorial theory of orbits in finite abelian groups developed in [Dutta and Prasad 2011] (the relevant part is recalled in Section 4A), well-known basic facts about Heisenberg groups and Weil representations which are recalled in Section 1C (of which simple proofs can be found in

[Prasad 2009]), and the standard combinatorial theory of partially ordered sets, as set out in Chapter 3 of the book [Stanley 1997].

1B. *Structure of the paper.* In Section 1C, we recall the definition of the Heisenberg group and its Schrödinger representation. Following Weil [1964], we deduce the existence of the Weil representation from the irreducibility and uniqueness of the Schrödinger representation. Section 1D contains a precise formulation of our main problem — the decomposition of the Weil representation associated to a finite abelian group into irreducible summands.

In Section 2, we use the primary decomposition of finite abelian groups to reduce the main problem to the case of primary finite abelian groups. In Section 3 we explain the relationship between the multiplicities of the summands in the decomposition of the Weil representation and the number of orbits for the action of the symplectic group on the quotient of the Heisenberg group by its center.

In Section 4A, we recall the combinatorial theory of orbits and characteristic subgroups in a finite abelian group developed in [Dutta and Prasad 2011]. An important order-reversing involution on the lattice of characteristic subgroups is introduced in Section 4B. The theory from that paper is extended to symplectic orbits on the quotient of the Heisenberg group modulo its center in Section 4C.

Section 5 contains the first major theorem of this article, namely that the decomposition of the Weil representation associated to a finite abelian group into simple representations is multiplicity-free (Theorem 5.5). This is achieved by computing the structure constants of its endomorphism algebra to show that this algebra is commutative (Lemma 5.3). It follows that the set of invariant subspaces of the Weil representation, partially ordered by inclusion, forms a Boolean lattice (Corollary 5.6).

The task of describing the irreducible components of the Weil representation is carried out using combinatorial analysis in Sections 6–9. In Section 6 two elementary types of invariant subspaces of the Weil representation are identified. The first type are the subspaces of $L^2(A)$ consisting of even and odd functions; the second type are associated to so-called small order ideals (these subspaces are far from being mutually disjoint and irreducible). In Section 7, we describe a tensor product decomposition of the invariant subspaces associated to small order ideals. In Section 8, we refine the invariant subspaces of Section 6 to construct a family of invariant subspaces, which as a poset under inclusion is described in terms of a combinatorially defined poset Q_λ. In Section 9 the irreducible subrepresentations of the Weil representations are extracted from the invariant subspaces of Section 8 (Theorem 9.3). The rest of Section 9 is devoted to the explicit computation of the dimensions of these subrepresentations as well as formulae for the orthogonal projections onto them in terms of a natural basis for the endomorphism algebra of the Weil representation.

Finally, in Section 10 we explain how to extend the ideas of this paper to analyze the Weil representation associated to any finite module over a Dedekind domain.

1C. ***Basic definitions.*** Let A be a finite abelian group of odd order. Let $\hat{A}$ denote the Pontryagin dual of A. This is the group of all homomorphisms $A \to U(1)$, where $U(1)$ denotes the group of unit complex numbers. Let $K = A \times \hat{A}$. For each $k = (x, \chi) \in K$, the unitary operator on $L^2(A)$ defined by

$$W_k f(u) = \chi(u - x/2) f(u - x) \quad \text{for all } f \in L^2(A),\ u \in A$$

is called a Weyl operator. These operators satisfy

$$W_k W_l = c(k, l) W_{k+l} \quad \text{for all } k, l \in K,$$

where, if $k = (x, \chi)$ and $l = (y, \lambda)$, then

$$c(k, l) = \chi(y/2)\lambda(x/2)^{-1}.$$

Observe that $c(k, l)$ is bimultiplicative; for example, $c(k, l + l') = c(k, l)c(k, l')$ for all $k, l, l' \in K$.

The subgroup

$$H = \{cW_k \mid c \in U(1), k \in K\}$$

of the group of unitary operators on $L^2(A)$ is called the Heisenberg group associated to A. This group is known to physicists as a generalized Pauli group or a Weyl–Heisenberg group. As defined here, it comes with a unitary representation on $L^2(A)$, called the Schrödinger representation. Mackey's generalization [1949, Theorem 1] of the Stone–von Neumann theorem applies:

Theorem 1.1. *The Schrödinger representation of H is irreducible. Let $U(\mathcal{H})$ be the group of unitary operators on a Hilbert space $\mathcal{H}$, and let $\rho : H \to U(\mathcal{H})$ be an irreducible unitary representation of H such that $\rho(cW_0) = c\,\mathrm{Id}_{\mathcal{H}}$ for every $c \in U(1)$. Then there exists, up to scaling, a unique isometry $W : L^2(A) \to \mathcal{H}$ such that*

$$WW_k = \rho(W_k)W \quad \textit{for all } k \in K.$$

If g is an automorphism of K such that

$$c(gk, gl) = c(k, l) \quad \text{for all } k, l \in K, \tag{1.2}$$

then $\rho_g : H \to U(L^2(A))$ defined by

$$\rho_g(cW_k) = cW_{g(k)} \quad \text{for all } c \in U(1),\ k \in K$$

is an irreducible unitary representation of H on $L^2(A)$ such that $\rho(cW_0) = c\,\mathrm{Id}_{L^2(A)}$. By Theorem 1.1, there exists a unitary operator W_g on $L^2(A)$ such that

$$W_g W_k = W_{g(k)} W_g \quad \text{for all } k \in K.$$

Writing W_g^* for the adjoint of the unitary operator W_g, we have:

$$W_g W_k W_g^* = W_{g(k)} \quad \text{for all } k \in K. \tag{1.3}$$

If g_1 and g_2 are two such automorphisms, both $W_{g_1 g_2}$ and $W_{g_1} W_{g_2}$ intertwine the Schrödinger representation with $\rho_{g_1 g_2}$, and hence must differ by a unitary scalar:

$$W_{g_1} W_{g_2} = c(g_1, g_2) W_{g_1 g_2} \quad \text{for some } c(g_1, g_2) \in U(1).$$

Let $\mathrm{Sp}(K)$ be the group of all automorphisms g of K which satisfy (1.2). We have shown that $g \mapsto W_g$ is a projective representation of $\mathrm{Sp}(K)$ on $L^2(A)$. This representation is known as the Weil representation.

Remark 1.4. The operators W_g, for $g \in \mathrm{Sp}(K)$ can be normalized in such a way that $c(g_1, g_2) = 1$ for all g_1, g_2 (see Remark 6.6). Thus the Weil representation can be taken to be an ordinary representation of $\mathrm{Sp}(K)$.

Remark 1.5. The overlap of notation between the Weyl operators and the Weil representation is suggested by (1.3), which implies that they can be combined to construct a representation of $H \rtimes \mathrm{Sp}(K)$. The operators in this representation are precisely the unitary operators which normalize H. The resulting group is sometimes known as a Clifford group or a Jacobi group. It plays a prominent role in the stabilizer formalism for quantum error-correcting codes (see Chapter X of [Nielsen and Chuang 2000]).

1D. ***Formulation of the problem.*** We investigate the decomposition

$$L^2(A) = \bigoplus_{\pi \in \mathrm{Sp}(K)^\wedge} m_\pi \mathcal{H}_\pi \tag{1.6}$$

into irreducible representations. Here $\mathrm{Sp}(K)^\wedge$ denotes the set of equivalence classes of irreducible unitary representations of $\mathrm{Sp}(K)$ and, for each $\pi : \mathrm{Sp}(K) \to U(\mathcal{H}_\pi)$ in $\mathrm{Sp}(K)^\wedge$, m_π denotes the multiplicity of π in the Weil representation. Although the Weil representation is defined only up to multiplication by a scalar representation, the multiplicities and dimensions of the irreducible representations occurring in the decomposition are invariant under such twists (see Remark 2.2). As explained in Section 1A, the outcome of this paper is an understanding of this decomposition.

2. Product decompositions

We shall recall and apply a well-known observation on Weil representations associated to a product of abelian groups (see [Gérardin 1977, Corollary 2.5]).

2A. ***Projective equivalence.*** Since Weil representations are defined only up to scalar factors, we use a definition of equivalence of representations that is weaker than unitary equivalence:

Definition 2.1 (Projective equivalence). Let G be a group and $\rho_i : G \to U(\mathcal{H}_i)$ for $i = 1, 2$ be two unitary representations of G. We say that ρ_1 and ρ_2 are projectively equivalent if there exists a homomorphism $\chi : G \to U(1)$ such that ρ_2 is unitarily equivalent to $\rho_1 \otimes \chi$.

Remark 2.2. If, as a representation of G,

$$\mathcal{H}_i = \bigoplus_{\pi \in \hat{G}} m_\pi^{(i)} \mathcal{H}_\pi$$

is the decomposition of $\mathcal{H}_i$ into irreducibles for representations as in Definition 2.1, then $m_{\pi\otimes\chi}^{(2)} = m_\pi^{(1)}$, so there is a bijection between the sets of irreducible representations of G that appear in $\mathcal{H}_1$ and $\mathcal{H}_2$ which preserves multiplicities and dimensions.

2B. ***Tensor product decomposition.*** If A admits a product decomposition $A = A' \times A''$, then

$$L^2(A) = L^2(A') \otimes L^2(A''). \tag{2.3}$$

Let $K' = A' \times (A')^\wedge$ and $K'' = A'' \times (A'')^\wedge$. Thus $K = K' \times K''$. Let S' and S'' be subgroups of $\mathrm{Sp}(K')$ and $\mathrm{Sp}(K'')$ respectively. Then $S = S' \times S''$ is a subgroup of $\mathrm{Sp}(K)$.

Theorem 2.4. *The Weil representation of S on $L^2(A)$ is projectively equivalent to the tensor product of the Weil representation of S' on $L^2(A')$ and the Weil representation of S'' on $L^2(A'')$.*

Proof. By (1.3), the Weil representations of S' and S'' satisfy

$$W_{g'} W_{k'} W_{g'}^* = W_{g'(k')} \quad \text{and} \quad W_{g''} W_{k''} W_{g''}^* = W_{g''(k'')}$$

for all $g' \in S'$, $g'' \in S''$, $k' \in K'$ and $k'' \in K''$, whence

$$(W_{g'} \otimes W_{g''})(W_{k'} \otimes W_{k''})(W_{g'} \otimes W_{g''})^* = W_{g'(k')} \otimes W_{g''(k'')}.$$

Since $W_{k'} \otimes W_{k''}$ coincides with $W_{(k',k'')}$ under the isomorphism (2.3), $W_{g'} \otimes W_{g''}$ satisfies the defining identity (1.3) for the Weil representation of S on $L^2(A)$. □

2C. ***Primary decomposition.*** A finite abelian group has a primary decomposition

$$A = \prod_{p \text{ prime}} A_p,$$

where A_p is the subgroup of elements of A annihilated by some power of p. Writing K_p for $A_p \times (A_p)^\wedge$,

$$K = \prod_p K_p \quad \text{and} \quad \mathrm{Sp}(K) = \prod_p \mathrm{Sp}(K_p).$$

Theorem 2.4, when applied to the primary decomposition, gives:

Corollary 2.5. *The Weil representation of* $\mathrm{Sp}(K)$ *on* $L^2(A)$ *is projectively equivalent to the tensor product over those primes* p *for which* $A_p \neq 0$ *of the Weil representations of* $\mathrm{Sp}(K_p)$ *on* $L^2(A_p)$.

In view of Corollary 2.5, it suffices to consider the case where A is a finite abelian p-group for some odd prime p.

3. Multiplicities and orbits

We now recall the relation between the decomposition of the Weil representation and orbits in K [Prasad 2009].

3A. *An orthonormal basis.*

Lemma 3.1. *The set* $\{W_k \mid k \in K\}$ *of Weyl operators is an orthonormal basis of* $\mathrm{End}_{\mathbb{C}}\, L^2(A)$.

Proof. For each $k \in K$ and $T \in \mathrm{End}_{\mathbb{C}}\, L^2(A)$, let

$$\tau(k)T = W_k T W_k^*.$$

Then $k \mapsto \tau(k)$ is a unitary representation of K on $\mathrm{End}_{\mathbb{C}}\, L^2(A)$. If $k = (x, \chi)$ and $l = (y, \lambda)$ are two elements of K, then

$$\begin{aligned}\tau(k)W_l &= W_k W_l W_k^* \\ &= W_k W_l (W_l W_k)^* W_l \\ &= c(k,l) W_{k+l} c(l,k)^{-1} W_{l+k}^* W_l \\ &= \chi(y)\lambda(x)^{-1} W_l.\end{aligned}$$

Thus the W_l are eigenvectors for the action of K with distinct eigencharacters. Therefore they form an orthonormal set of operators. Since $|K| = |A|^2 = \dim \mathrm{End}_{\mathbb{C}}\, L^2(A)$, this orthonormal set is a basis. □

3B. *Endomorphisms.*

By Lemma 3.1, every $T \in \mathrm{End}_{\mathbb{C}}\, L^2(A)$ has a unique expansion

$$T = \sum_{k\in K} T_k W_k, \quad \text{with each } T_k \in \mathbb{C}. \tag{3.2}$$

Theorem 3.3. *For every subgroup* S *of* $\mathrm{Sp}(K)$,

$$\mathrm{End}_S\, L^2(A) = \{T \in \mathrm{End}_{\mathbb{C}}\, L^2(A) \mid T_k = T_{g(k)} \text{ for all } g \in S,\ k \in K\}.$$

Proof. Note that $T \in \mathrm{End}_S\, L^2(A)$ if and only if $W_g T W_g^* = T$ for all $g \in S$. Expanding T as in (3.2) and using the defining identity (1.3) for W_g gives the theorem. □

Now suppose that as a representation of S, $L^2(A)$ has the decomposition

$$L^2(A) = \bigoplus_{\pi \in \hat{S}} m_{\pi,S} \mathcal{H}_\pi.$$

Then, together with Schur's lemma, Theorem 3.3 implies:

Corollary 3.4. *If $S \backslash K$ denotes the set of S-orbits in K,*

$$\sum_{\pi \in \hat{S}} m_{\pi,S}^2 = |S \backslash K|.$$

4. Orbits and characteristic subgroups

We first recall the theory of orbits (under the full automorphism group) and characteristic subgroups in a finite abelian group from [Dutta and Prasad 2011]. We then see how it applies to $\mathrm{Sp}(K)$-orbits in K.

4A. ***Orbits.*** Every finite abelian p-group is isomorphic to

$$A = \mathbb{Z}/p^{\lambda_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{\lambda_l}\mathbb{Z} \tag{4.1}$$

for a unique sequence $\lambda = (\lambda_1 \geq \cdots \geq \lambda_l)$ of positive integers (in other words, a partition). Henceforth, we assume that A is of the above form. For each partition λ, let

$$P_\lambda = \big\{(v, k) \mid k \in \{\lambda_1, \ldots, \lambda_l\},\ 0 \leq v < k\big\}.$$

Say that $(v, k) \geq (v', k')$ if and only if $v' \geq v$ and $k' - v' \leq k - v$. This relation is a partial order on P_λ. For $x \in \mathbb{Z}/p^k\mathbb{Z}$, let

$$v(x) = \max\{0 \leq v \leq k \mid x \in p^v\mathbb{Z}/p^k\mathbb{Z}\}.$$

For $a = (a_1, \ldots, a_l) \in A$, let $I(a)$ be the order ideal in P_λ generated by $(v(a_i), \lambda_i)$ with $a_i \neq 0$ in $\mathbb{Z}/p^{\lambda_i}\mathbb{Z}$.

Example 4.2. When $\lambda = (5, 4, 4, 1)$ and $a = (p^4, p^2, p^3, 0)$, the Hasse diagram of P_λ is shown on the left hand side of Figure 1. The ideal $I(a)$ is represented by black dots on the right hand side of Figure 1. Note that the elements of P_λ are arranged in such a way that k is constant along verticals and decreases from left to right.

Theorem 4.3 [Dutta and Prasad 2011, Theorem 4.1]. *For $a, b \in A$, the element b is the image of a under an endomorphism of A if and only if $I(b) \subset I(a)$.*

Given $x = (v, k) \in P_\lambda$, let $e(x)$ denote the element in A all of whose entries are zero except for the leftmost entry with $\lambda_i = k$, which is p^v. For an order ideal I in P_λ, denote by $\max I$ the set of maximal elements in I, and let

$$a(I) = \sum_{x \in \max I} e(x).$$

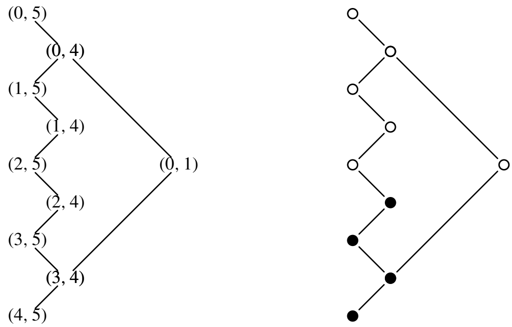


Figure 1. Left: the poset $P_{(5,4,4,1)}$. Right: the order ideal $I(p^4, p^2, p^3, 0)$.

Let G denote the group of all automorphisms of A.

Theorem 4.4 [Dutta and Prasad 2011, Theorem 5.4]. *The map $I \mapsto a(I)$ gives rise to a bijection from the set of order ideals in P_λ to the set of G-orbits in A.*

The elements $a(I)$, as I varies over the order ideals in P_λ, can be taken as representatives of the orbits. The inverse of the function of Theorem 4.4 is given by $a \mapsto I(a)$.

4B. ***Characteristic subgroups.*** For an order ideal $I \subset P_\lambda$,

$$A_I = \{a \in A \mid I(a) \subset I\}$$

is a characteristic subgroup of A of order $p^{[I]}$, where $[I]$ denotes the number of elements in I, counted with multiplicity (the multiplicity of (v, k) is the number of times that k occurs in the partition λ; see [ibid., Theorem 7.3]). Every characteristic subgroup of A is of the form A_I for some order ideal $I \subset P_\lambda$. In fact, $I \mapsto A_I$ is an isomorphism of the lattice of order ideals in P_λ onto the lattice of characteristic subgroups of A. Thus, the lattice of characteristic subgroups of A is a finite distributive lattice [Stanley 1997, Section 3.4]. If B is any group isomorphic to A, and $\phi : A \to B$ is an isomorphism, then since A_I is characteristic, the image $B_I = \phi(A_I)$ does not depend on the choice of ϕ. Consequently, it makes sense to talk of the subgroup $\hat{A}_I$ of $\hat{A}$, which is the image of A_I under any isomorphism $A \to \hat{A}$.

For each order ideal $I \subset P_\lambda$, its annihilator

$$A_I^\perp := \{\chi \in \hat{A} \mid \chi(a) = 1 \text{ for all } a \in A_I\}$$

is a characteristic subgroup of $\hat{A}$. Therefore, there exists an order ideal $I^\perp \subset P_\lambda$ such that $A_I^\perp = \hat{A}_{I^\perp}$. Clearly, $I \mapsto I^\perp$ is an order-reversing involution of the set of order ideals in P_λ. The Hasse diagram of P_λ has a horizontal axis of symmetry. $I^\perp$ can be visualized as the complement of the reflection of I about this axis of I (see Figure 2).

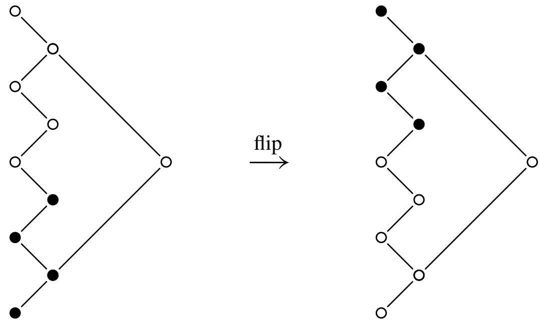


Figure 2. The involution on order ideals. Left: I (black dots). Right: $I^{\perp}$ (white dots).

4C. *Symplectic orbits.*

Theorem 4.5. *The map $I \mapsto (a(I), 0)$ (here 0 denotes the identity element of $\hat{A}$) gives rise to a bijection from the set of order ideals in P_λ to the set of $\mathrm{Sp}(K)$-orbits in K.*

Proof. We first show that each $\mathrm{Sp}(K)$-orbit in K intersects $A \times \{0\}$. Let $e_1, \dots, e_l$ denote the generators of A, so e_i is the element whose i-th coordinate is 1 and all other coordinates are 0. Each element $a \in A$ has an expansion

$$a = a_1 e_1 + \cdots + a_l e_l \quad \text{with } 0 \le a_i < p^{\lambda_i} \text{ for each } i \in \{1, \dots, l\}. \tag{4.6}$$

Let ϵ_j denote the unique element of $\hat{A}$ for which

$$\epsilon_j(e_k) = e^{2\pi i \delta_{jk} p^{-\lambda_j}}.$$

Then each element $\alpha \in \hat{A}$ has an expansion

$$\alpha = \alpha_1 \epsilon_1 + \cdots + \alpha_l \epsilon_l \quad \text{with } 0 \le \alpha_i < p^{\lambda_i} \text{ for each } i \in \{1, \dots, l\}. \tag{4.7}$$

Let $k = (a, \alpha) \in K$, with a and α as in (4.6) and (4.7), respectively. The automorphism of K which takes $e_i \mapsto \epsilon_i$ and $\epsilon_i \mapsto -e_i$ while preserving all other generators e_j and ϵ_j with $j \neq i$, lies in $\mathrm{Sp}(K)$. In terms of coordinates, it has the effect of interchanging a_i and α_i up to sign. Using this automorphism, we may arrange that $v(a_i) \le v(\alpha_i)$ for each i. Therefore, there exists an integer b_i such that

$$b_i a_i \equiv \alpha_i \mod p^{\lambda_i}.$$

Let $B_i : A \to \hat{A}$ be the homomorphism which takes e_i to $b_i \epsilon_i$ and all other generators e_j with $j \neq i$ to 0. Then the automorphism of K which takes (a, α) to $(a, \alpha - B_i(a))$ also lies in $\mathrm{Sp}(K)$. This has the effect of changing α_i to 0. Repeating this process for each i allows us to reduce (a, α) to $(a, 0)$ as claimed.

Now, for every automorphism g of A, the automorphism $(a, \alpha) \mapsto (g(a), \hat{g}^{-1}(\alpha))$ lies in $\mathrm{Sp}(K)$ (here $\hat{g}$ is the automorphism of $\hat{A}$ defined by $\hat{g}(\chi)(a) = \chi(g(a))$ for $a \in A$ and $\chi \in \hat{A}$). Such automorphisms can be used to reduce $(a, 0)$ further to an element of the form $(a(I), 0)$ for some order ideal $I \subset P_\lambda$. Since, for distinct I, these elements are in distinct $\mathrm{Aut}(K)$-orbits, they must also be in distinct $\mathrm{Sp}(K)$-orbits. $\square$

5. Multiplicity one

5A. ***Relation to commutativity.*** Suppose that the decomposition of the Weil representation into irreducible representations is given by

$$L^2(A) = \bigoplus_{\pi \in \mathrm{Sp}(K)^\wedge} m_\pi \mathcal{H}_\pi. \tag{5.1}$$

Then by Schur's lemma, the endomorphism algebra of $L^2(A)$ is a direct sum of matrix algebras:

$$\mathrm{End}_{\mathrm{Sp}(K)} L^2(A) = \bigoplus_{\pi \in \mathrm{Sp}(K)^\wedge} M_{m_\pi \times m_\pi}(\mathbb{C}).$$

It follows that $m_\pi \le 1$ for every $\pi \in \mathrm{Sp}(K)^\wedge$ if and only if the ring $\mathrm{End}_{\mathrm{Sp}(K)} L^2(A)$ of endomorphisms of the Weil representations is commutative. For each order ideal $I \subset P_\lambda$, let O_I denote the $\mathrm{Sp}(K)$-orbit of $(a(I), 0)$ in K and let

$$T_I = \sum_{k \in O_I} W_k.$$

By Theorems 3.3 and 4.5, the set of all T_I as I varies over the order ideals $I \subset P_\lambda$ is a basis of $\mathrm{End}_{\mathrm{Sp}(K)} L^2(A)$.

Let $K_I = A_I \times \hat{A}_I$ ($\hat{A}_I$ as in Section 4B) and define

$$\Delta_I = \sum_{k \in K_I} W_k. \tag{5.2}$$

Since $K_I = \bigsqcup_{J \subset I} O_J$,

$$\Delta_I = \sum_{J \subset I} T_J.$$

Thus, the elements Δ_I are obtained from the basis elements T_I of $\mathrm{End}_{\mathrm{Sp}(K)} L^2(A)$ by a unipotent upper-triangular transformation. Hence,

$$\{\Delta_I \mid I \in J(P_\lambda)\}$$

is also a basis of $\mathrm{End}_{\mathrm{Sp}(K)} L^2(A)$. Thus if Δ_I commutes with Δ_J for all $I, J \in J(P_\lambda)$, then $\mathrm{End}_{\mathrm{Sp}(K)} L^2(A)$ is a commutative algebra.

Therefore, in order to show that $m_\pi \le 1$ for each $\pi \in \mathrm{Sp}(K)^\wedge$, it suffices to show that for any two order ideals $I, J \subset P_\lambda$, Δ_I and Δ_J commute. This will follow from the calculation in Section 5B.

5B. ***Calculation of the product.***

Lemma 5.3. *For any two order ideals $I, J \subset P_\lambda$,*

$$\Delta_I \Delta_J = |K_{I\cap J}| \Delta_{(I\cap J)^\perp \cap (I\cup J)}.$$

Proof. The coefficient of W_k in $\Delta_I \Delta_J$ is

$$\sum_{\substack{x\in K_I, y\in K_J\\ x+y=k}} c(x, y). \tag{5.4}$$

From the definition of $I(a)$, it is easy to see that $I(a+b) \subset I(a)\cup I(b)$. Therefore, $A_I + A_J \subset A_{I\cup J}$ and hence $K_I + K_J \subset K_{I\cup J}$. It follows that the sum (5.4) is 0 unless $k \in K_{I\cup J}$. Suppose $x_0 \in K_I$ and $y_0 \in K_J$ are such that $x_0 + y_0 = k$. Then the sum (5.4) becomes

$$\begin{aligned}
\sum_{l\in K_I\cap K_J} c(x_0+l, y_0-l) &= c(x_0, y_0) \sum_{l\in K_I\cap K_J} c(l, y_0)c(l, x_0)\\
&= c(x_0, y_0) \sum_{l\in K_I\cap K_J} c(l, k)\\
&= \begin{cases} c(x_0, y_0)|K_I \cap K_J| & \text{if } k \in (K_I \cap K_J)^\perp,\\ 0 & \text{otherwise.}\end{cases}
\end{aligned}$$

Observe that $K_I \cap K_J = K_{I\cap J}$. It remains to show that, for every $k \in K_{I\cup J}$, there exist $x_0 \in K_I$ and $y_0 \in K_J$ such that $k = x_0+y_0$ and $c(x_0, y_0) = 1$. Since $\Delta_I\Delta_J$ is constant on $\mathrm{Sp}(K)$-orbits in K, we may use Theorem 4.5 to assume that $k = (a(I'), 0)$ for some order ideal $I' \subset I \cup J$. We have $\max I' \subset I \cup J$. Let I'_1 be the order ideal generated by $(\max I') \cap I$, and I'_2 be the order ideal generated by $(\max I') - I$. Then $a(I') = a(I'_1) + a(I'_2)$. Clearly $(a(I'_1), 0)$ and $(a(I'_2), 0)$ have the properties required of x_0 and y_0. □

5C. ***Multiplicity one.*** We have proved:

Theorem 5.5. *In the decomposition (5.1) of the Weil representation of* $\mathrm{Sp}(K)$, $m_\pi \le 1$ *for every isomorphism class π of irreducible representations of* $\mathrm{Sp}(K)$.

Every $\mathrm{Sp}(K)$-invariant subspace is completely determined by the subset of $\mathrm{Sp}(K)^\wedge$ consisting of representations that occur in it. Therefore:

Corollary 5.6. *The set of* $\mathrm{Sp}(K)$*-invariant subspaces of* $L^2(A)$, *partially ordered by inclusion, forms a finite Boolean lattice.*

6. Elementary invariant subspaces

In this section, we construct some elementary invariant subspaces for the Weil representation of $\mathrm{Sp}(K)$ on $L^2(A)$. In Section 8, we will use these subspaces and the results of Section 7 to construct enough invariant subspaces to carve out all the irreducible subspaces.

6A. *Small order ideals.*

Definition 6.1 (small order ideal). An order ideal $I \subset P_\lambda$ is said to be small if $I \subset I^\perp$, with $I^\perp$ as in Section 4B.

For example, the order ideal I in Figure 2 is small.

6B. *Interpreting some Δ_I.*

Lemma 6.2. *For each order ideal $I \subset P_\lambda$, let Δ_I be as in (5.2).*

(6.2.1) $\Delta_{P_\lambda} f(u) = |A| f(-u)$ *for all* $f \in L^2(A)$ *and* $u \in A$.

(6.2.2) *For every small order ideal $I \subset P_\lambda$, $|A_I|^{-2}\Delta_I$ is the orthogonal projection onto the subspace of $L^2(A)$ consisting of functions supported on $A_{I^\perp}$ and invariant under translations in A_I.*

Proof. For any order ideal $I \subset P_\lambda$, we have

$$\Delta_I f(u) = \sum_{x \in A_I} \sum_{\chi \in \hat{A}_I} \chi(u - x/2) f(u-x).$$

The inner sum is $f(u-x)$ times the sum of values of a character of $\hat{A}_I$, which vanishes if this character is nontrivial, namely if $u - x/2 \notin A_{I^\perp}$, and is $|A_I|$ otherwise. Therefore,

$$\Delta_I f(u) = |A_I| \sum_{x \in A_I \cap (2u + A_{I^\perp})} f(u-x) = |A_I| \sum_{x \in (u+A_I) \cap (-u + A_{I^\perp})} f(x). \tag{6.3}$$

Taking $I = P_\lambda$ in (6.3) gives (6.2.1).

Now suppose that $I \subset I^\perp$. If $u \notin A_{I^\perp}$ then $(u + A_I) \cap (-u + A_{I^\perp}) = \varnothing$, so that $\Delta_I f(u) = 0$. If $u \in A_{I^\perp}$, then the sum (6.3) is over $u + A_I$, so $|A_I|^{-2}\Delta_I$ is the averaging over A_I-cosets, from which (6.2.2) follows. □

6C. *Even and odd functions.*

Theorem 6.4. *The subspaces of $L^2(A)$ consisting of even and odd functions are invariant under* $\mathrm{Sp}(K)$.

Proof. By (6.2.1),

$$\left[\tfrac{1}{2}(\mathrm{Id}_{L^2(A)} \pm |A|^{-1}\Delta_{P_\lambda})\right] f(u) = \tfrac{1}{2}(f(u) \pm f(-u)). \tag{6.5}$$

These operators are the orthogonal projections onto the subspaces of even and odd functions in $L^2(A)$. Since these operators commute with $\mathrm{Sp}(K)$ (by Theorem 3.3), their images are $\mathrm{Sp}(K)$-invariant subspaces of $L^2(A)$. $\square$

Remark 6.6 (the Weil representation is an ordinary representation). The subspaces of even and odd functions on A have dimensions $(|A|+1)/2$ and $(|A|-1)/2$, respectively. For each $g \in \mathrm{Sp}(K)$, let W_g^+ and W_g^- denote the restrictions of W_g to these spaces. Taking the determinants of the identities

$$W_{g_1}^{\pm} W_{g_2}^{\pm} = c(g_1, g_2) W_{g_1 g_2}^{\pm}$$

gives the identities

$$\begin{aligned} \det W_{g_1}^{+} \det W_{g_2}^{+} &= c(g_1, g_2)^{(|A|+1)/2} \det W_{g_1 g_2}^{+}, \\ \det W_{g_1}^{-} \det W_{g_2}^{-} &= c(g_1, g_2)^{(|A|-1)/2} \det W_{g_1 g_2}^{-}. \end{aligned}$$

Dividing the first equation by the second and rearranging gives

$$c(g_1, g_2) = \frac{\alpha(g_1)\alpha(g_2)}{\alpha(g_1 g_2)}$$

for all $g_1, g_2 \in \mathrm{Sp}(K)$, where $\alpha : G \to U(1)$ is defined by

$$\alpha(g) = \frac{\det(W_g^+)}{\det(W_g^-)} \quad \text{for all } g \in \mathrm{Sp}(K).$$

Therefore, if each W_g is replaced by $\alpha(g)^{-1} W_g$, then $g \mapsto W_g$ is a representation of $\mathrm{Sp}(K)$ on $L^2(A)$. This argument seems to be well known. It has appeared before in [Adler and Ramanan 1996, Appendix I] and again in [Cliff et al. 2000].

6D. *Invariant spaces corresponding to small order ideals.* Since Δ_I commutes with $\mathrm{Sp}(K)$, its image is an $\mathrm{Sp}(K)$-invariant subspace of $L^2(A)$. An immediate consequence of (6.2.2) is the following theorem:

Theorem 6.7. *For each small order ideal $I \subset P_\lambda$, the subspace of $L^2(A)$ consisting of functions supported on $A_{I^\perp}$ which are invariant under translations in A_I is an $\mathrm{Sp}(K)$-invariant subspace of $L^2(A)$.*

Remark 6.8 (alternative description). For $f \in L^2(A)$, recall that its Fourier transform is the function on $\hat{A}$ defined by

$$\hat{f}(\chi) = \sum_{a \in A} f(a) \overline{\chi(a)} \quad \text{for each } \chi \in \hat{A}.$$

For any subgroup B of A, the Fourier transforms of functions invariant under translations in B are the functions supported on the annihilator subgroup $B^\perp$ of A (consisting of characters which vanish on B), and Fourier transforms of functions supported on B are the functions which are invariant under $B^\perp$. Therefore, the functions supported on $A_{I^\perp}$ which are invariant under A_I are precisely the functions supported on $A_{I^\perp}$ whose Fourier transforms are supported on $\hat{A}_{I^\perp}$. They are also the functions invariant under translations in A_I whose Fourier transforms are invariant under translations in $\hat{A}_I$.

Identify $L^2(A_{I^\perp}/A_I)$ with the space of functions in $L^2(A)$ which are supported on $A_{I^\perp}$ and invariant under translations in A_I. Let $K(I) = A_{I^\perp}/A_I \times (A_{I^\perp}/A_I)^\wedge$. $K(I)$ can be identified with $({A_I}^\perp \times \hat{A}_{I^\perp})/(A_I \times \hat{A}_I)$. Thus $K(I)$ is a quotient of one characteristic subgroup of K by another. Therefore the action $\mathrm{Sp}(K)$ on K descends to an action on $K(I)$, giving rise to a homomorphism $\mathrm{Sp}(K) \to \mathrm{Sp}(K(I))$. The defining condition (1.3) for the Weil representation ensures:

Theorem 6.9. *For every small order ideal $I \subset P_\lambda$, the Weil representation of* $\mathrm{Sp}(K)$ *on $L^2(A_{I^\perp}/A_I)$ is projectively equivalent to the representation obtained by composing the Weil representation of* $\mathrm{Sp}(K(I))$ *on $L^2(A_{I^\perp}/A_I)$ with the homomorphism* $\mathrm{Sp}(K) \to \mathrm{Sp}(K(I))$.

7. Component decomposition

7A. ***Connected components of a partially ordered set.*** A partially ordered set is said to be connected if its Hasse diagram is a connected graph. A connected component of a partially ordered set is a maximal connected induced subposet. Every partially ordered set can be written as the disjoint union of its connected components in the sense of [Stanley 1997, Section 3.2]. Denote the set of connected components of a poset P by $\pi_0(P)$.

7B. ***Connected components of $J - I$.*** Suppose that $I \subset J$ are two order ideals in P_λ. Each connected component $C \in \pi_0(J - I)$ determines a segment (namely, a contiguous set of integers) S_C in $\{1, \dots, l\}$:

$$S_C = \{1 \le k \le l \mid (v, k) \in C \text{ for some } v\}.$$

The S_C are pairwise disjoint, but their union may be strictly smaller than $\{1, \dots, l\}$. Write S_0 for the complement of $\bigsqcup_{C \in \pi_0(I^\perp - I)} S_C$ in $\{1, \dots, l\}$. It will be convenient to write

$$\tilde{\pi}_0(I^\perp - I) = \pi_0(I^\perp - I) \sqcup \{0\}.$$

Define partitions $\lambda(C) = (\lambda_k \mid k \in S_C)$ for each $C \in \tilde{\pi}_0(I^\perp - I)$. Then $P_{\lambda(C)}$ is the induced subposet of P_λ consisting of those pairs $(v, k) \in P_\lambda$ for which $k \in S_C$. Let $I(C)$ and $J(C)$ be the ideals in $P_{\lambda(C)}$ obtained by intersecting I and J, respectively, with $P_{\lambda(C)}$.

For example, if $\lambda = (5, 4, 4, 1)$ and I is the order ideal in the diagram on the left in Figure 2 and $J = I^\perp$, then $I^\perp - I$ is depicted in the diagram on the left in

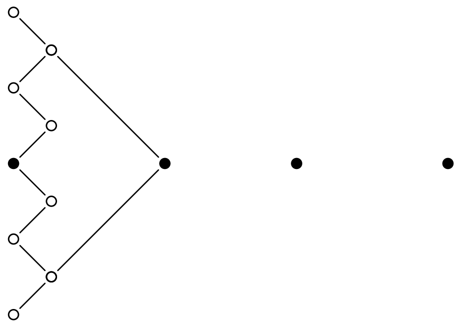


Figure 3. Left: $I^\perp - I$ inside $P_{5,4,4,1}$. Right: $I^\perp - I$ by itself.

Figure 3. As the diagram on the right shows, the induced subposet $I^\perp - I$ has two connected components, C_1 and C_2, with $\lambda(C_1) = (5)$ and $\lambda(C_2) = (1)$. Moreover, $\lambda(0) = (4, 4)$.

Lemma 7.1. *Let $I \subset J$ be two order ideals in P_λ. For each $C \in \pi_0(J - I)$ let $L(C)$ be an order ideal in C. Let*

$$L = I \sqcup \bigsqcup_{C \in \pi_0(J-I)} L(C).$$

Then L is an order ideal in P_λ.

Proof. Since $\bigsqcup_{C\in\pi_0(J-I)} L(C)$ is an order ideal in $J - I$, its union with I is an order ideal in P_λ. □

Corollary 7.2. *If $I \subset J$ are two order ideals in P_λ and C and D are distinct components of $J - I$, then the intersection with $P_{\lambda(C)}$ of the order ideal in P_λ generated by $J(D)$ is contained in $I(C)$.*

Proof. By Lemma 7.1, $J(D) \cup I$ is an order ideal in P_λ. Therefore, it contains the order ideal in P_λ generated by $J(D)$. If C and D are distinct connected components of $J - I$, then $(J(D) \cup I) \cap P_\lambda(C) = I(C)$. Therefore the intersection with $P_{\lambda(C)}$ of the order ideal in P_λ generated by $J(D)$ is contained in $I(C)$. □

7C. ***Decomposition of endomorphisms.*** Suppose that A has the form (4.1). Then define A_C to be the subgroup

$$A_C = \{(a_1, \dots, a_l) \mid a_k = 0 \text{ if } k \notin S_C\}.$$

Thus A_C is a finite abelian p-group of type $\lambda(C)$. We have a decomposition

(7.3) $$A = \prod_{C \in \tilde{\pi}_0(J-I)} A_C.$$

Denote the characteristic subgroups of A_C corresponding to $I(C)$ and $J(C)$

(which are order ideals in $P_{\lambda(C)}$) by $A_{I,C}$ and $A_{J,C}$ respectively. The decomposition (7.3) induces a decomposition

$$A_J/A_I = \prod_{C\in\pi_0(J-I)} A_{J,C}/A_{I,C}. \tag{7.4}$$

There is no contribution from A_0 since $A_{I,0} = A_{J,0}$.

With respect to the decomposition (7.3), every endomorphism of A can be written as a square matrix $\{\phi_{CD}\}$, where $\phi_{CD} : A_D \to A_C$ is a homomorphism.

Lemma 7.5. *Let $I \subset J$ be order ideals in P_λ. Then every endomorphism ϕ of A induces an endomorphism*

$$\bar{\phi} : A_J/A_I \to A_J/A_I$$

such that $\bar{\phi}(A_{J,C}/A_{I,C}) \subset A_{J,C}/A_{I,C}$ for each $C \in \pi_0(J-I)$, and

$$\bar{\phi} = \bigoplus_{C\in\pi_0(J-I)} \overline{\phi_{CC}},$$

where $\overline{\phi_{CC}}$ is the endomorphism of $A_{J,C}/A_{I,C}$ induced by ϕ_{CC}.

Proof. By Theorem 4.3 and Corollary 7.2, if $C \neq D$ then $\phi_{CD}(A_{J,D}) \subset A_{I,C}$. Therefore, $\bar{\phi}$ remains unchanged if ϕ_{CD} is replaced by 0 for all $C \neq D$. This amounts to replacing ϕ by $\oplus_C \phi_{CC}$, and the lemma follows. □

7D. ***Tensor product decomposition of invariant subspaces.*** Let I be a small order ideal. We shall use the notation of Section 7C with $J = I^\perp$. For each $C \in \pi_0(I^\perp - I)$ let $K_C = A_C \times \hat{A}_C$, and let $\mathrm{Sp}(K_C)$ be the corresponding symplectic group. Just as (by Theorem 6.7) $L^2(A_{I^\perp}/A_I)$ is an invariant subspace for the Weil representation of $\mathrm{Sp}(K)$ on $L^2(A)$, $L^2(A_{I^\perp,C}/A_{I,C})$ is an invariant subspace for the Weil representation of $\mathrm{Sp}(K_C)$ on $L^2(A_C)$.

Now, if $g \in \mathrm{Sp}(K)$, we may write

$$g = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$$

with respect to the decomposition $K = A \times \hat{A}$. For convenience, we identify $\hat{A}$ with A using $e_i \mapsto \epsilon_i$ for $i = 1, \dots, l$, where e_i and ϵ_i are as in Section 4C. Hence, we may think of each g_{ij} as an endomorphism of A. By Lemma 7.5, the resulting endomorphism $\bar{g}_{ij}$ of $A_{I^\perp}/A_I$ preserves $A_{I^\perp,C}/A_{I,C}$ for each C. It follows that the image of $\mathrm{Sp}(K)$ in $\mathrm{Sp}(K(I))$ (see Section 6D) is the product of the images of the $\mathrm{Sp}(K_C)$ in the $\mathrm{Sp}(K_C(I\cap C))$ as C ranges over $\pi_0(I^\perp - I)$.

Thus, by Theorems 2.4 and 6.9, we have:

Corollary 7.6. *The Weil representation of* $\mathrm{Sp}(K)$ *on* $L^2(A_{I^\perp}/A_I)$ *is projectively equivalent to the tensor product of the Weil representations of the* $\mathrm{Sp}(K_C)$ *on the* $L^2(A_{I^\perp,C}/A_{I,C})$ *as* C *ranges over* $\pi_0(I^\perp - I)$.

8. Poset of invariant subspaces

8A. ***The invariant subspaces.*** Let

$Q_\lambda = \{(I, \phi) \mid I \subset P_\lambda$ a small order ideal, $\phi : \pi_0(I^\perp - I) \to \mathbb{Z}/2\mathbb{Z}$ any function$\}$.

For each $(I, \phi) \in Q_\lambda$, use the decomposition of Corollary 7.6 to define $L^2(A)_{I,\phi}$ as the subspace of $L^2(A_{I^\perp}/A_I)$ given by

$$L^2(A)_{I,\phi} = \bigotimes_{C \in \pi_0(I^\perp - I)} L^2(A_{I^\perp,C}/A_{I,C})_{\phi(C)},$$

where $L^2(A_{I^\perp,C}/A_{I,C})_{\phi(C)}$ denotes the space of even or odd functions on the quotient $A_{I^\perp,C}/A_{I,C}$ when $\phi(C)$ is 0 or 1, respectively. In other words, $L^2(A)_{I,\phi}$ consists of functions on $A_{I^\perp}/A_I$ which, under the decomposition

$$A_{I^\perp}/A_I = \prod_{C \in \pi_0(I^\perp - I)} A_{I^\perp,C}/A_{I,C}, \tag{8.1}$$

are even in the components where $\phi(C) = 0$ and odd in the components where $\phi(C) = 1$. By Theorems 6.4 and 6.7, and by Corollary 7.6, $L^2(A)_{I,\phi}$ is an $\mathrm{Sp}(K)$-invariant subspace of $L^2(A)$ for each $(I, \phi) \in Q_\lambda$.

8B. ***The partial order.*** Clearly:

Lemma 8.2. *For* (I, ϕ) *and* (I', ϕ') *in* Q_λ, $L^2(A)_{I',\phi'} \subset L^2(A)_{I,\phi}$ *if and only if the following conditions are satisfied*:

(8.2.1) $I \subset I'$.

(8.2.2) *For each* $P \in \pi_0(I^\perp - I)$,

$$\phi(P) = \sum_{\substack{P' \in \pi_0(I'^\perp - I') \\ P' \subset P}} \phi'(P').$$

Thus, the conditions (8.2.1) and (8.2.2) define a partial order on Q_λ (which is obviously independent of p). Recall that the multiplicity $m(x)$ of an element $x = (v, k) \in P_\lambda$ is the number of times k occurs in the partition λ. For any subset $S \subset P_\lambda$, let $[S]$ denote the number of elements of S, counted with multiplicity:

$$[S] = \sum_{x \in S} m(x).$$

Lemma 8.3. *For each* $(I, \phi) \in Q_\lambda$,

$$\dim L^2(A)_{I,\phi} = \prod_{C \in \pi_0(I^\perp - I)} \frac{p^{[C]} + (-1)^{\phi(C)}}{2}.$$

9. Irreducible subspaces

9A. ***A bijection between $J(P_\lambda)$ and Q_λ.*** Let $J(P_\lambda)$ denote the lattice of order ideals in P_λ.

Lemma 9.1. *For each partition* λ, $|J(P_\lambda)| = |Q_\lambda|$.

Proof. We construct an explicit bijection $Q_\lambda \to J(P_\lambda)$. To $(I, \phi) \in Q_\lambda$, associate the ideal (see Lemma 7.1)

$$\Theta(I, \phi) = I \cup \bigsqcup_{C \in \pi_0(I^\perp - I)} I_{\phi(C)},$$

where

$$I_{\phi(C)} = \begin{cases} I \cap C & \text{if } \phi(C) = 0, \\ I^\perp \cap C & \text{if } \phi(C) = 1. \end{cases}$$

In the other direction, given an ideal $J \subset P_\lambda$, $I = J \cap J^\perp$ is a small order ideal. We have $I^\perp = J \cup J^\perp$. For each $C \in \pi_0(I^\perp - I)$, define

$$\phi_J(C) = \begin{cases} 0 & \text{if } I \cap C = J \cap C, \\ 1 & \text{if } I \cap C = J^\perp \cap C. \end{cases}$$

Define $\Psi : Q_\lambda \to J(P_\lambda)$ by $\Psi(J) = (J \cap J^\perp, \phi_J)$. It is easy to verify that Φ and Ψ are mutual inverses. □

9B. ***Existence lemma.***

Lemma 9.2. *For every* $(I, \phi) \in Q_\lambda$, *there exists* $f \in L^2(A)_{I,\phi}$ *such that* $f \notin L^2(A)_{I',\phi'}$ *for any* $(I', \phi') < (I, \phi)$.

Proof. Take as f the unique element in $L^2(A)_{I,\phi}$ whose value at $a(I^\perp) + A_I$ (using the notation of Section 4A) is 1, and which vanishes on all elements of $A_{I^\perp}/A_I$ not obtained from $a(I^\perp) + A_I$ by changing the signs of some of its components under the decomposition (8.1). □

9C. ***The irreducible invariant subspaces.*** The two lemmas above are enough to give us the main theorem:

Theorem 9.3. *For each* $(I, \phi) \in Q_\lambda$, *there is a unique irreducible subspace for the Weil representation of* $\mathrm{Sp}(K)$ *on* $L^2(A)$ *which is contained in* $L^2(A)_{I,\phi}$ *but not* $L^2(A)_{I',\phi'}$ *for any* $(I', \phi') < (I, \phi)$. *As p varies, the dimension of this representation is a polynomial in p of degree* $[I^\perp - I]$ *with leading coefficient* $2^{-|\pi_0(I^\perp - I)|}$ *and all coefficients in* $\mathbb{Z}_{(2)}$.

Proof. By Corollary 5.6, the $\mathrm{Sp}(K)$-invariant subspaces of $L^2(A)$ form a Boolean lattice Λ. Let R denote the set of minimal nontrivial $\mathrm{Sp}(K)$-invariant subspaces of $L^2(A)$. These are the atoms of Λ. By Corollary 3.4 and Theorem 4.5, the cardinality of R is the same as that of $J(P_\lambda)$. Each invariant subspace is determined by the atoms which are contained in it. The map $(I,\phi)\mapsto L^2(A)_{I,\phi}$ is an order-preserving map $Q_\lambda\to\Lambda$. Let $R_{I,\phi}$ be the set of atoms which occur in $L^2(A)_{I,\phi}$ but not in $L^2(A)_{I',\phi'}$ for any $(I',\phi')<(I,\phi)$. The subsets $R_{I,\phi}$ are $|Q_\lambda|$ pairwise disjoint subsets of R, and by Lemma 9.2, each of them is nonempty. Therefore, by Lemma 9.1, each of them must be a singleton, and these subspaces exhaust R. It follows that there is a unique irreducible representation of $\mathrm{Sp}(K)$ that occurs in $L^2(A)_{I,\phi}$ but not in $L^2(A)_{I',\phi'}$ for any $(I',\phi')<(I,\phi)$. Let $V_{I,\phi}$ denote this irreducible subspace.

By Lemma 8.3,

$$\sum_{(I',\phi')\leq(I,\phi)}\dim V_{I',\phi'}=\prod_{P\in\pi_0(I^\perp-I)}\frac{p^{[C]}+(-1)^{\phi(C)}}{2}.$$

By the Möbius inversion formula [Stanley 1997, Section 3.7],

$$\dim V_{I,\phi}=\sum_{(I',\phi')\leq(I,\phi)}\mu((I,\phi),(I',\phi'))\prod_{C\in\pi_0(I'^\perp-I')}\frac{p^{[C]}+(-1)^{\phi(C)}}{2},\tag{9.4}$$

where μ is the Möbius function of Q_λ. Since $\mu((I,\phi),(I,\phi))=1$ and the Möbius function is integer-valued, the right-hand side of (9.4) is indeed a polynomial in p with leading coefficient $2^{-|\pi_0(I^\perp-I)|}$. Clearly, the other coefficients do not have denominators other than powers of 2. $\square$

9D. *A combinatorial lemma.*

Lemma 9.5. *Let P be a poset and $J(P)$ be its lattice of order ideals. Let $m:P\to\mathbb{N}$ be any function (called the multiplicity function). For each subset $S\subset P$, let $[S]=\sum_{x\in S}m(x)$, the elements of S counted with multiplicity, and let* $\max S$ *denote the set of maximal elements of S. If $\alpha: J(P)\to\mathbb{C}[t]$ is a function such that*

$$\sum_{J\subset I}\alpha(J)=t^{[I]}\ \textit{for every order ideal } I\subset P,$$

then

$$\alpha(I)=t^{[I]}\prod_{x\in\max I}(1-t^{-m(x)}).\tag{9.6}$$

Proof. By the Möbius inversion formula for a finite distributive lattice [Stanley 1997, Example 3.9.6],

$$\alpha(I)=\sum_{I-\max I\subset J\subset I}(-1)^{|I-J|}t^{[J]}=t^{[I]}\sum_{S\subset\max I}(-1)^{|\max I-S|}t^{-[\max I-S]}.\tag{9.7}$$

Each term in the expansion of the product

$$\prod_{x \in \max I} (1 - t^{-m(x)})$$

is obtained by choosing a subset $S \subset \max I$ and taking

$$\prod_{x \notin S} (-t^{-m(x)}) = (-1)^{|\max I - S|} t^{-[\max I - S]}.$$

Therefore, the expression (9.7) for $\alpha(I)$ reduces to (9.6) as claimed. □

9E. ***Explicit formula for the dimension.*** Recall (from Section 9C) that for each $(I, \phi) \in Q_\lambda$, $V_{I,\phi}$ denotes the unique irreducible $\mathrm{Sp}(K)$-invariant subspace of $L^2(A)$ which lies in $L^2(A)_{I,\phi}$ but not in any proper subspace of the form $L^2(A)_{I',\phi'}$. We shall obtain a nice expression for $\dim V_{I,\phi}$ by applying Lemma 9.5 to the induced subposet of P_λ given by

$$P_\lambda^+ = \{(v, k) \in P_\lambda \mid v < (k-1)/2\}.$$

For each small order ideal $I \subset P_\lambda$, let $I^+ = I^\perp \cap P_\lambda^+$. Then $I \mapsto I^+$ is an order-reversing isomorphism from the partially ordered set of small order ideals in P_λ to the partially ordered set $J(P_\lambda^+)$ of all order ideals in P_λ^+.

Let

$$V_I = \bigoplus_{\phi : \pi_0(I^\perp - I) \to \mathbb{Z}/2\mathbb{Z}} V_{I,\phi}. \tag{9.8}$$

Denote by V_I^0 and V_I^1 the subspaces of even or odd functions in V_I respectively.

Lemma 9.9. *If $I \subset P_\lambda$ is a small order ideal, then for $\epsilon \in \{0, 1\}$,*

$$\dim V_I^\epsilon = \begin{cases} (p^{[I^\perp - I]} + (-1)^\epsilon)/2 & \text{if } I^+ = \varnothing, \\ p^{[I^\perp - I]} \prod_{x \in \max I^+} (1 - p^{-2m(x)})/2 & \text{otherwise.} \end{cases}$$

Proof. Suppose $I \subset P_\lambda$ is a small order ideal. Then

$$L^2(A_{I^\perp}/A_I) = \bigoplus_{\substack{J \supset I \\ J \text{ small}}} V_J = \bigoplus_{J^+ \subset I^+} V_J. \tag{9.10}$$

Define $\alpha : J(P_\lambda^+) \to \mathbb{C}$ by $\alpha(J^+) = \dim V_J$. Comparing dimensions,

$$\sum_{J^+ \subset I^+} \alpha(J^+) = p^{[I^\perp - I]}. \tag{9.11}$$

Let $E = \{(v, k) \in I^\perp - I \mid v = (k-1)/2\}$, the set of points in $I^\perp - I$ which lie on its axis of symmetry. Then $[I^\perp - I] = [E] + 2[I^+]$. Therefore (9.11) becomes

$$\sum_{J^+ \subset I^+} \alpha(J^+) = p^{[E]} p^{2[I^+]}.$$

Taking $P = P_\lambda^+$ and setting $t = p^2$ in Lemma 9.5 gives

$$\dim V_I = p^{[E]+2[I^+]} \prod_{x\in\max I^+} (1-p^{-2m(x)}) = p^{[I^\perp - I]} \prod_{x\in \max I^\perp} (1-p^{-2m(x)}).$$

In order to obtain Lemma 9.9, it remains to find the dimensions of the spaces of even and odd functions in V_I. If $I^+ = \varnothing$ then $E = I^\perp - I$. In this case, $V_{I,\phi}$ is just the set of even or odd functions in $L^2(A_{I^\perp}/A_I)$ and has dimension as claimed.

Otherwise, we proceed by induction on I^+. Thus, assume that Lemma 9.9 holds for small order ideals $I' \supsetneq I$. The space of even functions in $L^2(A_{I^\perp}/A_I)$ has dimension one more than the space of odd functions. Breaking up the spaces in (9.10) into even and odd functions, we see this difference is accounted for by the summand corresponding to $J^+ = \varnothing$, as discussed above. By the induction hypothesis, the dimensions of even and odd parts of the summands corresponding to $\varnothing \subsetneq J^+ \subsetneq I^+$ are equal. Therefore, the even and odd parts of V_I must have the same dimension. □

Theorem 9.12. *If $I \subset P_\lambda$ is a small order ideal, then*

$$\dim V_{I,\phi} = \prod_{C\in\pi_0(I^\perp - I)} \dim V_{I(C),\phi(C)},$$

where, since $I(C)^\perp - I(C)$ is connected, $\dim V_{I(C),\phi(C)}$ *is given by Lemma 9.9.*

9F. *Examples.* We begin with the case $A = (\mathbb{Z}/p^k\mathbb{Z})^l$, corresponding to the partition $\lambda = (k, \dots, k)$ (repeated l times). P_λ is then a linear order, with k points. Q_λ has two linear components, consisting of the even and odd parts. An informative way to display the decomposition of $L^2(A)$ is as the Hasse diagram of Q_λ, but with the dimension of the corresponding irreducible invariant subspace in place of each vertex. In this case we get

$$\begin{array}{ccc}
\dfrac{p^{lk}(1-p^{-2})}{2} & \qquad & \dfrac{p^{lk}(1-p^{-2})}{2} \\
\Big| & & \Big| \\
\dfrac{p^{l(k-2)}(1-p^{-2})}{2} & & \dfrac{p^{l(k-2)}(1-p^{-2})}{2} \\
\vdots & & \vdots \\
\dfrac{p^{(k-2(\lfloor k/2\rfloor-1))}(1-p^{-2})}{2} & & \dfrac{p^{(k-2(\lfloor k/2\rfloor-1))}(1-p^{-2})}{2} \\
\Big| & & \Big| \\
\dfrac{p^{(k-2\lfloor k/2\rfloor)}+1}{2} & & \dfrac{p^{(k-2\lfloor k/2\rfloor)}-1}{2}
\end{array}$$

The entry at the bottom right is zero when k is even and should be omitted. This is consistent with the previously known results in [Prasad 1998; Cliff et al. 2000]. The picture for $\lambda = (2, 1)$ is the same as that for $\lambda = (3)$. Perhaps the simplest nontrivial example is $\lambda = (3, 1)$ (it is the smallest example where $J(P_\lambda)$ is not a chain). We get

$$\begin{tikzcd}
 & \frac{p^4-p^2}{2} \arrow[dl, dash] \arrow[d, dash] & \frac{p^4-p^2}{2} \arrow[d, dash] \arrow[dr, dash] & \\
\frac{(p+1)^2}{4} & \frac{(p-1)^2}{4} & \frac{p^2-1}{4} & \frac{p^2-1}{4}
\end{tikzcd}$$

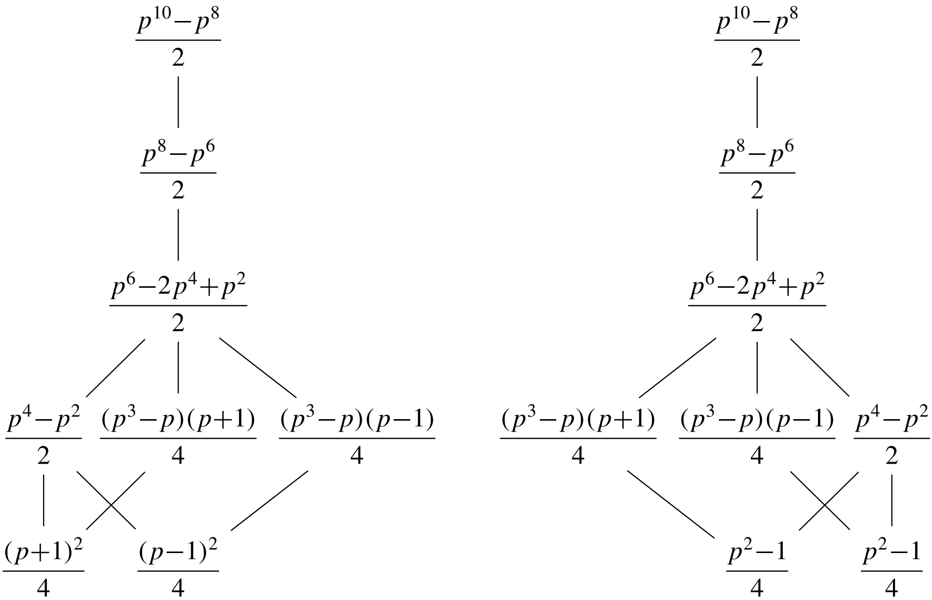
For $\lambda = (3, 2, 1)$, we get

$$\begin{tikzcd}
 & \frac{p^6-p^4}{2} \arrow[d, dash] & \frac{p^6-p^4}{2} \arrow[d, dash] & \\
 & \frac{p^4-p^2}{2} \arrow[dl, dash] \arrow[d, dash] & \frac{p^4-p^2}{2} \arrow[d, dash] \arrow[dr, dash] & \\
\frac{(p+1)^2}{4} & \frac{(p-1)^2}{4} & \frac{p^2-1}{4} & \frac{p^2-1}{4}
\end{tikzcd}$$

For $\lambda = (4, 2)$, we have

$$\begin{tikzcd}
 & \frac{p^6-p^4}{2} \arrow[d, dash] & \frac{p^6-p^4}{2} \arrow[d, dash] & \\
 & \frac{p^4-2p^2+1}{2} \arrow[dl, dash] \arrow[d, dash] & \frac{p^4-2p^2+1}{2} \arrow[d, dash] \arrow[dr, dash] & \\
\frac{p^2-1}{2} \arrow[dr, dash] & \frac{p^2-1}{2} \arrow[d, dash] & \frac{p^2-1}{2} & \frac{p^2-1}{2} \\
 & 1 & &
\end{tikzcd}$$

For $\lambda = (4, 3, 2, 1)$, we have



9G. ***Projections onto the irreducible subspaces.*** For each $(I, \phi) \in Q_\lambda$, let $E_{I,\phi}$ denote the projection operator onto $V_{I,\phi}$. Recall from Lemma 3.1 that the set of Weyl operators $\{W_k \mid k \in K\}$ is an orthonormal basis of $\operatorname{End}_{\mathbb{C}} L^2(A)$. Therefore, we may write

$$E_{I,\phi} = \sum_{k \in K} e_k(I, \phi) W_k$$

for some scalars $e_k(I, \phi)$. The goal of this section is to show that this expansion is completely combinatorial. More precisely, by Theorem 4.5, each $\mathrm{Sp}(K)$-orbit in K corresponds to an order ideal in P_λ. We shall show that if k lies in the $\mathrm{Sp}(K)$-orbit corresponding to the order ideal J, then $e_k(I, \phi)$ is a polynomial in p whose coefficients depend only on the combinatorial data I, ϕ, and J.

In Section 5A we saw that $\{\Delta_L \mid L \in J(P_\lambda)\}$ is a basis of $\operatorname{End}_{\mathrm{Sp}(K)} L^2(A)$. Therefore, we may write

$$E_{I,\phi} = \sum_{L \subset P_\lambda} \alpha_L(I, \phi) \Delta_L$$

for some constants $\alpha_L(I, \phi)$. If k lies in the orbit corresponding to J then

$$e_k(I, \phi) = \sum_{L \supset J} \alpha_L(I, \phi).$$

Therefore, it suffices to show that the $\alpha_L(I, \phi)$ are polynomials in p whose coefficients are determined by the combinatorial data L, I, and ϕ (Theorem 9.17). In fact, Theorems 9.18 and 9.19 compute $\alpha_L(I, \phi)$ explicitly.

To begin with, consider the case where $I^\perp - I$ is connected. If E_I is the projection operator onto V_I (defined by (9.8)), then by (6.2.2),

$$|A|^{-1} p^{[I^\perp - I]} \Delta_I = \sum_{J^+ \subset I^+} E_J.$$

Using Möbius inversion for a finite distributive lattice as in Section 9D,

$$|A| E_I = \sum_{I^+ - \max I^+ \subset J^+ \subset I^+} (-1)^{|I^+ - J^+|} p^{[J^\perp - J]} \Delta_J.$$

Since $V_{I,\phi}$ consists of even or odd functions in V_I (depending on whether $\phi(I^\perp - I)$ is 0 or 1), by (6.5), $E_{I,\phi}$ is given by

$$E_{I,\phi} = \tfrac{1}{2} E_I \bigl(\mathrm{Id}_{L^2(A)} + (-1)^{\phi(I^\perp - I)} |A|^{-1} \Delta_{P_\lambda}\bigr).$$

By Lemma 5.3,

$$(|A|^{-1} p^{[J^\perp - J]} \Delta_J)(|A|^{-1} \Delta_{P_\lambda}) = |A|^{-1} \Delta_{J^\perp}.$$

Therefore, when $I^\perp - I$ is connected,

$$2|A| E_{I,\phi} = \sum_{I^+ - \max I^+ \subset J^+ \subset I^+} (-1)^{|I^+ - J^+|} \bigl(p^{[J^\perp - J]} \Delta_J + (-1)^{\phi(I^\perp - I)} \Delta_{J^\perp}\bigr). \tag{9.13}$$

Now take $I \subset P_\lambda$ to be any small order ideal. The decomposition (7.3) gives

$$L^2(A) = \bigotimes_{C \in \tilde{\pi}_0(I^\perp - I)} L^2(A_C)$$

and

$$V_{I,\phi} = \Bigl(\bigotimes\nolimits_{C \in \pi_0(I^\perp - I)} V_{I(C),\phi(C)}\Bigr) \otimes L^2(A_{I^\perp(0)}/A_{I(0)}),$$

the last factor being one dimensional (since $I(0) = I^\perp(0)$). So we have

$$E_{I,\phi} = \Bigl(\bigotimes\nolimits_{C \in \pi_0(I^\perp - I)} E_{I(C),\phi(C)}\Bigr) \otimes \Delta_{I(0)}, \tag{9.14}$$

where, since $I(C)^\perp - I(C)$ is connected, $E_{I(C),\phi(C)}$ is determined by (9.13). A typical term in the expansion (9.14) will be of the form

$$\Bigl(\bigotimes\nolimits_{C \in \pi_0(I^\perp - I)} \Delta_{L(C)}\Bigr) \otimes \Delta_{I(0)}, \tag{9.15}$$

where for each $C \in \pi_0(I^\perp - I)$ we have $I(C) \subset L(C) \subset I^\perp(C)$, either $L(C)$ or $L(C)^\perp$ being a small order ideal in $P_{\lambda(C)}$. But this is just Δ_L, where

$$L = I \ \cup \bigsqcup_{C \in \pi_0(I^\perp - I)} L(C), \tag{9.16}$$

is an order ideal in P_λ by Lemma 7.1. We have this qualitative result:

Theorem 9.17. *For each* $(I,\phi)\in Q_\lambda$, $2^{|\pi_0(I^\perp-I)|}|A|\alpha_L(I,\phi)$ *is a polynomial in* p *whose coefficients are integers depending only on the combinatorial data* I, ϕ, L.

Let $I_L = L\cap L^\perp$. Examining (9.13) more carefully gives:

Theorem 9.18. *The coefficient* $\alpha_L(I,\phi)$ *is nonzero if and only if the following conditions hold*:

(9.18.1) *For each* $C\in\pi_0(I^\perp-I)$, *either* $L(C)$ *or* $L(C)^\perp$ *is a small order ideal in* $P_{\lambda(C)}$.

(9.18.2) $I^+ - \max I^+ \subset I_L^+ \subset I^+$.

Proof. For $\alpha_L(I,\phi)$ to be nonzero, it is necessary that L be of the form (9.16) for some order ideals $L(C)$ of $P_{\lambda(C)}$ which occur in the right hand side of (9.13). Furthermore, since each order ideal in $P_{\lambda(C)}$ appears at most once in the right hand side of (9.13), so each order ideal in P_λ appears only once in the expansion (9.14). In particular, no cancellation is possible, and for all such ideals $\alpha_L(I,\phi)\neq 0$.

Now $L(C)$ appears on the right hand side of (9.13) if and only if (9.18.1) holds, and $I(C)^+ - \max I(C)^+ \subset I_L(C)^+ \subset I(C)^+$. Since $\max I^+ = \bigsqcup_C \max I(C)^+$, this amounts to the condition (9.18.2). □

If these conditions do hold, then for each $C'\in\pi_0(I_L^\perp - I_L)$, there exists $C\in\pi_0(I^\perp-I)$ such that $C'\subset C$. Furthermore, $\phi_L(C')$ depends only on C, so we may denote its value by $\phi_L(C)$. For $I=I(C)$, the right hand side of (9.13) can be written as

$$\sum_{L(C)}(-1)^{|I(C)^+-L(C)^+|+\phi(C)\phi_L(C)}\,p^{[I_L(C)^\perp-L(C)]},$$

the sum being over an appropriate set of order ideals $L(C)\subset P_{\lambda(C)}$. Let

$$\langle\phi_1,\phi_2\rangle=\sum_{C\in\pi_0(I^\perp-I)}\phi_1(C)\phi_2(C)$$

for any functions $\phi_i:\pi_0(I^\perp-I)\to\mathbb{Z}/2\mathbb{Z}$. The additive nature of the exponents in the above expression allows us to get an exact expression for $\alpha_L(I,\phi)$:

Theorem 9.19. *If an order ideal* $L\subset P_\lambda$ *satisfies the conditions of Theorem 9.18, then*

$$2^{|\pi_0(I^\perp-I)|}|A|\alpha_L(I,\phi)=(-1)^{|I^+-I_L^+|+\langle\phi,\phi_L\rangle}\,p^{[I_L^\perp-L]}.$$

10. Finite modules over a Dedekind domain

Let F be a non-Archimedean local field with ring of integers R. Let P denote the maximal ideal of R. Assume that the residue field R/P is of odd order q. Fix a continuous character $\psi: F\to U(1)$ whose restriction to R is trivial, but whose restriction to $P^{-1}R$ is not (see, for example, Tate's thesis [1967]). Then if $\psi_x(y) = \psi(xy)$, the map $x \mapsto \psi_x$ is an isomorphism of F into $\hat{F}$. Under this isomorphism, R has image $R^\perp = (F/R)^\wedge$. More generally, P^{-n} has image $(P^n)^\perp = (F/P^n)^\wedge$ for every integer n (recall that for positive n, P^{-n} is the set of elements $x \in F$ such that $xP^n \subset R$). Thus, it gives rise to an isomorphism $P^{-n}/R \to (R/P^n)^\wedge$ for each positive integer n. Since P^{-n}/R inherits the structure of an R-module, this isomorphism also allows us to think of $(R/P^n)^\wedge$ as an R-module. Now suppose A is a finitely generated torsion module over R. Then

$$A = R/P^{\lambda_1} \times \cdots \times R/P^{\lambda_l} \tag{10.1}$$

for a unique partition λ. By the discussion above, $\hat{A}$ is also an R-module (noncanonically isomorphic to A). Let $K = A \times \hat{A}$, and $\mathrm{Sp}(K)$ be as in Theorem 1.1. Define $\mathrm{Sp}_R(K)$ to be the subgroup of $\mathrm{Sp}(K)$ consisting of R-module automorphisms.

The Weil representation of $\mathrm{Sp}_R(K)$ is simply the restriction of the Weil representation of $\mathrm{Sp}(K)$ on $L^2(A)$ to $\mathrm{Sp}_R(K)$. All the theorems and proofs in this article concerning finite abelian p-groups generalize to the Weil representation of $\mathrm{Sp}_R(K)$ on $L^2(A)$, so long as p is replaced by q in the formulas. Since every finitely generated torsion module over a Dedekind domain is a product of its primary components, and module automorphisms respect the primary decomposition, the reduction in Section 2C works for finite modules of odd order over Dedekind domains.

Singla [2010; 2011] has proved that the representation theory of $G(R/P^2)$, where G is a classical group, depends on R only through q, the order of the residue field. More precisely, if R and R' are two discrete valuation rings and an isomorphism between their residue fields is fixed (for example, take $R = \mathbb{Z}_p$, the ring of p-adic integers, and $R' = (\mathbb{Z}/p\mathbb{Z})[[t]]$, the ring of Laurent series with coefficients in $\mathbb{Z}/p\mathbb{Z}$), then there is a canonical bijection between the irreducible representations of $G(R/P^2)$ and $G(R'/P^2)$ which preserves dimensions. There is also a canonical bijection between their conjugacy classes which preserves sizes. All existing evidence points towards the existence of a similar correspondence for automorphism groups of modules of type λ (see, for example [Onn 2008, Conjecture 1.2]). The results in this paper also point in the same direction: for each partition λ, there is a canonical correspondence between the invariant subspaces of the Weil representations associated to the finitely generated torsion R-module of type λ and the finitely generated torsion R'-module of type λ which preserves dimensions.

Acknowledgment

We thank Uri Onn for helpful comments on a preliminary version of this paper. We thank the referee for suggestions to improve the readability of this paper.

References

[Adler and Ramanan 1996] A. Adler and S. Ramanan, *Moduli of abelian varieties*, Lecture Notes in Math. **1644**, Springer, Berlin, 1996. MR 2000b:14057 Zbl 0863.14024

[Aubert et al. 2010] A.-M. Aubert, U. Onn, A. Prasad, and A. Stasinski, "On cuspidal representations of general linear groups over discrete valuation rings", *Israel J. Math.* **175** (2010), 391–420. MR 2011h:20089 Zbl 1201.22012

[Cliff et al. 2000] G. Cliff, D. McNeilly, and F. Szechtman, "Weil representations of symplectic groups over rings", *J. London Math. Soc.* (2) **62**:2 (2000), 423–436. MR 2001i:20088 Zbl 1037.20044

[Cliff et al. 2003] G. Cliff, D. McNeilly, and F. Szechtman, "Clifford and Mackey theory for Weil representations of symplectic groups", *J. Algebra* **262**:2 (2003), 348–379. MR 2004h:20060 Zbl 1143.20307

[Deligne and Lusztig 1976] P. Deligne and G. Lusztig, "Representations of reductive groups over finite fields", *Ann. of Math.* (2) **103**:1 (1976), 103–161. MR 52 #14076 Zbl 0336.20029

[Dutta and Prasad 2011] K. Dutta and A. Prasad, "Degenerations and orbits in finite abelian groups", *J. Combin. Theory Ser. A* **118**:6 (2011), 1685–1694. MR 2012c:20156 Zbl 1229.20053

[Gérardin 1975] P. Gérardin, *Construction de séries discrètes p-adiques: sur les séries discrètes non ramifiées des groupes réductifs déployés p-adiques*, Lecture Notes in Math. **462**, Springer, Berlin, 1975. MR 53 #719 Zbl 0302.22002

[Gérardin 1977] P. Gérardin, "Weil representations associated to finite fields", *J. Algebra* **46**:1 (1977), 54–101. MR 57 #470 Zbl 0359.20008

[Green 1955] J. A. Green, "The characters of the finite general linear groups", *Trans. Amer. Math. Soc.* **80** (1955), 402–447. MR 17,345e Zbl 0068.25605

[Hill 1995a] G. Hill, "Regular elements and regular characters of $\mathrm{GL}_n(\mathcal{O})$", *J. Algebra* **174**:2 (1995), 610–635. MR 96i:20055 Zbl 0839.22021

[Hill 1995b] G. Hill, "Semisimple and cuspidal characters of $\mathrm{GL}_n(\mathcal{O})$", *Comm. Algebra* **23**:1 (1995), 7–25. MR 96c:22030 Zbl 0843.22022

[Kloosterman 1946] H. D. Kloosterman, "The behaviour of general theta functions under the modular group and the characters of binary modular congruence groups, I", *Ann. of Math.* (2) **47** (1946), 317–375. MR 9,12g Zbl 0063.03262

[Mackey 1949] G. W. Mackey, "A theorem of Stone and von Neumann", *Duke Math. J.* **16** (1949), 313–326. MR 11,10b Zbl 0036.07703

[Maktouf and Torasso 2012] K. Maktouf and P. Torasso, "Restriction de la représentation de Weil à un sous-groupe compact maximal", preprint, 2012. The same content can be found as part of arXiv preprint 1101.0560.

[Maktouf and Torasso 2014] K. Maktouf and P. Torasso, "Restriction de la représentation de Weil à un sous-groupe compact maximal", 2014, available at http://mathsoc.jp/publication/JMSJ/pdf/JMSJ6619.pdf. To appear in *J. Math. Soc. Japan*.

[Mœglin et al. 1987] C. Mœglin, M.-F. Vignéras, and J.-L. Waldspurger, *Correspondances de Howe sur un corps p-adique*, Lecture Notes in Math. **1291**, Springer, Berlin, 1987. MR 91f:11040 Zbl 0642.22002

[von Neumann 1931] J. von Neumann, "Die Eindeutigkeit der Schrödingerschen Operatoren", *Math. Ann.* **104**:1 (1931), 570–578. MR 1512685 Zbl 0001.24703

[Nielsen and Chuang 2000] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, 2000. MR 2003j:81038 Zbl 1049.81015

[Onn 2008] U. Onn, "Representations of automorphism groups of finite $\mathfrak{o}$-modules of rank two", *Adv. Math.* **219**:6 (2008), 2058–2085. MR 2009j:20061 Zbl 1185.20049

[Prasad 1998] D. Prasad, "A brief survey on the theta correspondence", pp. 171–193 in *Number theory* (Tiruchirapalli, 1996), edited by V. Kumar Murty and M. Waldschmidt, Contemp. Math. **210**, American Mathematical Society, Providence, RI, 1998. MR 99e:11063 Zbl 0922.11041

[Prasad 2009] A. Prasad, "On character values and decomposition of the Weil representation associated to a finite abelian group", *J. Anal.* **17** (2009), 73–85. MR 2012a:11053 Zbl 05924869

[Prasad 2011] A. Prasad, "An easy proof of the Stone–von Neumann–Mackey theorem", *Expo. Math.* **29**:1 (2011), 110–118. MR 2012f:22003 Zbl 1227.43009 arXiv 0912.0574

[Prasad et al. 2010] A. Prasad, I. Shapiro, and M. K. Vemuri, "Locally compact abelian groups with symplectic self-duality", *Adv. Math.* **225**:5 (2010), 2429–2454. MR 2012d:22008 Zbl 1245.22008

[Singla 2010] P. Singla, "On representations of general linear groups over principal ideal local rings of length two", *J. Algebra* **324**:9 (2010), 2543–2563. MR 2011h:20094 Zbl 1216.20035

[Singla 2011] P. Singla, "On representations of classical groups over finite local rings of length two", preprint, 2011. arXiv 1101.3696

[Srinivasan 1979] B. Srinivasan, "Weil representations of finite classical groups", *Invent. Math.* **51**:2 (1979), 143–153. MR 81d:20036 Zbl 0399.20040

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

[Stasinski 2009] A. Stasinski, "The smooth representations of $\mathrm{GL}_2(\mathfrak{o})$", *Comm. Algebra* **37**:12 (2009), 4416–4430. MR 2011f:20101 Zbl 1182.22008

[Stone 1930] M. H. Stone, "Linear transformations in Hilbert space, III: Operational methods and group theory", *Proc. Natl. Acad. Sci. USA* **16** (1930), 172–175. Zbl 56.0357.01

[Tanaka 1967a] S. Tanaka, "Construction and classification of irreducible representations of special linear group of the second order over a finite field", *Osaka J. Math.* **4**:1 (1967), 65–84. MR 36 #2714 Zbl 0245.20039

[Tanaka 1967b] S. Tanaka, "Irreducible representations of the binary modular congruence groups mod p^λ", *J. Math. Kyoto Univ.* **7**:2 (1967), 123–132. MR 37 #5311 Zbl 0219.20005

[Tate 1967] J. T. Tate, "Fourier analysis in number fields, and Hecke's zeta-functions", pp. 305–347 in *Algebraic number theory* (Brighton, 1965), edited by J. W. S. Cassels and A. Fröhlich, Academic Press, London, 1967. A second ed. with errata was published by the London Mathematical Society in 2010. MR 36 #121

[Weil 1964] A. Weil, "Sur certains groupes d'opérateurs unitaires", *Acta Math.* **111** (1964), 143–211. MR 29 #2324 Zbl 0203.03305

[Weyl 1949] H. Weyl, *The theory of groups and quantum mechanics*, Dover, New York, 1949. MR 56 #8744 Zbl 0041.56804

Received May 15, 2013. Revised May 21, 2014.

KUNAL DUTTA
MAX-PLANCK-INSTITUT FÜR INFORMATIK
D-66123 SAARBRÜCKEN
GERMANY
kdutta@mpi-inf.mpg.de

AMRITANSHU PRASAD
INSTITUTE OF MATHEMATICAL SCIENCES
CIT CAMPUS
CHENNAI 600113
INDIA
amri@imsc.res.in

PACIFIC JOURNAL OF MATHEMATICS
Vol. 275, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.275.325

COMPACT ANTI-DE SITTER 3-MANIFOLDS AND FOLDED HYPERBOLIC STRUCTURES ON SURFACES

FRANÇOIS GUÉRITAUD, FANNY KASSEL AND MAXIME WOLFF

We prove that any non-Fuchsian representation ρ of a surface group into $\mathrm{PSL}(2,\mathbb{R})$ is the holonomy of a folded hyperbolic structure on the surface, unless the image of ρ is virtually abelian. Using this idea, we establish that any non-Fuchsian representation ρ is strictly dominated by some Fuchsian representation j, in the sense that the hyperbolic translation lengths for j are uniformly larger than for ρ. Conversely, any Fuchsian representation j strictly dominates some non-Fuchsian representation ρ, whose Euler class can be prescribed. This has applications to the theory of compact anti-de Sitter 3-manifolds.

1. Introduction

Let Σ_g be a closed, connected, oriented surface of genus g, with fundamental group $\Gamma_g = \pi_1(\Sigma_g)$, and let $\mathrm{Rep}_g^{\mathrm{fd}}$ and $\mathrm{Rep}_g^{\mathrm{nfd}}$ be the sets of conjugacy classes of Fuchsian and non-Fuchsian representations of Γ_g into $\mathrm{PSL}(2,\mathbb{R})$, respectively. The letters "fd" stand for "faithful, discrete". By work of Goldman [1988], the space $\mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))$ of representations of Γ_g into $\mathrm{PSL}(2,\mathbb{R})$ has $4g-3$ connected components, indexed by the values of the Euler class

$$\mathsf{eu} : \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R})) \longrightarrow \{2-2g, \ldots, -1, 0, 1, \ldots, 2g-2\}.$$

In the quotient, $\mathrm{Rep}_g^{\mathrm{fd}}$ consists of the two connected components of extremal Euler class, and $\mathrm{Rep}_g^{\mathrm{nfd}}$ of all the other components of $\mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))/\mathrm{PSL}(2,\mathbb{R})$.

1A. ***Strictly dominating representations.*** For any $g \in \mathrm{PSL}(2,\mathbb{R})$, let

$$\lambda(g) := \inf_{p\in\mathbb{H}^2} d(p, g\cdot p) \geq 0 \tag{1-1}$$

be the translation length of g in the hyperbolic plane $\mathbb{H}^2$. This defines a function

The authors were partially supported by the Agence Nationale de la Recherche under the grants DiscGroup (ANR-11-BS01-013), ETTT (ANR-09-BLAN-0116-01), ModGroup (ANR-11-BS01-0020), SGT (ANR-11-BS01-0018), and through the Labex CEMPI (ANR-11-LABX-0007-01).
MSC2010: 20H10, 32G15, 53C50.
Keywords: representations of surface groups, folded hyperbolic structures, anti-de Sitter 3-manifolds.

$\lambda : \mathrm{PSL}(2,\mathbb{R}) \to \mathbb{R}^+$ which is invariant under conjugation. We say that an element $[j] \in \mathsf{Rep}_g^{\mathrm{fd}}$ *strictly dominates* an element $[\rho] \in \mathsf{Rep}_g^{\mathrm{nfd}}$ if

$$\sup_{\gamma \in \Gamma_g \smallsetminus \{1\}} \frac{\lambda(\rho(\gamma))}{\lambda(j(\gamma))} < 1. \tag{1-2}$$

Note that (1-2) can never hold when j and ρ are both Fuchsian [Thurston 1986]. In this paper we prove the following:

Theorem 1.1. *Any $[\rho] \in \mathsf{Rep}_g^{\mathrm{nfd}}$ is strictly dominated by some $[j] \in \mathsf{Rep}_g^{\mathrm{fd}}$. Any $[j] \in \mathsf{Rep}_g^{\mathrm{fd}}$ strictly dominates some $[\rho] \in \mathsf{Rep}_g^{\mathrm{nfd}}$, whose Euler class can be prescribed.*

The first statement of Theorem 1.1 has been simultaneously and independently obtained by Deroin and Tholozan [2013] using more analytical methods. Their paper deals, more generally, with representations of Γ_g into the isometry group of any complete, simply connected Riemannian manifold with sectional curvature at most -1. They also announce a version for general CAT(-1) spaces. The present methods, relying as they do on the Toponogov theorem (see Lemma 2.2 below), could likely extend to this general setting as well.

Our approach is constructive, using folded (or pleated) hyperbolic surfaces, as we now explain.

1B. *Folded hyperbolic surfaces.* Pleated hyperbolic surfaces were introduced by Thurston [1980] and play an important role in the theory of hyperbolic 3-manifolds. A *folded hyperbolic surface* is a pleated surface with all angles equal to 0 or π, whose holonomy takes values in $\mathrm{PSL}(2,\mathbb{R})$ (see Section 2B). It is easy to check (see [Thurston 1986, Proposition 2.1]) that the holonomy of a (nontrivially) folded hyperbolic structure on Σ_g belongs to $\mathsf{Rep}_g^{\mathrm{nfd}}$. In order to establish Theorem 1.1, we prove that the converse holds for representations whose image is not virtually abelian.

Theorem 1.2. *An element of $\mathsf{Rep}_g^{\mathrm{nfd}}$ is the holonomy of a folded hyperbolic structure on Σ_g if and only if its image is not virtually abelian.*

As usual, being *virtually abelian* means having an abelian subgroup of finite index. Besides abelian representations, Theorem 1.2 rules out dihedral representations, which preserve a geodesic line of $\mathbb{H}^2$ and contain order-two symmetries of that line.

This result seems to have been known to experts since the work of Thurston [1980], but to our knowledge it is neither stated nor proved in the literature. Note that another type of folded hyperbolic structure was previously investigated by Goldman [1987].

We construct the folded hyperbolic structures of Theorem 1.2 explicitly, folding along geodesic laminations that are the union of simple closed curves and of maximal laminations of some pairs of pants (Proposition 3.1). More precisely, given a non-Fuchsian representation ρ whose image is not virtually abelian, we use a result of Gallo, Kapovich, and Marden [Gallo et al. 2000] to find a pants decomposition of Σ_g such that the restriction of ρ to any pair of pants P is nonabelian and maps any cuff to a hyperbolic element. (The term *cuff*, always specific to a pair of pants, will in the sequel denote without distinction the homotopy class of a boundary component, or the geodesic in that class, or its length.) Folding along a certain maximal lamination in P then gives a simple dictionary between the representations of the fundamental group of P that have Euler class 0 and those that have Euler class ± 1 (Lemma 3.6). The converse direction in Theorem 1.2 is elementary (Observation 2.7).

1C. ***Idea of the proof of Theorem 1.1.*** If $[\rho] \in \mathrm{Rep}_g^{\mathrm{nfd}}$ is the holonomy of a folded hyperbolic structure on Σ_g, then the holonomy $[j_0] \in \mathrm{Rep}_g^{\mathrm{fd}}$ of the corresponding unfolded hyperbolic structure clearly dominates $[\rho]$ in the sense that $\lambda(\rho(\gamma)) \leq \lambda(j_0(\gamma))$ for all $\gamma \in \Gamma_g$. In fact,

$$\sup_{\gamma \in \Gamma_g \smallsetminus \{1\}} \frac{\lambda(\rho(\gamma))}{\lambda(j_0(\gamma))} = 1,$$

since any minimal component of the folding lamination can be approximated by simple closed curves. In order to prove Theorem 1.1 we need to make the domination *strict*.

To establish the first statement, the idea is, for $[\rho] \in \mathrm{Rep}_g^{\mathrm{nfd}}$, to consider the holonomy $[j_0] \in \mathrm{Rep}_g^{\mathrm{fd}}$ of the unfolded hyperbolic structure given by Theorem 1.2, and to lengthen the closed curves (close to being) contained in the folding lamination while simultaneously not shortening the other curves too much. To do this, we work independently in each "folded subsurface" of Σ_g, which is a compact surface with boundary endowed with a hyperbolic structure induced by j_0. In each such subsurface we use a *strip deformation* construction due to Thurston [1986], which consists in adding hyperbolic strips to obtain a new hyperbolic metric with longer boundary components. We then glue back along the boundary components, after making sure that the lengths agree.

The second statement is easier in that it does not rely on Theorem 1.2. Starting with an element $[j] \in \mathrm{Rep}_g^{\mathrm{fd}}$, we choose a pants decomposition of Σ_g along which to fold. To make sure that the cuffs of the pairs of pants will get contracted, we first deform j slightly by *negative strip deformations* into another element $[j_0] \in \mathrm{Rep}_g^{\mathrm{fd}}$ with shorter cuffs, in such a way that the other curves do not get much longer. Folding j_0 then gives an element $[\rho] \in \mathrm{Rep}_g^{\mathrm{nfd}}$ which is strictly dominated by $[j]$.

1D. ***An application to compact anti-de Sitter* 3*-manifolds.*** Theorem 1.1 has consequences for the theory of compact *anti-de Sitter* 3-manifolds. These are the compact Lorentzian 3-manifolds of constant negative curvature, i.e., the Lorentzian analogues of the compact hyperbolic 3-manifolds. They are locally modeled on the 3-dimensional anti-de Sitter space

$$\mathrm{AdS}^3 = \mathrm{PO}(2,2)/\,\mathrm{PO}(2,1),$$

which is identified with $\mathrm{PSL}(2,\mathbb{R})$ endowed with the natural Lorentzian structure induced by the Killing form of its Lie algebra. The identity component of the isometry group of AdS^3 is $\mathrm{PSL}(2,\mathbb{R}) \times \mathrm{PSL}(2,\mathbb{R})$, acting on $\mathrm{PSL}(2,\mathbb{R}) \simeq \mathrm{AdS}^3$ by right and left multiplication: $(g_1, g_2) \cdot g = g_2 g g_1^{-1}$. All compact anti-de Sitter 3-manifolds are geodesically complete [Klingler 1996]. By [Kulkarni and Raymond 1985] and the Selberg lemma [1960, Lemma 8], they are quotients of $\mathrm{PSL}(2,\mathbb{R})$ by torsion-free discrete subgroups $\mathbf{\Gamma}$ of $\mathrm{PSL}(2,\mathbb{R}) \times \mathrm{PSL}(2,\mathbb{R})$ acting properly discontinuously, up to a finite covering; moreover, the groups $\mathbf{\Gamma}$ are graphs of the form

$$\mathbf{\Gamma} = (\Gamma_g)^{j,\rho} := \{(j(\gamma), \rho(\gamma)) \mid \gamma \in \Gamma_g\}$$

for some $g \geq 2$, where j, $\rho \in \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))$ are representations and j is Fuchsian, up to switching the two factors of $\mathrm{PSL}(2,\mathbb{R}) \times \mathrm{PSL}(2,\mathbb{R})$. In particular, $\mathbf{\Gamma} \backslash \mathrm{AdS}^3$ is Seifert fibered over a hyperbolic base (see [Salein 1999, §3.4.2]).

Following [Salein 2000], we shall say that a pair $(j,\rho) \in \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))^2$ with j Fuchsian is *admissible* if the action of $(\Gamma_g)^{j,\rho}$ on AdS^3 is properly discontinuous. Note that (j,ρ) is admissible if and only if its conjugates under $\mathrm{PSL}(2,\mathbb{R}) \times \mathrm{PSL}(2,\mathbb{R})$ are. Therefore, in order to understand the moduli space of compact anti-de Sitter 3-manifolds, we need to understand, for any $g \geq 2$, the space

$$\mathrm{Adm}_g \subset \mathsf{Rep}_g^{\mathrm{fd}} \times \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))/\,\mathrm{PSL}(2,\mathbb{R})$$

of conjugacy classes of admissible pairs (j,ρ) with j Fuchsian.

Examples of admissible pairs are readily obtained by taking ρ to be trivial, or more generally with bounded image. The corresponding quotients of AdS^3 are called *standard*. The first nonstandard examples were constructed by Goldman [1985] by deformation of standard ones — a technique later generalized by Kobayashi [1998]. Salein [2000] constructed the first examples of admissible pairs (j,ρ) with $\mathsf{eu}(\rho) \neq 0$. He actually constructed examples where $\mathsf{eu}(\rho)$ can take any nonextremal value. A necessary and sufficient condition for admissibility was given in [Kassel 2009]: a pair (j,ρ) with j Fuchsian is admissible if and only if ρ is strictly dominated by j in the sense of (1-2). In particular, by [Thurston 1986],

$$\mathrm{Adm}_g \subset \mathsf{Rep}_g^{\mathrm{fd}} \times \mathsf{Rep}_g^{\mathrm{nfd}}.$$

This properness criterion was extended in [Guéritaud and Kassel 2013] to quotients of $\mathrm{PO}(n,1) = \mathrm{Isom}(\mathbb{H}^n)$ by discrete subgroups of $\mathrm{PO}(n,1) \times \mathrm{PO}(n,1)$ acting by left and right multiplication, for arbitrary $n \geq 2$ (recall that $\mathrm{PSL}(2,\mathbb{R}) \simeq \mathrm{PO}(2,1)_0$), and in [Guéritaud et al. 2015] to quotients of any simple Lie group G of real rank 1.

By completeness of compact anti-de Sitter manifolds [Klingler 1996], the Ehresmann–Thurston principle (see [Thurston 1980]) implies that Adm_g is open in $\mathsf{Rep}_g^{\mathrm{fd}} \times \mathsf{Rep}_g^{\mathrm{nfd}}$. Moreover, Adm_g has at least $4g-5$ connected components, as Salein's examples show. Using the fact that the two connected components of $\mathsf{Rep}_g^{\mathrm{fd}}$ are conjugate under $\mathrm{PGL}(2,\mathbb{R})$, we can reformulate Theorem 1.1 as follows:

Corollary 1.3. *The projections of* Adm_g *to* $\mathsf{Rep}_g^{\mathrm{fd}}$ *and to* $\mathsf{Rep}_g^{\mathrm{nfd}}$ *are both surjective. Moreover, for any connected components* $\mathscr{C}_1$ *of* $\mathsf{Rep}_g^{\mathrm{fd}}$ *and* $\mathscr{C}_2$ *of* $\mathsf{Rep}_g^{\mathrm{nfd}}$*, the projections of* $\mathrm{Adm}_g \cap (\mathscr{C}_1 \times \mathscr{C}_2)$ *to* $\mathscr{C}_1$ *and to* $\mathscr{C}_2$ *are both surjective.*

The topology of Adm_g is still unknown, but we believe that Corollary 1.3 (and the ideas behind its proof) could be used to prove that Adm_g is homeomorphic to $\mathsf{Rep}_g^{\mathrm{fd}} \times \mathsf{Rep}_g^{\mathrm{nfd}}$. Using the work of Hitchin [1987, Theorem 10.8 and Equation 10.6], this would give the homeomorphism type of the connected components of Adm_g corresponding to $\mathsf{eu}(\rho) \neq 0$.

Furthermore, it would be interesting to obtain a geometric and combinatorial description of the fibers of the second projection $\mathrm{Adm}_g \to \mathsf{Rep}_g^{\mathrm{nfd}}$. Such a description is given in [Danciger et al. 2014], in terms of the arc complex, in the different case that j and ρ are the holonomies of two convex cocompact hyperbolic structures on a given *noncompact* surface.

1E. *Organization of the paper.* In Section 2 we recall some facts about Lipschitz maps, folded hyperbolic structures, and the Euler class. Section 3 is devoted to the proof of Theorem 1.2, and Section 4 to that of Theorem 1.1.

2. Reminders and useful facts

2A. *Lipschitz maps and their stretch locus.* In the whole paper, we denote by d the metric on the real hyperbolic plane $\mathbb{H}^2$. For a Lipschitz map $f : \mathbb{H}^2 \to \mathbb{H}^2$ and a point $p \in \mathbb{H}^2$, we set:

- $\mathrm{Lip}(f) := \sup_{q \neq q'} d(f(q), f(q'))/d(q,q') \geq 0$ (the Lipschitz constant);
- $\mathrm{Lip}_p(f) := \inf_{\mathscr{U}} \mathrm{Lip}(f|_{\mathscr{U}}) \geq 0$, where $\mathscr{U}$ ranges over all neighborhoods of p in $\mathbb{H}^2$ (the local Lipschitz constant).

The function $p \mapsto \mathrm{Lip}_p(f)$ is upper semicontinuous:

$$\mathrm{Lip}_p(f) \geq \limsup_{n \to +\infty} \mathrm{Lip}_{p_n}(f)$$

for any sequence $(p_n)_{n \in \mathbb{N}}$ converging to p. The following is straightforward:

Remark 2.1. For any rectifiable path $\mathscr{L} \subset \mathbb{H}^2$,

$$\mathrm{length}(f(\mathscr{L})) \leq \sup_{p\in\mathscr{L}} \mathrm{Lip}_p(f) \cdot \mathrm{length}(\mathscr{L}).$$

In particular, if $\mathrm{Lip}_p(f) \leq C$ for all p in a convex set K, then $\mathrm{Lip}(f|_K) \leq C$.

2A1. *The stretch locus.* The following result is a particular case of [Guéritaud and Kassel 2013, Theorem 5.1]. It relies on the Toponogov theorem, a comparison theorem relating the curvature to the divergence rate of geodesics (see [Bridson and Haefliger 1999, Lemma II.1.13]).

Lemma 2.2 [Guéritaud and Kassel 2013]. *Let Γ be a torsion-free, finitely generated, discrete group and let $(j, \rho) \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))^2$ be a pair of representations with j convex cocompact. Suppose the infimum of Lipschitz constants for all (j, ρ)-equivariant maps $f : \mathbb{H}^2 \to \mathbb{H}^2$ is* 1, *and the space $\mathscr{F}$ of maps achieving this infimum is nonempty. Then there exists a nonempty, $j(\Gamma)$-invariant geodesic lamination $\widetilde{\Lambda}$ of $\mathbb{H}^2$ such that*:

- *any leaf of $\widetilde{\Lambda}$ is isometrically preserved by all maps $f \in \mathscr{F}$*;
- *any connected component of $\mathbb{H}^2 \smallsetminus \widetilde{\Lambda}$ is either isometrically preserved by all $f \in \mathscr{F}$, or consists entirely of points p at which $\mathrm{Lip}_p(f) < 1$ for some $f \in \mathscr{F}$* (*independent of p*).

Definition 2.3. The union of $\widetilde{\Lambda}$ and of the connected components of $\mathbb{H}^2 \smallsetminus \widetilde{\Lambda}$ that are isometrically preserved by all $f \in \mathscr{F}$ is called the *stretch locus* of (j, ρ).

By *convex cocompact* we mean that j is injective and discrete and that the group $j(\Gamma)$ does not contain any parabolic element. By (j, ρ)*-equivariant* we mean that $f(j(\gamma) \cdot p) = \rho(\gamma) \cdot f(p)$ for all $\gamma \in \Gamma$ and $p \in \mathbb{H}^2$. The space $\mathscr{F}$ is always nonempty, except possibly if $\rho(\Gamma)$ admits a unique fixed point in the boundary at infinity $\partial_\infty \mathbb{H}^2$ of $\mathbb{H}^2$ [Guéritaud and Kassel 2013, Lemma 4.11]. If j and ρ are conjugate under $\mathrm{PGL}(2, \mathbb{R})$, then the stretch locus of (j, ρ) is the preimage of the convex core of $j(\Gamma)\backslash\mathbb{H}^2$. (This preimage is by definition the smallest nonempty $j(\Gamma)$-invariant closed convex subset of $\mathbb{H}^2$.)

2A2. *Averaging Lipschitz maps.* We now describe a technical tool for understanding the stretch locus. It is a procedure for averaging Lipschitz maps (see [Guéritaud and Kassel 2013, §2.5]), under which Lip_p behaves as it would for the barycenter of maps between affine Euclidean spaces. In Section 3D, we shall use this procedure with a partition of unity, as follows.

Let $\psi_0, \dots, \psi_n : \mathbb{H}^2 \to [0, 1]$ be Lipschitz functions inducing a partition of unity on a subset X of $\mathbb{H}^2$, subordinated to an open covering $B_0 \cup \cdots \cup B_n \supset X$. For $0 \leq i \leq n$, let $\varphi_i : B_i \to \mathbb{H}^2$ be a Lipschitz map. For $p \in X$, let $I(p)$ be the collection of indices i such that $p \in B_i$. Let $\sum_{i=0}^{n} \psi_i \varphi_i : X \to \mathbb{H}^2$ be the map sending any $p \in X$ to the minimizer in $\mathbb{H}^2$ of

$$\sum_{i \in I(p)} \psi_i(p) d(\cdot, \varphi_i(p))^2.$$

Then the following holds:

Lemma 2.4 [Guéritaud and Kassel 2013, Lemma 2.13]. *The averaged map $\varphi := \sum_{i=0}^{n} \psi_i \varphi_i$ satisfies the "Leibniz rule"*

$$\mathrm{Lip}_p(\varphi) \leq \sum_{i \in I(p)} \left(\mathrm{Lip}_p(\psi_i) R(p) + \psi_i(p) \, \mathrm{Lip}_p(\varphi_i)\right)$$

for all $p \in X$, where $R(p)$ is the diameter of the set $\{\varphi_i(p) \mid i \in I(p)\}$.

2A3. *Admissibility.* For any discrete group Γ (not necessarily of the form Γ_g), we say that a pair of representations $(j, \rho) \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))^2$ is *admissible* if the group $\Gamma^{j,\rho} = \{(j(\gamma), \rho(\gamma)) \mid \gamma \in \Gamma\}$ acts properly discontinuously on AdS^3. In this case, at least one of j or ρ is injective and discrete [Kassel 2008].

Understanding the stretch locus has led to the following necessary and sufficient conditions for admissibility. We denote by Γ_s the set of nontrivial elements of Γ corresponding to simple closed curves on the surface $j(\Gamma)\backslash\mathbb{H}^2$.

Theorem 2.5 [Kassel 2009; Guéritaud and Kassel 2013]. *Let Γ be a torsion-free, finitely generated, discrete group and let $j, \rho \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))$ be two representations with j injective and discrete. The pair (j, ρ) is admissible if and only if the following condition holds up to switching j and ρ:*

(i) *There exists a (j, ρ)-equivariant map $f : \mathbb{H}^2 \to \mathbb{H}^2$ with* $\mathrm{Lip}(f) < 1$.

If j is convex cocompact or if the group $\rho(\Gamma)$ does not have a unique fixed point in $\partial_\infty \mathbb{H}^2$, then (i) *is equivalent to either of the following two conditions:*

(ii) *The representation ρ is strictly dominated by j:*

$$\sup_{\substack{\gamma \in \Gamma, \\ \lambda(j(\gamma)) > 0}} \frac{\lambda(\rho(\gamma))}{\lambda(j(\gamma))} < 1;$$

(iii) *The representation ρ is strictly dominated by j in restriction to simple closed curves:*

$$\sup_{\gamma \in \Gamma_s} \frac{\lambda(\rho(\gamma))}{\lambda(j(\gamma))} < 1.$$

The implication (iii) $\Rightarrow$ (i) is nontrivial and relies on Lemma 2.2. The implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) are immediate modulo the following easy remark (see [Guéritaud and Kassel 2013, Lemma 4.5]):

Remark 2.6. Let Γ be a discrete group and $(j, \rho) \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))^2$ a pair of representations. For any $\gamma \in \Gamma$ and any (j, ρ)-equivariant Lipschitz map $f : \mathbb{H}^2 \to \mathbb{H}^2$,

$$\lambda(\rho(\gamma)) \leq \mathrm{Lip}(f)\, \lambda(j(\gamma)).$$

2B. ***Pleated and folded hyperbolic structures.*** Let Σ be a connected, oriented surface of negative Euler characteristic, possibly with boundary, and let $\Gamma = \pi_1(\Sigma)$ be its fundamental group. Recall from [Bonahon 1996, §7] that a *pleated hyperbolic structure* on Σ is a quadruple (j, ρ, Υ, f) where:

- $j \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))$ is the holonomy of a hyperbolic structure on Σ;
- $\rho \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2, \mathbb{C}))$ is a representation;
- Υ is a geodesic lamination on Σ;
- $f : \mathbb{H}^2 \to \mathbb{H}^3$ is a (j, ρ)-equivariant, continuous map whose restriction to any connected component of $\mathbb{H}^2 \smallsetminus \widetilde{\Upsilon}$ is an isometric embedding. (Here we denote by $\widetilde{\Upsilon}$ the preimage in $\mathbb{H}^2$ of $\Upsilon \subset \Sigma \simeq j(\Gamma)\backslash\mathbb{H}^2$.)

The representation ρ is called the *holonomy* of the pleated hyperbolic structure. The closures of the connected components of $\mathbb{H}^2 \smallsetminus \widetilde{\Upsilon}$ are called the *plates*. Note that f is 1-Lipschitz. For any $g, h \in \mathrm{PGL}(2, \mathbb{R})$,

$$(g j(\cdot) g^{-1}, h\rho(\cdot)h^{-1}, \Upsilon, h \circ f \circ g^{-1})$$

is still a pleated hyperbolic structure on Σ.

Observation 2.7. *Suppose that Σ is compact. If (j, ρ, Υ, f) is a pleated hyperbolic structure on Σ, then the group $\rho(\Gamma)$ is not virtually abelian.*

Proof. We see Σ as the convex core of the hyperbolic surface $j(\Gamma)\backslash\mathbb{H}^2$. Consider a nondegenerate ideal triangle T of $\mathbb{H}^2$ which is entirely contained in the intersection of one plate with the preimage of Σ in $\mathbb{H}^2$. Let $(p_n)_{n\in\mathbb{N}}$ be a sequence of points of T going to infinity. Since Σ is compact, there exist $R > 0$ and a sequence $(\gamma_n)_{n\in\mathbb{N}}$ of elements of Γ such that $d(j(\gamma_n) \cdot p_0, p_n) \leq R$ for all $n \in \mathbb{N}$. Since f is (j, ρ)-equivariant and 1-Lipschitz,

$$d(\rho(\gamma_n) \cdot f(p_0), f(p_n)) \leq d(j(\gamma_n) \cdot p_0, p_n) \leq R$$

for all $n \in \mathbb{N}$. Applying this to sequences (p_n) converging to the three ideal vertices of T, and using the fact that the restriction of f to T is an isometry, we see that the limit set of $\rho(\Gamma)$ contains at least three points. In particular, $\rho(\Gamma)$ is not virtually abelian. $\square$

We shall also use the following elementary remark:

Remark 2.8. Let (j, ρ, Υ, f) be a pleated hyperbolic structure on Σ. If some leaf of Υ spirals to a boundary component of Σ corresponding to an element $\gamma \in \Gamma$, then $\lambda(j(\gamma)) = \lambda(\rho(\gamma))$, where $\lambda : \mathrm{PSL}(2, \mathbb{C}) \to \mathbb{R}^+$ is the translation length function in $\mathbb{H}^3$ extending (1-1).

Any pleated hyperbolic structure (j, ρ, Υ, f) on Σ defines a *bending cocycle*, i.e., a map β from the set of pairs of plates to $\mathbb{R}/2\pi\mathbb{Z}$ which is symmetric and additive:

$$\beta(P, Q) = \beta(Q, P) \quad \text{and} \quad \beta(P, Q) + \beta(Q, R) = \beta(P, R)$$

for all plates P, Q, R. Intuitively, $\beta(P, Q)$ is the total angle of pleating encountered when traveling from $f(P)$ to $f(Q)$ along $f(\mathbb{H}^2)$ in $\mathbb{H}^3$. Conversely, to any bending cocycle, Bonahon [1996, §8] associates a pleated surface.

In this paper we consider a special case of pleated surfaces (j, ρ, Υ, f), namely those for which f takes values in a copy of $\mathbb{H}^2$ inside $\mathbb{H}^3$ (i.e., in a totally geodesic plane) and ρ takes values in $\mathrm{Isom}^+(\mathbb{H}^2) = \mathrm{PSL}(2, \mathbb{R})$. In this case, we speak of a *folded hyperbolic structure* on Σ, and say that ρ is a *folding* of j. The map f defines a *coloring* of $\Sigma \smallsetminus \Upsilon$, i.e., a $j(\Gamma)$-invariant function $\tilde{c}$ from the set of plates to $\{-1, 1\}$. Namely, we set $\tilde{c}(P) = -1$ if the restriction of f to P is orientation-preserving, and $\tilde{c}(P) = 1$ otherwise. Note that the bending cocycle of a folded hyperbolic structure is valued in $\{0, \pi\}$: for all plates P and Q,

$$\beta(P, Q) = \tfrac{1}{2}(1 - \tilde{c}(P)\tilde{c}(Q))\pi \in \{0, \pi\}. \tag{2-1}$$

The coloring $\tilde{c}$ descends to a continuous, locally constant function c from $\Sigma \smallsetminus \Upsilon$ to $\{-1, 1\}$. Conversely, any such function, after lifting to a coloring $\tilde{c}$ from the set of connected components of $\mathbb{H}^2 \smallsetminus \widetilde{\Upsilon}$ to $\{-1, 1\}$, defines a bending cocycle on $\mathbb{H}^2 \smallsetminus \widetilde{\Upsilon}$ by the formula (2-1). This bending cocycle, in turn, defines a folded hyperbolic structure on Σ, by the work of Bonahon [1996].

2C. ***The Euler class.*** We now give a brief introduction to the Euler class, along the lines of [Wolff 2011, §2.3.3]. For details and complements we refer to [Ghys 2001] or [Calegari 2004, §2].

As in the introduction, let Σ_g be a closed, connected, oriented surface of genus $g \geq 2$ with fundamental group Γ_g. The Euler class of a representation $\rho \in \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2, \mathbb{R}))$ measures the obstruction to lifting ρ to the universal cover $\widetilde{\mathrm{PSL}}(2, \mathbb{R})$ of $\mathrm{PSL}(2, \mathbb{R})$, and its parity measures the obstruction to lifting ρ to $\mathrm{SL}(2, \mathbb{R})$. To define the Euler class, choose a set-theoretic section s of the covering map $\widetilde{\mathrm{PSL}}(2, \mathbb{R}) \to \mathrm{PSL}(2, \mathbb{R})$. Consider a triangulation of Σ_g with a vertex at the basepoint x_0 defining $\Gamma_g = \pi_1(\Sigma_g, x_0)$, and choose an orientation on every edge of the triangulation. Choose a maximal tree in the 1-skeleton of the triangulation and, for every oriented edge σ in this tree, set $\rho(\sigma) := 1 \in \mathrm{PSL}(2, \mathbb{R})$. Any other oriented edge σ' corresponds (by closing up in the unique possible way along the rooted tree) to an element $\gamma \in \Gamma_g$, and we set $\rho(\sigma') := \rho(\gamma) \in \mathrm{PSL}(2,\mathbb{R})$. The boundary of any oriented triangle τ of the triangulation can be written as $\sigma_1^{\varepsilon_1}\sigma_2^{\varepsilon_2}\sigma_3^{\varepsilon_3}$, where $\sigma_1, \sigma_2, \sigma_3$ are edges with the chosen orientation and $\varepsilon_i \in \{\pm 1\}$. We set

$$\mathsf{eu}(\rho)(\tau) := s(\rho(\sigma_1))^{\varepsilon_1} s(\rho(\sigma_2))^{\varepsilon_2} s(\rho(\sigma_3))^{\varepsilon_3}.$$

Summing over triangles τ, this defines an element of $H^2\big(\Sigma_g, \pi_1(\mathrm{PSL}(2,\mathbb{R}))\big)$, hence an element of $H^2(\Sigma_g, \mathbb{Z})$ under the identification $\pi_1(\mathrm{PSL}(2,\mathbb{R})) \simeq \mathbb{Z}$. This element $\mathsf{eu}(\rho) \in H^2(\Sigma_g, \mathbb{Z})$ is called the *Euler class* of ρ. Its evaluation on the fundamental class in $H_2(\Sigma_g, \mathbb{Z})$ is an integer, which we still call the Euler class of ρ. It is invariant under conjugation by $\mathrm{PSL}(2,\mathbb{R})$, and changes sign under conjugation by $\mathrm{PGL}(2,\mathbb{R}) \smallsetminus \mathrm{PSL}(2,\mathbb{R})$.

We can also define the Euler class for representations of the fundamental group of a compact, connected, oriented surface Σ with boundary, of negative Euler characteristic, provided that the boundary curves are sent to hyperbolic elements. Indeed, any hyperbolic element $g \in \mathrm{PSL}(2,\mathbb{R})$ has a canonical lift to $\widetilde{\mathrm{PSL}}(2,\mathbb{R})$ because it belongs to a unique one-parameter subgroup of $\mathrm{PSL}(2,\mathbb{R})$, which defines a path from the identity to g. Choose a section s of the projection $\widetilde{\mathrm{PSL}}(2,\mathbb{R}) \to \mathrm{PSL}(2,\mathbb{R})$ that maps any hyperbolic element to its canonical lift. Then the construction above, using triangulations of Σ containing exactly one vertex on each boundary component, defines an Euler class, independent of all choices.

For instance, let Σ be an oriented pair of pants with fundamental group $\Gamma = \langle \alpha, \beta, \gamma \mid \alpha\beta\gamma = 1\rangle$, where α, β, γ correspond to the three boundary curves, endowed with the orientation induced by the surface. For any representation $\rho \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2,\mathbb{R}))$ with $\rho(\alpha)$, $\rho(\beta)$, $\rho(\gamma)$ hyperbolic,

$$\mathsf{eu}(\rho) = s(\rho(\alpha))\, s(\rho(\beta))\, s(\rho(\gamma)) \in Z(\widetilde{\mathrm{PSL}}(2,\mathbb{R})) \simeq \mathbb{Z}. \tag{2-2}$$

In particular, $\mathsf{eu}(\rho) \in \{-1, 0, 1\}$, and $|\mathsf{eu}(\rho)| = 1$ if and only if ρ is the holonomy of a hyperbolic structure on Σ, possibly after reversing the orientation. If s' is a section of the projection $\mathrm{SL}(2,\mathbb{R}) \to \mathrm{PSL}(2,\mathbb{R})$ that maps any hyperbolic element to its lift of positive trace, then (2-2) implies

$$s'(\rho(\alpha))\, s'(\rho(\beta))\, s'(\rho(\gamma)) = (-\mathrm{Id})^{\mathsf{eu}(\rho)}. \tag{2-3}$$

By construction, the Euler class is *additive*: if Σ is the union of two subsurfaces Σ' and Σ'' glued along curves γ_i, and if $\rho \in \mathrm{Hom}(\pi_1(\Sigma), \mathrm{PSL}(2,\mathbb{R}))$ is a representation sending all the curves γ_i (and the boundary curves of Σ, if any) to hyperbolic elements of $\mathrm{PSL}(2,\mathbb{R})$, then $\mathsf{eu}(\rho)$ is the sum of the Euler classes of the restrictions of ρ to the fundamental groups of Σ' and Σ''. This implies that a folded hyperbolic structure defined by a coloring c from the set $\mathscr{P}$ of connected components of

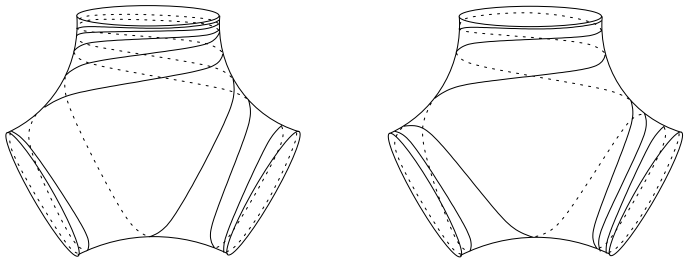


Figure 1. A pair of pants carries 24 maximal geodesic laminations containing a geodesic spiraling from a boundary component to itself (left), and 8 triskelion laminations (right).

$\Sigma \smallsetminus \Upsilon$ to $\{-1, 1\}$ has Euler class $\frac{1}{2\pi} \sum_{P \in \mathcal{P}} c(P)\mathcal{A}(P)$, where $\mathcal{A}(P)$ is the area of P.

We shall use the following terminology:

Definition 2.9. A representation $\rho \in \mathrm{Hom}(\pi_1(\Sigma), \mathrm{PSL}(2, \mathbb{R}))$ is *geometric* if it maps the boundary curves of Σ to hyperbolic elements of $\mathrm{PSL}(2, \mathbb{R})$ and has extremal Euler class or, equivalently, if it is the holonomy of a hyperbolic structure on Σ, possibly after reversing the orientation.

2D. ***Laminations in a pair of pants.*** A hyperbolic pair of pants Σ carries only finitely many geodesic laminations, because only 21 geodesics are simple — namely 3 closed geodesics (the boundary components), 6 geodesics spiraling from a boundary component to itself, and 12 geodesics spiraling from a boundary component to another. It admits 32 ideal triangulations, of which 24 contain a geodesic spiraling from a boundary component to itself and the other 8 do not (see Figure 1). We shall call the laminations corresponding to these 8 triangulations the *triskelion* laminations of Σ. They differ by the spiraling directions of the spikes of the triangles at each boundary component.

3. Holonomies of folded hyperbolic structures

Let $\lambda : \mathrm{PSL}(2, \mathbb{R}) \to \mathbb{R}^+$ be the translation length function (1-1). For any representation $\rho \in \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2, \mathbb{R}))$, we set

$$\lambda_\rho := \lambda \circ \rho : \Gamma_g \longrightarrow \mathbb{R}^+.$$

The function λ_ρ is identically zero if and only if the group $\rho(\Gamma_g)$ is unipotent or bounded. The goal of this section is to prove the following:

Proposition 3.1. *For any* $[\rho] \in \mathsf{Rep}_g^{\mathrm{nfd}}$ *with* $\lambda_\rho \not\equiv 0$, *there exist elements* $[j_0], [j_0']$ *of* $\mathsf{Rep}_g^{\mathrm{fd}}$ *and a decomposition* Π *of* Σ_g *into pairs of pants, each labeled* -1, 0, *or* 1, *with the following properties*:

(1) *For any representations* j_0, ρ *in the respective classes* $[j_0]$, $[\rho]$, *there is a* 1*-Lipschitz,* (j_0, ρ)*-equivariant map* $f : \mathbb{H}^2 \to \mathbb{H}^2$ *that is an orientation-preserving* (*resp. orientation-reversing*) *isometry in restriction to any connected subset of* $\mathbb{H}^2$ *projecting to a union of pants labeled* -1 (*resp.* 1) *in* $j_0(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$, *and that satisfies* $\mathrm{Lip}_p(f) < 1$ *for any* $p \in \mathbb{H}^2$ *projecting to the interior of a pair of pants labeled* 0;

(2) *For any representations* j_0', ρ *in the respective classes* $[j_0']$, $[\rho]$, *if the group* $\rho(\Gamma_g)$ *is not virtually abelian, then* ρ *is a folding of* j_0' *along a lamination* Υ *of* Σ_g *consisting of all the cuffs together with a triskelion lamination inside each pair of pants labeled* 0, *with the coloring* $c : \Sigma_g \smallsetminus \Upsilon \to \{-1, 1\}$ *taking the value* -1 (*resp.* 1) *on each pair of pants labeled* -1 (*resp.* 1), *and both values on each pair of pants labeled* 0;

(3) $[j_0]$ *and* $[j_0']$ *only differ by earthquakes along the cuffs of the pairs of pants of the decomposition.*

Property (1) is used to prove Theorem 1.1 in Section 4, while (2) is a more precise statement of Theorem 1.2. We refer to Section 2A for the notation $\mathrm{Lip}_p(f)$ and to Section 2D for triskelion laminations. By additivity (see Section 2C), the Euler class of ρ is the sum of the labels of the pairs of pants.

Proposition 3.1 is proved by choosing an appropriate pants decomposition (Section 3A) and understanding the representations of the fundamental group of a pair of pants (Section 3B). These ingredients are brought together in Section 3C. In Section 3D we present a variation on Proposition 3.1(1), which is later used to prove the second statement of Theorem 1.1.

3A. ***Pants decompositions.*** Our first ingredient is the following:

Lemma 3.2. *For any* $[\rho] \in \mathrm{Rep}_g^{\mathrm{nfd}}$ *with* $\lambda_\rho \not\equiv 0$, *there is a pants decomposition of* Σ_g *such that* ρ *maps any cuff to a hyperbolic element. If* $\rho(\Gamma_g)$ *is not virtually abelian, then we may assume that the restriction of* ρ *to the fundamental group of any pair of pants of the decomposition is nonabelian.*

Recall that $[\rho] \in \mathrm{Rep}_g^{\mathrm{nfd}}$ is said to be *elementary* if the group $\rho(\Gamma_g)$ admits a finite orbit in $\mathbb{H}^2$ or in $\partial_\infty\mathbb{H}^2$. In the case that $[\rho]$ is *not* elementary, Lemma 3.2 is contained in the following result of Gallo, Kapovich, and Marden:

Lemma 3.3 [Gallo et al. 2000, part A]. *For any nonelementary* $[\rho] \in \mathrm{Rep}_g^{\mathrm{nfd}}$, *there is a pants decomposition of* Σ_g *such that the fundamental group of any pair of pants maps injectively to a* 2*-generator Schottky group under* ρ.

We now treat the case that ρ is elementary.

Proof of Lemma 3.2 when ρ *is elementary.* By induction, Lemma 3.2 is a consequence of the following two claims:

Claim 3.4. *Let Σ be a connected compact surface of genus $g \geq 1$ with $k \geq 0$ boundary components such that $\chi(\Sigma) = 2-2g-k < 0$, and let $\rho \in \mathrm{Hom}(\pi_1(\Sigma), \mathrm{PSL}(2,\mathbb{R}))$ be an elementary representation with $\lambda_\rho \not\equiv 0$ sending each boundary curve of Σ (if any) to a hyperbolic element. Then we can cut Σ open along some nonseparating simple closed curve whose image under ρ is a hyperbolic element, yielding a new surface Σ' of genus $g-1$ and an induced representation $\rho' \in \mathrm{Hom}(\pi_1(\Sigma'), \mathrm{PSL}(2,\mathbb{R}))$ sending all $k+2$ boundary curves of Σ' to hyperbolic elements. If the image of ρ is not virtually abelian, then the image of ρ' is not virtually abelian.*

Claim 3.5. *Let Σ be a connected compact surface of genus $g = 0$ with $k \geq 4$ boundary components, and let $\rho \in \mathrm{Hom}(\pi_1(\Sigma), \mathrm{PSL}(2,\mathbb{R}))$ be an elementary representation sending each boundary curve of Σ to a hyperbolic element. Then we can cut Σ along some simple closed curve of Σ, not freely homotopic to a boundary component, whose image under ρ is a hyperbolic element, yielding two new surfaces Σ_1 and Σ_2 with lower complexity and two induced representations $\rho_i \in \mathrm{Hom}(\pi_1(\Sigma_i), \mathrm{PSL}(2,\mathbb{R}))$ sending each boundary curve to a hyperbolic element. If the image of ρ is nonabelian, then we can do this in such a way that the images of the ρ_i are nonabelian.*

Proof of Claim 3.4. We first observe that $\pi_1(\Sigma)$ is generated by elements representing nonseparating simple closed curves on Σ. Indeed, consider a standard presentation

$$\pi_1(\Sigma) = \langle a_1, b_1, \ldots, a_g, b_g, c_1, \ldots, c_k \mid [a_1, b_1] \cdots [a_g, b_g] c_1 \cdots c_k = 1 \rangle \tag{3-1}$$

of $\pi_1(\Sigma)$ by generators and relations, where a_i, b_i represent nonseparating simple closed curves and c_i a curve freely homotopic to a boundary component. Either $a_1 c_i$ represents a nonseparating simple closed curve for all i, or $a_1^{-1} c_i$ represents a nonseparating simple closed curve for all i. Thus we may take the generating set $\{a_1, b_1, \ldots, a_g, b_g, a_1^{\varepsilon} c_1, \ldots, a_1^{\varepsilon} c_k\}$ for some $\varepsilon \in \{-1, 1\}$.

Let us show that ρ sends some nonseparating simple closed curve of Σ to a hyperbolic element. Since $\lambda_\rho \not\equiv 0$, two mutually exclusive situations are possible:

- (T) The group $\rho(\pi_1(\Sigma))$ has a fixed point ξ in $\partial_\infty \mathbb{H}^2$; it is then conjugate to a group of triangular (possibly diagonal) matrices in $\mathrm{PSL}(2,\mathbb{R})$.
- (VA) The group $\rho(\pi_1(\Sigma))$ preserves a geodesic line ℓ of $\mathbb{H}^2$ and contains both translations along ℓ and order-two symmetries of ℓ reversing its orientation; it is then virtually abelian but not abelian.

Consider a system F of generators of $\pi_1(\Sigma)$ representing nonseparating simple closed curves. In case (T), some element of F is necessarily sent by ρ to a hyperbolic element: otherwise the group $\rho(\pi_1(\Sigma))$ would contain only parabolic elements and the identity, which would contradict the fact that $\lambda_\rho \not\equiv 0$. Suppose we are in case (VA) and ρ does not send any element of F to a hyperbolic element; it then sends some element $\gamma \in F$ to an order-two symmetry of ℓ (because it is not the constant homomorphism). We may complete γ into a new standard presentation of the form (3-1) with $\gamma = a_1$. Consider the generating set

$$F' = \{b_1, a_1 b_1, a_2^{-1} b_1, b_2 b_1, \dots, a_g^{-1} b_1, b_g b_1, c_1^{\varepsilon} b_1, \dots, c_k^{\varepsilon} b_1\},$$

where $\varepsilon \in \{-1, 1\}$. If ε is suitably chosen, then every $\gamma' \in F'$ represents a nonseparating simple closed curve, and γ' and $\gamma = a_1$ are standard generators of a one-holed torus embedded in Σ; it follows that $\gamma\gamma'$ represents a nonseparating simple closed curve as well. Necessarily, there exists $\gamma' \in F'$ such that $\rho(\gamma')$ does not commute with $\rho(\gamma)$: otherwise the group $\rho(\pi_1(\Sigma))$ would be contained in the centralizer of $\rho(\gamma)$, which is compact, and this would contradict the fact that $\lambda_\rho \not\equiv 0$. Either this $\rho(\gamma')$ is hyperbolic, or it is an order-two symmetry whose center is different from that of $\rho(\gamma)$, in which case $\rho(\gamma\gamma')$ is hyperbolic. In either case we have found a nonseparating simple closed curve whose image in $\pi_1(\Sigma)$ is mapped by ρ to a hyperbolic element.

Let Σ' be obtained by cutting Σ open along such a simple closed curve. If the image of the induced representation $\rho' \in \mathrm{Hom}(\pi_1(\Sigma'), \mathrm{PSL}(2, \mathbb{R}))$ is virtually abelian, then so is the image of ρ. Indeed, $\pi_1(\Sigma)$ is generated by $\pi_1(\Sigma')$ together with an element γ' that conjugates two elements of $\pi_1(\Sigma')$ with hyperbolic images under ρ'. If the image of ρ' is virtually abelian, preserving some geodesic line ℓ of $\mathbb{H}^2$, then $\rho(\gamma')$ has to preserve ℓ, and so does the whole image of ρ. Thus the image of ρ is virtually abelian. □

Proof of Claim 3.5. Since the boundary curves of Σ generate $\pi_1(\Sigma)$ and since they all have hyperbolic image under the elementary representation ρ, the group $\rho(\pi_1(\Sigma))$ has a fixed point ξ in $\partial_\infty \mathbb{H}^2$ (case (T) above). Choose a geodesic line ℓ of $\mathbb{H}^2$ with endpoint ξ. For any $\gamma \in \Gamma$ we may write in a unique way $\rho(\gamma) = a_\gamma u_\gamma$, where a_γ belongs to the stabilizer A of ξ and ℓ in $\mathrm{PSL}(2, \mathbb{R})$, and $u_\gamma \in \mathrm{PSL}(2, \mathbb{R})$ is unipotent or trivial. The map $\gamma \mapsto a_\gamma$ can be seen as a nonzero element ω of $H^1(\Sigma_g, \mathbb{R})$ after identifying A with $(\mathbb{R}, +)$. Consider a standard presentation

$$\pi_1(\Sigma) = \langle c_1, \dots, c_k \mid c_1 \cdots c_k = 1 \rangle$$

of $\pi_1(\Sigma)$ by generators and relations, where $c_1, \dots, c_k$ represent curves freely homotopic to the boundary components of Σ, and $c_i c_j$ represents a simple curve for any $i < j$. We claim that ρ sends one of the $c_i c_j$ to a hyperbolic element. Indeed, otherwise we would have $\omega(c_i) + \omega(c_j) = 0$ for all $i \neq j$; solving this linear system gives $\omega(c_i) = 0$ for all i, which would contradict the assumption that $\rho(c_i)$ is hyperbolic.

For $1 \le i \le k$, let $\xi_i \in \partial_\infty \mathbb{H}^2$ be the fixed point of $\rho(c_i)$ that is different from ξ. If the image of ρ is not abelian, then there exists i such that $\xi_i \neq \xi_{i+1}$ (with the convention that $\xi_{k+1} = \xi_1$). Precomposing ρ by a Dehn twist along a curve freely homotopic to $c_i c_{i+1}$ corresponds to conjugating $\rho(c_i)$ and $\rho(c_{i+1})$ by $\rho(c_i c_{i+1})$ while leaving all the other $\rho(c_j)$ unchanged. Applying a large enough power of this Dehn twist, with the appropriate sign if $\rho(c_i c_{i+1})$ is hyperbolic, pushes ξ_i and ξ_{i+1} to two distinct points arbitrarily close to ξ; in particular, we can make ξ_i and ξ_{i+1} distinct from the other points ξ_j. We then proceed similarly with the new point ξ_{i+1} and ξ_{i+2}, and so on, until all the points ξ_i are pairwise distinct. We then conclude as above: one of the $c_i c_j$ (with $i \neq j$) has hyperbolic image under ρ. It represents a curve cutting Σ into two smaller surfaces on which ρ induces nonabelian representations. $\square$

To prove Lemma 3.2, just make repeated use of Claim 3.4 to reduce to a surface of genus 0, then of Claim 3.5 to decompose it into pairs of pants. $\square$

3B. ***Representations of the fundamental group of a pair of pants.*** The following lemma gives a dictionary between the geometric and nongeometric representations (Definition 2.9) of the fundamental group of a pair of pants.

Lemma 3.6. *Let $\Gamma = \langle \alpha, \beta, \gamma \mid \alpha\beta\gamma = 1 \rangle$ be the fundamental group of a pair of pants Σ, with α, β, γ corresponding to the three boundary curves.*

For any $a, b, c > 0$ such that none is the sum of the other two, there are exactly two representations $\tau \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2,\mathbb{R}))$ satisfying

$$(\lambda_\tau(\alpha), \lambda_\tau(\beta), \lambda_\tau(\gamma)) = (a, b, c) \tag{3-2}$$

up to conjugation under $\mathrm{PGL}(2,\mathbb{R})$. One of them is geometric (with $|\mathrm{eu}(\tau)| = 1$). The other is nongeometric (with $\mathrm{eu}(\tau) = 0$), and is obtained from the geometric one by folding along any of the eight triskelion laminations of Σ.

For any $a, b, c > 0$ such that one is the sum of the other two, there are exactly four representations $\tau \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2,\mathbb{R}))$ satisfying (3-2), up to conjugation under $\mathrm{PGL}(2,\mathbb{R})$. One of them is geometric (with $|\mathrm{eu}(\tau)| = 1$). The other three are elementary (with $\mathrm{eu}(\tau) = 0$): two have an image that is not virtually abelian and the third one is their abelianization. Each of the two nonabelian elementary representations is obtained from the geometric one by folding along any of four different triskelion laminations of Σ.

When one of a, b, c is the sum of the other two, the images of the two nonabelian elementary representations $\tau \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2,\mathbb{R}))$ are conjugate to triangular matrices; their abelianization is by definition their projection to the group of diagonal matrices.

Proof. Fix a, b, $c > 0$. We first determine the number of conjugacy classes of representations τ satisfying (3-2). Set $(A, B, C) := (e^{a/2}, e^{b/2}, e^{c/2})$, and let

$\tau \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2,\mathbb{R}))$ satisfy (3-2). Up to conjugating τ by $\mathrm{PGL}(2,\mathbb{R})$, we can find lifts $\bar{\tau}(\alpha) \in \mathrm{SL}(2,\mathbb{R})$ of $\tau(\alpha)$ and $\bar{\tau}(\beta) \in \mathrm{SL}(2,\mathbb{R})$ of $\tau(\beta)$ of the form

$$\bar{\tau}(\alpha) = \begin{pmatrix} A & 0 \\ 0 & A^{-1} \end{pmatrix} \quad \text{and} \quad \bar{\tau}(\beta) = \begin{pmatrix} B+x & y \\ z & B^{-1}-x \end{pmatrix}$$

with x, y, $z \in \mathbb{R}$. Since α and β freely generate Γ, this determines a lift $\bar{\tau}$ of τ in $\mathrm{Hom}(\Gamma, \mathrm{SL}(2,\mathbb{R}))$. The sign $\varepsilon \in \{\pm 1\}$ of $\mathrm{Tr}(\bar{\tau}(\alpha))\,\mathrm{Tr}(\bar{\tau}(\beta))\,\mathrm{Tr}(\bar{\tau}(\gamma))$ does not depend on the choice of $\bar{\tau}(\alpha)$ and $\bar{\tau}(\beta)$. By (2-2), we have $\mathsf{eu}(\tau) \in \{-1, 0, 1\}$, with $|\mathsf{eu}(\tau)| = 1$ if and only if τ is geometric, and by (2-3)

$$\varepsilon = (-1)^{\mathsf{eu}(\tau)}.$$

The trace of $\bar{\tau}(\gamma) = \bar{\tau}(\alpha\beta)^{-1}$ is

$$A(B+x) + A^{-1}(B^{-1}-x) = \varepsilon(C + C^{-1}),$$

hence

$$x = \frac{\varepsilon(C + C^{-1}) - AB - (AB)^{-1}}{A - A^{-1}}$$

is uniquely determined by A, B, C, and ε. Let $\nu := (B+x)(B^{-1}-x)$. Since $\bar{\tau}(\beta) \in \mathrm{SL}(2,\mathbb{R})$, we have $yz = \nu - 1$. If $\nu \neq 1$, then any pair (y, z) of reals with product $\nu - 1$ can be obtained by conjugating $\bar{\tau}(\alpha)$ and $\bar{\tau}(\beta)$ by a diagonal matrix in $\mathrm{PGL}(2,\mathbb{R})$ (which does not change x). Thus τ is unique up to conjugation once we fix $\varepsilon \in \{-1, 1\}$. If $\nu = 1$, then $\bar{\tau}(\beta)$ is either upper or lower triangular, or both, hence there are three conjugacy classes for τ, with $\tau(\Gamma)$ consisting respectively of upper triangular, lower triangular, and diagonal matrices. The condition $\nu = 1$ amounts to $(B^{-1} - B - x)x = 0$, or equivalently to

$$\left(\frac{BC}{A} - \varepsilon\right)\left(\frac{AC}{B} - \varepsilon\right) \cdot \left(\frac{AB}{C} - \varepsilon\right)(ABC - \varepsilon) = 0;$$

in other words, $\varepsilon = 1$ and one of a, b, c is the sum of the other two.

Let $j \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2,\mathbb{R}))$ be geometric (Definition 2.9). For any folding ρ of j along a triskelion lamination Υ of Σ, the functions λ_j and λ_ρ agree on $\{\alpha, \beta, \gamma\}$ (Remark 2.8), and ρ is not conjugate to j under $\mathrm{PGL}(2,\mathbb{R})$ because the folding map f is not an isometry (see Section 2A). Therefore, $\mathsf{eu}(\rho) = 0$ by the above discussion.

If none of a, b, c is the sum of the other two, then ρ belongs to the unique conjugacy class of representations τ satisfying (3-2) and $\mathsf{eu}(\tau) = 0$.

If one of a, b, c is the sum of the other two, then ρ belongs to one of the two conjugacy classes of representations τ whose image is not virtually abelian and that satisfy (3-2) and $\varepsilon = 1$ (Observation 2.7). The representation ρ' obtained from j by folding along the image of Υ under the natural involution of the pair of pants belongs to the other conjugacy class of such representations. The abelianization of ρ or ρ' is not conjugate to j, hence satisfies (3-2) and $\varepsilon = 1$ as well. $\square$

Corollary 3.7. *Let $\Gamma = \langle \alpha, \beta, \gamma \mid \alpha\beta\gamma = 1 \rangle$ be the fundamental group of a pair of pants Σ, with α, β, γ corresponding to the three boundary curves. Consider two representations $j, \rho \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))$ with j geometric* (Definition 2.9), *with ρ nongeometric, and with*

$$(\lambda_j(\alpha), \lambda_j(\beta), \lambda_j(\gamma)) = (\lambda_\rho(\alpha), \lambda_\rho(\beta), \lambda_\rho(\gamma)).$$

Then there exists a 1-Lipschitz, (j, ρ)-equivariant map $f : \mathbb{H}^2 \to \mathbb{H}^2$ such that $\mathrm{Lip}_p(f) < 1$ for any $p \in \mathbb{H}^2$ projecting to a point of $j(\Gamma)\backslash\mathbb{H}^2$ off the boundary of the convex core.

Note that in this setting any (j, ρ)-equivariant map $f : \mathbb{H}^2 \to \mathbb{H}^2$ satisfies $\mathrm{Lip}(f) \geq 1$ by Remark 2.6, and if $\mathrm{Lip}(f) = 1$ then f is an isometry in restriction to the translation axes of $j(\alpha)$, $j(\beta)$, $j(\gamma)$ in $\mathbb{H}^2$. The convex core of $j(\Gamma)\backslash\mathbb{H}^2$ naturally identifies with Σ.

Proof. We first assume that the group $\rho(\Gamma)$ is nonabelian. By Lemma 3.6, the representation ρ is obtained from j by folding along any of at least four of the eight triskelion laminations of Σ. Let ℓ be an injectively immersed geodesic that spirals between two boundary components.

If the two boundary components are different, then ℓ is contained in only two triskelion laminations, and intersects the others transversely. If the two boundary components are the same, then ℓ intersects transversely all triskelion laminations of Σ. In both cases we see that a lift of ℓ to $\mathbb{H}^2$ cannot be isometrically preserved by all 1-Lipschitz, (j, ρ)-equivariant maps $f : \mathbb{H}^2 \to \mathbb{H}^2$ (such maps exist since ρ is a folding of j). This holds for any ℓ, which shows that the lamination $\widetilde{\Lambda} \subset \mathbb{H}^2$ of Lemma 2.2 is contained in (in fact, is equal to) the preimage of the boundary of the convex core of $j(\Gamma)\backslash\mathbb{H}^2$, which identifies with the boundary of Σ. By Lemma 2.2, this means that there exists a 1-Lipschitz, (j, ρ)-equivariant map $f : \mathbb{H}^2 \to \mathbb{H}^2$ such that $\mathrm{Lip}_p(f) < 1$ for any $p \in \mathbb{H}^2$ projecting to a point of $j(\Gamma)\backslash\mathbb{H}^2$ off the boundary of the convex core.

We now assume that $\rho(\Gamma)$ is abelian. By Lemma 3.6, the representation ρ is the abelianization of some representation ρ' that is a folding of j. The group $\rho'(\Gamma)$ fixes a point $\xi \in \partial_\infty\mathbb{H}^2$, and $\rho(\Gamma)$ preserves a geodesic line ℓ of $\mathbb{H}^2$ with endpoint ξ. By postcomposing any 1-Lipschitz, (j, ρ')-equivariant map with the projection onto ℓ along the horospheres centered at ξ, we obtain a 1-Lipschitz, (j, ρ)-equivariant map. Moreover, since 1 is the optimal Lipschitz constant (Remark 2.6), this shows that the stretch locus (Definition 2.3) of (j, ρ) is contained in that of (j, ρ'), and we conclude as above. $\square$

Remark 3.8. The nonabelian, nongeometric representations in Lemma 3.6 can also be obtained by folding along a nonmaximal geodesic lamination consisting of a unique leaf spiraling from a boundary component to itself. Folding along a maximal lamination which is not a triskelion gives a representation with values in $\mathrm{PGL}(2,\mathbb{R})$ and not $\mathrm{PSL}(2,\mathbb{R})$.

3C. ***Proof of Proposition 3.1.*** By Lemma 3.2, there is a pants decomposition Π of Σ_g such that ρ maps any cuff to a hyperbolic element, and such that if $\rho(\Gamma_g)$ is not virtually abelian then the restriction of ρ to the fundamental group of any pair of pants is nonabelian. Let $j \in \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))$ be a Fuchsian representation such that $\lambda_j(\gamma) = \lambda_\rho(\gamma)$ for all $\gamma \in \Gamma_g$ corresponding to cuffs of pants of Π. The twist parameters along the cuffs will be adjusted later; for the moment we choose them arbitrarily.

Let $\mathscr{C}$ be the $j(\Gamma_g)$-invariant (disjoint) union of all geodesics of $\mathbb{H}^2$ projecting to the cuffs in $j(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$. For each pair of pants P in Π, choose a subgroup Γ^P of Γ_g which is conjugate to $\pi_1(P)$. Then $j|_{\Gamma^P}$ is the holonomy of a hyperbolic metric on P with cuff lengths given by λ_ρ. Choose a lift $\widetilde{P} \subset \mathbb{H}^2$ of the convex core of $j(\Gamma^P)\backslash\mathbb{H}^2$. This lift is the closure of a connected component of $\mathbb{H}^2 \smallsetminus \mathscr{C}$. If the restrictions of j and ρ to Γ^P are conjugate by some isometry f^P of $\mathbb{H}^2$, then we give P the label -1 or 1, depending on whether f^P preserves the orientation or not. If the restrictions of j and ρ to Γ^P are not conjugate, then we give P the label 0. In this case:

- There is a 1-Lipschitz, $(j|_{\Gamma^P}, \rho|_{\Gamma^P})$-equivariant map $f^P : \widetilde{P} \to \mathbb{H}^2$ with $\mathrm{Lip}_p(f^P) < 1$ for all $p \notin \partial\widetilde{P}$, by Corollary 3.7.
- If $\rho(\Gamma_g)$ is not virtually abelian then $\rho|_{\Gamma^P}$ is a folding of $j|_{\Gamma^P}$ along some triskelion lamination of P, by Lemma 3.6; we denote by $F^P : \widetilde{P} \to \mathbb{H}^2$ the folding map.

Note that in restriction to any connected component of $\partial\widetilde{P}$ (a line), the maps f^P and F^P are both isometries, but they may disagree by a constant shift.

The collection of all maps f^P, extended (j,ρ)-equivariantly, piece together to yield a map $f^* : \mathbb{H}^2 \smallsetminus \mathscr{C} \to \mathbb{H}^2$. The obstruction to extending f^* by continuity on each geodesic $\ell \subset \mathscr{C}$ is that the maps on either side of ℓ may disagree by a constant shift along ℓ. This discrepancy $\delta(\ell) \in \mathbb{R}$ is the same on the whole $j(\Gamma_g)$-orbit of ℓ. To correct it, we postcompose j with an earthquake supported on the cuff associated with ℓ, of length $-\delta(\ell)$. We repeat for each $j(\Gamma_g)$-orbit in $\mathscr{C}$, and eventually obtain a new Fuchsian representation j_0. By construction, there is a 1-Lipschitz, (j_0,ρ)-equivariant map $f : \mathbb{H}^2 \to \mathbb{H}^2$, obtained simply by gluing together isometric translates of the f^P. This extension f satisfies Proposition 3.1(1).

If $\rho(\Gamma_g)$ is not virtually abelian, then similarly the maps f^P for P labeled ± 1 and F^P for P labeled 0 piece together to yield a map $F^* : \mathbb{H}^2 \smallsetminus \mathscr{C} \to \mathbb{H}^2$. As above, we can modify j by earthquakes into a new Fuchsian representation j_0', and F^* by piecewise isometries into a (j_0', ρ)-equivariant, continuous map $F : \mathbb{H}^2 \to \mathbb{H}^2$ which is a folding map. This proves Proposition 3.1(2).

Proposition 3.1(3) is satisfied by construction. □

3D. ***Uniform Lipschitz bounds.*** In order to prove the second claim of Theorem 1.1 in Section 4D, we shall use the following result, which gives Lipschitz bounds which are analogous to Proposition 3.1(1) but uniform.

Proposition 3.9. *For any decomposition Π of Σ_g into pairs of pants labeled $-1, 0, 1$ and any continuous family $(j_t)_{t\geq 0} \subset \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))$ of Fuchsian representations, there exist a family $(\rho_t)_{t\geq 0} \subset \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))$ of non-Fuchsian representations and, for any t in a small interval $[0, t_0]$, a 1-Lipschitz, (j_t, ρ_t)-equivariant map $\varphi_t : \mathbb{H}^2 \to \mathbb{H}^2$, with the following properties:*

- *φ_t is an orientation-preserving (resp. orientation-reversing) isometry in restriction to any connected subset of $\mathbb{H}^2$ projecting to a union of pants labeled -1 (resp. 1) in $j_t(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$;*
- *For any $\eta > 0$, there exists $C < 1$ such that $\mathrm{Lip}_p(\varphi_t) \leq C$ for all $t \in [0, t_0]$ and all $p \in \mathbb{H}^2$ whose image in $j_t(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$ lies inside a pair of pants P labeled 0, at distance at least η from the boundary of P.*

Proposition 3.9 is based on the following uniform version of Corollary 3.7:

Lemma 3.10. *Let $\Gamma = \langle \alpha, \beta, \gamma \mid \alpha\beta\gamma = 1\rangle$ be the fundamental group of a pair of pants Σ, with α, β, γ corresponding to the three boundary curves. Consider two continuous families $(j_t)_{t\geq 0}$, $(\rho_t)_{t\geq 0} \subset \mathrm{Hom}(\Gamma, \mathrm{PSL}(2,\mathbb{R}))$ of representations with j_t geometric (Definition 2.9), with ρ_t nongeometric, and with*

$$(\lambda_{j_t}(\alpha), \lambda_{j_t}(\beta), \lambda_{j_t}(\gamma)) = (\lambda_{\rho_t}(\alpha), \lambda_{\rho_t}(\beta), \lambda_{\rho_t}(\gamma))$$

for all $t \geq 0$. Then there exists a family of 1-Lipschitz, (j_t, ρ_t)-equivariant maps $\varphi_t : \mathbb{H}^2 \to \mathbb{H}^2$, defined for all t in a small interval $[0, t_0]$, with the following property: for any $\eta > 0$ there exists $C < 1$ such that $\mathrm{Lip}_p(\varphi_t) \leq C$ for all $t \in [0, t_0]$ and all $p \in \mathbb{H}^2$ whose image in $j_t(\Gamma)\backslash\mathbb{H}^2$ lies at distance at least η from the boundary of the convex core.

Proof of Lemma 3.10. By Corollary 3.7, there exists a 1-Lipschitz, (j_0, ρ_0)-equivariant map $f_0 : \mathbb{H}^2 \to \mathbb{H}^2$ such that $\mathrm{Lip}_p(f_0) < 1$ for any $p \in \mathbb{H}^2$ whose image in $j_0(\Gamma)\backslash\mathbb{H}^2$ does not belong to the boundary of the convex core. If $(j_t, \rho_t) = (j_0, \rho_0)$ for all t, then we may take $\varphi_t = f_0$. In the general case, we shall build φ_t as a small deformation of f_0 in restriction to the preimage of the convex core of $j_t(\Gamma)\backslash\mathbb{H}^2$.

Choose $\Delta > 0$ so that for all small $t \geq 0$, the 2Δ-neighborhoods of the boundary components of the convex core of the hyperbolic surface $j_t(\Gamma)\backslash\mathbb{H}^2$ are disjoint.

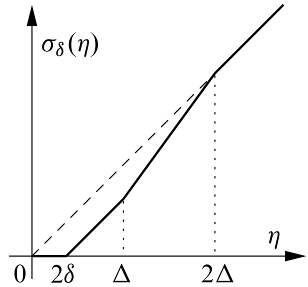


Figure 2. The function σ_δ in the proof of Lemma 3.10.

Choose a small $\delta \in (0, \Delta/2)$ and let $\sigma_\delta : \mathbb{R}^+ \to \mathbb{R}^+$ be the function that satisfies

$$\sigma_\delta(\eta) = \begin{cases} 0 & \text{for } 0 \le \eta \le 2\delta, \\ \Delta - 2\delta & \text{for } \eta = \Delta, \\ \eta & \text{for } \eta \ge 2\Delta \end{cases}$$

and is affine on $[2\delta, \Delta]$ and $[\Delta, 2\Delta]$ (Figure 2). Note that σ_δ is $(1+o(1))$-Lipschitz as $\delta \to 0$, and 1-Lipschitz away from $[\Delta, 2\Delta]$. For any $t \ge 0$, let $N_t \subset \mathbb{H}^2$ be the preimage of the convex core of $j_t(\Gamma)\backslash\mathbb{H}^2$, and let $\pi_t : \mathbb{H}^2 \to N_t$ be the closest-point projection, which is 1-Lipschitz. We set

$$\varphi_0 := f_0 \circ J_\delta \circ \pi_0,$$

where J_δ is the homotopy of $\mathbb{H}^2$ taking any point at distance $\eta \le 2\Delta$ from a boundary component ℓ_0 of N_0 to the point at distance $\sigma_\delta(\eta)$ from ℓ_0 on the same perpendicular ray to ℓ_0, leaving other points unchanged. By construction, in restriction to the 2δ-neighborhood of ∂N_0, the map φ_0 factors through the closest-point projection onto ∂N_0. The function $p \mapsto \mathrm{Lip}_p(f_0)$ is $j_0(\Gamma)$-invariant, upper semicontinuous, and less than 1 on $\mathbb{H}^2 \smallsetminus \partial N_0$, hence bounded away from 1 when $p \in N_0$ stays at distance at least $\Delta - 2\delta$ from ∂N_0. This implies that, if we have chosen δ small enough (which we shall assume from now on), then $\mathrm{Lip}(\varphi_0) = 1$ and $\mathrm{Lip}_p(\varphi_0) < 1$ for all p in the interior of N_0. For $t > 0$, we construct φ_t as a deformation of φ_0 via a partition of unity, as follows.

Let $\mathcal{U}_t^\delta \subset N_t$ be the δ-neighborhood of ∂N_t and $N_t^\delta := N_t \smallsetminus \mathcal{U}_t^\delta$ its complement in N_t; we define $\mathcal{U}_t^{2\delta}$ similarly. Choose a 1-Lipschitz, (j_t, ρ_t)-equivariant map $\varphi_t^0 : \mathcal{U}_t^{2\delta} \to \mathbb{H}^2$ factoring through the closest-point projection onto ∂N_t and taking any boundary component ℓ_t of N_t, stabilized by a cyclic subgroup $j_t(S)$ of $j_t(\Gamma)$, isometrically to the translation axis of $\rho_t(S)$ in $\mathbb{H}^2$. Up to postcomposing each φ_t^0 with an appropriate shift along the axis of $\rho_t(S)$, we may assume that $\varphi_t^0(p) \to \varphi_0(p)$ for any $p \in \mathcal{U}_0^{2\delta}$ as $t \to 0$ (recall that the restriction of φ_0 to any boundary component of N_0 is an isometry).

Let $B^1, \dots, B^n \subset N_0$ be balls of $\mathbb{H}^2$, each projecting injectively to $j_0(\Gamma)\backslash\mathbb{H}^2$, disjoint from a neighborhood of ∂N_0, and such that

$$N_0^\delta \subset j_0(\Gamma) \cdot \bigcup_{i=1}^n B^i.$$

For $1 \le i \le n$, let $\varphi_t^i : j_t(\Gamma) \cdot B^i \to \mathbb{H}^2$ be the (j_t, ρ_t)-equivariant map that agrees with φ_0 on B^i. By construction, for all $1 \le i \le n$ (resp. for $i = 0$) and for all $p \in j_0(\Gamma) \cdot B^i$ (resp. $p \in \mathcal{U}_0^{2\delta}$) we have $\varphi_t^i(p) \to \varphi_0(p)$ as $t \to 0$, uniformly for p in any compact set. However, the maps φ_t^i, for $0 \le i \le n$, may not agree at points where their domains overlap. The goal is to paste them together by the procedure described in Section 2A, using a $j_t(\Gamma)$-invariant partition of unity $(\psi_t^i)_{0 \le i \le n}$ that we now construct.

Let $\psi_t^0 : \mathbb{H}^2 \to [0, 1]$ be the function supported on $\mathcal{U}_t^{2\delta}$ that takes any point at distance η from ∂N_t to $\tau(\eta) \in [0, 1]$, where $\tau([0, \delta]) = 1$, $\tau([2\delta, +\infty)) = 0$, and τ is affine on $[\delta, 2\delta]$. Let $\psi^1, \dots, \psi^n : \mathbb{H}^2 \to [0, 1]$ be $j_0(\Gamma)$-invariant Lipschitz functions inducing a partition of unity on a neighborhood of N_0^δ, with ψ^i supported in $j_0(\Gamma) \cdot B^i$. Since N_t has a compact fundamental domain for $j_t(\Gamma)$ that varies continuously with t (for instance a right-angled octagon), for small enough t we have

$$N_t^\delta \subset j_t(\Gamma) \cdot \bigcup_{i=1}^n B^i.$$

For $1 \le i \le n$ and $t \ge 0$, let $\widehat{\psi}_t^i : \mathbb{H}^2 \to [0, 1]$ be the $j_t(\Gamma)$-invariant function supported on $j_t(\Gamma) \cdot B^i$ that agrees with ψ^i on B^i. Then $\sum_{i=1}^n \widehat{\psi}_t^i = 1 + o(1)$ as $t \to 0$, with an error term uniform on N_t^δ. Therefore the functions

$$\psi_t^0 \quad \text{and} \quad \psi_t^i := (1 - \psi_t^0) \frac{\widehat{\psi}_t^i}{\sum_{k=1}^n \widehat{\psi}_t^k} : \mathbb{H}^2 \longrightarrow [0, 1]$$

for $1 \le i \le n$ form a $j_t(\Gamma)$-invariant partition of unity of N_t, subordinated to the covering $\mathcal{U}_t^{2\delta} \cup j_t(\Gamma) \cdot B^1 \cup \dots \cup j_t(\Gamma) \cdot B^n \supset N_t$, and are all L-Lipschitz for some $L > 0$ independent of i and t.

For $t \ge 0$, let $\varphi_t := \sum_{i=0}^n \psi_t^i \varphi_t^i : N_t \to \mathbb{H}^2$ be the averaged map defined in Section 2A. This map is (j_t, ρ_t)-equivariant by construction. We extend it to a map $\varphi_t : \mathbb{H}^2 \to \mathbb{H}^2$ by precomposing with the closest-point projection $\pi_t : \mathbb{H}^2 \to N_t$. We claim that the maps φ_t satisfy the conclusion of Lemma 3.10. Indeed, by Lemma 2.4, for any $t \ge 0$ and p in the interior of N_t,

$$\operatorname{Lip}_p(\varphi_t) \le \sum_{i \in I_t(p)} \left(\operatorname{Lip}_p(\psi_t^i) R_t(p) + \psi_t^i(p) \operatorname{Lip}_p(\varphi_t^i)\right), \tag{3-3}$$

where $I_t(p)$ is the set of indices $0 \le i \le n$ such that p belongs to the support of ψ_t^i, and $R_t(p) \ge 0$ is the diameter of the set $\{\varphi_t^i(p) \mid i \in I_t(p)\}$. Let $\eta > 0$ be the distance from p to ∂N_t.

If $\eta < \delta$, then φ_t coincides on a neighborhood of p with φ_t^0, hence with the closest-point projection onto ∂N_t postcomposed with an isometry of $\mathbb{H}^2$, and the right-hand side of (3-3) reduces to

$$\mathrm{Lip}_p(\varphi_t^0) = \frac{1}{\cosh \eta} < 1$$

(see [Guéritaud and Kassel 2013, (A.9)], for instance).

If $\eta \ge \delta$, then the bound on $\mathrm{Lip}_p(\varphi_t^0)$ still holds, and $\mathrm{Lip}_p(\varphi_t^i)$ for $1 \le i \le n$ can also be uniformly bounded away from 1. Indeed, $\sup_{q \in B^i} \mathrm{Lip}_q(\varphi_t^i) < 1$ since B^i is disjoint from a neighborhood of ∂N_0 and the local Lipschitz constant is upper semicontinuous, and we argue by equivariance. Moreover, all the other contributions to (3-3) are small: $R_t(p) \to 0$ as $t \to 0$, uniformly in p, and $\mathrm{Lip}_p(\psi_t^i)$ is bounded independently of p, i, t (by L). Therefore, for small t there exists $C < 1$, independent of p and t, such that $\mathrm{Lip}_p(\varphi_t) \le C$.

This treats the case when $p \in N_t$. To conclude, we note that on a neighborhood of any $p \in \mathbb{H}^2 \smallsetminus N_t$ the map φ_t coincides with the closest-point projection onto ∂N_t postcomposed with an isometry of $\mathbb{H}^2$, hence $\mathrm{Lip}_p(\varphi_t) = 1/\cosh \eta < 1$, where $\eta = d(p, \partial N_t)$. $\square$

Proof of Proposition 3.9. Let Υ be a lamination of Σ_g consisting of all the cuffs of Π together with a triskelion lamination inside each pair of pants labeled 0. Let $c : \Sigma_g \smallsetminus \Upsilon \to \{-1, 1\}$ be a coloring taking the value -1 (resp. 1) on each pair of pants labeled -1 (resp. 1), and both values on each pair of pants labeled 0. For any $t \ge 0$, let ρ'_t be the folding of j_t along Υ with coloring c.

We now argue similarly to the proof of Proposition 3.1 in Section 3C. For each pair of pants P in Π, choose a subgroup Γ^P of Γ_g which is conjugate to $\pi_1(P)$, and for any $t \ge 0$ a lift $\widetilde{P}_t \subset \mathbb{H}^2$ of the convex core of $j_t(\Gamma^P)\backslash\mathbb{H}^2$.

If P is labeled -1 (resp. 1), then for any $t \ge 0$ the restrictions of j_t and ρ'_t to Γ^P are conjugate by some orientation-preserving (resp. orientation-reversing) isometry φ_t^P of $\mathbb{H}^2$.

If P is labeled 0, then, by Lemma 3.10, there is a family of 1-Lipschitz, $(j_t|_{\Gamma^P}, \rho'_t|_{\Gamma^P})$-equivariant maps $\varphi_t^P : \mathbb{H}^2 \to \mathbb{H}^2$, defined for all t in a small interval $[0, t_0]$, with the following property: for any $\eta > 0$, there exists $C < 1$ such that $\mathrm{Lip}_p(\varphi_t^P) \le C$ for all $t \in [0, t_0]$ and all $p \in \widetilde{P}_t$ at distance at least η from $\partial \widetilde{P}_t$.

The collection of all maps φ_t^P, extended (j_t, ρ'_t)-equivariantly, piece together to yield a map $\varphi_t^* : \mathbb{H}^2 \smallsetminus \mathscr{C}_t \to \mathbb{H}^2$, where $\mathscr{C}_t$ is the union of all geodesics of $\mathbb{H}^2$ projecting to cuffs of Π in $j_t(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$.

The obstruction to extending φ_t^* by continuity on each geodesic $\ell_t \subset \mathscr{C}_t$ is that the maps on either side of ℓ_t may disagree by a constant shift along ℓ_t if ℓ_t separates two pairs of pants labeled $(\pm 1, 0)$ or $(0, 0)$. This discrepancy $\delta(\ell_t) \in \mathbb{R}$ is the same on the whole $j_t(\Gamma_g)$-orbit of ℓ_t. To correct it, we precompose the folding ρ'_t of j_t with an earthquake, supported on the cuff associated with ℓ_t (in the j_t-metric), of length $-\delta(\ell_t)$. We repeat for each $j_t(\Gamma_g)$-orbit in $\mathscr{C}_t$, and eventually obtain a new folded representation ρ_t. By construction, there is a family of 1-Lipschitz, (j_t, ρ_t)-equivariant maps $\varphi_t : \mathbb{H}^2 \to \mathbb{H}^2$ satisfying Proposition 3.9, obtained simply by gluing together isometric translates of the φ_t^P. $\square$

4. Surjectivity of the two projections

In this section we prove Theorem 1.1. We first construct uniformly lengthening deformations of surfaces with boundary (Section 4A), then glue these together according to combinatorics given by Proposition 3.1 (Sections 4B and 4D). Section 4C is devoted to the proof of a technical lemma.

4A. ***Uniformly lengthening deformations of compact hyperbolic surfaces with boundary.*** Our two main tools to prove Theorem 1.1 are Proposition 3.1 and the following lemma:

Lemma 4.1. *Let Γ be the fundamental group and $j_0 \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))$ the holonomy of a compact, connected, hyperbolic surface Σ with nonempty geodesic boundary. Then there exist $t_0 > 0$ and a continuous family of representations $(j_t)_{0 \le t \le t_0}$ with the following properties*:

(a) $\lambda_{j_0}(\gamma) = (1-t)\lambda_{j_t}(\gamma)$ *for any $t \in [0, t_0]$ and any $\gamma \in \Gamma$ corresponding to a boundary component of Σ*;

(b) $\sup_{\gamma \in \Gamma \smallsetminus \{1\}} \lambda_{j_0}(\gamma)/\lambda_{j_t}(\gamma) < 1$ *for any $t \in (0, t_0]$*;

(c) $j_t(\gamma) = j_0(\gamma) + O(t)$ *for any $\gamma \in \Gamma$ as $t \to 0$, where both sides are seen as 2×2 real matrices with determinant 1 modulo $\pm \mathrm{Id}$*;

(d) *for any compact subset K of $\mathbb{H}^2$ projecting to the interior of the convex core of $j_0(\Gamma)\backslash\mathbb{H}^2$, there exists $L > 0$ such that*

$$d(p, f_t(p)) \le Lt$$

for any $p \in K$, any $t \in [0, t_0]$, and any 1-Lipschitz, (j_t, j_0)-equivariant map $f_t : \mathbb{H}^2 \to \mathbb{H}^2$.

As in Section 3B, the convex core of $j_0(\Gamma)\backslash\mathbb{H}^2$ naturally identifies with Σ. The idea is to construct the representations j_t as holonomies of hyperbolic surfaces obtained from $j_0(\Gamma)\backslash\mathbb{H}^2$ by *strip deformations*. This type of deformation was first

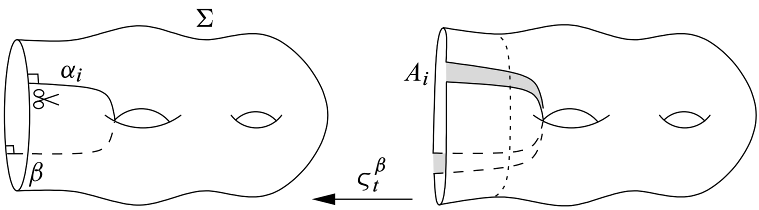


Figure 3. A strip deformation. In the source of the collapsing map ς_t^β we show the new peripheral geodesic, dotted.

introduced by Thurston [1986, proof of Lemma 3.4]. We refer to [Papadopoulos and Théret 2010; Danciger et al. 2014] for more details.

Proof. We first explain how to lengthen one boundary component β of Σ. Choose a finite collection of disjoint, biinfinite geodesic arcs $\alpha_1, \dots, \alpha_n \subset j_0(\Gamma)\backslash\mathbb{H}^2$, each crossing β orthogonally twice, that subdivide the convex core Σ into right-angled hexagons and one-holed right-angled bigons. Along each arc α_i, following [Thurston 1986], slice $j_0(\Gamma)\backslash\mathbb{H}^2$ open and insert a strip A_i of $\mathbb{H}^2$, bounded by two geodesics, with narrowest cross-section at the midpoint of $\alpha_i \cap \Sigma$ (see Figure 3).

This yields a new complete hyperbolic surface, with a compact convex core, equipped with a natural 1-Lipschitz map ς_t^β to $j_0(\Gamma)\backslash\mathbb{H}^2$ obtained by collapsing the strips A_i back to lines. Note that the image under ς_t^β of the new convex core is *strictly contained* in Σ (see Figure 3). The geodesic corresponding to β is longer in the new surface than in Σ. By adjusting the widths of the strips A_i, we may assume that the ratio of lengths is $1/(1-t)$. The appropriate widths for this ratio are in $O(t)$ as $t \to 0$. All lengths of geodesics corresponding to boundary components other than β are unchanged.

Repeat the construction, iteratively, for all boundary components $\beta_1, \dots, \beta_r$ of Σ, in some arbitrary order. We thus obtain a new complete hyperbolic surface $j_t(\Gamma)\backslash\mathbb{H}^2$, with a compact convex core Σ_t, such that j_t satisfies (a).

We claim that j_t also satisfies (b). Indeed, consider the 1-Lipschitz map $\varsigma_t := \varsigma_t^{\beta_r} \circ \dots \circ \varsigma_t^{\beta_1}$ from Σ_t to Σ. If 1 were its optimal Lipschitz constant, then by Lemma 2.2 there would exist a geodesic lamination of Σ_t whose leaves are isometrically preserved by ς_t. But this is not the case here since for every i, the map $\varsigma_t^{\beta_i}$ does not isometrically preserve any geodesic lamination except the boundary components other than β_i. Therefore ς_t has Lipschitz constant strictly less than 1, which implies (b) by Remark 2.6.

Up to replacing each j_t with a conjugate under $\mathrm{PSL}(2,\mathbb{R})$, we may assume that (c) holds. Indeed, it is well known that there exist elements $\gamma_1, \dots, \gamma_n \in \Gamma$ whose length functions form a smooth coordinate system for $\mathrm{Hom}(\Gamma, \mathrm{PSL}(2,\mathbb{R}))/\mathrm{PSL}(2,\mathbb{R})$ near $[j_0]$ (see [Goldman and Xia 2011, Theorem 2.1] for instance). For any i, the preimage under ς_t of the closed geodesic of Σ associated with γ_i is obtained by

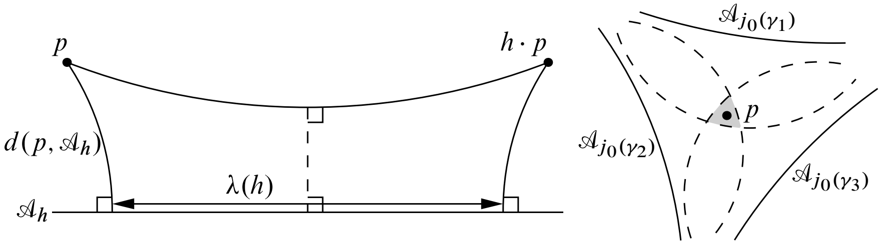


Figure 4. Illustration of the proof of Lemma 4.1. Left: a hyperbolic quadrilateral with two right angles. Right: the point $f_t(p)$ belongs to the shaded region.

expanding finitely many strips of width $O(t)$, hence $\lambda_{j_t}(\gamma_i) \leq \lambda_{j_0}(\gamma_i) + O(t)$ as $t \to 0$. On the other hand, $\lambda_{j_t}(\gamma_i) \geq \lambda_{j_0}(\gamma_i)$ due to the existence of the 1-Lipschitz map ς_t. Therefore, $d'(j_0, j_t) = O(t)$ for any smooth metric d' on a neighborhood of $[j_0]$ in $\mathrm{Hom}(\Gamma, \mathrm{PSL}(2,\mathbb{R}))/\,\mathrm{PSL}(2,\mathbb{R})$.

To check (d), we use a perturbative version of the argument that a $j_0(\Gamma)$-invariant, 1-Lipschitz map must be the identity on the preimage $N_0 \subset \mathbb{H}^2$ of the convex core Σ of $j_0(\Gamma)\backslash\mathbb{H}^2$. For any hyperbolic element $h \in \mathrm{PSL}(2,\mathbb{R})$, with translation axis $\mathscr{A}_h \subset \mathbb{H}^2$, and for any $p \in \mathbb{H}^2$, a classical formula gives

$$\sinh\bigl(\tfrac{1}{2} d(p, h\cdot p)\bigr) = \sinh\bigl(\tfrac{1}{2}\lambda(h)\bigr)\cdot \cosh d(p, \mathscr{A}_h) \tag{4-1}$$

(see Figure 4, left). Consider $p \in \mathbb{H}^2$ in the interior of N_0. We can find three translation axes $\mathscr{A}_{j_0(\gamma_1)}, \mathscr{A}_{j_0(\gamma_2)}, \mathscr{A}_{j_0(\gamma_3)} \subset \partial N_0$ of elements of $j_0(\Gamma)$ such that, if q_i denotes the projection of p to $\mathscr{A}_{j_0(\gamma_i)}$, then p belongs to the interior of the triangle $q_1q_2q_3$. For any $t \geq 0$ and any 1-Lipschitz, (j_t, j_0)-equivariant map $f_t : \mathbb{H}^2 \to \mathbb{H}^2$,

$$d(f_t(p), j_0(\gamma_i)\cdot f_t(p)) \leq d(p, j_t(\gamma_i)\cdot p),$$

which by (4-1) may be written as

$$\sinh\bigl(\tfrac{1}{2}\lambda_{j_0}(\gamma_i)\bigr)\cdot \cosh d(f_t(p), \mathscr{A}_{j_0(\gamma_i)}) \leq \sinh\bigl(\tfrac{1}{2}\lambda_{j_t}(\gamma_i)\bigr)\cdot \cosh d(p, \mathscr{A}_{j_t(\gamma_i)}).$$

Since $\lambda_{j_0}(\gamma_i) = \lambda_{j_t}(\gamma_i) + O(t)$ and $d(p, \mathscr{A}_{j_t(\gamma_i)}) = d(p, \mathscr{A}_{j_0(\gamma_i)}) + O(t)$ by (c), this implies

$$\cosh d(f_t(p), \mathscr{A}_{j_0(\gamma_i)}) \leq \cosh d(p, \mathscr{A}_{j_0(\gamma_i)}) + O(t),$$

where the error term does not depend on the choice of f_t. Since $d(p, \mathscr{A}_{j_0(\gamma_i)}) > 0$, we may invert the hyperbolic cosine:

$$d(f_t(p), \mathscr{A}_{j_0(\gamma_i)}) \leq d(p, \mathscr{A}_{j_0(\gamma_i)}) + O(t).$$

Applied to $i = 1, 2, 3$, this means that $f_t(p)$ belongs to a curvilinear triangle around p bounded by three hypercycles (curves at constant distance from a geodesic

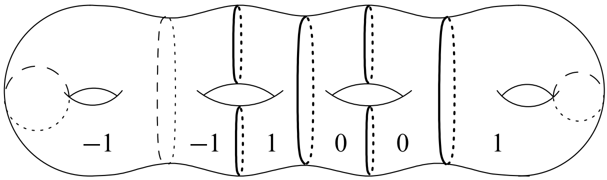


Figure 5. A labeled pants decomposition with $m = 5$. The boundary components of the Σ^i, $1 \leq i \leq 5$, are in bold.

line) expanding at rate $O(t)$ as t becomes positive, hence $d(p, f_t(p)) = O(t)$ (see Figure 4, right). All estimates $O(t)$ are robust under small perturbations of p, hence can be made uniform (and still independent of f_t) for p in a compact set K, yielding (d). □

4B. ***Gluing surfaces with boundary.*** We now prove the first claim of Theorem 1.1. Namely, given $[\rho] \in \mathrm{Rep}_g^{\mathrm{nfd}}$, we construct $[j] \in \mathrm{Rep}_g^{\mathrm{fd}}$ that strictly dominates $[\rho]$.

If $\lambda_\rho \equiv 0$, then any $[j] \in \mathrm{Rep}_g^{\mathrm{fd}}$ strictly dominates $[\rho]$. We now suppose $\lambda_\rho \not\equiv 0$. Proposition 3.1(1) then gives us an element $[j_0] \in \mathrm{Rep}_g^{\mathrm{fd}}$, a labeled pants decomposition Π of Σ_g, and, for any $j_0, \rho \in \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2, \mathbb{R}))$ in the respective classes $[j_0]$, $[\rho]$ (which we now fix), a 1-Lipschitz, (j_0, ρ)-equivariant map $f : \mathbb{H}^2 \to \mathbb{H}^2$ that is an orientation-preserving (resp. orientation-reversing) isometry in restriction to any connected subset of $\mathbb{H}^2$ projecting to a union of pants labeled -1 (resp. 1) in $j_0(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$ and that satisfies $\mathrm{Lip}_p(f) < 1$ for any $p \in \mathbb{H}^2$ projecting to the interior of a pair of pants labeled 0. Not all pairs of pants are labeled -1, and not all 1, since j_0 and ρ are not conjugate under $\mathrm{PGL}(2, \mathbb{R})$. By Remark 2.6, the class $[j_0]$ dominates $[\rho]$ in the sense that $\lambda(\rho(\gamma)) \leq \lambda(j_0(\gamma))$ for all $\gamma \in \Gamma_g$. Our goal is to use Lemma 4.1 to modify j_0 into a representation j such that $[j]$ *strictly* dominates $[\rho]$.

For this purpose, we erase all the cuffs that separate two pairs of pants of Π with labels $(-1, -1)$ or $(1, 1)$, and write

$$\Sigma_g = \Sigma^1 \cup \cdots \cup \Sigma^m,$$

where Σ^i, for any $1 \leq i \leq m$, is a compact surface with boundary that is one of:

- a pair of pants labeled 0,
- a full connected component of the subsurface of Σ_g made of pants labeled -1,
- or a full connected component of the subsurface of Σ_g made of pants labeled 1;

(see Figure 5). The boundary components of the Σ^i are the cuffs that separated two pairs of pants of Π with labels $(-1, 1)$, $(\pm 1, 0)$ or $(0, 0)$. Choose a small $\delta > 0$ such that, in all hyperbolic metrics on Σ_g which are close enough to that defined by j_0, any simple geodesic entering the δ-neighborhood of the geodesic representative of a cuff of Π crosses it. Let $\mathscr{C}_0 \subset \mathbb{H}^2$ be the union of all geodesic lines of $\mathbb{H}^2$ projecting to boundary components of the Σ^i in $j_0(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$, let $N_0^\delta \subset \mathbb{H}^2$ be the complement of the δ-neighborhood of $\mathscr{C}_0$, and let $K \subset \mathbb{H}^2 \smallsetminus \mathscr{C}_0$ be a compact set whose interior contains a fundamental domain of N_0^δ for the action of $j_0(\Gamma_g)$, with m connected components projecting respectively to $\Sigma^1, \dots, \Sigma^m$.

We apply Lemma 4.1 to $\Gamma^i := \pi_1(\Sigma^i)$ and $j_0^i := j_0|_{\Gamma^i}$ and obtain continuous families $(j_t^i)_{0 \le t \le t_0} \subset \mathrm{Hom}(\Gamma^i, \mathrm{PSL}(2,\mathbb{R}))$ of representations for $1 \le i \le m$ satisfying properties (a)–(d) of Lemma 4.1 with a uniform constant $L > 0$ for the compact set $K \subset \mathbb{H}^2 \smallsetminus \mathscr{C}_0$. For any $t \in [0, t_0]$, using (a), we can glue together the (compact) convex cores of the $j_t^i(\Gamma^i)\backslash\mathbb{H}^2$ following the same combinatorics as the Σ^i. This gives a closed hyperbolic surface of genus g, hence a holonomy representation $j_t \in \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))$. By (c), up to adjusting the twist parameters, we may assume that

$$j_t(\gamma) = j_0(\gamma) + O(t) \tag{4-2}$$

for any $\gamma \in \Gamma_g$ as $t \to 0$, where both sides are seen as 2×2 real matrices with determinant 1 modulo $\pm\,\mathrm{Id}$.

To complete the proof of the first statement of Theorem 1.1, it is sufficient to prove that for small enough $t > 0$,

$$\sup_{\gamma \in (\Gamma_g)_s} \frac{\lambda_\rho(\gamma)}{\lambda_{j_t}(\gamma)} < 1, \tag{4-3}$$

where $(\Gamma_g)_s$ is the set of nontrivial elements of Γ_g corresponding to simple closed curves on Σ_g; then $[j] := [j_t]$ will strictly dominate $[\rho]$ by Theorem 2.5. Note that $\lambda(j_t(\gamma)) = \lambda(j_t^i(\gamma))$ for all γ in Γ^i, seen as a subgroup of Γ_g. Thus (b) gives the control required in (4-3) for simple closed curves *contained in one of the* Σ^i. We now explain why the lengths of the other simple closed curves also decrease uniformly, based on (b), (c), and (d).

For any $t \in (0, t_0]$, let $\mathscr{C}_t \subset \mathbb{H}^2$ be the union of the lifts to $\mathbb{H}^2$ of the simple closed geodesics of $j_t(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$ corresponding to $\mathscr{C}_0$ and let N_t^δ be the complement of the δ-neighborhood of $\mathscr{C}_t$ in $\mathbb{H}^2$. For t small enough, we can find a fundamental domain K_t of N_t^δ for the action of $j_t(\Gamma_g)$ that is contained in K and has m connected components. By (b) and Theorem 2.5, for any $1 \le i \le m$ and $t \in (0, t_0]$ there exists a $(j_t|_{\Gamma^i}, j_0|_{\Gamma^i})$-equivariant map $f_t^i : \mathbb{H}^2 \to \mathbb{H}^2$ with $\mathrm{Lip}(f_t^i) < 1$. For small $t > 0$, we choose a (j_t, j_0)-equivariant map $f_t : (N_t^\delta \cup \mathscr{C}_t) \to \mathbb{H}^2$ such that:

- $f_t = f_t^i$ on the component of K_t projecting to Σ^i for all $1 \le i \le m$;
- f_t takes any geodesic line in $\mathscr{C}_t$ to the corresponding line in $\mathscr{C}_0$, multiplying all distances on it by the uniform factor $(1 - t)$.

We choose the f_t so that, in addition, for any compact set $K' \subset \mathbb{H}^2$ there exists $L_1 \geq 0$ such that $d(x', f_t(x')) \leq L_1 t$ for all small enough $t > 0$ and all $x' \in \mathscr{C}_t \cap K'$. Consider the (j_t, ρ)-equivariant map

$$F_t := f \circ f_t : (N_t^\delta \cup \mathscr{C}_t) \longrightarrow \mathbb{H}^2,$$

where $f : \mathbb{H}^2 \to \mathbb{H}^2$ is the (j_0, ρ)-equivariant map from the beginning of the proof. In order to prove (4-3), it is sufficient to establish the following:

Lemma 4.2. *For small enough $t > 0$, there exists $C < 1$ such that for all $p, q \in \partial N_t^\delta$ lying at distance δ from a line $\ell_t \subset \mathscr{C}_t$, on opposite sides of ℓ_t,*

$$d(F_t(p), F_t(q)) \leq C d(p, q).$$

Indeed, fix a small $t > 0$. Any geodesic segment $I = [p, q]$ of $\mathbb{H}^2$ projecting to a closed geodesic of $j_t(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$ may be decomposed into subsegments $I_1, \ldots, I_n$ contained in N_t^δ alternating with subsegments $I'_1, \ldots, I'_n$ crossing connected components of $\mathbb{H}^2 \smallsetminus N_t^\delta$ (indeed, any simple closed curve that enters one of these components crosses it, by the choice of δ). By construction, the map F_t has Lipschitz constant strictly less than 1 on each connected component of N_t^δ, hence moves the endpoints of each I_k closer together by a uniform factor (independent of I). Lemma 4.2 ensures that the same holds for the I'_k. Thus the ratio $d(F_t(p), F_t(q))/d(p, q)$ is bounded by some factor $C' < 1$ independent of I, and the corresponding element $\gamma \in \Gamma_g$ satisfies $\lambda(\rho(\gamma)) \leq C'\lambda(j_t(\gamma))$. This proves (4-3), hence completes the proof of the first statement of Theorem 1.1.

4C. *Proof of Lemma 4.2.*

We first make the following observation:

Observation 4.3. *There exists $L' \geq 0$ such that, for any small enough $t > 0$, any $p \in \partial N_t^\delta$ at distance δ from a geodesic $\ell_t \subset \mathscr{C}_t$, and any $x \in \ell_t$,*

$$d(f_t(p), f_t(x)) \leq (1 - t)\, d(p, x) + L't.$$

Proof. Since f_t is (j_t, j_0)-equivariant and $\mathscr{C}_0$ has only finitely many connected components modulo $j_0(\Gamma_g)$, we may fix a geodesic $\ell_0 \subset \mathscr{C}_0$ and prove the observation only for the geodesics $\ell_t \subset \mathscr{C}_t$ corresponding to ℓ_0. For any $t > 0$, the map f_t takes ℓ_t linearly to ℓ_0, multiplying all distances by the uniform factor $1 - t$. Let $h_t : \mathbb{H}^2 \to \mathbb{H}^2$ be the orientation-preserving map that coincides with f_t on ℓ_t, takes any line orthogonal to ℓ_t to a line orthogonal to ℓ_0, and multiplies all distances by $1 - t$ on such lines. At distance η from ℓ_t, the differential of h_t has principal values $1 - t$ and $(1 - t) \cosh((1 - t)\eta)/ \cosh \eta \leq 1 - t$ (see [Guéritaud and Kassel 2013, (A.9)]), hence $\mathrm{Lip}(h_t) \leq 1 - t$ and

$$d(f_t(x), h_t(p)) = d(h_t(x), h_t(p)) \leq (1 - t)\, d(x, p)$$

for all $x \in \ell_t$ and $p \in \mathbb{H}^2$. By the triangle inequality, it is enough to find $L' \geq 0$ such that $d(h_t(p), f_t(p)) \leq L't$ for all small enough $t > 0$ and all $p \in \partial N_t^\delta$ at distance δ from ℓ_t. Since f_t and h_t are both (j_t, j_0)-equivariant under the stabilizer S of ℓ_0 in Γ_g, and $j_t(S)$ acts cocompactly on the set $\overline{\mathcal{U}}_t$ of points at distance at most δ from ℓ_t, we may restrict to p in a compact fundamental domain of $\overline{\mathcal{U}}_t$ for $j_t(S)$. Let $K' \subset \mathbb{H}^2$ be a compact set containing such fundamental domains for all $t \in [0, t_0]$. By construction of f_t, there exists $L_1 \geq 0$ such that $d(x', f_t(x')) \leq L_1 t$ for all small enough $t > 0$ and all $x' \in \ell_t \cap K'$. By definition of h_t, this implies the existence of $L_2 \geq 0$ such that $d(p, h_t(p)) \leq L_2 t$ for all small enough $t > 0$ and all $p \in K'$. On the other hand, condition (d) of Lemma 4.1 (applied to the Γ^i and j_0^i as in Section 4B) implies the existence of $L_3 \geq 0$ such that $d(p, f_t(p)) \leq L_3 t$ for all t and $p \in \partial N_t^\delta \cap K'$. By the triangle inequality, we may take $L' = L_2 + L_3$. □

Proof of Lemma 4.2. As in the proof of Observation 4.3, we may fix a geodesic $\ell_0 \subset \mathscr{C}_0$ and restrict to the geodesics $\ell_t \subset \mathscr{C}_t$ corresponding to ℓ_0. Fix a small $t > 0$ and consider $p, q \in \partial N_t^\delta$ lying at distance δ from ℓ_t on opposite sides of ℓ_t. The segment $[p, q]$ can be subdivided at its intersection point x with ℓ_t into two subsegments to which Observation 4.3 applies, yielding

$$\begin{cases} d(f_t(p), f_t(x)) \leq (1-t)d(p, x) + L't, \\ d(f_t(x), f_t(q)) \leq (1-t)d(x, q) + L't. \end{cases} \tag{4-4}$$

Up to switching p and q, we may assume that either $[p, x]$ projects to a pair of pants labeled 0 in $j_t(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$, or $[p, x]$ projects to a pair of pants labeled -1 and $[x, q]$ to a pair of pants labeled 1.

Suppose that $[p, x]$ projects to a pair of pants labeled 0 in $j_t(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$. We first observe that, if t is small enough (independently of p), then

$$d(f_t(p), \ell_0) \geq \frac{3\delta}{4}. \tag{4-5}$$

Indeed, as in the proof of Observation 4.3, the inequality is true for $p \in \partial N_t^\delta$ in a fixed compact set K' independent of t, by condition (d) of Lemma 4.1 and (4-2), and we then use the fact that f_t is (j_t, j_0)-equivariant under the stabilizer S of ℓ_0 in Γ_g, which acts cocompactly (by j_t) on the set of points at distance δ from ℓ_t. By (4-5), if t is small enough (independently of p), then the segment $[f_t(p), f_t(x)]$ spends at least $\delta/4$ units of length in the complement $N_0^{\delta/2}$ of the $\delta/2$-neighborhood of $\mathscr{C}_0$. The point is that $\mathrm{Lip}_y(f) < 1$ for all $y \in \mathbb{H}^2 \smallsetminus \mathscr{C}_0$ projecting to a pair of pants labeled 0 in $j_0(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$, and this bound is uniform in restriction to $N_0^{\delta/2}$ since the function $p \mapsto \mathrm{Lip}_p(f)$ is upper semicontinuous and $j_0(\Gamma_g)$-invariant. Remark 2.1 thus implies the existence of a constant $\varepsilon > 0$, independent of t, ℓ_t, p, x, such that

$$d(f \circ f_t(p), f \circ f_t(x)) \leq d(f_t(p), f_t(x)) - \varepsilon. \tag{4-6}$$

Using the triangle inequality and the fact that f is 1-Lipschitz, together with (4-4) and (4-6), we find

$$\begin{aligned} d(F_t(p), F_t(q)) &\le d(f \circ f_t(p), f \circ f_t(x)) + d(f \circ f_t(x), f \circ f_t(q)) \\ &\le (1-t)d(p,x) + L't - \varepsilon + (1-t)d(x,q) + L't, \end{aligned}$$

which is bounded by $(1-t)d(p,q)$ as soon as $t \le \varepsilon/(2L')$.

Suppose that $[p,x]$ projects to a pair of pants labeled -1 and $[x,q]$ to a pair of pants labeled 1. We then use the fact that the continuous map f folds along $\ell_0 = f_t(\ell_t)$. In restriction to the connected component of $\mathbb{H}^2 \smallsetminus \mathscr{C}_0$ containing $f_t(p)$ (resp. $f_t(q)$), it is an isometry preserving (resp. reversing) the orientation. In particular, $d(F_t(p), F_t(q)) < d(f_t(p), f_t(q))$. Moreover, this inequality can be made uniform in the following sense: there exists $\varepsilon > 0$ such that

$$d(F_t(p), F_t(q)) \le d(f_t(p), f_t(q)) - \varepsilon$$

whenever $f_t(p)$ and $f_t(q)$ lie at distance at least $3\delta/4$ from ℓ_0 (which is the case for t small enough by (4-5)) and at distance at most $3L'$ from each other. By (4-4),

$$d(f_t(p), f_t(q)) \le (1-t)d(p,q) + 2L't, \tag{4-7}$$

which implies

$$d(F_t(p), F_t(q)) \le (1-t)d(p,q)$$

for $d(p,q) \le 3L'$ as soon as $t \le \varepsilon/(2L')$ is small enough. If $d(p,q) \ge 3L'$, then applying the 1-Lipschitz map f to (4-7) directly gives

$$d(F_t(p), F_t(q)) \le (1-t)d(p,q) + 2L't \le \left(1 - \tfrac{1}{3}t\right)d(p,q). \qquad \square$$

4D. ***Folding a given surface.*** We now prove the second statement of Theorem 1.1. Namely, given $[j_0] \in \mathsf{Rep}_g^{\mathrm{fd}}$ and an integer $k \in (-2g+2, 2g-2)$, we construct $[\rho] \in \mathsf{Rep}_g^{\mathrm{nfd}}$ with $\mathsf{eu}(\rho) = k$ that is strictly dominated by $[j_0]$.

It is easy to find $[\rho]$ with $\mathsf{eu}(\rho) = k$ such that $\lambda_\rho(\gamma) \le \lambda_{j_0}(\gamma)$ for all $\gamma \in \Gamma_g$: just decompose Σ_g into pairs of pants and assign arbitrary values $0, 1, -1$ to each so that the sum is k. Consider a lamination Υ of Σ_g consisting of all the cuffs together with a triskelion lamination inside each pair of pants labeled 0, and let $c : \Sigma_g \smallsetminus \Upsilon \to \{-1, 1\}$ be a coloring taking the value -1 (resp. 1) on each pair of pants labeled -1 (resp. 1), and both values on each pair of pants labeled 0. Folding along Υ with the coloring c gives an element $[\rho] \in \mathsf{Rep}_g^{\mathrm{nfd}}$ with $\lambda_\rho(\gamma) \le \lambda_{j_0}(\gamma)$ for all $\gamma \in \Gamma_g$. However, we need a *strict* domination. The idea is to obtain ρ by folding not j_0 but a small deformation of j_0. For this purpose, we use the following result, which is analogous to Lemma 4.1.

Lemma 4.4. *Let Γ be the fundamental group and $j_0 \in \mathrm{Hom}(\Gamma, \mathrm{PSL}(2,\mathbb{R}))$ the holonomy of a compact, connected hyperbolic surface Σ with nonempty geodesic* *boundary. Then there exist $t_0 > 0$ and a continuous family of representations $(j_t)_{0\leq t\leq t_0}$ with the following properties*:

(a) $\lambda_{j_t}(\gamma) = (1-t)\lambda_{j_0}(\gamma)$ *for any* $t \in [0, t_0]$ *and any* $\gamma \in \Gamma$ *corresponding to a boundary component of* Σ;

(b) $\sup_{\gamma\in\Gamma\smallsetminus\{1\}} \lambda_{j_t}(\gamma)/\lambda_{j_0}(\gamma) < 1$ *for any* $t \in (0, t_0]$;

(c) $j_t(\gamma) = j_0(\gamma) + O(t)$ *for any* $\gamma \in \Gamma$ *as* $t \to 0$, *where both sides are seen as* 2×2 *real matrices with determinant* 1 *modulo* $\pm\,\mathrm{Id}$;

(d) *for any compact subset* K *of* $\mathbb{H}^2$ *projecting to the interior of the convex core of* $j_0(\Gamma)\backslash\mathbb{H}^2$, *there exists* $L > 0$ *such that*

$$d(p, f_t(p)) \leq Lt$$

for any $p \in K$, *any* $t \in [0, t_0]$, *and any* 1*-Lipschitz*, (j_0, j_t)*-equivariant map* $f_t : \mathbb{H}^2 \to \mathbb{H}^2$.

As in the proof of Lemma 4.1, we construct the representations j_t as holonomies of hyperbolic surfaces obtained from $j_0(\Gamma)\backslash\mathbb{H}^2$ by deformation. Now the deformation needs to be shortening instead of lengthening, so we use *negative* strip deformations.

Proof of Lemma 4.4. We view Σ as the convex core of $j_0(\Gamma)\backslash\mathbb{H}^2$. To shorten one boundary component β of Σ, choose a finite collection of disjoint, biinfinite geodesic arcs $\alpha_1, \dots, \alpha_n \subset j_0(\Gamma)\backslash\mathbb{H}^2$, each crossing β orthogonally twice, subdividing Σ into right-angled hexagons and one-holed right-angled bigons. Near each α_i, choose a second geodesic arc α'_i, also crossing β twice, such that α_i, α'_i are closest at some points $p_i,\ p'_i \in \Sigma$. We take all arcs to be pairwise disjoint. For every i, delete the hyperbolic strip A_i bounded by α_i and α'_i and glue the arcs back together isometrically, identifying p_i with p'_i. This yields a new complete hyperbolic surface with a compact convex core, equipped with a natural 1-Lipschitz map ς_t^β *from* $j_0(\Gamma)\backslash\mathbb{H}^2$ obtained by collapsing the strips A_i to lines. The set $\varsigma_t^\beta(\Sigma)$ is strictly contained in the new convex core. The geodesic corresponding to β is shorter in the new surface than in Σ. By adjusting the widths of the strips A_i, we may assume that the ratio of lengths is $1/(1-t)$. Note that the appropriate widths for this ratio are in $O(t)$ as $t \to 0$. All lengths of geodesics corresponding to boundary components other than β are unchanged.

Repeat the construction, iteratively, for all boundary components $\beta_1, \dots, \beta_r$ of Σ, in some arbitrary order. We thus obtain a new complete hyperbolic surface $j_t(\Gamma)\backslash\mathbb{H}^2$, with a compact convex core Σ_t, such that j_t satisfies (a). As in the proof of Lemma 4.1, up to replacing each j_t with a conjugate under $\mathrm{PSL}(2,\mathbb{R})$, we may assume that (c) is satisfied. To see that (b) and (d) also hold, we use the

1-Lipschitz map $\varsigma_t := \varsigma_t^{\beta_r} \circ \cdots \circ \varsigma_t^{\beta_1}$ from Σ to Σ_t and argue as in the proof of Lemma 4.1, switching j_t and j_0. $\square$

As in Section 4B, we write $\Sigma_g = \Sigma^1 \cup \cdots \cup \Sigma^m$, where Σ^i, for any $1 \leq i \leq m$, is a compact surface with boundary that is one of:

- a pair of pants labeled 0,
- a full connected component of the subsurface of Σ_g made of pants labeled -1,
- or a full connected component of the subsurface of Σ_g made of pants labeled 1.

Choose a small $\delta > 0$ such that, in all hyperbolic metrics on Σ_g which are close enough to that defined by j_0, any simple geodesic entering the δ-neighborhood of the geodesic representative of a cuff of our chosen pants decomposition crosses the cuff. We use again the notation $\mathscr{C}_0$, N_0^δ, K from Section 4B. Applying Lemma 4.4 to $\Gamma^i := \pi_1(\Sigma^i)$ and $j_0^i := j_0|_{\Gamma^i}$, we obtain continuous families of representations $(j_t^i)_{0 \leq t \leq t_0}$ for $1 \leq i \leq m$ satisfying (a)–(d), with a uniform constant $L > 0$ for the compact set $K \subset \mathbb{H}^2 \smallsetminus \mathscr{C}_0$. For any $t \geq 0$, using (a), we can glue together the convex cores of the $j_t^i(\Gamma^i)\backslash\mathbb{H}^2$ following the same combinatorics as the Σ^i. This gives a closed hyperbolic surface of genus g, hence a holonomy representation $j_t \in \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))$. By (c), up to adjusting the twist parameters, we may assume that $j_t(\gamma) = j_0(\gamma) + O(t)$ for any $\gamma \in \Gamma_g$ as $t \to 0$, where both sides are seen as 2×2 real matrices with determinant 1 modulo $\pm\,\mathrm{Id}$.

Recall the notation $\mathscr{C}_t$, N_t^δ from Section 4B. By Proposition 3.9, there exist a family $(\rho_t)_{0 \leq t \leq t_0} \subset \mathrm{Hom}(\Gamma_g, \mathrm{PSL}(2,\mathbb{R}))$ of non-Fuchsian representations and, for any $t \in [0, t_0]$, a 1-Lipschitz, (j_t, ρ_t)-equivariant map $\varphi_t : \mathbb{H}^2 \to \mathbb{H}^2$ that is an orientation-preserving (resp. orientation-reversing) isometry in restriction to any connected subset of $\mathbb{H}^2$ projecting to a union of pants labeled -1 (resp. 1) in $j_t(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$, such that

$$\mathrm{Lip}_p(\varphi_t) \leq C^* < 1 \tag{4-8}$$

for all $t \in [0, t_0]$ and all $p \in N_t^\delta$ that project to a pair of pants labeled 0 in $j_t(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$, for some $C^* < 1$ independent of p and t.

We claim that, for $t > 0$ small enough,

$$\sup_{\gamma \in (\Gamma_g)_s} \frac{\lambda_{\rho_t}(\gamma)}{\lambda_{j_0}(\gamma)} < 1, \tag{4-9}$$

which by Theorem 2.5 is enough to prove that $[\rho_t]$ is strictly dominated by $[j_0]$. Indeed, by (b) and Theorem 2.5, for any $1 \leq i \leq m$ and $t \in (0, t_0]$, there exists a $(j_t|_{\Gamma^i}, j_0|_{\Gamma^i})$-equivariant map $f_t^i : \mathbb{H}^2 \to \mathbb{H}^2$ with $\mathrm{Lip}(f_t^i) < 1$. Let f_t be a (j_0, j_t)-equivariant map $(N_0^\delta \cup \mathscr{C}_0) \to \mathbb{H}^2$ such that:

- $f_t = f_t^i$ on the component of K projecting to Σ^i for all $1 \leq i \leq m$;

- f_t takes any geodesic line in $\mathscr{C}_0$ to the corresponding line in $\mathscr{C}_t$, multiplying all distances by the uniform factor $(1-t)$, and $d(x, f_t(x)) \leq L_1 t$ for all $x \in \mathscr{C}_0 \cap K$, for some $L_1 \geq 0$ independent of x and t.

Consider the (j_0, ρ_t)-equivariant map

$$G_t := \varphi_t \circ f_t : (N_0^\delta \cup \mathscr{C}_0) \longrightarrow \mathbb{H}^2.$$

Any geodesic segment $I = [p,q]$ of $\mathbb{H}^2$ that projects to a closed geodesic of $j_0(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$ may be decomposed into subsegments $I_1, \dots, I_n$ contained in N_0^δ alternating with subsegments $I_1', \dots, I_n'$ crossing connected components of $\mathbb{H}^2 \smallsetminus N_0^\delta$. By contractivity of f_t, the map G_t has Lipschitz constant strictly less than 1 on each connected component of N_0^δ, hence moves the endpoints of each I_k closer together by a uniform factor (independent of I). The subsegments I_k' are treated by the following lemma, which implies (4-9) and therefore completes the proof of the second statement of Theorem 1.1.

Lemma 4.5 (analogue of Lemma 4.2). *For small enough $t > 0$, there exists $C < 1$ such that, for all $p, q \in \partial N_0^\delta$ lying at distance δ from a line $\ell_0 \subset \mathscr{C}_0$ on opposite sides of ℓ_0,*

$$d(G_t(p), G_t(q)) \leq C d(p,q).$$

The proof of Lemma 4.5 uses the following observation, which is identical to Observation 4.3 after exchanging j_0 and j_t.

Observation 4.6. *There exists $L' \geq 0$ such that, for any small enough $t > 0$, any $p \in \partial N_0^\delta$ at distance δ from a geodesic $\ell_0 \subset \mathscr{C}_0$, and any $x \in \ell_0$,*

$$d(f_t(p), f_t(x)) \leq (1-t) d(p,x) + L't. \tag{4-10}$$

Proof of Lemma 4.5. We argue as in the proof of Lemma 4.2, but switch j_0 and j_t and use (4-8) to obtain the analogue

$$d(\varphi_t \circ f_t(p), \varphi_t \circ f_t(x)) \leq d(f_t(p), f_t(x)) - \varepsilon$$

of (4-6) when $[p,x]$ projects to a pair of pants labeled 0 in $j_0(\Gamma_g)\backslash\mathbb{H}^2 \simeq \Sigma_g$. $\square$

Acknowledgements

We would like to thank Bertrand Deroin and Nicolas Tholozan for sharing their thoughts on the subject with us. The third author is grateful to Antonin Guilloux, Julien Marché, and Richard Wentworth for their support and for helpful conversations.

References

[Bonahon 1996] F. Bonahon, "Shearing hyperbolic surfaces, bending pleated surfaces and Thurston's symplectic form", *Ann. Fac. Sci. Toulouse Math.* (6) **5**:2 (1996), 233–297. MR 97i:57011 ZBL 0880.57005

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

[Calegari 2004] D. Calegari, "Circular groups, planar groups, and the Euler class", pp. 431–491 in *Proceedings of the Casson Fest* (Fayetteville, AR/Austin, TX, 2003), edited by C. Gordon and Y. Rieck, Geometry and Topology Monographs **7**, Mathematical Sciences Publishers, Coventry, 2004. MR 2006i:57038 Zbl 1181.57022

[Danciger et al. 2014] J. Danciger, F. Guéritaud, and F. Kassel, "Margulis spacetimes via the arc complex", preprint, 2014. arXiv 1407.5422

[Deroin and Tholozan 2013] B. Deroin and N. Tholozan, "Dominating surface group representations by Fuchsian ones", preprint, 2013. arXiv 1311.2919

[Gallo et al. 2000] D. Gallo, M. Kapovich, and A. Marden, "The monodromy groups of Schwarzian equations on closed Riemann surfaces", *Ann. of Math.* (2) **151**:2 (2000), 625–704. MR 2002j:57029 Zbl 0977.30028

[Ghys 2001] É. Ghys, "Groups acting on the circle", *Enseign. Math.* (2) **47**:3-4 (2001), 329–407. MR 2003a:37032 Zbl 1044.37033

[Goldman 1985] W. M. Goldman, "Nonstandard Lorentz space forms", *J. Differential Geom.* **21**:2 (1985), 301–308. MR 87h:53093 Zbl 0591.53051

[Goldman 1987] W. M. Goldman, "Projective structures with Fuchsian holonomy", *J. Differential Geom.* **25**:3 (1987), 297–326. MR 88i:57006 Zbl 0595.57012

[Goldman 1988] W. M. Goldman, "Topological components of spaces of representations", *Invent. Math.* **93**:3 (1988), 557–607. MR 89m:57001 Zbl 0655.57019

[Goldman and Xia 2011] W. M. Goldman and E. Z. Xia, "Ergodicity of mapping class group actions on SU(2)-character varieties", pp. 591–608 in *Geometry, rigidity, and group actions* (Chicago, 2007), edited by B. Farb and D. Fisher, University of Chicago Press, 2011. MR 2012k:22028 Zbl 1283.22016 arXiv 0901.1402

[Guéritaud and Kassel 2013] F. Guéritaud and F. Kassel, "Maximally stretched laminations on geometrically finite hyperbolic manifolds", preprint, 2013. arXiv 1307.0250

[Guéritaud et al. 2015] F. Guéritaud, O. Guichard, F. Kassel, and A. Wienhard, "Anosov representations and proper actions", preprint, 2015. To appear in *Geom. Topol.* arXiv 1502.03811

[Hitchin 1987] N. J. Hitchin, "The self-duality equations on a Riemann surface", *Proc. London Math. Soc.* (3) **55**:1 (1987), 59–126. MR 89a:32021 Zbl 0634.53045

[Kassel 2008] F. Kassel, "Proper actions on corank-one reductive homogeneous spaces", *J. Lie Theory* **18**:4 (2008), 961–978. MR 2010g:22019 Zbl 1173.22009

[Kassel 2009] F. Kassel, *Quotients compacts d'espaces homogènes réels ou p-adiques*, Ph.D. thesis, Université Paris-Sud 11, 2009, Available at http://math.univ-lille1.fr/~kassel/These.pdf.

[Klingler 1996] B. Klingler, "Complétude des variétés Lorentziennes à courbure constante", *Math. Ann.* **306**:2 (1996), 353–370. MR 97g:53082 Zbl 0862.53048

[Kobayashi 1998] T. Kobayashi, "Deformation of compact Clifford–Klein forms of indefinite-Riemannian homogeneous manifolds", *Math. Ann.* **310**:3 (1998), 395–409. MR 99b:53074 ZBL 0891.22014

[Kulkarni and Raymond 1985] R. S. Kulkarni and F. Raymond, "3-dimensional Lorentz space-forms and Seifert fiber spaces", *J. Differential Geom.* **21**:2 (1985), 231–268. MR 87h:53092 Zbl 0563.57004

[Papadopoulos and Théret 2010] A. Papadopoulos and G. Théret, "Shortening all the simple closed geodesics on surfaces with boundary", *Proc. Amer. Math. Soc.* **138**:5 (2010), 1775–1784. MR 2011d:30111 Zbl 1197.32006

[Salein 1999] F. Salein, *Variétés anti-de Sitter de dimension* 3, Ph.D. thesis, École Normale Supérieure de Lyon, 1999, Available at http://www.umpa.ens-lyon.fr/∼zeghib/these.salein.pdf.

[Salein 2000] F. Salein, "Variétés anti-de Sitter de dimension 3 exotiques", *Ann. Inst. Fourier* (*Grenoble*) **50**:1 (2000), 257–284. MR 2001h:57020 Zbl 0951.53047

[Selberg 1960] A. Selberg, "On discontinuous groups in higher-dimensional symmetric spaces", pp. 147–164 in *Contributions to function theory* (Bombay, 1960), Tata Institute of Fundamental Research, Bombay, 1960. Reprinted in *Collected papers*, A. Selberg, Vol. 1, pp. 475–492, Springer, Berlin, 1989. MR 24 #A188 Zbl 0201.36603

[Thurston 1980] W. P. Thurston, "The geometry and topology of three-manifolds", 1980, Available at http://library.msri.org/books/gt3m. Lecture notes.

[Thurston 1986] W. P. Thurston, "Minimal stretch maps between hyperbolic surfaces", preprint, 1986. arXiv math/9801039

[Wolff 2011] M. Wolff, "Connected components of the compactification of representation spaces of surface groups", *Geom. Topol.* **15**:3 (2011), 1225–1295. MR 2825313 Zbl 1226.57027

Received November 16, 2013. Revised July 7, 2014.

FRANÇOIS GUÉRITAUD
CNRS AND UNIVERSITÉ LILLE 1
LABORATOIRE PAUL PAINLEVÉ, UNIVERSITÉ LILLE 1
59655 VILLENEUVE D'ASCQ CEDEX
FRANCE

francois.gueritaud@math.univ-lille1.fr

FANNY KASSEL
CNRS AND UNIVERSITÉ LILLE 1
LABORATOIRE PAUL PAINLEVÉ, UNIVERSITÉ LILLE 1
59655 VILLENEUVE D'ASCQ CEDEX
FRANCE

fanny.kassel@math.univ-lille1.fr

MAXIME WOLFF
INSTITUT DE MATHÉMATIQUES DE JUSSIEU-PARIS RIVE GAUCHE, CNRS
UNIVERSITÉ PIERRE ET MARIE CURIE - PARIS 6
4 PLACE JUSSIEU - CASE 247
75005 PARIS 05
FRANCE

maxime.wolff@imj-prg.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 275, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.275.361

CIRCULAR HANDLE DECOMPOSITIONS OF FREE GENUS ONE KNOTS

FABIOLA MANJARREZ-GUTIÉRREZ,
VÍCTOR NÚÑEZ AND ENRIQUE RAMÍREZ-LOSADA

We determine the structure of the circular handle decompositions of the family of free genus one knots. Namely, if k is a free genus one knot, then the handle number $h(k) = 0$, 1 or 2, and, if k is not fibered (that is, if $h(k) > 0$), then k is almost fibered. For this, we develop *practical* techniques to construct circular handle decompositions of knots with free Seifert surfaces in the 3-sphere (and compute handle numbers of many knots), and, also, we characterize the free genus one knots with more than one Seifert surface. These results are obtained through analysis of spines of surfaces on handlebodies. Also we show that there are infinite families of free genus one knots with either $h(k) = 1$ or $h(k) = 2$.

1. Introduction

In the study of the topology of a given 3-manifold, M, it has been useful to consider regular real-valued Morse functions $f : M \to \mathbb{R}$, where M has some smooth structure. A regular real-valued Morse function on M corresponds to a handle decomposition of M of the form $M = b_0 \cup B_1 \cup P_1 \cup \cdots \cup B_r \cup P_r \cup b_3$, where b_0 is a collection of 0-handles, B_j is a collection of 1-handles, P_j is a collection of 2-handles, and b_3 is a collection of 3-handles, in such a way that the i-handles of the decomposition are neighborhoods of the critical points of index i of the Morse function ($j = 1, \ldots, r$, and $i = 0, 1, 2, 3$). In the celebrated paper [Scharlemann and Thompson 1994], the concept of *thin position* for 3-manifolds is introduced; the idea is to build the manifold as described above (that is, step by step: adding to the set b_0 the set B_1, and then adding P_1, and then adding B_2, and so on) with a sequence of sets of 1-handles and sets of 2-handles chosen to keep the boundaries of the intermediate steps as simple as possible.

Now if a 3-manifold M satisfies $H^1(M; \mathbb{Q}) \neq 0$, then there are essential (non-nullhomotopic) regular Morse functions $f : M \to S^1$, and one can always find

MSC2010: 57M25.
Keywords: circular thin position, free genus, free genus one knots, Seifert surfaces, handle decompositions, almost fibered.

such functions having only critical points of index 1 and 2 (see Section 2B). Such a function corresponds to a *circular handle decomposition*

$$M = F \times [0,1] \cup B_1 \cup P_1 \cup \cdots \cup B_r \cup P_r,$$

where F is a properly embedded surface in M, B_j is a collection of 1-handles, and P_j is a collection of 2-handles (the handles are glued along, say, $F \times \{1\}$), and, as above, the set of i-handles of the decomposition corresponds to the critical points of index i of the Morse function. With this kind of circular handle decomposition we may also require that the intermediate steps be as simple as possible: that requirement leads to the notion of thin position for circular handle decompositions. The existence of these decompositions gives rise to numerical topological invariants such as the (circular) handle number $h(M) = \sum_{1=1}^{r} \#(B_i)$, where the sum $\sum \#(B_i)$ is minimal among all circular handle decompositions; also, when the decomposition is in thin position, we obtain the circular width $\mathrm{cw}(M)$ (see Section 2D).

Outstanding examples of manifolds that admit circular handle decompositions are the exteriors of links in S^3. In this case the interesting intermediate surfaces in the decomposition are Seifert surfaces for the given link. (These intermediate surfaces have no closed components, and, if the decomposition is in thin position, they are a sequence of Seifert surfaces which are alternately incompressible and weakly incompressible. See [Manjarrez 2009, Theorem 3.2], where there is a statement for knots, but its proof works verbatim for links.)

If the exterior of a link ℓ in S^3 admits a circular decomposition of the form $E(\ell) = F \times [0,1] \cup B_1 \cup P_1$, and this decomposition is in thin position, we say that ℓ is an *almost fibered link*. One may regard the set of almost fibered knots as the set of knots with the simplest nontrivial circular handle structure.

Thus, an interesting problem of this theory is to determine the set of all almost fibered knots. We solve this problem for the family of free genus one knots. In fact, we show that all free genus one knots are almost fibered (Theorem 6.7).

Also it is interesting to find explicit constructions of circular handle decompositions of the exterior of a given link which are minimal (that is, that realize the handle number), or that are in thin position. In [Goda 1993], although in a different context, explicit minimal circular handle decompositions of the exterior of the 250 knots in Rolfsen's table are given. Of these knots, 117 are fibered and 132 have handle number one. As far as we know, there are no other previously published explicit constructions of circular handle decompositions of exteriors of links in the 3-sphere.

As mentioned above, in this paper we are interested mainly in the circular handle structures of the family of free genus one knots.

In the first part of this work (Section 3) we develop techniques to construct explicit circular decompositions of link exteriors for links that admit a free Seifert surface;

these decompositions are interesting, of course, when the free Seifert surface used in the construction is of minimal genus for the link. The information needed to construct these decompositions for the exterior of a given link is encoded in some spine of a free Seifert surface of the link. In this sense, the techniques developed in Section 3 (and throughout this paper) could be regarded as elements for a possible theory of spines of surfaces on handlebodies that might be worthy of consideration. As applications we construct minimal circular decompositions for all rational knots and links and, also, for a family of pretzel knots, namely, pretzel knots of the form $P(\pm 3, q, r)$ with $|q|, |r|$ odd integers ≥ 3. These circular decompositions for both families of links are all minimal and have handle number one; they are also in thin position, giving also the circular width of each link considered. This last family gives examples of nonfibered knots whose handle number is strictly less than their tunnel number (Remark 3.10). Also, it is shown that free genus one knots have handle number at most 2 (Corollary 3.5).

Secondly (Section 4), we construct circular handle decompositions for the exteriors of all pretzel knots of the form $P(p, q, r)$ with $|p|, |q|, |r|$ odd integers ≥ 5, and we show that these decompositions are minimal with handle number two (Theorem 4.1), and are also in thin position, giving the circular width equal to 6 for each of these knots. These examples answer a question posed in [Veber et al. 2001] (Remark 4.5).

Next, in Section 5, we give a characterization of the free genus one knots that admit at least two different (nonparallel) Seifert surfaces of genus one. This characterization is given in terms of the existence of a special spine for the given genus one free Seifert surface of the knot (see Theorem 5.2).

Using the characterization given in Section 5 we show, in the final part of this work, that all (nonfibered) free genus one knots are almost fibered (Theorem 6.7).

It follows from the proof of Theorem 6.7 that the free genus one knots with handle number two have a unique minimal-genus Seifert surface (that is, free genus one knots with at least two genus one Seifert surfaces have handle number one). It is an interesting open problem to determine the family of free genus one knots with handle number two.

2. Preliminaries

Unless explicitly stated, we will use the word "knot" for a knot or a link in S^3. That is, we will emphasize connectedness if needed. Otherwise, we will admit nonconnected knots.

Let X be a manifold and let $Y \subset X$ be a subcomplex. We write $E(Y) = \overline{X - \mathcal{N}(Y)}$ for the *exterior* of Y in X, where $\mathcal{N}(Y)$ is a regular neighborhood of Y in X.

Let X be a manifold and let $Y \subset X$ be a properly embedded submanifold. Y is called *∂-parallel* in X, or *parallel into* ∂X, if there is an embedding

$$e : (Y, \partial Y) \times I \to (X, \partial X)$$

such that $e_0 : Y \to Y$ is the identity, and $e_1(Y) \subset \partial X$. If Y is ∂-parallel in X with embedding $e : (Y, \partial Y) \times I \to (X, \partial X)$, then the submanifold $e(Y \times I)$ is called a *∂-parallelism for* Y. Notice that if Y is disconnected with components $Y_1, \dots, Y_n$, and Y is ∂-parallel in X with a ∂-parallelism W, then W is a disjoint union of ∂-parallelisms $W_1, \dots, W_n$ for $Y_1, \dots, Y_n$, respectively.

2A. ***Seifert surfaces.*** Let $k \subset S^3$ be a knot, and let F be a Seifert surface for k; that is, F is an orientable surface and $\partial F = k$. Then, by drilling out a small neighborhood $\mathcal{N}(k)$ of k, the surface $\widehat{F} = F \cap E(k)$ is a properly embedded surface in $E(k)$, the exterior of k in S^3, and one may assume that $\partial\widehat{F}$ is parallel to k in $\mathcal{N}(k)$. Usually we identify F with $\widehat{F}$; but, more appropriately, we start with a Seifert surface $F \subset E(k)$ for k. Seifert surfaces may be disconnected, but they are not allowed to contain closed components. The *genus* $g(k)$ of a knot k is the minimal genus among all Seifert surfaces for k.

A surface $F \subset S^3$ is called *free* if $E(F)$ is a handlebody. The *free genus* $g_f(k)$ of a knot k is the minimal genus among all free Seifert surfaces for k.

In this work we will be interested mainly in free genus one knots.

2B. ***Handle decompositions of rel-∂ cobordisms.*** Let W be a cobordism rel ∂ between surfaces $\partial_+ W$ and $\partial_- W$ with no closed components. A *moderate handle decomposition of* W is a decomposition of the form

$$W \cong (\partial_+ W \times I) \cup (\text{1-handles}) \cup (\text{2-handles}).$$

Given W, a cobordism rel ∂ between surfaces $\partial_+ W$ and $\partial_- W$ with no closed components, it is easy to find a moderate decomposition as above by considering a triangulation of the exterior $E(\partial_+ W) = \overline{W - \mathcal{N}(\partial_+ W)}$.

Given a cobordism W and a moderate handle decomposition for W, one can find a regular Morse function $f : W \to I$ which realizes the handle decomposition of W. That is, f only has critical points of index 1 and 2, neighborhoods of the critical points of f correspond to the 1- and 2-handles of W, and the preimage of each regular value of f is a properly embedded surface in W. We will call such a Morse function a *moderate Morse function* (see [Veber et al. 2001]).

2C. ***Circular decompositions.*** Let k be a knot in S^3. Since $H_1(E(k))$ is a free abelian group of positive rank, we can always find an essential (non-nullhomotopic) moderate Morse function $f : E(k) \to S^1$. Any such Morse function, as in

Section 2B, induces a decomposition

$$E(k) = (F \times I) \cup B \cup P,$$

where $F \subset E(k)$ is a Seifert surface for k, B is a set of n 1-handles glued along, say, $F \times \{1\}$, and P is a set of the same number, n, of 2-handles glued along the same side.

We call such a decomposition a *circular handle decomposition of* $E(k)$ *based on* F, and write $h(F) = n$, the *handle number of* F, where n is the minimal number of 1-handles among all circular handle decompositions of $E(k)$ based on F. The *circular handle number* $h(k)$ *of* k, or simply the *handle number of* k, is the minimal $h(F)$ among all Seifert surfaces $F \subset E(k)$. Notice that $h(k) = 0$ if and only if k is a fibered knot.

By rearranging the critical points of a moderate Morse function $f : E(k) \to S^1$, we can thin a circular handle decomposition of $E(k)$:

$$E(k) = (F \times I) \cup B_1 \cup P_1 \cup B_2 \cup P_2 \cup \cdots \cup B_\ell \cup P_\ell,$$

where B_i is a set of 1-handles glued along $F \times \{1\}$ and P_i is a set of 2-handles, $i = 1, \ldots, \ell$ (of course, it is not always possible to thin a given circular handle decomposition).

For $i = 1, \ldots, \ell$, the set $W_i = \left(F \times \left[\frac{1}{2}, 1\right]\right) \cup B_1 \cup P_1 \cup \cdots \cup B_i$ gives a moderate handle decomposition for the rel-∂ cobordism W_i with $\partial_+ W_i = F \times \left\{\frac{1}{2}\right\}$. Write $S_i = \partial_- W_i$. Now we define

$$c(S_i) = \sum_{j=1}^{n_i} (1 - \chi(G_{i,j})),$$

where χ stands for Euler characteristic and $G_{i,1}, \ldots, G_{i,n_i}$ are the components of S_i. (Notice that there are no closed components of S_i, for F has no closed components and the handle decomposition is moderate). Order the surfaces $S_{\sigma(1)}, \ldots, S_{\sigma(\ell)}$ in such a way that $c(S_{\sigma(i)}) \geq c(S_{\sigma(i+1)})$ for $i = 1, \ldots, \ell - 1$, where σ is a permutation on the symbols $1, \ldots, \ell$. Then the *circular width* of this decomposition is the tuple $(c(S_{\sigma(1)}), \ldots, c(S_{\sigma(\ell)}))$. The *circular width* $\mathrm{cw}(k)$ *of* k is the minimal circular width, with respect to lexicographic order, among all thinned circular decompositions of $E(k)$ based on all possible Seifert surfaces for k.

Let $k \subset S^3$ be a knot whose circular width has the form $\mathrm{cw}(k) = (n)$. Then we write $\mathrm{cw}(k) = n$, or $\mathrm{cw}(k) \in \mathbb{Z}$. If k is a nonfibered knot and $\mathrm{cw}(k) \in \mathbb{Z}$, then k is said to be an *almost fibered* knot.

Remark 2.1 (equivalence of knots). Let $k, \ell \subset S^3$ be two knots. If the pairs (S^3, k) and (S^3, ℓ) are homeomorphic, then their exteriors also are homeomorphic, i.e., $E(k) \cong E(\ell)$; therefore, the exteriors of k and ℓ have homeomorphic handle decompositions. We regard two knots as being *equivalent* if their corresponding pairs are homeomorphic.

Remark 2.2 (construction of circular decompositions). To describe, or, rather, to actually *construct* a decomposition

$$E(k) = (F \times I) \cup B \cup P,$$

where B is a set of 1-handles and P is a set of 2-handles, it is convenient to write

$$E(k) = \left(F \times \left[\tfrac{1}{2}, 1\right]\right) \cup B \cup P \cup \left(F \times \left[0, \tfrac{1}{2}\right]\right).$$

Then, to obtain (describe) this circular decomposition, we have dual options:

(1) Start with a regular neighborhood $\mathcal{N}(F)$ of F in $E(k)$. Then add a number of 1-handles to $\mathcal{N}(F)$ (the elements of B) on one side, say $F \times \{1\}$, and then add the same number of 2-handles (the elements of P) on the same side. The complement of the union above is a regular neighborhood of $F \times \{0\}$ in $E(k)$. Or,

(2) Start with $E(F)$, the exterior of F in $E(k)$. Then drill a number of 2-handles (the elements of B) out of $E(F)$. Now drill the same number of 1-handles (the elements of P) out of $E(F)$. Here one should be careful that the drilled-out 2-handles intersect $\partial E(F)$ on the same side, say $F \times \{1\}$, and that the following drilled-out 1-handles intersect the remaining boundary of $E(F)$ on the same side. The result of this drilling is a regular neighborhood of $F \times \{0\}$ in $E(k)$.

Of course, in (1) above, $\mathcal{N}(F)$ stands for $F \times \left[\frac{1}{2}, 1\right]$, and in (2), $E(F)$ stands for the exterior

$$\overline{E(k) - F \times \left[\tfrac{1}{2}, 1\right]}.$$

To describe a thinned circular decomposition, one proceeds similarly, but now there will be several steps. Note that in a thinned decomposition the number of 1-handles and the number of 2-handles at each step are not necessarily the same.

We emphasize that the main use of the program outlined in (1) is to describe an explicit circular handle decomposition of some given example.

Remark 2.3 (decompositions of non-almost-fibered knots). Now start with a circular decomposition

$$E(k) = \left(F \times \left[\tfrac{1}{2}, 1\right]\right) \cup B_1 \cup P_1 \cup B_2 \cup P_2 \cup \cdots \cup B_\ell \cup P_\ell \cup \left(F \times \left[0, \tfrac{1}{2}\right]\right)$$

which realizes $\mathrm{cw}(k)$, the circular width of k. For $i = 1, \ldots, \ell$, the set

$$V_i = (F \times \left[\tfrac{1}{2}, 1\right]) \cup B_1 \cup P_1 \cup \cdots \cup B_i \cup P_i$$

gives a moderate handle decomposition for the rel-∂ cobordism V_i with $\partial_+ V_i = F \times \left\{\frac{1}{2}\right\}$. Write $T_i = \partial_- V_i$. Then the ℓ disjoint surfaces $T_1, T_2, \ldots, T_\ell = F$ are incompressible in $E(k)$ and are pairwise nonparallel (see [Manjarrez 2009, Theorem 3.2]; as noted in the introduction, the theorem also holds for nonconnected knots). That is, *if k is nonfibered and not an almost fibered knot, then k has at least two nonparallel incompressible Seifert surfaces.*

Remark 2.4 (decompositions of pairs). Let $k \subset S^3$ be a knot with Seifert surface $F \subset E(k)$. There is a copy $F_0 \subset \partial E(F)$ of F such that $E(F)$ is a cobordism rel ∂ between $F_0 = \partial_+ E(F)$ and $\partial_- E(F)$. We commit an abuse of notation by identifying F with F_0. To find a circular decomposition of $E(k)$ based on F is the same as finding a moderate handle decomposition of the rel-∂ cobordism $E(F)$. A *handle decomposition of the pair* $(E(F), F)$ is, by definition, a handle decomposition of the rel-∂ cobordism $E(F)$.

Now let $\ell \subset S^3$ be another knot with Seifert surface $G \subset E(\ell)$. If there is a homeomorphism of pairs $(E(F), F) \cong (E(G), G)$, then the handle decompositions of the pairs $(E(F), F)$ and $(E(G), G)$ (as well as those of $E(F)$ and $E(G)$ as rel-∂ cobordisms) are in one-to-one correspondence via the given homeomorphism. That is, *to find circular decompositions of $E(k)$ based on F, we need only to construct moderate handle decompositions of the homeomorphism class of the pair $(E(F), F)$. In particular, it is not necessary to regard $E(F)$ as embedded in S^3.*

This remark is very helpful in the search for circular decompositions.

2D. *Spines.* Let X be either a handlebody or a surface with boundary. A *spine* of X is a graph $\Gamma \subset X$ such that X is a regular neighborhood of Γ. In this work we mainly consider spines of the form $\Gamma \cong \bigvee_{i=1}^{n} S^1$, a wedge of circles. We write $\Gamma = a_1 \vee \cdots \vee a_n$ to emphasize the circles involved, and we assume that the curves a_i carry a given orientation. Notice that it is allowed for Γ to be a single simple closed curve.

Let $k \subset S^3$ be a knot and let $F \subset E(k)$ be a Seifert surface for k. A regular neighborhood $\mathcal{N}(F)$ of F in $E(k)$ admits a product structure $\mathcal{N}(F) = F \times I$, where $\partial F \times I = \mathcal{N}(k) \cap \mathcal{N}(F)$. A spine $\Gamma \subset F \times \{0\}$, $\Gamma \cong \bigvee_{i=1}^{n} S^1$, is also a spine for $\mathcal{N}(F)$, and the graph Γ induces a product structure $\mathcal{N}(F) = G \times I$, where, say, $G \times \{0\}$ is a regular neighborhood of Γ in $\partial\mathcal{N}(F)$ (here, of course, G is isotopic to F in $\partial\mathcal{N}(F)$). A spine $\Gamma \subset F \times \{0\}$ is also a graph $\Gamma \subset \partial E(F)$. A spine for F, $\Gamma \subset F \times \{0\}$ (or $\Gamma \subset F \times \{1\}$), is called a *spine for F on $\partial\mathcal{N}(F)$*. Also, we say that Γ is a *spine for F on $\partial E(F)$*.

If Γ is a spine for F on $\partial E(F)$, and G is a regular neighborhood of Γ in $\partial E(F)$, then a *handle decomposition for the pair* $(E(F), \Gamma)$ is, by definition, a handle decomposition for the pair $(E(F), G)$.

Let $\Gamma = a_1 \vee \cdots \vee a_n$ be a spine for F on $\partial E(F)$, and let $t(a_i)$ be a Dehn twist on F along the curve a_i. If $\widetilde{\Gamma}$ is the graph obtained from Γ by replacing the curve a_j by the curve $t(a_i)(a_j)$, then $\widetilde{\Gamma}$ is also a spine for F. The graph $\widetilde{\Gamma}$ is called *the spine for F obtained from* Γ *by sliding a_j along $a_i^{\pm 1}$* $(i, j \in \{1, \dots, n\})$.

Remark 2.5. Notice that if $\widetilde{\Gamma}$ is another spine for F on $\partial E(F)$, and $\widetilde{G}$ is a regular neighborhood of $\widetilde{\Gamma}$ in $\partial E(F)$, then the pairs $(E(F),\Gamma)$ and $(E(F),\widetilde{\Gamma})$ usually are not homeomorphic, but the pairs $(E(F),F)$ and $(E(F),\widetilde{G})$ are homeomorphic. Thus, *to find circular decompositions of $E(k)$ based on F, we need only to construct moderate handle decompositions of the homeomorphism class of a pair $(E(F),\Gamma)$ for some spine Γ for F on $\partial E(F)$.*

Remark 2.6. Let $F \subset S^3$ be a connected orientable surface with boundary $k = \partial F$. If a spine Γ for F on $\partial\mathcal{N}(F)$ is also a spine for $E(F)$, then k is a fibered knot with fiber F. Indeed, $E(F)$ is a handlebody (for it is an irreducible 3-manifold with connected boundary and with free fundamental group), and both $\mathcal{N}(F)$ and $E(F)$ admit a product structure of the form $G \times I$, where G is a regular neighborhood of Γ in $\partial\mathcal{N}(F) = \partial E(F)$.

2E. *Whitehead diagrams.* Let H be a genus-g handlebody, and let $x_1,\dots,x_g$ be a system of meridional disks for H. The exterior $E(x_1\cup\cdots\cup x_g)$ is a 3-ball with $2g$ fat vertices $x_1,\bar{x}_1,\dots,x_g,\bar{x}_g$ on its boundary, where $x_i = x_i\times\{0\}$ and $\bar{x}_i = x_i\times\{1\}$ are the copies of x_i in the product structure $\mathcal{N}(x_i) = x_i\times I \subset H$, $i = 1,\dots,g$.

There is a one-to-one correspondence between isotopy classes of systems of meridional disks $\{x_1,\dots,x_g\}$ for H and homotopy classes of spines of the form $a_1\vee\cdots\vee a_g \subset H$ such that $\#(a_i\cap x_i) = 1$ and $a_i\cap x_j = \varnothing$ for $i\neq j$, $i,j = 1,\dots,g$. It is convenient to commit an abuse of notation and write both $\{x_1,\dots,x_g\}$ for a meridional system of disks for H, and $\{x_1,\dots,x_g\}$ for the corresponding basis of $\pi_1(H)$ represented by the curves $a_1,\dots,a_g$ in the one-to-one correspondence above. Throughout this paper we adhere to this abuse of notation.

A graph $\Gamma = a_1\vee\cdots\vee a_n \subset \partial H$ intersects $E(x_1\cup\cdots\cup x_g)$ in a set of subarcs of the curves a_i; some of these arcs intersect in the base point of Γ. These arcs, together with $x_1,\bar{x}_1,\dots,x_g,\bar{x}_g$, form a graph G with $2g$ fat vertices immersed in $\partial E(x_1\cup\cdots\cup x_g)$. The base point of Γ appears in the drawing on $\partial E(x_1\cup\cdots\cup x_g)$ as the intersection of some edges of G, but the base point of Γ is not considered a vertex of G. We require that the graph G has no loops, that is, that there are no edges with ends in the same fat vertex of G. In our examples, we will be able to realize this assumption — no loops in G — through the use of some isotopies of H. For each i we number the ends of the arcs in x_i and $\bar{x}_i$ in such a way that the gluing homeomorphisms, which recover H from $E(x_1\cup\cdots\cup x_g)$, identify equally numbered points. The immersion of the graph G in $\partial E(x_1\cup\cdots\cup x_g)$, together with these numberings, is called *the Whitehead diagram of the pair* (H,Γ) associated to the system of meridional disks $x_1,\dots,x_g \subset H$ (see Figure 1). The graph G is called the *Whitehead graph* of the corresponding Whitehead diagram.

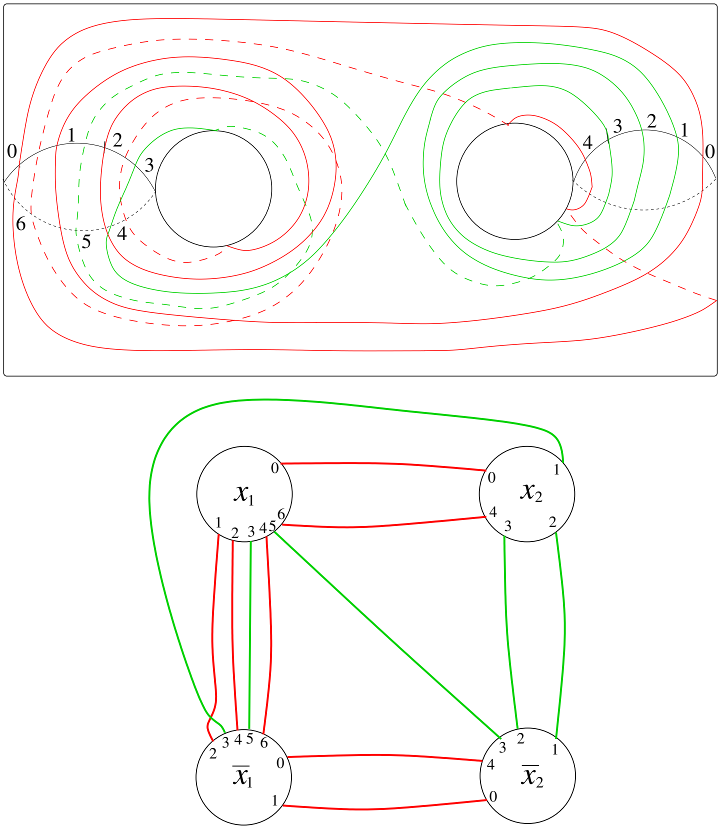


Figure 1. A Whitehead diagram associated to the exterior of the pretzel knot $p(5, 5, 5)$.

Let X be a graph and let e, f be two edges of X; we say that e and f are *parallel* if they connect the same pair of vertices of X. The *simple graph associated to* X is the graph obtained from X by replacing each parallelism class of edges of X by a single edge and deleting each loop in X (if any).

If X is a connected graph, a vertex v of X is called a *cut vertex* of X if $X - \{v\}$ is not connected. Notice that a loopless graph X contains a cut vertex if and only if the simple graph associated to X contains a cut vertex.

Let $\mathcal{F}$ be a free group with basis Y and let A be a set of cyclically reduced words on $Y \cup Y^{-1}$, regarded as elements of $\mathcal{F}$. The *genuine Whitehead graph of* A is the graph Γ with vertices $Y \cup Y^{-1}$, and for each $\alpha \in A$ and $v_1, v_2 \in Y \cup Y^{-1}$, an edge from v_1 to v_2^{-1} if α contains the word of length two $v_1 v_2$, up to a cyclic shift of α. If α is of length 1, $\alpha = v$, then there is an edge from v to v^{-1}. If A is a set of elements of $\mathcal{F}$, we can replace A with a set A' of cyclically reduced words representing the conjugacy classes of the elements of A, and then the *genuine*

Whitehead graph of A is, by definition, the genuine Whitehead graph of A'. The genuine Whitehead graph of a set of elements of $\mathcal{F}$ is regarded as being embedded in 3-space and also contains no loops.

Let $\mathcal{F}$ be a free group and let A be a set of elements of $\mathcal{F}$. Then A is called *separable* if there exists a nontrivial splitting $\mathcal{F} \cong \mathcal{F}_1 * \mathcal{F}_2$ such that each $\alpha \in A$ represents, up to conjugacy, an element of $\mathcal{F}_j$ for some j.

Theorem 2.7 [Stallings 1999, Theorem 2.4]. *Let A be a set of elements of a free group $\mathcal{F}$ with genuine Whitehead graph Γ. If Γ is connected and if A is separable in $\mathcal{F}$, then there is a cut vertex in Γ.*

The next result follows from Theorem 2.7 and is included here for future reference.

Corollary 2.8. *Let $\Gamma = a_1 \vee \cdots \vee a_n$ be a wedge of n simple closed curves embedded in the boundary of a handlebody H. Assume that the Whitehead graph for some Whitehead diagram of the pair (H, Γ) is connected and has no cut vertex. Then Γ intersects every essential disk of H.*

Proof. Let G be the Whitehead graph of the pair (H, Γ) with respect to some system of meridional disks $\{x_1, \dots, x_g\}$ such that G has no cut vertex and is connected. In particular, G has no loops. If we regard G as a graph G' embedded in 3-space so that the base point of Γ vanishes, then G' is the genuine Whitehead graph of the set of elements of $\pi_1(H)$ represented by $\{a_1, \dots, a_n\}$ with respect to the basis $\{x_1, \dots, x_g\}$. Since G is connected and has no cut vertex, it follows that G' is also connected and has no cut vertex (recall that the base point of Γ is not part of G; then G and G' are isomorphic graphs). If there is an essential disk in H disjoint with Γ, then the set of elements of $\pi_1(H)$ represented by $\{a_1, \dots, a_n\}$ clearly is separable, and by Theorem 2.7, G' has a cut vertex or is disconnected. Since G' is connected and has no cut vertex, it follows that Γ intersects every essential disk of H. $\square$

2F. *Handle slides.* Handle slides in a handlebody are conveniently visualized when translated into a Whitehead diagram. Figure 2 shows the effect of sliding the handle corresponding to the disk x_2 along the handle corresponding to x_1. But, of course, in the final step, the meridional disks $x_1, \bar{x}_1, x_2, \bar{x}_2$ in the drawing are no longer the same disks, but are their images after the handle slide in the handlebody (The effect of such a handle slide in the fundamental group of the handlebody is a *Whitehead automorphism*. See [Stallings 1999]).

2F1. *∂-parallel arcs in handlebodies.* Let $k \subset S^3$ be a knot, and let $F \subset E(k)$ be a free Seifert surface for k. Also let Γ be a spine for F on $\partial E(F)$. In Remark 2.2(2) a program is outlined to construct a circular decomposition for $E(k)$. It starts by drilling some 2-handles out of $E(F)$ disjoint from F. A 2-handle $P \subset E(F)$ is a product $P = D^2 \times I$ such that $(D^2 \times I) \cap \partial E(F) = D^2 \times \{0, 1\}$, and it is determined

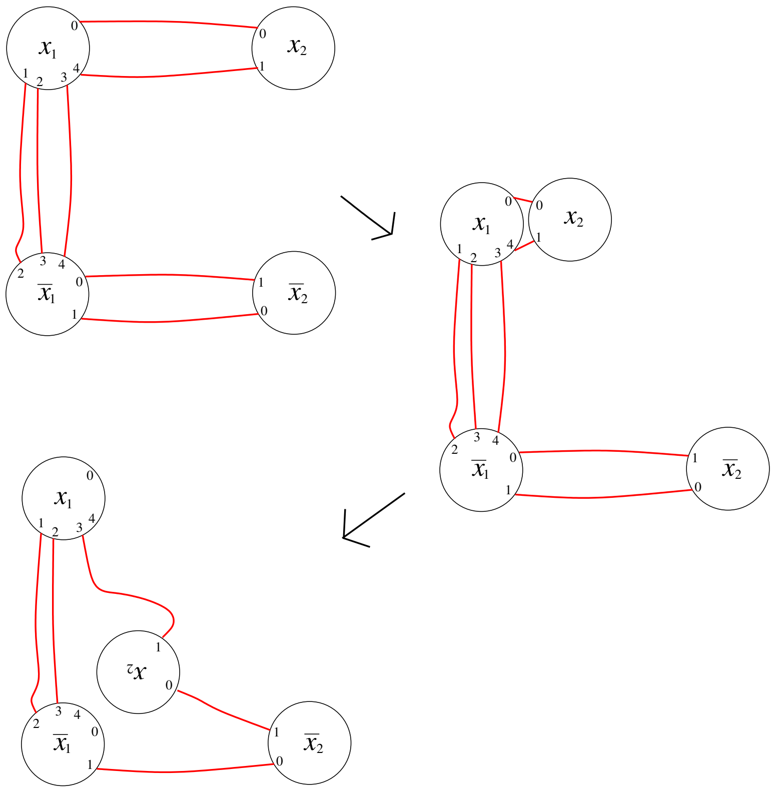


Figure 2. A handle slide.

by its "cocore" $\gamma = \{0\} \times I$. This cocore, γ, can be visualized in $E(F)$ as a properly embedded arc with ends disjoint from Γ.

Given two properly embedded arcs γ and γ' in $E(F)$ disjoint from Γ, if the triples $(E(F), \Gamma, \gamma)$ and $(E(F), \Gamma, \gamma')$ are homeomorphic, then the pairs $(E(\gamma), \Gamma)$ and $(E(\gamma'), \Gamma)$ are homeomorphic, and, therefore, have homeomorphic handle decompositions. In this sense, we say that γ and γ' *induce homeomorphic handle decompositions* of $(E(F), \Gamma)$. Also we say, as an abuse of language, that γ and γ' are *equivalent 2-handles*.

Let k be a knot with $h(k) = 1$ and let $F \subset E(k)$ be a free Seifert surface for k which realizes a one-handled circular decomposition of $E(F)$. Let $\gamma \subset E(F)$ be a properly embedded arc disjoint from $F \times \{0\}$. If the arc γ is the cocore of the single 2-handle of the one-handled circular decomposition of $E(F)$, then γ is called *the arc of the handle decomposition*. Note that in this case we know that γ is parallel into $\partial E(F)$ (see Corollary 4.3 below).

2F2. *Criterion for one-handledness.* We will establish a criterion to determine if an arc is the arc of some one-handled decomposition.

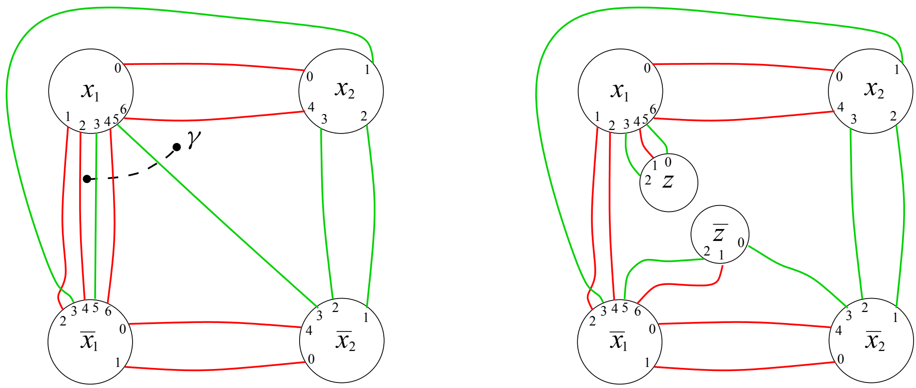


Figure 3. Drilling out a 2-handle.

Let k be a knot with $h(k) = 1$ and let $F \subset E(k)$ be a free Seifert surface for k which realizes a one-handled circular decomposition of $E(F)$. Let $\gamma \subset E(F)$ be a ∂-parallel properly embedded arc disjoint from $F \times \{0\}$.

Consider a system of meridional disks $x_1, \ldots, x_g \subset E(F)$. Let z be a ∂-parallelism disk for γ. After an isotopy of $E(F)$ which keeps Γ fixed pointwise, we may assume that z is disjoint from the disks $x_1, \ldots, x_g$. Then γ can be visualized in the Whitehead diagram of $(E(F), \Gamma)$ with respect to $x_1, \ldots, x_g \subset E(F)$ as a properly embedded arc in $E(x_1 \cup \cdots \cup x_g)$ disjoint from G, where G is the corresponding Whitehead graph. After drilling out the 2-handle, which is a regular neighborhood of γ, we are "adding a new handle" to $E(F)$; that is, the exterior $E(\gamma) \subset E(F)$ is homeomorphic to $E(F)$ plus one 1-handle. We obtain a new Whitehead diagram for $(E(\gamma), \Gamma)$ with respect to $x_1, \ldots, x_g, z$, adding two fat vertices z and $\bar{z}$ as in Figure 3.

Define the complexity of a Whitehead graph as the sum of all valences of the fat vertices of the graph. The new Whitehead diagram obtained in the last paragraph may contain a cut vertex v. For example, $v = x_1$ in Figure 3. When there is a cut vertex v in G, this vertex decomposes the graph G into two nontrivial graphs X_1 and X_2. One of these graphs, say X_1, does not contain $\bar{v}$. Then we can slide the part corresponding to the graph X_1 along the handle defined by the disk v. If cut vertices appear after sliding, we continue sliding along some cut vertex on and on. See Figures 4 and 5. Since each such handle slide lowers the complexity of the graph, eventually we end up with either

(1) *A disconnected diagram*, or

(2) *A connected diagram with no cut vertices.*

In case (1) (see the last drawing of Figure 5) there are obvious essential disks in $E(\gamma)$ disjoint from Γ (more precisely, disjoint from the *image* of Γ on the

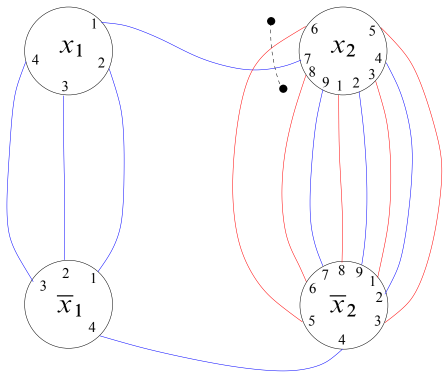


Figure 4

diagram after the slides); the boundaries of these essential disks are curves that separate the components of the current Whitehead graph. Assume a neighborhood of one of these disks is a 1-handle B inside $E(\gamma)$ such that, after drilling out B, $E(\gamma \cup B)$ is a regular neighborhood of $F = F \times \{0\}$. (See the last drawing in Figure 5, where the disk labeled x_1 corresponds to B.) Then we have found a circular one-handled decomposition of $E(k)$ based on F according to the program outlined in Remark 2.2(2), and γ is the arc of this handle decomposition. Otherwise, we have to restart the program, choosing a different arc to drill out.

In case (2), by Corollary 2.8, the chosen arc is not part of a one-handled circular decomposition. Again, we have to restart the program, choosing a different arc to drill out.

2F3. *Some definitions.* Now let γ and γ' be two ∂-parallel properly embedded arcs in $E(F)$ disjoint from Γ, with ∂-parallelism disks z and z', respectively; let $\{x_1, \ldots, x_g\}$ be a meridional system of disks for $E(F)$, and let G be the corresponding Whitehead graph with respect to this system of disks. Then, by an isotopy of $E(F)$, we may assume that z and z' are contained in $E(x_1 \cup \cdots \cup x_g)$ and (the images of) γ and γ' are disjoint from G.

Assume that for two faces of G (that is, two connected components $A, B \subset \partial E(x_1 \cup \cdots \cup x_g) - G$) the face A contains an endpoint of γ and one of γ', and the face B contains the other two endpoints of γ and γ'. Then there is an isotopy of $E(x_1 \cup \cdots \cup x_g)$ that fixes G pointwise and sends γ onto γ'. Such an isotopy exists because, since γ and γ' are ∂-parallel, they are unknotted properly embedded arcs in the 3-ball $E(x_1 \cup \cdots \cup x_g)$, and the isotopy can be chosen to fix G, for the endpoints of the arcs are, by pairs, in components of $\partial E(x_1 \cup \cdots \cup x_g) - G$. Then we see that a class of "equivalent" 2-handles in the Whitehead diagram of $(E(F), \Gamma)$ with respect to $x_1, \ldots, x_g$ is determined by pairs of faces of G in $\partial E(x_1 \cup \cdots \cup x_g)$ (and conversely). That is, for ∂-parallel properly embedded arcs $\gamma, \gamma' \subset E(x_1 \cup \cdots \cup x_g)$,

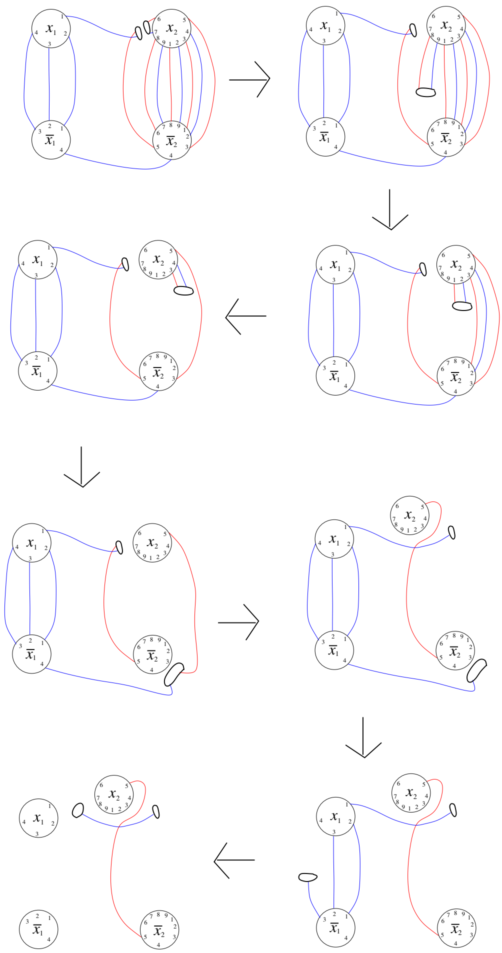


Figure 5

the triples $(E(x_1 \cup \cdots \cup x_g), G, \gamma)$ and $(E(x_1 \cup \cdots \cup x_g), G, \gamma')$ are homeomorphic if and only if γ and γ' connect the same pair of faces of G.

This is a very useful fact. To search for a one-handled decomposition, one must only test a finite number of ∂-parallel arcs in some Whitehead diagram, and analyze as above: there are as many ∂-parallel arcs to check as pairs of faces of the corresponding Whitehead graph.

We end this section with some definitions. Assume the arc γ is boundary-parallel into $\partial E(F)$. Let z be a ∂-parallelism disk for γ such that $\partial z = \gamma \cup \gamma_z^{\boldsymbol{B}}$, where $\gamma_z^{\boldsymbol{B}}$ is an arc in $\partial E(F)$. Then, after a small isotopy of z, if necessary, $\gamma_z^{\boldsymbol{B}}$ intersects the edges of Γ transversely in a finite number of points. If $e_1, \ldots, e_n$ are the edges of Γ that intersect $\gamma_z^{\boldsymbol{B}}$ and each e_i intersects only once with $\gamma_z^{\boldsymbol{B}}$, we say that

γ *encircles the edges* $e_1, \dots, e_n$. If γ encircles the edges $e_1, \dots, e_n$, and all e_i are incident in the vertex ξ of Γ, we say that the arc γ *is around the vertex* ξ. Notice that if $e_1, \dots, e_n, e_{n+1}, \dots, e_{n+m}$ are all the edges incident in the vertex ξ of Γ, and γ is around vertex ξ encircling the edges $e_1, \dots, e_n$, then γ also encircles the edges $e_{n+1}, \dots, e_{n+m}$. The *length* of γ in Γ is the minimal number of intersection points of γ_z^B and Γ among all ∂-parallelism disks z for γ.

3. Primitive elements in spines

Let $\mathcal{F}$ be a free group. An element $x \in \mathcal{F}$ is called *primitive* if x is part of some basis of $\mathcal{F}$. A set of elements $x_1, x_2, \dots, x_k \in \mathcal{F}$ are called *associated primitive elements* if they are contained in some basis of $\mathcal{F}$.

Let H be a genus-g handlebody. A simple closed curve $\alpha \subset H$ represents a primitive element in $\pi_1(H)$ if and only if there is an essential properly embedded disk $D \subset H$ such that $\alpha \cap D$ consists of a single point. A set of simple closed curves $\alpha_1, \dots, \alpha_k \subset H$ represents a set of associated primitive elements in $\pi_1(H)$ if and only if there is a system of meridional disks $D_1, D_2, \dots, D_g \subset H$ such that, up to renumbering, $\alpha_i \cap D_i$ consists of a single point, and $\alpha_i \cap D_j = \varnothing$ for $i \neq j$, $i = 1, \dots, k$, and $j = 1, \dots, g$.

Theorem 3.1. *Let $k \subset S^3$ be a knot and let $F \subset E(k)$ be a free Seifert surface for k. Assume $E(F)$ is a handlebody of genus g.*

If there exists a graph $\Gamma = a_1 \vee \cdots \vee a_g$ such that Γ is a spine for F on $\partial E(F)$, and the ℓ curves $a_1, \dots, a_\ell$ represent associated primitive elements of $\pi_1(E(F))$, then the handle number $h(F)$ is at most $g - \ell$.

Proof. We follow the plan in Remark 2.2(2): we will exhibit a system of properly embedded arcs (the arcs β_j^I below) which are the cocores of $(g-\ell)$ 2-handles to be drilled out of $E(F)$, and a system of $(g-\ell)$ 2-disks ($D_{\ell+1}, \dots, D_g$ below) which define the cocores of $(g-\ell)$ 1-handles to be drilled out of $E(F \cup \bigcup_j \beta_j^I)$

Let $D_1, D_2, \dots, D_g \subset E(F)$ be a system of meridional disks for $E(F)$ such that $|a_i \cap D_i| = 1$ and $a_i \cap D_j = \varnothing$ for $i \neq j$, $i = 1, \dots, \ell$, and $j = 1, \dots, g$. This system of meridional disks exists since $a_1, \dots, a_\ell$ represent associated primitive elements of $\pi_1(E(F))$.

Let $P \subset E(F)$ be a regular neighborhood of the base point $x_0 \in \partial E(F)$ (x_0 is also the base point of the graph Γ). We visualize P as a $2g$-gonal prism (see Figure 6). For $i = 1, \dots, g$, let T_i be a regular neighborhood of a_i in $E(F)$ such that $T_i \cap T_j = P$ if $i \neq j$. Write $\widehat{T}_i = \overline{T_i - P}$; then $\widehat{T}_i$ is a 3-ball. The intersection $\widehat{T}_i \cap P = d_i^+ \cup d_i^-$ is the disjoint union of two 2-disks d_i^+ and d_i^- (see Figure 6). Also, write $\partial d_i^+ = \beta_i^B \cup \beta_i^I$, where β_i^B is an arc in $\partial E(F)$ and β_i^I is a properly embedded arc in $E(F)$. Finally, write

$$A_i = \overline{\partial T_i - (d_i^+ \cup d_i^- \cup \partial E(F))},$$

which is a 2-disk.

The arcs $\beta_{\ell+1}^I, \ldots, \beta_g^I$ are the cocores of 2-handles in $E(F)$ to be drilled out, according to the plan in Remark 2.2(2).

Notice that the exterior $E(\beta_i^I)$ of each β_i^I satisfies

$$E(\beta_i^I) = \overline{E(F) - \mathcal{N}(\beta_i^I)} \cong \overline{E(F) - \mathcal{N}(A_i)},$$

and this homeomorphism is the identity map outside a small neighborhood of A_i.

Consider

$$V = \overline{E(F) - (\widehat{T}_{\ell+1} \cup \widehat{T}_{\ell+2} \cup \cdots \cup \widehat{T}_g)}.$$

Then V is a genus-g handlebody and $E(F)$ is a regular neighborhood of V. We see that

$$\overline{E(F) - \bigcup_{\ell+1}^{g} \mathcal{N}(\beta_i^I)} \cong \overline{E(F) - \bigcup_{\ell+1}^{g} \mathcal{N}(A_i)} \cong V \cup (g - \ell \text{ 1-handles}),$$

where the $(g-\ell)$ 1-handles are the $(g-\ell)$ balls $\widehat{T}_i$ attached along the disks d_i^+, d_i^-, $i = \ell+1, \ldots, g$.

By the choice of the disks $\{D_i\}$, we see that $\overline{V - \bigcup_{\ell+1}^{g} \mathcal{N}(D_i \cap V)}$ is a regular neighborhood of $a_1 \vee \cdots \vee a_\ell$. Then

$$\overline{E(F) - \Big(\bigcup_{\ell+1}^{g} \mathcal{N}(\beta_i^I) + \bigcup_{\ell+1}^{g} \mathcal{N}(D_i \cap V)\Big)}$$

is a regular neighborhood of Γ. In other words,

$$\mathcal{N}(F) \cup \{\mathcal{N}(\beta_i^I) : i = \ell+1, \ldots, g\} \cup \{\mathcal{N}(D_i \cap V) : i = \ell+1, \ldots, g\}$$

determines a circular handle decomposition of $E(k)$ based on F, as in Remark 2.2(2). Therefore $h(F) \le g - \ell$. □

Corollary 3.2 (the case $\ell = g$). *Let $k \subset S^3$ be a knot and let F be a free Seifert surface for k. Assume that $E(F)$ is a handlebody of genus g.*

If there exists a graph $\Gamma = a_1 \vee a_2 \vee \cdots \vee a_g$ such that Γ is a spine for F on $\partial E(F)$, and the curves $a_1, \ldots, a_g$ form a basis of $\pi_1(E(F))$, then k is a fibered knot with fiber F.

Proof. In this case $h(F) = 0$, therefore $E(F)$ admits a product structure $E(F) = F \times I$ induced by Γ, and k is fibered with fiber F. □

Corollary 3.3 (the case $\ell = 0$). *Let $k \subset S^3$ be a knot and let $F \subset E(k)$ be a free Seifert surface for k. Assume that $E(F)$ is a handlebody of genus g.*

Then $h(k) \le g$.

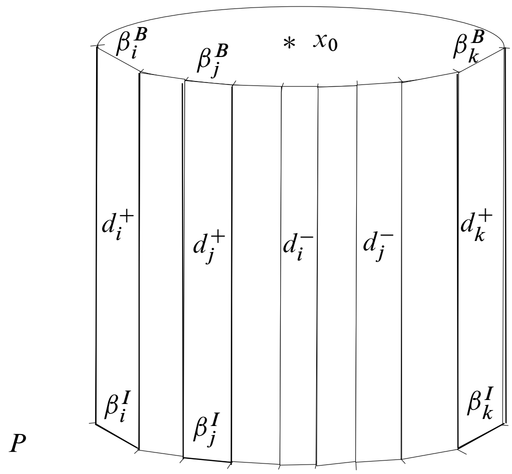


Figure 6. The neighborhood of x_0.

Proof. By Theorem 3.1, since $\ell = 0$, we have $h(F) \leq g$. Therefore $h(k) \leq g$. □

Remark 3.4. Corollary 3.3 asserts that $h(k) \leq 2g_f(k)$ for a connected knot k. See [Hirasawa and Rudolph 2003] for another proof of this fact (therein called the "free genus estimate").

Corollary 3.5. *If k is a connected free genus one knot, then $h(k) = 0, 1,$ or 2.* □

Remark 3.6. Let k be a connected free genus one knot in S^3 such that k is not fibered (that is, $k \neq 3_1, 4_1$). At this point we can give some estimates for $\mathrm{cw}(k)$.

If k is almost fibered, it follows from Corollary 3.5 that $\mathrm{cw}(k) = 4$ or $\mathrm{cw}(k) = 6$. In any case, that is, if k is almost fibered or not, $\mathrm{cw}(k) \leq 6$.

If k is not almost fibered, consider a circular decomposition

$$E(k) = (F \times I) \cup B_1 \cup P_1 \cup B_2 \cup P_2 \cup \cdots \cup B_n \cup P_n,$$

with $n > 1$ and $B_i, P_i \neq \varnothing$, which realizes $\mathrm{cw}(k)$. Then there are Seifert surfaces

$$T_1, \ldots, T_n = F, \quad S_1, \ldots, S_n \subset E(k)$$

for k such that S_i is obtained from T_{i-1} by adding the 1-handles B_i, T_i is obtained from S_i by adding the 2-handles P_i, and $\mathrm{cw}(k) = (c(S_{\sigma(1)}), \ldots, c(S_{\sigma(n)}))$ with $c(S_{\sigma(1)}) \geq \cdots \geq c(S_{\sigma(n)})$, where $c(S) = 1 - \chi(S)$.

Now, all T_i are incompressible (Remark 2.3), and of genus one, for if some T_j is of genus at least two, then S_j is of genus at least three, and the complexity $c(S_j)$ is at least 6. But then, since $n > 1$, $\mathrm{cw}(k) = (c(S_{\sigma(1)}), \ldots, c(S_{\sigma(n)})) > 6$, a contradiction. It follows that $\mathrm{cw}(k) = (4, \ldots, 4)$.

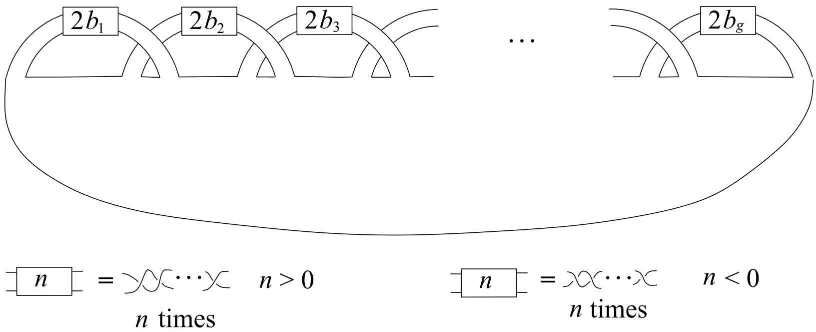


Figure 7. A minimal-genus Seifert surface for the knot $k = [2b_1, 2b_2, \ldots, 2b_n]$.

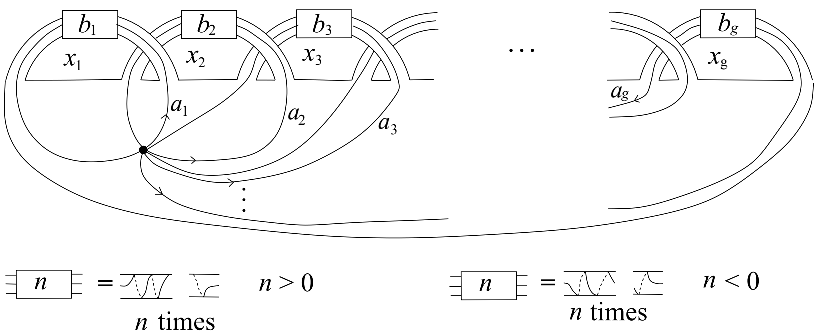


Figure 8. A spine for $k = [2b_1, 2b_2, \ldots, 2b_g]$ in $\partial\mathcal{N}(F)$.

That is, *if k is a connected nonfibered free genus one knot, then* $\mathrm{cw}(k) = 4, 6$*, or* $(4, \ldots, 4)$.

As was mentioned in the introduction, a connected nonfibered free genus one knot in S^3 is almost fibered (Theorem 6.7 below). It follows that $\mathrm{cw}(k) \in \{4, 6\}$.

Example 3.7 (rational knots). *If $k \subset S^3$ is a nonfibered rational knot, then $h(k) = 1$. Also* $\mathrm{cw}(k) = 4g(k)$ *if k is connected, and* $\mathrm{cw}(k) = 4g(k) + 1$ *otherwise.*

Let $k \subset S^3$ be a rational knot. Then k is encoded with a continued fraction of the form $[2b_1, 2b_2, \ldots, 2b_g]$ where g is even or odd if k is connected or not, respectively. Here $b_1, \ldots, b_g$ are nonzero integers. Now k has a minimal-genus Seifert surface F as in Figure 7 (see [Gabai 1986, Answer 1.19]). This surface is free. Note that $g(F) = g/2$ if k is connected, and $g(F) = (g-1)/2$ otherwise.

In a neighborhood V of this surface we can find a spine $\Gamma \subset F \times \{0\} \subset \partial V$ with $\Gamma = a_1 \vee a_2 \vee \cdots \vee a_g$, as in Figure 8. For the obvious meridional disks $x_1, x_2, \ldots, x_g$, of the handlebody $E(F)$ corresponding to a basis $\{x_1, x_2, \ldots, x_g\}$

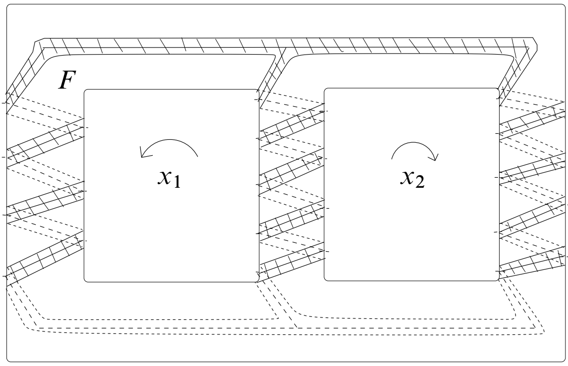


Figure 9. Black surface for $P(7,9,9)$.

of $\pi_1(E(F))$, the curves $a_1, a_2, \ldots, a_g$ represent the elements

$$x_1^{b_1}, x_2^{b_2}x_1, x_3^{b_3}x_2, \ldots, x_{g-1}^{b_{g-1}}x_{g-2}, x_g^{b_g}x_{g-1}$$

of $\pi_1(E(F))$, respectively.

If each $|b_i| = 1$, then $a_1, a_2, \ldots, a_g$ represent a basis of $\pi_1(E(F))$, and, by Corollary 3.2, k is fibered with fiber F.

If some $|b_i| \geq 2$, then $\{x_g, x_2^{b_2}x_1, x_3^{b_3}x_2 \ldots, x_{g-1}^{b_{g-1}}x_{g-2}, x_g^{b_g}x_{g-1}\}$ is a basis for $\pi_1(E(F))$; it follows that the curves $a_2, a_3, \ldots, a_g \subset \Gamma$ represent associated primitive elements of $\pi_1(E(F))$, and, by Theorem 3.1, $h(k) \leq h(F) = 1$. By the second part of the statement of Answer 1.19 of [Gabai 1986], k is not fibered. Therefore, $0 < h(k) = h(F) = 1$, and $\mathrm{cw}(k) = 2g$ if k is connected, and $\mathrm{cw}(k) = 2g + 1$ otherwise.

Remark 3.8. In Theorem 3.21 of [Goda 2006] it is claimed that the result in Example 3.7, the one-handledness of rational knots, is known, but unpublished.

Example 3.9 (pretzel knots). *The pretzel knot $k = P(\pm 3, q, r)$ with $|q|, |r|$ odd integers ≥ 3, has $h(k) = 1$ and, therefore,* $\mathrm{cw}(k) = 4$.

Let k be the pretzel knot $P(p, q, r)$ with p, q, r odd integers. Then k is a connected knot, and the "black surface" F of a standard projection of k is a free genus one Seifert surface for k. See Figure 9. If $|p|, |q|, |r| \geq 3$, the following facts are known:

(1) k has a unique incompressible Seifert surface (see [Goda and Ishiwata 2006]), namely, the free black surface F of genus one;

(2) k has tunnel number two (see [Klimenko and Sakuma 1998]);

(3) $h(k) \leq 2$ (see Corollary 3.5);

(4) since $t(k) \neq 1$, k is not a rational knot;

(5) k is not fibered (that is, $k \neq 3_1, 4_1$).

For any permutation s, t, u of p, q, r, the pair (S^3, k) is homeomorphic to the pair (S^3, ℓ), where ℓ is the pretzel knot $P(s, t, u)$. Also, by a reflection, $P(p, q, r)$ is equivalent to $P(-p, -q, -r)$. Then, by Remark 2.1, we may assume that either $p, q, r > 0$ (case 1) or $p < 0$ and $q, r > 0$ (case 2).

There is a spine shown in Figure 9 for the surface $F \times \{0\} \subset \partial\mathcal{N}(F)$. This spine is a θ-graph. To obtain a wedge of circles as a spine $\Gamma = a_1 \vee a_2 \subset F \times \{0\} \subset \partial\mathcal{N}(F)$, we slide the middle edge of the θ-graph to the left. Now we examine the two cases separately.

Case 1: $(p, q, r > 0)$ After sliding the middle edge of the graph to the left, we obtain the left part of Figure 10. Writing $\pi_1(E(F)) \cong \langle x_1, x_2 : -\rangle$, we see that the curves a_1 and a_2 represent the elements $x_2^{(r+1)/2} x_1^{-(p-1)/2}$ and $x_1^{(p+1)/2}(x_2x_1)^{(q-1)/2}$, respectively, in $\pi_1(E(F))$.

Now assume that $3 \in \{|p|, q, r\}$. Using a homeomorphism of S^3, we may assume that $p = 3$. In this case, the curve $a_1 \simeq x_2^{(r+1)/2} x_1^{-1}$ represents a primitive element of $\pi_1(E(F))$, for the set $\{x_2^{(r+1)/2} x_1^{-1}, x_2\}$ is a basis of $\pi_1(E(F))$. Therefore, by Theorem 3.1, $h(k) = h(F) = 1$, and $\operatorname{cw}(k) = 4$.

Case 2: $(p < 0$ and $q, r > 0)$ After sliding the middle edge of the graph to the left and using an isotopy to avoid unnecessary intersections of the curve a_2 with the disk x_1, we obtain the right part of Figure 10. Writing $\pi_1(E(F)) \cong \langle x_1, x_2 : -\rangle$, we see that the curves a_1 and a_2 represent the elements $x_2^{(r+1)/2} x_1^{(|p|+1)/2}$ and $x_1^{-(|p|-3)/2}(x_2x_1)^{(q-3)/2}x_2$, respectively, in $\pi_1(E(F))$.

Now assume that $3 \in \{|p|, q, r\}$. If $p = -3$, then the curve $a_2 \simeq (x_2x_1)^{(q-3)/2}x_2$ represents a primitive element of $\pi_1(E(F))$, for the set $\{(x_2x_1)^{(q-3)/2}x_2, x_2x_1\}$ is a basis of $\pi_1(E(F))$. If $q = 3$ or $r = 3$, we may assume that $q = 3$, and then the curve $a_2 \simeq x_1^{(|p|-3)/2}x_2$ represents a primitive element of $\pi_1(E(F))$, for the set $\{x_1^{-(|p|-3)/2}x_2, x_1\}$ is a basis of $\pi_1(E(F))$.

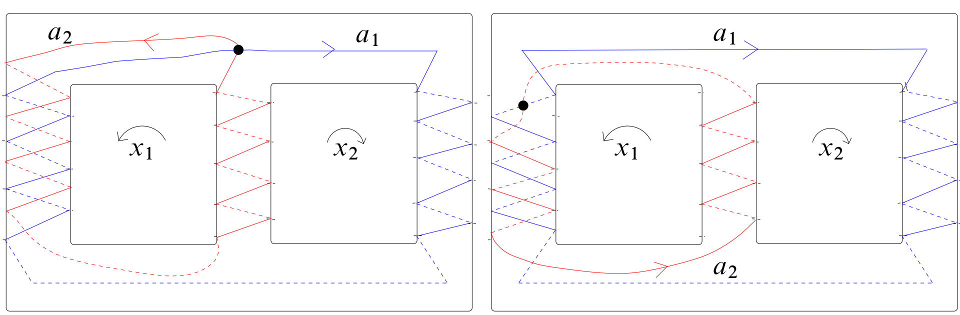


Figure 10. Spines for $P(p, q, r)$.

In both cases ($p = -3$, or q or $r = 3$) we conclude by Theorem 3.1 that $h(k) = h(F) = 1$, and $\mathrm{cw}(k) = 4$.

Remark 3.10. If $|q|, |r|$ are odd integers ≥ 3, then $k = P(\pm 3, q, r)$ has tunnel number two. Then the family of pretzel knots $\{P(\pm 3, q, r) : |q|, |r| \text{ odd integers } \geq 3\}$ is a family of examples of nonfibered knots k for which the strict inequality $h(k) < t(k)$ holds (compare with [Pajitnov 2010], where it is proved that $h(k) \leq t(k)$).

4. Pretzel knots: the case $|p|, |q|, |r| \geq 5$

In this section we show:

Theorem 4.1. *The free genus one Seifert surface for a pretzel knot $P(p, q, r)$ with $|p|, |q|, |r| \geq 5$ has handle number two.*

As noted in Example 3.9, when dealing with the pretzel knot $k = P(p, q, r)$ we may assume that either $p, q, r > 0$ (case 1) or $p < 0$ and $q, r > 0$ (case 2).

4A. *Handle decompositions of $E(P(p, q, r))$.*

Lemma 4.2. *Let V be a handlebody and let $\alpha \subset V$ be a properly embedded arc. If the exterior $E(\alpha) \subset V$ is a handlebody, then α is parallel into ∂V.*

Proof. By hypothesis, $\pi_1(E(\alpha))$ is a finitely generated free group. If $\mathcal{N}(\alpha) = D^2 \times I$ is a regular neighborhood of α in V, let $\mu = \partial D^2 \times \{\frac{1}{2}\}$ be a meridian of $\mathcal{N}(\alpha)$. If $N\langle\mu\rangle$ denotes the normal closure of the element represented by μ in $\pi_1(E(\alpha))$, then $\pi_1(E(\alpha))/N\langle\mu\rangle$ is isomorphic to the fundamental group of the space obtained from $E(\alpha)$ by adding a 2-handle along μ. Then $\pi_1(E(\alpha))/N\langle\mu\rangle \cong \pi_1(V)$ is a free group. It follows that μ represents a primitive element in $\pi_1(E(\alpha))$ (see [Whitehead 1936, Theorem 4]). Thus, there is an essential disk $\delta \subset E(\alpha)$ such that the number of points $\#(\delta \cap \mu)$ is equal to 1. After an isotopy, we may assume that $\partial\delta \cap \partial N(\alpha) = \gamma$ is an arc and $\partial\delta = \beta \cup \gamma$, where β is an arc contained in ∂V.

There is a product 2-disk $Z = (\text{radius of } D^2) \times I$ between γ and α, with $Z \subset \mathcal{N}(\alpha)$ for some product structure $D^2 \times I$ of $\mathcal{N}(\alpha)$. Then δ can be extended to a disk $\delta' = Z \cup \delta$ whose boundary is a union of arcs $\alpha \cup \beta'$ with $\beta' \subset \partial V$ (and $\beta \subset \beta'$). Therefore α is parallel into ∂V. $\square$

Corollary 4.3. *Let F be a free Seifert surface for a knot k. Suppose F has handle number one, and let α be the core of the 1-handle of a one-handled circular decomposition of $E(k)$ based on F. Then α is parallel into $\partial E(F)$.*

Proof. As in Remark 2.2(2), the one-handled decomposition of the pair $(E(F), F)$ is constructed by first drilling a 2-handle out of $E(F)$ disjoint from, say, $F \times \{1\}$. This 2-handle has as cocore the arc α of the statement (see Section 2F1). Secondly, after drilling out α, we drill one 1-handle B out of the exterior $E(\alpha) \subset E(F)$, with B disjoint from $F \times \{1\}$. The result of this drilling is a regular neighborhood of the surface $F \times \{0\}$ in $E(k)$, which is a handlebody. Therefore, the exterior $E(\alpha)$ in $E(F)$ is the union of the neighborhood of $F \times \{0\}$ and the 1-handle B; that is, $E(\alpha)$ is a handlebody. By Lemma 4.2 we conclude that α is parallel into $\partial E(F)$. $\square$

Proof of Theorem 4.1. Let F be the free genus one Seifert surface for $k = P(p, q, r)$ with $|p|, |q|, |r|$ odd integers ≥ 5.

For the sake of contradiction, we assume that F has handle number one. By Corollary 4.3, the core γ of the 1-handle of the circular decomposition of $E(k)$ based on F is parallel into $\partial E(F)$. By assumption, there is also a 2-handle $B \cong I \times D^2$ that completes the decomposition, such that the exterior $E(\gamma \cup B) \subset E(\gamma)$ is a regular neighborhood of F in $E(k)$, and ∂B is disjoint from F. In particular, the core $\{\frac{1}{2}\} \times D^2$ of B is an essential disk in $E(\gamma)$ disjoint from F. We will show

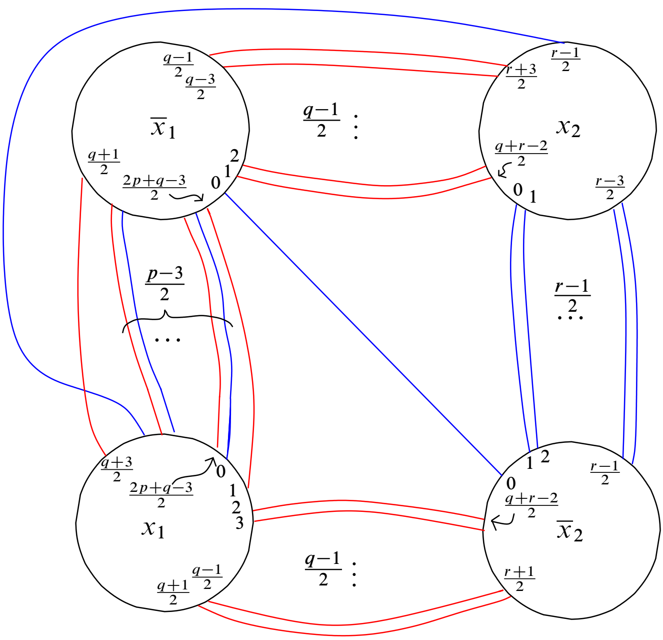


Figure 11

that any essential disk in $E(\gamma)$ intersects F, obtaining the desired contradiction.

Case 1: $(p, q, r > 0)$ Let $\Gamma = a_1 \vee a_2$ be the spine for F given in Example 3.9. By Remark 2.5, we only need to analyze the handle decompositions of $(E(F), \Gamma)$. There is an obvious system of meridional disks $x_1, x_2 \subset E(F)$ as depicted in the upper part of Figure 10. The Whitehead diagram for $(E(F), \Gamma)$ with respect to x_1, x_2 looks like Figure 11.

In the corresponding Whitehead graph G, we see:

- Four fat vertices corresponding to the meridional disks x_1 and x_2.

- There are $(q-1)/2$ horizontal edges connecting $\bar{x}_1$ and x_2, and $(q-1)/2$ horizontal edges connecting x_1 and $\bar{x}_2$; all these horizontal arcs belong to the curve a_2.
- There are $(r-1)/2$ vertical edges connecting x_2 and $\bar{x}_2$, one diagonal edge connecting x_1 and x_2, and one diagonal edge connecting $\bar{x}_1$ and $\bar{x}_2$; all these vertical and diagonal edges belong to the curve a_1.
- Finally, connecting x_1 with $\bar{x}_1$, we find, going from right to left in Figure 11, first an arc belonging to a_2, and then $(p-3)/2$ pairs of arcs belonging consecutively to a_1 and a_2, and a last arc belonging to a_2 which crosses with the diagonal arc from x_1 to x_2 on the base point of Γ.

Claim 0. *Let z be a ∂-parallelism disk for the arc γ in $E(F)$. Then the disk z contains at least one point of a_1 and one point of a_2.*

Proof of Claim 0. For $i=1,2$, let G_i be the Whitehead graph of the pair $(E(F), a_i)$ with respect to x_1, x_2 (see Figure 12). After sliding the handle defined by the disk x_2 along the handle defined by $\bar{x}_1$ on the right side of Figure 12, the image of the graph G_2 looks like Figure 13. Since these graphs are connected and contain no cut vertex, it follows from Corollary 2.8 that any essential disk in $E(F)$ intersects a_i $(i=1,2)$. Now, the exterior $E(\gamma)$ can be regarded as a copy of $E(F)$ plus one 1-handle defined by the disk z. Assume $z \cap a_2 \neq \varnothing$. If $z \cap a_1 = \varnothing$, then a_1 is contained in the copy of $E(F) \subset E(\gamma)$. By hypothesis, there is an essential disk $\Delta \subset E(\gamma)$ such that $\Delta \cap (a_1 \cup a_2) = \varnothing$. Now, $\Delta \cap z \neq \varnothing$, otherwise Δ is a subset of the copy of $E(F) \subset E(\gamma)$ missing the extra 1-handle, and $\Delta \cap a_1 = \varnothing$, contradicting that any essential disk in $E(F)$ intersects a_1. Through isotopies, we may assume that $\Delta \cap z$ is a set of disjoint arcs. Then the intersection of Δ with the copy of $E(F) \subset E(\gamma)$, that is, the set $\Delta \cap \overline{(E(\gamma) - \mathcal{N}(z))}$, is a set of disjoint properly embedded disks $\Delta_1, \ldots, \Delta_n \subset E(F)$. Since Δ is not parallel to z in $E(\gamma)$, at least one Δ_i is essential in $E(F)$, otherwise Δ would be parallel into $\partial E(\gamma)$. We obtain again an essential disk in $E(F)$ disjoint from a_1, which is a contradiction, as above, and therefore $z \cap a_1 \neq \varnothing$. □

The arc γ, being ∂-parallel in $E(F)$ by Corollary 4.3, can be isotoped into this Whitehead diagram as a properly embedded arc with ends disjoint from G (that is, after an isotopy of $E(F)$, we may assume that γ is disjoint from the system of disks x_1 and x_2). Recall that we are assuming that γ is the core of a 1-handle of a one-handled circular decomposition of $E(k)$ based on F. Therefore, after drilling out γ, there is an essential disk in $E(\gamma)$ disjoint from Γ; that is, after drilling out γ and obtaining a new Whitehead diagram with six fat vertices with Whitehead graph G', there is a sequence of handle slides of $E(\gamma)$ that disconnect the graph G', giving an essential disk in $E(F)$ disjoint from Γ (see Section 2F).

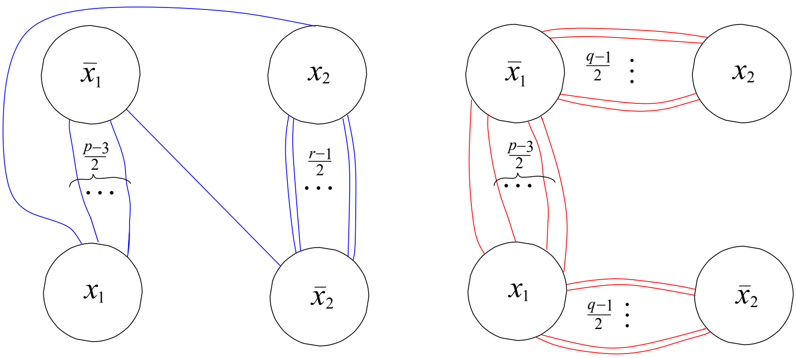


Figure 12. The graphs of curves a_1 and a_2.

Let G_i be the Whitehead graph of the pair $(E(F), a_i)$ with respect to x_1, x_2 (see Figure 12). After drilling out the arc γ from the diagram of G_i, we obtain a new Whitehead diagram for $(E(\gamma), a_i)$ with six fat vertices, corresponding to x_1, x_2, and z, and with Whitehead graph G_i'. Performing the handle slides of $E(\gamma)$ as above, the image of the graph G_i' will be also disconnected, giving an essential disk in $E(\gamma)$ disjoint from a_i $(i = 1, 2)$.

Notice that, if we drill out an arc of length one in G_i and perform handle slides, the image of G_i is disconnected (it contains four isolated fat vertices, $i = 1, 2$). We deal with this kind of arc after Claims 1 and 2.

Claim 1. *Let α be a properly embedded arc in $(E(F), a_2)$, disjoint from a_2, such that α is parallel into $\partial E(F)$ and α has length at least two in G_2. Then any essential disk in $E(\alpha)$ intersects a_2.*

Proof of Claim 1. The arc α *minimally* encircles a number of edges of the graph G_2. For example, the arc that encircles the two diagonal edges in Figure 13 actually has length 0.

Now, after sliding the handle defined by the disk x_2 along the handle defined by $\bar{x}_1$ on the right side of Figure 12, the image of the graph G_2 looks like Figure 13. The fat vertices of this graph are also obtained from the images of the disks x_1 and x_2 after the slide. We still call the new graph and new disks G_2 and x_1, x_2, respectively. This graph has $(q-3)/2$ vertical edges connecting x_2 with $\bar{x}_2$, one diagonal edge connecting x_2 with $\bar{x}_1$, one diagonal edge connecting x_1 with $\bar{x}_2$, and there are $(p-1)/2$ vertical arcs connecting x_1 with $\bar{x}_1$.

Let z be a minimal ∂-parallelism disk for α in $E(F)$, and let G be the Whitehead graph of $(E(\alpha), a_2)$ with respect to x_1, x_2, and z, which is obtained from G_2 by cutting along z and adding two fat vertices z and $\bar{z}$.

We now treat two separate cases: (i) α has length two, and (ii) α has length at least three.

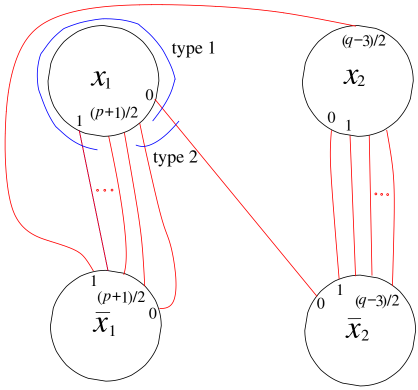


Figure 13

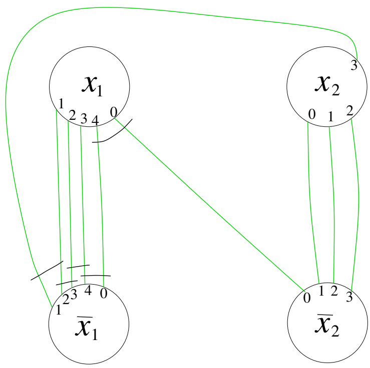


Figure 14

Case (i): (α has length two) Since $p \geq 5$, there are at least two vertical edges connecting x_1 and $\bar{x}_1$. Then there are two types of arcs of length two for the edges of G_2 around x_1 as in Figure 13, for any arc encircling two consecutive edges of G_2 connecting x_1 and $\bar{x}_1$ can be slid in $E(F)$ into an arc of type 1 or type 2. See Figure 14, where the arcs that can be slid in $E(F)$ into an arc of type 2 are shown.

After drilling out the arc α, if α is of type 1 or of type 2, the new Whitehead graph contains a cut vertex (see Figure 15).

After sliding handles, as in Section 2F, we end up with a graph G_2' with its simple associated graph a cycle of six vertices and six edges; that is, this simple graph contains no cut vertex. Therefore, G_2' contains no cut vertex, and by Corollary 2.8, a_2 intersects every essential disk of $E(\alpha)$.

If $q \geq 7$, there are at least two vertical edges connecting x_2 and $\bar{x}_2$. Then, by symmetry, the analysis of arcs of length two around x_2 and $\bar{x}_2$ is the same as for arcs of length two around x_1 and $\bar{x}_1$.

If $q = 5$, there is a single vertical edge connecting x_2 and $\bar{x}_2$, and then there are no arcs of length two around x_2 or $\bar{x}_2$.

For arcs not around a vertex of G_2, there are two more types of arcs of length two, as in Figure 16, but, after drilling out the arc α of type 3 or 4, the new Whitehead graph contains no cut vertex, and then, by Corollary 2.8, a_2 intersects every essential disk of $E(\alpha)$.

Case (ii): (α has length at least three) If α is an arc around x_i, we may assume that the length of α in G_2 is between 3 and degree$(x_i)/2$ (see last paragraph of Section 2F1), and α contains a subarc of type 1 or 2. After drilling out the arc α and sliding, if cut vertices appear, we end up with a graph with its simple associated graph a cycle with six vertices and six edges. Therefore, a_2 again intersects every essential disk of $E(\alpha)$.

If α is of length at least 3 and α contains a subarc of type 3 or 4 then, after drilling out the arc α, the new Whitehead graph contains no cut vertex, and by Corollary 2.8 we conclude that a_2 intersects every essential disk of $E(\alpha)$.

By the final remarks of Section 2F1, the arcs of types 1–4 exhaust all arcs to be considered as arcs of a one-handled decomposition for G_2. □

Claim 2. *Let α be a properly embedded arc in $(E(F), a_1)$, disjoint from a_1, such that α is parallel into $\partial E(F)$ and α has length at least two in G_1. Then any essential disk in $E(\alpha)$ intersects a_1.*

Proof of Claim 2. The Whitehead graph G_1 of $(E(F), a_1)$ has a shape as in Figure 13, but with $(r-1)/2$ vertical edges connecting x_2 with $\bar{x}_2$, one diagonal edge connecting x_2 with $\bar{x}_1$, one diagonal edge connecting x_1 with $\bar{x}_2$, and there are $(p-3)/2$ vertical arcs connecting x_1 with $\bar{x}_1$.

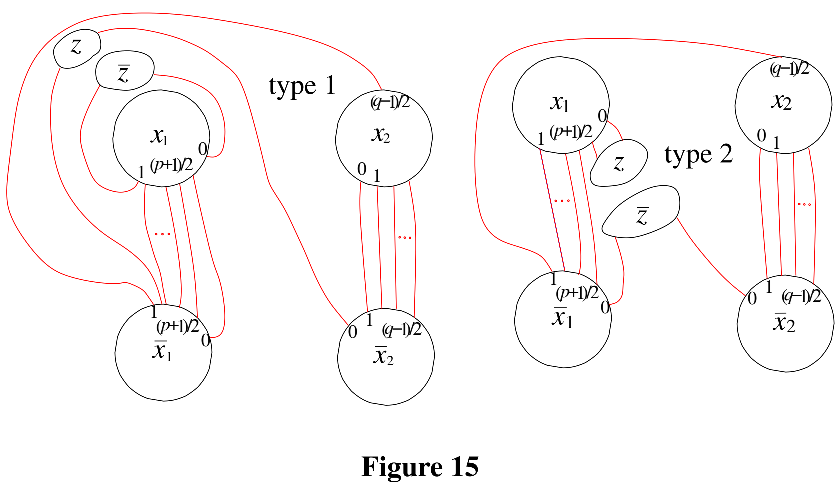


Figure 15

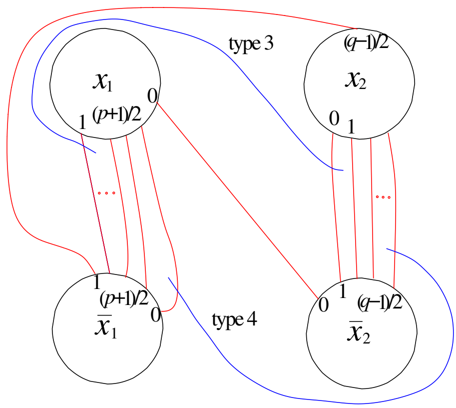


Figure 16

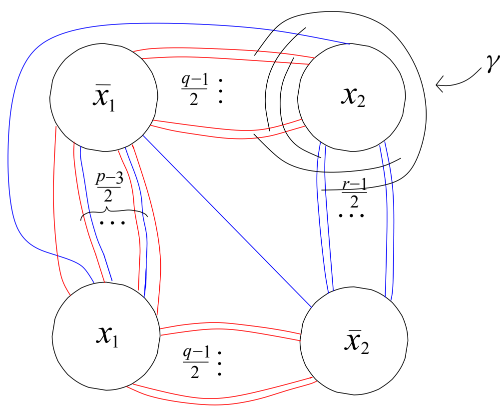


Figure 17

A similar (symmetric) analysis as in Claim 1 gives that a_1 intersects every essential disk of $E(\alpha)$. □

We are assuming that, after drilling out the arc γ, there is a set of handle slides of $E(\gamma)$ that disconnect the graph G', giving an essential disk in $E(F)$ disjoint from Γ.

By Claims 1 and 2, γ is of length one in G_1 and of length one in G_2. If γ is around one fat vertex ξ of G, it might happen that γ encircles exactly one edge of G_1, and all but one edge of G_2, or vice versa. In this case, γ is around either x_2 or $\bar{x}_2$. There are four arcs around x_2, and four arcs around $\bar{x}_2$ of this kind. The four arcs with this property around $\bar{x}_2$ can be slid in $E(F)$ and become equivalent to the four arcs around x_2 in Figure 17; see Section 2F1. After drilling out γ, there is a cut vertex in the new Whitehead graph, and a single handle slide produces a graph G' with no cut vertices. By Corollary 2.8, there are no essential disks disjoint from G in $E(\gamma)$. Another possibility is that γ encircles all but one edge of G_1 and all but one edge of G_2, but in this case γ also encircles exactly one edge of G_1 and exactly one edge of G_2.

There are four types of arcs of length two encircling exactly one edge of G_1 and exactly one edge of G_2 (see Figure 18). Again, any arc encircling two edges of G, one of G_1 and one of G_2 can be slid in $E(F)$ into an arc of one of the four types; see Section 2F1.

After drilling out the arc γ, if γ is of one of the four types, the new Whitehead graph contains a cut vertex. After sliding, we end up with a graph G' whose simple associated graph is one of the drawings in Figure 19. Since these graphs contain no cut vertex, by Corollary 2.8 we conclude that any essential disk in $E(\gamma)$ intersects G, and therefore intersects $\Gamma \subset F$. This contradiction shows that $h(F) \neq 1$. Since $k = P(p,q,r)$ is not fibered and $h(F) \leq 2$, by Corollary 3.5 it follows that $h(F) = 2$ when $p,q,r \geq 5$.

This finishes Case 1.

Case 2: ($p < 0$ and $q,r > 0$) As in Example 3.9, we construct a spine $\Gamma = a_1 \vee a_2$ for F starting with the spine shown in Figure 9, but now we slide the middle edge of the θ-graph rightwards. The spine Γ looks like Figure 20, and the Whitehead diagram for $(E(F),\Gamma)$ with respect to the system of disks x_1, x_2 is as in Figure 21. By Remark 2.5, we only need to analyze the handle decompositions of $(E(F),\Gamma)$.

The Whitehead graphs G_1 and G_2 of the pairs $(E(F),a_1)$ and $(E(F),a_2)$, respectively, are shown in Figure 22. Although these diagrams are similar to the diagrams in Figure 12 of Case 1, the configuration of the diagram for a_1 here is not the same as the configuration of the positive case (Case 1); that is, the corresponding Whitehead diagrams are not isomorphic.

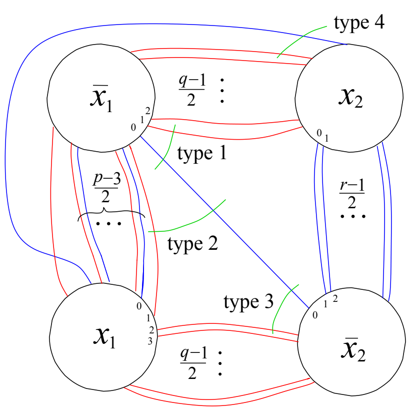


Figure 18

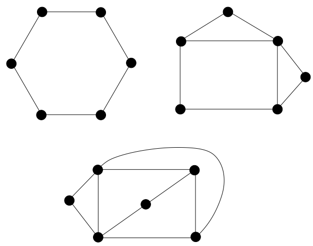


Figure 19

However, the analysis of the different properly embedded arcs in the Whitehead diagrams of $(E(F), a_1)$, $(E(F), a_2)$, and $(E(F), \Gamma)$, giving rise to a possible one-handled decomposition, is completely similar to that of Case 1.

The Whitehead diagram for $(E(F), a_2)$ is isomorphic to the corresponding Whitehead diagram of Case 1. Then:

Claim 1. *Let α be a properly embedded arc in $(E(F), a_2)$, disjoint from a_2, such that α is parallel into $\partial E(F)$, and α has length at least two in G_2. Then any essential disk in $E(\alpha)$ intersects a_2.* □

Claim 2. *Let α be a properly embedded arc in $(E(F), a_1)$, disjoint from a_1, such that α is parallel into $\partial E(F)$ and α has length at least two in G_1. Then any essential disk in $E(\alpha)$ intersects a_1.*

Proof. We first analyze arcs of length 2 in G_1. The arcs around vertices x_1 and $\bar{x}_1$ are shown in Figure 23. There are only two types after sliding the arcs in $E(F)$. After drilling out the arc α, if α is of type 1 or of type 2 the new Whitehead graph contains a cut vertex, but after sliding handles, as in Section 2F1, we end up with a graph G_1' whose simple associated graph is a cycle of six vertices and six edges; that is, this simple graph contains no cut vertex. Therefore G_1' contains no cut vertex, and, by Corollary 2.8, a_2 intersects every essential disk of $E(\alpha)$.

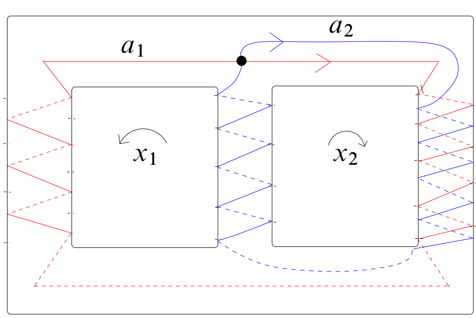


Figure 20

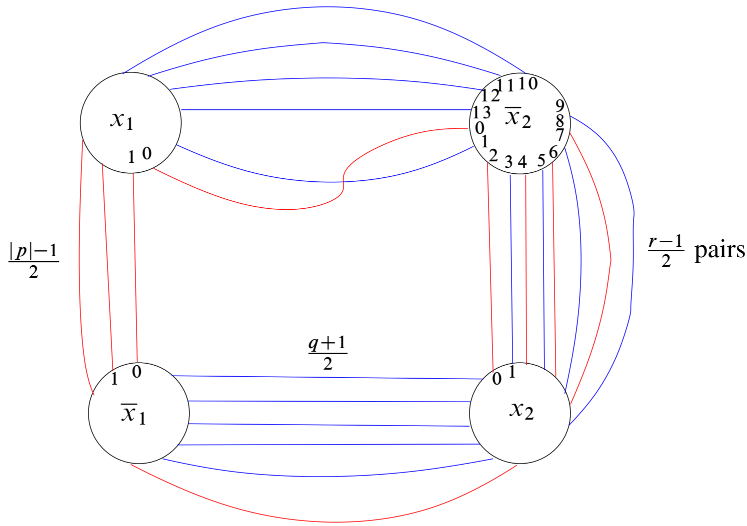


Figure 21

For arcs of length 2 around the vertices x_2 and $\bar{x}_2$, the analysis is identical to Case 1.

For arcs not around a vertex of G_1, there are two more types of arcs of length two, as in Figure 24, but, after drilling out the arc α of type 3 or 4, the new Whitehead graph contains no cut vertex, and then, by Corollary 2.8, a_2 intersects every essential disk of $E(\alpha)$.

For arcs of length at least three, we follow the same argument as in Case 1, and conclude that a_2 intersects every essential disk of $E(\alpha)$. □

Recall that we are assuming that γ is the core of a 1-handle of a one-handled circular decomposition of $E(k)$ based on F. In view of Claims 1 and 2, as in Case 1, we see that the arc γ encircles exactly one edge of G_1 and exactly one edge of G_2.

There are four types of arcs of length two encircling exactly one edge of G_1 and exactly one edge of G_2 (see Figure 25). For, such an arc can be slid in $E(F)$ into an arc of type 1, type 2, type 3, or type 4 (Section 2F1).

After drilling out the arc γ, if γ is of one of the four types, the new Whitehead graph contains a cut vertex. After sliding, we end up with a graph G' whose simple associated graph is one of the drawings in Figure 26. Since these graphs contain no cut vertex, by Corollary 2.8 we conclude that any essential disk in $E(\gamma)$ intersects G, and, therefore, intersects $\Gamma \subset F$. Thus $h(F) \neq 1$. Since $k = P(p,q,r)$ is not fibered and $h(F) \leq 2$, by Corollary 3.5 it follows that $h(F) = 2$ when $p \leq -5$ and $q, r \geq 5$.

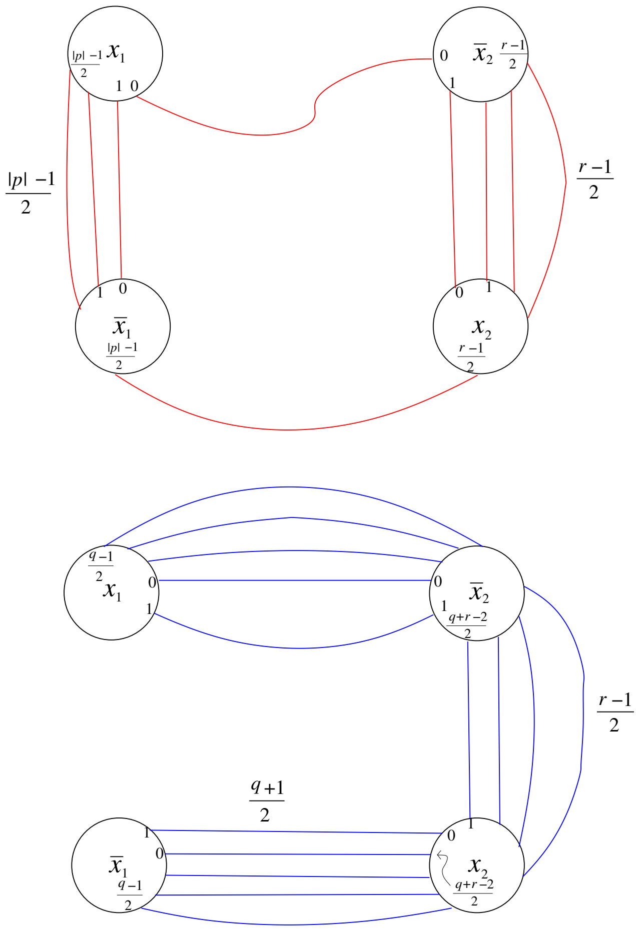


Figure 22

This finishes Case 2, and also the proof of Theorem 4.1. □

Corollary 4.4. *Let k be the pretzel knot $P(p,q,r)$ with $|p|,|q|,|r| \geq 5$. Then* $\mathrm{cw}(k) = 6$.

Proof. Since k has a unique incompressible Seifert surface, by Remark 2.3 it follows that $\mathrm{cw}(k) \in \mathbb{Z}$. By Theorem 4.1, $\mathrm{cw}(k) = 6$. □

Remark 4.5. Theorem 4.1 gives a family of knots of genus one and handle number two. This answers in the affirmative a question in [Hirasawa and Rudolph 2003]: does there exist a knot k with $h(k) > g(k)$?

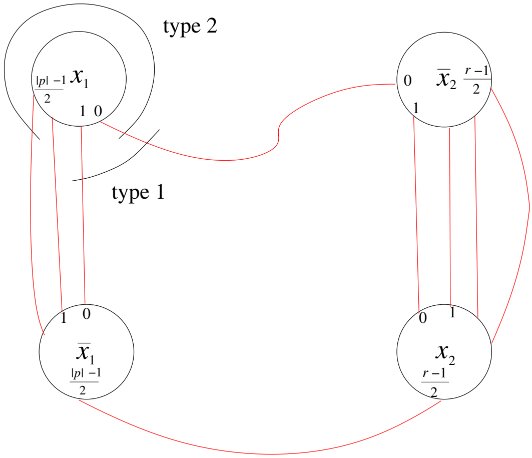


Figure 23

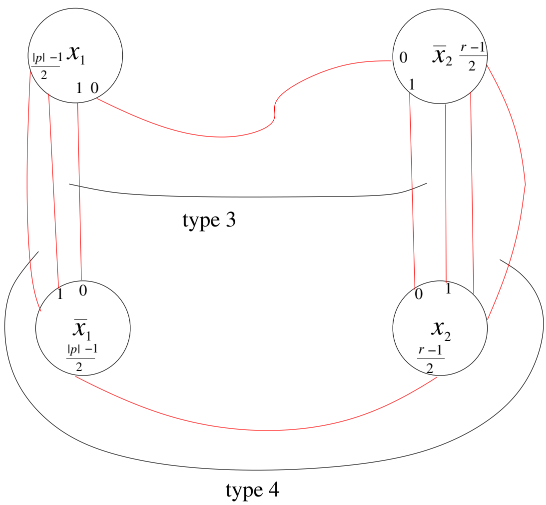


Figure 24

5. Genus one essential surfaces and powers of primitive elements

In this section we show that if k is a free genus one knot with at least two nonisotopic Seifert surfaces, then the free Seifert surface of k admits a special type of spine. This result is essential to prove the main theorem of Section 6 (Theorem 6.7).

Lemma 5.1. *Let H be a handlebody of genus $g \geq 2$ and let $\alpha \subset \partial H$ be a simple closed curve. Assume that there is a primitive element $p \in \pi_1(H)$ such that α*

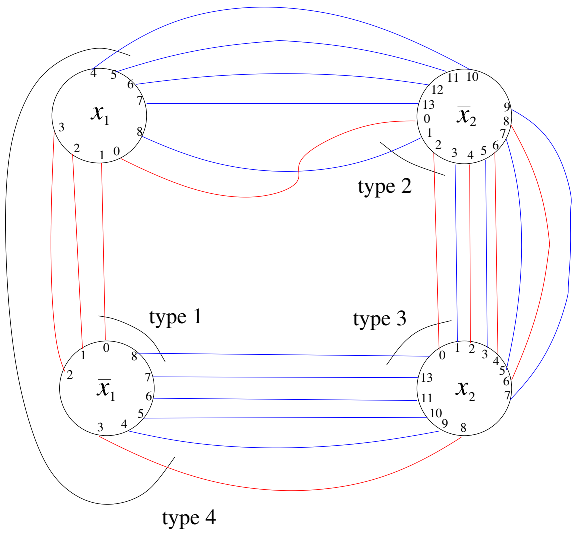


Figure 25

represents an element conjugate to p^n for some $n \in \mathbb{Z}$, $n \neq 0$. Then there is an essential 2-disk $D \subset H$ such that $D \cap \alpha = \varnothing$.

Proof. Let $\{p, q_2, \ldots, q_g\}$ be a basis for $\pi_1(H)$. Then $\pi_1(H) = \langle p \rangle * \langle q_2, \ldots, q_g \rangle$ is a nontrivial splitting, and α is conjugate to $p^n \in \langle p \rangle$. Then $\{\alpha\}$ is separable in $\pi_1(H)$, and the disk D is obtained by Theorem 3.2 of [Stallings 1999]. □

Let $\Gamma \cong a_1 \vee a_2$ be a graph in the boundary of a genus-two handlebody H. We say that a_2 *spoils disks for* a_1 if, for any essential disk $D \subset H$ such that $D \cap a_1 = \varnothing$, the number of points $\#(D \cap a_2)$ is at least 2.

Theorem 5.2. *Let $k \subset S^3$ be a nontrivial connected knot, and let $F \subset E(k)$ be a free genus one Seifert surface for k. Then*:

There is another genus-one Seifert surface for k which is not equivalent to F if and only if there exists a spine $\Gamma = a_1 \vee a_2$ for F in $\partial \mathcal{N}(F)$ such that a_1 represents an element conjugate to g^n with $n \geq 2$ for some primitive element $g \in \pi_1(E(F))$, and a_2 spoils disks for a_1.

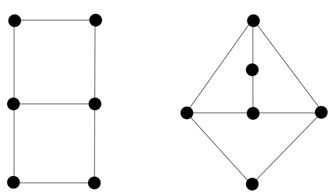


Figure 26

Proof. Let $\Gamma = a_1 \vee a_2$ be a spine for F such that a_1 represents an element conjugate to g^n with $n \geq 2$ for some primitive element $g \in \pi_1(E(F))$, and a_2 spoils disks for a_1.

Let $D \subset E(F)$ be an essential properly embedded disk such that $a_1 \cap D = \varnothing$, which is given by Lemma 5.1. We may assume that $H_1 = \overline{E(F) - \mathcal{N}(D)}$ is a solid torus. Let A_1 be a regular neighborhood of a_1 in $\partial E(F)$; then $A_1 \subset \partial H_1$. Write $B_1 = \overline{\partial H_1 - A_1}$. Since $|n| \geq 2$, the annuli A_1 and B_1 are nonparallel in H_1. We push $\mathrm{Int}(B_1)$ into H_1 to obtain B_1', a properly embedded annulus in H_1.

Let $\mathcal{N}(a_2) \subset \partial E(F)$ be a regular neighborhood of a_2 such that $A_1 \cap \mathcal{N}(a_2)$ is a rectangle; then $B_2 = \overline{\mathcal{N}(a_2) - A_1}$ is a "band" (that is, a 2-disk) such that $B_2 \cap A_1 = B_2 \cap B_1'$ is a pair of arcs in $\partial B_1'$. Then $G = B_2 \cup B_1'$ is a genus-one Seifert surface for k (we push $\mathrm{Int}(G)$ slightly into $E(F)$ to get a properly embedded surface in $E(F)$).

Now, $\widehat{G} = G \cap H_1$ is the union of the annulus B_1' with the disk components of $\widehat{B}_2 = B_2 \cap H_1$. Notice that $\partial \widehat{B}_2 \subset B_1 \subset \partial H_1$.

By hypothesis, $\#(a_2 \cap D) \geq 2$; thus, $\widehat{G}$ is disconnected, and the components of $\widehat{G}$ are $B_1' \cup$ (two 2-disks of $\widehat{B}_2$), and at least one subdisk $z \subset \widehat{B}_2$ with $\partial z \subset \mathrm{Int}(B_1)$.

Since $|n| \geq 2$, we cannot push B_1' onto A_1 in H_1. Then a ∂-parallelism for $\widehat{G}$ in H_1 contains a ∂-parallelism W for B_1' onto B_1, but then W contains the 2-disk $z \subset \widehat{G}$. Therefore, $\widehat{G}$ is not parallel into ∂H_1. We conclude that G is not boundary-parallel in $E(F)$, for a ∂-parallelism for G induces a ∂-parallelism for $\widehat{G}$. It follows that G and F are not equivalent. This finishes sufficiency.

Now, if there is another genus-one Seifert surface for k which is not equivalent to F, we can find still another nonequivalent genus-one Seifert surface $G \subset E(k)$ for k such that G and F have disjoint interiors; see [Scharlemann and Thompson 1988]. We write $k = G \cap \partial E(F)$.

The surface G splits $E(F)$ into two handlebodies, $H_0 \cup H_1 = \overline{E(F) - \mathcal{N}(G)}$, of genus two, for H_0 and H_1 are irreducible and, since G is π_1-injective into H_0 and H_1, it follows that H_0 and H_1 are π_1-injective into $E(F)$; therefore, H_0 and H_1 have free fundamental groups. We assume $\partial H_i = G \cup (F \times \{i\})$ plus a neighborhood of k ($i = 0, 1$). By considering a system of disks for the handlebody $E(F)$, we see that there is a disk $D \subset E(F)$ that ∂-compresses G in $E(F)$, and D is contained in, say, H_0, and is properly embedded in H_0.

Then k is a $((1,0),(n,m))$-curve in ∂H_0 (Lemma 4.3 of [Tsutsumi 2003]) with $|k \cap D| = 2$. See Figure 27.

Cutting H_0 along D, we obtain a solid torus $V \subset H_0$ such that $\widehat{G} = G \cap V$ is an (n,m)-torus annulus in ∂V; and the complementary annulus

$$\widehat{F} = \overline{\partial V - \widehat{G}}$$

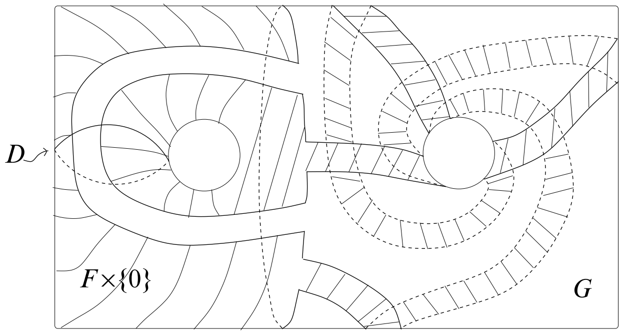


Figure 27. Surfaces G and $F\times\{0\}$ in H_0.

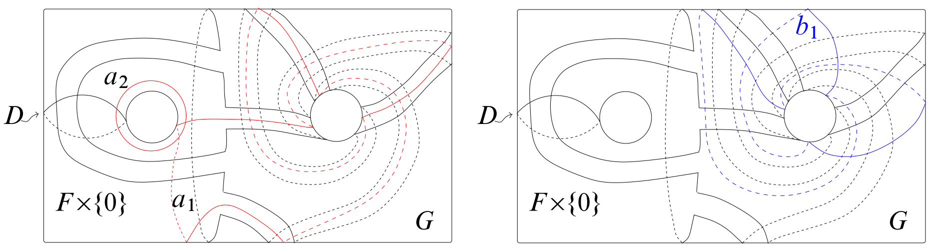


Figure 28. $\Gamma = a_1 \vee a_2$ and b_1.

contains, and is isotopic to, $(F\times\{0\})\cap V$ in ∂V with an isotopy fixed outside a regular neighborhood of D.

Let $a_1 \subset F\times\{0\}$ be the core of the annulus $\widehat{F}$, and let $b_1 \subset \widehat{G}$ be the core of the annulus $\widehat{G}$.

Let $C' \subset \partial V$ be a 2-disk that contains the pair of disks $\partial V \cap \mathcal{N}(D)$, and let $C \subset H_0$ be a properly embedded disk with $\partial C = \partial C'$. Now let $Z \subset H_0$ be a meridional disk such that $Z \cap C = \varnothing$. Then $\widetilde{F} = (F\times\{0\}) \cap (\overline{H_0 - \mathcal{N}(Z)})$ contains a (1,0)-annulus A in the solid torus $\overline{H_0 - \mathcal{N}(Z)}$. Let $a_2 \subset \mathrm{Int}(F\times\{0\})$ be the core of A, where we can arrange that $a_1 \cap a_2$ is just one point. Then $\Gamma = a_1 \vee a_2$ is a spine for F; see Figure 28.

The curve a_2 spoils disks for a_1 in $E(F)$, for otherwise there is an essential disk $D \subset E(F)$ such that $D \cap a_1 = \varnothing$, and the number of points $\#(D\cap a_2)$ is less than 2. If $D \cap a_2 = \varnothing$, since Γ is a spine for F, the surface F is contained in the solid torus $E(D) \subset E(F)$; it follows that F is compressible in $E(D)$, and thus F is compressible in $E(F)$. But, since k is nontrivial and $g(F) = 1$, F is incompressible in $E(k)$. Then $D \cap a_2$ is just one point, and $D \cap \partial F$ is a set of two points. We may assume that D intersects $k = \partial G$ in exactly two points. Since G is incompressible, we may arrange that $D \cap G$ is just one arc. Now, this arc is essential in G, for otherwise we can slide G along D and obtain G' homotopic to G in $E(F)$

such that G' is contained in the solid torus $E(D)$; then G' is not π_1-injective, and, since G and G' are homotopic embeddings, thus, G is not π_1-injective; but that makes G compressible. Then $\widehat{G} = G \cap E(D)$ is an annulus, therefore, $\widehat{G}$ is parallel into $\partial E(D)$. Using the disk D we can extend this parallelism to a parallelism of G into $\partial E(F)$, contradicting that G is essential in $(E(F), k)$.

Now, $a_1 \subset F \times \{0\}$ represents, up to conjugacy, the same element as $b_1 \subset G$ in $\pi_1(H_0)$ for, they are disjoint curves on a torus, and therefore, parallel.

Observe that, since G is not parallel to $F \times \{0\}$, we have $|n| \geq 2$. In particular, $\widehat{G}$ and $\widehat{F}$ are not parallel in V.

We now explore H_1.

Recall that D is a ∂-compression disk for G in $E(F)$; in particular, $D \cap \partial E(F)$ is an arc. It follows that, to recover $E(F)$ from $\overline{E(F) - \mathcal{N}(D)}$, we attach to $\overline{E(F) - \mathcal{N}(D)}$ the 3-ball $\mathcal{N}(D)$ along a disk. Then $\overline{E(F) - \mathcal{N}(D)}$ is a genus-two handlebody. In fact, $E(F)$ is a regular neighborhood of $\overline{E(F) - \mathcal{N}(D)}$. In particular, the inclusion induces an isomorphism $\pi_1(\overline{E(F) - \mathcal{N}(D)}) \to \pi_1(E(F))$.

Since $\overline{E(F) - \mathcal{N}(D)} = H_1 \cup_{\widehat{G}} V$, then $H_1 \cup_{\widehat{G}} V$ is a genus-two handlebody. Therefore, the core b_1 of $\widehat{G}$ represents a primitive element $\beta_1 \in \pi_1(H_1)$, for if $\pi_1(V) = \langle v : - \rangle$ then b_1 represents v^n, which is not primitive in V. The element β_1 is part of a basis, say, $\pi_1(H_1) = \langle w, \beta_1 : - \rangle$. By Seifert–van Kampen,

$$\pi_1(E(F)) \cong \pi_1(H_1 \cup_{\widehat{G}} V) = \langle w, \beta_1, v : \beta_1 = v^n \rangle \cong \langle w, v : - \rangle.$$

That is, v is primitive in $\pi_1(E(F))$, and b_1 represents v^n. □

6. Free genus one knots are almost fibered

In this section we show that all free genus one knots are almost fibered. We outline here the plan of the proof.

Start with a nonfibered free genus one knot k with a genus-one free Seifert surface $F \subset E(k)$. If k has a unique Seifert surface, then k is almost fibered (Remark 2.3). If k were not almost fibered, then, as in Remark 3.6, k has a genus-one Seifert surface not isotopic to F. By Theorem 5.2, there is a spine $\Gamma = a_1 \vee a_2$ for F in $\partial\mathcal{N}(F)$ such that a_1 represents an element conjugate to g^p with $p \geq 2$ for some primitive element $g \in \pi_1(E(F))$, and a_2 spoils disks for a_1. By Lemma 5.1, we can find an essential disk $\Delta \subset E(F)$ with $\Delta \cap a_1 = \varnothing$, and the exterior $E(\Delta) = \overline{E(F) - \mathcal{N}(\Delta)}$ is the disjoint union of two solid tori V_0 and V_1 with, say, $a_1 \subset \partial V_0$. We regard $\Delta \subset \partial V_0$. Then $\Gamma \cap V_0$ consists of the curve a_1, which is a (p, q)-curve in V_0, and an arc with endpoints on $\partial\Delta$ intersecting a_1 in exactly one point, and a set of parallel arcs with endpoints on $\partial\Delta$ which are disjoint from a_1; see Figure 32.

In Section 6A we show how to find a properly embedded arc in V_0 disjoint from Γ which, in Section 6B, is shown to be the core of the 1-handle of a one-handled circular decomposition for $E(k)$ based on F. In this analysis, the disk Δ is regarded as "unreachable", and should be thought of as very near the point at infinity. That is, all homeomorphisms in this subsection will fix pointwise the disk Δ.

6A. ***Handles for torus manifolds.*** Let p and q be a pair of coprime integers. Consider the points $\{s_\ell\}_{\ell=1}^p \subset S^1$ with $s_\ell = e^{2\pi i\ell/p}$; also let $\widetilde{V}$ be the cylinder $D^2 \times I$, and write $s_\ell^I = s_\ell \times I \subset \widetilde{V}$. The rotation ρ_q of angle $2\pi q/p$ on D^2 gives a quotient $P : \big(\widetilde{V}, \bigcup_{\ell=1}^p s_\ell^I\big) \to (V, \alpha)$, where V is the solid torus obtained from $\widetilde{V}$ by identifying $(z, 0)$ with $(\rho_q(z), 1)$ for each $z \in D^2$, and α is the simple closed curve on ∂V obtained as the image of the union $\bigcup_{\ell=1}^p s_\ell^I$ in this quotient. The rotation ρ_q acts on $\{s_\ell\}_{\ell=1}^p$ as the cyclic permutation of order p such that $\rho_q(s_i) = s_{i+q}$, where subindices are taken mod p. We consider also a fixed point $\infty \in \alpha$, the "point at infinity". The homeomorphism type of the pair (V, α) is called *the* (p, q)-*torus sutured manifold*, or simply *the* (p, q)-*manifold*. Throughout this section, we assume that $0 < q < p$. Notice that the (p, q)-torus sutured manifold (V, α) is not a sutured manifold, but α is a spine of a small regular neighborhood $\mathcal{N}(\alpha) \subset \partial V$, and the pair $(V, \mathcal{N}(\alpha))$ is a true sutured manifold with suture α.

In the following, we perform several operations on the (p, q)-manifold (drilling of arcs, homeomorphisms, etc.), and it will be done in such a way that the point at infinity of the manifold will remain fixed.

Let $x \subset V$ be the meridional disk $P(D^2 \times \{0\})$. From the pair $\big(\widetilde{V}, \bigcup_{\ell=1}^p s_\ell^I\big)$ we give a Whitehead diagram for the (p, q)-manifold (V, α) associated to x as follows.

We regard $\widetilde{V} = D^2 \times I$ as the exterior $E(x) \subset V$, and write x and $\bar{x}$ for $D^2 \times \{0\}$ and $D^2 \times \{1\}$, respectively. The arcs $s_1^I, \dots, s_p^I$ are the edges of G, the corresponding Whitehead graph with fat vertices x and $\bar{x}$. To obtain a Whitehead diagram, we have to number the endpoints of $s_1^I, \dots, s_p^I$. In a plane projection of the graph G, we assume that the unbounded face of G contains the edges s_q^I and s_{q+1}^I; see Figure 29. The point at infinity is either the middle point of s_q^I or the middle point of s_{q+1}^I. If $\infty \in s_q^I$, then we rename $v_j = (s_j, 0)$ and $\bar{v}_j = (\rho_q(s_j), 1) = (s_{j+q}, 1)$; if $\infty \in s_{q+1}^I$, we rename $v_j = (s_{j+q}, 0)$ and $\bar{v}_j = (\rho_q(s_{j+q}), 1) = (s_{j+2q}, 1)$, where subindices are taken mod p. In any case, we number the point v_i with the number i, and the point $\bar{v}_j$ with the number j $(i, j = 1, \dots, p)$. Also, we write α_i for the edge of G such that $v_i \in \alpha_i$. This diagram and the corresponding Whitehead graph are called *the* (p, q)-*diagram* and *the* (p, q)-*graph*, respectively. Notice that the edge α_1 connecting x with $\bar{x}$ starting at the point numbered $1 \in x$ ends at the point numbered $p - q + 1 \in \bar{x}$.

Remark 6.1. Consider a Whitehead diagram of a pair (V, α) associated to x, where V is a solid torus, α is a simple closed curve on ∂V, and x is a meridional

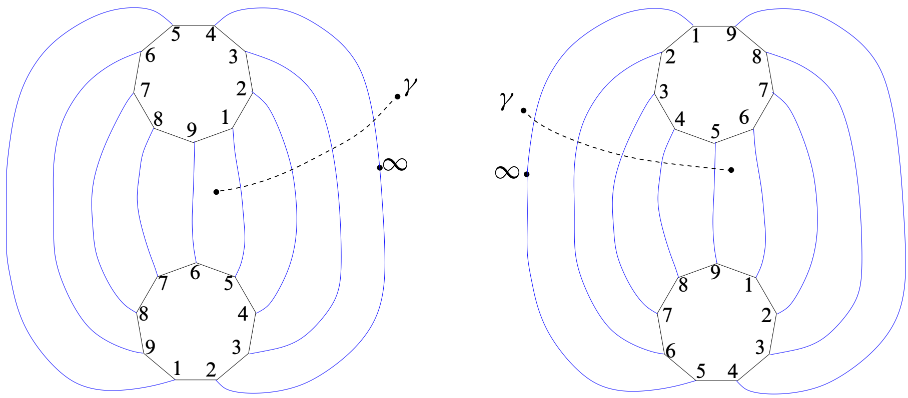


Figure 29. Whitehead diagrams for the (9,4)-manifold and the (9,5)-manifold.

disk of V. If, in the fat vertices of the Whitehead diagram of (V,α), the points corresponding to ends of edges are numbered with elements of the set $\{1,\ldots,p\}$ consecutively in the positive (negative) direction on x (on $\bar{x}$), in a compatible way with the gluing homeomorphism to recover the V, then if the edge connecting x with $\bar{x}$ starting at the point numbered $1 \in x$ ends at the point numbered $t \in \bar{x}$, then $t = p - q + 1$; that is, the Whitehead diagram corresponds to the (p,q)-torus sutured manifold with $q = p - t + 1$.

Let (V,α) be the (p,q)-torus sutured manifold, and let G be the Whitehead graph of (V,α) with respect to a meridional disk $x \subset V$. Let γ be a properly embedded arc in V, such that γ is around the vertex x in the Whitehead diagram of (V,α) with respect to x, and γ encircles the edges $\alpha_1,\ldots,\alpha_q$. Also, assume that γ lies "above" the point $\infty \in \alpha$, that is, γ is between ∞ and x; see Figure 29. The arc γ is called *the canonical 2-handle of length q* for the (p,q)-manifold. Note that the arc γ is the cocore of a 2-handle in V.

If we drill out the canonical 2-handle of length q, we obtain a Whitehead diagram with respect to the system of disks $x, z \subset E(\gamma) \subset V$, where z is the obvious ∂-parallelism disk for γ; see Figure 30. We refer to this Whitehead diagram as the *Whitehead diagram obtained by drilling out the canonical 2-handle of length q* of the (p,q)-manifold. Notice that the arc g in Figure 30 is a "longitude" for the handle defined by z. That is, if we glue back the disks z and $\bar{z}$ and kill the longitude g with a 2-handle, we recover the Whitehead diagram of the (p,q)-manifold. In practice, we just join the ends of the edges in z with the ends of the edges in $\bar{z}$ with parallel arcs on the diagram, and delete the disks z and $\bar{z}$ from the picture, and we get the Whitehead diagram of the (p,q)-manifold back.

Let G be the graph of the Whitehead diagram obtained by drilling out the canonical 2-handle of length q of the (p,q)-manifold. Then G is a graph with four

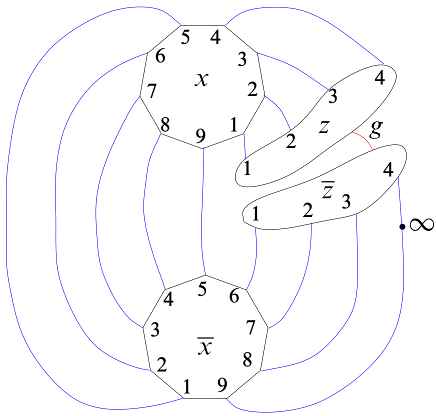


Figure 30. Whitehead diagram for the (9,4)-torus sutured manifold.

fat vertices $x, \bar{x}, z$, and $\bar{z}$; there are q edges connecting z and x; there are q edges connecting $\bar{z}$ and $\bar{x}$; and there are $p-q$ edges connecting x with $\bar{x}$. Compare with Figure 30. Note that x is a cut vertex of G (and z and $\bar{z}$ are *not* cut vertices); then we can slide the handle corresponding to z along the handle defined by x

After sliding, if the new disk x is still a cut vertex, we can again slide the new disk z along the new disk x, and so on. Let G' be the image of the graph G after κ handle slides of z along x. The graph G' is called *the κ-slid graph obtained from the (p, q)-graph G.*

Lemma 6.2. *Let p, q be a pair of coprime integers*, $0 < q < p$, *and assume that*

$$p = \kappa_1 q + r_1, \quad \textit{with } 0 \le r_1 < q \textit{ and } \kappa_1 \ge 1.$$

Let G be the graph of the Whitehead diagram obtained by drilling out the canonical 2-handle of length q of the (p, q)-manifold, and let G' be the κ_1-slid graph obtained from the (p, q)-graph G. Then G' is the graph of the Whitehead diagram obtained by drilling out the canonical 2-handle of length r_1 of the (q, r_1)-manifold. The point at infinity is a fixed point of these handle slides.

Proof. In the Whitehead graph G, the ends of the edges connecting the disk z with the disk x are numbered $1, 2, \ldots, q$ in the disk x; these ends are the points $v_1, v_2, \ldots, v_q$ in ∂x. Then, after sliding z along x, the new disk z carries the edges with ends that were numbered $1, 2, \ldots, q$ in $\bar{x}$. Thus, now the ends of the edges connecting z and x, after the slide, have ends which are the images of the points $v_1, v_2, \ldots, v_q$ under the rotation ρ_q of angle $2\pi q/p$; that is, the ends are the points $v_{q+1}, v_{q+2}, \ldots, v_{2q}$, which are numbered $q+1, q+2, \ldots, 2q$ in x.

We see that, after sliding $\kappa_1 - 1$ times z along x, the ends of the edges connecting z and x are numbered $(\kappa_1 - 1)q + 1, (\kappa_1 - 1)q + 2, \ldots, \kappa_1 q$ in x. Then, after

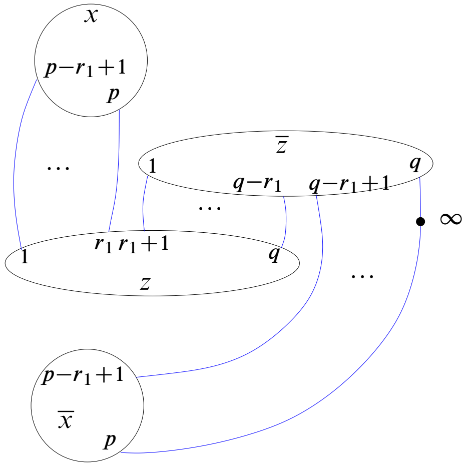


Figure 31. After sliding z along x.

sliding κ_1 times z along x, the points still connected by edges in x are numbered $\kappa_1 q+1, \kappa_1 q+2, \ldots, p$. Now, by hypothesis $p = \kappa_1 q + r_1$, so $\kappa_1 q + 1 = p - r_1 + 1$, which means that there are r_1 points left in x. That is (see Figure 31) we have a graph, the image of G after the slides, with fat vertices $x, \bar{x}, z, \bar{z}$; there are r_1 edges connecting x with z; there are r_1 edges connecting $\bar{x}$ with $\bar{z}$; and there are $q - r_1$ edges connecting z with $\bar{z}$. Now, the edge with one end in z numbered with 1 has the other end numbered with $p - r_1 + 1$ in x; and the edge with one end in $\bar{x}$ numbered with $p - r_1 + 1$ has the other end in $\bar{z}$ numbered with $q - r_1 + 1$.

Therefore, the new diagram is the Whitehead diagram obtained by drilling out the canonical 2-handle of length r_1 of the (q, r_1)-manifold. Since the disks $\bar{x}$ and $\bar{z}$ were never touched, the point at infinity is a fixed point of the handle slides.

Notice that if $q = 1$, then $\kappa_1 = p$, and $r_1 = 0$, and everything is easier: the image graph G above, in this case, replacing the values of q and r_1, has four fat vertices $x, \bar{x}, z, \bar{z}$; there are 0 edges connecting x with z; there are 0 edges connecting $\bar{x}$ with $\bar{z}$; and there is 1 edge connecting z with $\bar{z}$. That is, after canceling the handle defined by x, we obtain the (1,0)-manifold. $\square$

Corollary 6.3. *Let r_1, r_2 be a pair of coprime integers, $0 < r_2 < r_1$. Assume that*

$$\begin{aligned}
r_1 &= \kappa_1 r_2 + r_3, & 0 &< r_3 < r_2,\\
r_2 &= \kappa_2 r_3 + r_4, & 0 &< r_4 < r_3,\\
&\vdots & &\vdots\\
r_{n-1} &= \kappa_{n-1} r_n + 1, & 0 &< 1 < r_n,\\
r_n &= \kappa_n,
\end{aligned}$$

with $\kappa_i \geq 1$, $i = 1, \ldots, n$.

Let G be the graph of the Whitehead diagram obtained by drilling out the canonical 2-handle of length r_2 of the (r_1, r_2)-manifold. Let G_1 be the κ_1-slid graph obtained from the (r_1, r_2)-graph G. For $i = 1, \ldots, n-1$, let G_{i+1} be the κ_{i+1}-slid graph obtained from the (r_i, r_{i+1})-graph G_i.

Then G_n is the graph of the Whitehead diagram obtained by drilling out the canonical 2-handle of length 0 of the $(1,0)$-manifold (V, α).

The point at infinity is a fixed point of these handle slides. □

Remark 6.4. The graph G_i in the statement of Corollary 6.3 is the graph of the Whitehead diagram obtained by drilling out the canonical 2-handle of length r_{i+2} of the (r_{i+1}, r_{i+2})-manifold. Then G_i is a graph with four fat vertices $\xi, \bar{\xi}, \zeta$, and $\bar{\zeta}$. The symbols ξ and ζ stand for the symbols x and z in some order (that is, the sets $\{\xi, \zeta\}$ and $\{x, z\}$ are equal, but just as unordered sets). There are r_{i+2} edges connecting ζ and ξ, there are r_{i+2} edges connecting $\bar{\zeta}$ and $\bar{\xi}$ and there are $r_{i+1} - r_{i+2}$ edges connecting ξ with $\bar{\xi}$.

Remark 6.5. Let p, q be a pair of coprime integers, and assume that $p/q = [\kappa_1, \ldots, \kappa_n]$ as a continued fraction, with $\kappa_i \geq 1$ for each i.

(1) Write $p_i/q_i = [\kappa_1, \ldots, \kappa_i]$, with p_i, q_i coprime. Write $p_0 = 1$, $p_{-1} = 0$, and $q_0 = 0$, $q_{-1} = 1$. It is well known that $p_i = \kappa_i p_{i-1} + p_{i-2}$ and $q_i = \kappa_i q_{i-1} + q_{i-2}$; also $p_i q_{i-1} - p_{i-1} q_i = (-1)^i$ for $i \geq 1$ [Hall and Knight 1946, Articles 337 and 338]. Since $\kappa_i \geq 1$, one easily shows $p_i > q_i > 0$ for $i \geq 1$. In particular, $p > q > 0$. Note also that $p_{i+1} > p_i$.

(2) Let r, s be the two coprime integers p_{n-1}, q_{n-1}, respectively, and let (V, α) be the (p, q)-manifold. Then the (r, s)-torus curve can be drawn on ∂V as a simple closed curve β, which intersects α exactly at the point at infinity for $ps - qr = \pm 1$. Note that, if n is even, then the point at infinity is at the right in the Whitehead diagram, and if n is odd, it is at the left, as in Figure 29. The curve β can be visualized on the Whitehead diagram of the (p, q)-manifold as a set of new edges connecting the fat vertices, and disjoint from the Whitehead graph, and a single new edge intersecting the Whitehead graph at the point at infinity. Conversely, the curve α can be visualized in a similar way on the Whitehead diagram of the (r, s)-manifold.

   Notice that between two edges of α there is at most one edge of β, for $p > r$.

Theorem 6.6. *Assume $p/q = [\kappa_1, \ldots, \kappa_n]$ with p, q coprime, and $\kappa_i \geq 1$ for each i. Let r, s be the pair of coprime integers such that $r/s = [\kappa_1, \ldots, \kappa_{n-1}]$. Let (V, α) be the (p, q)-manifold, and let $\beta \subset \partial V$ be the (r, s)-torus curve such that α intersects β exactly at the point at infinity.*

If $\gamma \subset V$ is the canonical 2-handle of length q of the (p, q)-manifold, then the exterior $E(\gamma)$ is a regular neighborhood of $\alpha \cup \beta$.

Proof. Let G be the graph of the Whitehead diagram obtained by drilling out the canonical 2-handle of length q of the (p,q)-manifold, but including the arcs of the curve β. Call *α-edges* the edges of G corresponding to the (p,q)-torus curve α, and *β-edges* the edges of G corresponding to the (r,s)-torus curve β.

Writing $r_1 = p$ and $r_2 = q$, the statement $p/q = [\kappa_1, \dots, \kappa_n]$ with $\kappa_i \geq 1$ means that there are integers $r_3, \dots, r_n$ such that

$$\begin{aligned}
r_1 &= \kappa_1 r_2 + r_3, & 0 &< r_3 < r_2, \\
r_2 &= \kappa_2 r_3 + r_4, & 0 &< r_4 < r_3, \\
&\ \vdots & &\ \vdots \\
r_{n-1} &= \kappa_{n-1} r_n + 1, & 0 &< 1 < r_n, \\
r_n &= \kappa_n.
\end{aligned}$$

See Remark 6.5(1). Writing $\rho_1 = r$ and $\rho_2 = s$, the statement $r/s = [\kappa_1, \dots, \kappa_{n-1}]$ means that there are integers $\rho_3, \dots, \rho_{n-1}$ such that

$$\begin{aligned}
\rho_1 &= \kappa_1 \rho_2 + \rho_3, & 0 &< \rho_3 < \rho_2, \\
\rho_2 &= \kappa_2 \rho_3 + \rho_4, & 0 &< \rho_4 < \rho_3, \\
&\ \vdots & &\ \vdots \\
\rho_{n-2} &= \kappa_{n-2} \rho_{n-1} + 1, & 0 &< 1 < \rho_{n-1}, \\
\rho_{n-1} &= \kappa_{n-1}.
\end{aligned}$$

Notice that the canonical 2-handle of length q for the (p,q)-manifold is the canonical 2-handle of length q for the α-edges of G, but it is also the canonical 2-handle of length s for the β-edges of G. Then the graph G_{n-1} of Corollary 6.3 (Remark 6.4) contains four fat vertices $\xi, \bar{\xi}, \zeta$, and $\bar{\zeta}$. Note that $r_{n+1} = 1$; then there is a single α-edge connecting ζ and ξ, there is a single α-edge connecting $\bar{\zeta}$ and $\bar{\xi}$, and there are $r_n - 1$ α-edges connecting ξ with $\bar{\xi}$. Note that $\rho_n = 1$ and $\rho_{n+1} = 0$; then there is a single β-edge connecting ξ with $\bar{\xi}$ intersecting the α-edge connecting $\bar{\zeta}$ and $\bar{\xi}$ at the point at infinity, and there are no more β-edges. The graph G_n is obtained by sliding ζ through ξ the number $\kappa_n = r_n$ of times. Then G_n has a single α-edge connecting ξ with $\bar{\xi}$ and a single β-edge connecting ζ with $\bar{\zeta}$ intersecting at the point at infinity. The theorem follows.

Notice that when $q = 1$, then $n = 1$ and the graph G_{n-1} coincides with G. $\square$

6B. *One-handledness of knots.*

Theorem 6.7. *If k is a nonfibered free genus one knot in S^3, then k is almost fibered.*

Proof. Let $k \subset S^3$ be a knot and let $F \subset E(k)$ be a genus one free Seifert surface for k. Assume k is not almost fibered. Then, as in Remark 3.6, k has another genus-one Seifert surface disjoint from and not equivalent to F. By Theorem 5.2 there is a spine $\Gamma = a_1 \vee a_2$ for F in $\partial\mathcal{N}(F)$ such that a_1 represents an element conjugate to g^p with $p \geq 2$, for some primitive element $g \in \pi_1(E(F))$, and a_2 spoils the disks of a_1. We shall show that the existence of such a graph Γ implies $h(F) = 1$, and, since F is of minimal genus, therefore, $\mathrm{cw}(k) = 4$. This contradiction gives the theorem.

By Lemma 5.1, there is an essential 2-disk $\Delta \subset E(F)$ such that $\Delta \cap a_1 = \varnothing$. We may assume that the exterior $E(\Delta) \subset E(F)$ is not connected, and is the union of two solid tori H_0 and H_1, and $a_1 \subset H_0$. There is a copy of Δ in ∂H_0; then $a_1 \subset \partial H_0 - \Delta$. Write $T = \overline{\partial H_0 - \Delta}$; T is a once-punctured torus. A properly embedded arc $\alpha \subset T$ is called a rel-Δ curve in ∂H_0, and is visualized as the arc α union a properly embedded arc in Δ with the same ends as α. Or, rather, we may regard Δ as a point at infinity of the torus $T/\partial\Delta$.

We have that a_1 is a (p, q)-torus curve in H_0 for some q (this implies that we have fixed a longitude-meridian pair in ∂H_0; by changing the longitude-meridian pair, we may assume that $0 < q < p$). The intersection $a_2 \cap H_0 = a_2 \cap \partial H_0$ is a set of disjoint arcs $c \cup b_1 \cup \cdots \cup b_m \subset \partial H_0$ with ends in $\partial\Delta$ and such that $b_i \cap a_1 = \varnothing$ for each i, and the set $c \cap a_1$ is a single point, the base point of Γ.

Regarding c as a rel-Δ curve, c is an (r, s)-torus rel-Δ curve in H_0 with $ps - qr = \pm 1$. Since $ps - qr = \pm 1$, any other pair (r', s') such that $ps' - qr' = \pm 1$ is of the form $(r', s') = (r + \ell p, s + \ell q)$ for some integer ℓ. Then by sliding a_2 along $a_1^{\pm 1}$ several times, we obtain a new spine for F. By Remark 2.5, we may assume that the arc c is an (r, s)-torus rel-Δ curve in H_0 where, if $p/q = [\kappa_1, \ldots, \kappa_n]$ as a continued fraction with terms $\kappa_i \geq 1$, then $r/s = [\kappa_1, \ldots, \kappa_{n-1}]$.

Since $b_1, \ldots, b_m \subset \partial H_0 - (\mathrm{Int}(\Delta) \cup a_1 \cup c) \cong D^2$, then each of $b_1, \ldots, b_m$ are rel-Δ curves parallel to a_1.

Now consider the graph G of the Whitehead diagram of the (p, q)-manifold (H_0, a_1), and include in G the edges induced by the rel-∂ curves $c, b_1, \ldots, b_m$. By deforming the diagram, we may assume that Δ is contained in a small neighborhood of the point at infinity, which is the base point of Γ, the point of intersection of c and a_1. Let γ be the canonical 2-handle of length q for (H_0, a_1). In the Whitehead diagram, we place γ in such a way that it starts by encircling the arc c coming from infinity, and then encircles the q edges belonging to a_1 and whatever is in the middle, and nothing more (that is, after encircling the last edge belonging to a_1, the arc γ does not encircle any arc belonging to c or $b_1, \ldots, b_m$). See Figure 32, where the dotted line is a set of parallel arcs. We drill out γ and, by Theorem 6.6, if we slide handles in the Whitehead diagram obtained by drilling γ out of H_0, we

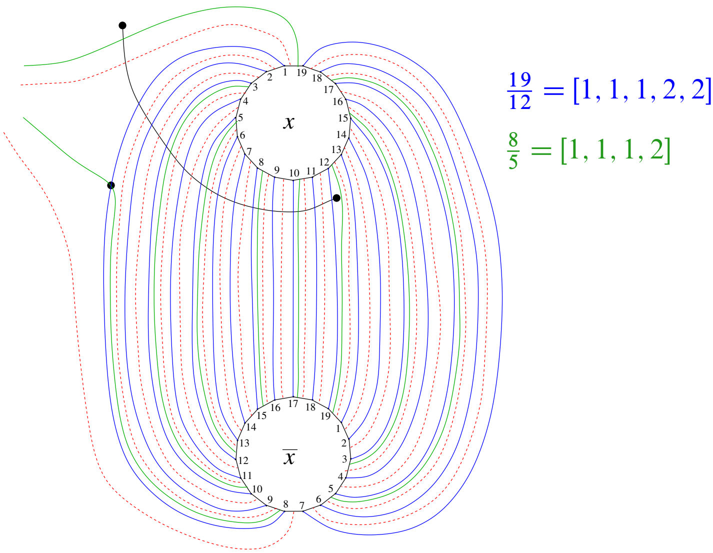


Figure 32. The (19,12) and (8,5)-torus curves.

obtain a sequence of diagrams as in Figures 33–35. All handle slides fix pointwise the small neighborhood of the point at infinity, and, thus, also the disk Δ.

The resulting Whitehead graph on ∂H_0 consists of four fat vertices $\xi, \bar{\xi}, \zeta, \bar{\zeta}$; there is a single a_1-edge connecting ξ and $\bar{\xi}$, and a single c-edge connecting ζ with $\bar{\zeta}$ intersecting in the base point of Γ (in Figure 32, $\xi = z$ and $\zeta = x$). Notice that the c-arc is actually two arcs, one connecting ζ with $\partial\Delta$, and the other connecting $\partial\Delta$ with $\bar{\zeta}$. Without loss of generality, this last arc contains the base point of Γ.

Let v be a meridional disk for H_1 disjoint from Δ. Then ξ, ζ and v form a system of meridional disks for the handlebody $E(\gamma)$. Write $\pi_1(E(\gamma)) = \langle \xi, \zeta, v : - \rangle$. Then a_1 represents the element ξ and a_2 represents an element $\bar{\zeta} \cdot W(\xi, v)$, where $W(\xi, v)$ is a word in the letters ξ and v. Since $\{\xi, \bar{\zeta} \cdot W(\xi, v), \zeta\}$ is a basis for $\pi_1(E(\gamma))$, it follows that a_1 and a_2 represent associated primitive elements. Then we can find a system of disks D_1, D_2, D_3 for $E(\gamma)$ such that $a_i \cap D_i$ is exactly one point, and $a_i \cap D_j = \varnothing$ for $i \neq j$, $i = 1, 2$, and $j = 1, 2, 3$. Therefore $\overline{E(\gamma) - \mathcal{N}(D_3)}$ is a regular neighborhood of $\Gamma = a_1 \vee a_2$. We conclude that D_3 is the cocore of a 1-handle that, together with γ, gives a one-handled circular decomposition for $E(k)$ as in Remark 2.2(2). Since k is not fibered, it follows that $h(k) = 1$, and that k is almost fibered. This contradiction finishes the proof of the theorem. □

Remark 6.8. By [Pajitnov 2010], a tunnel number one knot admits a one-handled circular decomposition based on some unspecified surface. In [Scharlemann 2004], genus-one knots with tunnel number one were classified, and it turns out that these knots are free genus one knots. Let k be a nonfibered genus-one knot with tunnel

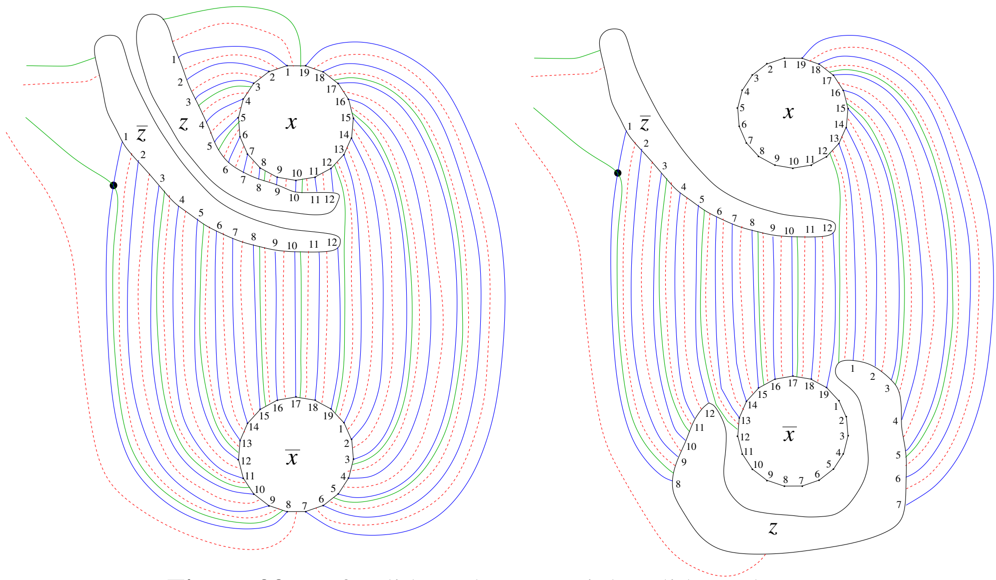


Figure 33. Left: slide z along x. Right: slide x along z.

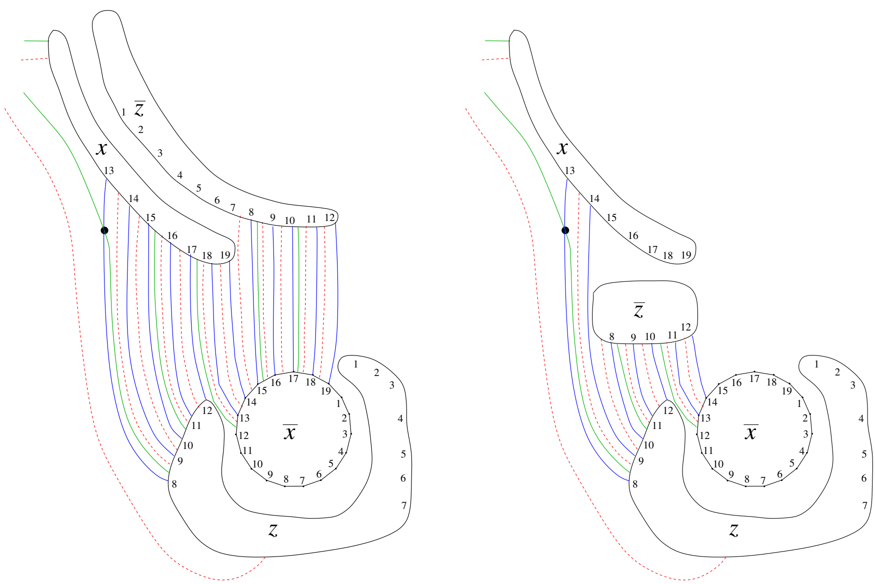


Figure 34. Left: slide $\bar{z}$ along $\bar{x}$. Right: slide $\bar{x}$ twice along $\bar{z}$.

number one. In Example 3.7, we considered the case that k is simple, and in the proof of Theorem 6.7, we considered the case that k is not simple. It follows that, for these knots, their circular width is realized with a one-handled circular decomposition based on a minimal (genus-one) free Seifert surface.

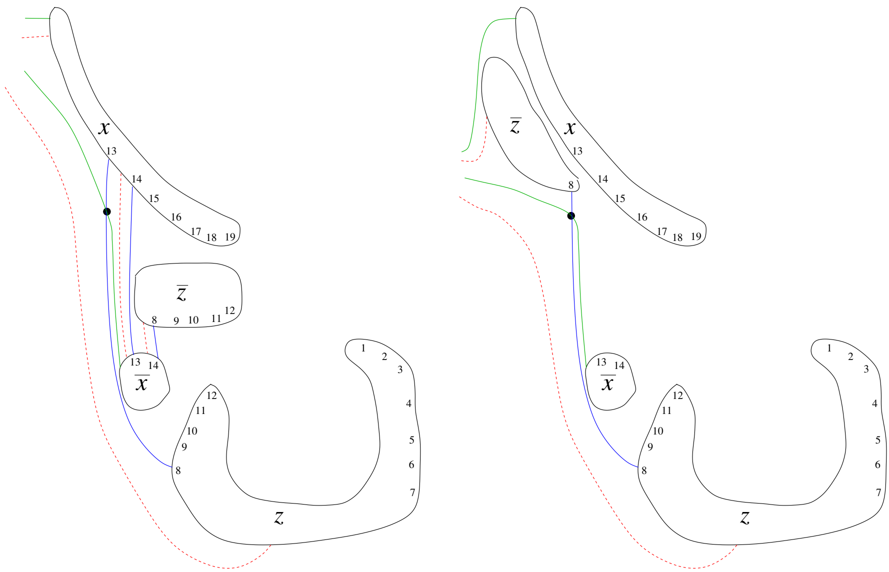


Figure 35. Left: slide twice $\bar{z}$ along $\bar{x}$. Right: A long slide of x deletes the curve.

References

[Gabai 1986] D. Gabai, "Genera of the arborescent links", pp. 1–98 Mem. Amer. Math. Soc. **339**, Amer. Math. Soc., Providence, RI, 1986. MR 87h:57010 Zbl 0585.57003

[Goda 1993] H. Goda, "On handle number of Seifert surfaces in S^3", *Osaka J. Math.* **30**:1 (1993), 63–80. MR 93k:57010 Zbl 0814.57004

[Goda 2006] H. Goda, "Circle valued Morse theory for knots and links", pp. 71–99 in *Floer homology, gauge theory, and low-dimensional topology*, edited by D. A. Ellwood et al., Clay Math. Proc. **5**, Amer. Math. Soc., Providence, RI, 2006. MR 2007e:57007 Zbl 1105.57010

[Goda and Ishiwata 2006] H. Goda and M. Ishiwata, "A classification of Seifert surfaces for some pretzel links", *Kobe J. Math.* **23**:1-2 (2006), 11–28. MR 2007k:57038 Zbl 1170.57006

[Hall and Knight 1946] H. S. Hall and S. R. Knight, *Higher algebra: A sequel to elementary algebra*, 4th ed., Macmillan, London, 1946. JFM 19.0061.06

[Hirasawa and Rudolph 2003] M. Hirasawa and L. Rudolph, "Constructions of Morse maps for knots and links, and upper bounds on the Morse–Novikov number", preprint, 2003. arXiv math/0311134

[Klimenko and Sakuma 1998] E. Klimenko and M. Sakuma, "Two-generator discrete subgroups of Isom($\mathbb{H}^2$) containing orientation-reversing elements", *Geom. Dedicata* **72**:3 (1998), 247–282. MR 2000a:20111 Zbl 0928.20040

[Manjarrez 2009] F. Manjarrez-Gutiérrez, "Circular thin position for knots in S^3", *Algebr. Geom. Topol.* **9**:1 (2009), 429–454. MR 2010i:57020 Zbl 1171.57005

[Pajitnov 2010] A. Pajitnov, "On the tunnel number and the Morse–Novikov number of knots", *Algebr. Geom. Topol.* **10**:2 (2010), 627–635. MR 2011c:57044 Zbl 1196.57008

[Scharlemann 2004] M. Scharlemann, "There are no unexpected tunnel number one knots of genus one", *Trans. Amer. Math. Soc.* **356**:4 (2004), 1385–1442. MR 2005e:57031 Zbl 1042.57003

[Scharlemann and Thompson 1988] M. Scharlemann and A. Thompson, "Finding disjoint Seifert surfaces", *Bull. London Math. Soc.* **20**:1 (1988), 61–64. MR 89a:57007 Zbl 0654.57005

[Scharlemann and Thompson 1994] M. Scharlemann and A. Thompson, "Thin position for 3-manifolds", pp. 231–238 in *Geometric topology* (Haifa, 1992), edited by C. Gordon et al., Contemp. Math. **164**, Amer. Math. Soc., Providence, RI, 1994. MR 95e:57032 Zbl 0818.57013

[Stallings 1999] J. R. Stallings, "Whitehead graphs on handlebodies", pp. 317–330 in *Geometric group theory down under* (Canberra, 1996), edited by J. Cossey et al., de Gruyter, Berlin, 1999. MR 2001i:57028 Zbl 1127.57300

[Tsutsumi 2003] Y. Tsutsumi, "Universal bounds for genus one Seifert surfaces for hyperbolic knots and surgeries with non-trivial JSJT-decompositions", *Interdiscip. Inform. Sci.* **9**:1 (2003), 53–60. MR 2004k:57013 Zbl 1055.57011

[Veber et al. 2001] K. Veber, A. Pazhitnov, and L. Rudolf, "The Morse–Novikov number for knots and links", *Algebra i Analiz* **13**:3 (2001), 105–118. In Russian; translated in *St. Petersburg Math. J.* **13**:3 (2002), 417–426. MR 2002k:57040 Zbl 1006.57003

[Whitehead 1936] J. H. C. Whitehead, "On equivalent sets of elements in a free group", *Ann. of Math.* (2) **37**:4 (1936), 782–800. MR 1503309 Zbl 0015.24804

Received January 14, 2014. Revised April 28, 2014.

FABIOLA MANJARREZ-GUTIÉRREZ
CIMAT
A.P. 402
36000 GUANAJUATO
MEXICO
fabiola@cimat.mx

VÍCTOR NÚÑEZ
CIMAT
A.P. 402
36000 GUANAJUATO
MEXICO
victor@cimat.mx

ENRIQUE RAMÍREZ-LOSADA
CIMAT
A.P. 402
36000 GUANAJUATO
MEXICO
kikis@cimat.mx

PACIFIC JOURNAL OF MATHEMATICS
Vol. 275, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.275.409

A POINTWISE A-PRIORI ESTIMATE FOR THE $\bar{\partial}$-NEUMANN PROBLEM ON WEAKLY PSEUDOCONVEX DOMAINS

R. MICHAEL RANGE

The main result is a pointwise a-priori estimate for the $\bar{\partial}$-Neumann problem that holds on an *arbitrary* weakly pseudoconvex domain D. It is shown that for $(0,q)$-forms f in the domain of the adjoint $\bar{\partial}^*$ of $\bar{\partial}$, the pointwise growth of the derivatives of each coefficient of f with respect to $\overline{z_j}$ and in complex tangential directions is carefully controlled by the sum of the suprema of f, $\bar{\partial} f$, and $\bar{\partial}^* f$ over D. These estimates provide a pointwise analog of the classical basic estimate in the L^2 theory that has been the starting point for all major work in this area involving L^2 and Sobolev norm estimates for the complex Neumann and related operators.

1. Introduction

The L^2 theory of the $\bar{\partial}$-Neumann problem on pseudoconvex domains has been highly developed for many years. In particular, J. J. Kohn [1979] introduced the technique of subelliptic multipliers which led to the proof of subelliptic estimates in the case where the boundary is of finite type [D'Angelo 1982; Catlin 1987; Siu 2010]. The starting point for these and other investigations has been the following basic estimate, valid on any smoothly bounded pseudoconvex domain D (see [Folland and Kohn 1972; Kohn 1979] for more details). Let us fix a point $P \in bD$ and a smooth orthonormal frame for $(1,0)$-forms $\omega_1, \omega_2, \dots, \omega_n$ on a small neighborhood U of P with $\omega_n = \gamma(\zeta)\partial r$, where r is a defining function for D. Let $L_1, \dots, L_n$ be the corresponding dual frame for $(1,0)$ vector fields. One defines

$$\mathfrak{D}_q(D) = C^\infty_{(0,q)}(\bar{D}) \cap \operatorname{dom} \bar{\partial}^*,$$

and one denotes by $\mathfrak{D}_{qU}$ those forms in $\mathfrak{D}_q(D)$ which have compact support in $\bar{D} \cap U$. Then $f \in \mathfrak{D}_{qU}$ can be written as $\sum'_J f_J \bar{\omega}^J$, where the summation is over strictly increasing q-tuples J. Since $f \in \operatorname{dom} \bar{\partial}^*$, one has $f_J = 0$ on $bD \cap U$

MSC2010: primary 32W05, 35N15; secondary 32A26, 32T27.
Keywords: a-priori estimates, $\bar{\partial}$ Neumann problem, integral representations, weakly pseudoconvex domains.

whenever $n \in J$. In case $q = 1$, the "L^2 basic estimate" states that there exists a constant C such that

$$\sum_{j,k} \|\overline{L_j} f_k\|^2 + \int_{bD\cap U} \mathcal{L}(r,\zeta; f^\sharp)\, dS(\zeta) \le C\big[\|\bar\partial f\|^2 + \|\bar\partial^* f\|^2 + \|f\|^2\big]$$

for all $f = \sum f_k \overline{\omega_k} \in \mathfrak{D}_{1U}$, where $f^\sharp = (f_1, \ldots, f_n)$. The norms here are the standard L^2 norms over $D \cap U$, and $\mathcal{L}$ is the Levi form of the defining function r with respect to the frame $\{L_1, \ldots, L_n\}$. Since $f \in \operatorname{dom} \bar\partial^*$, one has $f_n = 0$ on bD, so that pseudoconvexity implies that $\mathcal{L}(r,\zeta; f^\sharp) \ge 0$ on bD. Furthermore, it readily follows from $f_n|_{bD\cap U} = 0$ that one then also has the estimate

$$\|f_n\|_1^2 \le C_1 \|\bar\partial f_n\|^2 \le C_2\big[\|\bar\partial f\|^2 + \|\bar\partial^* f\|^2 + \|f\|^2\big].$$

Here $\|f_n\|_1$ is the full 1-Sobolev norm, i.e., $\|f_n\|_1^2$ is the sum of the squares of the L^2 norms of all first order derivatives of f_n.

Over the years there has been much interest in obtaining corresponding results involving pointwise and Hölder estimates. Techniques of integral representations have been most successful on *strictly* pseudoconvex domains, where the Levi polynomial provides a simple explicit local holomorphic support function (see [Range 1986] for a systematic exposition). Holomorphic support functions also exist on convex domains, and some results have been obtained in that setting in the case of finite type [Cumenge 1997; Diederich et al. 1999]. However, it has long been known that in general there are no analogous holomorphic support functions, even in very simple pseudoconvex domains of finite type [Kohn and Nirenberg 1973]. This obstruction has blocked any progress on these questions in the case of more general pseudoconvex domains.

Recently the author has introduced a *nonholomorphic* modification of the Levi polynomial to obtain new Cauchy–Fantappié kernels on arbitrary weakly pseudoconvex domains which reflect the complex geometry of the boundary and satisfy some significant partial estimates [Range 2013]. In this paper we use the new kernels in the integral representation formula developed by I. Lieb and the author in the strictly pseudoconvex case (see [Lieb and Range 1983; 1986]) to prove a pointwise analog of the classical basic L^2-estimate, as follows. This result was already announced in [Range 2011]. We define

$$\mathfrak{D}_q^k(D) = C^k_{(0,q)}(\overline{D}) \cap \operatorname{dom} \bar\partial^*$$

for $k = 1, 2, \ldots$, and we denote by $\mathfrak{D}^k_{qU}$ those forms in $\mathfrak{D}_q^k(D)$ that have compact support in $\overline{D} \cap U$. We shall use the frames $\omega_1, \omega_2, \ldots, \omega_n$ and $L_1, \ldots, L_n$ as above. Vector fields V act on forms coefficientwise, i.e., if $f = \sum_J' f_J \bar\omega^J$, then

$V(f) = \sum_J{}' V(f_J)\bar\omega^J$. For a C^1-form f of type $(0,q)$ on $\bar D$ we define the norm

$$Q_0(f) = |f|_0 + |\bar\partial f|_0 + |\vartheta f|_0,$$

where ϑ is the formal adjoint of $\bar\partial$, and $|\varphi|_0$ denotes the sum of the supremum norms over D of the coefficients of φ. For $0 < \delta \le 1$, $|\varphi|_\delta$ denotes the corresponding Hölder norm of order δ.

Main Theorem. *There exists an integral operator $S^{bD} : C_{(0,q)}(bD) \to C^\infty_{(0,q)}(D)$ which has the following properties. If bD is (Levi) pseudoconvex in a neighborhood U of the point $P \in bD$ and if U is sufficiently small, there exist constants C_δ depending on $\delta > 0$, so that one has the following uniform estimates for all $f \in \mathcal{D}^1_{qU}$, $1 \le q \le n$, and $z \in D \cap U$:*

(i) $|f - S^{bD}(f)|_\delta \le C_\delta Q_0(f)$ *for any* $\delta < 1$.

(ii) $|\overline{L_j} S^{bD}(f)(z)| \le C_\delta \operatorname{dist}(z, bD)^{\delta-1} Q_0(f)$ *for* $j = 1, \dots, n$ *and any* $\delta < \frac{1}{2}$.

(iii) $|L_j S^{bD}(f)(z)| \le C_\delta \operatorname{dist}(z, bD)^{\delta-1} Q_0(f)$ *for* $j = 1, \dots, n-1$ *and any* $\delta < \frac{1}{3}$.

Furthermore, if $f_J \bar\omega^J$ is a normal component of f with respect to the frame $\bar\omega_1, \dots, \bar\omega_n$, one has

$$|f_J|_\delta \le C_\delta Q_0(f) \quad \textit{for any } \delta < \tfrac{1}{2} \textit{ if } n \in J.$$

Note that if one also had an estimate analogous to (iii) for the normal derivative $L_n S^{bD}(f)(z)$ for *some* $\delta > 0$ (with $\delta < \frac{1}{3}$), standard results would imply the Hölder estimate $|S^{bD}(f)|_\delta \le C_\delta Q_0(f)$; by using (i) one therefore would obtain an estimate

$$|f|_\delta \le C_\delta Q_0(f),$$

i.e., the Hölder analog of a subelliptic estimate. It is known that such an estimate does not hold on arbitrary pseudoconvex domains. On the other hand, the Main Theorem provides a starting point in a general setting which, combined with additional suitable properties of the boundary such as finite type, might be useful to obtain appropriate estimates for $L_n S^{bD}(f)$. In particular, the author is investigating analogs of Kohn's subelliptic multipliers in the integral representation setting underlying the Main Theorem (see [Range 2011] for an outline of such potential applications).

2. Integral representations

We briefly recall some fundamentals of the integral representation machinery. We follow the terminology and notation from [Range 1986], where full details may be found. A (kernel) generating form $W(\zeta, z)$ for the smoothly bounded domain

$D \subset \mathbb{C}^n$ is a $(1,0)$-form $W = \sum_{j=1}^n w_j \, d\zeta_j$ defined on $bD \times D$ with coefficients of class C^1 which satisfies $\sum w_j(\zeta_j - z_j) = 1$. For $0 \le q \le n-1$, the associated Cauchy–Fantappié (= CF) form of order q is defined by

$$\Omega_q(W) = c_{nq} W \wedge (\overline{\partial_\zeta} W)^{n-q-1} \wedge (\overline{\partial_z} W)^q.$$[1]

$\Omega_q(W)$ is a double form on $bD \times D$ of type $(n, n-q-1)$ in ζ and type $(0,q)$ in z. One also sets $\Omega_{-1}(W) = \Omega_n(W) = 0$.

With $\beta = |\zeta - z|^2$, the form $B = \partial\beta/\beta = \sum_{j=1}^n \overline{(\zeta_j - z_j)}/|\zeta - z|^2 \, d\zeta_j$ is the generating form for the Bochner–Martinelli–Koppelman (= BMK) kernels. One has the following BMK formula (here and in the following, the integration variable is always ζ): if $f \in C^1_{(0,q)}(\bar{D})$ then for $z \in D$,

$$(1) \quad f(z) = \int_{bD} f(\zeta) \wedge \Omega_q(B) - \overline{\partial_z} \int_D f(\zeta) \wedge \Omega_{q-1}(B) - \int_D \overline{\partial_\zeta} f(\zeta) \wedge \Omega_q(B).$$

The next formula, due to W. Koppelman, describes how to replace $\Omega_q(B)$ by some other CF kernel $\Omega_q(W)$ on the boundary bD. Since $\Omega_n(B) \equiv 0$, we shall assume $q < n$ from now on. Given any generating form W on $bD \times D$, one has

$$f(z) = \int_{bD} f(\zeta) \wedge \Omega_q(W) + \overline{\partial_z} T_q^W(f) + T_{q+1}^W(\bar{\partial} f) \quad \text{for } f \in C^1_{(0,q)}(\bar{D}), z \in D.$$

Here the integral operator $T_q^W : C^1_{(0.q)}(\bar{D}) \to C_{(0,q-1)}(D)$ is defined by

$$T_q^W(f) = \int_{bD} f \wedge \Omega_{q-1}(W, B) - \int_D f(\zeta) \wedge \Omega_{q-1}(B)$$

for any $0 \le q < n$, where the "transition" kernels $\Omega_{q-1}(W, B)$ involve explicit expressions in terms of W and B which will be recalled later on.

Remark. For D strictly pseudoconvex, Henkin and Ramirez have constructed a generating form $W^{HR}(\zeta, z)$ that is holomorphic in z, so $\Omega_q(W^{HR}) = 0$ on bD for $q \ge 1$. Consequently, if f is a $\bar{\partial}$-closed $(0,q)$-form on $\bar{D}$, one has $f = \overline{\partial_z} T_q^{W^{HR}}(f)$, with an explicit solution operator $T_q^{W^{HR}}$. Based on the critical information that the Levi form of the boundary is positive definite in this case, it is well known that this solution operator is bounded from L^∞ into $\Lambda_{1/2}$. Furthermore, one also has the a-priori Hölder estimate $|f|_{1/2} \le C Q_0(f)$ for all $f \in \mathcal{D}^1_{qU}$ (see [Lieb and Range 1986]). Attempts to prove corresponding estimates on more general domains ultimately run into the obstruction of the example by Kohn and Nirenberg [1973] mentioned above, i.e., in general it is not possible to find a corresponding reasonably explicit *holomorphic* generating form on weakly pseudoconvex domains—even if of finite type—except under very restrictive geometric conditions.

[1] $c_{nq} = ((-1)^{q(q-1)/2}/(2\pi i)^n)\binom{n-1}{q}$

In case $f \in \mathfrak{D}^1_q(D)$, one may transform formula (1) into

$$f = \int_{bD} f \wedge \Omega_q(B) + (\bar\partial f, \bar\partial\omega_q) + (\vartheta f, \vartheta\omega_q),$$

where ω_q denotes the fundamental solution of $\Box$ on $(0,q)$-forms, ϑ denotes the formal adjoint of $\bar\partial$, so that $\vartheta f = \bar\partial^* f$, and $(\cdot,\cdot)$ denotes the standard L^2 inner product of forms over D (see [LR 1983] and [Range 1986]). The fundamental solution ω_q is an isotropic kernel whose regularity properties are well understood. In particular, the operator

$$S^{\text{iso}} : f \to S^{\text{iso}}(f) = (\bar\partial f, \bar\partial\omega_q) + (\vartheta f, \vartheta\omega_q)$$

satisfies a Hölder estimate

$$(2) \qquad |S^{\text{iso}}(f)|_\delta \le C_\delta Q_0(f) \quad \text{for all } f \in C^1_{(0,q)}(D) \text{ and any } \delta < 1.$$

Consequently, the essential information regarding all pointwise estimations is contained in the boundary integral $S^{bD}(f) = \int_{bD} f \wedge \Omega_q(B)$. The kernel of $\Omega_q(B)$ is isotropic; it treats derivatives in all directions equally, and direct differentiation under the integral in $\int_{bD} f \wedge \Omega_q(B)$ leads to an expression that will in general blow up like $\text{dist}(z, bD)^{-1}$. So this general representation of the operator S^{bD} does not provide any useful information.

Note that since $\Omega_n(B) \equiv 0$, the Main Theorem holds trivially with $S^{bD} \equiv 0$ when $q = n$.

By the Koppelman formulas, given any generating form W on $bD \times D$, one can transform $S^{bD}(f)$ into

$$(3) \quad S^{bD}(f) = \int_{bD} f \wedge \Omega_q(W) + \int_{bD} \bar\partial f \wedge \Omega_q(W, B) + \int_{bD} f \wedge \overline{\partial_z}\,\Omega_{q-1}(W, B).$$

The proof of the Main Theorem relies on formula (3) on a weakly pseudoconvex domain $D \Subset \mathbb{C}^n$, applied to the *nonholomorphic* generating form $W^{\mathcal{L}}(\zeta, z)$ introduced in [Range 2013]. Let us briefly recall the key properties of $W^{\mathcal{L}}(\zeta, z)$. Given a sufficiently small neighborhood $U = U(P)$, on $(bD \cap U) \times (D \cap U)$ the form $W^{\mathcal{L}}(\zeta, z)$ is represented explicitly by

$$W^{\mathcal{L}}(\zeta, z) = \frac{\sum_{j=1}^n g_j(\zeta, z)\, d\zeta_j}{\Phi_K(\zeta, z)},$$

where $\Phi_K(\zeta, z) = \sum_{j=1}^n g_j(\zeta, z)(\zeta_j - z_j)$ for $\zeta \in bD$. The (nonholomorphic) support function Φ_K is defined by

$$\Phi_K(\zeta, z) = F^{(r)}(\zeta, z) - r(\zeta) + K|\zeta - z|^3,$$

where $F^{(r)}(\zeta, z)$ is the Levi polynomial of a suitable defining function r, and $K > 0$ is a suitably chosen large constant. We note that $W^{\mathcal{L}}(\zeta, z)$ is C^∞ in z for $z \neq \zeta$. Recall from [Range 2013] that the neighborhood U, the constant K, and $\varepsilon > 0$ can be chosen so that for all $\zeta, z \in \overline{D} \cap U$ with $|\zeta - z| < \varepsilon$, one has

$$|\Phi_K(\zeta, z)| \gtrsim \bigl[|\operatorname{Im} F^{(r)}(\zeta, z)| + |r(\zeta)| + |r(z)| + \mathcal{L}(r, \zeta; \pi^t_\zeta(\zeta - z)) + K|\zeta - z|^3\bigr]. \tag{4}$$

Here $\pi^t_\zeta(\zeta - z)$ denotes the projection of $(\zeta - z)$ onto the complex tangent space of the level surface $M_{r(\zeta)}$ through the point ζ, and $\mathcal{L}(r, \zeta; \pi^t_\zeta(\zeta - z))$ denotes the Levi form of r at the point ζ. As shown in [Range 2013], the defining function r can be chosen so that $\mathcal{L}(r, \zeta; \pi^t_\zeta(\zeta - z)) \geq 0$ for all $\zeta \in \overline{D} \cap U$. We also recall that—as in the classical strictly pseudoconvex case—$r(\zeta)$ and $\operatorname{Im} F^{(r)}(\zeta, z)$ can be used as (real) coordinates in a neighborhood of a fixed point z.

Note that since $|\zeta - z|^3$ is real and symmetric in ζ and z, it follows from the known case $K = 0$ (see [Range 1986], for example) that if one defines $\Phi^*_K(\zeta, z) = \overline{\Phi_K(z, \zeta)}$, one has the approximate symmetry

$$\Phi^*_K - \Phi_K = \mathcal{E}_3.\text{[2]} \tag{5}$$

In the following, we simplify the notation by dropping the subscript K, i.e., we will write Φ instead of Φ_K.

For $0 \leq q < n$ we thus consider the integral representation formula

$$f = S^{bD}(f) + S^{\text{iso}}(f) \quad \text{for } f \in \mathfrak{D}^1_{0,q}(\overline{D}), \tag{6}$$

where the boundary operator S^{bD} is given by

$$\int_{bD} f \wedge \Omega_q(W^{\mathcal{L}}) + \int_{bD} \bar\partial f \wedge \Omega_q(W^{\mathcal{L}}, B) + \int_{bD} f \wedge \bar\partial_z \Omega_{q-1}(W^{\mathcal{L}}, B), \tag{7}$$

for $f \in \mathfrak{D}^1_{0,q}(\overline{D})$. Corresponding formulas hold locally on $U \cap bD$ whenever the boundary is Levi pseudoconvex in U. It is then clear that property (i) in the Main Theorem is satisfied. The main difficulty involves establishing the estimates (ii) and (iii).

The proof of the Main Theorem involves a careful analysis of the boundary integrals in formula (7). In contrast to [Lieb and Range 1983], which we henceforth abbreviate [LR 1983], for the most part we deal directly with the integrals over bD, thereby simplifying the analysis. However, for the most critical terms we will need to apply Stokes' theorem and introduce the Hodge $*$ operator as in [LR 1983] to transform the integrals into standard L^2 inner products of forms over $D \cap U$, and exploit certain approximate symmetries in the kernels.

[2] $\mathcal{E}_j$ denotes a smooth expression which satisfies $|\mathcal{E}_j| \leq C|\zeta - z|^j$.

3. The integral $\int_{bD} f \wedge \Omega_q(W^{\mathcal{L}})$

When D is strictly pseudoconvex, $W^{\mathcal{L}}$ can be chosen to be holomorphic in z for ζ close to z, so that the estimations become trivial if $q \geq 1$, since then $\overline{\partial_z} W = 0$ near the singularity. In the general case considered here, this integral needs to be carefully estimated as well. The analysis of this integral involves straightforward modifications of the case $q = 0$ discussed in [Range 2013], as follows.

We only consider ζ, z with $|\zeta - z| < \frac{1}{2}\varepsilon$, so that we can use the explicit form of $W^{\mathcal{L}} = g/\Phi$ recalled above, and the local frames $\{\omega_1, \dots, \omega_n\}$ and $\{L_1, \dots, L_n\}$. Recall that for $j = 0, 1, 2$, an expression $\mathcal{E}_j^\sharp$ denotes a form which is smooth for $\zeta \neq z$ and that satisfies a uniform estimate $|\mathcal{E}_j^\sharp| \lesssim |\zeta - z|^j$, and whose precise formula may change from place to place. While Φ is not holomorphic in z, one has $\overline{\partial_z} \Phi = \mathcal{E}_2^\sharp$; furthermore, one has $L_j^z \Phi = \mathcal{E}_1^\sharp$ for $j < n$, while $L_n^z \Phi \neq 0$ at $\zeta = z$.

By the properties of CF forms, on bD one has

$$\Omega_q(W^{\mathcal{L}}) = c_{nq} \frac{g \wedge (\overline{\partial_\zeta} g)^{n-q-1} \wedge (\overline{\partial_z} g)^q}{\Phi^n}.$$

The coefficients g_j of $g = \sum g_j \, d\zeta_j$ are given by

$$g_j = \frac{\partial r}{\partial \zeta_j} - \frac{1}{2} \sum_k \frac{\partial^2 r}{\partial \zeta_j \partial \zeta_k} (\zeta_k - z_k) + \mathcal{E}_2^\sharp.$$

The form of g implies that

$$g = \partial r(\zeta) + \mathcal{E}_1 + \mathcal{E}_2^\sharp, \quad \bar\partial_\zeta g = \bar\partial \partial r(\zeta) + \mathcal{E}_1^\sharp, \quad \text{and} \quad \bar\partial_z g = \mathcal{E}_1^\sharp.$$

It follows readily that for $0 \leq t \leq n-1$ one has

$$g \wedge (\overline{\partial_\zeta} g)^t = \partial r(\zeta) \wedge \sum_{k=0}^{t} [\bar\partial \partial r(\zeta)]^k (\mathcal{E}_1^\sharp)^{t-k} + \sum_{k=0}^{t} [\bar\partial \partial r(\zeta)]^k (\mathcal{E}_1^\sharp)^{t-k+1},$$

where $(\mathcal{E}_1^\sharp)^s$ denotes a generic form of appropriate degree whose coefficients are products of s terms of type $\mathcal{E}_1^\sharp$ in the case $s \geq 1$, or a term of type $\mathcal{E}_0^\sharp$ for $s = 0, -1$.

Note that since $\iota^*(\omega_n \wedge \overline{\omega_n}) = 0$ on bD, the pullback of $\partial r(\zeta) \wedge [\bar\partial \partial r(\zeta)]^k$ to bD involves only *tangential* components $\tan[\bar\partial \partial r(\zeta)]$, while the pullback of $[\bar\partial \partial r(\zeta)]^k$ alone will involve exterior products of at least $k-1$ different tangential components.

When estimating integrals involving these expressions, we make use of the fact that — in suitable z-diagonalizing coordinates (see [Range 2013]) — each *tangential* component $\tan[\bar\partial \partial r(\zeta)]$ in the numerator of the kernel reduces the order of the vanishing of the corresponding factor Φ in the denominator from three to an estimate $\gtrsim |\zeta_l - z_l|^2$, i.e.,

$$|\tan[\bar\partial \partial r(\zeta)]/\Phi| \lesssim 1/(|r(z)| + |\zeta_l - z_l|^2), \tag{8}$$

where ζ_l is an appropriate complex tangential coordinate. Similarly,

$$|\mathcal{E}_1^\sharp/\Phi| \lesssim 1/(|r(z)| + |\zeta - z|^2).$$

In order to keep track of these estimates, we introduce forms $\mathcal{L}[\mu]$ of *Levi weight* μ as follows. If $\mu \geq 1$, we say $\mathcal{L}[\mu]$ has Levi weight μ if each summand of $\mathcal{L}[\mu]$ contains at least μ factors which either are (different) purely tangential components of $\bar{\partial}\partial r(\zeta)$, or of type $\mathcal{E}_1^\sharp$. We also set $\mathcal{L}[\mu] = \mathcal{E}_0$ if $\mu \leq 0$. It then follows that

$$g \wedge (\overline{\partial_\zeta} g)^t = \mathcal{L}[t]$$

on the boundary, and consequently the numerator of $\Omega_q(W^{\mathcal{L}})$ satisfies

$$g \wedge (\overline{\partial_\zeta} g)^{n-q-1} \wedge (\overline{\partial_z} g)^q = \mathcal{L}[n-1]. \tag{9}$$

Proposition 1. *For any q with $0 \leq q \leq n-1$, the operator*

$$T_q^{\mathcal{L}} : C_{(0,q)}(bD) \to C^\infty_{(0,q)}(D),$$

defined by

$$T_q^{\mathcal{L}} f(z) = \int_{bD} f(\zeta) \wedge \Omega_q(W^{\mathcal{L}})(\zeta, z),$$

satisfies the estimates

$$\left|\overline{L_j^z}(T_q^{\mathcal{L}} f(z))\right| \leq C_\delta |f|_0 \operatorname{dist}(z, bD)^{\delta-1} \qquad \text{for } \delta < \tfrac{2}{3} \text{ and } 1 \leq j \leq n, \tag{10}$$

$$\left|L_j^z(T_q^{\mathcal{L}} f(z))\right| \leq C_\delta |f|_0 \operatorname{dist}(z, bD)^{\delta-1} \qquad \text{for } \delta < \tfrac{1}{3} \text{ and } j \leq n-1, \tag{11}$$

for suitable constants C_δ.

Given the estimation (9) of the numerator of $\Omega_q(W^{\mathcal{L}})$, the proof given in [Range 2013] for the case $q = 0$ and for the derivatives $\overline{L_j^z}$ carries over to the general case. To prove the estimate (11), one uses $L_j^z \Phi = \mathcal{E}_1^\sharp$ for $j \leq n-1$, which implies that $|L_j^z \Phi/\Phi| \preceq \operatorname{dist}(z, bD)^{-2/3}$. The estimations then proceed as in [Range 2013].

Remark. There is no corresponding estimate for the differentiation with respect to L_n^z, i.e., in the normal direction, since $L_n^z \Phi \neq 0$ for $\zeta = z$; therefore the operator $T_q^{\mathcal{L}}$ is not smoothing, i.e., there is no Hölder estimate

$$|T_q^{\mathcal{L}} f|_\delta \leq C_\delta |f|_0 \quad \text{for any } \delta > 0.$$

Proposition 1 provides a *partial* smoothing property:

Definition 2. A kernel $\Gamma(\zeta, z)$, or the integral operator $T_\Gamma : C_*(\bar{D}) \to C^1_*(D)$ defined by it, is $\bar{z}$-smoothing of order $\delta > 0$ if T_Γ satisfies the estimates (10). Similarly, we say that Γ (or T_Γ) is tangentially smoothing of order δ if T_Γ satisfies the estimates (11) for L_j^z and $\overline{L_j^z}$ for $j = 1, \ldots, n-1$.

Here C_* denotes spaces of forms of appropriate type.

4. Boundary admissible kernels

Before proceeding with the analysis of the integrals involving the transition kernels, we introduce admissible kernels and their weighted order by suitably modifying corresponding notions from [LR 1983]. We say that a kernel $\Gamma(\zeta, z)$ defined on

$$bD \times \bar{D} - \{(\zeta, \zeta) : \zeta \in bD\}$$

is simple *admissible* if for each $P \in bD$, there exists a neighborhood U of P, such that on $(bD \cap U) \times (\bar{D} \cap U)$ there is a representation of the form

$$\Gamma = \frac{\mathcal{L}[\mu](\mathcal{E}_1^\sharp)^j}{\Phi^{t_1}\beta^{t_0}},$$

where all exponents are ≥ 0. Note that j may be a noninteger, in which case $(\mathcal{E}_1^\sharp)^j$ denotes a form which is estimated by $C|\zeta - z|^j$. Such a representation is said to have (weighted) *boundary* order $\geq \lambda$ ($\lambda \in \mathbb{R}$) provided:

i) if $t_1 \geq 1$ and $\mu \geq 1$, then

$$2n - 1 + j - 1 - 2\max(0, \min(t_1 - 1, \mu)) - 3\max(t_1 - 1 - \mu, 0) - 2t_0 \geq \lambda;$$

while if $\mu \leq 0$ then

$$2n - 1 + j - 1 - 3\max(t_1 - 1, 0) - 2t_0 \geq \lambda;$$

or

ii) if $t_1 = 0$, then

$$2n - 1 + j - 2t_0 \geq \lambda.$$

This definition of order takes into account that the dimension of bD is $2n - 1$, and that *one* factor Φ may be counted with weight 1, since by estimate (4) one has $|\Phi| \gtrsim |\operatorname{Im} F^r|$, and $\operatorname{Im} F^r(\cdot, z)$ serves as a local coordinate on the boundary in a neighborhood of z.

A kernel Γ is admissible of boundary order $\geq \lambda$ if it is a finite sum of simple admissible kernels with representations of boundary order $\geq \lambda$.

The results in the previous section show that $\Omega_q(W^{\mathcal{L}})$ is admissible of boundary order ≥ 0.

As in the strictly pseudoconvex case considered in [LR 1983], an admissible kernel Γ of boundary order $\lambda \geq 1$ is smoothing of some positive order δ. This follows from an estimate

$$|V^z T_\Gamma(f)(z)| \leq C_\delta |f|_0 \operatorname{dist}(z, bD)^{-1+\delta},$$

for any vector field V^z of unit length acting in z. On the other hand, admissible kernels of boundary order $\lambda = 0$ are not smoothing in general.

More precisely, we have:

Theorem 3. *Let Γ_λ be an admissible kernel of boundary order $\geq \lambda$, and let*

$$J_\lambda(z) = \int_{bD} |\Gamma_\lambda(\zeta, z)|\, dS(\zeta).$$

$J_\lambda(z)$ has the following properties:

(a) *If $\lambda > 0$, then $\sup_D J_\lambda(z) < \infty$.*

(b) *If $\lambda = 0$, then $J_0(z) \lesssim \operatorname{dist}(z, bD)^{-\alpha}$ for any $\alpha > 0$.*

(c) *If $\lambda \geq 1$, then Γ_λ is smoothing of order δ for any $\delta < \frac{1}{3}$, tangentially smoothing of order $\delta < \frac{2}{3}$, and $\bar{z}$-smoothing of order $\delta < 1$.*

(d) *If $\lambda \geq 2$, then Γ_λ is smoothing of order δ for any $\delta < \frac{2}{3}$.*

Proof. Part (a) was essentially proved in [Range 2013] for the kernel $\Omega_0(W^{\mathcal{L}})$. The general case follows by the same arguments. Part (b) follows from (a) by noting that $\operatorname{dist}(z, bD)^\alpha \leq |\zeta - z|^\alpha$ for $\zeta \in bD$. For (c), note that given a vector field V^z, all terms in $V^z \Gamma_\lambda$ are of boundary order $\geq \lambda - 1 \geq 0$, except those where differentiation is applied to Φ; in that case use $V^z(\Phi^{-s}) = (\Phi^{-s})[\mathcal{E}_0^\sharp/\Phi]$ and $|1/\Phi| \lesssim |r(z)|^{-2/3}/|\zeta - z)|$ to see that $V^z\Gamma$ is estimated by $|r(z)|^{-2/3}$ multiplied with a kernel of order ≥ 0. Similarly, in the case where V^z is tangential, one can replace $|r(z)|^{-2/3}$ by $|r(z)|^{-1/3}$, and in the case $V^z = \overline{L_j^z}$, one uses $\overline{L_j^z}(\Phi^{-s}) = (\Phi^{-s})[\mathcal{E}_2^\sharp/\Phi]$ to see that $\overline{L_j^z}\Gamma_\lambda$ is of order ≥ 0. The required estimates then follow from (b). Finally, (d) follows by appropriately modifying the proof of (c). $\square$

The most significant part of this paper is the analysis of the kernels of order zero. Such kernels are not smoothing in general. However, as we saw for $\Omega_q(W^{\mathcal{L}})$, it turns out that in many cases they are at least $\bar{z}$-smoothing and tangentially smoothing of some positive order. On the other hand, one readily checks that kernels of type such as $\mathcal{E}_1^\sharp/\beta^n$ (e.g., those appearing in the BMK kernels) or $1/(\Phi\beta^{n-1})$, which are of boundary order zero, do not give preference to tangential or $\bar{z}$-derivatives, and consequently such kernels are not $\bar{z}$-smoothing of any positive order. Therefore one needs to analyze the kernels of boundary order ≥ 0 that arise in the current setting more carefully in order to obtain the estimates stated in the Main Theorem. It will be convenient to introduce the following notation:

Definition 4. The symbol Γ_λ denotes an admissible kernel of (boundary) order $\geq \lambda$. We denote by $\Gamma^{\bar{z}}_{0,1/2}$ (resp. $\Gamma^{\bar{z}}_{0,2/3}$) an admissible kernel of order ≥ 0 which is $\bar{z}$-smoothing of any order $\delta < \frac{1}{2}$ (resp. $\delta < \frac{2}{3}$) and tangentially smoothing of any order $\delta < \frac{1}{3}$.

According to this definition, Proposition 1 states that $\Omega_q(W^{\mathcal{L}})$ is a kernel of type $\Gamma^{\bar{z}}_{0,2/3}$. Similarly, we note that by Theorem 3 a kernel of type Γ_1 is (better than) of type $\Gamma^{\bar{z}}_{0,2/3}$, and, in fact, is smoothing of order $\delta < \frac{1}{3}$ in *all* directions.

5. The integrals $\int_{bD} \bar\partial f \wedge \Omega_q(W^{\mathcal{L}}, B)$ and $\int_{bD} f \wedge \overline{\partial_z}\Omega_{q-1}(W^{\mathcal{L}}, B)$

We recall (see [LR 1983] and [Range 1986], for example) that for $0 \le q \le n-2$ the transition kernels $\Omega_q(W^{\mathcal{L}}, B)$ are defined by

$$\text{(12)}\quad \Omega_q(W^{\mathcal{L}}, B) = (2\pi i)^{-n} \sum_{\mu=0}^{n-q-2} \sum_{k=0}^{q} a_{\mu,k,q} W^{\mathcal{L}} \wedge B \wedge (\overline{\partial_\zeta} W^{\mathcal{L}})^{\mu} \wedge (\overline{\partial_\zeta} B)^{n-q-2-\mu} \wedge (\overline{\partial_z} W^{\mathcal{L}})^{k} \wedge (\overline{\partial_z} B)^{q-k},$$

where the coefficients $a_{\mu,k,q}$ are certain rational numbers, while $\Omega_{n-1}(W^{\mathcal{L}}, B) \equiv 0$. Again, it is enough to consider $|\zeta - z| \le \frac{1}{2}\varepsilon$, so that $W^{\mathcal{L}} = g/\Phi$. It then follows from (12) and standard results about CF form that on bD the form $\Omega_q(W^{\mathcal{L}}, B)$ is given by a sum of terms

$$\text{(13)}\quad A_{q,\mu k} = \frac{a_{\mu,k,q}}{(2\pi i)^n} \frac{g \wedge \partial\beta \wedge (\overline{\partial_\zeta} g)^{\mu} \wedge (\overline{\partial_z} g)^{k} \wedge (\overline{\partial_\zeta}\partial\beta)^{n-q-2-\mu} \wedge (\overline{\partial_z}\partial\beta)^{q-k}}{\Phi^{1+\mu+k}\ \beta^{n-\mu-k-1}},$$

where $a_{\mu,k,q} \in \mathbb{Q}$, $0 \le \mu \le n-q-2$ and $0 \le k \le q$. As in the case of the kernel $\Omega_q(W^{\mathcal{L}})$, it follows that

$$A_{q,\mu k}(W^{\mathcal{L}}, B) = \frac{\mathcal{L}[\mu](\mathcal{E}_1^{\sharp})^k}{\Phi^{1+\mu+k}} \frac{\mathcal{E}_1}{\beta^{n-\mu-k-1}}.$$

Consequently the kernels $A_{q,\mu k}(W^{\mathcal{L}}, B)$ are admissible of boundary order ≥ 1. The integral $\int_{bD} \bar\partial f \wedge \Omega_q(W^{\mathcal{L}}, B)$ is therefore covered by Theorem 3(c); in particular, its kernel is smoothing of order $\delta < \frac{1}{3}$.

Next, one checks that

$$\overline{\partial_z} A_{q,\mu k}(W^{\mathcal{L}}, B) = \frac{\mathcal{L}[\mu](\mathcal{E}_1^{\sharp})^k}{\Phi^{1+\mu+k}} \frac{\mathcal{E}_0}{\beta^{n-\mu-k-1}} + + \frac{\mathcal{L}[\mu](\mathcal{E}_1^{\sharp})^k \mathcal{E}_2^{\sharp}}{\Phi^{1+\mu+k+1}} \frac{\mathcal{E}_1}{\beta^{n-\mu-k-1}} + \frac{\mathcal{L}[\mu](\mathcal{E}_1^{\sharp})^k}{\Phi^{1+\mu+k}} \frac{\mathcal{E}_2}{\beta^{n-\mu-k}}.$$

This shows that the kernels $\overline{\partial_z} A_{q,\mu k}(W^{\mathcal{L}}, B)$ are admissible, and one easily verifies that $\overline{\partial_z} A_{q,\mu k}$ is of boundary order ≥ 0.

Recall that $0 \le \mu \le n-q-2$ and $0 \le k \le q$, so that $0 \le \mu + k \le n-2$. We first consider the case $\mu + k \ge 1$, which occurs only when $n \ge 3$.

Lemma 5. *Suppose $\mu + k \ge 1$. Then $\overline{\partial_z} A_{q,\mu k}(W^{\mathcal{L}}, B)$ is a kernel of type $\Gamma_{0,1/2}^{\bar z}$.*

Proof. We saw that $\overline{\partial_z} A_{q,\mu k}(W^{\mathcal{L}}, B)$ is of order ≥ 0. Applying a derivative with respect to $\overline{z_j}$ to a factor $1/\Phi$ in any of the summands of $\overline{\partial_z} A_{q,\mu k}(W^{\mathcal{L}}, B)$ results in a term estimated by a kernel Γ_0 of order ≥ 0 multiplied by $|\mathcal{E}_2^{\sharp}/\Phi|$, and since $|\Phi| \gtrsim |r(z)|^{1-\delta}|\zeta - z|^{3\delta}$, the factor $|\mathcal{E}_2^{\sharp}/\Phi|$ will be bounded by $|\zeta - z|^{\alpha}|r(z)|^{-1+\delta}$ for some $\alpha > 0$ if $\delta < \frac{2}{3}$. By Theorem 3(a), the kernel $|\zeta - z|^{\alpha}|\Gamma_0|$ is integrable uniformly in z if $\alpha > 0$. For the other differentiations, note that since $\mu + k \ge 1$, each summand in $\overline{\partial_z} A_{q,\mu k}$ has at least one factor $\mathcal{L}[1]/\Phi$ or $\mathcal{E}_1^\sharp/\Phi$ with weight ≥ -2 in addition to the one factor $1/\Phi$ which is counted with weight ≥ -1 in the calculation of the order of $\overline{\partial_z} A_{q,\mu k}$. Differentiating the numerator of any such factor of weight ≥ -2 results in a term $\mathcal{E}_0/\Phi$, which can be estimated by $(1/|\zeta - z|^{3\delta})(1/|r(z)|^{1-\delta})$, where the factor $1/|\zeta - z|^{3\delta}$ is of weight > -2 for any $\delta < \frac{2}{3}$. If the differentiation is applied to any of the remaining factors in each of the summands, the order of the kernel decreases at most by one without affecting any of the factors $\mathcal{L}[1]/\Phi$ or $\mathcal{E}_1^\sharp/\Phi$. In order to compensate for this decrease, one must extract a factor $\mathcal{E}_1^\sharp$ from such a factor of weight -2, leaving a factor of weight -2 multiplied with a suitable power $|r(z)|^{-1+\delta}$. This follows as before for a factor $\mathcal{E}_1^\sharp/\Phi$, since $1/\Phi$ is estimated by $1/(|\zeta - z|^2|r(z)|^{1/3})$. For factors $\mathcal{L}[1]/\Phi$, note that according to (8) — after introducing z-diagonalizing coordinates — one may estimate $|\tan \bar{\partial}\partial r/\Phi|$ by terms of the type

$$\frac{1}{|r(z)| + |\zeta_l - z_l|^2} \lesssim \frac{1}{|\zeta_l - z_l|^{2\delta}|r(z)|^{1-\delta}} \lesssim \frac{1}{|\zeta_l - z_l|^{2\delta+1}|r(z)|^{1-\delta}} \mathcal{E}_1^\sharp$$

for a suitable $l \leq n-1$ (see [Range 2013] for details). Here the factor $1/|\zeta_l - z_l|^{2\delta+1}$ is of weight ≥ -2 for any $\delta < \frac{1}{2}$. Altogether we thus proved that $\overline{\partial_z} A_{q,\mu k}$ is $\bar{z}$-smoothing of any order $\delta < \frac{1}{2}$. Finally, if one considers a tangential derivative L_j^z, $j \leq n-1$, the same arguments apply as long as $\delta < \frac{1}{3}$. □

Remark. Note that this last argument restricts the order of $\bar{z}$-smoothing to $\delta < \frac{1}{2}$, while in all other previous instances one has the stronger estimates of order $\delta < \frac{2}{3}$.

6. The kernel $\overline{\partial_z} A_{q-1,00}$

We are thus left with $\overline{\partial_z} A_{q-1,00}$. This is the critical and most delicate case. Note that this kernel contains a term $\mathcal{E}_0/(\Phi\beta^{n-1})$ (of order ≥ 0); however, differentiation with respect to $\overline{z_j}$ will result, among others, in a term $\mathcal{E}_1/(\Phi\beta^n)$ which is estimated, at best, by $|\mathcal{E}_1/\beta^n||r(z)|^{-1}$. We see that $\overline{\partial_z} A_{q-1,00}$ contains terms which are not $\bar{z}$-smoothing of any positive order $\delta > 0$. In order to proceed we need to identify these critical terms and exploit certain approximate symmetries in analogy to the method introduced in [LR 1983].

We begin by applying Stokes' theorem to replace the integral

$$\int_{bD} f \wedge \overline{\partial_z} A_{q-1,00}$$

by an integral over D. For this purpose we first extend $A_{q-1,00}$ — that is, $W^{\mathcal{L}}$ and B — from the boundary into D without introducing any new singularities, as follows. By the estimate (4) for $|\Phi|$, as long as $-\varepsilon < r(\zeta) < 0$, one has $|\Phi| \gtrsim |r(\zeta)|$. Choose a C^∞ function φ on $\bar{D}$ so that $\varphi(\zeta) \equiv 1$ for $-\frac{1}{2}\varepsilon \leq r(\zeta)$ and $\varphi(\zeta) \equiv 0$ for

$r(\zeta) \le -\frac{3}{4}\varepsilon$. We then define the $(0,1)$-form

$$\widehat{W^{\mathcal{L}}}(\zeta,z) = \varphi(\zeta) W^{\mathcal{L}}(\zeta,z)$$

on $\bar D \times \bar D - \{(\zeta,\zeta) : \zeta \in bD\}$, so that $\widehat{W^{\mathcal{L}}}(\zeta,z) = W^{\mathcal{L}}(\zeta,z)$ for $\zeta \in bD$.

We also define

$$\widehat{B}(\zeta,z) = \frac{\partial_\zeta \beta}{\mathcal{P}(\zeta,z)}, \quad \text{where } \mathcal{P}(\zeta,z) = \beta(\zeta,z) + \frac{2r(\zeta)r(z)}{\|\partial r(\zeta)\|\|\partial r(z)\|}$$

on $\bar D \times \bar D - \{(\zeta,\zeta) : \zeta \in bD\}$. (Note that $r(\zeta)r(z) \ge 0$ for $\zeta, z \in D$.) Clearly $\widehat{B}(\zeta,z) = B(\zeta,z)$ for $\zeta \in bD$. By replacing $W^{\mathcal{L}}$ with $\widehat{W^{\mathcal{L}}}$ and B with $\widehat{B}$ in $A_{q-1,00}$, one can therefore assume that $A_{q-1,00}$ extends to $\bar D \times \bar D - \{(\zeta,\zeta) : \zeta \in bD\}$ without any singularities.

It then follows that

$$\int_{bD} f \wedge \overline{\partial_z} A_{q-1,00} = \int_D \overline{\partial_\zeta} f \wedge \overline{\partial_z} A_{q-1,00} + (-1)^q \int_D f \wedge \overline{\partial_\zeta}\,\overline{\partial_z} A_{q-1,00}.$$

Remark. When one considers kernels that are integrated over D, the definition of admissible kernels and of their weighted order needs to be modified appropriately. First of all, the dimension of the domain of integration is now $2n$, which leads to an increase in order by one. Also, each factor $r(z)$ or $r(\zeta)$ in the numerator increases the order by at least one. Furthermore, since both $r(\zeta)$ and $\operatorname{Im} F^{(r)}(\zeta,z)$ are used as coordinates in a neighborhood of z for $\operatorname{dist}(z,bD) < \varepsilon$, the weighted order is adjusted to account for the fact that by estimate (4), now up to *two* factors Φ are counted with weight ≥ -1 (see [LR 1983, Definition 4.2] for more details). In particular, it follows that the (extended) kernel $\overline{\partial_z} A_{q-1,00}$, which is admissible of boundary order ≥ 0, is admissible of order ≥ 1 over D.

It is straightforward to prove the analogous version of Theorem 3 for kernels integrated over D. In particular, one then obtains:

Lemma 6. *The operator $T_{q-1,00} : C_{(0,q+1)}(\bar D) \to C^1_{(0,q)}(D)$ defined by*

$$T_{q-1,00}(\psi) = \int_D \psi \wedge \overline{\partial_z} A_{q-1,00}$$

is smoothing of any order $\delta < \frac{1}{3}$ and $\bar z$-smoothing of any order $\delta < 1$.

Note that because of the term $K|\zeta - z|^3$ contained in Φ, the kernel $\overline{\partial_z} A_{q-1,00}$ is only of class C^1 jointly in (ζ,z) near points where $\zeta = z$.

We are left with estimating integrals of the kernel $\overline{\partial_\zeta}\,\overline{\partial_z} A_{q-1,00} = \overline{\partial_z}\,\overline{\partial_\zeta} A_{q-1,00}$ over D. This kernel is readily seen to be of order ≥ 0, but it contains terms that are not $\bar z$-smoothing of any order $\delta > 0$.

Proceeding as in [LR 1983], we introduce the kernel $\overline{L_{q-1}} = (-1)^q * A_{q-1,00}$ for $1 \le q \le n-1$, where $*$ is the Hodge operator acting on the variable ζ with respect to the standard inner product of forms in $\mathbb{C}^n$. Note that in [LR 1983] the definition of L_{q-1} involved $\Omega_{q-1}(\widehat{W^{\mathcal{L}}}, \widehat{B})$, while here we only take those summands $A_{q,\mu k}$ with $\mu + k = 0$. Since $A_{q-1,00} = -*\overline{L_{q-1}}$, one then obtains

$$(-1)^q \overline{\partial_\zeta} A_{q-1,00} = **\overline{\partial_\zeta} A_{q-1,00} = *(*\overline{\partial_\zeta})(-*\overline{L_{q-1}}) = *\overline{\vartheta_\zeta}\,\overline{L_{q-1}},$$

where $\vartheta_\zeta = -*\partial_\zeta *$ is the (formal) adjoint of $\bar\partial$. It follows that

$$(-1)^q \int_D f \wedge \bar\partial_\zeta A_{q-1,00} = \int_D f \wedge *\overline{\vartheta_\zeta}\,\overline{L_{q-1}} = (f, \vartheta_\zeta L_{q-1})_D,$$

where the inner product is taken by integrating the pointwise inner product of forms over D. Since $\overline{\partial_z}$ commutes with $*_\zeta$, one has

$$(-1)^q \int_D f \wedge \overline{\partial_z}\,\overline{\partial_\zeta} A_{q-1,00} = (f, \partial_z \vartheta_\zeta L_{q-1})_D.$$

Let us introduce the Hermitian transpose K^* of a double form $K = K(\zeta, z)$ by

$$K^*(\zeta, z) = \overline{K(z, \zeta)}.$$

Note that $K^*(\zeta, z)$ is the kernel of the adjoint T^* of $T : f \to (f, K(\cdot\,, z))_D$, i.e., $T^*(f) \to (f, K^*(\cdot\,, z))_D$.

One now writes

$$(f, \partial_z \vartheta_\zeta L_{q-1})_D = (f, \partial_z \vartheta_\zeta L_{q-1} - [\partial_z \vartheta_\zeta L_{q-1}]^*)_D + (f, [\partial_z \vartheta_\zeta L_{q-1}]^*)_D.$$

Since $[\partial_z \vartheta_\zeta L_{q-1}]^* = \overline{\partial_\zeta}\,\overline{\vartheta_z} L^*_{q-1}$, if $f \in \operatorname{dom} \bar\partial^*$, one may integrate by parts in the second inner product, resulting in $(f, \overline{\partial_\zeta}\,\overline{\vartheta_z} L^*_{q-1})_D = (\bar\partial^* f, \overline{\vartheta_z} L^*_{q-1})_D$.

Expanding the definition of admissible kernels to allow for factors Φ^*, with corresponding definition of order, one verifies that the kernel $\vartheta_z \overline{L^*_{q-1}}$ is admissible of order ≥ 1, and consequently it is smoothing of order $\delta < \frac{1}{3}$, and $\bar z$-smoothing of order $\delta < 1$.

7. The critical singularities

We now carefully examine $\partial_z \vartheta_\zeta L_{q-1} - [\partial_z \vartheta_\zeta L_{q-1}]^*$ and verify that there is a cancellation of critical terms, so that the conjugate of this kernel is partially smoothing as required for the Main Theorem.

We use the standard orthonormal frame $\omega_1, \ldots, \omega_n$ for $(1,0)$-forms on a neighborhood U of $P \in bD$, with $\omega_n = \partial r(\zeta)/\|\partial r(\zeta)\|$. After shrinking ε, we may assume that $B(P, \varepsilon) \subset U$. As usual, we shall focus on estimating integrals for fixed $z \in D$ with $|z - P| < \frac{1}{2}\varepsilon$, and integration over $\zeta \in D \cap U$ with $r(\zeta) \geq -\frac{1}{2}\varepsilon$ and $|\zeta - z| < \frac{1}{2}\varepsilon$. Let $L_1, \ldots, L_n$ be the corresponding dual frame of $(1,0)$-vectors, acting on the ζ variables. If $L = \sum_{k=1}^n a_k(\zeta)\partial/\partial\zeta_k$, we denote by $L^z = L^z_\zeta = \sum_{k=1}^n a_k(\zeta)\partial/\partial z_k$

the corresponding vector field acting on the z variables. A similar convention applies to ω_j^z, which we denote by θ_j.

One then readily verifies the following equations:

(a) $\bar\partial r(\zeta) = \overline{\omega_n}\|\partial r(\zeta)\|$;

(b) $\partial\beta = \sum_{j=1}^n (L_j\beta)\omega_j$ and $\bar\partial\beta = \sum_{j=1}^n (\overline{L_j}\beta)\overline{\omega_j}$;

(c) $\bar\partial\partial\beta = 2\sum_{j=1}^n \overline{\omega_j}\wedge\omega_j + \mathcal{E}_1$;

(d) $\overline{\partial_z}\partial\beta = -2\sum_{j=1}^n \overline{\theta_j}\wedge\omega_j + \mathcal{E}_1$;

(e) $L_j^z\beta = -L_j\beta + \mathcal{E}_2$ and $L_jL_k\beta = \mathcal{E}_1$;

(f) $L_j\mathcal{P} = \mathcal{E}_1$ and $L_j^z\mathcal{P} = \mathcal{E}_1$ for $j < n$.

Somewhat more delicate are the following two formulas. They are the analogs of [LR 1983, Lemmas 5.9 and 5.35], with the differences due to the fact that in the present paper the defining function is not normalized, as it is restricted to a special form so that its level surfaces remain pseudoconvex. The definition of the extension $\mathcal{P}$ has been modified accordingly. Since both formulas require *exact* identification of the leading terms, we include the details of the proofs.

Lemma 7. $L_n^z\mathcal{P} = -\dfrac{2}{\|\partial r(\zeta)\|}\overline{\Phi} + \mathcal{E}_0 r(\zeta)r(z) + \mathcal{E}_1 r(\zeta) + \mathcal{E}_2.$

Proof. We fix $\zeta\in U$. After a unitary change of coordinates in the ζ variables, one may assume that $\partial r/\partial\zeta_j(\zeta) = 0$ for $j<n$ and $\partial r/\partial\zeta_n(\zeta) > 0$, so that $\|\partial r(\zeta)\| = \sqrt{2}\partial r/\partial\zeta_n(\zeta)$ and $(L_n^z)_z = \sqrt{2}\partial/\partial z_n + \mathcal{E}_1$. Then $L_n^z r(z) = \sqrt{2}\partial r/\partial z_n(\zeta) + \mathcal{E}_1 = \|\partial r(\zeta)\| + \mathcal{E}_1$. In this coordinate system one has

$$\begin{aligned}\sqrt{2}\overline{\Phi}(\zeta,z) &= \sqrt{2}\frac{\partial r}{\partial\overline{\zeta_n}}(\zeta)(\overline{\zeta_n - z_n}) + \mathcal{E}_2 - \sqrt{2}r(\zeta)\\ &= \|\partial r(\zeta)\|(\overline{\zeta_n - z_n}) - \sqrt{2}r(\zeta) + \mathcal{E}_2,\end{aligned}$$

and

$$\begin{aligned}L_n^z P(\zeta,z) &= -\sqrt{2}(\overline{\zeta_n - z_n}) + \mathcal{E}_2 + \frac{2r(\zeta)}{\|\partial r(\zeta)\|}\frac{L_n^z r(z)}{\|\partial r(z)\|} + \mathcal{E}_0 r(\zeta)r(z)\\ &= -\sqrt{2}(\overline{\zeta_n - z_n}) + \mathcal{E}_2 + \frac{2r(\zeta)}{\|\partial r(\zeta)\|}[1+\mathcal{E}_1] + \mathcal{E}_0 r(\zeta)r(z)\\ &= -\sqrt{2}(\overline{\zeta_n - z_n}) + \frac{2r(\zeta)}{\|\partial r(\zeta)\|} + \mathcal{E}_2 + \mathcal{E}_1 r(\zeta) + \mathcal{E}_0 r(\zeta)r(z)\\ &= -\frac{\sqrt{2}}{\|\partial r(\zeta)\|}[\|\partial r(\zeta)\|(\overline{\zeta_n - z_n}) - \sqrt{2}r(\zeta)]\\ &\qquad + \mathcal{E}_2 + \mathcal{E}_1 r(\zeta) + \mathcal{E}_0 r(\zeta)r(z).\end{aligned}$$

The proof is completed by combining these two equations. ☐

Lemma 8. $2\mathcal{P}-\sum_{j=1}^{n-1}|L_j\beta|^2=\dfrac{4}{\|\partial r(\zeta)\|\|\partial r(z)\|}|\Phi|^2+\mathcal{E}_3+\mathcal{E}_2 r(\zeta).$

Proof. Here we fix z, and after a unitary coordinate change in ζ we may assume that $L_j=\sqrt{2}\partial/\partial\zeta_j+\mathcal{E}_1$, so that $L_j\beta=\sqrt{2}(\overline{\zeta_j-z_j})+\mathcal{E}_2$. Hence $\sum_{j=1}^{n-1}|L_j\beta|^2=2\sum_{j=1}^{n-1}|\zeta_j-z_j|^2+\mathcal{E}_3$, and therefore

$$2\mathcal{P}-\sum_{j=1}^{n-1}|L_j\beta|^2=2|\zeta_n-z_n|^2+\frac{4r(\zeta)r(z)}{\|\partial r(\zeta)\|\|\partial r(z)\|}+\mathcal{E}_3. \tag{14}$$

Furthermore,

$$\begin{aligned}\sqrt{2}\overline{\Phi}(\zeta,z)&=\sqrt{2}\frac{\partial r}{\partial\overline{\zeta_n}}(z)(\overline{\zeta_n-z_n})+\mathcal{E}_2-\sqrt{2}r(\zeta)\\&=\|\partial r(z)\|(\overline{\zeta_n-z_n})-\sqrt{2}r(\zeta)+\mathcal{E}_2,\end{aligned}$$

It follows that

$$\begin{aligned}&2|\Phi(\zeta,z)|^2\\&\quad=\|\partial r(z)\|(\overline{\zeta_n-z_n})\sqrt{2}\Phi-2r(\zeta)\Phi+\mathcal{E}_2\Phi\\&\quad=\|\partial r(z)\|^2|\zeta_n-z_n|^2+\|\partial r(z)\|(\overline{\zeta_n-z_n})[-\sqrt{2}r(\zeta)+\mathcal{E}_2]-2r(\zeta)\Phi+\mathcal{E}_2\Phi\\&\quad=\|\partial r(z)\|^2|\zeta_n-z_n|^2-\sqrt{2}r(\zeta)[\|\partial r(z)\|(\overline{\zeta_n-z_n})+\sqrt{2}\Phi]+\mathcal{E}_3+\mathcal{E}_2\Phi.\end{aligned}$$

By (5) one has

$$\begin{aligned}\sqrt{2}\Phi=\sqrt{2}\Phi^*+\mathcal{E}_3&=\|\partial r(\zeta)\|(\overline{z_n-\zeta_n})-\sqrt{2}r(z)+\mathcal{E}_2\\&=-\|\partial r(z)\|(\overline{\zeta_n-z_n})-\sqrt{2}r(z)+\mathcal{E}_2,\end{aligned}$$

where we used that $\|\partial r(\zeta)\|=\|\partial r(z)\|+\mathcal{E}_1$. Inserting this equation into the previous one and using $\mathcal{E}_2\Phi=\mathcal{E}_3+\mathcal{E}_2 r(\zeta)$, results in

$$\begin{aligned}2|\Phi(\zeta,z)|^2&=\|\partial r(z)\|^2|\zeta_n-z_n|^2-\sqrt{2}r(\zeta)(-\sqrt{2}r(z)+\mathcal{E}_2)+\mathcal{E}_3+\mathcal{E}_2 r(\zeta)\\&=\|\partial r(\zeta)\|\|\partial r(z)\||\zeta_n-z_n|^2+2r(\zeta)r(z)+\mathcal{E}_3+\mathcal{E}_2 r(\zeta)\\&=\tfrac{1}{2}\|\partial r(\zeta)\|\|\partial r(z)\|\left[2|\zeta_n-z_n|^2+\frac{4r(\zeta)r(z)}{\|\partial r(\zeta)\|\|\partial r(z)\|}\right]+\mathcal{E}_3+\mathcal{E}_2 r(\zeta).\end{aligned}$$

The lemma follows after inserting (14) and rearranging. □

We now identify precisely the kernel L_{q-1} and the critical highest order singularity of $\partial_z\vartheta_\zeta L_{q-1}$ with respect to the standard frames introduced above. To simplify notation we replace $q-1$ with q and consider $A_{q,00}$ and L_q for $0\le q\le n-2$. The computations follow closely those for the case $\mu=0$ in [LR 1983]; therefore we just state the relevant formulas, and provide more details only where critical differences arise.

From (13) one sees that

$$A_{q,00} = \frac{a_q}{(2\pi i)^n}\frac{g\wedge\partial\beta\wedge(\overline{\partial_\zeta}\partial\beta)^{n-q-2}\wedge(\overline{\partial_z}\partial\beta)^q}{\Phi\mathcal{P}^{n-1}}$$
$$= \frac{a_q}{(2\pi i)^n}\frac{\partial r\wedge\partial\beta\wedge(\overline{\partial_\zeta}\partial\beta)^{n-q-2}\wedge(\overline{\partial_z}\partial\beta)^q+\mathcal{E}_2}{\Phi\mathcal{P}^{n-1}}.$$

Then

$$\overline{L_q} = (-1)^{q+1} * A_{q,00} = \overline{C_q} + \frac{\mathcal{E}_2}{\Phi\mathcal{P}^{n-1}},$$

where

$$\overline{C_q} = (-1)^{q+1}\frac{a_q}{(2\pi i)^n} *_\zeta \frac{\partial r\wedge\partial\beta\wedge(\overline{\partial_\zeta}\partial\beta)^{n-q-2}\wedge(\overline{\partial_z}\partial\beta)^q}{\Phi\mathcal{P}^{n-1}}.$$

It follows that

$$\overline{\partial_z}\overline{\vartheta_\zeta}\overline{L_q} = \overline{\partial_z}\overline{\vartheta_\zeta}\overline{C_q} + \overline{\partial_z}\overline{\vartheta_\zeta}\frac{\mathcal{E}_2}{\Phi\mathcal{P}^{n-1}}.$$

Remark. Note that in contrast to [LR 1983], the kernels $L_q = C_q + \mathcal{E}_2/(\bar\Phi\mathcal{P}^{n-1})$ and $\partial_z\vartheta_\zeta L_q$ analyzed here only involve the term corresponding to $\mu = 0$ in the same reference. Since in this paper we are concerned with $\bar z$-smoothing, we need to consider the conjugates $\overline{L_q}$, $\overline{C_q}$, and $\overline{\partial_z}\overline{\vartheta_\zeta}\overline{L_q}$, which are the kernels that appear in the integral $\int f\wedge *\overline{\partial_z}\overline{\vartheta_\zeta}\overline{L_q} = (f, \partial_z\vartheta_\zeta L_q)_D$.

Lemma 9. $\overline{\partial_z}\overline{\vartheta_\zeta}\dfrac{\mathcal{E}_2}{\Phi\mathcal{P}^{n-1}}$ *is of type* Γ_1.

Proof. The proof of this lemma involves a straightforward verification. Note that $\mathcal{E}_2/(\Phi\mathcal{P}^{n-1})$ is of order ≥ 3. Differentiation with respect to ζ reduces the order by one only, since after differentiating $1/\Phi$ the resulting factor $1/\Phi^2$ has weight -2. Similarly, since $\overline{\partial_z}\Phi = \mathcal{E}_2^\sharp$, subsequent application of $\overline{\partial_z}$ also reduces the order by one only. □

Next we represent $\overline{C_q}$ in terms of the local orthonormal frames. By utilizing the various formulas recalled above, it follows that

$$\overline{C_q} = \frac{\gamma_q\|\partial r(\zeta)\|}{i^n\Phi\mathcal{P}^{n-1}} *_\zeta \sum_{\substack{|Q|=q\\ j<n}} \omega_n\wedge(L_j\beta)\omega_j\wedge(\bar\omega\wedge\omega)^J\wedge\omega^Q\wedge\bar\theta^Q + \frac{\mathcal{E}_2}{\Phi\mathcal{P}^{n-1}},$$

where γ_q is a *real* constant. The summation is over all strictly increasing q-tuples Q with $n\notin Q$, over $j<n$ with $j\notin Q$, and J is the ordered $(n-q-2)$-tuple complementary to njQ in $\{1,\dots,n\}$. Since $*[\omega^{njQ}\wedge(\bar\omega\wedge\omega)^J] = b_q i^n\omega^{njQ}$, where b_q is real, it follows that

$$\overline{C_q} = \widetilde{\gamma_q}\|\partial r(\zeta)\| \sum_{\substack{|Q|=q\\ j<n}} \frac{L_j\beta}{\Phi\mathcal{P}^{n-1}}\omega^{njQ}\wedge\bar\theta^Q + \frac{\mathcal{E}_2}{\Phi\mathcal{P}^{n-1}},$$

for another *real* constant $\widetilde{\gamma_q}$. By using Lemma 9, it then follows that

$$\overline{\partial_z}\,\overline{\vartheta_\zeta}\,\overline{C_q} = \widetilde{\gamma_q}\|\partial r(\zeta)\|\overline{\partial_z}\,\overline{\vartheta_\zeta}\Bigg[\sum_{\substack{|Q|=q\\ j<n}} \frac{L_j\beta}{\Phi\mathcal{P}^{n-1}}\omega^{njQ}\wedge\bar{\theta}^Q\Bigg]+\Gamma_1.$$

Let us introduce

$$\overline{C_q^0} = \sum_{\substack{|Q|=q\\ j<n}} \frac{L_j\beta}{\Phi\mathcal{P}^{n-1}}\omega^{njQ}\wedge\bar{\theta}^Q.$$

The heart of the analysis of $\partial_z\vartheta_\zeta L_q - [\partial_z\vartheta_\zeta L_q]^*$ is contained in:

Theorem 10. *For $0\le q\le n-2$ the kernel*

$$\overline{\partial_z}\,\overline{\vartheta_\zeta}\,\overline{C_q^0} - [\overline{\partial_z}\,\overline{\vartheta_\zeta}\,\overline{C_q^0}]^*$$

is of type $\Gamma^{\bar{z}}_{0,2/3}$.

Corollary 11. *The operator*

$$f\to(f,\partial_z\vartheta_\zeta L_{q-1}-[\partial_z\vartheta_\zeta L_{q-1}]^*)_D$$

is $\bar{z}$-smoothing of order $\delta<\frac{2}{3}$ and tangentially smoothing of order $\delta<\frac{1}{3}$ for $1\le q\le n-1$.

Proof. This follows from Theorem 10 by using Lemma 9 and also by observing that the differentiation of $\|\partial r(\zeta)\|$ results in an error term of type Γ_1. Similarly, when considering the difference $[\cdots]-[\cdots]^*$, the substitution $\|\partial r(\zeta)\| = \|\partial r(z)\|+\mathcal{E}_1$ leads to an error term of the same type. $\square$

Remark. In order to be consistent with the notation and formulas in [LR 1983], in the proof of Theorem 10 we analyze $\Delta_q = \partial_z\vartheta_\zeta C_q^0-[\partial_z\vartheta_\zeta C_q^0]^*$. In the end we must verify that its *conjugate* $\overline{\Delta_q}$ is of type $\Gamma^{\bar{z}}_{0,2/3}$.

Since C_q^0 is a double form of type $(0,q+2)$ in ζ and type $(q,0)$ in z, the form $\partial_z\vartheta_\zeta C_q^0$ is of type $(0,q+1)$ in ζ and type $(q+1,0)$ in z. Consequently,

$$\partial_z\vartheta_\zeta C_q^0 = \vartheta_\zeta\partial_z C_q^0 = \sum_{|L|=q+1}\Bigg(\sum_{|K|=q+1} A_{KL}\bar{\omega}^K\Bigg)\wedge\theta^L,$$

where the sums are taken over all strictly increasing $(q+1)$-tuples L and K. It follows that

$$\partial_z\vartheta_\zeta C_q^0-[\partial_z\vartheta_\zeta C_q^0]^* = \sum_{|L|=q+1}\Bigg(\sum_{|K|=q+1}[A_{KL}-(A_{LK})^*]\bar{\omega}^K\Bigg)\wedge\theta^L.$$

In the next section we identify the coefficients A_{KL} precisely in order to verify that $[A_{KL}-(A_{LK})^*]$ is of order ≥ 0 and that its *conjugate* is at least of type $\Gamma^{\bar{z}}_{0,2/3}$.

8. The approximate symmetries

The computation of $\vartheta_\zeta \partial_z C_q^0$ uses the expressions for ∂_z and ϑ_ζ in terms of the standard adapted boundary frames plus error terms which do not involve differentiation. These error terms — in the end — result in kernels which are conjugates of admissible kernels of order ≥ 1, and hence will be ignored in the discussion that follows.

As usual ε^L_{lQ} denotes the sign of the permutation which carries the ordered $(q+1)$-tuple lQ into the ordered $(q+1)$-tuple L if $lQ = L$ as sets, and $\varepsilon^L_{lQ} = 0$ otherwise. We introduce $m_{lj} = (1/\overline{\Phi}) L^z_l (\overline{L_j}\beta/\mathcal{P}^{n-1})$ for $1 \leq j < n$ and $1 \leq l \leq n$.

Lemma 12. *For any K, L one has*

$$A_{KL} = -\sum_{\substack{Q\\ j,l,k}} \varepsilon^L_{lQ}\varepsilon^{njQ}_{kK}(L_k m_{lj}) + \overline{\Gamma^{\bar z}_{0,2/3}} + \Gamma_1.$$

Proof. This lemma is the analogue of [LR 1983, Lemma 5.6]. In the present case, the computation shows that the error term is of the form $\mathcal{E}_3/[\overline{\Phi}^3\mathcal{P}^{n-1}] + \Gamma_1$, where the first term is only of order zero. (In the case where D is strictly pseudoconvex, and hence $|\Phi| \gtrsim |\zeta - z|^2$, this term is Γ_1 as well.) However, one readily checks that its conjugate $\mathcal{E}_3/[\Phi^3\mathcal{P}^{n-1}]$ is of type $\Gamma^{\bar z}_{0,2/3}$. $\square$

We are therefore left with

$$A^{(0)}_{KL} = -\sum_{\substack{Q\\ j,l,k}} \varepsilon^L_{lQ}\varepsilon^{njQ}_{kK}(L_k m_{lj}). \tag{15}$$

Only terms with $j < n$ appear with nonzero coefficients. In the following it will be assumed that $j < n$.

For $l < n$ one has

$$m_{lj} = \frac{-2\delta_{lj}}{\overline{\Phi}\mathcal{P}^{n-1}} - (n-1)\frac{L^z_l\beta\overline{L_j}\beta}{\overline{\Phi}\mathcal{P}^n} + \frac{\mathcal{E}_1}{\overline{\Phi}\mathcal{P}^{n-1}}, \tag{16}$$

while by using Lemma 7 one obtains

$$\begin{aligned} m_{nj} &= \frac{1}{\overline{\Phi}}\left[\frac{L^z_n\overline{L_j}\beta}{\mathcal{P}^{n-1}} - (n-1)\frac{(L^z_n\mathcal{P})\overline{L_j}\beta}{\mathcal{P}^n}\right] \\ &= \frac{2(n-1)}{\|\partial r(\zeta)\|}\frac{\overline{L_j}\beta}{\mathcal{P}^n} + \frac{\mathcal{E}_1}{\overline{\Phi}\mathcal{P}^{n-1}} + \frac{\mathcal{E}_1 r(\zeta)r(z) + \mathcal{E}_2 r(\zeta) + \mathcal{E}_3}{\overline{\Phi}\mathcal{P}^n}. \end{aligned} \tag{17}$$

Note that all the error terms are of order ≥ 2, and that they have only *one* factor Φ in the denominator. Consequently, applying L_k to the error terms results in Γ_1 terms. So only the leading terms of m_{lj} identified above need to be considered in the following analysis.

Proof of Theorem 10. Since all relevant error terms are of type $\Gamma^{\bar{z}}_{0,2/3}$ or better, it is enough to examine

$$\overline{A^{(0)}_{KL}} - \overline{A^{(0)*}_{LK}}, \quad \text{where } \overline{A^{(0)}_{KL}} = -\sum_{\substack{Q\\ j,l,k}} \varepsilon^{lQ}_{L}\,\varepsilon^{njQ}_{kK}\,(\overline{L_k m_{lj}}).$$

We need to consider separate cases, depending on whether n is in K (resp. L), or not:

Case 1. $n \in K$ and $n \in L$.

In this case the computations in [LR 1983] apply without further changes, subject to the adjustments due to the fact that the defining function r is not normalized. Combined with $\|\partial r(\zeta)\| = \|\partial r(z)\| + \mathcal{E}_1$, it follows that

$$\overline{A^{(0)}_{KL} - A^{(0)*}_{LK}} = \Gamma_1.$$

Case 2. $n \notin K$ and $n \notin L$.

In this case $\varepsilon^{njQ}_{kK} \neq 0$ only for $k = n$. Hence

$$A^{(0)}_{KL} = -\sum_{\substack{Q\\ j,l}} \varepsilon^{L}_{lQ}\varepsilon^{njQ}_{nK}(L_n m_{lj}) + \Gamma_1 = -\sum_{\substack{l\in L\\ j\in K}} \varepsilon^{L}_{lQ}\varepsilon^{jQ}_{K}(L_n m_{lj}) + \Gamma_1.$$

Lemma 5.21 in [LR 1983] needs to be replaced by:

Lemma 13. $L_n m_{lj} - (L_n m_{jl})^* = \overline{\Gamma^{\bar{z}}_{0,2/3}}$ *for* $j, l < n$.

Assuming the lemma, one obtains (after replacing j with l in the last equation)

$$\begin{aligned} A^{(0)*}_{LK} &= -\sum_{j,l<n} \varepsilon^{K}_{lQ}\varepsilon^{jQ}_{L}(L_n m_{lj})^* + \Gamma_1 = -\sum_{\substack{j\in L\\ l\in K}} \varepsilon^{K}_{lQ}\varepsilon^{jQ}_{L}(L_n m_{jl}) + \overline{\Gamma^{\bar{z}}_{0,2/3}} \\ &= A^{(0)}_{KL} + \overline{\Gamma^{\bar{z}}_{2/3}}. \end{aligned}$$

To prove the lemma, one uses (16). The calculation of $L_n m_{lj}$ proceeds as in [LR 1983] with the obvious changes. By using $\|\partial r(\zeta)\| = \|\partial r(z)\| + \mathcal{E}_1$ one obtains

$$\begin{aligned} &(L_n m_{jl})^* - L_n m_{lj} \\ &= -\frac{2\delta_{lj}\|\partial r(\zeta)\|}{\mathcal{P}^{n-1}}\left[\frac{1}{\overline{\overline{\Phi^{*2}}}} - \frac{1}{\overline{\overline{\Phi^2}}}\right] - 4\frac{\delta_{lj}(n-1)}{\|\partial r(\zeta)\|\mathcal{P}^n}\left[\frac{\Phi^*}{\overline{\overline{\Phi^*}}} - \frac{\Phi}{\overline{\overline{\Phi}}}\right] \\ &\quad -(n-1)\frac{L^z_j\beta\overline{L_l\beta}\|\partial r(\zeta)\|}{\mathcal{P}^n}\left[\frac{1}{\overline{\overline{\Phi^{*2}}}} - \frac{1}{\overline{\overline{\Phi^2}}}\right] - \frac{2n(n-1)}{\|\partial r(\zeta)\|}\frac{L^z_j\beta\overline{L_l\beta}}{\mathcal{P}^{n+1}}\left[\frac{\Phi^*}{\overline{\overline{\Phi^*}}} - \frac{\Phi}{\overline{\overline{\Phi}}}\right] + \Gamma_1. \end{aligned}$$

In the strictly pseudoconvex case the differences in $[\cdots]$ are of higher order than the terms individually, resulting in $(L_n m_{jl})^* - L_n m_{lj} = \Gamma_1$. In the present case, only a weaker result holds, as follows. Note that — after taking conjugates — one has

$$\frac{1}{\Phi^{*2}} - \frac{1}{\Phi^2} = \frac{(\Phi - \Phi^*)(\Phi + \Phi^*)}{\Phi^{*2}\Phi^2} = \frac{\mathcal{E}_3}{\Phi^{*2}\Phi} + \frac{\mathcal{E}_3}{\Phi^*\Phi^2},$$

where we have used the approximate symmetry (5) in the second equation. It now readily follows that $\mathcal{E}_3/[\mathcal{P}^{n-1}\Phi^{*2}\Phi]$ and $\mathcal{E}_3/[\mathcal{P}^{n-1}\Phi^{*}\Phi^2]$ (while of order zero, and hence not smoothing as in the strictly pseudoconvex case) are in fact of type $\Gamma^{\bar z}_{0,2/3}$. Since $\mathcal{E}_2/\mathcal{P}^n$ is estimated by $\mathcal{E}_0/\mathcal{P}^{n-1}$, the same argument works for the third term above. For the conjugate of the second term, note that

$$\frac{1}{\mathcal{P}^n}\left[\frac{\overline{\Phi^*}}{\Phi^*}-\frac{\overline{\Phi}}{\Phi}\right]=\frac{1}{\mathcal{P}^n}\left[\frac{\Phi\overline{\Phi^*}-\Phi^*\overline{\Phi}}{\Phi^*\Phi}\right]=\frac{1}{\mathcal{P}^n}\frac{\mathcal{E}_3}{\Phi^*\Phi}=\Gamma_1.$$

The fourth term is estimated the same way by first estimating $\mathcal{E}_2/\mathcal{P}^{n+1}$ by $\mathcal{E}_0/\mathcal{P}^n$.

Case 3. The mixed case $n\in K$ and $n\notin L$.

As in [LR 1983], this is — computationally — the most complicated case. On the other hand, aside from the differences as noted, for example, in Lemmas 7 and 8, the details of the proof essentially carry over from [LR 1983] to the case considered here, with the result that one has

$$\overline{A^{(0)}_{KL}}-\overline{A^{(0)*}_{LK}}=\Gamma_1.$$

In more detail, since $n\in K$, there is exactly one ordered q-tuple J such that $K=J\cup\{n\}$, and one then has $\varepsilon^K_{nJ}A^{(0)}_{KL}=A^{(0)}_{(nJ)L}$.

Note that we need to identify the leading terms of both $\overline{A^{(0)}_{KL}}$ and $\overline{A^{(0)}_{LK}}$. Let us first consider the simpler term $\overline{A^{(0)}_{LK}}$. After interchanging L and K in (15), one has

$$A^{(0)}_{LK}=-\sum_{\substack{Q\\ j,l,k}}\varepsilon^K_{lQ}\varepsilon^{njQ}_{kL}(L_k m_{lj})+\Gamma_1.$$

Since $n\notin L$, the factor $\varepsilon^K_{lQ}\varepsilon^{njQ}_{kL}\neq 0$ only if $k=n$ and $l=n$, and furthermore $Q=J$. Therefore the leading term of $A^{(0)}_{LK}$, i.e., the sum, is different from zero only if $J\subset L$ so that $\varepsilon^{jJ}_L\neq 0$ only for that unique j for which $L=J\cup\{j\}$. It follows that for $j<n$ one has

$$A^{(0)}_{(jJ)(nJ)}=\varepsilon^{jJ}_L\,\varepsilon^K_{nJ}A^{(0)}_{LK}+\Gamma_1=-L_n m_{nj}+\Gamma_1.$$

Since $L_n\mathcal{P}=(\overline{L^z_n\mathcal{P}})^*$, Lemma 7 implies that

$$L_n\mathcal{P}=-\frac{2}{\|\partial r(z)\|}\Phi^*+\mathcal{E}_0 r(\zeta)r(z)+\mathcal{E}_1 r(z)+\mathcal{E}_2.$$

By using this equation and (17), it follows that

$$A^{(0)}_{(jJ)(nJ)}=-\frac{4n(n-1)}{\|\partial r(\zeta)\|\|\partial r(z)\|}\frac{\overline{L_j\beta}\Phi^*}{\mathcal{P}^{n+1}}+\Gamma_1. \tag{18}$$

Finally we calculate $\overline{A^{(0)}_{KL}}$. With J as before, one has

$$A^{(0)}_{KL} = -\sum_{\substack{Q\\ j,l,k}} \varepsilon^{L}_{lQ}\varepsilon^{njQ}_{kK}(L_k m_{lj}) + \Gamma_1 = \varepsilon^{nJ}_{K} \sum_{\substack{n\notin Q\\ j,l,k<n}} \varepsilon^{L}_{lQ}\varepsilon^{jQ}_{kJ}(L_k m_{lj}) + \Gamma_1.$$

Continuing with the intricate calculations as in [LR 1983, pp. 237–239, Case Id)] and using Lemma 8 above in place of [LR 1983, Lemma 5.35], one obtains

$$A^{(0)}_{KL} = -\varepsilon^{nJ}_{K} \frac{4n(n-1)}{\|\partial r(\zeta)\|\|\partial r(z)\|} \sum_{l<n} \varepsilon^{L}_{lJ} \frac{(L^z_l\beta)\Phi}{\mathcal{P}^{n+1}} + \Gamma_1.$$

Here the only nonzero term in the sum arises for that unique l, for which $L = J\cup\{l\}$. Consequently, the last formula implies

$$A^{(0)}_{(nJ)(lJ)} = -\frac{4n(n-1)}{\|\partial r(\zeta)\|\|\partial r(z)\|} \frac{(L^z_l\beta)\Phi}{\mathcal{P}^{n+1}} + \Gamma_1. \tag{19}$$

Let us now consider $\overline{A^{(0)}_{KL}} - \overline{A^{(0)*}_{LK}}$. As the preceding formulas show, each summand is of type Γ_1 except in the case that for the unique q-tuple $J\subset\{1,2,\dots,n-1\}$ with $K = J\cup\{n\}$, L satisfies $L = J\cup\{l\}$ as sets for some unique $l<n$. In this latter case, equations (19) and (18) imply

$$\overline{A^{(0)}_{(nJ)(lJ)}} - \overline{A^{(0)*}_{(lJ)(nJ)}} = -\frac{4n(n-1)}{\|\partial r(\zeta)\|\|\partial r(z)\|} \frac{1}{\mathcal{P}^{n+1}} \overline{[(L^z_l\beta)\Phi - (\overline{L_l\beta})^*\Phi]} + \Gamma_1 = \Gamma_1.$$

The last equation holds because $(\overline{L_l\beta})^* = L^z_l\beta$.

Case 4. $n\notin K$ and $n\in L$.

This is reduced to Case 3 by noting that

$$\overline{A^{(0)}_{KL}} - \overline{A^{(0)*}_{LK}} = -\big(\overline{A^{(0)}_{LK}} - \overline{A^{(0)*}_{KL}}\big)^*.$$

It thus follows that for all K and L one has $\overline{A^{(0)}_{KL}} - \overline{A^{(0)*}_{LK}} = \Gamma^{\bar z}_{0,2/3}$. This completes the proof of Theorem 10. □

9. Proof of the Main Theorem

In Sections 4–8, we have analyzed the integrals that appear in the representation (3) of the boundary operator S^{bD}. By combining these results, it follows that for all $f\in\mathcal{D}^1_{qU}$ and $z\in D\cap U$ one has the estimates

$$|\overline{L^z_j} S^{bD}(f)(z)| \le C_\delta \operatorname{dist}(z,bD)^{\delta-1} Q_0(f) \quad \text{for } j=1,\dots,n \text{ and any } \delta<\tfrac12,$$
$$|L^z_j S^{bD}(f)(z)| \le C_\delta \operatorname{dist}(z,bD)^{\delta-1} Q_0(f) \quad \text{for } j=1,\dots,n-1 \text{ and any } \delta<\tfrac13.$$

We have thus completed the proof of claims (ii) and (iii) of the Main Theorem. As noted earlier, (i) follows trivially from the classical estimate (2) for S^{iso}.

Finally we prove the statement about the normal components of f. We use the following lemma, which is a routine variation of classical estimates for the BMK kernel. For $0 \le \alpha < 1$, set $C^{-\alpha}_{(0,1)}(D) = \{g \in C_{(0,1)}(D) : \sup_{z\in D} |g(z)| \operatorname{dist}(z, bD)^\alpha < \infty\}$, with the norm $|g|_{-\alpha}$ defined by the relevant supremum.

Lemma 14. *The operator $T^{BM} : C^{-\alpha}_{(0,1)}(D) \to C(D)$ defined by*

$$T^{BM}(g) = \int_D g(\zeta) \wedge \Omega_0(B)$$

satisfies the estimate

$$\left|T^{BM}(g)\right|_{1-\alpha'} \lesssim |g|_{-\alpha} \quad \textit{for any } \alpha' > \alpha. \tag{20}$$

Now suppose $f \in \mathcal{D}^1_{qU}$ and let f_J be a normal component of f, so that $f_J\,|_{bD} = 0$. Decompose $f_J = h_J + [S^{\text{iso}}(f)]_J$, where $h = S^{bD}(f)$. We already know by estimate (2) that $|S^{\text{iso}}(f)|_\alpha \lesssim Q_0(f)$ for any $\alpha < 1$. Note that on $bD \cap U$ one has $h_J = -[S^{\text{iso}}(f)]_J$, so that $|(h_J\,|_{bD\cap U})|_\alpha \lesssim |S^{\text{iso}}(f)|_\alpha \lesssim Q_0(f)$ as well. By standard properties of the BM kernel $\Omega_0(B)$, it follows that $\int_{bD} h_J \Omega_0(B)$ satisfies the same estimate on $\bar D \cap U$ if $\alpha > 0$. By the case $q = 0$ of the BMK representation formula (1) applied to h_J, one has

$$h_J = \int_{bD} h_J\, \Omega_0(B) - \int_D \bar\partial h_J \wedge \Omega_0(B). \tag{21}$$

Given $\delta < \frac{1}{2}$, choose δ' with $\delta < \delta' < \frac{1}{2}$. By part (ii) of the Main Theorem, $\bar\partial h_J \in C^{-(1-\delta')}_{(0,1)}(D)$, with $|\bar\partial h_J|_{-(1-\delta')} \le C_{\delta'} Q_0(f)$. It then follows from Lemma 14 that

$$\left|T^{BM}(\bar\partial h_J)\right|_\delta \lesssim Q_0(f).$$

Each summand in the representation (21) therefore satisfies the desired Hölder estimate, so that

$$|h_J|_{\Lambda^\delta(\bar D\cap U)} \lesssim Q_0(f).$$

Since $f_J = h_J + [S^{\text{iso}}(f)]_J$, the required estimate $|f_J|_\delta \lesssim Q_0(f)$ holds as well. □

References

[Catlin 1987] D. Catlin, "Subelliptic estimates for the $\bar\partial$-Neumann problem on pseudoconvex domains", *Ann. of Math.* (2) **126**:1 (1987), 131–191. MR 88i:32025 Zbl 0627.32013

[Cumenge 1997] A. Cumenge, "Estimées Lipschitz optimales dans les convexes de type fini", *C. R. Acad. Sci. Paris Sér. I Math.* **325**:10 (1997), 1077–1080. MR 99i:32002 Zbl 0904.32013

[D'Angelo 1982] J. P. D'Angelo, "Real hypersurfaces, orders of contact, and applications", *Ann. of Math.* (2) **115**:3 (1982), 615–637. MR 84a:32027 Zbl 0488.32008

[Diederich et al. 1999] K. Diederich, B. Fischer, and J. E. Fornæss, "Hölder estimates on convex domains of finite type", *Math. Z.* **232**:1 (1999), 43–61. MR 2001f:32065 Zbl 0932.32008

[Folland and Kohn 1972] G. B. Folland and J. J. Kohn, *The Neumann problem for the Cauchy–Riemann complex*, Annals of Mathematics Studies **75**, Princeton University Press, 1972. MR 57 #1573 Zbl 0247.35093

[Kohn 1979] J. J. Kohn, "Subellipticity of the $\bar{\partial}$-Neumann problem on pseudo-convex domains: sufficient conditions", *Acta Math.* **142**:1-2 (1979), 79–122. MR 80d:32020 Zbl 0395.35069

[Kohn and Nirenberg 1973] J. J. Kohn and L. Nirenberg, "A pseudo-convex domain not admitting a holomorphic support function", *Math. Ann.* **201** (1973), 265–268. MR 48 #8850 Zbl 0248.32013

[Lieb and Range 1983] I. Lieb and R. M. Range, "On integral representations and a priori Lipschitz estimates for the canonical solution of the $\bar{\partial}$-equation", *Math. Ann.* **265**:2 (1983), 221–251. MR 85f:32033 Zbl 0504.32015

[Lieb and Range 1986] I. Lieb and R. M. Range, "Integral representations and estimates in the theory of the $\bar{\partial}$-Neumann problem", *Ann. of Math.* (2) **123**:2 (1986), 265–301. MR 88j:32022a Zbl 0589.32034

[Range 1986] R. M. Range, *Holomorphic functions and integral representations in several complex variables*, Graduate Texts in Mathematics **108**, Springer, New York, 1986. Corr. 2nd printing, 1998. MR 87i:32001 Zbl 0591.32002

[Range 2011] R. M. Range, "A pointwise basic estimate and Hölder multipliers for the $\bar{\partial}$-Neumann problem on pseudoconvex domains", preliminary report, 2011. arXiv 1106.3132v1

[Range 2013] R. M. Range, "An integral kernel for weakly pseudoconvex domains", *Math. Ann.* **356**:2 (2013), 793–808. MR 3048615 Zbl 1293.32006

[Siu 2010] Y.-T. Siu, "Effective termination of Kohn's algorithm for subelliptic multipliers", *Pure Appl. Math. Q.* **6**:4, Part 2 (2010), 1169–1241. MR 2012e:32063 Zbl 1235.32030

Received April 7, 2014.

R. Michael Range
Department of Mathematics
State University of New York at Albany
1400 Washington Avenue
Albany, NY 12222
United States

range@math.albany.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 275, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.275.433

EXPLICIT HILBERT–KUNZ FUNCTIONS OF 2 × 2 DETERMINANTAL RINGS

MARCUS ROBINSON AND IRENA SWANSON

Let $k[X] = k[x_{i,j} : i = 1, \ldots, m;\ j = 1, \ldots, n]$ be the polynomial ring in mn variables $x_{i,j}$ over a field k of arbitrary characteristic. Denote by $I_2(X)$ the ideal generated by the 2×2 minors of the generic $m \times n$ matrix $[x_{i,j}]$. We give a closed polynomial formulation for the dimensions of the k-vector space $k[X]/(I_2(X) + (x_{1,1}^q, \ldots, x_{m,n}^q))$ as q varies over all positive integers, i.e., we give a closed polynomial form for the generalized Hilbert–Kunz function of the determinantal ring $k[X]/I_2(X)$. We also give a closed formulation of dimensions of other related quotients of $k[X]/I_2(X)$. In the process we establish a formula for the numbers of some compositions (ordered partitions of integers), and we give a proof of a new binomial identity.

1. Introduction

Throughout, let m, n, q be nonnegative integers, and let k, $k[X]$, and $I_2(X)$ be as in the abstract. We write $\mathbb{N}$ for the set of nonnegative integers.

The generalized Hilbert–Kunz function of $R = k[X]/I_2(X)$ is the function $\mathrm{HK}_{R,X} : \mathbb{N} \to \mathbb{N}$ given by

$$\mathrm{HK}_{R,X}(q) = \left(\frac{k[X]}{I_2(X) + (x_{1,1}^q, \ldots, x_{m,n}^q)}\right).$$

Namely, $k[X]/(I_2(X) + (x_{1,1}^q, \ldots, x_{m,n}^q))$ is a finite-dimensional k-vector space, and length measures that dimension. The standard Hilbert–Kunz function is only defined when k has positive prime characteristic p and when q varies over powers of p, whereas the generalized Hilbert–Kunz function is defined for arbitrary field k, regardless of the characteristic. While the Hilbert–Kunz function is not necessarily a polynomial function, it has a well-defined normalized leading coefficient. The normalized leading coefficient of the generalized Hilbert–Kunz function has been studied for example in [Conca 1996; Eto 2002; Eto and Yoshida 2003], while [Miller and Swanson 2013] studied the whole generalized Hilbert–Kunz function. Miller and Swanson gave a recursive formulation for $\mathrm{HK}_{R,X}$ and proved that it is a

MSC2010: primary 13D40; secondary 05A15, 05A10.
Keywords: Hilbert–Kunz function, multiplicity, combinatorial identity.

polynomial function. They gave closed formulations in the case $m \le 2$. This paper is an extension of [Miller and Swanson 2013].

The main result of this paper, Theorem 3.3, is the closed formulation of $\mathrm{HK}_{R,X}$ for arbitrary positive integers m, n. We also give, in Theorem 3.1, an explicit formula for the length of

$$\frac{k[X]}{I_2(X) + (x_{i,j}^q : i, j) + \sum_{j=1}^n (x_{1,j}, \dots, x_{m,j})^q}.$$

In Lemma 2.5 and Corollary 3.4 we give some explicit formulas for the number of tuples of specific length of nonnegative integers that sum up to at most a fixed number and whose first few entries are at most another fixed number. (In other words, we give formulas for the numbers of some specific compositions of integers.)

2. Set-up

Our proofs are based on the following result:

Theorem 2.1 [Miller and Swanson 2013, Theorem 2.4]. *The quotient ring*

$$\frac{k[X]}{I_2(X) + (x_{1,1}^q, \dots, x_{m,n}^q)}$$

has a k-vector space basis consisting precisely of monomials $\prod_{i,j} x_{i,j}^{p_{i,j}}$ with the following properties:

(1) *Whenever $p_{i,j} > 0$ and $i' < i$, $j < j'$, we have $p_{i',j'} = 0$.* (Monomials satisfying this property will be called *staircase monomials*. The name comes from the southwest-northeast staircase-like shape of the nonzero entries $p_{i,j}$ in the $m \times n$ matrix of all the $p_{i,j}$.)

(2) *Either $\sum_j p_{i,j} < q$ for all $i = 1, \dots, m$ or $\sum_i p_{i,j} < q$ for all $j = 1, \dots, n$.* □

Thus, to compute the Hilbert–Kunz function, we need to be able to count such monomials. The recursive formulations for this function in [Miller and Swanson 2013], as well as the explicit formulations below, require counting related sets of monomials:

Definition 2.2 [Miller and Swanson 2013, Section 3]. Let $r_1, \dots, r_m, c_1, \dots, c_n \in \mathbb{N} \cup \{\infty\}$. (In general we think of the r_i as the row sums and the c_j as the column sums.) Define $N_q(m, n; r_1, \dots, r_m; c_1, \dots, c_n)$ to be the number of monomials $\prod_{i,j} x_{i,j}^{p_{i,j}}$ with the following properties:

(1) $\prod_{i,j} x_{i,j}^{p_{i,j}}$ is a staircase monomial, i.e., whenever $p_{i,j} > 0$ and $i' < i$, $j < j'$, we have $p_{i',j'} = 0$.

(2) $\sum_j p_{i,j} \le r_i$ for all $i \in \{1, \dots, m\}$ and $\sum_i p_{i,j} \le c_j$ for all $j \in \{1, \dots, n\}$.

(3) Either $\sum_j p_{i,j} < q$ for all $i \in \{1, \dots, m\}$ or $\sum_i p_{i,j} < q$ for all $j \in \{1, \dots, n\}$.

For ease of notation, for any $c \in \mathbb{N} \cup \{\infty\}$ we let $\overline{c}$ denote a repetition of cs, where the number of occurrences depends on the context. For example, $N_q(m, n; \overline{\infty}, \overline{\infty})$ stands for $N_q(m, n; \infty, \dots, \infty; \infty, \dots, \infty)$, with m occurrences of ∞ in the first instance and n in the second. By convention, $N_q(0, n; ; c_1, \dots, c_n) = 1$.

It was proved in [Miller and Swanson 2013, Section 3] that

$$N_q(m, n; r_1, \dots, r_m; c_1, \dots, c_n) = \text{ length of} \left(\frac{k[X]}{I_2(X) + (x_{i,j}^q : i, j) + \sum_{i=1}^m (x_{i,1}, \dots, x_{i,n})^{r_i+1} + \sum_{j=1}^n (x_{1,j}, \dots, x_{m,j})^{c_j+1}} \right),$$

where for an ideal I, we set I^∞ to be the 0 ideal. Thus, in particular,

$$N_q(m, n; \overline{\infty}; \overline{\infty}) = \mathrm{HK}_{K[X]/I_2(X), X}(q).$$

Our main result in this paper relies on the count of the following monomials as well:

Definition 2.3. Let $r_1, \dots, r_m, c_1, \dots, c_n \in \mathbb{N} \cup \{\infty\}$. (We think of r_i as the i-th row sum, and of c_j as the j-th column sum.) Define $M_q(m, n; r_1, \dots, r_m; c_1, \dots, c_n)$ to be the number of monomials $\prod_{i,j} x_{i,j}^{p_{i,j}}$ such that:

(1) $\prod_{i,j} x_{i,j}^{p_{i,j}}$ is a staircase monomial, i.e., whenever $p_{i,j} > 0$ and $i' < i$, $j < j'$, we have $p_{i',j'} = 0$.

(2) $\sum_j p_{i,j} \le \min\{r_i, q-1\}$ for all $i \in \{1, \dots, m\}$.

(3) There exists $j \in \{1, \dots, n\}$ such that $\sum_i p_{i,j} > c_j$.

The following lemma says that mn exponents $p_{i,j}$ of a staircase monomial can be identified by $m+n$ or even $m+n-1$ numbers:

Lemma 2.4. *Suppose that $r_1, \dots, r_m, c_1, \dots, c_n$ are nonnegative integers and that $\sum_i r_i = \sum_j c_j$. Then there exists a unique staircase monomial $\prod_{i,j} x_{i,j}^{p_{i,j}}$ such that $r_i = \sum_j p_{i,j}$ for all $i = 1, \dots m$ and $c_j = \sum_i p_{i,j}$ for all $j = 1, \dots, n$.*

Proof. If $m = 1$, then clearly $p_{1,j} = c_j$, which is uniquely determined. If $n = 1$, necessarily $p_{i,1} = r_i$.

In general, for arbitrary m and n, knowing c_1 and r_m is enough information to uniquely determine $p_{m,1}$: if $p_{m,1} < \min\{c_1, r_m\}$, then the m-th row has a nonzero number beyond the first entry and the first column has a nonzero number in the first $m-1$ rows, which then makes the corresponding monomial nonstaircase and is not allowed. So necessarily $p_{m,1} = \min\{c_1, r_m\}$. If $p_{m,1} = c_1$, then no more nonzero exponents appear in the first column, and it remains to fill in the remaining $m \times (n-1)$

matrix of $p_{i,j}$ with the remaining numbers $r_1, \dots, r_{m-1}, r_m - c_1, c_2, \dots, c_n$. If instead $p_{m,1} = r_m$, then no more nonzero exponents appear in the last row, and it remains to fill in the remaining $(m-1) \times n$ matrix of $p_{i,j}$ with the remaining numbers $r_1, \dots, r_{m-1}, c_1 - r_m, c_2, \dots, c_n$. $\square$

Lemma 2.5. *Let a, b, w, z be integers with $a \le b$. The number of b-tuples of nonnegative integers that sum up to at most w and for which the first a entries are strictly smaller than z equals*

$$\sum_{i=0}^{a} (-1)^i \binom{a}{i} \binom{w-iz+b}{b}.$$

Proof. (This proof was suggested by the referee.) Let E be the set of all b-tuples $(v_1, \dots, v_b)$ of nonnegative integers that sum up to at most w. It is well known that $|E| = \binom{w+b}{b}$. Let E_j be the subset E of those tuples for which $v_i \ge z$. Then $|E_j| = \binom{w-z+b}{b}$, and more generally, $|E_{j_1} \cap \dots \cap E_{j_i}| = \binom{w-iz+b}{b}$. The desired cardinality is $|E \setminus (E_1 \cup \dots \cup E_a)|$, which, by the inclusion-exclusion principle, equals

$$\begin{aligned} |E| - \sum_{i=1}^{a} (-1)^{i-1} \sum_{1 \le j_1 < \dots < j_i \le a} |E_{j_1} \cap \dots \cap E_{j_i}| \\ &= \binom{w-iz+b}{b} - \sum_{i=1}^{a} (-1)^{i-1} \binom{a}{i} \binom{w-iz+b}{b} \\ &= \sum_{i=0}^{a} (-1)^i \binom{a}{i} \binom{w-iz+b}{b}. \end{aligned}$$ $\square$

Lemma 2.6. *Let a, b, w, z be integers with $a < b$. The following numbers are the same*:

(1) *The number of b-tuples of nonnegative integers that sum to at most $a(z-1) - w$ and for which the first a entries are strictly smaller than z.*

$$\sum_{i=0}^{a} (-1)^i \binom{a}{i} \binom{a(z-1)-w-iz+b}{b}. \tag{2}$$

(3) *The number of b-tuples of nonnegative integers for which the first a entries are strictly smaller than z and the sum of the first a entries is greater than or equal to w plus the sum of the remaining entries.*

Proof. The first two numbers are the same by Lemma 2.5.

Let

$$E = \left\{ (v_1, \ldots, v_b) \in \mathbb{N}^b : v_1, \ldots, v_a < z, \sum_i v_i \le a(z-1) - w \right\},$$

$$F = \left\{ (u_1, \ldots, u_b) \in \mathbb{N}^b : u_1, \ldots, u_a < z, \sum_{i=1}^{a} u_i \ge \sum_{i>a} u_i + w \right\},$$

so $|E|$ is the number in (1) and $|F|$ is the number in (3). Define $\varphi : E \to F$ by

$$\varphi(v_1, \ldots, v_b) = (z-1-v_1, \ldots, z-1-v_a, v_{a+1}, \ldots, v_b).$$

Certainly this image is in $\mathbb{N}^b$, each of the first a entries is strictly smaller than z, and the sum of the first a entries is $a(z-1)-\sum_{i=1}^{a} v_i = a(z-1)-\sum_{i=1}^{b} v_i + \sum_{i=a+1}^{b} v_i \ge w + \sum_{i=a+1}^{b} v_i$, so that the range of φ is in F. The proof of surjectivity is similar, and injectivity is clear. Thus φ is bijective, which proves that the numbers in (1) and (3) are the same. □

3. Main theorems

In this section we give explicit (nonrecursive) formulas for $N_q(m, n; \overline{\infty}; \overline{q-1})$, $M_q(m, n; \overline{q-1}; \overline{q-1})$, and $N_q(m, n; \overline{\infty}; \overline{\infty})$ for arbitrary positive integers m, n.

Theorem 3.1. *For all nonnegative integers m, n, q, the k-length of the quotient ring $k[X]/\big(I_2(X) + (x_{i,j}^q : i, j) + \sum_{j=1}^{n}(x_{1,j}, \ldots, x_{m,j})^q\big)$ equals*

$$N_q(m, n; \overline{\infty}; \overline{q-1}) = \sum_{i=0}^{n} (-1)^{n-i} \binom{n}{i} \binom{iq+m-1}{m+n-1},$$

and furthermore, this number equals the number of $(m+n-1)$-tuples of nonnegative integers that sum up to at most $n(q-1)$ and for which the first n entries are strictly smaller than q.

Proof. Let $T_{m,n,q}$ be the set of all staircase monomials $\prod_{i,j} x_{i,j}^{p_{i,j}}$ such that $\sum_i p_{i,j} < q$ for all $j = 1, \ldots, n$. By [Miller and Swanson 2013, Section 3],

$$\begin{aligned} |T_{m,n,q}| &= N_q(m, n; \overline{\infty}; \overline{q-1}) \\ &= \left(\frac{k[X]}{I_2(X) + (x_{i,j}^q : i, j) + \sum_{j=1}^{n}(x_{1,j}, \ldots, x_{m,j})^q} \right). \end{aligned}$$

Let W be the set of $(m+n-1)$-tuples of nonnegative integers such that the first n entries are strictly smaller than q, and the sum of the first n entries is greater than or equal to the sum of the remaining entries. To each element $\prod_{i,j} x_{i,j}^{p_{i,j}}$ in $T_{m,n,q}$ we associate the $(m+n-1)$-tuple $\big(\sum_i p_{i1}, \ldots, \sum_i p_{in}, \sum_j p_{2j}, \ldots, \sum_j p_{mj}\big)$ of nonnegative integers. This is an element of W. For any $(c_1, \ldots, c_n, r_2, \ldots, r_m) \in W$, set $r_1 = \sum_j c_j - \sum_{i=2}^m r_m \in \mathbb{N}$. By Lemma 2.4, there is a unique element of $T_{m,n,q}$ that corresponds to $(c_1, \dots, c_n, r_2, \dots, r_m)W$. Thus $|T_{m,n,q}| = |W|$, and by Lemma 2.6 applied with $w = 0$,

$$\begin{aligned} N_q(m, n; \overline{\infty}; \overline{q-1}) &= |W| \\ &= \sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{n(q-1)-iq+m+n-1}{m+n-1} \\ &= \sum_{i=0}^{n} (-1)^i \binom{n}{n-i} \binom{(n-i)q+m-1}{m+n-1} \\ &= \sum_{i=0}^{n} (-1)^{n-i} \binom{n}{i} \binom{iq+m-1}{m+n-1}. \end{aligned}$$

□

Theorem 3.2. *For all positive integers m, n,*

$$M_q(m, n; \overline{q-1}; \overline{q-1}) = \sum_{i=1}^{n} \sum_{j=0}^{m} (-1)^{m-j+i-1} \binom{n}{i} \binom{m}{j} \binom{jq-iq+n-1}{m+n-1}.$$

Proof. Let S be the set of all $(m+n)$-tuples $(r_1, \dots, r_m, c_1, \dots, c_n)$ of nonnegative integers such that $\sum_i r_i = \sum_j c_j$, $r_1, \dots, r_m < q$, and there exists j such that $c_j \ge q$. By Lemma 2.4, each such tuple uniquely determines a staircase monomial, and by definition of M_q, the number of these monomials is precisely $M_q(m, n; \overline{q-1}; \overline{q-1})$. We will count these monomials via the tuples.

We define the function $f : S \to 2^{\{1,\dots,n\}}$ as $f(r_1, \dots, r_m, c_1, \dots, c_n) = \{j : c_j \ge q\}$. By S_k we denote the set of all those $x \in S$ for which $|f(x)| = k$. Consider the set A of all (L, x) for which $x \in S$ and L is a nonempty subset of $f(x)$. For each $l = 1, \dots, n$, the number of $(L, x) \in A$ with $|L| = l$ equals

$$\sum_{k=l}^{n} \binom{k}{l} |S_k|.$$

In other words, (L, x) only arises if $L \subseteq f(x)$, and each $x \in S_k$, generates $\binom{k}{l}$ distinct elements (L, x) in A with $|L| = l$.

We count the elements of A another way. Put $(L, x) = (L, r_1, \dots, r_m, c_1, \dots, c_n)$ with $|L| = l$. Since $\sum_i r_i = \sum_j c_j$, one of the c_j is redundant, and we remove c_s, where $s = \min L$. Furthermore, we lose no information if we subtract q from each c_l with $l \in L$. Set $c'_j = c_j - q$ if $j \in L$ and $c'_j = c_j$ otherwise. Thus, to count all (L, x), it suffices to count all $(L, r_1, \dots, r_m, c'_1, \dots, c'_{s-1}, c'_{s+1}, \dots, c'_n)$. But the set of all such $(m+n)$-tuples equals $P_l(n) \times W$, where $P_l(n)$ is the set of all l-element subsets of $\{1, \dots, n\}$, and W is the set of all $(m+n-1)$-tuples of nonnegative integers whose first m entries are strictly smaller than q, and the sum of the first

m entries is greater than or equal to lq plus the sum of the remaining entries. By Lemma 2.6, W has cardinality

$$\sum_{i=0}^{m}(-1)^i\binom{m}{i}\binom{m(q-1)-lq-iq+m+n-1}{m+n-1} = \sum_{i=0}^{m}(-1)^i\binom{m}{m-i}\binom{(m-i)q-lq+n-1}{m+n-1} = \sum_{i=0}^{m}(-1)^{m-i}\binom{m}{i}\binom{iq-lq+n-1}{m+n-1}.$$

Using $|P_l(n)| = \binom{n}{l}$, this gives a system of n linear equations with matrix form:

$$\begin{bmatrix} \binom{1}{1} & \binom{2}{1} & \binom{3}{1} & \cdots & \binom{n}{1} \\ \binom{1}{2} & \binom{2}{2} & \binom{3}{2} & \cdots & \binom{n}{2} \\ \binom{1}{3} & \binom{2}{3} & \binom{3}{3} & \cdots & \binom{n}{3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \binom{1}{n} & \binom{2}{n} & \binom{3}{n} & \cdots & \binom{n}{n} \end{bmatrix} \begin{bmatrix} |S_1| \\ |S_2| \\ |S_2| \\ \vdots \\ |S_n| \end{bmatrix} = \begin{bmatrix} \binom{n}{1}\sum_{i=0}^{m}(-1)^{m-i}\binom{m}{i}\binom{iq-1q+n-1}{m+n-1} \\ \binom{n}{2}\sum_{i=0}^{m}(-1)^{m-i}\binom{m}{i}\binom{iq-2q+n-1}{m+n-1} \\ \binom{n}{3}\sum_{i=0}^{m}(-1)^{m-i}\binom{m}{i}\binom{iq-3q+n-1}{m+n-1} \\ \vdots \\ \binom{n}{n}\sum_{i=0}^{m}(-1)^{m-i}\binom{m}{i}\binom{iq-nq+n-1}{m+n-1} \end{bmatrix}.$$

Note that the matrix $\big[\binom{j}{i}\big]_{i,j}$ is upper triangular with determinant 1. Its inverse is the upper triangular matrix $\big[(-1)^{i+j}\binom{j}{i}\big]_{i,j}$, as for all $i \le j$,

$$\begin{aligned}\sum_{k=1}^{n}(-1)^{i+k}\binom{k}{i}\binom{j}{k} &= \sum_{k=i}^{j}(-1)^{i+k}\frac{k!}{i!(k-i)!}\frac{j!}{k!(j-k)!} \\ &= \sum_{k=i}^{j}(-1)^{i+k}\frac{j!}{i!(j-i)!}\frac{(j-i)!}{(k-i)!(j-k)!} \\ &= \binom{j}{i}\sum_{k=i}^{j}(-1)^{i+k}\binom{j-i}{k-i} = \binom{j}{i}\sum_{k=0}^{j-i}(-1)^k\binom{j-i}{k},\end{aligned}$$

which is 0 if $j > i$ and is 0 if $j = i$. Thus, by Cramer's rule,

$$\begin{aligned}
M_q(m,n;\overline{q-1};\overline{q-1}) &= \sum_{k=1}^{n} |S_k| \\
&= \sum_{k=1}^{n}\sum_{j=1}^{n}(-1)^{k+j}\binom{j}{k}\binom{n}{j}\sum_{i=0}^{m}(-1)^{m-i}\binom{m}{i}\binom{iq-jq+n-1}{m+n-1} \\
&= \sum_{j=1}^{n}(-1)^{j}\binom{n}{j}\sum_{k=1}^{n}(-1)^{k}\binom{j}{k}\sum_{i=0}^{m}(-1)^{m-i}\binom{m}{i}\binom{iq-jq+n-1}{m+n-1} \\
&= \sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j}\sum_{i=0}^{m}(-1)^{m-i}\binom{m}{i}\binom{iq-jq+n-1}{m+n-1} \\
&= \sum_{j=1}^{n}\sum_{i=0}^{m}(-1)^{m-i+j-1}\binom{n}{j}\binom{m}{i}\binom{iq-jq+n-1}{m+n-1} \\
&= \sum_{j=1}^{n}\sum_{i=1}^{m}(-1)^{m-i+j-1}\binom{n}{j}\binom{m}{i}\binom{iq-jq+n-1}{m+n-1}. \qquad \square
\end{aligned}$$

The main theorem on the generalized Hilbert–Kunz function now follows:

Theorem 3.3. *For all positive integers m, n, the Hilbert function $\mathrm{HK}_{R,X}(q)$ of $k[X]/I_2(X)$ at q, i.e., the length of $k[X]/\big(I_2(X)+(x_{1,1}^q,\dots,x_{m,n}^q)\big)$, equals*

$$\begin{aligned}
N_q(m,n;\overline{\infty};\overline{\infty}) &= N_q(m,n;\overline{\infty};\overline{q-1}) + M_q(m,n;\overline{q-1};\overline{q-1}) \\
&= \sum_{i=1}^{n}(-1)^{n-i}\binom{n}{i}\binom{iq+m-1}{m+n-1} \\
&\qquad + \sum_{i=1}^{n}\sum_{j=1}^{m}(-1)^{m-j+i-1}\binom{n}{i}\binom{m}{j}\binom{jq-iq+n-1}{m+n-1}.
\end{aligned}$$

Proof. By Definition 2.2, $N_q(m,n;\overline{\infty};\overline{\infty})$ counts all the staircase monomials $\prod_{i,j} x_{i,j}^{p_{i,j}}$ with the property that either $\sum_i p_{i,j} < q$ for all j or $\sum_j p_{i,j} < q$ for all i.

The number $N_q(m,n;\overline{\infty};\overline{q-1})$ counts those monomials in the previous paragraph for which $\sum_i p_{i,j} < q$ for all j, and $M_q(m,n;\overline{q-1};\overline{q-1})$ counts those monomials for which $\sum_i p_{i,j} \geq q$ for some j. Thus

$$N_q(m,n;\overline{\infty};\overline{\infty}) = N_q(m,n;\overline{\infty};\overline{q-1}) + M_q(m,n;\overline{q-1};\overline{q-1}),$$

and by Theorems 3.1 and 3.2, this is equal to the claimed sums of binomial coefficients. $\square$

In particular, comparison with Theorem 4.4 in [Miller and Swanson 2013] when $m = 2$ gives:

Corollary 3.4. *The number of $(n+1)$-tuples of nonnegative integers that sum up to at most $n(q-1)$ and for which the first n entries are strictly smaller than q equals*

$$\sum_{i=1}^{n}(-1)^{n-i}\binom{n}{i}\binom{iq+1}{n+1}=\frac{nq^{n+1}-(n-2)q^{n}}{2}.$$

Proof. According Theorem 4.4 in to [Miller and Swanson 2013],

$$N_q(2,n;\overline{\infty};\overline{\infty})=\frac{nq^{n+1}-(n-2)q^{n}}{2}+n\binom{q+n-1}{n+1},$$

and by Theorem 3.3,

$$\begin{aligned}N_q(2,n;\overline{\infty};\overline{\infty})&=\sum_{i=1}^{n}(-1)^{n-i}\binom{n}{i}\binom{iq+1}{n+1}+2\sum_{i=1}^{n}(-1)^{i}\binom{n}{i}\binom{q-iq+n-1}{n+1}\\&\qquad+\sum_{i=1}^{n}(-1)^{i-1}\binom{n}{i}\binom{2q-iq+n-1}{n+1}\\&=\sum_{i=1}^{n}(-1)^{n-i}\binom{n}{i}\binom{iq+1}{n+1}+\binom{n}{1}\binom{2q-1q+n-1}{n+1}\\&=\sum_{i=1}^{n}(-1)^{n-i}\binom{n}{i}\binom{iq+1}{n+1}+n\binom{q+n-1}{n+1}.\end{aligned}$$

Thus $\sum_{i=1}^{n}(-1)^{n-i}\binom{n}{i}\binom{iq+1}{n+1}=\big(nq^{n+1}-(n-2)q^{n}\big)/2$. By Theorem 3.1, this number is the number of $(n+1)$-tuples of nonnegative integers that sum up to at most $n(q-1)$ and for which the first n entries are strictly smaller than q. $\square$

We remark here that we know of no other proof of the equality in the last corollary. Natural first attempts would be induction and Gosper's algorithm, and neither of these is successful, as for one thing, the summands depend not only on the summing index i but also on n. The challenge remains to establish a closed-form expression for $N_q(m,n;\overline{\infty};\overline{\infty})$ and $N_q(m,n;\overline{\infty};\overline{q-1})$ for higher m.

Acknowledgment

We thank the anonymous referee for all comments, especially for the improved proof of Lemma 2.5.

References

[Conca 1996] A. Conca, "Hilbert–Kunz function of monomial ideals and binomial hypersurfaces", *Manuscripta Math.* **90**:3 (1996), 287–300. MR 97g:13006 Zbl 0882.13019

[Eto 2002] K. Eto, "Multiplicity and Hilbert–Kunz multiplicity of monoid rings", *Tokyo J. Math.* **25**:2 (2002), 241–245. MR 2003m:13015 Zbl 1096.13519

[Eto and Yoshida 2003] K. Eto and K.-i. Yoshida, "Notes on Hilbert–Kunz multiplicity of Rees algebras", *Comm. Algebra* **31**:12 (2003), 5943–5976. MR 2004k:13040 Zbl 1096.13021

[Miller and Swanson 2013] L. E. Miller and I. Swanson, "Hilbert–Kunz functions of 2×2 determinantal rings", *Illinois J. Math.* **57**:1 (2013), 251–277. MR 3224570 Zbl 06322992

Received March 4, 2014. Revised December 16, 2014.

MARCUS ROBINSON
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH
155 S 1400 E
JOHN WIDTSOE BUILDING 233
SALT LAKE CITY, UT 84112
UNITED STATES

robinson@math.utah.edu

IRENA SWANSON
DEPARTMENT OF MATHEMATICS
REED COLLEGE
3203 SE WOODSTOCK BLVD
PORTLAND, OR 97202-8199
UNITED STATES

iswanson@reed.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 275, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.275.443

THE JOHNSON–MORITA THEORY FOR THE RING OF FRICKE CHARACTERS OF FREE GROUPS

TAKAO SATOH

For a free group F_n of rank n, we consider the ring $\mathfrak{X}_{\mathbb{Q}}(F_n)$ of $\mathrm{SL}(2, \mathbb{C})$-characters of F_n generated by all Fricke characters $\operatorname{tr} x$ for $x \in F_n$. Its ideal J generated by $\operatorname{tr} x - 2$ for all $x \in F_n$ is $\operatorname{Aut} F_n$-invariant. We denote by $\mathcal{E}_n(1)$ the subgroup of the automorphism group $\operatorname{Aut} F_n$ of F_n consisting of all automorphisms which act on J/J^2 trivially. The group $\mathcal{E}_n(1)$ is regarded as a Fricke character analogue of the IA-automorphism group of F_n and the Torelli subgroup of the mapping class group of a surface. In our previous work, we constructed a homomorphism η_1 from $\mathcal{E}_n(1)$ into $\mathrm{Hom}_{\mathbb{Q}}(J/J^2, J^2/J^3)$ as a Fricke character analogue of the first Johnson homomorphisms of the mapping class group and $\operatorname{Aut} F_n$.

In this paper, according to Morita's work for the extension of the first Johnson homomorphism of the mapping class group, we extend η_1 to $\operatorname{Aut} F_n$ as a crossed homomorphism. We see that the obtained crossed homomorphism η is not null cohomologous. We also compute the images of Nielsen's generators of $\operatorname{Aut} F_n$ by η.

1.	Introduction	443
2.	Preliminaries	447
3.	An extension of η_1 as a crossed homomorphism	452
4.	Nontriviality of η as a 1-cocycle	457
5.	The image of the crossed homomorphism η	458
	Acknowledgments	460
	References	460

1. Introduction

In a series of works, Dennis Johnson [1980; 1983; 1985a; 1985b] established a remarkable method to investigate the group structure of the Torelli subgroup of the mapping class group of a surface. In particular, he constructed in [Johnson 1985b] a homomorphism τ to determine the abelianization of the Torelli subgroup.

MSC2010: primary 20F28; secondary 20J06.

Keywords: ring of Fricke characters, automorphism group of free groups, IA-automorphism group, Andreadakis–Johnson filtration, Johnson homomorphisms.

Today, his homomorphism τ is called the first Johnson homomorphism and has been generalized to those of higher degrees. Over the last two decades, good progress was made in the study of the Johnson homomorphisms of mapping class groups through the work of many authors including Morita [1993a], Hain [1997] and others.

Let F_n be a free group generated by $x_1, x_2, \ldots, x_n$. As is well known, for any $g \geq 1$, the mapping class group $\mathcal{M}_{g,1}$ of a compact oriented surface $\Sigma_{g,1}$ with one boundary component can be embedded into $\mathrm{Aut}\, F_{2g}$ by a classical work of Dehn and Nielsen. This embedding is induced from the action of $\mathcal{M}_{g,1}$ on the fundamental group of $\Sigma_{g,1}$. The definition of the Johnson homomorphisms of $\mathcal{M}_{g,1}$ can be naturally generalized to those of $\mathrm{Aut}\, F_n$. Let H be the abelianization of F_n. The kernel of the homomorphism $\mathrm{Aut}\, F_n \to \mathrm{Aut}\, H \cong \mathrm{GL}(n, \mathbb{Z})$ induced from the action of $\mathrm{Aut}\, F_n$ on H, is called the IA-automorphism group of F_n and is denoted by IA_n. The group IA_n is a free group analogue of the Torelli subgroup $\mathcal{I}_{g,1}$ of $\mathcal{M}_{g,1}$. Andreadakis [1965] introduced a central filtration

$$\mathrm{IA}_n = \mathcal{A}_n(1) \supset \mathcal{A}_n(2) \supset \cdots$$

of IA_n, and showed that each graded quotient $\mathrm{gr}^k \mathcal{A}_n := \mathcal{A}_n(k)/\mathcal{A}_n(k+1)$ is a free abelian group of finite rank. We call the above filtration the Andreadakis–Johnson filtration of $\mathrm{Aut}\, F_n$. Johnson studied this type of filtration for the mapping class groups in 1980s. The general linear group $\mathrm{GL}(n, \mathbb{Z})$ naturally acts on each $\mathrm{gr}^k \mathcal{A}_n$. In order to investigate the $\mathrm{GL}(n, \mathbb{Z})$-module structure of $\mathrm{gr}^k \mathcal{A}_n$, the k-th Johnson homomorphism

$$\tau_k : \mathrm{gr}^k \mathcal{A}_n \to H^* \otimes_{\mathbb{Z}} \mathcal{L}_n(k+1)$$

is a powerful and useful tool. However, even the $\mathrm{GL}(n, \mathbb{Q})$-structure of $(\mathrm{gr}^k \mathcal{A}_n) \otimes_{\mathbb{Z}} \mathbb{Q}$ is not determined in general (see Section 2C for notation, and [Satoh 2009; 2013b] for basic materials concerning the Andreadakis–Johnson filtration and the Johnson homomorphisms of $\mathrm{Aut}\, F_n$).

Now, we study a Fricke character analogue of the Andreadakis–Johnson filtration and the Johnson homomorphisms of $\mathrm{Aut}\, F_n$. Let $R(F_n)$ be the set of all $\mathrm{SL}(2, \mathbb{C})$-representations of F_n, and $\mathcal{F}(R(F_n), \mathbb{C})$ the set of all complex-valued functions on $R(F_n)$. Then $\mathcal{F}(R(F_n), \mathbb{C})$ naturally has the $\mathbb{C}$-algebra structure coming from the pointwise product, and $\mathrm{Aut}\, F_n$ naturally acts on $\mathcal{F}(R(F_n), \mathbb{C})$ from the right. For any $x \in F_n$, define $\mathrm{tr}\, x \in \mathcal{F}(R(F_n), \mathbb{C})$ by

$$(\mathrm{tr}\, x)(\rho) := \mathrm{tr}\, \rho(x), \qquad \rho \in R(F_n).$$

Here tr on the right-hand side means the usual trace of 2×2 matrix $\rho(x)$. The element $\mathrm{tr}\, x$ is called the Fricke character of $x \in F_n$. Classically, Fricke characters were introduced by Fricke and Klein [1897] to study the moduli space of compact

Riemann surfaces. In this paper, however, we focus on purely algebraic properties of the Fricke characters. Let $\mathfrak{X}_{\mathbb{Q}}(F_n)$ be the $\mathbb{Q}$-subalgebra of $\mathcal{F}(R(F_n), \mathbb{C})$ generated by all $\operatorname{tr} x$ for $x \in F_n$. We call $\mathfrak{X}_{\mathbb{Q}}(F_n)$ the ring of Fricke characters of F_n over $\mathbb{Q}$. Horowitz [1972] showed that $\mathfrak{X}_{\mathbb{Q}}(F_n)$ is finitely generated by

$$\{\operatorname{tr} x_{i_1} \cdots x_{i_l} \mid 1 \le l \le 3,\ 1 \le i_1 < i_2 < \cdots < i_l \le n\}.$$

In order to establish the Johnson–Morita theory for $\mathfrak{X}_{\mathbb{Q}}(F_n)$, to begin with, we need a descending filtration of $\mathfrak{X}_{\mathbb{Q}}(F_n)$ consisting of $\operatorname{Aut} F_n$-invariant ideals. Consider an ideal

$$J := (\operatorname{tr}' x_{i_1} \cdots x_{i_l} \mid 1 \le l \le 3,\ 1 \le i_1 < i_2 < \cdots < i_l \le n) \subset \mathfrak{X}_{\mathbb{Q}}(F_n),$$

where $\operatorname{tr}' x := \operatorname{tr} x - 2$ for any $x \in F_n$. Magnus [1980] showed that J is $\operatorname{Aut} F_n$-invariant for the case where $n = 3$, and studied a representation of the quotient group of IA_3 by the inner automorphism group $\operatorname{Inn} F_3$ of F_3. On the other hand, it is easily seen that J is $\operatorname{Aut} F_n$-invariant for general $n \ge 3$. In this paper we use this ideal and the descending filtration

$$J \supset J^2 \supset J^3 \supset \cdots$$

in order to define a descending filtration of $\operatorname{Aut} F_n$ as an analogy of the Andreadakis–Johnson filtration. Although each graded quotient $\operatorname{gr}^k J := J^k / J^{k+1}$ is an $\operatorname{Aut} F_n$-invariant finite-dimensional $\mathbb{Q}$-vector space, by combinatorial complexities, it is quite difficult to give a basis of $\operatorname{gr}^k J$ for $n \ge 3$ in general. In [Hatakenaka and Satoh 2014] — henceforth abbreviated [HS 2014] — we explicitly give bases of $\operatorname{gr}^k J$ for $k = 1$ and 2 (see Section 4 for details).

For any $k \ge 1$, set

$$\mathcal{E}_n(k) := \ker\bigl(\operatorname{Aut} F_n \to \operatorname{Aut}(J/J^{k+1})\bigr),$$

where the homomorphism is induced by the action of $\operatorname{Aut} F_n$ on J/J^{k+1}. The groups $\mathcal{E}_n(k)$ define a descending filtration

$$\mathcal{E}_n(1) \supset \mathcal{E}_n(2) \supset \cdots$$

of $\operatorname{Aut} F_n$. In [HS 2014], we showed this filtration is central, and $\mathcal{A}_n(2k) \subset \mathcal{E}_n(k)$ for any $k \ge 1$. Furthermore, we determined the first term $\mathcal{E}_n(1)$ to be $\operatorname{Inn} F_n \cdot \mathcal{A}_n(2)$ where $\operatorname{Inn} F_n$ is the inner automorphism group of F_n. Thus, each graded quotient $\operatorname{gr}^k \mathcal{E}_n := \mathcal{E}_n(k)/\mathcal{E}_n(k+1)$ is an abelian group for any $k \ge 1$. In order to study the structures of $\operatorname{gr}^k \mathcal{E}_n$, we have introduced homomorphisms

$$\eta_k : \operatorname{gr}^k \mathcal{E}_n \to \operatorname{Hom}_{\mathbb{Q}}(\operatorname{gr}^1 J, \operatorname{gr}^{k+1} J)$$

defined by

$$\sigma \pmod{\mathcal{E}_n(k+1)} \mapsto \bigl(f \pmod{J^2} \mapsto f^{\sigma} - f \pmod{J^{k+1}}\bigr).$$

The homomorphism η_k is a Fricke character analogue of the k-th Johnson homomorphism τ_k. In [HS 2014], we showed that each η_k is an $\mathrm{Aut}\,F_n/\mathcal{E}_n(1)$-equivariant injective homomorphism. This implies that each of $\mathrm{gr}^k\mathcal{E}_n$ is torsion-free, and that $\dim_{\mathbb{Q}}(\mathrm{gr}^k\mathcal{E}_n \otimes_{\mathbb{Z}} \mathbb{Q}) < \infty$.

In this paper, we concentrate on the homomorphism

$$\tilde{\eta}_1 : \mathcal{E}_n(1) \to \mathrm{gr}^1\mathcal{E}_n \xrightarrow{\eta_1} \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J).$$

Morita [1993b] showed that the composition

$$\mathcal{I}_{g,1} \xrightarrow{\tau_1} H^* \otimes_{\mathbb{Z}} \mathcal{L}_{2g}(k+1) \to (H^* \otimes_{\mathbb{Z}} \mathcal{L}_{2g}(k+1)) \otimes_{\mathbb{Z}} \mathbb{Z}\big[\tfrac{1}{2}\big]$$

of the first Johnson homomorphism of the mapping class group $\mathcal{M}_{g,1}$ the natural projection naturally extends to $\mathcal{M}_{g,1}$ as a crossed homomorphism. He also showed that this extension is unique up to 1-coboundary. Here $\mathbb{Z}\big[\frac{1}{2}\big]$ means a subring of $\mathbb{Q}$ obtained from $\mathbb{Z}$ by attaching $\frac{1}{2}$. The analogous result for $\mathrm{Aut}\,F_n$ was obtained by Kawazumi [2006], who showed the composition

$$\mathrm{IA}_n \to \mathrm{gr}^1\mathcal{A}_n \xrightarrow{\tau_1} H^* \otimes_{\mathbb{Z}} \mathcal{L}_n(k+1) \to (H^* \otimes_{\mathbb{Z}} \mathcal{L}_n(k+1)) \otimes_{\mathbb{Z}} \mathbb{Z}\big[\tfrac{1}{2}\big]$$

of the first Johnson homomorphism of $\mathrm{Aut}\,F_n$ with the natural projection naturally and uniquely extends to $\mathrm{Aut}\,F_n$ as a crossed homomorphism up to 1-coboundary. Furthermore, very recently, Day [2013] showed that each Johnson homomorphism τ_k of $\mathrm{Aut}\,F_n$ can be extended to $\mathrm{Aut}\,F_n$ as a crossed homomorphism.

The main purpose of the paper is to give a Fricke character analogue of these results. Namely, according to Morita's work, we extend $\tilde{\eta}_1$ to $\mathrm{Aut}\,F_n$ as a crossed homomorphism:

Theorem 3.6. *There is a crossed homomorphism $\eta : \mathrm{Aut}\,F_n \to \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$ such that the restriction of η to $\mathcal{E}_n(1)$ is η_1.*

At the present stage, we do not know whether the extension η is unique up to 1-coboundary or not since we cannot determine the first cohomology group of $\mathrm{Aut}\,F_n$ with coefficients in $\mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$ due to the combinatorial complexity. By using his extended crossed homomorphisms, Kawazumi [2006] constructed twisted higher cocycles of $\mathrm{Aut}\,F_n$, and investigated its restriction to the mapping class group. In particular, he expressed the Morita–Mumford classes as a cup product of the twisted cocycles. In order to study the twisted cohomology groups of $\mathrm{Aut}\,F_n$ with coefficients in modules of the Fricke characters, we are convinced that our work establishes a foothold as a first step.

In Section 2, we review the definitions of Fricke characters, the Johnson homomorphisms τ_k and the homomorphisms η_k. In Section 3, we extend the homomorphism η_1 to $\mathrm{Aut}\,F_n$ as a crossed homomorphism. In Section 4, we show that the crossed homomorphism η is not null cohomologous in the first cohomology group of $\mathrm{Aut}\, F_n$ with coefficients in $\mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$. Furthermore, in Section 5, we calculate the image of Nielsen's generators of $\mathrm{Aut}\, F_n$ by η.

2. Preliminaries

2A. ***Notation and conventions.*** Throughout the paper, we use the following notation and conventions: Let G be a group and N a normal subgroup of G.

- The abelianization of G is denoted by G^{ab} unless otherwise noted.
- The automorphism group $\mathrm{Aut}\, G$ of G acts on G from the right. For any $\sigma \in \mathrm{Aut}\, G$ and $x \in G$, the action of σ on x is denoted by x^σ.
- For an element $g \in G$, we also denote the coset class of g in G/N by g if there is no confusion in context.
- For any elements x and y of G, the commutator bracket $[x, y]$ of x and y is defined to be $[x, y] := xyx^{-1}y^{-1}$.

2B. ***The rings of Fricke characters.*** In this subsection, we review the rings of Fricke characters of the free group F_n. Let $R(F_n)$ be the set $\mathrm{Hom}(F_n, \mathrm{SL}(2, \mathbb{C}))$ of all $\mathrm{SL}(2, \mathbb{C})$-representations of F_n, and $\mathcal{F}(R(F_n), \mathbb{C})$ the set of all complex-valued functions on $R(F_n)$. Then $\mathcal{F}(R(F_n), \mathbb{C})$ naturally has a $\mathbb{C}$-algebra structure by the operations defined by

$$\begin{aligned}(\chi + \chi')(\rho) &:= \chi(\rho) + \chi'(\rho),\\ (\chi\chi')(\rho) &:= \chi(\rho)\chi'(\rho),\\ (\lambda\chi)(\rho) &:= \lambda(\chi(\rho)),\end{aligned}$$

for any $\chi, \chi' \in \mathcal{F}(R(F_n), \mathbb{C})$, $\lambda \in \mathbb{C}$ and $\rho \in R(F_n)$. The group $\mathrm{Aut}\, F_n$ naturally acts on $R(F_n)$ and $\mathcal{F}(R(F_n), \mathbb{C})$ from the right by

$$\begin{aligned}\rho^\sigma(x) &:= \rho(x^{\sigma^{-1}}) \quad \text{for } \rho \in R(F_n) \text{ and } x \in F_n,\\ \chi^\sigma(\rho) &:= \chi(\rho^{\sigma^{-1}}) \quad \text{for } \chi \in \mathcal{F}(R(F_n), \mathbb{C}) \text{ and } \rho \in R(F_n),\end{aligned}$$

for any $\sigma \in \mathrm{Aut}\, F_n$. For any $x \in F_n$, we define an element $\mathrm{tr}\, x$ of $\mathcal{F}(R(F_n), \mathbb{C})$ by

$$(\mathrm{tr}\, x)(\rho) := \mathrm{tr}\, \rho(x)$$

for any $\rho \in R(F_n)$. Here tr on the right-hand side means the trace of 2×2 matrix $\rho(x)$. The element $\mathrm{tr}\, x$ is called the Fricke character of $x \in F_n$. The action of an element $\sigma \in \mathrm{Aut}\, F_n$ on $\mathrm{tr}\, x$ is given by $\mathrm{tr}\, x^\sigma$. We have the following well-known formulae:

(1) $\mathrm{tr}\, xy + \mathrm{tr}\, xy^{-1} = (\mathrm{tr}\, x)(\mathrm{tr}\, y)$,

(2) $\mathrm{tr}\, xyz + \mathrm{tr}\, yxz = (\mathrm{tr}\, x)(\mathrm{tr}\, yz) + (\mathrm{tr}\, y)(\mathrm{tr}\, xz) + (\mathrm{tr}\, z)(\mathrm{tr}\, xy) - (\mathrm{tr}\, x)(\mathrm{tr}\, y)(\mathrm{tr}\, z)$,

(3) $\mathrm{tr}[x, y] = (\mathrm{tr}\, x)^2 + (\mathrm{tr}\, y)^2 + (\mathrm{tr}\, xy)^2 - (\mathrm{tr}\, x)(\mathrm{tr}\, y)(\mathrm{tr}\, xy) - 2$,

$$(4)\quad 2\,\mathrm{tr}\,xyzw = (\mathrm{tr}\,x)(\mathrm{tr}\,yzw)+(\mathrm{tr}\,y)(\mathrm{tr}\,zwx)+(\mathrm{tr}\,z)(\mathrm{tr}\,wxy)+(\mathrm{tr}\,w)(\mathrm{tr}\,xyz)$$
$$+(\mathrm{tr}\,xy)(\mathrm{tr}\,zw)-(\mathrm{tr}\,xz)(\mathrm{tr}\,yw)+(\mathrm{tr}\,xw)(\mathrm{tr}\,yz)$$
$$-(\mathrm{tr}\,x)(\mathrm{tr}\,y)(\mathrm{tr}\,zw)-(\mathrm{tr}\,y)(\mathrm{tr}\,z)(\mathrm{tr}\,xw)-(\mathrm{tr}\,x)(\mathrm{tr}\,w)(\mathrm{tr}\,yz)$$
$$-(\mathrm{tr}\,z)(\mathrm{tr}\,w)(\mathrm{tr}\,xy)+(\mathrm{tr}\,x)(\mathrm{tr}\,y)(\mathrm{tr}\,z)(\mathrm{tr}\,w)$$

for any $x, y, z, w \in F_n$. Equations (2) and (4) are due to Vogt [1889]. (For details, see [Maclachlan and Reid 2003, Section 3.4] for example.) The point of (2) is that $\mathrm{tr}\,yxz$ can be written as a sum of $-\mathrm{tr}\,xyz$ and a polynomial in $\mathrm{tr}\,v$, with v a word in x, y, z of length at most two. Similarly, the point of (4) is that $\mathrm{tr}\,xyzw$ can be written as a polynomial in $\mathrm{tr}\,v$ with v a word in x, y, z, w of length at most three.

Let $\mathfrak{X}_{\mathbb{Q}}(F_n)$ be the $\mathbb{Q}$-vector subspace of $\mathcal{F}(R(F_n),\mathbb{C})$ generated by all $\mathrm{tr}\,x$ for $x \in F_n$. From (1), it turns out that $\mathfrak{X}_{\mathbb{Q}}(F_n)$ is a $\mathbb{Q}$-subalgebra of $\mathcal{F}(R(F_n),\mathbb{C})$. We call $\mathfrak{X}_{\mathbb{Q}}(F_n)$ the ring of Fricke characters of F_n over $\mathbb{Q}$. Let $\mathbb{Q}[t]$ be the rational polynomial ring

$$\mathbb{Q}[t_{i_1\cdots i_l} \mid 1 \le l \le 3,\ 1 \le i_1 < i_2 < \cdots < i_l \le n]$$

of $n+\binom{n}{2}+\binom{n}{3}$ indeterminates. Consider the ring homomorphism

$$\pi : \mathbb{Q}[t] \to \mathcal{F}(R(F_n),\mathbb{C})$$

defined by

$$\pi(1) := \tfrac{1}{2}(\mathrm{tr}\,1_{F_n}), \qquad \pi(t_{i_1\cdots i_l}) := \mathrm{tr}\,x_{i_1}\cdots x_{i_l}.$$

Clearly, $\mathrm{Im}\,\pi \subset \mathfrak{X}_{\mathbb{Q}}(F_n)$. By a classical result due to Horowitz [1972], we have:

Theorem 2.1. *For any $n \ge 2$, the homomorphism $\pi : \mathbb{Q}[t] \to \mathfrak{X}_{\mathbb{Q}}(F_n)$ is surjective.*

More precisely, Horowitz obtained a generating set of the ring of Fricke characters of G over $\mathbb{Z}$. Using this and (4), we can obtain the above theorem. Set

$$I := \mathrm{Ker}\,\pi = \{f \in \mathbb{Q}[t] \mid f(\mathrm{tr}\,\rho(x_{i_1}\cdots x_{i_l})) = 0 \text{ for any } \rho \in R(F_n)\}.$$

Then π induces an isomorphism $\mathbb{Q}[t]/I \to \mathfrak{X}_{\mathbb{Q}}(F_n)$. In this paper, we always identify $\mathbb{Q}[t]/I$ with $\mathfrak{X}_{\mathbb{Q}}(F_n)$ through this isomorphism, and also call $\mathbb{Q}[t]/I$ the ring of Fricke characters of F_n over $\mathbb{Q}$. We define an action of $\mathrm{Aut}\,F_n$ on $\mathbb{Q}[t]/I$ such that the isomorphism $\mathbb{Q}[t]/I \xrightarrow{\cong} \mathfrak{X}_{\mathbb{Q}}(F_n)$ is $\mathrm{Aut}\,F_n$-equivariant.

We remark that the structure of the ideal I is quite complicated in general. It is an open problem to find a generating set for I when $n \ge 4$. Horowitz [1972] showed that $I = (0)$ for $n = 1$ and 2, and that I is the principal ideal generated by a quadratic polynomial

$$t_{123}^2 - P_{123}(t)t_{123} + Q_{123}(t),$$

where

$$P_{abc}(t) := t_{ab}t_c + t_{ac}t_b + t_{bc}t_a - t_a t_b t_c,$$
$$Q_{abc}(t) := t_a^2 + t_b^2 + t_c^2 + t_{ab}^2 + t_{ac}^2 + t_{bc}^2 - t_a t_b t_{ab} - t_a t_c t_{ac} - t_b t_c t_{bc} + t_{ab} t_{bc} t_{ac} - 4.$$

For $n \geq 4$, Whittemore [1973] showed that the ideal I is not principal. However, very little is known further in this case.

Although we can obtain a representation of Aut F_n by the action of Aut F_n on $\mathfrak{X}_{\mathbb{Q}}(F_n) \cong \mathbb{Q}[t]/I$, it is an infinite-degree representation, and hence it is not so easy to handle. In order to construct finite-dimensional representations of Aut F_n, we consider a descending filtration of Aut F_n-invariant ideals of $\mathbb{Q}[t]/I$, and take its graded quotients. Set $t'_{i_1\cdots i_l} := t_{i_1\cdots i_l} - 2 \in \mathbb{Q}[t]$. For simplicity, we also denote by $t'_{i_1\cdots i_l}$ its coset class in $\mathbb{Q}[t]/I$ by abuse of notation. Consider the ideal

$$J := (t'_{i_1\cdots i_l} \mid 1 \leq l \leq 3,\ 1 \leq i_1 < i_2 < \cdots < i_l \leq n) \subset \mathbb{Q}[t]/I$$

of $\mathbb{Q}[t]/I$ generated by all $t'_{i_1\cdots i_l}$s. Then, we have a descending filtration

$$J \supset J^2 \supset J^3 \supset \cdots$$

of Aut F_n-invariant ideals of $\mathbb{Q}[t]/I$. Set $\operatorname{gr}^k J := J^k/J^{k+1}$ for each $k \geq 1$. Then each $\operatorname{gr}^k J$ is an Aut F_n-invariant finite-dimensional $\mathbb{Q}$-vector space. In order to describe the action of Aut F_n on $\operatorname{gr}^k J$ precisely, we have to find a basis of it. In general, however, by combinatorial complexities, it is quite a difficult problem. In [HS 2014], we explicitly gave bases of $\operatorname{gr}^k J$ for $k = 1$ and 2 (see Section 4). For $k \geq 3$, even the dimension of $\operatorname{gr}^k J$ is not determined.

In [Hatakenaka and Satoh 2015], we studied the rings of Fricke characters of free abelian groups by a parallel argument as above. We review that work briefly. Let H be the abelianization of F_n. The (coset classes of) $x_1, x_2, \ldots, x_n$ form a basis of H. Let $\mathfrak{X}_{\mathbb{Q}}(H)$ be the ring of Fricke characters of H over $\mathbb{Q}$. Since

$$2\operatorname{tr} xyz = (\operatorname{tr} x)(\operatorname{tr} yz) + (\operatorname{tr} y)(\operatorname{tr} xz) + (\operatorname{tr} z)(\operatorname{tr} xy) - (\operatorname{tr} x)(\operatorname{tr} y)(\operatorname{tr} z)$$

in $\mathfrak{X}_{\mathbb{Q}}(H)$ for any $x, y, z \in H$, by using Horowitz's result as mentioned above, we can see that $\mathfrak{X}_{\mathbb{Q}}(H)$ is generated by $\operatorname{tr} x_i$ for $1 \leq i \leq n$ and $\operatorname{tr} x_i x_j$ for $1 \leq i < j \leq n$. Let J_H be the ideal generated by all $\operatorname{tr}' x_i := \operatorname{tr} x_i - 2$ for $1 \leq i \leq n$ and $\operatorname{tr}' x_i x_j := \operatorname{tr} x_i x_j - 2$ for $1 \leq i < j \leq n$. Then we have a descending filtration $J_H \supset J_H^2 \supset \cdots$. We obtained a basis of the graded quotient $\operatorname{gr}^k J_H$ as a $\mathbb{Q}$-vector space for any $k \geq 1$:

Theorem 2.2 [Hatakenaka and Satoh 2015]. *For any $n \geq 2$ and $k \geq 1$,*

$$\bigcup_{l=0}^{k} \Big\{ (\operatorname{tr}' x_{p_1} x_{q_1}) \cdots (\operatorname{tr}' x_{p_l} x_{q_l})(\operatorname{tr}' x_{i_{l+1}}) \cdots (\operatorname{tr}' x_{i_k}) \\ \Big|\ 1 \leq p_1 < q_1 < \cdots < p_l < q_l \leq n,\ 1 \leq i_{l+1} \leq \cdots \leq i_k \leq n \Big\}$$

is a basis of $\operatorname{gr}^k J_H$.

As a corollary, we obtain a lower bound on the dimension of $\mathrm{gr}^k J$ by

$$\dim_{\mathbb{Q}}(\mathrm{gr}^k J) \geq \dim_{\mathbb{Q}}(\mathrm{gr}^k J_H) = \sum_{l=0}^{k} \binom{n}{2l} \binom{n+k-l-1}{k-l}$$

through the surjective homomorphism $\mathrm{gr}^k J \to \mathrm{gr}^k J_H$ induced from the abelianization $F_n \to H$.

2C. *Johnson homomorphisms of* Aut F_n. In this subsection, we review the Johnson homomorphisms of $\mathrm{Aut}\, F_n$ (for details, see [Satoh 2013b], for example). Let $\rho : \mathrm{Aut}\, F_n \to \mathrm{Aut}\, H$ be the natural homomorphism induced from the abelianization of F_n. We identify $\mathrm{Aut}\, H$ with the general linear group $\mathrm{GL}(n, \mathbb{Z})$ by fixing the basis $x_1, \ldots, x_n$ of H. The kernel IA_n of ρ is called the IA-automorphism group of F_n. Let $\Gamma_n(1) \supset \Gamma_n(2) \supset \cdots$ be the lower central series of a free group F_n defined by the rule

$$\Gamma_n(1) := F_n, \qquad \Gamma_n(k) := [\Gamma_n(k-1), F_n], \quad k \geq 2.$$

We denote by $\mathcal{L}_n(k) := \Gamma_n(k)/\Gamma_n(k+1)$ the graded quotient of the lower central series of F_n. For each $k \geq 1$, the natural action of $\mathrm{Aut}\, F_n$ on $\mathcal{L}_n(k)$ induces that of $\mathrm{GL}(n, \mathbb{Z})$ since IA_n acts on $\mathcal{L}_n(k)$ trivially.

For each $k \geq 1$, the action of $\mathrm{Aut}\, F_n$ on the nilpotent quotient group $F_n/\Gamma_n(k+1)$ induces a homomorphism

$$\mathrm{Aut}\, F_n \to \mathrm{Aut}(F_n/\Gamma_n(k+1)).$$

We denote its kernel by $\mathcal{A}_n(k)$. Then the groups $\mathcal{A}_n(k)$ define a descending central filtration

$$\mathrm{IA}_n = \mathcal{A}_n(1) \supset \mathcal{A}_n(2) \supset \mathcal{A}_n(3) \supset \cdots$$

of IA_n. We call this the Andreadakis–Johnson filtration of $\mathrm{Aut}\, F_n$; historically, it was introduced by Andreadakis [1965], who showed:

Theorem 2.3. (1) *For any $k, l \geq 1$, $\sigma \in \mathcal{A}_n(k)$ and $x \in \Gamma_n(l)$, $x^{-1}x^{\sigma} \in \Gamma_n(k+l)$.*

(2) *For any k and $l \geq 1$, $[\mathcal{A}_n(k), \mathcal{A}_n(l)] \subset \mathcal{A}_n(k+l)$.*

(3) $\bigcap_{k \geq 1} \mathcal{A}_n(k) = 1$.

In the 1980s, Dennis Johnson studied this type of filtration for the mapping class group of a surface, and this became known as the Johnson filtration of the mapping class group.

For each $k \geq 1$, the group $\mathrm{Aut}\, F_n$ acts on $\mathcal{A}_n(k)$ by conjugation, and it naturally induces an action of $\mathrm{GL}(n, \mathbb{Z}) = \mathrm{Aut}\, F_n/\mathrm{IA}_n$ on the graded quotients $\mathrm{gr}^k\mathcal{A}_n := \mathcal{A}_n(k)/\mathcal{A}_n(k+1)$ by Theorem 2.3(2). In order to study the $\mathrm{GL}(n, \mathbb{Z})$-module structure of $\mathrm{gr}^k \mathcal{A}_n$, we consider the Johnson homomorphisms of Aut F_n. For each $k \geq 1$, define a homomorphism $\tilde{\tau}_k : \mathcal{A}_n(k) \to \mathrm{Hom}_{\mathbb{Z}}(H, \mathcal{L}_n(k+1))$ by

$$\sigma \mapsto (x \mapsto x^{-1} x^{\sigma}), \quad x \in H.$$

Then the kernel of $\tilde{\tau}_k$ is just $\mathcal{A}_n(k+1)$. Hence it induces an injective homomorphism

$$\tau_k : \mathrm{gr}^k \mathcal{A}_n \hookrightarrow \mathrm{Hom}_{\mathbb{Z}}(H, \mathcal{L}_n(k+1)) = H^* \otimes_{\mathbb{Z}} \mathcal{L}_n(k+1).$$

The homomorphisms $\tilde{\tau}_k$ and τ_k are called the k-th Johnson homomorphisms of Aut F_n. We can see each τ_k is $\mathrm{GL}(n, \mathbb{Z})$-equivariant.

Now, we have a question to ask whether the first Johnson homomorphism can be extended to Aut F_n or not. Such study and results were given by Morita [1993b] for the mapping class group of a surface, and by Kawazumi [2006] for the automorphism group of a free group as mentioned in the introduction. In particular, Kawazumi showed the first rational Johnson homomorphism

$$\tilde{\tau}_1 : \mathrm{IA}_n \to H^* \otimes_{\mathbb{Z}} \Lambda^2 H \to (H^* \otimes_{\mathbb{Z}} \Lambda^2 H) \otimes_{\mathbb{Z}} \mathbb{Q}$$

can be extended to Aut F_n as a crossed homomorphism. In [Satoh 2013a], we showed that

$$H^1(\mathrm{Aut}\, F_n, (H^* \otimes_{\mathbb{Z}} \Lambda^2 H) \otimes_{\mathbb{Z}} \mathbb{Q}) \cong \mathbb{Q}^{\oplus 2},$$

and that the extension of the first Johnson homomorphism and a crossed homomorphism obtained from the Magnus representation form a basis of the above first cohomology group. These results are free group analogues of Morita's work [1993b] for the mapping class group of a surface.

2D. *Homomorphisms η_k.* In [HS 2014], we introduced homomorphisms η_k which are Fricke character analogue of the Johnson homomorphisms τ_k. Here we review the definition of η_k.

For any $k \geq 1$, let

$$\mathcal{E}_n(k) := \mathrm{Ker}(\mathrm{Aut}\, F_n \to \mathrm{Aut}(J/J^{k+1}))$$

be the kernel of a homomorphism Aut $F_n \to \mathrm{Aut}(J/J^{k+1})$ induced from the action of Aut F_n on J/J^{k+1}. Then the groups $\mathcal{E}_n(k)$ define a descending filtration

$$\mathcal{E}_n(1) \supset \mathcal{E}_n(2) \supset \cdots \supset \mathcal{E}_n(k) \supset \cdots$$

of Aut F_n.

Theorem 2.4 [HS 2014]. *For any $n \geq 3$, we have*:

(1) *For any $k, l \geq 1$, $[\mathcal{E}_n(k), \mathcal{E}_n(l)] \subset \mathcal{E}_n(k+l)$.*

(2) $\mathcal{E}_n(1) = \mathrm{Inn}\, F_n \cdot \mathcal{A}_n(2)$, *where* Inn F_n *is the inner automorphism group of F_n.*

(3) *For any $k \geq 1$, $\mathcal{A}_n(2k) \subset \mathcal{E}_n(k)$.*

From Theorem 2.4(1), the graded quotients $\mathrm{gr}^k\mathcal{E}_n := \mathcal{E}_n(k)/\mathcal{E}_n(k+1)$ are abelian groups for any $k \geq 1$. Since each $\mathcal{E}_n(k)$ is a normal subgroup of $\mathrm{Aut}\,F_n$, the group $\mathrm{Aut}\,F_n$ naturally acts on $\mathrm{gr}^k\mathcal{E}_n$ by the conjugation from the right. Namely, for any $\sigma \in \mathrm{Aut}\,F_n$ and $\tau \in \mathcal{E}_n(k)$, the action of σ on $\tau \pmod{\mathcal{E}_n(k+1)}$ is given by

$$(\tau \pmod{\mathcal{E}_n(k+1)}) \cdot \sigma := \sigma^{-1}\tau\sigma \pmod{\mathcal{E}_n(k+1)}.$$

Furthermore, since $\{\mathcal{E}_n(k)\}$ is a central filtration, the action of $\mathcal{E}_n(1)$ on $\mathrm{gr}^k\mathcal{E}_n$ is trivial. Hence we can consider each $\mathrm{gr}^k\mathcal{E}_n$ as an $\mathrm{Aut}\,F_n/\mathcal{E}_n(1)$-module. In order to investigate the $\mathrm{Aut}\,F_n/\mathcal{E}_n(1)$-module structure of $\mathrm{gr}^k\mathcal{E}_n$, we have introduced

$$\eta_k : \mathrm{gr}^k\mathcal{E}_n \to \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^{k+1} J)$$

defined by

$$\sigma \pmod{\mathcal{E}_n(k+1)} \mapsto \big(f \pmod{J^2} \mapsto f^{\sigma} - f \pmod{J^{k+1}}\big).$$

In [HS 2014], we showed that each η_k is an $\mathrm{Aut}\,F_n/\mathcal{E}_n(1)$-equivariant injective homomorphism. However, the structure of the image of η_k is not well-understood even in the case where $k = 1$.

3. An extension of η_1 as a crossed homomorphism

In this section, we extend the homomorphism

$$\tilde{\eta}_1 : \mathcal{E}_n(1) \to \mathrm{gr}^1\mathcal{E}_n \xrightarrow{\eta_1} \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$$

to $\mathrm{Aut}\,F_n$ as a crossed homomorphism, following [Morita 1993b]. Furthermore, according to the usual convention in homological algebra, for any group G and G-module M, we consider that G acts on M from the left unless otherwise noted. Hence, the right actions mentioned above are read as left actions in the natural way. For example, for any $\sigma \in \mathrm{Aut}\,F_n$ and $x \in F_n$, the left action of σ on the Fricke character $\mathrm{tr}\,x$ is given by

$$\sigma \cdot (\mathrm{tr}\,x) = \mathrm{tr}\,x^{\sigma^{-1}}.$$

For basic materials for cohomology of associative algebras, see [Cartan and Eilenberg 1999, Chapter IX], for example.

First, consider an extension

$$0 \to J^2/J^3 \to \mathfrak{X}_{\mathbb{Q}}(F_n)/J^3 \to \mathfrak{X}_{\mathbb{Q}}(F_n)/J^2 \to 0 \tag{5}$$

of associative $\mathbb{Q}$-algebras. For an associative ring R, we denote by $\mathrm{Aut}_{(\mathrm{Ring})}(R)$ the ring automorphism group of R. For a $\mathbb{Q}$-vector space M, we denote by $\mathrm{Aut}(M)$ the $\mathbb{Q}$-linear automorphism group of M.

Proposition 3.1. *The natural homomorphism*

$$\Phi : \mathrm{Aut}_{(\mathrm{Ring})}(\mathfrak{X}_{\mathbb{Q}}(F_n)/J^3) \to \mathrm{Aut}_{(\mathrm{Ring})}(\mathfrak{X}_{\mathbb{Q}}(F_n)/J^2)$$

is surjective.

Proof. First, observe the characteristic class $\theta(\mathrm{id}_{J/J^2})$ of the extension (5), where

$$\theta : \mathrm{Hom}_{E(\mathfrak{X}_{\mathbb{Q}}(F_n)/J^2)}(J^2/J^3, J^2/J^3) \to H^2(\mathfrak{X}_{\mathbb{Q}}(F_n)/J^2, J^2/J^3)$$

is the connecting homomorphism, and $E(\mathfrak{X}_{\mathbb{Q}}(F_n)/J^2)$ is the enveloping algebra of $\mathfrak{X}_{\mathbb{Q}}(F_n)/J^2$. Choose a 2-cocycle c of the algebra $\mathfrak{X}_{\mathbb{Q}}(F_n)/J^2$ with coefficients in J^2/J^3, which represents the cohomology class $\theta(\mathrm{id}_{J/J^2})$. Then $\mathfrak{X}_{\mathbb{Q}}(F_n)/J^3$ can be explicitly described as the product $J^2/J^3 \times \mathfrak{X}_{\mathbb{Q}}(F_n)/J^2$ equipped with the multiplication given by

$$(\xi, \tau)(\xi', \tau') = (\xi\tau' + \tau\xi' + c(\tau, \tau'), \tau\tau')$$

for any $\xi, \xi' \in J^2/J^3$ and $\tau, \tau' \in \mathfrak{X}_{\mathbb{Q}}(F_n)/J^2$. We denote by Λ_c this associative algebra.

For any $\alpha \in \mathrm{Aut}_{(\mathrm{Ring})}(\mathfrak{X}_{\mathbb{Q}}(F_n)/J^2)$, since the cohomology class of c is the characteristic class induced from id_{J/J^2}, the 2-cocycle $\alpha^\sharp(c)$ should be cohomologous to c where $\alpha^\sharp$ is the induced homomorphism from α. Hence there exists a 1-chain $d : \mathfrak{X}_{\mathbb{Q}}(F_n)/J^2 \to J^2/J^3$ such that $\alpha^\sharp(c) - c = \delta d$, where δ is the coboundary homomorphism. Namely, we have

$$\alpha^{-1} \cdot c(\alpha(\tau), \alpha(\tau')) - c(\tau, \tau') = \tau d(\tau') - d(\tau\tau') + d(\tau)\tau'$$

for any $\tau, \tau' \in \mathfrak{X}_{\mathbb{Q}}(F_n)/J^2$. Define the map $\tilde{\alpha} : \Lambda_c \to \Lambda_c$ to be

$$(\xi, \tau) \mapsto (\alpha(\xi) - \alpha(d(\tau)), \alpha(\tau)).$$

Then $\tilde{\alpha} \in \mathrm{Aut}_{(\mathrm{Ring})}(\mathfrak{X}_{\mathbb{Q}}(F_n)/J^3)$ under the identification $\Lambda_c = \mathfrak{X}_{\mathbb{Q}}(F_n)/J^3$, and $\Phi(\tilde{\alpha}) = \alpha$. □

For any $f \in J$, we denote the coset class of f in J/J^k by $[f]_k$. For any $k \geq 2$, set

$$\overline{\mathrm{Aut}}(J/J^k) := \{\sigma \in \mathrm{Aut}(J/J^k) \mid \sigma([\gamma\gamma']_k) = \sigma([\gamma]_k)\,\sigma([\gamma']_k),\ \gamma, \gamma' \in J\}.$$

Note that $\overline{\mathrm{Aut}}(J/J^2) = \mathrm{Aut}(J/J^2) = \mathrm{GL}_{\mathbb{Q}}(J/J^2)$.

Lemma 3.2. *The group homomorphism*

$$\Psi : \mathrm{Aut}_{(\mathrm{Ring})}(\mathfrak{X}_{\mathbb{Q}}(F_n)/J^k) \to \overline{\mathrm{Aut}}(J/J^k),$$

defined by $\sigma \mapsto \sigma|_{J/J^k}$, *is an isomorphism.*

Proof. Consider the polynomial ring

$$\mathbb{Q}[t'] := \mathbb{Q}[t'_{i_1\cdots i_l} \mid 1 \le l \le 3,\ 1 \le i_1 < i_2 < \cdots < i_l \le n]$$

with indeterminates $t'_{i_1\cdots i_l}$, and the natural surjection $\pi' : \mathbb{Q}[t'] \to \mathfrak{X}_{\mathbb{Q}}(F_n)$ given by $1 \mapsto \frac{1}{2}(\mathrm{tr}\, 1_{F_n})$ and $t'_{i_1\cdots i_l} \mapsto \mathrm{tr}'\, x_{i_1} \cdots x_{i_l}$. The kernel I' of π' is contained in the ideal J' generated by all $t'_{i_1\cdots i_l}$.

We construct the inverse map of Ψ. For any $\beta \in \overline{\mathrm{Aut}}(J/J^k)$, define the $\mathbb{Q}$-algebra homomorphism $\tilde{\beta} : \mathbb{Q}[t'] \to \mathfrak{X}_{\mathbb{Q}}(F_n)/J^k$ to be

$$\tilde{\beta}(1) := \left[\tfrac{1}{2}(\mathrm{tr}\, 1_{F_n})\right]_k, \quad \tilde{\beta}(t'_{i_1\cdots i_l}) := \beta([\pi'(t'_{i_1\cdots i_l})]_k).$$

Since $I' \subset \mathrm{Ker}\, \tilde{\beta}$ and $\tilde{\beta}(J') = J$, the above $\tilde{\beta}$ induces a ring homomorphism $\mathfrak{X}_{\mathbb{Q}}(F_n)/J^k \to \mathfrak{X}_{\mathbb{Q}}(F_n)/J^k$, say $\tilde{\beta}$, by abuse of notation. Since β is an automorphism, so is $\tilde{\beta}$. Then we have the homomorphism $\Psi' : \overline{\mathrm{Aut}}(J/J^k) \to \mathrm{Aut}(\mathfrak{X}_{\mathbb{Q}}(F_n)/J^k)$ defined by $\beta \mapsto \tilde{\beta}$, and see that Ψ' is the inverse of Ψ. □

From Proposition 3.1 and Lemma 3.2, we obtain the induced surjective homomorphism

$$\varphi : \overline{\mathrm{Aut}}(J/J^3) \to \overline{\mathrm{Aut}}(J/J^2).$$

Next, we consider an embedding of $\mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$ into $\overline{\mathrm{Aut}}(J/J^3)$. For any $f \in \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$, define the map $\tilde{f} : J/J^3 \to J/J^3$ by

$$\tilde{f}([\gamma]_3) := [\gamma]_3 + f([\gamma]_2)$$

for any $\gamma \in J$.

Proposition 3.3. *With the above notation, for any $f \in \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$, we see $\tilde{f} \in \overline{\mathrm{Aut}}(J/J^3)$, and the map*

$$\iota : \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J) \to \overline{\mathrm{Aut}}(J/J^3),$$

defined by $f \mapsto \tilde{f}$, is injective.

Proof. First, we show $\tilde{f}$ is a homomorphism. For any $\gamma, \gamma' \in J$,

$$\begin{aligned}\tilde{f}([\gamma]_3 + [\gamma']_3) &= \tilde{f}([\gamma + \gamma']_3) = [\gamma + \gamma']_3 + f([\gamma + \gamma']_2)\\ &= ([\gamma]_3 + f([\gamma]_2)) + ([\gamma']_3 + f([\gamma']_2))\\ &= \tilde{f}([\gamma]_3) + \tilde{f}([\gamma']_3).\end{aligned}$$

Thus $\tilde{f}$ is a homomorphism. Furthermore, $\tilde{f}$ satisfies

$$\begin{aligned}\tilde{f}([\gamma]_3[\gamma']_3) &= \tilde{f}([\gamma\gamma']_3) = [\gamma\gamma']_3 + f([\gamma\gamma']_2) = [\gamma]_3[\gamma']_3\\ &= ([\gamma]_3 + f([\gamma]_2))([\gamma']_3 + f([\gamma']_2)) = \tilde{f}([\gamma]_3)\tilde{f}([\gamma']_3)\end{aligned}$$

for any $\gamma, \gamma' \in J$. On the other hand, we have $\widetilde{f+g} = \tilde{f} \circ \tilde{g}$ for any $f, g \in \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$. In fact, for any $\gamma \in J$,

$$\begin{aligned}
(\widetilde{f+g})([\gamma]_3) &= [\gamma]_3 + f([\gamma]_2) + g([\gamma]_2), \\
(\tilde{f} \circ \tilde{g})([\gamma]_3) &= \tilde{f}([\gamma]_3 + g([\gamma]_2)) = [\gamma]_3 + g([\gamma]_2) + f([\gamma]_2).
\end{aligned}$$

This shows that ι is a homomorphism. On the other hand, for the zero map $0 \in \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$, it is obvious that $\tilde{0} = \mathrm{id}_{J/J^3}$. Hence each $\tilde{f}$ has its inverse map $\tilde{f}^{-1} = \widetilde{-f}$. This means $\tilde{f}$ an automorphism on J/J^3.

Finally, we show that ι is injective. Assume that $\tilde{f} = \mathrm{id}_{J/J^3}$ for some $f \in \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$. Then for any $[\gamma]_3 \in J/J^3$, we have

$$\tilde{f}([\gamma]_3) = [\gamma]_3 + f([\gamma]_2) = [\gamma]_3.$$

Hence $f([\gamma]_2) = 0$ for any $[\gamma]_2 \in J/J^2$, and $f = 0$. This shows ι is injective. $\square$

Proposition 3.4. *The sequence*

$$0 \to \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J) \xrightarrow{\iota} \overline{\mathrm{Aut}}(J/J^3) \xrightarrow{\varphi} \overline{\mathrm{Aut}}(J/J^2) \to 1 \tag{6}$$

is a split group extension.

Proof. First, we show the above sequence is exact. Namely, it suffices to show $\mathrm{Im}\,\iota = \mathrm{Ker}\,\varphi$. The fact that $\mathrm{Im}\,\iota \subset \mathrm{Ker}\,\varphi$ follows from

$$\begin{aligned}
(\varphi \circ \iota)(f)([\gamma]_2) &= \varphi(\tilde{f})([\gamma]_2) = [\gamma]_3 + f([\gamma]_2) \pmod{J^2} \\
&= [\gamma]_2
\end{aligned}$$

for any $f \in \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$ and $\gamma \in J$. To show $\mathrm{Im}\,\iota \supset \mathrm{Ker}\,\varphi$, take any $f \in \mathrm{Ker}\,\varphi$. Define $g \in \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$ to be

$$g([\gamma]_2) := f([\gamma]_3) - [\gamma]_3$$

for any $\gamma \in J$. The map g is well-defined. In fact, for any $\gamma, \gamma' \in J$ such that $[\gamma]_2 = [\gamma']_2$, if we set $\gamma' - \gamma = \varepsilon \in J^2$, then we have

$$\begin{aligned}
g([\gamma']_2) &= f([\gamma']_3) - [\gamma']_3 = f([\gamma + \varepsilon]_3) - [\gamma + \varepsilon]_3 \\
&= f([\gamma]_3) - [\gamma]_3 + f([\varepsilon]_3) - [\varepsilon]_3 = f([\gamma]_3) - [\gamma]_3 \\
&= g([\gamma]_2).
\end{aligned}$$

Here we remark that $f([\varepsilon]_3) - [\varepsilon]_3 = 0 \in J^2/J^3$ since $\varepsilon \in J^2$ and $f \in \mathrm{Ker}\,\varphi$. It is easy to show that g is a homomorphism. Furthermore, for any $\gamma \in J$,

$$\tilde{g}([\gamma]_3) = [\gamma]_3 + g([\gamma]_2) = f([\gamma]_3).$$

This shows $f = \tilde{g} = \iota(g) \in \mathrm{Im}\,\iota$.

Finally, we construct the section $s : \overline{\mathrm{Aut}}(J/J^2) \to \overline{\mathrm{Aut}}(J/J^3)$ of (6). Take elements $\gamma_1, \gamma_2, \dots, \gamma_p \in J$, $\gamma_{p+1}, \dots, \gamma_{p+q} \in J^2$ such that $([\gamma_1]_2, [\gamma_2]_2, \dots, [\gamma_p]_2)$ and $([\gamma_{p+1}]_3, \dots, [\gamma_{p+q}]_3)$ form bases of $\mathrm{gr}^1 J$ and $\mathrm{gr}^2 J$, respectively. Then $([\gamma_1]_3, [\gamma_2]_3, \dots, [\gamma_{p+q}]_3)$ is a basis of J/J^3.

For any $\beta \in \overline{\mathrm{Aut}}(J/J^2)$, there exists an element $\tilde{\beta} \in \overline{\mathrm{Aut}}(J/J^3)$ such that $\varphi(\tilde{\beta}) = \beta$. In general, for any $1 \le j \le p$, the image $\tilde{\beta}([\gamma_j]_3)$ can be written as

$$\tilde{\beta}([\gamma_j]_3) = a_{1j}[\gamma_1]_3 + \cdots + a_{pj}[\gamma_p]_3 + a_{p+1,j}[\gamma_{p+1}]_3 + \cdots + a_{p+q,j}[\gamma_{p+q}]_3$$

for some $a_{ij} \in \mathbb{Q}$. Since $\beta \in \mathrm{Aut}(J/J^2)$, if for any $1 \le j \le p$ we set

$$v_j := a_{1j}[\gamma_1]_2 + \cdots + a_{pj}[\gamma_p]_2,$$

then $(v_1, v_2, \dots, v_p)$ is a basis of $\mathrm{gr}^1 J$. Let $\delta = \delta_{\tilde{\beta}} : \mathrm{gr}^1 J \to \mathrm{gr}^2 J$ be the $\mathbb{Q}$-linear map given by

$$\delta(v_j) = -(a_{p+1,j}[\gamma_{p+1}]_3 + \cdots + a_{p+q,j}[\gamma_{p+q}]_3)$$

for any $1 \le j \le p$. Then we obtain

$$\begin{aligned}
&(\tilde{\delta} \circ \tilde{\beta})([\gamma_j]_3) \\
&\quad = \tilde{\delta}(a_{1j}[\gamma_1]_3 + \cdots + a_{pj}[\gamma_p]_3 + a_{p+1,j}[\gamma_{p+1}]_3 + \cdots + a_{p+q,j}[\gamma_{p+q}]_3) \\
&\quad = a_{1j}[\gamma_1]_3 + \cdots + a_{pj}[\gamma_p]_3 + a_{p+1,j}[\gamma_{p+1}]_3 + \cdots + a_{p+q,j}[\gamma_{p+q}]_3 + \delta(v_j) \\
&\quad = a_{1j}[\gamma_1]_3 + \cdots + a_{pj}[\gamma_p]_3
\end{aligned}$$

for any $1 \le j \le p$. Consider the map $s : \overline{\mathrm{Aut}}(J/J^2) \to \overline{\mathrm{Aut}}(J/J^3)$ defined by $\beta \mapsto \tilde{\delta} \circ \tilde{\beta}$. We can see that s is a homomorphism and is the required section. Hence the exact sequence (6) splits. □

Now, we construct a crossed homomorphism of $\mathrm{Aut}\, F_n$ which is an extension of $\tilde{\eta}_1$. We provide an easy exercise:

Lemma 3.5. *Let*

$$0 \to K \to G \to N \to 1$$

be a split extension of groups over N with an additive abelian group K. For any g, there exist unique elements $k_g \in K$ and $n_g \in N$ such that $g = k_g n_g$. Then the map $k : G \to K$ defined by $g \mapsto k_g$ is a crossed homomorphism.

By using the above lemma, we obtain:

Theorem 3.6. *There is a crossed homomorphism $\eta : \mathrm{Aut}\, F_n \to \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$ such that the restriction of η to $\mathcal{E}_n(1)$ is $\tilde{\eta}_1$.*

Proof. By applying Lemma 3.5 to the split extension (6), we obtain a crossed homomorphism $k : \overline{\mathrm{Aut}}(J/J^3) \to \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$. Hence by composing k and the natural homomorphism $\mathrm{Aut}\, F_n \to \mathrm{Aut}(J/J^3)$, we obtain a crossed homomorphism

$$\eta : \mathrm{Aut}\, F_n \to \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J).$$

This is the required homomorphism. □

4. Nontriviality of η as a 1-cocycle

In this section, we give a few remarks about the image of η_1 and the nontriviality of η as a 1-cocycle.

The first Johnson homomorphism $\tau_1 : \mathrm{gr}^1 \mathcal{A}_n \to H^* \otimes_{\mathbb{Z}} \mathcal{L}_n(l+1)$ is surjective, and hence an isomorphism. We can easily see this fact by calculating the images of Magnus's generators of IA_n, given by

$$K_{ij} : \begin{cases} x_i \mapsto {x_j}^{-1} x_i x_j, \\ x_t \mapsto x_t \quad (t \neq i), \end{cases}$$

for distinct $i, j \in \{1, 2, \dots, n\}$, and

$$K_{ijk} : \begin{cases} x_i \mapsto x_i x_j x_k {x_j}^{-1} {x_k}^{-1}, \\ x_t \mapsto x_t \quad (t \neq i), \end{cases}$$

for distinct $i, j, k \in \{1, 2, \dots, n\}$ such that $j < k$. From a viewpoint of a comparative study, it is a natural problem for us to determine the image of η_1. However, this is complicated for two reasons. One is that we do not have any generating set of $\mathcal{E}_n(1)$. From our result $\mathcal{E}_n(1) = \mathrm{Inn}\, F_n \cdot \mathcal{A}_n(2)$, it suffices to obtain a generating set of $\mathcal{A}_n(2)$. This seems, however, quite difficult in general. The other is that the basis of $\mathrm{gr}^2 J$ obtained in [HS 2014] is too lengthy to handle. In fact, consider

$$T_1 := \{t'_i \mid 1 \le i \le n\} \cup \{t'_{ij} \mid 1 \le i < j \le n\} \cup \{t'_{ijk} \mid 1 \le i < j < k \le n\} \subset J$$

and

$$\begin{aligned} T_2 := {}& \{t'_i t'_j \mid 1 \le i \le j \le n\} \cup \{t'_i t'_{ab} \mid 1 \le i \le n,\ 1 \le a < b \le n\} \\ & \cup \{t'_i t'_{abc} \mid 1 \le i \le n,\ 1 \le a < b < c \le n\} \\ & \cup \{t'_{ij} t'_{ab} \mid 1 \le i < j \le n,\ 1 \le a < b \le n,\ (i, j) \le (a, b)\} \\ & \cup \{t'_{ab} t'_{abc}, t'_{ac} t'_{abc}, t'_{bc} t'_{abc} \mid 1 \le a < b < c \le n\} \\ & \cup \{t'_{ia} t'_{abc}, t'_{ib} t'_{abc}, t'_{ic} t'_{abc}, t'_{ia} t'_{ibc}, t'_{ab} t'_{iac}, t'_{ab} t'_{ibc}, t'_{ac} t'_{ibc}, t'_{ib} t'_{iac} \mid \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 1 \le i < a < b < c \le n\} \\ & \cup \{t'_{ja} t'_{ibc}, t'_{jb} t'_{iac}, t'_{jc} t'_{iab}, t'_{ab} t'_{ijc}, t'_{ac} t'_{ijb}, t'_{bc} t'_{ija} \mid 1 \le i < j < a < b < c \le n\} \\ & \subset J^2. \end{aligned}$$

We showed in [HS 2014] that the T_k form a basis of $\mathrm{gr}^k J$ for $k = 1, 2$. We cannot write down the total image of η_1 explicitly by hand, because T_2 consists of too many monomials.

Finally, we remark on the nontriviality of the coset class of η in the first cohomology group of $\mathrm{Aut}\, F_n$ with coefficients in $\mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)$. Consider the automorphism $\sigma := [K_{21}, K_{23}] \in \mathcal{A}_n(2) \subset \mathcal{E}_n(1)$. It satisfies

$$x_i^\sigma = \begin{cases} [x_1^{-1}, x_3^{-1}]x_2[x_3^{-1}, x_1^{-1}] & \text{if } i = 2, \\ x_i & \text{if } i \neq 2. \end{cases}$$

Hence we see that σ maps $t'_{123} \in \mathbb{Q}[t]/I$ to

$$\begin{aligned} t'_{123} - 2(t'_1)^2 - 4t'_1t'_2 + 4t'_2t'_3 + 2(t'_3)^2 + 2t'_{12}t'_1 - 2t'_{12}t'_3 + 6t'_{13}t'_1 + 6t'_{13}t'_2 \\ + 2t'_{13}t'_3 + 2t'_{23}t'_1 - 2t'_{23}t'_3 - 4(t'_{13})^2 - 6t'_{12}t'_{13} - 2t'_{13}t'_{23} + 4t'_{13}t'_{123} \end{aligned}$$

modulo J^3 by a hand calculation. By using the basis T_2 of $\mathrm{gr}^2 J$, we can see that

$$\tilde{\eta}_1(\sigma)(t'_{123}) = (t'_{123})^\sigma - t'_{123} \neq 0 \in \mathrm{gr}^2 J.$$

Thus the restriction $\tilde{\eta}_1$ of η to $\mathcal{E}_n(1)$ is a nontrivial homomorphism, and so is the cohomology class of η in

$$H^1(\mathrm{Aut}\, F_n, \mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J)).$$

It seems to be a natural and interesting problem to determine the above first cohomology group, and show whether η generates it or not.

5. The image of the crossed homomorphism η

In order to calculate the image of η, it is important to know $\eta(\sigma)$ with σ generators of $\sigma \in \mathrm{Aut}\, F_n$. In this section, we calculate the images of Nielsen's generators of $\mathrm{Aut}\, F_n$ by the crossed homomorphism η.

For any $k \geq 2$, let $\rho_k : \mathrm{Aut}\, F_n \to \overline{\mathrm{Aut}}(J/J^k)$ be the natural homomorphism induced from the action of $\mathrm{Aut}\, F_n$ on J/J^k. For any $\sigma \in \mathrm{Aut}\, F_n$, by identifying $\overline{\mathrm{Aut}}(J/J^3)$ with the semidirect product $\mathrm{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathrm{gr}^2 J) \rtimes \overline{\mathrm{Aut}}(J/J^2)$, we have

$$\rho_3(\sigma) = (\eta(\sigma), \rho_2(\sigma)) = (\eta(\sigma), 1)(0, \rho_2(\sigma)) = (\eta(\sigma), 1)s(\rho_2(\sigma)) \in \overline{\mathrm{Aut}}(J/J^3),$$

and hence

$$(\eta(\sigma), 1) = \rho_3(\sigma)s(\rho_2(\sigma))^{-1}.$$

From this, in order to compute $\eta(\sigma)$ for any $\sigma \in \mathrm{Aut}\, F_n$, it suffices to compute $\rho_3(\sigma)s(\rho_2(\sigma))^{-1}$. Let $\gamma_1, \gamma_2, \ldots, \gamma_p \in J$ and $\gamma_{p+1}, \ldots, \gamma_{p+q} \in J^2$ be elements such that $([\gamma_1]_2, [\gamma_2]_2, \dots, [\gamma_p]_2)$ and $([\gamma_{p+1}]_3, \dots, [\gamma_{p+q}]_3)$ form bases of $\mathrm{gr}^1 J$ and $\mathrm{gr}^2 J$. By recalling the arguments in the previous section, we see that for any $1 \le j \le p$, if $\rho_3(\sigma)$ satisfies

$$\rho_3(\sigma)([\gamma_j]_3) = a_{1j}[\gamma_1]_3 + \cdots + a_{pj}[\gamma_p]_3 + a_{p+1,j}[\gamma_{p+1}]_3 + \cdots + a_{p+q,j}[\gamma_{p+q}]_3$$

for some $a_{ij} \in \mathbb{Q}$, then

$$s(\rho_2(\sigma))([\gamma_j]_3) = a_{1j}[\gamma_1]_3 + \cdots + a_{pj}[\gamma_p]_3$$

for any $1 \le j \le n$. By using these facts, we can compute the images $\eta(\sigma)$ for any $\sigma \in \operatorname{Aut} F_n$.

Here let us recall Nielsen's generators of $\operatorname{Aut} F_n$. Let P, Q, S and U be the automorphisms of F_n defined as follows:

	x_1	x_2	x_3	$\cdots$	x_{n-1}	x_n
P	x_2	x_1	x_3	$\cdots$	x_{n-1}	x_n
Q	x_2	x_3	x_4	$\cdots$	x_n	x_1
S	x_1^{-1}	x_2	x_3	$\cdots$	x_{n-1}	x_n
U	x_1x_2	x_2	x_3	$\cdots$	x_{n-1}	x_n

Namely, P is the automorphism of F_n induced from the permutation of x_1 and x_2, Q is induced from the cyclic permutation of the basis, and so on. Nielsen [1924] showed that $\operatorname{Aut} F_n$ is generated by P, Q, S and U for any $n \ge 2$, and also gave finitely many relations among the generators P, Q, S and U. This is the first finite presentation for $\operatorname{Aut} F_n$.

By direct computation, we can see that $\rho_3(P)s(\rho_2(P))^{-1}$ satisfies

$$\begin{aligned} t_i' &\mapsto t_i' \quad \text{for any } 1 \le i \le n, \\ t_{ij}' &\mapsto t_{ij}' \quad \text{for any } 1 \le i < j \le n, \\ t_{ijk}' &\mapsto \begin{cases} t_{12k}' - \{t_1't_{2k}' + t_2't_{1k}' + t_k't_{12}' - 2(t_1't_2' + t_1t_k' + t_2t_k')\} & \text{if } (i,j) = (1,2), \\ t_{ijk}' & \text{otherwise,} \end{cases} \end{aligned}$$

and hence obtain

$$\eta(P) = -\sum_{k=3}^{n} (t_{12k}')^* \otimes \{t_1't_{2k}' + t_2't_{1k}' + t_k't_{12}' - 2(t_1't_2' + t_1t_k' + t_2t_k')\},$$

where $(t_\bullet')^*$ means the dual basis in $\operatorname{Hom}_{\mathbb{Q}}(\mathrm{gr}^1 J, \mathbb{Q})$ of $t_\bullet'$ in $\mathrm{gr}^1 J$. Similarly, we can obtain the equalities

$$\eta(Q) = 0,$$

$$\eta(S) = -\sum_{j=2}^{n} (t'_{ij})^* \otimes t'_1 t'_j - \sum_{2 \le j < k \le n} (t'_{1jk})^* \otimes t'_1 t'_{jk},$$

$$\begin{aligned}\eta(U) = &-(t'_{12})^* \otimes t'_1 t'_2 - \sum_{k=3}^{n} (t'_{12k})^* \otimes t'_2 t'_{1k} \\ &+ \sum_{3 \le j < k \le n} (t'_{1jk})^* \otimes \big\{ -(t'_1 t'_k + t'_2 t'_j + t'_j t'_k + 2t'_1 t'_j + 2t'_2 t'_k) \\ &\qquad + (t'_1 t'_{2j} + t'_1 t'_{jk} - t'_2 t'_{1j} + t'_2 t'_{jk} + t'_j t'_{12} + t'_j t'_{1k} + t'_k t'_{12} + t'_k t'_{2j}) \\ &\qquad - \tfrac{1}{2}(t'_1 t'_{2jk} - t'_2 t'_{1jk} + t'_j t'_{12k} + t'_k t'_{12j}) - \tfrac{1}{2}(t'_{12} t'_{jk} - t'_{1j} t'_{2k} + t'_{1k} t'_{2j}) \big\}.\end{aligned}$$

Since P, Q, S and U generate $\mathrm{Aut}\, F_n$, by using the Leibniz rule, we can calculate $\eta(\sigma)$ for any $\sigma \in \mathrm{Aut}\, F_n$.

Acknowledgments

The author would like to thank the referee for the careful reading and helpful comments for the original paper. The author is supported by Grant-in-Aid for Young Scientists (B) by JSPS.

References

[Andreadakis 1965] S. Andreadakis, "On the automorphisms of free groups and free nilpotent groups", *Proc. London Math. Soc.* (3) **15** (1965), 239–268. MR 32 #5746 Zbl 0135.04502

[Cartan and Eilenberg 1999] H. Cartan and S. Eilenberg, *Homological algebra*, Princeton University Press, 1999. MR 2000h:18022 Zbl 0933.18001

[Day 2013] M. B. Day, "Extensions of Johnson's and Morita's homomorphisms that map to finitely generated abelian groups", *J. Topol. Anal.* **5**:1 (2013), 57–85. MR 3043442 Zbl 1270.57006

[Fricke and Klein 1897] R. Fricke and F. Klein, *Vorlesungen über die Theorie der automorphen Funktionen, 1: Die gruppentheoretischen Grundlagen*, Teubner, Leipzig, 1897. MR 32 #1348 JFM 28.0334.01

[Hain 1997] R. Hain, "Infinitesimal presentations of the Torelli groups", *J. Amer. Math. Soc.* **10**:3 (1997), 597–651. MR 97k:14024 Zbl 0915.57001

[Hatakenaka and Satoh 2014] E. Hatakenaka and T. Satoh, "On the graded quotients of the ring of Fricke characters of a free group", 2014. arXiv 1206.1500

[Hatakenaka and Satoh 2015] E. Hatakenaka and T. Satoh, "On the rings of Fricke characters of free abelian groups", *J. Commut. Alg.* (online publication November 2015). To appear.

[Horowitz 1972] R. D. Horowitz, "Characters of free groups represented in the two-dimensional special linear group", *Comm. Pure Appl. Math.* **25** (1972), 635–649. MR 47 #3542 Zbl 1184.20009

[Johnson 1980] D. Johnson, "An abelian quotient of the mapping class group $\mathcal{I}_g$", *Math. Ann.* **249**:3 (1980), 225–242. MR 82a:57008 Zbl 0409.57009

[Johnson 1983] D. Johnson, "The structure of the Torelli group, I: A finite set of generators for $\mathcal{I}$", *Ann. of Math.* (2) **118**:3 (1983), 423–442. MR 85a:57005 Zbl 0549.57006

[Johnson 1985a] D. Johnson, "The structure of the Torelli group, II: A characterization of the group generated by twists on bounding curves", *Topology* **24**:2 (1985), 113–126. MR 86i:57011 Zbl 0571.57009

[Johnson 1985b] D. Johnson, "The structure of the Torelli group, III: The abelianization of $\mathcal{I}$", *Topology* **24**:2 (1985), 127–144. MR 87a:57016 Zbl 0571.57010

[Kawazumi 2006] N. Kawazumi, "Cohomological aspects of Magnus expansions", 2006. arXiv math. GT/0505497

[Maclachlan and Reid 2003] C. Maclachlan and A. W. Reid, *The arithmetic of hyperbolic* 3*-manifolds*, Graduate Texts in Mathematics **219**, Springer, New York, 2003. MR 2004i:57021 Zbl 1025.57001

[Magnus 1980] W. Magnus, "Rings of Fricke characters and automorphism groups of free groups", *Mathematische Zeitschrift* **170**:1 (1980), 91–103.

[Morita 1993a] S. Morita, "Abelian quotients of subgroups of the mapping class group of surfaces", *Duke Math. J.* **70**:3 (1993), 699–726. MR 94d:57003 Zbl 0801.57011

[Morita 1993b] S. Morita, "The extension of Johnson's homomorphism from the Torelli group to the mapping class group", *Invent. Math.* **111**:1 (1993), 197–224. MR 93j:57001 Zbl 0787.57008

[Nielsen 1924] J. Nielsen, "Die Isomorphismengruppe der freien Gruppen", *Math. Ann.* **91**:3-4 (1924), 169–209. MR 1512188 JFM 50.0078.04

[Satoh 2009] T. Satoh, "The cokernel of the Johnson homomorphisms of the automorphism group of a free metabelian group", *Trans. Amer. Math. Soc.* **361**:4 (2009), 2085–2107. MR 2009i:20073 Zbl 1171.20024

[Satoh 2013a] T. Satoh, "First cohomologies and the Johnson homomorphisms of the automorphism group of a free group", *J. Pure Appl. Algebra* **217**:1 (2013), 137–152. MR 2965912 Zbl 1286.20040

[Satoh 2013b] T. Satoh, "A survey of the Johnson homomorphisms of the automorphism groups of free groups and related topics", preprint, 2013. Submitted to *Handbook of Teichmüller theory, volume V*. arXiv 1204.0876

[Vogt 1889] H. Vogt, "Sur les invariants fondamentaux des équations différentielles linéaires du second ordre", *Ann. Sci. École Norm. Sup.* (3) **6** (1889), 3–71. MR 1508833 JFM 21.0314.01

[Whittemore 1973] A. Whittemore, "On special linear characters of free groups of rank $n \geq 4$", *Proc. Amer. Math. Soc.* **40** (1973), 383–388. MR 48 #428 Zbl 0248.20031

Received July 21, 2014. Revised October 24, 2014.

TAKAO SATOH
DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE DIVISION II
TOKYO UNIVERSITY OF SCIENCE
1-3 KAGURAZAKA
SHINJUKU-KU
TOKYO 162-8601
JAPAN

takao@rs.tus.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 275, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.275.463

GLOBAL REPRESENTATIONS OF THE CONFORMAL GROUP AND EIGENSPACES OF THE YAMABE OPERATOR ON $S^1 \times S^n$

MARK R. SEPANSKI AND JOSE A. FRANCO

Using parabolic induction, a global representation of a double cover of the conformal group $\mathrm{SO}(2, n+1)_0$ is constructed. Its space of finite vectors is realized as a direct sum of eigenspaces of the Yamabe operator on $S^1 \times S^n$. The explicit form of the corresponding eigenvalues is obtained. An explicit basis of K-finite eigenvectors is used to study its structure as a representation of the Lie algebra of the conformal group.

1. Introduction

M. Hunziker, M. Sepanski, and R. Stanke [Hunziker et al. 2012] used parabolic induction to construct a representation $I_{m,r}$ of a twofold cover $\widetilde{G}$ of the conformal group $G := \mathrm{SO}(2, n+1)_0$ of $\mathbb{R}^{2,n+1}$ and studied the kernel of a distinguished central element Ω in the universal enveloping algebra of the Lie algebra $\mathfrak{g}$ of $\widetilde{G}$. It was shown that this kernel carries the minimal representation of $\widetilde{G}$ as a positive energy representation and that elements in this kernel correspond to solutions of the wave equation. In this article, we study the structure of $I_{m,r}$ and its compact picture $I''_{m,r}$ as $(\mathfrak{g}_{\mathbb{C}}, \widetilde{K})$-modules, where $\widetilde{K}$ denotes the maximal compact subgroup of $\widetilde{G}$. We generalize the results of [loc. cit.] by considering the space of $\widetilde{K}$-finite vectors $\ker(\Omega-\mu)_{\widetilde{K}}$ of $\ker(\Omega-\mu)$ where $\mu \in \mathbb{R}$. We explicitly determine the conditions on μ such that $\ker(\Omega-\mu)_{\widetilde{K}}$ is nonzero. We show that Ω can be realized as the Yamabe operator $\widetilde{\Delta}_{S^1\times S^n}$ acting on $S^1 \times S^n$ embedded in the Minkowski space $\mathbb{R}^{2,n+1}$. Using this realization, we show that the space of $\widetilde{K}$-finite vectors $(I''_{m,r})_{\widetilde{K}}$ of $I''_{m,r}$ is isomorphic to a direct sum of eigenspaces of the Yamabe operator $\widetilde{\Delta}_{S^1\times S^n}$. An explicit basis of eigenvectors vectors for $(I''_{m,r})_{\widetilde{K}}$ in terms of harmonic and Gegenbauer polynomials is constructed. We show that each of these eigenspaces is invariant under the action of $\widetilde{K}$, but not invariant under the action of all of $\mathfrak{g}$ with the exception of the null eigenspace. While $\mathfrak{g}$ does not preserve each individual eigenspace, it preserves the direct sum of them, $(I''_{m,r})_{\widetilde{K}}$. The zero eigenspace is left invariant under the action of $\mathfrak{g}$ and $(I''_{m,r})_{\widetilde{K}}/(\ker\Omega)_{\widetilde{K}}$ decomposes as the direct

MSC2010: 22E46, 22E70.
Keywords: globalizations, Yamabe operator, conformal Laplace operator, Lie group, conformal group, parabolic induction.

sum of two irreducible representations explicitly identified in Theorem 1. The elements in the zero eigenspace correspond to solutions to the wave equation in the noncompact picture.

1.1. ***Related work.*** This article is motivated by [Hunziker et al. 2012], where the case $\mu = 0$ was studied. This work is similar in spirit to [Franco and Sepanski 2013], which provides a generalization of [Sepanski and Stanke 2010]. Since the minimal representation of conformal groups, the wave equation, and the Yamabe operator are heavily studied topics in mathematics there exists substantial overlap with existing literature. See [Hunziker et al. 2012] for a more extensive set of references. Of particular note, for $n = 3$, some of the results and formulas in this article appear in [Segal et al. 1981] and the references therein. In terms of the minimal representation of the conformal group G, B. Binegar and R. Zierau [1991] realized the minimal representation of $\mathrm{SO}(p,q)_0$ with p, q of the same parity. A very detailed study of the minimal representation of $O(p,q)$ is given by T. Kobayashi and B. Ørsted in [2003a; 2003b; 2003c].

1.2. ***Organization of the work.*** We introduce most of the objects that will be fundamental to the study of our problem in Section 2. In particular, the induced, noncompact, and compact pictures are introduced in this section. Since there is a considerable overlap with existing literature, most of the proofs are omitted. In Section 3 we realize the Yamabe operator on $S^1 \times S^n$ as a central element of the universal enveloping algebra of the Lie algebra of $\widetilde{G}$ and study the invariance of the eigenspaces of this operator. In Section 4 we introduce the space of $\widetilde{K}$-finite vectors in the compact picture and give an explicit basis consisting of eigenvectors of the Yamabe operator. In Section 5 we study the structure of the space of $\widetilde{K}$-finite vectors and close with Theorem 1, where the main results are summarized.

2. Preliminary constructions

This section contains a substantial overlap with [Hunziker et al. 2012] due to the similarity of the problems studied. Therefore, most of this section will be dedicated to a quick survey of the results that will be useful for the study of our problem.

2.1. ***Group and subgroups.*** Let $G = \mathrm{SO}(2, n+1)_0$ and let $\mathfrak{g} := \mathfrak{so}(2, n+1)$ denote its Lie algebra. The group G acts naturally on the space $\mathbb{R}^{2,n+1}$. Since G is a group of linear transformations that preserve the signed quadratic form on $\mathbb{R}^{2,n+1}$, the action of G descends to an action on the subspace

$$C^{2,n+1} := \{(a,b) \in \mathbb{R}^{2,n+1} \mid \|a\| = \|b\| \neq 0\}.$$

$C^{2,n+1}$ is a cone in the sense that it is invariant under the action of $\mathbb{R}^{\times}$. Moreover, since the actions of G and $\mathbb{R}^{\times}$ commute, G acts on the projectivized cone

$\mathbb{P}(C^{2,n+1}) \cong C^{2,n+1}/\mathbb{R}^\times$. For $[0,1,\pm 1,0,\dots,0] \in \mathbb{P}(C^{2,n+1})$, their stabilizers $Q^\pm := \mathrm{Stab}_G([0,1,\pm 1,0,\dots,0])$ are isomorphic to the minimal parabolic subgroups with Langlands decompositions $Q^\pm = MAN^\pm$. The corresponding parabolic subalgebras $\mathfrak{q}^\pm$ have Langlands decomposition $\mathfrak{q}^\pm = \mathfrak{m} \oplus \mathfrak{a} \oplus \mathfrak{n}^\pm$. We describe these subalgebras and subgroups in more detail. The nilpotent subalgebras are

$$\mathfrak{n}^\pm = \{N^\pm_{t,x} \mid (t,x) \in \mathbb{R}^{1,n}\}, \quad \text{where} \quad N^\pm_{t,x} := \left(\begin{array}{ccc|c} 0 & t & \mp t & 0 \\ -t & 0 & 0 & x \\ \mp t & 0 & 0 & \pm x \\ \hline 0 & x^T & \mp x^T & 0_n \end{array}\right),$$

and the maximal abelian subalgebra is

$$\mathfrak{a} = \{H_s \mid s \in \mathbb{R}\}, \quad \text{where} \quad H_s := \left(\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 0 & 0 & s & 0 \\ 0 & s & 0 & 0 \\ \hline 0 & 0 & 0 & 0_n \end{array}\right),$$

and if we denote the $+1$ eigenspace of the Cartan involution on $\mathfrak{g}$ by $\mathfrak{k}$, then the centralizer of $\mathfrak{a}$ in $\mathfrak{k} \cong \mathfrak{so}(2) \times \mathfrak{so}(n+1)$ is

$$\mathfrak{m} = \{L_{A,b} \mid A \in \mathfrak{so}(n) \text{ and } b \in \mathbb{R}^n\}, \quad \text{where} \quad L_{A,b} := \left(\begin{array}{ccc|c} 0 & 0 & 0 & b \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline b^T & 0 & 0 & A \end{array}\right).$$

The corresponding groups in G are

$$N^\pm = \{n^\pm_{t,x} \mid (t,x) \in \mathbb{R}^{1,n}\}, \text{ where } n^\pm_{t,x} := \left(\begin{array}{ccc|c} 1 & t & \mp t & 0 \\ -t & 1+\frac{1}{2}q(t,x) & \mp\frac{1}{2}q(t,x) & x \\ \mp t & \pm\frac{1}{2}q(t,x) & 1-\frac{1}{2}q(t,x) & \pm x \\ \hline 0 & x^T & \mp x^T & I_n \end{array}\right),$$

and

$$A = \{h_s \mid s \in \mathbb{R}\}, \quad \text{where} \quad h_s := \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & \cosh(s) & \sinh(s) & 0 \\ 0 & \sinh(s) & \cosh(s) & 0 \\ \hline 0 & 0 & 0 & I_n \end{array}\right),$$

with $q(t,x) = -t^2 + \|x\|^2$. If $\mathrm{SO}(1,n)_0$ denotes the identity component of $\mathrm{SO}(1,n)$ and $\mathrm{SO}(1,n)_1$ denotes the remaining component, then

$$M = \{m_{\epsilon,Y} \mid \epsilon = \pm 1\}$$

where

$$m_{\epsilon,Y} = \left(\begin{array}{ccc|c} a & 0 & 0 & b \\ 0 & \epsilon & 0 & 0 \\ 0 & 0 & \epsilon & 0 \\ \hline c & 0 & 0 & d \end{array}\right) \text{ with } Y = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \begin{cases} \mathrm{SO}(1,n)_0 & \text{if } \epsilon = +1, \\ \mathrm{SO}(1,n)_1 & \text{if } \epsilon = -1. \end{cases}$$

As in [Hunziker et al. 2012], we will look at the representations induced from a character on Q^-. To construct this character, let M_0 denote the connected component of M where $\epsilon = +1$ and let M_1 denote the component where $\epsilon = -1$. For $g_j \in M_j$ define $\mu_M : M \to \mathbb{C}$ by $\mu_M(g_j) := (-1)^j$. Define $\nu_A : A \to \mathbb{C}$ by $\nu_A(h_s) = e^s$. Then, the family of characters from which we will induce our representations is defined by

$$\chi_{m,r}(q^-) = \nu_A(q_A^-)^r \, \mu_M(q_M^-)^m$$

with $q^- = q_M^- q_A^- q_N^- \in Q^-$, $r \in \mathbb{C}$, and $m \in \mathbb{Z}_2$.

2.2. ***Double cover and induced representations.*** For technical reasons it is necessary to work in a double cover of G that we will denote by $\widetilde{G}$. The maximal compact group of G is $K \cong \mathrm{SO}(2) \times \mathrm{SO}(n+1)$. The double cover $\widetilde{K}$ of K is such that the cover map $\pi : \widetilde{K} \to K$ is given by

$$\pi \begin{pmatrix} R_{\varphi/2} & \\ & u_{n+1} \end{pmatrix} = \begin{pmatrix} R_{\varphi} & \\ & u_{n+1} \end{pmatrix}. \tag{2-1}$$

Up to isomorphism, $\widetilde{K}$ extends uniquely to a group $\widetilde{G}$ that is a double cover of G. Letting $\widetilde{Q}^{\pm}$ denote the parabolic subgroups of $\widetilde{G}$ that cover $Q^{\pm}$, we have that they have Langlands decomposition $\widetilde{Q}^{\pm} = \widetilde{M}\widetilde{A}\widetilde{N}^{\pm}$ and

$$\widetilde{M} \cap \widetilde{K} = \{z_{j,k} \mid k \in O(n), \det k = (-1)^j\} \quad \text{where} \quad z_{j,k} = \left(\begin{array}{cc|c} R_{\pi/2}^j & 0 & 0 \\ 0 & (-1)^j & 0 \\ \hline 0 & 0 & k \end{array}\right).$$

In particular, $\widetilde{M}$ has four connected components. To define a character on $\widetilde{Q}^-$, we define $\gamma_{\widetilde{M}} : \widetilde{M} \to \mathbb{C}$ by

$$\gamma_{\widetilde{M}}\big|_{\widetilde{M}_j} := i^j$$

and $\nu_{\widetilde{A}} : \widetilde{A} \to \mathbb{C}$ by

$$\nu_{\widetilde{A}}(\widetilde{h}_s) = e^s.$$

In [Hunziker et al. 2012], it is shown that the character defined by

$$\widetilde{\chi}_{m,r}(\widetilde{q}^-) := \nu_{\widetilde{A}}(\widetilde{q}_{\widetilde{A}}^-)^r \, \gamma_{\widetilde{M}}(\widetilde{q}_{\widetilde{M}}^-)^m$$

with $\widetilde{q}^- = \widetilde{q}_{\widetilde{M}}^- \widetilde{q}_{\widetilde{A}}^- \widetilde{q}_{\widetilde{N}}^- \in \widetilde{Q}^-$, $r \in \mathbb{C}$, and $m \in \mathbb{Z}_4$, satisfies $\widetilde{\chi}_{m,r} = \chi_{m,r} \circ \pi$.

The representation $\widetilde{\chi}_{m,r}$ of $\widetilde{Q}^-$ is used to induce a representation $\mathrm{Ind}_{Q^-}^{\widetilde{G}}(\widetilde{\chi}_{m,r})$ of $\widetilde{G}$. This representation is defined by

$$\text{(2-2)}\quad \mathrm{Ind}_{Q^-}^{\widetilde{G}}(\widetilde{\chi}_{m,r}) = \{\phi \in C^\infty(\widetilde{G}) \mid \varphi(\widetilde{g}\widetilde{q}^-) = \widetilde{\chi}_{m,r}^{-1}(\widetilde{q}^-)\phi(\widetilde{g}) \text{ for } \widetilde{g} \in \widetilde{G}, \widetilde{q}^- \in \widetilde{Q}^-\}.$$

2.3. ***Noncompact picture.*** For $r \in \mathbb{C}$ and $m \in \mathbb{Z}_4$, define

$$I'_{m,r} := \{f \in C^\infty(\mathbb{R}^{1,n}) \mid f(t,x) = \phi(\widetilde{n}_{t,x}) \text{ for some } \phi \in \mathrm{Ind}_{Q^-}^{\widetilde{G}}(\widetilde{\chi}_{m,r}) \text{ and all } (t,x) \in \mathbb{R}^{1,n}\}.$$

It follows from [Hunziker et al. 2012, Proposition 3.13] that the restriction map is a linear isomorphism and with the appropriate $\widetilde{G}$-action, $I'_{m,r} \cong \mathrm{Ind}_{Q^-}^{\widetilde{G}}(\widetilde{\chi}_{m,r})$ as $\widetilde{G}$-representations.

The action of the group $\widetilde{G}$ and of the corresponding Lie algebra $\mathfrak{g}$ on $I'_{m,r}$ are calculated in [Hunziker et al. 2012, Section 4]. We will record the actions of the Lie algebra for future use.

Proposition 1. *The elements of the Lie algebra $\mathfrak{g}$ act on $I'_{m,r}$ by*

$$\text{(2-3a)}\quad H_s = s(r - t\partial_t - x\partial_x^T),$$

$$\text{(2-3b)}\quad L_{A,b} = -bx^T\partial_t + (xA - tb)\partial_x^T,$$

$$\text{(2-3c)}\quad N_{s,y}^+ = -s\partial_t - y\partial_x^T, \text{ and}$$

$$\text{(2-3d)}\quad N_{s,y}^- = 2(st - yx^T)(r - t\partial_t - x\partial_x^T) - q(t,x)(s\partial_t + y\partial_x^T).$$

Proof. See [Hunziker et al. 2012]. □

A distinguished copy of $\mathfrak{so}(2,1) \cong \mathfrak{sl}_2(\mathbb{R})$ can be embedded in the upper left corner of $\mathfrak{g}$. A standard $\mathfrak{sl}_2$-basis for this Lie algebra is

$$\text{(2-4)}\qquad H := H_2 \qquad E := N_{1,0}^+ \qquad F := N_{1,0}^-.$$

The difference of the Casimir element $\Omega_{\mathrm{SL}(2)}$ corresponding to this copy of $\mathfrak{sl}_2(\mathbb{R})$ with the Casimir element $\Omega_{\mathrm{SO}(n)}$ corresponding to $\mathfrak{so}(n)$ (embedded in the lower right corner) will play a special role in this article. The following corollary follows from Proposition 1.

Corollary 1. *The operator $\Omega_{\mathrm{SL}(2)} - \Omega_{\mathrm{SO}(n)}$ acts by*

$$\Omega_{\mathrm{SL}(2)} - \Omega_{\mathrm{SO}(n)} = \|x\|^2\Box + (1 - n - 2r)\mathscr{E} + r(r+1)$$

on $I'_{m,r}$, where $\Box$ is the wave operator on $\mathbb{R}^{1,n}$ and $\mathscr{E} := \sum_{i=1}^n x_i\partial_{x_i}$ is the Euler operator on $\mathbb{R}^n$.

2.4. ***Compact picture.*** In [Hunziker et al. 2012] it is shown that the space $I''_{m,r}$ of functions $F \in C^\infty(\mathbb{R}\times\mathbb{R}\times S^{n-1})$ that satisfy

$$\begin{aligned} F(\varphi,\theta+\pi,-\hat{x}) &= F(\varphi,\theta,\hat{x}) \\ F(\varphi+\pi,\theta+\pi,\hat{x}) &= i^{-m}F(\varphi,\theta,\hat{x}) \\ F(\varphi,0,\hat{x}) &= F(\varphi,0,\hat{x}') \end{aligned}$$

for all $\varphi,\theta\in R$ and $\hat{x},\hat{x}'\in S^{n-1}$, is a $\widetilde{K}$-representation. Moreover, this representation is isomorphic to $\operatorname{Ind}_{Q^-}^{\widetilde{G}}(\widetilde{\chi}_{m,r})$, hence to $I'_{m,r}$. If $f\in I'_{m,r}$ and $F\in I''_{m,r}$ correspond under the canonical isomorphism between $I'_{m,r}$ and $I''_{m,r}$, then they are related by

$$F(\varphi,\theta,\hat{x}) = i^{mj}\left|\frac{\cos\varphi+\cos\theta}{2}\right|^r f\left(\frac{\sin\varphi}{\cos\varphi+\cos\theta},\frac{\hat{x}\sin\theta}{\cos\varphi+\cos\theta}\right), \tag{2-5}$$

where j is given by

$$j=\begin{cases} 0 & \text{if } \cos\varphi-\cos\theta>0 \quad\text{and}\quad \frac{\varphi}{2}\in(-\frac{\pi}{2},\frac{\pi}{2}) \pmod{2\pi},\\ 1 & \text{if } \cos\varphi-\cos\theta<0 \quad\text{and}\quad \frac{\varphi}{2}\in(0,\pi) \pmod{2\pi},\\ 2 & \text{if } \cos\varphi-\cos\theta>0 \quad\text{and}\quad \frac{\varphi}{2}\in(-\frac{\pi}{2},\frac{3\pi}{2}) \pmod{2\pi},\\ 3 & \text{if } \cos\varphi-\cos\theta<0 \quad\text{and}\quad \frac{\varphi}{2}\in(\pi,2\pi) \pmod{2\pi},\end{cases}$$

and

$$f(t,x)=\lambda(t,x)^r F\left(\operatorname{sgn} t\,\cos^{-1}\frac{1+q(t,x)}{\lambda(t,x)},\ \cos^{-1}\frac{1-q(t,x)}{\lambda(t,x)},\ \frac{x}{\|x\|}\right) \tag{2-6}$$

where $\lambda(t,x)=(4t^2+(1+q(t,x))^2)^{1/2}$ and $q(t,x)=-t^2+\|x\|^2$.

Define a function $\widetilde{F}\in C^\infty(\mathbb{R}\times\mathbb{R}\times(\mathbb{R}^n\setminus\{0\}))$ by

$$\widetilde{F}(\varphi,\theta,\hat{x}) := F\Big(\varphi,\theta,\frac{x}{\|x\|}\Big),$$

associated to $F\in C^\infty(\mathbb{R}\times\mathbb{R}\times S^{n-1})$, and also the derivative

$$\partial_{\hat{x}_i}F(\varphi,\theta,\hat{x}) = \partial_{x_i}F(\varphi,\theta,x)\big|_{x=\hat{x}}.$$

With this convention, the actions of the Lie algebra are given by the formulas in the following proposition.

Proposition 2. *The Lie algebra action of* $\mathfrak{g}$ *on* $I''_{m,r}$ *is given by*:

(2-7a) $$H_s = s(r\cos\theta\cos\varphi - \cos\theta\sin\varphi\partial_\varphi - \sin\theta\cos\varphi\partial_\theta),$$

(2-7b) $$L_{A,b} = -bx^T(r\sin\varphi\sin\theta + \cos\varphi\sin\theta\partial_\varphi + \sin\varphi\cos\theta\partial_\theta) + (\hat{x}A - \frac{\sin\varphi}{\sin\theta}b)\partial_{\hat{x}}^T,$$

(2-7c) $$\begin{aligned}N_{s,y}^+ = {} & -r(y\hat{x}^T\sin\theta\cos\varphi + s\cos\theta\sin\varphi)\\ & + (y\hat{x}^T\sin\theta\sin\varphi - s(\cos\theta\cos\varphi + 1))\partial_\varphi\\ & - (y\hat{x}^T(\cos\theta\cos\varphi + 1) - s\sin\theta\sin\varphi)\partial_\theta - \frac{\cos\varphi + \cos\theta}{\sin\theta}y\partial_{\hat{x}}^T,\end{aligned}$$

(2-7d) $$\begin{aligned}N_{s,y}^- = {} & -r(y\hat{x}^T\sin\theta\cos\varphi - s\cos\theta\sin\varphi)\\ & + (y\hat{x}^T\sin\theta\sin\varphi + s(\cos\theta\cos\varphi - 1))\partial_\varphi\\ & - (y\hat{x}^T(\cos\theta\cos\varphi - 1) + s\sin\theta\sin\varphi)\partial_\theta - \frac{\cos\varphi - \cos\theta}{\sin\theta}y\partial_{\hat{x}}^T\end{aligned}$$

Proof. See [Hunziker et al. 2012]. □

Now that we have introduced the spaces, mappings, and groups that we will use in the rest of the article, we can start studying the problem that concerns us.

3. Yamabe operator

Recall that the maximal compact group $\widetilde{K}$ is isomorphic to $\mathrm{SO}(2) \times \mathrm{SO}(n+1)$ with $\mathrm{SO}(2)$ embedded in the upper left corner and $\mathrm{SO}(n+1)$ embedded in the lower right corner. It will prove profitable to investigate the action of the Casimir operators associated to these groups.

Proposition 3. *When* $r = (1-n)/2$, *the element*

$$\Omega := \Omega_{\mathrm{SO}(2)} - \Omega_{\mathrm{SO}(n+1)} - r^2$$

of $\mathfrak{g}$ *acts on* $I''_{m,r}$ *as the Yamabe operator* $\widetilde{\Delta}_{S^1\times S^n}$ *on the manifold* $S^1 \times S^n$ *as a subset of* $\mathbb{R}^{2,n+1}$. *In particular, when* $r = (1-n)/2$, $\ker(\Omega - \mu) = \ker(\widetilde{\Delta}_{S^1\times S^n} - \mu)$ *as subsets of* $I''_{m,r}$, *for any constant* $\mu \in \mathbb{R}$.

Proof. The explicit form of the Yamabe operator on $S^1 \times S^n$ embedded in the Minkowski space $\mathbb{R}^{2,n+1}$ is calculated in [Kobayashi and Ørsted 2003a] and is equal to

$$\widetilde{\Delta}_{S^1\times S^n} = \Delta_{S^1} - \Delta_{S^n} - \frac{(n-1)^2}{4}.$$

The Casimir elements in the universal enveloping algebras of $\mathrm{SO}(2)$ and $\mathrm{SO}(n+1)$ act on $I''_{m,r}$ by

$$\Omega_{\mathrm{SO}(2)} = -\tfrac{1}{4}(N_{1,0}^+ + N_{1,0}^-)^2 = -\partial_\varphi^2$$

and

$$\Omega_{\mathrm{SO}(n+1)} = -\tfrac{1}{4}\sum_{i=1}^{n}(N^{+}_{0,e_i} - N^{-}_{0,e_i})^2 - \sum_{1\le i<j\le n}(L_{E_{i,j}-E_{j,i},0})^2$$
$$= -\sum_{i=1}^{n}(\hat{x}_i\partial_\theta + \cot\theta\,\partial_{\hat{x}_i})^2 - \sum_{1\le i<j\le n}(\hat{x}_i\partial_{\hat{x}_j} - \hat{x}_j\partial_{\hat{x}_i})^2 = -\Delta_{S^n}$$

by Equations (2-7) and the recursion formula

$$\Delta_{S^n} = \partial_\theta^2 + (n-1)\cot\theta\,\partial_\theta - \csc^2\theta\,\Delta_{S^{n-1}}.$$

Then, Ω acts on $I''_{m,r}$ by the Yamabe operator $\widetilde{\Delta}_{S^1\times S^n}$. Therefore, the solution space of the equation

$$\widetilde{\Delta}_{S^1\times S^n}F(\varphi,\theta,\hat{x}) = \mu F(\varphi,\theta,\hat{x})$$

is equal to $\ker(\Omega-\mu)$ in $I''_{m,r}$. □

For the rest of the article, unless otherwise specified, let

$$r = \frac{1-n}{2}.$$

We will now determine the maximal subgroup of $\widetilde{G}$ that leaves $\ker(\Omega-\mu)\subset I''_{m,r}$ invariant. To do that, we will use the fact that, on $I''_{m,r}$,

$$\Omega_{\mathrm{SL}(2)} - \Omega_{\mathrm{SO}(n)} - r(r+1) = \sin^2\theta\,(\Omega_{\mathrm{SO}(2)} - \Omega_{\mathrm{SO}(n+1)} - r^2)$$

(see [Hunziker et al. 2012]). From this, it follows that

$$\ker(\Omega_{\mathrm{SL}(2)} - \Omega_{\mathrm{SO}(n)} - r(r+1) - \mu\sin^2\theta) = \ker(\Omega-\mu)$$

when viewed as subspaces of $I''_{m,r}$. Using the canonical isomorphism between $I''_{m,r}$ and $I'_{m,r}$, more specifically (2-6), we obtain that

$$\ker(\Omega-\mu) = \ker\left(\Omega_{\mathrm{SL}(2)} - \Omega_{\mathrm{SO}(n)} - r(r+1) - \frac{4\mu\|x\|^2}{\lambda(t,x)^2}\right) = \ker(\lambda(t,x)^2\square - 4\mu)$$

as subspaces of $I'_{m,r}$, where $\square$ is the wave operator on $\mathbb{R}^{1,n}$. We will use this fact to show the invariance of $\ker(\Omega-\mu)$.

Proposition 4. *If* $\mu\neq 0$, *then the stabilizer of* $\ker(\Omega-\mu)\subset I''_{m,r}$ *in* $\widetilde{G}$ *is the maximal compact subgroup* $\widetilde{K}$.

Proof. Since $\widetilde{G}$ is connected, it suffices to find the maximal subalgebra of $\mathfrak{g}$ that leaves $\ker(\Omega-\mu)\subset I''_{m,r}$ invariant. Since $I'_{m,r}$ and $I''_{m,r}$ are isomorphic, it suffices to determine the maximal subalgebra that leaves $\ker(\lambda(t,x)^2\square - 4\mu)\subset I'_{m,r}$ invariant.

A necessary and sufficient condition for $\ker(\lambda(t,x)^2\Box - 4\mu)$ to be invariant under the action of an element $X \in \mathfrak{g}$ is that

$$[X, \lambda(t,x)^2\Box - 4\mu] = \beta(t,x)(\lambda(t,x)^2\Box - 4\mu) \tag{3-1}$$

for some function $\beta(t,x)$. Straightforward calculations using the formulas in Proposition 1 give the following equations:

$$\begin{aligned}
[H_1, \lambda(t,x)^2\Box - 4\mu] &= 2(1 - q(t,x)^2)\Box, \\
[L_{E_{ij}-E_{ji},0}, \lambda(t,x)^2\Box - 4\mu] &= 0, \\
[L_{0,e_i}, \lambda(t,x)^2\Box - 4\mu] &= -8x_i t\Box, \\
[N^+_{1,0}, \lambda(t,x)^2\Box - 4\mu] &= 4t(q(t,x) - 1)\Box, \\
[N^+_{0,e_i}, \lambda(t,x)^2\Box - 4\mu] &= -4x_i(q(t,x) + 1)\Box, \\
[N^-_{1,0}, \lambda(t,x)^2\Box - 4\mu] &= 4t(q(t,x) - 1)\Box, \\
[N^-_{0,e_i}, \lambda(t,x)^2\Box - 4\mu] &= -4x_i(q(t,x) + 1)\Box.
\end{aligned}$$

From these equations, it can be seen that the condition (3-1) is only satisfied when $\beta(t,x) = 0$. Therefore, the maximal invariance subalgebra of $\ker(\Omega - \mu)$ is spanned by the set $\{L_{E_{ij}-E_{ji}} \mid 1 \le i < j \le n\} \cup \{N^+_{1,0} + N^-_{0,1}\} \cup \{N^+_{0,e_i} - N^-_{0,e_i} \mid 1 \le i \le n\}$ which is isomorphic to the maximal compact subalgebra $\mathfrak{k}$. This yields the desired result. □

It is worth remarking that in the noncompact picture,

$$\ker(\Omega - \mu) = \ker\Big(\Box - \frac{4\mu}{\lambda(t,x)^2}\Big)$$

Therefore, $\ker(\Omega - \mu)$ corresponds to the space of solutions of

$$-u_{tt} + \Delta_n u = \frac{4\mu}{\lambda(t,x)^2} u$$

in $I'_{m,r}$. In particular, $\ker\Omega$ corresponds to the space of solutions of the wave equation in $I'_{m,r}$.

4. $\widetilde{K}$-finite vectors

In this section we will determine the space $\widetilde{K}$-finite vectors of the representation $\ker(\Omega - \mu) \subset I''_{m,r}$. In order to do this, we will first determine the $\widetilde{K}$-finite vectors in $I''_{m,r}$ explicitly by using the following realization of $I''_{m,r}$:

$$I''_{m,r} \cong \{\phi \in C^\infty(S^1 \times S^n) \mid \phi(c \cdot w) = i^{-m}\phi(c) \text{ for every } c \in S^1 \times S^n\},$$

where

$$w = \begin{pmatrix} R_{\pi/2} & \\ & -I_{n+1} \end{pmatrix}.$$

Let the space of $\widetilde{K}$-finite vectors in $C^\infty(S^1 \times S^n)$ be denoted by $C^\infty(S^1 \times S^n)_{\tilde{K}}$. Then it is well known that

$$C^\infty(S^1 \times S^n)_{\tilde{K}} \cong \bigoplus_{\substack{p,k\in\mathbb{Z}\\ k\geq 0}} \mathbb{C}e^{ip\varphi/2} \otimes \mathcal{H}_k(S^n).$$

Where $\mathcal{H}_k(S^n)$ denotes the space of homogeneous harmonic polynomials of degree k on $\mathbb{R}^{n+1}$ restricted to S^n. Then,

$$\mathcal{H}_k(S^n) = \{h \in C^\infty(S^n) \mid \Omega_{\mathrm{SO}(n+1)} h = k(k+n-1)h\}.$$

Since $\phi \in I''_{m,r}$ must satisfy $\phi(c \cdot w) = i^{-m}\phi(c)$, the space of $\widetilde{K}$ finite vectors in $I''_{m,r}$ is given by

$$(I''_{m,r})_{\widetilde{K}} \cong \bigoplus_{\substack{(p,k)\in\mathbb{Z}\times\mathbb{Z}^{\geq 0}\\ p+2k\equiv -m \pmod 4}} \mathbb{C}e^{ip\varphi/2} \otimes \mathcal{H}_k(S^n). \tag{4-1}$$

Proposition 5. *With* $r = (1-n)/2$, *let* $\ker(\Omega-\mu)_{\widetilde{K}}$ *denote the space of* $\widetilde{K}$*-finite vectors in* $\ker(\Omega-\mu) \subset I''_{m,r}$. *Then,*

$$\ker(\Omega-\mu)_{\widetilde{K}} \cong \bigoplus_{\substack{(p,k)\in\mathbb{Z}\times\mathbb{Z}^{\geq 0}\\ p+2k\equiv -m \pmod 4\\ \mu=(p/2)^2-(k-r)^2}} \mathbb{C}e^{ip\frac{\varphi}{2}} \otimes \mathcal{H}_k(S^n).$$

Proof. From the decomposition (4-1) and the well-known fact that $\Omega_{\mathrm{SO}(n)}$ acts on $\mathcal{H}_k(S^n)$ by $k(k-1+n)$, it is easy to see that the operator $\Omega-\mu$ acts on $\mathbb{C}e^{ip\varphi/2} \otimes \mathcal{H}_k(S^n)$ by

$$\left(\frac{p}{2}\right)^2 - k(k-2r) - r^2 = \left(\frac{p}{2}\right)^2 - (k-r)^2.$$

The proposition follows from this and (4-1). □

The following lemma is proved in [Hunziker et al. 2012] and will be used to write a basis for $(I''_{m,r})_{\widetilde{K}}$ explicitly.

Lemma 1. *Let* $\mathrm{SO}(n) \subset \mathrm{SO}(n+1)$ *be the stabilizer of* $(\pm 1, 0, \ldots, 0) \in S^n$. *Then, as an* $\mathrm{SO}(n)$*-module,*

$$\mathcal{H}_k(S^n) \cong \bigoplus_{l=0}^{k} \mathcal{H}_l(S^{n-1}),$$

where the isomorphism is given by

$$(h_0(\hat{x}), \dots, h_k(\hat{x})) \mapsto \sum_{l=0}^{k} \widetilde{C}_{k-l}^{(l-r)}(\cos\theta)\sin^l\theta\, h_l(\hat{x}),$$

where $r = (1-n)/2$ *and* $\widetilde{C}_d^{(\lambda)}$ *is the degree* d *normalized Gegenbauer polynomial with parameter* λ.

Let $\{h_{l,j}\}$ be a basis for the homogeneous harmonic polynomials on $\mathbb{R}^n$ of degree l such that when restricted to S^{n-1} they form an orthonormal basis for $L^2(S^{n-1})$. Then, Lemma 1 implies that the functions of the form

$$F_{p,d,l,j}(\varphi,\theta,\hat{x}) := e^{ip/2\varphi} C_d^{(l-r)}(\cos\theta)\sin^l\theta\, h_{l,j}(\hat{x}) \tag{4-2}$$

with $p, d, l, j \in \mathbb{Z}$, and $d, l, j \geq 0$ such that $p + 2(d+l) \equiv -m \pmod 4$, form a basis of $(I''_{m,r})_{\widetilde{K}}$, where $r = (1-n)/2$. For later use, we note that if we define

$$k := d + l,$$

then the function $C_d^{(l-r)}(\cos\theta)\sin^l\theta\, h_{l,j}(\hat{x}) \in \mathcal{H}_k(S^n)$.

By Proposition 5 we know that $F_{p,d,l,j}$ is an eigenvector for Ω with eigenvalue

$$\mu_{p,d,l} := \left(\frac{p}{2}\right)^2 - \left(\frac{2d+2l-n+1}{2}\right)^2.$$

Definition. We will say that $\mu \in \mathbb{R}$ is an *admissible* eigenvalue of Ω if and only if $\mu = \mu_{p,d,l}$ for some $p, d, l \in \mathbb{Z}$ such that $d \geq 0$, $l \geq 0$, and $p+2(d+l) \equiv -m \pmod 4$. Let S be the set of admissible eigenvalues of Ω.

Then, it is clear that

$$(I''_{m,r})_{\widetilde{K}} \cong \bigoplus_{\mu\in S} \ker(\Omega-\mu)_{\widetilde{K}}.$$

Since $(I''_{m,r})_{\widetilde{K}}$ has the structure of a $(\mathfrak{g}, \widetilde{K})$-module, the direct sum of all eigenspaces of Ω with admissible eigenvalues does too, and as such it can be completed to a representation of $\widetilde{G}$.

One last note to close this section is to contrast the functions $F_{p,l,j}$ constructed in [Hunziker et al. 2012] and the functions $F_{p,l,j,d}$ that we just defined. Even though they are very similar, the main difference is that here we are allowing the parameter $d \geq 0$ to vary, in contrast with $F_{p,l,j}$ where the parameter $d = |p|/2 + r - l$ was fixed for each choice of l. In a sense, the introduction of the parameter d is counting for the fact that different eigenvalues $\mu_{p,d,l} \neq 0$ of the Yamabe operator are being admitted. In [loc. cit.], the only eigenvalue that was considered admissible was $\mu = 0$, in this sense we are generalizing their result.

5. Structure theorems

In this section, we will study the structure of $\oplus_{\mu\in S}\ker(\Omega-\mu)_{\widetilde{K}}$ as a representation of $\mathfrak{g}$. To do this, we will introduce a new basis $\{\kappa, e^+, e^-\}$ for the complexification of the distinguished Lie algebra $\mathfrak{sl}(2,\mathbb{R})$ introduced in Section 2.3. This basis is defined by

$$\kappa := i(E-F), \quad e^+ := \tfrac{1}{2}(H - i(E+F)), \quad e^- := \tfrac{1}{2}(H+i(E+F)). \tag{5-1}$$

Using (2-4) and (2-7) it can be shown that this $\mathfrak{sl}(2,\mathbb{C})$-triple acts on $I''_{m,r}$ by

$$\kappa = -2i\,\partial_\varphi \qquad e^\pm = e^{\pm i\varphi}(r\cos\theta \pm i\cos\theta\,\partial_\varphi - \sin\theta\,\partial_\theta). \tag{5-2}$$

Lemma 2. *The differential operators e^+ and e^- on $I''_{m,r}$ can be written as linear combinations of the form $e^\pm = A^\pm + B^\pm$ such that $A^\pm F_{p,d,l,j}$ and $B^\pm F_{p,d,l,j}$ are all eigenvectors of Ω.*

Proof. Notice that $\Omega_{\mathrm{SO}(2)}$ and $e^\pm$ commute. Since $F_{p,d,l,j}$ is an eigenvector for $\Omega_{\mathrm{SO}(2)}$, so are $e^\pm F_{p,d,l,j}$. Moreover, if $A^\pm$ and $B^\pm$ are a linear combinations of $\{r\cos\theta, \cos\theta\,\partial_\varphi, \sin\theta\,\partial_\theta\}$, then $A^\pm F_{p,d,l,j}$ and $B^\pm F_{p,d,l,j}$ are all eigenvectors of $\Omega_{\mathrm{SO}(2)}$. So, by the definition of Ω, it suffices to determine $A^\pm$ and $B^\pm$ so that $A^\pm F_{p,d,l,j}$ and $B^\pm F_{p,d,l,j}$ are eigenvectors of $\Omega_{\mathrm{SO}(n+1)}$.

Now, $A^\pm F_{p,d,l,j}$ and $B^\pm F_{p,d,l,j}$ are eigenvectors of $\Omega_{\mathrm{SO}(n+1)}$ if and only if $F_{p,d,l,j}$ is an eigenvector of $[\Omega_{\mathrm{SO}(n+1)}, A^\pm]$ and of $[\Omega_{\mathrm{SO}(n+1)}, B^\pm]$. Writing these conditions out explicitly gives the following form for the operators $A^\pm$ and $B^\pm$:

$$A^\pm = e^{\pm i\varphi}\Big(\frac{r}{2}\cos\theta - \frac{4r^2+2(d+l)(2(d+l)\mp p-4r)}{4p(d+l-r)} i\cos\theta\,\partial_\varphi \\ - \frac{2(d+l)\mp p-2r}{4((d+l)-r)}\sin\theta\,\partial_\theta\Big) \tag{5-3}$$

and

$$B^\pm = e^{\pm i\varphi}\Big(\frac{r}{2}\cos\theta + \frac{4r(r\mp p)+2(d+l)(2(d+l)\pm p-4r)}{4p(d+l-r)} i\cos\theta\,\partial_\varphi \\ - \frac{2(d+l)\pm p-2r}{4(d+l-r)}\sin\theta\,\partial_\theta\Big). \tag{5-4}$$

These operators satisfy the required conditions. □

By making the change of variables $s=\cos\theta$ and considering the fact that $\partial_\varphi F_{p,d,l,j} = ip/2$, when restricted to the basis elements

$$F_{p,d,l,j}(\varphi, s, \hat{x}) = e^{ip\varphi/2}(1-s^2)^{l/2} C_d^{(l-r)}(s)\, h_{l,j}(\hat{x}),$$

the operators in (5-3) and (5-4) act by

$$A^{\pm} = \frac{\mp p+2(d+l-r)}{4(d+l-r)} e^{\pm i\varphi}((d+l)s + (1-s^2)\partial_s) \tag{5-5}$$

and

$$B^{\pm} = \frac{\pm p-2(d+l-r)}{4(d+l-r)} e^{\pm i\varphi}((2r-d-l)s + (1-s^2)\partial_s), \tag{5-6}$$

respectively. We are now in a position to calculate the actions of $e^{\pm}$ on the functions $F_{p,d,l,j}(\varphi, s, \hat{x})$. We start with the following proposition.

Proposition 6. *Let $F_{p,d,l,j}$ be defined as in (4-2) and let $A^{\pm}$ and $B^{\pm}$ be defined by (5-3) and (5-4) respectively. Then,*

$$A^{\pm} F_{p,d,l,j} = \frac{\mp p+2(d+l-r)}{4(d+l-r)} (d+2(l-r)-1) F_{p\pm 2,d-1,l,j} \tag{5-7}$$

and

$$B^{\pm} F_{p,d,l,j} = \frac{\mp p-2(d+l-r)}{4(d+l-r)} (d+1) F_{p\pm 2,d+1,l,j}. \tag{5-8}$$

Proof. Firstly, we will need to state two well-known identities for the Gegenbauer polynomials:

$$(1-s^2)\frac{d}{ds}C_d^{(\lambda)}(s) = -dsC_d^{(\lambda)}(s) + (d+2\lambda-1)C_{d-1}^{(\lambda)}(s) \tag{5-9}$$

and

$$(1-s^2)\frac{d}{ds}C_d^{(\lambda)}(s) = (d+2\lambda)sC_d^{(\lambda)}(s) - (d+1)C_{d+1}^{(\lambda)}(s) \tag{5-10}$$

(see [Abramowitz and Stegun 1964, Formulas 22.8.2 and 22.7.3]). Now, (5-7) follows from (5-5) and (5-9), and analogously (5-8) is obtained by combining (5-6) and (5-10). □

Corollary 2. *Let $F_{p,d,l,j}$ be defined as in (4-2). Then,*

$$e^{+} F_{p,d,l,j} = \frac{-p+2(d+l-r)}{4(d+l-r)} (d+2(l-r)-1) F_{p+2,d-1,l,j} + \frac{-p-2(d+l-r)}{4(d+l-r)} (d+1) F_{p+2,d+1,l,j} \tag{5-11}$$

and

$$e^{-} F_{p,d,l,j} = \frac{p+2(d+l-r)}{4(d+l-r)} (d+2(l-r)-1) F_{p-2,d-1,l,j} + \frac{p-2(d+l-r)}{4(d+l-r)} (d+1) F_{p-2,d+1,l,j}, \tag{5-12}$$

with the convention that $F_{p,-1,l,j} \equiv 0$.

Proof. The corollary follows at once from Proposition 6. □

Now we can read much information from the coefficients in (5-11) and (5-12). Firstly, by definition $d, l \geq 0$ and $r < 0$, so $d + l - r > 0$. Therefore, these equations are well defined for every $F_{p,d,l,j}$ as in (4-2). Secondly, we can look for highest/lowest weight vectors for the action of $\mathfrak{sl}(2, \mathbb{C})$. In order for $F_{p,d,l,j}$ to be a highest or lowest weight vector, we would need either e^+ or e^- to annihilate $F_{p,d,l,j}$ respectively. By inspecting the coefficients we conclude that this can only occur whenever $p = 2(d+l-r)$ or $p = -2(d+l-r)$. However, this can occur only if $\mu = 0$. In this case, the actions of $e^\pm$ would have only one term depending on the sign of p. This corresponds with the action calculated in [Hunziker et al. 2012]. Moreover, the highest and lowest weight vectors are precisely the ones calculated therein.

In more detail, if $p = -2(d+l-r)$, then

$$e^+ F_{-2(d+l-r),d,l,j} = \frac{-p+2(d+l-r)}{4(d+l-r)}(d+2(l-r)-1)F_{-2(d+l-r-1),d-1,l,j}.$$

In particular,

$$e^+ F_{-2(l-r),0,l,j} = 0.$$

Similarly,

$$e^- F_{2(l-r),0,l,j} = 0$$

for $l, j \in \mathbb{Z}^{\geq 0}$. So, for each combination of parameters $l, j \in \mathbb{Z}^{\geq 0}$ there exists a highest weight $\mathfrak{sl}(2, \mathbb{C})$-module in $(I''_{m,r})_{\widetilde{K}}$ with highest weight vector $F_{-2(l-r),0,l,j}$. There also exists a lowest weight module with lowest weight vector $F_{2(l-r),0,l,j}$.

In Figure 1 we show the projection to the plane $l = 0$ of the representation $(I''_{m,r})_{\widetilde{K}}$ pictorially. However, the pictorial representation of all of $(I''_{m,r})_{\widetilde{K}}$ would require a third axis for the l parameter. In this picture, the span of highest weight vector $F_{2r,0,0,0}$ of $(\mathscr{H}^-)_{\widetilde{K}}$ corresponds to the point $(0, 2r)$. For a fixed $l \in \mathbb{Z}^{\geq 0}$, the span of the highest weight vectors $F_{-2(l-r),0,l,j}$ corresponds to the dot in the position $(0, -2(l-r))$ in that particular plane. To illustrate, Figure 2 shows the projection onto the plane $l = 1$. The highest weight vectors there are $F_{2r-2,0,1,j}$ and represented by the point $(0, 2r-2)$.

In Figure 1 we also describe the actions of $\mathfrak{sl}(2, \mathbb{C})$. In this figure, a dot at the (d, p) coordinate represents the span of $\{F_{p,d,l,j} \mid l, j \in \mathbb{Z}^{\geq 0}\}$. As shown in Corollary 2, the action of e^+ sends a multiple of $F_{p,d,l,j}$ into the span of $F_{p+2,d\pm1,l,j}$, thus the northeast and northwest arrows. Similarly, the action of e^- sends a multiple of $F_{p,d,l,j}$ into the span of $F_{p-2,d\pm1,l,j}$, hence the arrows pointing in the southeast and southwest directions. Lastly, the semisimple element κ acts by the scalar p on $F_{p,d,l,j}$, thus leaving each point fixed.

To finish this analysis we introduce two spaces that correspond to the positive and negative energy representations for the zero eigenspace. These representations

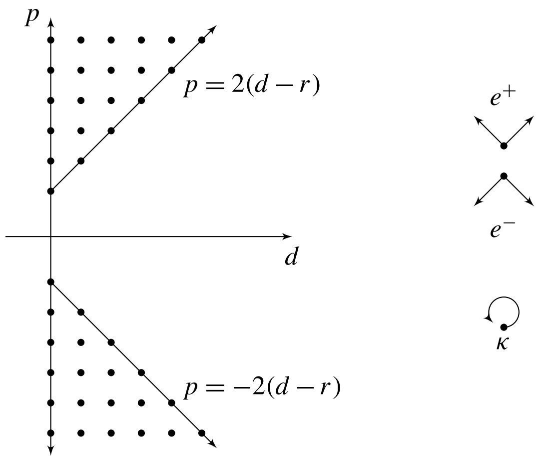


Figure 1. Summary of the action of $\mathfrak{sl}(2, \mathbb{C})$ on $(I''_{m,r})_{\widetilde{K}}$ and $l = 0$.

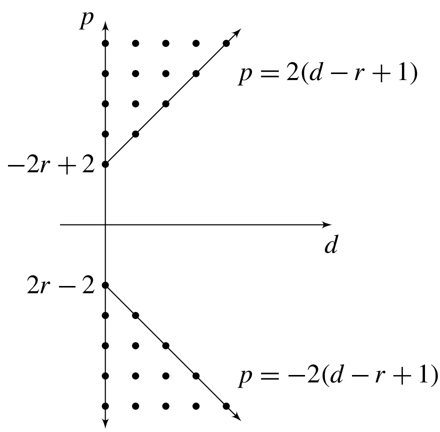


Figure 2. Pictorial representation of the plane $l = 1$ of $(I''_{m,r})_{\widetilde{K}}$.

are related to positive and negative energy solutions of the wave equation. For more information in this direction, see [Hunziker et al. 2012]. Let

$$(\mathcal{H}^{-})_{\widetilde{K}} := \bigoplus_{\substack{p \in \mathbb{Z}, k \in \mathbb{Z}^{\geq 0} \\ p = -2(k - r) \\ p + 2k \equiv -m \pmod 4}} \mathbb{C}e^{ip\varphi/2} \otimes \mathcal{H}_k(S^n)$$

and

$$(\mathcal{H}^{+})_{\widetilde{K}} := \bigoplus_{\substack{p \in \mathbb{Z}, k \in \mathbb{Z}^{\geq 0} \\ p = 2(k - r) \\ p + 2k \equiv -m \pmod 4}} \mathbb{C}e^{ip\varphi/2} \otimes \mathcal{H}_k(S^n).$$

In the pictorial representation, $(\mathscr{H}^-)_{\widetilde{K}}$ would live in the $p=-2(d+l-r)$ plane and $(\mathscr{H}^+)_{\widetilde{K}}$ would live in the plane $p=2(d+l-r)$. Each projection onto a fixed l would look essentially the same as Figure 1, with the intercepts at $\pm 2(l-r)=\pm(2l-n+1)$.

So far we have studied the structure of $(I''_{m,r})_{\widetilde{K}}$ as an $\mathfrak{sl}(2,\mathbb{C})$-module. To analyze the structure of $(I''_{m,r})_{\widetilde{K}}$ as a $(\mathfrak{g}_{\mathbb{C}},\widetilde{K})$-module, we start by fixing a Cartan subalgebra $\mathfrak{h}_{\mathbb{C}}$ of $\mathfrak{k}_{\mathbb{C}}=\mathfrak{so}(2)\times\mathfrak{so}(n+1)$ given by

$$\mathfrak{h}_{\mathbb{C}} := \{H_{h_1,\dots,h_\ell} \mid h_1,\dots,h_\ell \in \mathbb{C}\},$$

where

$$H_{h_1,\dots,h_\ell} = \left(\begin{array}{cc|cccccc} 0 & ih_0 & & & & & & \\ -ih_0 & 0 & & & & & & \\ \hline & & \ddots & & & & & \\ & & & 0 & ih_2 & & & \\ & & & -ih_2 & 0 & & & \\ & & & & & & 0 & ih_1 \\ & & & & & & -ih_1 & 0 \end{array}\right)$$

where $\ell=[(n+1)/2]$. Let $\epsilon_j:\mathfrak{h}_{\mathbb{C}}\to\mathbb{C}$ be the functional such that $\epsilon_j(H)=h_j$. As it turns out, the roots $\epsilon_0+\epsilon_1$ and $\epsilon_0-\epsilon_1$ are the noncompact simple root and the highest root respectively. The corresponding root vectors $X_{\epsilon_0\pm\epsilon_1}$ are given by

$$X_{\epsilon_0\pm\epsilon_1} := L_{0,e_{n-1}\mp ie_n} + \tfrac{1}{2}(N^+_{0,-ie_{n-1}\mp e_n} + N_{0,-ie_{n-1}\mp e_n}) \in \mathfrak{p}^+$$

and the complex conjugates $\overline{X}_{\epsilon_0\pm\epsilon_1}$ are the root vectors for the respective negative roots. The proof of [Hunziker et al. 2012, Proposition 9.6] implies that for $k\geq 0$

$$(X_{\epsilon_0\pm\epsilon_1})^k \cdot F_{-2r,0,0,0} \in \operatorname{span}\{F_{-2r+2k,0,k,j} \mid j\in\mathbb{Z}^{\geq 0}\}$$

and

$$(\overline{X}_{\epsilon_0\pm\epsilon_1})^k \cdot F_{2r,0,0,0} \in \operatorname{span}\{F_{2r-2k,0,k,j} \mid j\in\mathbb{Z}^{\geq 0}\}.$$

Moreover, they show that the functions $(X_{\epsilon_0+\epsilon_1})^k F_{-2r,0,0,0}$ are $\mathfrak{k}_{\mathbb{C}}$-lowest weight vectors and $(\overline{X}_{\epsilon_0-\epsilon_1})^k F_{2r,0,0,0}$ are $\mathfrak{k}_{\mathbb{C}}$-lowest weight vectors. Since these vectors are also annihilated by e^+ and e^- respectively, the vectors $F_{-2r,0,0,0}$ and $F_{2r,0,0,0}$ are $\mathfrak{g}_{\mathbb{C}}$-lowest and highest weight vectors respectively. Intuitively, in the pictorial representation of $(I''_{m,r})_{\widetilde{K}}$ these root vectors $X_{\epsilon_0\pm\epsilon_1}$ and $\overline{X}_{\epsilon_0\pm\epsilon_1}$ allow us to move from one l plane to the preceding $l-1$ and superseding $l+1$ planes.

Putting all this information together, we obtain our main result.

Theorem 1. *Let $r=(1-n)/2$. Let $(I''_{m,r})_{\widetilde{K}}$ denote the space of $\widetilde{K}$-finite vectors of $I''_{m,r}$ and let*

$$(I''_{m,r})^{\pm}_{\widetilde{K}} := \bigoplus_{\substack{\pm p\in\mathbb{Z}^{>0}\\ k\in\mathbb{Z}^{\geq 0}\\ p+2k\equiv -m \pmod 4}} \mathbb{C}e^{ip\varphi/2}\otimes\mathscr{H}_k(S^n).$$

Then, as $(\mathfrak{g}_{\mathbb{C}}, \widetilde{K})$-modules:

(1) *The submodules $(\mathscr{H}^{\pm})_{\widetilde{K}}$ are irreducible lowest/highest weight modules with weight vectors $F_{\pm 2r,0,0,0}$ respectively.*

(2) *The quotient modules $(I''_{m,r})^{+}_{\widetilde{K}}/(\mathscr{H}^{+})_{\widetilde{K}}$ and $(I''_{m,r})^{-}_{\widetilde{K}}/(\mathscr{H}^{-})_{\widetilde{K}}$ are irreducible lowest/highest weight modules with the weight vectors being the cosets corresponding to $F_{-2r+2,0,0,0}$ and $F_{2r-2,0,0,0}$.*

(3) *The following is a composition series of $(I''_{m,r})_{\widetilde{K}}$:*

$$\{0\} \subset (\mathscr{H}^{\pm})_{\widetilde{K}} \subset (\mathscr{H}^{-})_{\widetilde{K}} \oplus (\mathscr{H}^{+})_{\widetilde{K}} \subset (I''_{m,r})_{\widetilde{K}}.$$

Proof. The only statement left to be shown is (2), as this implies the composition series in (3). This statement follows from the fact that for $k > 0$ the nonzero vectors

$$(X_{\epsilon_0 \pm \epsilon_1})^k \cdot F_{-2r+2,0,0,0} \in \operatorname{span}\{F_{-2(r-k-1),0,k,j} \mid j \in \mathbb{Z}^{\geq 0}\}$$

and

$$(\overline{X}_{\epsilon_0 \pm \epsilon_1})^k \cdot F_{2r-2,0,0,0} \in \operatorname{span}\{F_{2(r-k-1),0,k,j} \mid j \in \mathbb{Z}^{\geq 0}\},$$

which is proved using the explicit actions of $X_{\epsilon_0 \pm \epsilon_1}$ and $\overline{X}_{\epsilon_0 \pm \epsilon_1}$ and induction. □

References

[Abramowitz and Stegun 1964] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, National Bureau of Standards Applied Mathematics Series **55**, U.S. Government Printing Office, Washington, DC, 1964. Reprinted by Dover, New York, 1974. MR 29 #4914 Zbl 0171.38503

[Binegar and Zierau 1991] B. Binegar and R. Zierau, "Unitarization of a singular representation of $\mathrm{SO}(p,q)$", *Comm. Math. Phys.* **138**:2 (1991), 245–258. MR 92h:22027 Zbl 0748.22009

[Franco and Sepanski 2013] J. A. Franco and M. R. Sepanski, "Global representations of the heat and Schrödinger equation with singular potential", *Electron. J. Differential Equations* **2013**:154 (2013), 1–16. MR 3084634 Zbl 1286.22016 arXiv 1110.1030

[Hunziker et al. 2012] M. Hunziker, M. R. Sepanski, and R. J. Stanke, "The minimal representation of the conformal group and classical solutions to the wave equation", *J. Lie Theory* **22**:2 (2012), 301–360. MR 2976923 Zbl 1250.22015

[Kobayashi and Ørsted 2003a] T. Kobayashi and B. Ørsted, "Analysis on the minimal representation of $\mathrm{O}(p,q)$, I: Realization via conformal geometry", *Adv. Math.* **180**:2 (2003), 486–512. MR 2004k:22018a Zbl 1046.22004

[Kobayashi and Ørsted 2003b] T. Kobayashi and B. Ørsted, "Analysis on the minimal representation of $\mathrm{O}(p,q)$, II: Branching laws", *Adv. Math.* **180**:2 (2003), 513–550. MR 2004k:22018b Zbl 1049.22006

[Kobayashi and Ørsted 2003c] T. Kobayashi and B. Ørsted, "Analysis on the minimal representation of $\mathrm{O}(p,q)$, III: Ultrahyperbolic equations on $\mathbb{R}^{p-1,q-1}$", *Adv. Math.* **180**:2 (2003), 551–595. MR 2004k:22018c Zbl 1039.22005

[Segal et al. 1981] I. E. Segal, H. P. Jakobsen, B. Ørsted, S. M. Paneitz, and B. Speh, "Covariant chronogeometry and extreme distances: elementary particles", *Proc. Nat. Acad. Sci. USA* **78**:9 (1981), 5261–5265. MR 82i:81150

[Sepanski and Stanke 2010] M. R. Sepanski and R. J. Stanke, "Global Lie symmetries of the heat and Schrödinger equation", *J. Lie Theory* **20**:3 (2010), 543–580. MR 2012a:22013 Zbl 05792649

Received July 8, 2014. Revised November 10, 2014.

MARK R. SEPANSKI
DEPARTMENT OF MATHEMATICS
BAYLOR UNIVERSITY
ONE BEAR PLACE #97328
WACO, TX 76798-7328
UNITED STATES
mark_sepanski@baylor.edu

JOSE A. FRANCO
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NORTH FLORIDA
1 UNF DRIVE
JACKSONVILLE, FL 32224
UNITED STATES
jose.franco@unf.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 275, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.275.481

ROTA–BAXTER OPERATORS ON THE POLYNOMIAL ALGEBRA, INTEGRATION, AND AVERAGING OPERATORS

SHANGHUA ZHENG, LI GUO AND MARKUS ROSENKRANZ

The concept of a Rota–Baxter operator is an algebraic abstraction of integration. Following this classical connection, we study the relationship between Rota–Baxter operators and integrals in the case of the polynomial algebra $\boldsymbol{k}[x]$. We consider two classes of Rota–Baxter operators, monomial ones and injective ones. For the first class, we apply averaging operators to determine monomial Rota–Baxter operators. For the second class, we make use of the double product on Rota–Baxter algebras.

1. Introduction 481
2. General concepts and properties 483
3. Monomial Rota–Baxter operators on $\boldsymbol{k}[x]$ 485
4. Injective Rota–Baxter operators on $\boldsymbol{k}[x]$ 497
Acknowledgements 505
References 505

1. Introduction

Rota–Baxter operators are deeply rooted in analysis. Their study originated from the work of G. Baxter [1960] on Spitzer's identity [1956] in fluctuation theory. More fundamentally, the notion of Rota–Baxter operator is an algebraic abstraction of the *integration by parts formula* of calculus. Throughout the 1960s, Rota–Baxter operators were studied by well-known analysts such as Atkinson [1963]. In the 1960s and 1970s, the works of Rota [1969a; 1969b] and Cartier [1972] led the study of Rota–Baxter operators into algebra and combinatorics. In the 1980s, the Rota–Baxter operator for Lie algebras was independently discovered by mathematical physicists as the operator form of the classical Yang–Baxter equation [Semenov-Tian-Shansky 1983]. In the late 1990s, the operator appeared again as a fundamental algebraic structure in the work of Connes and Kreimer [2000] on renormalization of quantum field theory. The present century has witnessed a remarkable renaissance of

MSC2010: primary 16W99; secondary 45N05, 47G10, 12H20.
Keywords: Rota–Baxter operator, averaging operator, integration, monomial linear operator.

Rota–Baxter operators through systematic algebraic studies with wide applications to combinatorics, number theory, operads and mathematical physics [Connes and Kreimer 2000; Guo and Keigher 2000; Aguiar 2001; Ebrahimi-Fard et al. 2004; 2006; Bai 2007; Guo and Zhang 2008; Bai et al. 2010; 2013]. See [Guo 2009] for a brief introduction and [Guo 2012] for a more detailed treatment.

Recently, structures related to Rota–Baxter operators, including differential Rota–Baxter algebras [Guo and Keigher 2008] and integro-differential algebras [Rosenkranz and Regensburger 2008], were introduced in the algebraic study of calculus, especially in boundary problems for linear differential equations [Gao et al. 2014; Guo et al. 2014]. The upshot is that the Green's operator of such a boundary problem can be represented by suitable operator rings based on an integro-differential algebra.

In this paper, we revisit the analysis origin of Rota–Baxter operators to study how their algebraic properties are linked with their analytic appearance. We focus on the polynomial algebra $\mathbb{R}[x]$, which plays a central role both in analysis, where it is taken as approximation of analytic functions, and in algebra, where it is the free object in the category of commutative algebras. This algebra, together with the standard integral operator, is also the free commutative Rota–Baxter algebra on the empty set or, in other words, the initial object in the category of commutative Rota–Baxter algebras. Thus it provides an ideal testing ground for the interaction between analytically defined and algebraically defined Rota–Baxter operators.

One natural question in this regard is when an algebraically defined Rota–Baxter operator on $\mathbb{R}[x]$ can be realized in analysis. It is a classical fact that the Riemann integral with variable upper limit is a Rota–Baxter operator of weight zero on $\mathbb{R}[x]$. This remains true when any polynomial is multiplied before the integral operator is applied. We might call these Rota–Baxter operators on $\mathbb{R}[x]$ *analytically modeled*. It is easy to see that such operators are injective. We conjecture that all injective Rota–Baxter operators on $\mathbb{R}[x]$ are indeed analytically modeled. We provide evidence for this conjecture by exploring two classes of such operators, one class being special but interesting on its own right and the other one being speculatively the general case.

The first comprises what we call *monomial Rota–Baxter operators* over an arbitrary integral domain $\boldsymbol{k}$ of characteristic zero, meaning Rota–Baxter operators P with $P(x^n) = ax^k$, where both $a \in \boldsymbol{k}$ and $k \in \mathbb{N}$ may depend on n. We classify monomial Rota–Baxter operators on $\boldsymbol{k}[x]$ and show that all injective monomial Rota–Baxter operators are analytically modeled. The second class is restricted to $\boldsymbol{k} = \mathbb{R}$ and contains those operators that satisfy a *differential law* $\partial \circ P = r$, where the right-hand side denotes the multiplication operator induced by an arbitrary monomial $r \in \mathbb{R}[x]$. We show that any injective Rota–Baxter operator is of this form and, provided r is monomial, analytically modeled.

In Section 2, we discuss general algebraic properties of Rota–Baxter operators that will be used in subsequent sections. In Section 3, we focus on monomial Rota–Baxter operators. While determining these operators, we prove that all injective monomial Rota–Baxter operators are analytically modeled. In Section 4, we study injective Rota–Baxter operators in general (on the real polynomial ring). We first show that injective Rota–Baxter operators are precisely those that satisfy a differential law. Then we prove that, in the monomial case, they are analytically modeled.

2. General concepts and properties

Notation. Let M be a monoid with zero element 0_M. Set $M^\times = \{x \in M \mid x \neq 0_M\}$. If $M^\times$ is closed under the multiplication of M, then it is a subsemigroup of M. In particular, the monoid of natural numbers (nonnegative integers) is denoted by $\mathbb{N}$, so $\mathbb{N}^\times$ is the semigroup of positive integers. The notation $l \mid k$ signifies that l is a divisor of k.

We use $\boldsymbol{k}$ to denote a commutative ring with identity 1 unless otherwise specified. All $\boldsymbol{k}$-algebras in this paper are assumed to be commutative and with a unit 1_A that will be identified with $1_{\boldsymbol{k}}$ through the structure map $\boldsymbol{k} \to A$.

We start by collecting some general properties of Rota–Baxter operators for later use. First we give the definition of a Rota–Baxter $\boldsymbol{k}$-algebra of arbitrary weight:

Definition 2.1 [Baxter 1960; Rota 1995; Guo and Keigher 2000]. Let λ be a given element of $\boldsymbol{k}$. A *Rota–Baxter* $\boldsymbol{k}$*-algebra of weight* λ, or simply an *RBA of weight* λ, is a pair (R, P) consisting of a $\boldsymbol{k}$-algebra R and a linear operator $P : R \to R$ that satisfies the *Rota–Baxter equation*

$$P(u)P(v) = P(uP(v)) + P(P(u)v) + \lambda P(uv) \quad \text{for all } u, v \in R. \tag{1}$$

In this case P is called a *Rota–Baxter operator of weight* λ. If R is only assumed to be a nonunitary $\boldsymbol{k}$-algebra, we call R a nonunitary Rota–Baxter $\boldsymbol{k}$-algebra of weight λ.

Observe first that the standard integration operator $J_0 : \boldsymbol{k}[x] \to \boldsymbol{k}[x]$, given by $x^n \mapsto x^{n+1}/(n+1)$, is a (prototypical) Rota–Baxter operator of weight 0. Of course the choice of initialization point is irrelevant, so for any $a \in \boldsymbol{k}$ there is another weight-0 Rota–Baxter operator $J_a : \boldsymbol{k}[x] \to \boldsymbol{k}[x]$, given by $x^n \mapsto (x^{n+1} - a^{n+1})/(n+1)$. In this paper we shall only be concerned with the weight-0 case, so from now on the term "Rota–Baxter operator" is to be understood as "Rota–Baxter operator of weight 0".

Recall that from a derivation δ on a commutative $\boldsymbol{k}$-algebra R one can produce a new derivation $r\delta$ by post-multiplying with any $r \in R$. Analogously, from a Rota–Baxter operator P on R one obtains a new Rota–Baxter operator Pr by *pre-multiplying* with any $r \in R$. Indeed,

$$\begin{aligned}(Pr)(u)(Pr)(v) = P(ru)P(rv) &= P(ruP(rv)) + P(P(ru)rv) \\ &= (Pr)(u(Pr)(v)) + (Pr)((Pr)(u)v)\end{aligned}$$

for any $u, v \in R$. Applying this to $R = \boldsymbol{k}[x]$, we obtain the family $J_a r$ of *analytically modeled* Rota–Baxter operators on $\boldsymbol{k}[x]$, where $a \in \boldsymbol{k}$ and $r \in \boldsymbol{k}[x]$ are arbitrary. As we will show in Theorem 4.9, in the case of monomials r, this family exhausts the injective monomial Rota–Baxter operators.

Let $\mathrm{End}(R) := \mathrm{End}_{\boldsymbol{k}}(R)$ denote the $\boldsymbol{k}$-module of linear operators on R. Then the subset $\mathrm{RBO}(R)$ of $\mathrm{End}(R)$ consisting of Rota–Baxter operators $P : R \to R$ is closed under *multiplications by scalars* $c \in \boldsymbol{k}$, since in that case $Pc = cP$. In the case of derivations on R more is true, since they form a $\boldsymbol{k}$-module (in fact a Lie algebra) while in general the sum of two Rota–Baxter operators is not a Rota–Baxter operator. This motivates the following terminology:

Definition 2.2. (a) We call two Rota–Baxter operators $P_1, P_2 \in \mathrm{RBO}(R)$ *compatible* if $c_1 P_1 + c_2 P_2$ are in $\mathrm{RBO}(R)$ for all $c_1, c_2 \in \boldsymbol{k}$.

(b) Let $P \in \mathrm{RBO}(R)$. Then $Q \in \mathrm{End}(R)$ is called *consistent* with P if $P - Q$ is in $\mathrm{RBO}(R)$.

(c) For $P, Q \in \mathrm{End}(R)$, we define the bilinear form $\mathrm{RB}(P, Q) : R \otimes R \to R$ by

$$\mathrm{RB}(P, Q)(u, v) := P(u)Q(v) - P(uQ(v)) - Q(P(u)v), \quad u, v \in R.$$

Thus $P \in \mathrm{RBO}(R)$ means that $\mathrm{RB}(P, P) = 0$ on $R \otimes R$.

Recall that for a Rota–Baxter algebra (R, P) the multiplication

$$\begin{aligned}&\star_P : R \otimes R \to R,\\ &\quad u \star_P v := P(u)v + uP(v) \quad \text{for all } u, v \in R,\end{aligned}$$

is an associative product on R, called the *double multiplication* [Guo 2012, Theorem 1.1.17]. Moreover, $P : (R, \star_P) \to R$ is then a homomorphism of nonunitary Rota–Baxter algebras.

If A is a $\boldsymbol{k}$-module, its (linear) *dual* is denoted by A^*. If A is moreover a $\boldsymbol{k}$-algebra, we use the notation

$$A^\bullet := \{\phi \in A^* \mid \phi(uv) = \phi(u)\phi(v)\}$$

for the set of *multiplicative functionals*. Through the structure map $\boldsymbol{k} \to A$ we may also view the elements of A^* as $\boldsymbol{k}$-linear operators from A to A, and those of $A^\bullet$ as $\boldsymbol{k}$-algebra homomorphisms from A to $\boldsymbol{k}$.

Proposition 2.3. (a) *Two Rota–Baxter operators* $P_1, P_2 \in \mathrm{RBO}(R)$ *are compatible if and only if* $\mathrm{RB}(P_1, P_2) + \mathrm{RB}(P_2, P_1) = 0$. *This will be the case in particular when*

$$P_1(u)P_2(v) = P_1(uP_2(v)) + P_2(P_1(u)v),$$
$$P_2(u)P_1(v) = P_2(uP_1(v)) + P_1(P_2(u)v),$$

hold for all $u, v \in R$.

(b) *Let* $P \in \mathrm{RBO}(R)$ *and* $Q \in \mathrm{End}(R)$ *be given. Then* Q *is consistent with* P *if and only if*

$$\mathrm{RB}(Q, Q) = \mathrm{RB}(P, Q) + \mathrm{RB}(Q, P).$$

(c) *Let* P *be in* $\mathrm{RBO}(R)$. *The set of* $f \in R^*$ *that are consistent with* P *equals* $(R, \star_P)^\bullet$.

Proof. (a) For arbitrary $c_1, c_2 \in \boldsymbol{k}$, the bilinear form $\mathrm{RB}(c_1P_1 + c_2P_2, c_1P_1 + c_2P_2)$ is given by

$$c_1^2\,\mathrm{RB}(P_1, P_1) + c_1c_2(\mathrm{RB}(P_1, P_2) + \mathrm{RB}(P_2, P_1)) + c_2^2\,\mathrm{RB}(P_2, P_2),$$

which simplifies to $c_1c_2(\mathrm{RB}(P_1, P_2) + \mathrm{RB}(P_2, P_1))$ since $P_1, P_2 \in \mathrm{RBO}(R)$.

(b) Since $P \in \mathrm{RBO}(R)$, we obtain

$$\mathrm{RB}(P - Q, P - Q) = -\mathrm{RB}(P, Q) - \mathrm{RB}(Q, P) + \mathrm{RB}(Q, Q),$$

and hence the conclusion.

(c) Using that P is a linear operator and f a linear functional, we have

$$\mathrm{RB}(f, f) = -f(u)f(v),$$
$$\mathrm{RB}(f, P)(u, v) = -f(uP(v)), \quad \mathrm{RB}(P, f)(u, v) = -f(P(u)v).$$

Thus by (b) we conclude that f is consistent with P if and only if

$$f(u)f(v) = f(P(u)v + uP(v)) = f(u \star_P v). \qquad \square$$

3. Monomial Rota–Baxter operators on $\boldsymbol{k}[x]$

In this section, we determine the Rota–Baxter operators on $\boldsymbol{k}[x]$ that send monomials to monomials, called monomial Rota–Baxter operators. We show that the nondegenerate monomial Rota–Baxter operators are analytically modeled. For their independent interest in studying Rota–Baxter operators on $\boldsymbol{k}[x]$, we also consider the degenerate ones. Throughout this section, we assume that $\boldsymbol{k}$ is an integral domain containing $\mathbb{Q}$.

3A. ***General properties.*** We first give general criteria for a monomial linear operator to be a Rota–Baxter operator before specializing in the following sections to the two cases of nondegenerate and degenerate operators.

Definition 3.1. (a) A linear operator P on $\boldsymbol{k}[x]$ is called *monomial* if for each $n \in \mathbb{N}$

$$P(x^n) = \beta(n)x^{\theta(n)} \quad \text{with } \beta : \mathbb{N} \to \boldsymbol{k} \text{ and } \theta : \mathbb{N} \to \mathbb{N}. \tag{2}$$

If $\beta(n) = 0$, the value of $\theta(n)$ does not matter; by convention we set $\theta(n) = 0$ in this case.

(b) A monomial operator is called *degenerate* if $\beta(n) = 0$ for some $n \in \mathbb{N}$.

Let A be a nonempty set and let B be a set containing a distinguished element 0. For a map $\phi : A \to B$ we define its *zero set* as $\mathcal{Z}_\phi := \{a \in A \mid \phi(a) = 0\}$ and its *support* as $\mathcal{S}_\phi := A \setminus \mathcal{Z}_\phi$. Thus a monomial linear operator P on $\boldsymbol{k}[x]$ is nondegenerate if and only if $\mathcal{Z}_\beta = \varnothing$. As the following lemma shows, for a Rota–Baxter operator P, degeneracy at $n \in \mathbb{N}$ occurs whenever P is constant on the corresponding monomial:

Lemma 3.2. *Let P be a monomial Rota–Baxter operator on $\boldsymbol{k}[x]$ and let $n \in \mathbb{N}$. If $P(x^n)$ is in $\boldsymbol{k}$, then $P(x^n) = 0$. In other words, $\mathcal{S}_\beta = \mathcal{S}_\theta$, and hence $\mathcal{Z}_\beta = \mathcal{Z}_\theta$.*

Proof. If $P(x^n) = c$ is a nonzero constant, then

$$P(x^n)P(x^n) = c^2 \neq 2c^2 = 2P(x^n P(x^n)).$$

Hence P is not a Rota–Baxter operator, and we must have $c = 0$. □

Theorem 3.3. *Let P be a monomial linear operator on $\boldsymbol{k}[x]$ defined by $P(x^n) = \beta(n)x^{\theta(n)}$, $n \in \mathbb{N}$. Then P is a Rota–Baxter operator if θ and β satisfy the following conditions*:

(a) $\mathcal{Z}_\beta + \theta(\mathcal{S}_\beta) \subseteq \mathcal{Z}_\beta$.

(b) *We have*

$$\theta(m) + \theta(n) = \theta(m + \theta(n)) = \theta(\theta(m) + n), \tag{3}$$

$$\beta(m)\beta(n) = \beta(m + \theta(n))\beta(n) + \beta(n + \theta(m))\beta(m), \tag{4}$$

for all $m, n \in \mathcal{S}_\beta$.

Under the assumption that $\mathcal{S}_\beta + \theta(\mathcal{S}_\beta) \subseteq \mathcal{S}_\beta$, if P is a Rota–Baxter operator then the above conditions hold.

Proof. Since P is a monomial linear operator on $\boldsymbol{k}[x]$, the Rota–Baxter relation in (1) is equivalent to

$$\begin{aligned}\beta(m)\beta(n)x^{\theta(m)+\theta(n)} = {}& \beta(m + \theta(n))\beta(n)x^{\theta(m+\theta(n))} \\ & + \beta(\theta(m) + n)\beta(m)x^{\theta(\theta(m)+n)} \quad \text{for all } m, n \in \mathbb{N}.\end{aligned} \tag{5}$$

Suppose (a) and (b) hold. Since $\mathbb{N}$ is the disjoint union of $\mathcal{Z}_\beta$ and $\mathcal{S}_\beta$, we can verify (5) by considering the following four cases:

$$m, n \in \mathcal{Z}_\beta, \quad m \in \mathcal{Z}_\beta, n \in \mathcal{S}_\beta, \quad m \in \mathcal{S}_\beta, n \in \mathcal{Z}_\beta, \quad m, n \in \mathcal{S}_\beta.$$

In the first case we have $\beta(m) = \beta(n) = 0$. Thus (5) holds. In the second case, we have $\beta(m) = 0$ and so (5) becomes $\beta(m + \theta(n))\beta(n) = 0$. Then (5) follows from condition (a). The third case can be treated similarly. In the last case, (5) follows from (3) and (4). Thus P is a Rota–Baxter operator on $\boldsymbol{k}[x]$.

Now assume that $\mathcal{S}_\beta + \theta(\mathcal{S}_\beta) \subseteq \mathcal{S}_\beta$ and that P is a Rota–Baxter operator. Then (5) holds. Taking $m \in \mathcal{Z}_\beta$ and $n \in \mathcal{S}_\beta$, we obtain $0 = \beta(m + \theta(n))\beta(n)x^{\theta(m+\theta(n))}$. Since $\beta(n) \neq 0$, we must have $\beta(m + \theta(n)) = 0$, proving (a). Taking $m, n \in \mathcal{S}_\beta$, then $\beta(m + \theta(n)) \neq 0$ and $\beta(\theta(m) + n) \neq 0$ by the assumption. Then all the coefficients in (5) are nonzero. Thus the degrees of the monomials must be the same; this yields (3), and (4) follows. □

By symmetry, only one of the two identities in (3) is needed. Note also that by definition $A + \varnothing = \varnothing$ for any set A, so that $\mathcal{S}_\beta + \theta(\mathcal{S}_\beta) \subseteq \mathcal{S}_\beta$ and $\mathcal{Z}_\beta + \theta(\mathcal{S}_\beta) \subseteq \mathcal{Z}_\beta$ are automatic in the nondegenerate case. Otherwise, we obtain the following constraint on $\mathcal{S}_\beta$:

Lemma 3.4. *If P is a degenerate monomial Rota–Baxter operator on $\boldsymbol{k}[x]$, then $\mathcal{S}_\beta$ is either empty or infinite. The same applies to $\mathcal{Z}_\beta$.*

Proof. Suppose $\mathcal{S}_\beta \neq \varnothing$ and $|\mathcal{S}_\beta| = t < \infty$. Then we may assume that

$$\mathcal{S}_\beta = \{m_i \in \mathbb{N} \mid 1 \leqslant i \leqslant t, m_1 < \cdots < m_t\}.$$

By (5), $\beta(m_t)^2 = 2\beta(m_t)\beta(m_t + \theta(m_t))$. Since $\beta(m_t) \neq 0$, we have $\beta(m_t) = 2\beta(m_t + \theta(m_t))$, and so $\beta(m_t + \theta(m_t)) \neq 0$. Thus $m_t + \theta(m_t)$ is in $\mathcal{S}_\beta$. By Lemma 3.2, $\theta(m_t) \geqslant 1$. Then $m_t + \theta(m_t) > m_t$, a contradiction. Thus either $\mathcal{S}_\beta = \varnothing$ or $|\mathcal{S}_\beta| = \infty$.

On the other hand, let $\mathcal{Z}_\beta \neq \varnothing$. If $\mathcal{Z}_\beta = \mathbb{N}$, then it is certainly infinite. If $\mathcal{Z}_\beta \neq \mathbb{N}$, then take $k \in \mathcal{S}_\beta$. Since $\mathcal{S}_\theta = \mathcal{S}_\beta$ by Lemma 3.2, we have $\theta(k) > 0$. By Theorem 3.3(a), we obtain $\mathcal{Z}_\beta + \theta(k) \subseteq \mathcal{Z}_\beta$. This implies that $\mathcal{Z}_\beta$ is infinite. □

We now give a general setup for constructing monomial Rota–Baxter operators on $\boldsymbol{k}[x]$. This setup will be applied in Section 3B to construct nondegenerate monomial Rota–Baxter operators and in Section 3C to construct degenerate monomial Rota–Baxter operators.

Theorem 3.5. *Let $\mathcal{S}$ be a subset of $\mathbb{N}$.*

(a) *Let the maps $\theta : \mathcal{S} \to \mathbb{N}^\times$ and $\beta : \mathcal{S} \to \boldsymbol{k}^\times$ satisfy the following conditions*:

(i) $\mathbb{N} \setminus \mathcal{S} + \theta(\mathcal{S}) \subseteq \mathbb{N} \setminus \mathcal{S}$.

(ii) *Equations (3) and (4) are fulfilled for all $m, n \in \mathcal{S}$.*

Extend θ and β to $\mathbb{N}$ by defining $\theta(n) = 0$ and $\beta(n) = 0$ for $n \in \mathbb{N} \setminus \mathcal{S}$. Then $P : \boldsymbol{k}[x] \to \boldsymbol{k}[x]$, defined by $P(x^n) = \beta(n)x^{\theta(n)}$ for $n \in \mathbb{N}$, is a Rota–Baxter operator on $\boldsymbol{k}[x]$.

(b) *Let $\theta : \mathcal{S} \to \mathbb{N}^\times$ satisfy (3) and $\mathbb{N}\setminus\mathcal{S}+\theta(\mathcal{S}) \subseteq \mathbb{N}\setminus\mathcal{S}$. Extend θ to $\mathbb{N}$ by defining $\theta(n) = 0$ for $n \in \mathbb{N} \setminus \mathcal{S}$. For any $c \in \boldsymbol{k}^\times$, define $\beta : \mathbb{N} \to \boldsymbol{k}$ by*

$$\beta(n) = \begin{cases} c/\theta(n) & n \in \mathcal{S}, \\ 0 & n \notin \mathcal{S}. \end{cases} \tag{6}$$

Then $P : \boldsymbol{k}[x] \to \boldsymbol{k}[x]$ defined by $P(x^n) = \beta(n)x^{\theta(n)}$ is a Rota–Baxter operator on $\boldsymbol{k}[x]$.

Proof. (a) This follows from Theorem 3.3.

(b) Under the assumption, for $m, n \in \mathcal{S}$,

$$\begin{aligned} \beta(m+\theta(n))\beta(n) + \beta(\theta(m)+n)\beta(m) \\ &= \frac{c^2}{\theta(m+\theta(n))\theta(n)} + \frac{c^2}{\theta(\theta(m)+n)\theta(m)} \\ &= \frac{c^2}{(\theta(m)+\theta(n))\theta(n)} + \frac{c^2}{(\theta(m)+\theta(n))\theta(m)} \\ &= \frac{c}{\theta(m)}\frac{c}{\theta(n)} = \beta(m)\beta(n) \end{aligned}$$

holds. Thus θ and β satisfy the conditions in Theorem 3.3 for P to be a Rota–Baxter operator on $\boldsymbol{k}[x]$. □

3B. *Nondegenerate case.* As mentioned earlier, for a nondegenerate monomial linear operator P on $\boldsymbol{k}[x]$, the conditions $\mathcal{S}_\beta + \theta(\mathcal{S}_\beta) \subseteq \mathcal{S}_\beta$ and $\mathcal{Z}_\beta + \theta(\mathcal{S}_\beta) \subseteq \mathcal{Z}_\beta$ are automatic. Thus we obtain a characterization of nondegenerate monomial Rota–Baxter operators from Theorems 3.3 and 3.5:

Corollary 3.6. (a) *Let P be a nondegenerate monomial linear operator on $\boldsymbol{k}[x]$ as in (2). Then P is a Rota–Baxter operator if and only if the sequences θ and β satisfy (3) and (4) for all $m, n \in \mathbb{N}$. In this case, $\theta(n) \neq 0$ for all $n \in \mathbb{N}$.*

(b) *If a sequence $\theta : \mathbb{N} \to \mathbb{N}$ is nonzero and satisfies (3), then for any $c \in \boldsymbol{k}^\times$, the map $\beta : \mathbb{N} \to \boldsymbol{k}$ given by $\beta(n) := c/\theta(n)$ satisfies (4) and hence gives a Rota–Baxter operator on $\boldsymbol{k}[x]$.*

Equation (3) characterizes θ as an averaging operator defined as follows:

Definition 3.7. (a) A map $\theta : S \to S$ on a semigroup S is called an *averaging operator* if

$$\theta(m\theta(n)) = \theta(m)\theta(n) = \theta(\theta(m)+n) \quad \text{for all } m, n \in S.$$

(b) A linear map $\Theta : R \to R$ on a $\boldsymbol{k}$-algebra R is called an *averaging operator* if Θ is an averaging operator on the multiplicative semigroup of R.

The study of averaging operators can be tracked back to Reynolds [1895] and Birkhoff [1950]. We refer the reader to [Pei and Guo 2014] and the references therein for further details.

By Corollary 3.6, a nondegenerate monomial operator P on $\boldsymbol{k}[x]$ is a Rota–Baxter operator if and only if the map θ is an averaging operator on the semigroup $(\mathbb{N}, +)$, and the corresponding $\boldsymbol{k}$-linear operator $\Theta : x^n \mapsto x^{\theta(n)}$ makes $(\boldsymbol{k}[x], \Theta)$ into an averaging algebra. We write $\mathcal{A}$ for the set of all nondegenerate averaging operators, i.e., sequences $\theta : \mathbb{N} \to \mathbb{N}^\times$ satisfying (3). We describe $\mathcal{A}$ as the first step to determine nondegenerate monomial Rota–Baxter operators on $\boldsymbol{k}[x]$. We denote the free semigroup over $\mathbb{N}^\times$ by $S(\mathbb{N}^\times)$, so the elements $\sigma \in S(\mathbb{N}^\times)$ are *finite sequences* $(\sigma_0, \dots, \sigma_{d-1})$ of positive numbers having any length $d > 0$.

Theorem 3.8. *There is a bijective correspondence $\Phi : \mathcal{A} \to S(\mathbb{N}^\times)$ given by*

$$\Phi(\theta) = (\theta(0), \dots, \theta(d-1))/d$$

with

$$d := \min\{j \in \mathbb{N}^\times \mid \theta(r + j) = \theta(r) + j \text{ for all } r \in \mathbb{N}\},$$

whose inverse maps $\sigma := (\sigma_0, \dots, \sigma_{d-1}) \in S(\mathbb{N}^\times)$ to the map $\theta : \mathbb{N} \to \mathbb{N}^\times$ defined by $\theta(n) = (\ell + \sigma_j)d$ for $n = \ell d + j$ with $\ell \in \mathbb{N}$ and $0 \leqslant j < d$. Moreover, we have $\operatorname{im}\theta = d\mathbb{N}_{\geqslant s}$ *for $s := \min(\sigma)$.*

Proof. First consider $\theta \in \mathcal{A}$. Defining the map $\tilde{\theta} := \theta - \mathrm{id}_{\mathbb{N}} : \mathbb{N} \to \mathbb{Z}$, one obtains from (3) that $\tilde{\theta}(m + \theta(n)) = \tilde{\theta}(m)$ for all $m, n \in \mathbb{N}$. Hence $\tilde{\theta}$ is periodic, and d is well-defined as the primitive period of $\tilde{\theta}$. Since every $\theta(n)$ is also a period of $\tilde{\theta}$, this implies $\operatorname{im}\theta \subseteq d\mathbb{N}^\times$ so that the given map $\Phi : \mathcal{A} \to S(\mathbb{N}^\times)$ is well-defined.

Next let us write Ψ for the assignment $\sigma \mapsto \theta$ defined above. By checking (3) one sees that this yields a well-defined map $\Psi : S(\mathbb{N}^\times) \to \mathcal{A}$.

Now we prove $\Phi \circ \Psi = \mathrm{id}_{S(\mathbb{N}^\times)}$. So let $\theta : \mathbb{N} \to \mathbb{N}^\times$ be the map defined as above by a given sequence $(\sigma_0, \dots, \sigma_{d-1}) \in S(\mathbb{N}^\times)$. Since $\tilde{\theta}(n) = \sigma_j d - j$ for $n = \ell d + j$, we see that d is a period of the map $\tilde{\theta}$. Assume d is greater than its primitive period d'. Then $d = kd'$ for $k > 1$, and

$$\sigma_0 k d' = \theta(0) = \tilde{\theta}(0) = \tilde{\theta}(d') = \sigma_{d'} d - d' = (\sigma_{d'} k - 1)\, d'$$

implies $\sigma_0 k = \sigma_{d'} k - 1$, which contradicts $k > 1$. We conclude that d is the primitive period of $\tilde{\theta}$, so the definition of Φ recovers the correct value of d. Moreover, for $j = 0, \dots, d-1$ we have $\theta(j) = \sigma_j d$, which implies $\Phi(\theta) = \sigma$ as required.

It remains to prove the converse relation $\Psi \circ \Phi = \mathrm{id}_{\mathcal{A}}$. Taking an arbitrary $\theta \in \mathcal{A}$, we must prove that it coincides with the sequence θ' defined by $\theta'(\ell d + j) = (\ell + \theta(j)/d)\, d = \ell d + \theta(j)$ for any $\ell \in \mathbb{N}$ and $0 \leqslant j < d$. For these values we must then show that $\theta(\ell d + j) = \ell d + \theta(j)$, which is equivalent to $\tilde{\theta}(\ell d + j) = \tilde{\theta}(j)$. The latter is ensured since we know that $\tilde{\theta}$ has primitive period d.

As noted above, $\operatorname{im} \theta \subseteq d\mathbb{N}^{\times}$ so $\theta/d : \mathbb{N} \to \mathbb{N}^{\times}$ is well-defined. We must show $\operatorname{im}(\theta/d) = \mathbb{N}_{\geqslant s}$. The inclusion from left to right follows since $(\theta/d)(\ell d + j) = \ell + \sigma_j \geqslant \sigma_j \geqslant s$. Now let $n \geqslant s$ be given and write $s = \sigma_j$ for some $j = 0, \ldots, d-1$. Then $\ell := n - \sigma_j \in \mathbb{N}$ is such that $(\theta/d)(\ell d + j) = n$, which establishes the inclusion from right to left. □

As sequences, the relation between $\theta : \mathbb{N} \to \mathbb{N}^{\times}$ and $\sigma : \{0, \ldots, d-1\} \to \mathbb{N}^{\times}$ can be written as $\theta/d = (\sigma, \sigma + 1, \sigma + 2, \ldots)$, where $1, 2, \ldots$ designate constant sequences of length d. More precisely, we have

$$\theta/d = (\sigma_0, \ldots, \sigma_{d-1}, \sigma_0 + 1, \ldots, \sigma_{d-1} + 1, \ldots).$$

Theorem 3.8 yields a construction algorithm for the map θ from a nondegenerate monomial Rota–Baxter operator:

Algorithm 3.9. Every sequence $\theta : \mathbb{N} \to \mathbb{N}$ corresponding to a nondegenerate monomial Rota–Baxter operator on $\boldsymbol{k}[x]$ can be generated as follows:

(a) Let $d \in \mathbb{N}^{\times}$ be given. For each $j = 0, \ldots, d-1$, fix $\sigma_j \in \mathbb{N}^{\times}$.

(b) For $n \in \mathbb{N}$ with $n = \ell d + \bar{n}$, where $\bar{n} \in \{0, \ldots, d-1\}$ is the remainder of n modulo d, define

$$\theta(n) := n + \sigma_{\bar{n}} d - \bar{n} = \ell d + \sigma_{\bar{n}} d.$$

We consider two *extreme cases* of Algorithm 3.9 of particular interest:

Case 1. If $d = 1$ one can only choose $\theta(0) \neq 0$ so that $\theta(n) = n + \theta(0)$ for all $n \in \mathbb{N}$.

Case 2. For $d > 1$ and $\sigma_j = 1, 0 \leqslant j \leqslant d-1$, we obtain $\theta(n) = n + d - \bar{n} = (\ell + 1)d$ with $n = \ell d + \bar{n}$.

Example 3.10. Setting $d = 2$ and $\sigma_0 = \sigma_1 = 1$, we choose the sequence β according to Corollary 3.6(b) with $c = 2$. Then the $\boldsymbol{k}$-linear map $P : \boldsymbol{k}[x] \to \boldsymbol{k}[x]$ defined by

$$P(x^{2k}) = \frac{x^{2k+2}}{k+1} \quad \text{and} \quad P(x^{2k+1}) = \frac{x^{2k+2}}{k+1}$$

is a nondegenerate Rota–Baxter operator on $\boldsymbol{k}[x]$.

Next we determine all β for the sequences θ coming from the two extreme cases above.

Theorem 3.11. (a) *Let $d = 1$ with $\theta(n) = n+k$ for some $k \in \mathbb{N}^\times$. Then $\beta : \mathbb{N} \to \boldsymbol{k}$ satisfies (4) if and only if $\beta(n) = c/\theta(n)$ for some $c \in \boldsymbol{k}^\times$.*

(b) *Let $d > 1$ be given with $\theta(n) = n + d - \bar{n}$. Then $\beta : \mathbb{N} \to \boldsymbol{k}$ satisfies (4) if and only if it is defined as follows: Fix $c_j \in \boldsymbol{k}^\times$ and assign $\beta(j) := 1/c_j$ for $0 \leqslant j \leqslant d-1$. Then for any $n \in \mathbb{N}$ with $n = \ell d + \bar{n}$ define $\beta(n) = \beta(\bar{n})/(\ell+1)$.*

Proof. (a) For a θ of the given form, by (4), we have

$$\beta(n)\beta(0) = \beta(n+k)(\beta(0)+\beta(n)). \tag{7}$$

Set $\beta(0) := a$ for some $a \in \boldsymbol{k}^\times$ and write $c := ka$. Then $\beta(0) = c/k$ and c is in $\boldsymbol{k}^\times$. We next prove $\beta(n) = c/(n+k)$ by induction on $n \geqslant 0$. The base case $n = 0$ is true. Assume $\beta(n) = c/(n+k)$ has been proved for $n \geqslant 0$. By (7), we obtain

$$\beta(n+1-k)\beta(0) = \beta(n+1)(\beta(0)+\beta(n+1-k)). \tag{8}$$

Since $k \geqslant 1$, we obtain $n+1-k \leqslant n$. By the induction hypothesis, we get $\beta(n+1-k) = c/(n+1)$. Then by (8) we have

$$\beta(n+1) = \frac{c^2/(k(n+1))}{c/k + c/(n+1)} = \frac{c}{n+1+k}.$$

This completes the induction. Thus $\beta(n) = c/\theta(n)$ for some $c \in \boldsymbol{k}^\times$ and all $n \in \mathbb{N}$. The converse follows from Corollary 3.6(b).

(b) Taking $\gamma(n) = 1/\beta(n)$, (4) is equivalent to

$$\frac{\gamma(m)}{\gamma(m+\theta(n))} + \frac{\gamma(n)}{\gamma(\theta(m)+n)} = 1. \tag{9}$$

Thus we just need to show that, for a fixed sequence θ in the theorem, a sequence $\gamma : \mathbb{N} \to \boldsymbol{k}$ satisfies (9) if and only if γ is defined by $\gamma(n) = (\ell+1)\gamma(\bar{n})$ if $n = \ell d + \bar{n}$, where the $\gamma(\bar{n}) \in \boldsymbol{k}^\times$ for $\bar{n} \in \{0, \dots, d-1\}$ are arbitrarily preassigned.
($\Rightarrow$) Take $m = 0$ and $n = \ell d$ with $\ell \geqslant 0$ in (9). After simplifying we obtain

$$\gamma((\ell+1)d) = \gamma(\ell d) + \gamma(0).$$

Then by an induction on ℓ, we obtain

$$\gamma(\ell d) = (\ell+1)\gamma(0). \tag{10}$$

Next note that, for $n = \ell d + \bar{n}$,

$$\theta(n) = \ell d + d \tag{11}$$

holds. Then for $j \in \{0, \dots, d-1\}$, taking $m = 0$ and $n = \ell d + j$ in (9) we obtain

$$1 = \frac{\gamma(0)}{\gamma(\theta(\ell d + j))} + \frac{\gamma(\ell d + j)}{\gamma(\theta(0) + \ell d + j)} = \frac{\gamma(0)}{\gamma(\ell d + d)} + \frac{\gamma(\ell d + j)}{\gamma(d + \ell d + j)}.$$

This gives

$$\gamma((\ell+1)d+j)=\frac{\ell+2}{\ell+1}\gamma(\ell d+j),$$

and recursively yields

$$\gamma(\ell d+j)=(\ell+1)\gamma(j).$$

($\Leftarrow$) Conversely, suppose a sequence β is given by $\gamma(n)=(\ell+1)\gamma(\bar{n})$ if $n=\ell d+\bar{n}$, for preassigned $\gamma(\bar{n})$ as specified above. Then for any $m,n\in\mathbb{N}$ with $m=kd+\overline{m}$ and $n=\ell d+\bar{n}$, by (11) we obtain

$$\begin{aligned}\frac{\gamma(m)}{\gamma(m+\theta(n))}&+\frac{\gamma(n)}{\gamma(\theta(m)+n)}\\&=\frac{\gamma(kd+\overline{m})}{\gamma(kd+\overline{m}+\theta(\ell d+\bar{n}))}+\frac{\gamma(\ell d+\bar{n})}{\gamma(\theta(kd+\overline{m})+\ell d+\bar{n})}\\&=\frac{\gamma(kd+\overline{m})}{\gamma(kd+\overline{m}+\ell d+d)}+\frac{\gamma(\ell d+\bar{n})}{\gamma(kd+d+\ell d+\bar{n})}\\&=\frac{(k+1)\gamma(\overline{m})}{(k+\ell+2)\gamma(\overline{m})}+\frac{(\ell+1)\gamma(\bar{n})}{(k+\ell+2)\gamma(\bar{n})}=1.\end{aligned}$$

This is (9). □

In the special case of *polynomial sequences* $\theta:\mathbb{N}\to\mathbb{N}$ and $\alpha=1/\beta:\mathbb{N}\to\boldsymbol{k}$, the range of possibilities can be drastically narrowed down.

Theorem 3.12. *Suppose $\boldsymbol{k}$ is a field containing $\mathbb{Q}$. Let $P:\boldsymbol{k}[x]\to\boldsymbol{k}[x]$ be a nondegenerate monomial linear operator with $P(x^n)=(1/\alpha(n))x^{\theta(n)}$ for $n\in\mathbb{N}$, and assume $\theta(n)$ as well as $\alpha(n)$ are polynomials. Then P is a Rota–Baxter operator if and only if*

$$\theta(n)=n+k\quad\textit{and}\quad\alpha(n)=c(n+k),\tag{12}$$

for some $k\in\mathbb{N}^\times$ and some $c\in\boldsymbol{k}^\times$.

Proof. By Corollary 3.6, the operator P defined by (12) is a Rota–Baxter operator. So we just need to show that any Rota–Baxter operator given by (2) with polynomial sequences $\theta(n)$ and $\alpha(n)$ must satisfy the conditions in (12). Since P is a Rota–Baxter operator, (3) gives the characteristic relation $2\theta(n)=\theta(\theta(n)+n)$. But θ and α are polynomials with $\deg\theta$ and $\deg\alpha$ respectively. Checking degrees, let us first assume $\deg\theta\geqslant 2$. In this case we have

$$\deg 2\theta=\deg\theta<(\deg\theta)^2=\deg\theta(\theta(n)+n),$$

which contradicts the characteristic relation. Thus, $\deg\theta\leqslant 1$, and we can write $\theta(n)=sn+k$ for some $s,k\in\mathbb{N}$. The characteristic relation becomes $2(sn+k)=s(sn+n+k)+k$, or equivalently $(sn+k)(s-1)=0$. If $s\neq 1$, we obtain

$sn+k=0$ for all $n \in \mathbb{N}$. But then $s=k=0$, and P is the zero operator, which contradicts the hypothesis that P is nondegenerate. Therefore $s=1$ and hence $\theta(n)=n+k$ as claimed in (12).

For deriving the second condition of (12), we specialize (4) to obtain $2\alpha(n)=\alpha(\theta(n)+n)$ and hence the recursion $2\alpha(n)=\alpha(2n+k)$. Set $\ell=\deg\alpha$ and suppose the leading coefficient of α is $c \in \boldsymbol{k}^{\times}$. Now taking leading coefficients of the recursion, we get $2c=2^{\ell}c$ and thus $\ell=1$. This means we can write $\alpha(n)=cn+c_0$ for some $c \in \boldsymbol{k}^{\times}$ and $c_0 \in \boldsymbol{k}$. Substituting this into the recursion leads to $2(cn+c_0)=c(2n+k)+c_0$ and hence $\alpha(n)=c(n+k)$ as claimed in (12). It remains to show that $k \neq 0$. But this follows because $P(1)=x^k/ck$ so that necessarily $ck \neq 0$. □

Next we investigate injective monomial Rota–Baxter operators and show them to be analytically modeled. We note first that if P is degenerate, then there exists $n_0 \in \mathbb{N}$ such that $\beta(n_0)=0$, and then $P(x^{n_0})=0$. Thus $\ker P \neq \{0\}$ and P is not injective. Thus any injective monomial Rota–Baxter operator is nondegenerate.

Theorem 3.13. *Let P be a monomial Rota–Baxter operator on $\boldsymbol{k}[x]$. The following statements are equivalent*:

(a) *The operator P is injective.*

(b) *The θ in (2) from P satisfies $\theta(n)=n+k$ for some $k \in \mathbb{N}^{\times}$.*

(c) *There are $k \in \mathbb{N}^{\times}$ and $c \in \boldsymbol{k}^{\times}$ such that $P(x^n)=c\int_0^x t^{n+k-1}\,dt$ and hence $P=cJ_0x^{k-1}$.*

Proof. (a) ⇒ (b) Assume that P is an injective monomial Rota–Baxter operator. Then P is nondegenerate. By Algorithm 3.9, there are $d \geqslant 1$ and $\sigma_j \in \mathbb{N}^{\times}$ for $j \in \{0,\dots,d-1\}$ such that $\theta(n)=\ell d+\sigma_{\bar{n}}d$, where $n=\ell d+\bar{n}$ and $\bar{n}$ is the remainder of n modulo d. Suppose $d>1$. Without loss of generality, we may assume $\sigma_0 \geqslant \sigma_1$ so that $n:=(\sigma_0-\sigma_1)d+1>0$. Since $\theta(0)=\sigma_0 d$, we obtain

$$\theta(n)=\theta((\sigma_0-\sigma_1)d+1)=(\sigma_0-\sigma_1)d+\sigma_1 d=\theta(0),$$

hence θ is not injective. This forces $d=1$. Then, by the first case considered after Algorithm 3.9, we have $\theta(n)=n+k$ for fixed $k \geqslant 1$.

(b) ⇒ (c) For a θ of the given form, by Theorem 3.11(a), we have $\beta(n)=c/\theta(n)$ for some $c \in \boldsymbol{k}^{\times}$. Thus

$$P(x^n)=\beta(n)x^{\theta(n)}=\frac{c}{n+k}x^{n+k}=c\int_0^x t^{n+k-1}\,dt,$$

as needed.

(c) ⇒ (a) Since $P(x^n)=c\displaystyle\int_0^x t^{n+k-1}\,dt=(c/(n+k))x^{n+k}$ for all $n \in \mathbb{N}$, the operator P is injective. □

3C. *Degenerate case.* We next apply Theorem 3.5 to construct degenerate monomial Rota–Baxter operators on $\boldsymbol{k}[x]$ when $\mathcal{S}_\beta$ is either $k\mathbb{N}$, where $k \geqslant 1$, or $\mathbb{N}\backslash(k\mathbb{N})$, where $k \geqslant 2$.

Proposition 3.14. *Let $P(x^n) = \beta(n)x^{\theta(n)}$, $n \in \mathbb{N}$ and define a monomial linear operator on $\boldsymbol{k}[x]$ such that $\mathcal{S}_\beta = k\mathbb{N}$ for some $k > 0$. Then P is a Rota–Baxter operator on $\boldsymbol{k}[x]$ if and only if $\theta(km) = \tilde{\theta}(m) \in \mathcal{S}_\beta{}^\times$ and $\beta(km) = \tilde{\beta}(m), m \geqslant 0$ for maps $\tilde{\theta} : \mathbb{N} \to \mathbb{N}$ and $\tilde{\beta} : \mathbb{N} \to \boldsymbol{k}$ that satisfy*

$$\tilde{\theta}(m_1) + \tilde{\theta}(m_2) = \tilde{\theta}\left(m_1 + \frac{1}{k}\tilde{\theta}(m_2)\right) = \tilde{\theta}\left(\frac{1}{k}\tilde{\theta}(m_1) + m_2\right) \tag{13}$$

and

$$\tilde{\beta}(m_1)\tilde{\beta}(m_2) = \tilde{\beta}\left(m_1 + \frac{1}{k}\tilde{\theta}(m_2)\right)\tilde{\beta}(m_2) + \tilde{\beta}\left(m_2 + \frac{1}{k}\tilde{\theta}(m_1)\right)\tilde{\beta}(m_1) \tag{14}$$

for all $m_1, m_2 \in \mathbb{N}$.

Proof. Since $\mathcal{S}_\beta = \{km \mid m \in \mathbb{N}\}$, $\mathcal{Z}_\beta = \{km+i \mid 1 \leqslant i \leqslant k-1, m \in \mathbb{N}\}$. Suppose P is a Rota–Baxter operator on $\boldsymbol{k}[x]$. Then by (1), we have $P(x^{km_1+i}P(x^{km_2})) = 0$ for all $m_1, m_2 \in \mathbb{N}$ and $1 \leqslant i \leqslant k-1$. Thus $\beta(km_2)\beta(km_1 + i + \theta(km_2)) = 0$. Since $\beta(km_2) \neq 0$, we obtain $\beta(km_1 + i + \theta(km_2)) = 0$. Then $km_1 + i + \theta(km_2)$ is in $\mathcal{Z}_\beta$, and then $i + \theta(km_2)$ is in $\mathcal{Z}_\beta$ for $1 \leqslant i \leqslant k-1$. Suppose that there exists $m_0 \in \mathbb{N}$ such that $\theta(km_0) \not\equiv 0 \pmod{k}$. Then there exists $1 \leqslant i_0 \leqslant k-1$ such that $i_0 + \theta(km_0) \equiv 0 \pmod{k}$. So $i_0 + \theta(km_0)$ is in $\mathcal{S}_\beta$ by the definition of $\mathcal{S}_\beta$. This is a contradiction to the fact proved above that $i + \theta(km_2)$ is in $\mathcal{Z}_\beta$ for $1 \leqslant i \leqslant k-1$.

Thus $\theta(km)$ is in $\mathcal{S}_\beta{}^\times$ for all $m \in \mathbb{N}$. So $kn + \theta(km)$ is in $\mathcal{S}_\beta$ for all $n, m \in \mathbb{N}$. By Theorem 3.3, equations (3) and (4) hold. Let $\tilde{\theta}(m) := \theta(km)$ and let $\tilde{\beta}(m) := \beta(km)$, $m \in \mathbb{N}$. Thus $\tilde{\beta}(m) \neq 0$ for all $m \in \mathbb{N}$. Then by (3) and (4), equations (13) and (14) hold. This is what we want. The converse follows from Theorem 3.5(a). $\square$

Proposition 3.14 gives a large class of monomial Rota–Baxter operators on $\boldsymbol{k}[x]$ with $\mathcal{S}_\beta = k\mathbb{N}$, reducing to Corollary 3.6 for $k = 1$. On the other hand, Theorem 3.5 also gives the following result on monomial Rota–Baxter operators on $\boldsymbol{k}[x]$, where $\mathcal{S}_\beta$ is now complementary to Proposition 3.14:

Proposition 3.15. *Let $P(x^n) = \beta(n)x^{\theta(n)}$ be a monomial linear operator on $\boldsymbol{k}[x]$ with $\mathcal{S}_\beta = \mathbb{N} \setminus k\mathbb{N}$ for some $k \geqslant 2$.*

(a) *For any $t \in \mathbb{N}^\times$ one obtains a degenerate monomial RBO by setting $\theta(km+i) = k(m+t)$ and $\theta(km) = 0$ for $m \in \mathbb{N}$ and $1 \leqslant i \leqslant k-1$, choosing β as in Theorem 3.5(b).*

(b) *Assume that $\theta(i) = k$ for $1 \leqslant i \leqslant k-1$. Then θ corresponds to a degenerate monomial RBO on $\boldsymbol{k}[x]$ if and only if $\theta(km+i) = k(m+1)$ for all $m \in \mathbb{N}$ and $1 \leqslant i \leqslant k-1$.*

Proof. (a) By our assumption on $\mathcal{S}_\beta$, $\mathcal{Z}_\beta = \{km \mid m \in \mathbb{N}\}$. By assumption $\theta(km+i) = k(m+t)$ for all $m \in \mathbb{N}$ and $1 \leqslant i \leqslant k-1$, hence we obtain $\mathcal{Z}_\beta + \theta(\mathcal{S}_\beta) \subseteq \mathcal{Z}_\beta$. Since

$$\theta(km_1+i_1)+\theta(km_2+i_2) = k(m_1+m_2+2t),$$
$$\theta(km_1+i_1+\theta(km_2+i_2)) = \theta(k(m_1+m_2+t)+i_1) = k(m_1+m_2+2t),$$

for all $m_1, m_2 \in \mathbb{N}$ and $1 \leqslant i_1, i_2 \leqslant k-1$, we obtain (3). Thus we may apply Theorem 3.5(b) to obtain a degenerate RBO P on $\boldsymbol{k}[x]$.

(b) Assume first that P is a monomial RBO on $\boldsymbol{k}[x]$. Then by (5), for all $m \in \mathbb{N}$ and $1 \leqslant i \leqslant k-1$, we obtain $\beta(km+i)\beta(km+\theta(km+i)) = 0$. Since $\beta(km+i) \neq 0$, we have $\beta(km+\theta(km+i)) = 0$, so $km+\theta(km+i)$ is in $\mathcal{Z}_\beta$. From $\mathcal{Z}_\beta = k\mathbb{N}$ we infer $\theta(km+i) \in \mathcal{Z}_\beta$. Thus $\mathcal{S}_\beta + \theta(\mathcal{S}_\beta) \subseteq \mathcal{S}_\beta$. By Theorem 3.3, (3) holds. We now prove that $\theta(km+i) = k(m+1)$ by induction on $m \geqslant 0$. The base case $m = 0$ is immediate from our assumption. Assume that $\theta(km+i) = k(m+1)$ has been proved for $m \geqslant 0$. By (3), we have

$$\theta(k(m+1)+i) = \theta(km+i+\theta(i)) = \theta(km+i)+\theta(i).$$

By the induction hypothesis, we get $\theta(k(m+1)+i) = k(m+2)$. This completes the proof.

Conversely, by $\theta(km+i) = k(m+1)$ and (a), we obtain a degenerate RBO P on $\boldsymbol{k}[x]$. □

Example 3.16. Taking $k = 2$ in Propositions 3.14 and 3.15, we obtain the following degenerate monomial Rota–Baxter operators on $\boldsymbol{k}[x]$:

(a) $P(x^{2k}) = x^{2(k+1)}/(k+1)$ and $P(x^{2k+1}) = 0$ for all $k \in \mathbb{N}$.

(b) $P(x^{2k}) = 0$ and $P(x^{2k+1}) = x^{2(k+1)}/(k+1)$ for all $k \in \mathbb{N}$.

The above examples may also be regarded as special cases of the following result:

Proposition 3.17. *Let $P_0 \in \mathrm{RBO}(R)$ for a $\boldsymbol{k}$-algebra R. Assume that ϕ is a $\boldsymbol{k}$-linear operator on R such that $E := P_0(\operatorname{im}\phi)$ is a nonunitary $\boldsymbol{k}$-subalgebra. If ϕ is a homomorphism of the E-module R, then $P_0 \circ \phi$ is also a Rota–Baxter operator on R.*

Proof. This follows immediately since

$$\begin{aligned}(P_0\circ\phi)(a)(P_0\circ\phi)(b) &= P_0(\phi(a))P_0(\phi(b)) \\ &= P_0\big(\phi(a)P_0(\phi(b))\big) + P_0(P_0(\phi(a))\phi(b)) \\ &= (P_0\circ\phi)(a(P_0\circ\phi)(b)) + (P_0\circ\phi)((P_0\circ\phi)(a)b)\end{aligned}$$

for all $a, b \in R$. □

For $R = \boldsymbol{k}[x]$, let $\phi : f(x) \mapsto \frac{1}{2}(f(x) + f(-x))$ be the projector onto the $\boldsymbol{k}$-subspace spanned by the even monomials, and set $P_0 = 2J_0 x$. Then

$$(P_0 \circ \phi)(x^n) := \begin{cases} x^{2(k+1)}/(k+1) & \text{if } n = 2k, \\ 0 & \text{if } n = 2k+1, \end{cases}$$

for all $k \in \mathbb{N}$ so that $P_0 \circ \phi$ is the same as P in Example 3.16(a). On the other hand, choosing ϕ as the projector $f(x) \mapsto \frac{1}{2}(f(x) - f(-x))$ onto the space of odd monomials and setting $P_0 = 2J_0$ yields

$$(P_0 \circ \phi)(x^n) := \begin{cases} 0 & \text{if } n = 2k, \\ x^{2(k+1)}/(k+1) & \text{if } n = 2k+1, \end{cases}$$

for all $k \in \mathbb{N}$ so that $P_0 \circ \phi$ is the same as P in Example 3.16(b). In both cases, E is the nonunitary algebra of nonconstant even monomials.

Proposition 3.18. *Let $P(x^n) = \beta(n) x^{\theta(n)}$ be a nonzero degenerate monomial RBO on $\boldsymbol{k}[x]$ satisfying the condition $\mathcal{S}_\beta + \theta(\mathcal{S}_\beta) \subseteq \mathcal{S}_\beta$.*

(a) *There exists a map $\sigma : \mathbb{N} \to \mathcal{S}_\beta$ such that $P_0(x^n) := P(x^{\sigma(n)})$ defines a nondegenerate monomial RBO on $\boldsymbol{k}[x]$.*

(b) *We have the disjoint union*

$$\mathcal{S}_\beta = C \uplus (s_1 + e\mathbb{N}) \uplus \cdots \uplus (s_k + e\mathbb{N}), \tag{15}$$

where $C \subset \mathbb{N}$ is finite, $k < e \in \mathbb{N}^\times$, and $s_1, \ldots, s_k \in \mathcal{S}_\beta$ are incongruent modulo e (in the sense that $x - y \notin e\mathbb{Z}$) such that $s_1 - e, \ldots, s_k - e \notin C$. Moreover, there exists a finite set $E \subset \mathcal{S}_\beta$ such that θ is determined uniquely by its values on E.

Proof. Since P is nonzero, both $\mathcal{S}_\beta \neq \varnothing$ and $\mathcal{Z}_\beta \neq \varnothing$ are infinite by Lemma 3.4. From (3) and the condition $\mathcal{S}_\beta + \theta(\mathcal{S}_\beta) \subseteq \mathcal{S}_\beta$ we see that $T := \theta(\mathcal{S}_\beta)$ is additively closed. As in the proof of Theorem 3.8, one checks that $\theta - \mathrm{id}_\mathbb{N}$ is periodic on $\mathcal{S}_\beta$ with primitive period d and $T \subseteq d\mathbb{N}^\times$ so that $d \mid e := \gcd(T)$. Hence T/e is a numerical semigroup [Rosales and García-Sánchez 1999, Proposition 10.1], meaning a subsemigroup of $\mathbb{N}^\times$ with a finite complement $G \subseteq \mathbb{N}^\times$ of so-called gaps. Thus we obtain $T = e\mathbb{N}^\times \setminus eG$. We write $f \in \mathbb{N}$ for the Frobenius number of T/e, meaning the greatest element of G for $G \neq \varnothing$ and $f = 0$ otherwise.

(a) Fix an element s of $\mathcal{S}_\beta$. We define $\sigma : \mathbb{N} \to \mathcal{S}_\beta$ as follows. For $n \in \mathbb{N}$, write $n = \ell e + r$ with $\ell \geqslant 0$ and $0 \leqslant r < e$. Define $\sigma(\ell e + r) := (f + \ell + 1)e + s$. Then $\sigma : \mathbb{N} \to \mathcal{S}_\beta$ follows from the condition $\mathcal{S}_\beta + T \subseteq \mathcal{S}_\beta$ since $(f + \ell + 1)e \in T$ for all $\ell \geqslant 0$. We show now that

$$\sigma\big(n + \theta(\sigma(m))\big) = \sigma(n) + \theta(\sigma(m)) \tag{16}$$

for all $m, n \in \mathbb{N}$. We have $\theta(\sigma(m)) = te \in T$ for some $t \notin G$, and we may write $n = \ell e + r$ with $0 \leqslant r < e$ and $\ell \geqslant 0$. Then one computes $\sigma(r) + (\ell + t)e$ for both sides of (16).

Let us now prove that P_0 satisfies (1), or equivalently $\mathrm{RB}(P_0, P_0) = 0$. Since the latter is a symmetric bilinear form and $\boldsymbol{k}[x]$ has characteristic zero, the polarization identity implies that it suffices to prove $\mathrm{RB}(P_0, P_0)(u, u) = 0$ for all $u \in \boldsymbol{k}[x]$. Of course we may restrict ourselves to the canonical basis $u = x^n$, so it remains to show $P(x^{\sigma(n)})^2 = 2P_0(x^n P(x^{\sigma(n)}))$. Applying the definition of P, this is equivalent to

$$\beta(\sigma(n))^2 x^{2\theta(\sigma(n))} = 2\beta(\sigma(n)) P_0(x^{n+\theta(\sigma(n))}),$$

and we may use (16) to expand the right-hand side further to

$$2\beta(\sigma(n))\beta\big(\sigma(n) + \theta(\sigma(n))\big)x^{\theta(\sigma(n)+\theta(\sigma(n)))}.$$

But now we may apply Equations (3) and (4) from Theorem 3.3 to conclude that this is equal to the left-hand side. Hence P_0 is indeed a monomial RBO on $\boldsymbol{k}[x]$. Clearly $P_0(x^n) = P(x^{\sigma(n)}) \neq 0$ since $\sigma(n) \in \mathcal{S}_\beta$, so P_0 is nondegenerate.

(b) For defining $s_1, \dots, s_k$, consider first the sets $\Sigma_i := \mathcal{S}_\beta \cap (i + e\mathbb{N})$ for $i \in \{0, \dots, e-1\}$. Suppressing the empty ones, we reindex the rest as $\Sigma_1, \dots, \Sigma_k$ for $1 \leqslant k \leqslant e$. Then for any $i \in \{1, \dots, k\}$ there exists $\sigma_i \in \Sigma_i$ such that $\sigma_i + e\mathbb{N} \subseteq \Sigma_i$. Indeed, one may choose $\sigma_i = \sigma_i' + (f+1)e$ for any $\sigma_i' \in \Sigma_i$ since then $(f+1)e \in T$, and the hypothesis $\mathcal{S}_\beta + T \subseteq \mathcal{S}_\beta$ implies the required condition $\sigma_i + e\mathbb{N} \subseteq \Sigma_i$. Let $s_i \in \Sigma_i$ be minimal such that the condition is satisfied; this implies in particular $s_i - e \notin \mathcal{S}_\beta$. Then clearly $\Sigma_i = C_i \uplus (s_i + e\mathbb{N})$ for finite sets $C_i \subset \mathbb{N}$. Now define $C := C_1 \cup \dots \cup C_k$ to obtain the decomposition (15). We must have $k < e$ since otherwise $\mathcal{Z}_\beta \subseteq \{0, \dots, \max(s_1, \dots, s_e)\}$ is finite, contradicting Lemma 3.4. Finally, note that $E := \mathcal{S}_\beta \setminus (\mathcal{S}_\beta + T)$ is bounded by $\max(s_1, \dots, s_k) + (f+1)e$ and hence finite. Clearly, θ is determined on $\mathcal{S}_\beta \setminus E$ by (3). □

4. Injective Rota–Baxter operators on $\boldsymbol{k}[x]$

For now let $\boldsymbol{k}$ be an arbitrary field of characteristic zero. An important subclass of Rota–Baxter operators P on $\boldsymbol{k}[x]$ are those associated with the standard derivation ∂ in the sense that $\partial \circ P = 1_{\boldsymbol{k}[x]}$. We generalize this for arbitrary $r \in \boldsymbol{k}[x]^\times$ to the *differential law* $\partial \circ P = r$, where r denotes the corresponding multiplication operator. Thus we define

$$\text{(17)} \qquad \mathrm{RBO}_r(\boldsymbol{k}[x]) := \{P \in \mathrm{RBO}(\boldsymbol{k}[x]) \mid \partial \circ P = r\}.$$

Let us now show that the class of all operators satisfying a differential law actually coincides with the class of all injective operators, which we denote by $\mathrm{RBO}_*(\boldsymbol{k}[x])$.

Theorem 4.1.
$$\mathrm{RBO}_*(\boldsymbol{k}[x]) = \bigcup_{r \in \boldsymbol{k}[x]^\times} \mathrm{RBO}_r(\boldsymbol{k}[x]).$$

Proof. The inclusion from right to left is simple as $P(f) = 0$ implies $\partial(P(f)) = rf = 0$ and hence $f = 0$ since $\boldsymbol{k}[x]$ is an integral domain.

Now let $P : \boldsymbol{k}[x] \to \boldsymbol{k}[x]$ be an injective Rota–Baxter operator. Then there exists a linear map $D : \mathrm{im}\, P \to \boldsymbol{k}[x]$ with $D \circ P = 1_{\boldsymbol{k}[x]}$. Adjoining $\boldsymbol{k}$ as constants, one can immediately check that D is a derivation on the unitary subalgebra $J := \boldsymbol{k} + \mathrm{im}\, P$. Note that $P(1) \notin \boldsymbol{k}$ since $P(1) = c$ implies $c^2 = P(1)^2 = 2P(P(1)) = 2c^2$ and hence $c = 0$, contradicting injectivity. This means $\boldsymbol{k} \subsetneq J$. Since $J \subseteq \boldsymbol{k}[x]$ is an integral domain, D extends uniquely to a derivation on the fraction field $K \subseteq \boldsymbol{k}(x)$ of the ring J. By Lüroth's theorem [Cohn 2003, Theorem 11.3.4], the intermediate field $\boldsymbol{k} \subset K \subseteq \boldsymbol{k}(x)$ is a simple transcendental extension of $\boldsymbol{k}$, so there exists $\phi \in \boldsymbol{k}(x) \setminus \boldsymbol{k}$ with $K = \boldsymbol{k}(\phi)$. But then $K \subseteq \boldsymbol{k}(x)$ is an algebraic field extension [van der Waerden 1993, § 73], so the derivation D extends uniquely to $\boldsymbol{k}(x)$ according to [Cohn 2003, Theorem 11.5.3]. But it is well known [Nowicki 1994, Proposition 1.3.2] that every $\boldsymbol{k}$-derivation on $\boldsymbol{k}[x]$ is a multiple of the canonical derivation, so we must have $D = \psi\partial$ for $\psi := D(x)$. Then $D \circ P = 1$ on $\boldsymbol{k}[x]$ implies that $\psi \cdot P(1)' = 1$, so we obtain $D = r^{-1}\partial$ with $r := P(1)' \in \boldsymbol{k}[x]$ and then also $\partial \circ P = r$. ☐

Thus the study of injective Rota–Baxter operators on $\boldsymbol{k}[x]$ reduces to the study of $\mathrm{RBO}_r(\boldsymbol{k}[x])$. As noted in Section 2, all standard integral operators J_a are in $\mathrm{RBO}_1(\boldsymbol{k}[x])$; more generally, the analytically modeled operators $J_a r$ are in $\mathrm{RBO}_r(\boldsymbol{k}[x])$. It is thus tempting to speculate that $\mathrm{RBO}_r(\boldsymbol{k}[x])$ is exhausted by the $J_a r$. For the special case $\boldsymbol{k} = \mathbb{R}$ and $r = x^k$ this will be proved at the end of this section in Theorem 4.9. For the moment, let $\boldsymbol{k}$ be an arbitrary field containing $\mathbb{Q}$.

From integration over the reals, it is well known that the difference between two indefinite integrals is always a definite integral, which may be interpreted as a *measure*. This generalizes to the algebraic setting in the following way:

Lemma 4.2. *Let $r \in \boldsymbol{k}[x]^\times$ and $a \in \boldsymbol{k}$ be arbitrary. Then $P \in \mathrm{End}(\boldsymbol{k}[x])$ satisfies the differential law $\partial \circ P = r$ if and only if $J_a r - P \in \boldsymbol{k}[x]^*$.*

Proof. Since $\partial \circ J_a r = r$, a linear operator P on $\boldsymbol{k}[x]$ satisfies $\partial \circ P = r$ if and only if $\partial \circ \mu = 0$ for $\mu := J_a r - P$. The latter identity holds if and only if $\mathrm{im}\, \mu$ is contained in $\ker \partial = \boldsymbol{k}$. ☐

In analogy to the reals, we call the above linear functional μ the *associated measure* of P. Then the lemma says that the linear operators satisfying the differential law are classified by their associated measures in the sense that

$$\{P \in \mathrm{End}(\boldsymbol{k}[x]) \mid \partial \circ P = r\} = J_a r + \boldsymbol{k}[x]^*,$$

where the initialized point a may be chosen arbitrarily (typically $a = 0$). But in the real case, a measure is more than an arbitrary linear functional; for the algebraic situation this is captured in the following result. Here and henceforth we employ the abbreviation $\star_{r,a}$ for $\star_{J_a r}$, and $\star_r$ for $\star_{r,0}$.

Theorem 4.3. *Let $r \in \boldsymbol{k}[x]^\times$ and $a \in \boldsymbol{k}$ be arbitrary. Then the map defined by $P \mapsto J_a r - P$ is a bijection between* $\mathrm{RBO}_r(\boldsymbol{k}[x])$ *and* $(\boldsymbol{k}[x], \star_{r,a})^\bullet$.

Proof. By Lemma 4.2 and Proposition 2.3(c), we obtain an surjective map

$$\mathrm{RBO}_r(\boldsymbol{k}[x]) \to (\boldsymbol{k}[x], \star_{r,a})^\bullet, \quad P \mapsto J_a r - P.$$

The map is injective since $J_a r - P = J_a r - \widetilde{P}$ implies $P = \widetilde{P}$. □

Thus the preceding classification of operators satisfying a differential law may be refined to

$$\mathrm{RBO}_r(\boldsymbol{k}[x]) = J_a r - (\boldsymbol{k}[x], \star_{r,a})^\bullet.$$

For working out a more explicit description, we specialize to the monomial case $r = x^k$, where we use the abbreviation $\star_k$ for $\star_{x^k}$. To this end, we will determine $(\boldsymbol{k}[x], \star_k)^\bullet$, starting with $k = 0$.

Theorem 4.4. (a) *For any $k \in \mathbb{N}$, the isomorphism $(\boldsymbol{k}[x], \star_k) \cong x^{k+1}\boldsymbol{k}[x]$ of nonunitary algebras holds.*

(b) *There is a bijection $(\boldsymbol{k}[x], \star_0)^\bullet \to \boldsymbol{k}$ that sends μ to $\mu(1)$. In particular, the value $a := \mu(1) \in \boldsymbol{k}$ determines μ uniquely by*

$$\mu(x^n) = \frac{1}{n+1} a^{n+1} \tag{18}$$

for all $n \in \mathbb{N}$. Moreover, the codimension of $\ker \mu$ *equals* 0 *for $a = 0$, and* 1 *for $a \neq 0$.*

Proof. (a) Note that $\{u_n := n x^{n-k-1} \mid n \geqslant k+1\}$ is a $\boldsymbol{k}$-linear basis of $\boldsymbol{k}[x]$ with

$$\begin{aligned} u_m \star_k u_n &= m x^{m-k-1} J_0(x^k \cdot n x^{n-k-1}) + n x^{n-k-1} J_0(x^k \cdot m x^{m-k-1}) \\ &= m x^{m-k-1} x^n + n x^{n-k-1} x^m = (m+n) x^{m+n-k-1} = u_{m+n}. \end{aligned}$$

Thus the $\boldsymbol{k}$-linear map induced by $u_n \mapsto x^n$ $(n \geqslant k+1)$ is an isomorphism $(\boldsymbol{k}[x], \star_k) \to x^{k+1}\boldsymbol{k}[x]$ of nonunitary $\boldsymbol{k}$-algebras as claimed.

(b) Since $x\boldsymbol{k}[x]$ is the free nonunitary commutative $\boldsymbol{k}$-algebra on x, so is $(\boldsymbol{k}[x], \star_0)$ by the isomorphism from (a). Then the bijection follows from the universal property of free nonunitary commutative $\boldsymbol{k}$-algebra on x. Note that under the isomorphism from (a), the generator x of $x\boldsymbol{k}[x]$ corresponds to the generator $1 = u_1$ of $(\boldsymbol{k}[x], \star_0)$.

To prove (18), we use induction on n. For the base case $n = 0$, we get $\mu(1) = a$ by the definition of a. Now suppose (18) has been proved for a fixed n. Since

$$1 \star_0 x^n = J_0(x^n) + x^n J_0(1) = \frac{n+2}{n+1} x^{n+1}$$

and μ is a $\boldsymbol{k}$-algebra homomorphism, we have

$$\mu\left(\frac{n+2}{n+1}x^{n+1}\right) = \mu(1 \star_0 x^n) = \mu(1)\mu(x^n) = \frac{1}{n+1}a^{n+2},$$

applying the induction hypothesis in the last step. Thus we obtain $\mu(x^{n+1}) = (1/(n+2))a^{n+2}$, and the induction is complete. The last statement follows since the codimension of $\ker\mu$ equals the dimension of $\operatorname{im}\mu$ and μ is surjective if and only if $\mu(1) \neq 0$. □

At this juncture, the results accumulated are sufficient for classifying all Rota–Baxter operators P satisfying the differential relation $\partial \circ P = 1_{\boldsymbol{k}[x]}$. This is an important special case since it states that all *indefinite integrals* are analytically modeled.

Theorem 4.5. *We have* $\mathrm{RBO}_1(\boldsymbol{k}[x]) = \{J_a \mid a \in \boldsymbol{k}\}$.

Proof. The inclusion from right to left is clear, so assume $P \in \mathrm{RBO}_1(\boldsymbol{k}[x])$. By Theorem 4.3, there exists $\mu \in (\boldsymbol{k}[x], \star_0)^\bullet$ such that $P = J_0 - \mu$. Setting now $a := \mu(1)$, Theorem 4.4 asserts that $\mu(x^n) = (1/(n+1))a^{n+1}$ for all $n \in \mathbb{N}$. Then

$$P(x^n) = J_0(x^n) - \mu(x^n) = \frac{x^{n+1} - a^{n+1}}{n+1} = J_a(x^n),$$

so that $P = J_a$; and the inclusion from left to right is established. □

For classifying the Rota–Baxter operators P with $\partial \circ P = x^k$ $(k > 0)$ we must determine all algebra homomorphisms μ with respect to the multiplication $\star_k$. At this point, we have to restrict ourselves to the field $\boldsymbol{k} = \mathbb{R}$ since we shall make use of the order on the *reals* in the next two lemmas.

Lemma 4.6. *Let* $\mu : (\mathbb{R}[x], \star_{2\ell+1}) \to \mathbb{R}$ *be an* $\mathbb{R}$*-algebra homomorphism with* $\ell \geqslant 0$. *Then we get* $\mu(1) \geqslant 0$.

Proof. Since $1 \star_{2\ell+1} 1 = 2J_0(x^{2\ell+1}) = x^{2\ell+2}/(\ell+1)$ and μ is an $\mathbb{R}$-algebra homomorphism, we obtain $c^2 = \mu(1 \star_{2\ell+1} 1) = \mu(x^{2\ell+2}/(\ell+1))$, where we have set $c := \mu(1)$. Hence we get the relation $\mu(x^{2\ell+2}) = (\ell+1)c^2$. We have also

$$1 \star_{2\ell+1} x^{2\ell+2} = J_0(x^{4\ell+3}) + x^{2\ell+2} J_0(x^{2\ell+1}) = \frac{3}{4\ell+4} x^{4\ell+4},$$

which implies by the $\mathbb{R}$-algebra homomorphism property and the previous relation that

$$\mu(x^{4\ell+4}) = \tfrac{4}{3}(\ell+1)^2 c^3. \tag{19}$$

Next we observe that $x^{\ell+1} \star_{2\ell+1} x^{\ell+1} = 2x^{\ell+1} J_0(x^{3\ell+2}) = \frac{2}{3}x^{4\ell+4}/(\ell+1)$. Setting $\tilde{c} := \mu(x^{\ell+1})$, this yields yet another relation

$$\mu(x^{4\ell+4}) = \tfrac{3}{2}(\ell+1)\tilde{c}^2. \tag{20}$$

Combining (19) and (20), we obtain $\frac{4}{3}(\ell+1)^2c^3 = \frac{3}{2}(\ell+1)\tilde{c}^2$ and thus $c = \sqrt[3]{(9/(8(\ell+1)))\tilde{c}^2} \geqslant 0$. ☐

Lemma 4.7. *Let $\mu : (\mathbb{R}[x], \star_k) \to \mathbb{R}$ be an $\mathbb{R}$-algebra homomorphism for $k \in \mathbb{N}$. Then there exists a number $a \in \mathbb{R}$ such that $\mu(1) = a^{k+1}/(k+1)$.*

Proof. We set $c := \mu(1)$ and $a := \sqrt[k+1]{(k+1)c}$. If $k = 2\ell+1$ with $\ell \in \mathbb{N}$, Lemma 4.6 implies that $c \geqslant 0$ and we may extract an even root to obtain $a \in \mathbb{R}$. If on the other hand $k = 2\ell$ for $\ell \in \mathbb{N}$, the root in a is odd and hence clearly $a \in \mathbb{R}$ also in this case. ☐

The number a ensured by the previous lemma serves to characterize the associated measure μ of the Rota–Baxter operator underlying the double product $\star_k$. Analytically speaking, $\mu(1)$ is the Riemann integral over $[0, a]$.

Proposition 4.8. *Let $\mu : (\mathbb{R}[x], \star_k) \to \mathbb{R}$ be an $\mathbb{R}$-algebra homomorphism. Then there exists a number $a \in \mathbb{R}$ such that $\mu(x^n) = a^{n+k+1}/(n+k+1)$ for all $n \in \mathbb{N}$. In particular, μ is uniquely determined by a.*

Proof. We prove the claim by induction on $n \in \mathbb{N}$. In the base case $n = 0$, Lemma 4.7 yields $\mu(1) = a^{k+1}/(k+1)$. Now suppose the claim has been proved up to a fixed n. Since

$$1 \star_k x^{n-k} = J_0(x^n) + x^{n-k} J_0(x^k) = \frac{n+k+2}{(n+1)(k+1)} x^{n+1}$$

and μ is an $\mathbb{R}$-algebra homomorphism, we obtain

$$\begin{aligned} \mu\left(\frac{n+k+2}{(n+1)(k+1)} x^{n+1}\right) &= \mu(1 \star_k x^{n-k}) = \mu(1)\mu(x^{n-k}) \\ &= \frac{1}{(n+1)(k+1)} a^{n+k+2}, \end{aligned}$$

where we have applied the induction hypothesis in the last step since $n - k \leqslant n$. But this gives immediately $\mu(x^{n+1}) = a^{n+k+2}/(n+k+2)$, which completes the induction. ☐

Since the number a of the proposition above characterizes the associated measures, we obtain now the *desired classification* of the Rota–Baxter operators P on $\mathbb{R}[x]$ that satisfy the differential relation $\partial \circ P = x^k$. The number a plays the role of the initialization point of the integral (we regain the standard integral J_0 for $a = 0$ since then the associated measure is zero).

Theorem 4.9. *We have* $\mathrm{RBO}_{x^k}(\mathbb{R}[x]) = \{J_a x^k \mid a \in \mathbb{R}\}$ *for any* $k \in \mathbb{N}$.

Proof. The inclusion from right to left is clear, so assume $P \in \mathrm{RBO}_{x^k}(\mathbb{R}[x])$. Then Theorem 4.3 yields an $\mathbb{R}$-algebra homomorphism $\mu : (\mathbb{R}[x], \star_k) \to \mathbb{R}$ such that $P = J_0 x^k - \mu$. By Proposition 4.8, there exists a number $a \in \mathbb{R}$ such that $\mu(x^n) = a^{n+k+1}/(n+k+1)$. Thus

$$P(x^n) = J_0(x^{n+k}) - \mu(x^n) = \frac{x^{n+k+1} - a^{n+k+1}}{n+k+1} = J_a(x^{n+k}),$$

so that $P = J_a x^k$, and the inclusion from left to right is established. □

As mentioned earlier, it is tempting to generalize the above result from monomials to *arbitrary polynomials*. Together with Theorem 4.1, this would imply that:

Conjecture 4.10. $\mathrm{RBO}_*(\mathbb{R}[x]) = \bigcup_{r \in \mathbb{R}[x]^\times} \mathrm{RBO}_r(\mathbb{R}[x]) = \{J_a r \mid a \in \mathbb{R},\ r \in \mathbb{R}[x]^\times\}.$

For this we only need to verify

$$\mathrm{RBO}_r(\mathbb{R}[x]) \subseteq \{J_a r \mid a \in \mathbb{R}\} \quad \text{for any } r \in \mathbb{R}[x]^\times. \tag{21}$$

In the rest of this paper, we add some preliminary results in support of this conjecture. Let us call a Rota–Baxter operator P on $\mathbb{R}[x]$ *initialized* at a point $a \in \mathbb{R}$ if $\mathrm{ev}_a \circ P$ is the zero operator, where $\mathrm{ev}_a : \mathbb{R}[x] \to \mathbb{R}[x]$ denotes evaluation at a. The typical case is when $P = J_a r$. It is easy to see that Conjecture 4.10 is equivalent to the claim that all Rota–Baxter operators in $\mathrm{RBO}_r(\mathbb{R}[x])$ are initialized. Indeed, if P is initialized at a, then we may multiply the differential law $\partial \circ P = r$ by J_a from the left to obtain $P = J_a r$ since we have $J_a \partial = 1_{\mathbb{R}[x]} - \mathrm{ev}_a$. So for proving Conjecture 4.10 one has to determine the initialization point a from a given Rota–Baxter operator P and $r \in \mathbb{R}[x]^\times$. If P is already known to be of the form $J_a r$, this can be done as follows:

Lemma 4.11. *For the Rota–Baxter operator $P = J_a r$ with $a \in \mathbb{R}$ and $r \in \mathbb{R}[x]^\times$, we have*

$$a = \frac{P(2xr' + r) - xr^2}{P(2r') - r^2}, \tag{22}$$

provided $r(a) \neq 0$. On the other hand, if $r(a) = 0$ then $P = (r - J_a r') \circ J_0$.

Proof. Let us first consider the generic case $r(a) \neq 0$. Using the differential law $\partial \circ P = r$, one sees immediately that numerator and denominator are both constants since they vanish under ∂. Moreover, the denominator cannot be zero since

$$P(2r') = \int_a^x (r^2)' = r^2 - r(a)^2 \neq r^2,$$

by the assumption of genericity. Integrating $(r^2 r^{(i)})' = 2rr'r^{(i)} + r^2 r^{(i+1)}$ from a to x, we obtain

$$r^2 r^{(i)} - r(a)^2 r^{(i)}(a) = P(2r'r^{(i)} + rr^{(i+1)}). \tag{23}$$

Assuming r has degree n, we can write

$$r = 1 + r_1 x + r_2 \frac{x^2}{2!} + \cdots + r_n \frac{x^n}{n!},$$

so that $r^{(n-1)} = r_{n-1} + r_n x$ and $r^{(n)} = r_n$. Substituting $i = n-1$ and $r(a)^2 = r^2 - P(2r')$ in (23), we obtain the relation

$$(r_{n-1} + r_n x)r^2 - (r^2 - P(2r'))(r_{n-1} + r_n a) = P(2r_{n-1}r' + 2r_n xr' + r_n r), \tag{24}$$

which simplifies to $(x-a)r^2 = P(2xr' - 2ar' + r)$. Solving this for a gives (22).

Now assume $r(a) = 0$. Then for $f \in \mathbb{R}[x]$ we obtain

$$Pf' = \int_a^x rf' = [rf]_a^x - \int_a^x r'f = rf - J_a r' f,$$

and hence by $(J_0 f)' = f$, the required identity $Pf = (rJ_0)f - (J_a r' J_0)f = (r - J_a r')J_0(f)$. □

Lemma 4.11 suggests the following strategy for proving Conjecture 4.10. Given an arbitrary $P \in \mathrm{RBO}_r(\mathbb{R}[x])$, we determine first the denominator of (22). If it vanishes, we try to find $\widetilde{P} \in \mathrm{RBO}_{r'}(\mathbb{R}[x])$ with $P = (r - \widetilde{P}) \circ J_0$, and we use induction on the degree of r to handle $\widetilde{P}$. In the generic case of nonvanishing denominator, we compute the value of a from (22), and it suffices to prove that P is initialized at a. For doing this, the first step would be to ascertain that $r(a)^2 = r^2 - P(2r')$. This would imply that $P(r')$ vanishes at $x = a$ and hence also $P(2xr' + r)$ by (24). Using the Rota–Baxter axiom and the above relations, one can produce polynomials p such that $P(p)$ vanishes at $x = a$. If this is done for sufficiently many polynomials p to generate $\mathbb{R}[x]$ as a real vector space, we are done. Here is an example of a class of polynomials where one can infer vanishing at $x = a$, provided $r(a)^2 = r^2 - P(2r')$ has been established. For $P = J_a r$, it recovers the fact that $J_a(r'r^{2k+1}) = (2k+2)^{-1} J_a((r^{2k+2})') = (2k+2)^{-1}(r^{2k+2} - r(a)^{2k+2})$.

Lemma 4.12. *Let $P \in \mathrm{RBO}_r(\mathbb{R}[x])$ be arbitrary. Then we obtain $P(r'r^{2k}) = (2k+2)^{-1}(r^{2k+2} - c^{k+1})$ for $c := r^2 - P(2r') \in \mathbb{R}$ and all $k \geqslant 0$.*

Proof. We use induction on k. The base case $k = 0$ is immediate from the definition of c. Now assume the claim for all degrees below a fixed $k > 0$; we prove it for k. By the Rota–Baxter axiom and the definition of c we find

$$\begin{aligned} P(r')^{k+1} &= (k+1)!\, P_{r'}^{k+1}(1) = (k+1)!\, P(r' P_{r'}^{k}(1)) = (k+1) P(r' P(r')^k) \\ &= 2^{-k}(k+1) P(r'(r^2 - c)^k), \end{aligned}$$

where $P_{r'} : \mathbb{R}[x] \to \mathbb{R}[x]$ is defined by $P_{r'}(p) := P(r'p)$. Substituting the defining relation of c on the left-hand side, we obtain $(r^2-c)^{k+1} = 2(k+1)\,P(r'(r^2-c)^k)$, so the binomial theorem yields

$$(2k+2)\,P(r'r^{2k}) = (r^2-c)^{k+1} - 2(k+1)\sum_{l=0}^{k-1}\binom{k}{l}(-c)^{k-l}\,P(r'r^{2l}).$$

Applying the induction hypothesis leads to

$$\begin{aligned}
&(2k+2)\,P(r'r^{2k})\\
&\quad = (r^2-c)^{k+1} - (k+1)\sum_{l=0}^{k-1}\binom{k}{l}\frac{(-c)^{k-l}}{l+1}((r^2)^{l+1} - c^{l+1})\\
&\quad = (r^2-c)^{k+1} + (r^{2k+2} - c^{k+1}) - (k+1)\sum_{l=0}^{k}\binom{k}{l}\frac{(-c)^{k-l}}{l+1}((r^2)^{l+1} - c^{l+1}).
\end{aligned}$$

For evaluating the above sum, just note that integrating $(x-c)^k$ from α to β and using the binomial theorem gives

$$\frac{(\beta-c)^{k+1} - (\alpha-c)^{k+1}}{k+1} = \sum_{l=0}^{k}\binom{k}{l}\frac{(-c)^{k-l}}{l+1}(\beta^{l+1} - \alpha^{l+1}),$$

which may be evaluated at $(\alpha,\beta) = (c, r^2)$ in the previous sum to obtain

$$(2k+2)\,P(r'r^{2k}) = (r^2-c)^{k+1} + r^{2k+2} - c^{k+1} - (r^2-c)^{k+1} = r^{2k+2} - c^{k+1},$$

which completes the induction. □

We conclude with a simple result about the double product $\star$ in the general case of $J_a r$. This lemma is a kind of analogy (though not a generalization) of Theorem 4.4(a). In fact, the two results coincide for $r = x$.

Lemma 4.13. *Let $\star$ be the double product corresponding to the Rota–Baxter operator $J_a r$ and set $\rho = r(a)$. Then the nonunitary subalgebra of $(\boldsymbol{k}[x], \star)$ generated by $u_n = nr^{n-2}r'$ $(n \geqslant 2)$ is isomorphic to the nonunitary subalgebra of $(\boldsymbol{k}[x], \cdot)$ generated by $x^n - \rho^n$ $(n \geqslant 2)$.*

Proof. The double product of the basis elements u_m $(m \geqslant 2)$ and u_n $(n \geqslant 2)$ is given by

$$\begin{aligned}
u_m \star u_n &= mnr^{m-2}r'J_a r^{n-1}r' + mnr^{n-2}r'J_a r^{m-1}r'\\
&= mr^{m-2}r'(r^n - \rho^n) + nr^{n-2}r'(r^m - \rho^m)\\
&= u_{m+n} - \rho^n u_m - \rho^m u_n,
\end{aligned}$$

so the $\boldsymbol{k}$-linear map ϕ defined by $\phi(u_m) = x^m - \rho^m$ is a homomorphism of nonunitary $\boldsymbol{k}$-algebras since

$$(x^m - \rho^m)(x^n - \rho^n) = (x^{m+n} - \rho^{m+n}) - \rho^n(x^m - \rho^m) - \rho^m(x^n - \rho^n).$$

The map ϕ is clearly bijective as it maps a $\boldsymbol{k}$-basis to a $\boldsymbol{k}$-basis. □

Acknowledgements

This work was supported by the National Science Foundation of US (Grant No. DMS1001855), the Engineering and Physical Sciences Research Council of UK (Grant No. EP/I037474/1) and the National Natural Science Foundation of China (Grant No. 11371178). L. Guo and S. Zheng also thank the Kavli Institute for Theoretical Physics China and the Morning Center of Mathematics in Beijing for hospitality and support.

References

[Aguiar 2001] M. Aguiar, "On the associative analog of Lie bialgebras", *J. Algebra* **244**:2 (2001), 492–532. MR 2003c:17035 Zbl 0991.16033

[Atkinson 1963] F. V. Atkinson, "Some aspects of Baxter's functional equation", *J. Math. Anal. Appl.* **7** (1963), 1–30. MR 27 #5135 Zbl 0118.12903

[Bai 2007] C. Bai, "A unified algebraic approach to the classical Yang–Baxter equation", *J. Phys. A* **40**:36 (2007), 11073–11082. MR 2009b:17014 Zbl 1118.17008

[Bai et al. 2010] C. Bai, L. Guo, and X. Ni, "Nonabelian generalized Lax pairs, the classical Yang–Baxter equation and PostLie algebras", *Comm. Math. Phys.* **297**:2 (2010), 553–596. MR 2011i:17032 Zbl 1206.17020

[Bai et al. 2013] C. Bai, O. Bellier, L. Guo, and X. Ni, "Splitting of operations, Manin products, and Rota–Baxter operators", *Int. Math. Res. Not.* **2013**:3 (2013), 485–524. MR 3021790

[Baxter 1960] G. Baxter, "An analytic problem whose solution follows from a simple algebraic identity", *Pacific J. Math.* **10** (1960), 731–742. MR 22 #9990 Zbl 0095.12705

[Birkhoff 1950] G. Birkhoff, "Moyennes des fonctions bornées", pp. 143–153 in *Algèbre et Théorie des Nombres*, Colloques Internationaux du Centre National de la Recherche Scientifique **24**, Centre National de la Recherche Scientifique, Paris, 1950. MR 13,361d Zbl 0041.44306

[Cartier 1972] P. Cartier, "On the structure of free Baxter algebras", *Advances in Math.* **9** (1972), 253–265. MR 49 #2807 Zbl 0267.60052

[Cohn 2003] P. M. Cohn, *Basic algebra: Groups, rings and fields*, Springer, London, 2003. MR 2003m: 00001 Zbl 1003.00001

[Connes and Kreimer 2000] A. Connes and D. Kreimer, "Renormalization in quantum field theory and the Riemann–Hilbert problem, I: The Hopf algebra structure of graphs and the main theorem", *Comm. Math. Phys.* **210**:1 (2000), 249–273. MR 2002f:81070 Zbl 1032.81026

[Ebrahimi-Fard et al. 2004] K. Ebrahimi-Fard, L. Guo, and D. Kreimer, "Spitzer's identity and the algebraic Birkhoff decomposition in pQFT", *J. Phys. A* **37**:45 (2004), 11037–11052. MR 2006b:81332 Zbl 1062.81113

[Ebrahimi-Fard et al. 2006] K. Ebrahimi-Fard, L. Guo, and D. Manchon, "Birkhoff type decompositions and the Baker–Campbell–Hausdorff recursion", *Comm. Math. Phys.* **267**:3 (2006), 821–845. MR 2008c:17020 Zbl 1188.17020

[Gao et al. 2014] X. Gao, L. Guo, and S. Zheng, "Construction of free commutative integro-differential algebras by the method of Gröbner–Shirshov bases", *J. Algebra Appl.* **13**:5 (2014), 1350160. MR 3190089 Zbl 06303780

[Guo 2009] L. Guo, "What is. . .a Rota–Baxter algebra?", *Notices Amer. Math. Soc.* **56**:11 (2009), 1436–1437. MR 2590727 Zbl 1184.16038

[Guo 2012] L. Guo, *An introduction to Rota–Baxter algebra*, Surveys of Modern Mathematics **4**, International Press, Somerville, MA, 2012. MR 3025028 Zbl 1271.16001

[Guo and Keigher 2000] L. Guo and W. Keigher, "Baxter algebras and shuffle products", *Adv. Math.* **150**:1 (2000), 117–149. MR 2001g:05015 Zbl 0947.16013

[Guo and Keigher 2008] L. Guo and W. Keigher, "On differential Rota–Baxter algebras", *J. Pure Appl. Algebra* **212**:3 (2008), 522–540. MR 2008i:16037 Zbl 1185.16038

[Guo and Zhang 2008] L. Guo and B. Zhang, "Renormalization of multiple zeta values", *J. Algebra* **319**:9 (2008), 3770–3809. MR 2009b:11155 Zbl 1165.11071

[Guo et al. 2014] L. Guo, G. Regensburger, and M. Rosenkranz, "On integro-differential algebras", *J. Pure Appl. Algebra* **218**:3 (2014), 456–473. MR 3124211 Zbl 06244927

[Nowicki 1994] A. Nowicki, *Polynomial derivations and their rings of constants*, Uniwersytet Mikołaja Kopernika, Toruń, 1994. MR 2553232 Zbl 1236.13023

[Pei and Guo 2014] J. Pei and L. Guo, "Averaging algebra, Schröder numbers and rooted trees", *J. Algebraic Combinatorics* (December 2014).

[Reynolds 1895] O. Reynolds, "On the dynamical theory of incompressible viscous fluids and the determination of the criterion", *Phil. Trans. Roy. Soc. A* **186** (1895), 123–164. JFM 26.0872.02

[Rosales and García-Sánchez 1999] J. C. Rosales and P. A. García-Sánchez, *Finitely generated commutative monoids*, Nova Science Publishers, Commack, NY, 1999. MR 2000d:20074 Zbl 0966.20028

[Rosenkranz and Regensburger 2008] M. Rosenkranz and G. Regensburger, "Solving and factoring boundary problems for linear ordinary differential equations in differential algebras", *J. Symbolic Comput.* **43**:8 (2008), 515–544. MR 2009d:34067 Zbl 1151.34008

[Rota 1969a] G.-C. Rota, "Baxter algebras and combinatorial identities, I", *Bull. Amer. Math. Soc.* **75** (1969), 325–329. MR 39 #5387 Zbl 0192.33801

[Rota 1969b] G.-C. Rota, "Baxter algebras and combinatorial identities, II", *Bull. Amer. Math. Soc.* **75** (1969), 330–334. MR 39 #5387 Zbl 0319.05008

[Rota 1995] G.-C. Rota, *Gian-Carlo Rota on combinatorics: Introductory papers and commentaries*, edited by J. P. S. Kung, Contemporary Mathematicians, Birkhäuser, Boston, MA, 1995. MR 99b:01027 Zbl 0841.01031

[Semenov-Tian-Shansky 1983] M. A. Semenov-Tian-Shansky, "What is a classical r-matrix?", *Funct. Ana. Appl.* **17**:4 (1983), 259–272. Zbl 0535.58031

[Spitzer 1956] F. Spitzer, "A combinatorial lemma and its application to probability theory", *Trans. Amer. Math. Soc.* **82** (1956), 323–339. MR 18,156e Zbl 0071.13003

[van der Waerden 1993] L. B. van der Waerden, *Algebra, I*, 9th ed., Springer, Berlin, 1993. In German. Zbl 0781.12002

Received May 30, 2014.

SHANGHUA ZHENG
DEPARTMENT OF MATHEMATICS
JIANGXI NORMAL UNIVERSITY
NANCHANGE, JIANGXI 330022
CHINA
zheng2712801@163.com

LI GUO
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
RUTGERS UNIVERSITY AT NEWARK
216 SMITH HALL
101 WARREN STREET
NEWARK, NJ 07102
UNITED STATES
liguo@rutgers.edu

MARKUS ROSENKRANZ
SCHOOL OF MATHEMATICS
STATISTICS AND ACTUARIAL SCIENCE
UNIVERSITY OF KENT
CANTERBURY
CT2 7NF
UNITED KINGDOM
M.Rosenkranz@kent.ac.uk

PACIFIC JOURNAL OF MATHEMATICS
Vol. 275, No. 2, 2015
dx.doi.org/10.2140/pjm.2015.275.509

CORRECTION TO THE ARTICLE QUIVER GRASSMANNIANS, QUIVER VARIETIES AND THE PREPROJECTIVE ALGEBRA

ALISTAIR SAVAGE AND PETER TINGLEY

Volume **251**:2 (2011), 393–429

For quivers not of finite or affine type, certain isomorphisms asserted in the article under correction do not hold, as pointed out by Sarah Scherotzke. This note describes the affected results briefly. A corrected version of the paper can be found at arXiv 0909.3746.

The original published version of this paper contained the following errors. We thank Sarah Scherotzke for bringing this to our attention.

Error 1. If $\mathfrak{g}$ is not of finite or affine type, then the Nakajima quiver variety $\Lambda(v, w)$ is not actually isomorphic to the variety $\mathrm{Gr}(v, q^w)$ of all v-dimensional subrepresentations of the injective module q^w. In fact, beyond affine type, $\mathrm{Gr}(v, q^w)$ does not have a natural variety structure, or, at least, is not finite-dimensional. This is because there are continuous families of nonisomorphic modules, all of which have a nontrivial extension with some one-dimensional simple module S_i.

There are two ways to modify the statement to make it true, and, with either of these modifications, the work in the original paper does prove the correct result. One must replace $\mathrm{Gr}(v, q^w)$ with either the variety $\mathrm{NGr}(v, q^w)$ of nilpotent v-dimensional subrepresentations of q^w, or with the variety $\mathrm{Gr}(v, \tilde{q}^w)$ of all v-dimensional subrepresentations, but where the injective hull q^w in the category of all representations of the preprojective algebra has been replaced with the injective hull $\tilde{q}^w$ in the category of locally nilpotent representations. Our work shows that these are naturally isomorphic, and are also isomorphic to $\Lambda(v, w)$.

Error 2. Lemma 2.9 (which essentially asserted that $\mathrm{Gr}(v, q^w)$ and $\mathrm{NGr}(v, q^w)$ were isomorphic) is false beyond affine type, and should be removed. The proof is simply incorrect. In fact, this caused most of the issues in Error 1.

MSC2010: 16G20.
Keywords: quiver grassmannian, quiver variety, preprojective algebra.

A corrected version of the paper that addresses these points can be found at arXiv 0909.3746. We show that both of the fixes to Error 1 discussed above work, although we mainly work with $\mathrm{Gr}(v, \tilde{q}^w)$. The reason is that $\tilde{q}^w$ is a direct limit of finite-dimensional varieties, and each quiver grassmannian is contained in one of these. Thus, with this viewpoint, the quiver grassmannians are naturally subvarieties of ordinary grassmannians, which we find helpful.

Received May 6, 2014.

ALISTAIR SAVAGE
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF OTTAWA
585 KING EDWARD AVE
OTTAWA ON K1N 6N5
CANADA
alistair.savage@uottawa.ca

PETER TINGLEY
DEPARTMENT OF MATHEMATICS AND STATISTICS
LOYOLA UNIVERSITY CHICAGO
CHICAGO, IL 60660
UNITED STATES
ptingley@luc.edu

CONTENTS

Volume 275, no. 1 and no. 2

Ola A. **Ashour**, Mourad E. H. Ismail and Zeinab S. Mansour: *On certain dual q-integral equations* 63

Florent **Balacheff**: *Measurements of Riemannian two-disks and two-spheres* 167

Nilton **Barroso** and Pedro Roitman: *Surfaces in* $\mathbb{R}^3_+$ *with the same Gaussian curvature induced by the Euclidean and hyperbolic metrics* 19

Marcin **Bownik**, Kurt Luoto and Edward Richmond: *A combinatorial characterization of tight fusion frames* 257

Tapas **Chatterjee** and M. Ram Murty: *On a conjecture of Erdős and certain Dirichlet series* 103

Kunal **Dutta** and Amritanshu Prasad: *Combinatorics of finite abelian groups and Weil representations* 295

Jose A. **Franco** with Mark R. Sepanski 463

Xianghong **Gong** and Jiří Lebl: *Normal forms for CR singular codimension-two Levi-flat submanifolds* 115

François **Guéritaud**, Fanny Kassel and Maxime Wolff: *Compact anti-de Sitter 3-manifolds and folded hyperbolic structures on surfaces* 325

Li **Guo** with Shanghua Zheng and Markus Rosenkranz 481

Qilong **Guo** with Ruifeng Qiu and Yanqing Zou 231

Mourad E. H. **Ismail** with Ola A. Ashour and Zeinab S. Mansour 63

Fanny **Kassel** with François Guéritaud and Maxime Wolff 325

Jiří **Lebl** with Xianghong Gong 115

Kurt **Luoto** with Marcin Bownik and Edward Richmond 257

Fabiola **Manjarrez-Gutiérrez**, Víctor Núñez and Enrique Ramírez-Losada: *Circular handle decompositions of free genus one knots* 361

Zeinab S. **Mansour** with Ola A. Ashour and Mourad E. H. Ismail 63

M. Ram **Murty** with Tapas Chatterjee 103

Víctor **Núñez** with Fabiola Manjarrez-Gutiérrez and Enrique Ramírez-Losada 361

Oscar M. **Perdomo**: *Constant-speed ramps* 1

Dan **Petersen**: *Cohomology of local systems on the moduli of principally polarized abelian surfaces* 39

Amritanshu **Prasad** with Kunal Dutta 295

Ruifeng **Qiu**, Yanqing Zou and Qilong Guo: *The Heegaard distances cover all nonnegative integers* 231

Enrique **Ramírez-Losada** with Fabiola Manjarrez-Gutiérrez and Víctor Núñez 361

R. Michael **Range**: *A pointwise a-priori estimate for the* $\bar{\partial}$*-Neumann problem on weakly pseudoconvex domains* 409

Edward **Richmond** with Marcin Bownik and Kurt Luoto 257

Marcus **Robinson** and Irena Swanson: *Explicit Hilbert–Kunz functions of* 2×2 *determinantal rings* 433

Pedro **Roitman** with Nilton Barroso 19

Markus **Rosenkranz** with Shanghua Zheng and Li Guo 481

Takao **Satoh**: *The Johnson–Morita theory for the ring of Fricke characters of free groups* 443

Alistair **Savage** and Peter Tingley: *Correction to the article Quiver grassmannians, quiver varieties and the preprojective algebra* 509

Mark R. **Sepanski** and Jose A. Franco: *Global representations of the conformal group and eigenspaces of the Yamabe operator on* $S^1 \times S^n$ 463

Claus M. **Sorensen**: *Eigenvarieties and invariant norms* 191

Irena **Swanson** with Marcus Robinson 433

Peter **Tingley** with Alistair Savage 509

Jianming **Wan**: *Harmonic maps from* $\mathbb{C}^n$ *to Kähler manifolds* 183

Maxime **Wolff** with François Guéritaud and Fanny Kassel 325

Shanghua **Zheng**, Li Guo and Markus Rosenkranz: *Rota–Baxter operators on the polynomial algebra, integration, and averaging operators* 481

Yanqing **Zou** with Ruifeng Qiu and Qilong Guo 231

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 275 No. 2 June 2015

A combinatorial characterization of tight fusion frames 257
MARCIN BOWNIK, KURT LUOTO and EDWARD RICHMOND

Combinatorics of finite abelian groups and Weil representations 295
KUNAL DUTTA and AMRITANSHU PRASAD

Compact anti-de Sitter 3-manifolds and folded hyperbolic structures on surfaces 325
FRANÇOIS GUÉRITAUD, FANNY KASSEL and MAXIME WOLFF

Circular handle decompositions of free genus one knots 361
FABIOLA MANJARREZ-GUTIÉRREZ, VÍCTOR NÚÑEZ and ENRIQUE RAMÍREZ-LOSADA

A pointwise a-priori estimate for the $\bar{\partial}$-Neumann problem on weakly pseudoconvex domains 409
R. MICHAEL RANGE

Explicit Hilbert–Kunz functions of 2×2 determinantal rings 433
MARCUS ROBINSON and IRENA SWANSON

The Johnson–Morita theory for the ring of Fricke characters of free groups 443
TAKAO SATOH

Global representations of the conformal group and eigenspaces of the Yamabe operator on $S^1 \times S^n$ 463
MARK R. SEPANSKI and JOSE A. FRANCO

Rota–Baxter operators on the polynomial algebra, integration, and averaging operators 481
SHANGHUA ZHENG, LI GUO and MARKUS ROSENKRANZ

Correction to the article Quiver grassmannians, quiver varieties and the preprojective algebra 509
ALISTAIR SAVAGE and PETER TINGLEY