Pacific Journal of Mathematics

Volume 247 No. 2 October 2010

PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

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

See inside back cover or www.pjmath.org for submission instructions.

The subscription price for 2010 is US $420/year for the electronic version, and $485/year for print and electronic.
Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. Prior back issues are obtainable from Periodicals Service Company, 11 Main Street, Germantown, NY 12526-5635. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 969 Evans Hall, Berkeley, CA 94720-3840, is published monthly except July and August. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

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

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS
at the University of California, Berkeley 94720-3840
A NON-PROFIT CORPORATION
Typeset in LaTeX

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

A FAMILY OF REPRESENTATIONS OF BRAID GROUPS ON SURFACES

BYUNG HEE AN AND KI HYOUNG KO

We propose a family of homological representations of the braid groups on surfaces. This family extends linear representations of the braid groups on a disc, such as the Burau representation and the Lawrence–Krammer–Bigelow representation.

1. Introduction

1A. ***Preliminaries and history.*** Let $\Sigma(g, p)$ be a compact, connected, orientable 2-dimensional manifold of genus g with p boundary components. Set $\Sigma = \Sigma(g, p)$. Let $\{z_1^0, \dots, z_n^0\}$ be a set of n preferred distinct points in Σ for $n \geq 0$, and let $\Sigma_n = \Sigma - \{z_1^0, \dots, z_n^0\}$. We call Σ_n the surface Σ with n punctures.

For integers $n, k \geq 0$, we consider three types of *configuration spaces* as follows: The space of k-tuples of distinct points in Σ_n is denoted by

$$P_{n,k}(\Sigma) = \{(z_1, \dots, z_k) \in \Sigma_n \times \cdots \times \Sigma_n \mid z_i \neq z_j \text{ for } i \neq j\},$$

the space of subsets of k elements in Σ_n is denoted by

$$B_{n,k}(\Sigma) = \{\{z_1, \dots, z_k\} \subset \Sigma_n\},$$

and the space $B_{n;k}(\Sigma)$ of pairs of disjoint subsets of n elements and k elements in Σ is denoted by

$$B_{n;k}(\Sigma) = \{(\{z_1, \dots, z_n\}, \{z_{n+1}, \dots, z_{n+k}\}) \mid z_i \in \Sigma,\ z_i \neq z_j \text{ for } i \neq j\}.$$

It is easy to see that $B_{n,k}(\Sigma) = P_{n,k}(\Sigma)/\mathbf{S}_k$ and $B_{n;k}(\Sigma) = P_{0,n+k}(\Sigma)/\mathbf{S}_n \times \mathbf{S}_k$, where the symmetric group $\mathbf{S}_k$ acts on $P_{n,k}(\Sigma)$ by permuting components of a k-tuple and similarly $\mathbf{S}_n \times \mathbf{S}_k \subset \mathbf{S}_{n+k}$ acts on $P_{0,n+k}(\Sigma)$.

The braid groups on a surface Σ are defined by the fundamental groups of configuration spaces. Choose a basepoint $\{z_{n+1}^0, \dots, z_{n+k}^0\}$ in $\partial\Sigma$ if $\partial\Sigma \neq \varnothing$. If Σ is

---

MSC2000: 20F36, 57M07, 57M10.

Keywords: homological representation, surface braid group.

This work was supported by the Korea Science and Engineering Foundation grant funded by the Korean government, MOST number R01-2006-000-10152-0.

closed, then place it anywhere in Σ_n. The *pure k-braid group on* Σ_n is defined and denoted by

$$\mathbf{P}_{n,k}(\Sigma) = \pi_1(P_{n,k}(\Sigma), (z^0_{n+1}, \dots, z^0_{n+k})).$$

Similarly, the (*full*) *k-braid group on* Σ_n is given by

$$\mathbf{B}_{n,k}(\Sigma) = \pi_1(B_{n,k}(\Sigma), \{z^0_{n+1}, \dots, z^0_{n+k}\}),$$

and the *intertwining* (n, k)*-braid group on* Σ is given by

$$\mathbf{B}_{n;k}(\Sigma) = \pi_1(B_{n;k}(\Sigma), (\{z^0_1, \dots, z^0_n\}, \{z^0_{n+1}, \dots, z^0_{n+k}\})).$$

It is sometimes easier to understand if these groups are regarded as subgroups of $\mathbf{B}_{0,n+k}(\Sigma)$. The intertwining (n, k)-braid group $\mathbf{B}_{n;k}(\Sigma)$ is the preimage of $\mathbf{S}_n \times \mathbf{S}_k$ under the canonical projection: $\mathbf{B}_{0,n+k}(\Sigma) \to \mathbf{S}_{n+k}$. In addition, $\mathbf{B}_{n,k}(\Sigma)$ is the subgroup of $(n+k)$-braids in $\mathbf{B}_{n;k}(\Sigma)$ that become trivial by forgetting the last k strands, and $\mathbf{P}_{n,k}(\Sigma)$ is the subgroup of $(n+k)$-braids in $\mathbf{B}_{n,k}(\Sigma)$ that are pure, that is, the induced permutation is trivial. If the surface Σ is the 2-disc D, we will call the braid groups *classical*. For example, $\mathbf{B}_{0,n}(D)$ denotes the classical n-braid group studied by E. Artin.

In the 60s and 70s, presentations for braid groups on various surfaces were found, on the 2-sphere and the projective plane in [Fadell and van Buskirk 1962; Van Buskirk 1966], on the torus in [Birman 1969], and on all closed surfaces in [Scott 1970]. The study of braid groups on surfaces has been revived recently. González-Meneses [2001] found new presentations of the braid groups on surfaces, and the authors of [Bellingeri 2004; Bellingeri and Godelle 2007] found positive presentations of the braid groups $\mathbf{B}_{n,k}(\Sigma)$ for all surfaces Σ, with or without boundary. Here, we are interested in braid groups on surfaces with nonempty boundary and will use Bellingeri's presentations.

Boundary components of a surface can be traded with punctures when we consider braid groups. Let $\Sigma = \Sigma(g, p)$ and $\Sigma' = \Sigma(g, p+q)$. Then there are continuous maps $i : \Sigma_q \to \Sigma'$ and $j : \Sigma' \to \Sigma_q$ that are homotopy inverses each other. The induced maps $\bar{i} : B_{n+q,k}(\Sigma) \to B_{n,k}(\Sigma')$ and $\bar{j} : B_{n,k}(\Sigma') \to B_{n+q,k}(\Sigma)$ on configuration spaces are also homotopy inverses each other and induce isomorphisms $\bar{i}_*$ and $\bar{j}_*$ on fundamental groups [Bellingeri 2004; Paris and Rolfsen 1999]. Therefore we may assume $\Sigma = \Sigma(g, 1)$ by treating all but one boundary component as a puncture whenever we deal with a surface with nonempty boundary.

We use Bellingeri's presentation [2004] for the braid group $\mathbf{B}_{n,k}(\Sigma(g, 1))$:

- The generators are $\sigma_1, \dots, \sigma_{k-1}, a_1, \dots, a_g, b_1, \dots, b_g, \zeta_1, \dots, \zeta_n$.
- The relations are

(BR$_1$) $[\sigma_i, \sigma_j]$ for $|i - j| \geq 2$;

(BR$_2$) $\sigma_i\sigma_j\sigma_i = \sigma_j\sigma_i\sigma_j$ for $|i - j| = 1$;

(CR$_1$) $[a_r, \sigma_i],\ [b_r, \sigma_i],\ [\zeta_t, \sigma_i]$ for $i > 1$;

(CR$_2$) $[a_r, \sigma_1 a_r \sigma_1],\ [b_r, \sigma_1 b_r \sigma_1],\ [\zeta_t, \sigma_1 \zeta_t \sigma_1]$;

(CR$_3$) $[a_r, \sigma_1^{-1} a_s \sigma_1],\ [a_r, \sigma_1^{-1} b_s \sigma_1],\ [b_r, \sigma_1^{-1} a_s \sigma_1],\ [b_r, \sigma_1^{-1} b_s \sigma_1]$ for $r < s$, $[a_r, \sigma_1^{-1} \zeta_u \sigma_1], [b_r, \sigma_1^{-1} \zeta_u \sigma_1], [\zeta_t, \sigma_1^{-1} \zeta_u \sigma_1]$ for $t < u$;

(SCR) $\sigma_1 b_r \sigma_1 a_r \sigma_1 = a_r \sigma_1 b_r$.

The corresponding result for configuration spaces of pure braids by Fadell and Neuwirth can be generalized to show that the projection $B_{n;k}(\Sigma) \to B_{0,n}(\Sigma)$ is a fiber bundle with fiber $B_{n,k}(\Sigma)$. Except for the cases $\Sigma = S^2$ and $\Sigma = \mathbb{R}P^2$, Gonçalves and Guaschi [2003] completely determined when the short exact sequences of braid groups derived from Fadell–Neuwirth fibrations split. In particular, the short exact sequence derived from the fibration above, that is,

$$1 \to \mathbf{B}_{n,k}(\Sigma) \to \mathbf{B}_{n;k}(\Sigma) \to \mathbf{B}_{0,n}(\Sigma) \to 1,$$

always splits for $k \geq 1$ if Σ has nonempty boundaries.

The braid groups are closely related to the mapping class groups. Birman [1974] determined when surface braid group embeds into the corresponding mapping class group. In particular, if $\partial\Sigma$ is nonempty, $\mathbf{B}_{0,n}(\Sigma)$ embeds into the mapping class group on Σ_n and so an n-braid on Σ can be regarded as a homeomorphism of Σ that preserves the set of n punctures.

The classical braid groups have various representations that can be as simple as taking exponent sums or taking induced permutations. The braid action on the punctured disk D_n gives rise to a faithful representation into automorphism groups of free groups, and a characterization of automorphisms coming from braid actions is possible. Each representation serves its own purpose. It is common to try to construct a linear representation to have a better understanding of a given group via matrices over a certain commutative ring and their multiplications.

For the classical braid groups, linear representations are abundant. Burau in 1936 and Gassner in 1961 found linear representations of $\mathbf{B}_{0,n}(D)$ and $\mathbf{P}_{0,n}(D)$, respectively. These representations are derived from braid actions on homologies of appropriate coverings of D_n. These representations take the form of $(n-1)\times(n-1)$ matrices that can also be computed via Fox's free differential calculus on automorphisms of free groups mentioned above. The Burau representation is faithful for $n \leq 3$ but not for $n \geq 5$ [Bigelow 1999]. The faithfulness of the Gassner representation is known only for $n \leq 3$.

Lawrence [1990] discovered a family of linear representations of $\mathbf{B}_{0,n}(D)$ via a monodromy on a vector bundle over $P_{n,k}(D)$. Krammer [2000] defined a free $\mathbb{Z}[q^{\pm 1}, t^{\pm 1}]$-module V using *forks* and relations between them, and he proved using an algebraic and combinatorial argument that the braid group acts on V faithfully for braid index 4. This representation is essentially the same as the one considered by Lawrence for $k=2$ but uses the configuration space $B_{n,2}(D)$ instead of $P_{n,2}(D)$ and is now called the Lawrence–Krammer–Bigelow representation. Bigelow reinterpreted this representation using covering spaces and covering transformation groups instead of vector bundles and local coefficients. Then the monodromy corresponds to the braid action on homology groups of covering spaces, as it did for the Burau representation and the Gassner representation. Bigelow [2001] constructed a linear representation using homology group $H_2(\tilde{B}_{n,2}(D))$ of the covering space $\tilde{B}_{n,2}(D)$ whose covering transformation group is $\langle q\rangle \oplus \langle t\rangle$, and he proved that $\mathbb{R}\otimes V$ is isomorphic to $\mathbb{R}\otimes H_2(\tilde{B}_{n,2}(D))$. Also, Krammer [2002] and Bigelow [2001] independently proved that the Lawrence–Krammer–Bigelow representation is faithful for all $n \geq 1$, and so the classical braid groups are linear. Bardakov [2005] applied this linearity to show that the braid groups of the sphere and projective plane are linear. Bigelow and Budney [2001] proved using the Lawrence–Krammer–Bigelow representation and a suitable branched covering that the mapping class group of genus 2 surface has a faithful linear representation. However, Paoluzzi and Paris showed that there is a difference between V and $H_2(\tilde{B}_{n,2})(D)$ as a $\mathbb{Z}[q^{\pm 1}, t^{\pm 1}]$-module for $n \geq 3$ and a found basis for a $\mathbb{Z}[q^{\pm 1}, t^{\pm 1}]$-module $H_2(\tilde{B}_{n,2}(D))$; so the exact definition of "Lawrence–Krammer–Bigelow representation" became somewhat ambiguous.

For any $k \geq 1$, Bigelow [2004] considered the braid action on the Borel–Moore homology group $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))$. He obtained a family of representations via the induced action on the image of $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))$ in $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D), \partial\tilde{B}_{n,k}(D))$. For simplicity, we will consider the braid action on the free module $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))$, whose basis can be easily described by forks to obtain a linear representation. We will call these representations *homology linear representations*. The Burau representation, and the Lawrence–Krammer–Bigelow representation of $\mathbf{B}_{0,n}(D)$ are homology linear representations

$$\Phi_k : \mathbf{B}_{0,n}(D) \to \mathrm{GL}\left(\binom{n+k-2}{k}, \mathbb{Z}[q^{\pm 1}, t^{\pm 1}]\right)$$

obtained from the braid action on homologies of covers of $B_{n,k}(D)$ when $k=1$ ($t=1$ in this case) and $k=2$, respectively [Bigelow 2004]. For $k \geq 3$, Zheng [2005] proved that Φ_k is faithful for all $n \geq 1$.

Overview. We construct a family of homological representations of braid groups on a surface with nonempty boundary; these extend the homological linear representations of the classical braid group. In Section 2, we first try to follow how the homology linear representations of $\mathbf{B}_{0,n}(D)$ were constructed via a covering of the configuration space $B_{n,k}(D)$. In the case of the disk, the braid action automatically commutes with covering transformations, or in other words, braids act trivially on local coefficients. However, in the case of surfaces of genus ≥ 1, this condition forces the variable q to equal 1; see Lemma 2.6. Then the braid action becomes almost trivial. For example, if $k=1$, the action of σ_i^2 is trivial. To get around this problem, we introduce in Section 3 the intertwining braid group $\mathbf{B}_{n;k}(\Sigma)$ to replace $\mathbf{B}_{n,k}(\Sigma)$. As we mentioned earlier, this group is a semidirect product of $\mathbf{B}_{n,k}(\Sigma)$ and $\mathbf{B}_{0,n}(\Sigma)$. Although the braid action does not preserve the local coefficient given by the $\mathbf{B}_{n,k}(\Sigma)$ factor, the $\mathbf{B}_{0,n}(\Sigma)$ factor of $\mathbf{B}_{n;k}(\Sigma)$ can adjust the coefficient so that the braid action becomes compatible. We will extend the coefficient ring for homology representations to give room to control the braid action, at the expense of giving up its commutativity, so that it becomes more interesting and still preserves coefficients. Eventually we obtain in Theorem 3.2 representations of braid groups on surfaces that extend homology linear representations of the classical braid group. Also we explicitly compute the representations in the form of matrices using a geometric argument. We extend the intersection pairing between $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))$ and its dual space $H_k(\tilde{B}_{n,k}(D), \partial\tilde{B}_{n,k}(D))$ and use bases for the two spaces that are described by *forks* and *noodles*; see Theorem 3.4.

In Section 4, we argue that the construction of our representations is natural and that there are no other alternatives if one wants to obtain an extension of the homological representation using covers of the configuration space $B_{n,k}(D)$. We show that the intertwining braid group $\mathbf{B}_{n;k}(\Sigma)$ is the normalizer of $\mathbf{B}_{n,k}(\Sigma)$ in $\mathbf{B}_{0,n+k}(\Sigma)$ so that the intertwining braid group $\mathbf{B}_{n;k}(\Sigma)$ is such a group that is unique and maximal up to a meaningless extension; see Theorem 4.2. The coefficient ring for our representations is the integral group ring of a quotient group of $\mathbf{B}_{n;k}(\Sigma)$. Theorem 4.3 shows that for $k \geq 3$ the quotient group is uniquely determined if one wants to extend homology linear representations of the classical braid group. For $k=1, 2$, the quotient group is the simplest that serves our purpose. Theorem 4.4 shows that the braid action on the quotient group is virtually unique.

Our construction involving the group extension $\mathbf{B}_{n;k}(\Sigma)$ of $\mathbf{B}_{n,k}(\Sigma)$ is purely algebraic, without a good geometric interpretation. Thus, some useful geometric tools are not available. For example, the intersection pairing mentioned above is not invariant under the braid group action. This seems to make it difficult to discuss properties of our representations such as faithfulness and irreducibility. Although the corresponding representation of the classical braid group is faithful for $k=2$ and irreducible for $k \leq 2$ [Jones 1987; Zinno 2001], the faithfulness and irreducibility of our representations are beyond the scope of this article.

2. Homology linear representations

We first review the construction of homology linear representations of the classical braid group $\mathbf{B}_{0,n}(D)$ using the configuration space $B_{n,k}(D)$; we then discuss the difficulty in extending these homology linear representations to the braid group $\mathbf{B}_{0,n}(\Sigma)$ on a surface Σ with nonempty boundary. As we noted earlier, boundary components can be traded with punctures. From now on, we assume that Σ denotes a compact, connected, oriented surface with exactly one boundary component and that n and k are positive integers.

Homology linear representations of classical braid group. Let $\phi : \mathbf{B}_{n,k}(D) \to G$ be an epimorphism onto a group G. Consider the covering $p : \tilde{B}_{n,k}(D) \to B_{n,k}(D)$ corresponding to $\operatorname{Ker}\phi$. Since the classical braid group embeds into the mapping class group of the punctured disk D_n, we may assume we have a homeomorphism $\bar{\beta} : B_{n,k}(D) \to B_{n,k}(D)$ for each $\beta \in \mathbf{B}_{0,n}$. By the lifting criteria, $\bar{\beta}$ lifts to $\tilde{\beta} : \tilde{B}_{n,k}(D) \to \tilde{B}_{n,k}(D)$ if and only if $\bar{\beta}_*(\operatorname{Ker}\phi) \subset \operatorname{Ker}\phi$. Equivalently, there is an induced automorphism $\beta_\sharp$ on G such that $\beta_\sharp\phi = \phi\bar{\beta}_*$.

Now we consider *Borel–Moore homology* [Borel and Moore 1960; Hughes and Ranicki 1996] defined by

$$H_\ell^{\mathrm{BM}}(\tilde{B}_{n,k}(D)) = \varprojlim H_\ell(\tilde{B}_{n,k}(D), p^{-1}(B_{n,k}(D) \setminus A)),$$

where the inverse limit is taken over all compact subsets A of $B_{n,k}(D)$.

The middle-dimensional homology group $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))$ is a free $\mathbb{Z}[G]$-module of rank $\binom{n+k-2}{k}$ (see [Bigelow 2004]) and $\tilde{\beta}$ induces a map $\tilde{\beta}_* : H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D)) \to H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))$ such that

$$\tilde{\beta}_*(yc) = \beta_\sharp(y)\tilde{\beta}_*(c) \quad \text{for } y \in G \text{ and } c \in H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D)).$$

Thus the map $\tilde{\beta}_*$ is a $\mathbb{Z}[G]$-module homomorphism if and only if $\beta_\sharp(y) = y$ for all $y \in G$ if and only if

$$(*) \qquad \phi = \phi\bar{\beta}_* \quad \text{for all } \beta \in \mathbf{B}_{0,n}.$$

Notice that the condition $(*)$ also implies $\bar{\beta}_*(\operatorname{Ker}\phi) \subset \operatorname{Ker}\phi$. Here we need to know that the induced homomorphism $\tilde{\beta}_*$ depends only on the isotopy class of the homeomorphism β. In fact, since D has a boundary, we choose the basepoint $\{z_{n+1}^0, \dots, z_{n+k}^0\}$ of $B_{n,k}(D)$ in ∂D, and then the isotopy preserves the basepoint and gives the same induced map $\tilde{\beta}_*$. Consequently, if we choose a group G and an epimorphism $\phi : \mathbf{B}_{n,k}(D) \to G$ satisfying $(*)$, we obtain a family of representations Φ_k from $\mathbf{B}_{0,n}(D)$ to $\operatorname{Aut}_{\mathbb{Z}[G]}(H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D)))$, the group of $\mathbb{Z}[G]$-module automorphisms on $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))$; the Φ_k are defined by

$$\Phi_k(\beta) = \tilde{\beta}_* : H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D)) \to H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D)).$$

Because we want to get a linear representation, G should be abelian. By the presentation given in Section 1A, $\mathbf{B}_{n,k}(D)$ is generated by $\zeta_1, \dots, \zeta_n, \sigma_1, \dots, \sigma_{k-1}$.

Suppose that $\phi : \mathbf{B}_{n,k}(D) \to G$ is an epimorphism such that $(*)$ holds and G is abelian. Each generator σ_i of $\mathbf{B}_{0,n}(D)$ acts trivially on $\mathbf{B}_{n,k}(D)$ except for

$$(\bar{\sigma}_i)_*(\zeta_i) = \zeta_i \zeta_{i+1} \zeta_i^{-1} \quad \text{and} \quad (\bar{\sigma}_i)_*(\zeta_{i+1}) = \zeta_i.$$

Then the condition $(*)$ implies that $\phi(\zeta_i) = \phi((\bar{\sigma}_i)_*(\zeta_{i+1})) = \phi(\zeta_{i+1})$. Hence for $k = 1$, G is a quotient of $\langle q \rangle$, and $\phi(\zeta_i) = q$ for $i = 1, \dots, n$. For $k \geq 2$, G is a quotient of $\langle q \rangle \oplus \langle t \rangle$, and $\phi(\zeta_i) = q$ and $\phi(\sigma_j) = t$ for $i = 1, \dots, n$ and $j = 1, \dots, k-1$.

We define a group G_D and an epimorphism $\phi_D : \mathbf{B}_{n,k}(D) \to G_D$ depending only on k as follows:

$$\phi_D : \mathbf{B}_{n,k}(D) \to G_D = \begin{cases} \langle q \rangle & \text{if } k = 1, \\ \langle q \rangle \oplus \langle t \rangle & \text{if } k \geq 2. \end{cases}$$

Theorem 2.1 [Bigelow 2004; Lawrence 1990]. *Let $\phi_D : \mathbf{B}_{n,k}(D) \to G_D$ be the epimorphism defined above. Then there is a homomorphism*

$$\Phi_k : \mathbf{B}_{0,n}(D) \to \operatorname*{Aut}_{\mathbb{Z}[G_D]} (H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))).$$

In fact, Φ_1 is the Burau representation and Φ_2 is the Lawrence–Krammer–Bigelow representation.

Naive extension to braid groups on surfaces. Let Σ be a surface of genus $g \geq 1$ having one boundary component. The assumption $\partial\Sigma \neq \varnothing$ is necessary for another reason besides the two mentioned at the end of Section 1A. Suppose that $\partial\Sigma = \varnothing$ and $\beta \in \mathbf{B}_{0,n}(\Sigma)$ uniquely determines the isotopy class of a homeomorphism $\bar{\beta} : B_{n,k}(\Sigma) \to B_{n,k}(\Sigma)$. Then we must choose the basepoint $\{z_{n+1}^0, \dots, z_{n+k}^0\}$ in the interior of Σ. We can easily find a homeomorphism $\bar{\beta} : B_{n,k}(\Sigma) \to B_{n,k}(\Sigma)$ that is isotopic to the identity via an isotopy that does not preserve the basepoint. Then β represents the identity element in $\mathbf{B}_{0,n}(\Sigma)$ but $\tilde{\beta}_* : H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma)) \to H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$ may be nontrivial. Thus no representation can be obtained in this way if $\partial\Sigma = \varnothing$.

We need to define what it means to say that a representation of the braid group $\mathbf{B}_{0,n}(\Sigma)$ extends homology linear representations of the classical braid groups.

Definition 2.2. Given a ring R, let M be an R-module on which the braid group $\mathbf{B}_{0,n}(\Sigma)$ acts as R-module isomorphisms. The R-module M is an *extension* of homology linear representations of the classical braid groups $\mathbf{B}_{0,n}(D)$ if there exists a $\mathbb{Z}[G_D]$-submodule M' of M such that

(i) M' is invariant under the action by the subgroup $\mathbf{B}_{0,n}(D)$ of $\mathbf{B}_{0,n}(\Sigma)$; and

(ii) for some $k \geq 1$, R contains $\mathbb{Z}[G_D]$ as a subring and there is a $\mathbb{Z}[G_D]$-isomorphism from $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))$ to M' that commutes with the $\mathbf{B}_{0,n}(D)$ action.

As in the classical braid cases, we have to look at the action of $\mathbf{B}_{0,n}(\Sigma)$ on $\mathbf{B}_{n,k}(\Sigma)$. The following lemma helps us to observe the action we want.

Lemma 2.3 [Birman 1974; Fadell and Neuwirth 1962; Gonçalves and Guaschi 2003]. *Let $\pi_n : B_{n;k}(\Sigma) \to B_{0,n}(\Sigma)$ be the projection onto the first n coordinates. Then the space $B_{n;k}(\Sigma)$ is a fiber bundle with fiber $B_{n,k}(\Sigma)$, and the induced short exact sequence*

$$1 \to \mathbf{B}_{n,k}(\Sigma) \longrightarrow \mathbf{B}_{n;k}(\Sigma) \xrightarrow{(\pi_n)_*} \mathbf{B}_{0,n}(\Sigma) \to 1$$

splits for all $k \geq 1$.

This lemma shows us how to decompose a braid $\beta \in \mathbf{B}_{n;k}(\Sigma)$ into a product $\beta = \beta_1\beta_2$ for $\beta_1 \in \mathbf{B}_{0,n}(\Sigma)$ and $\beta_2 \in \mathbf{B}_{n,k}(\Sigma)$. Let $\imath : \mathbf{B}_{0,n}(\Sigma) \to \mathbf{B}_{n;k}(\Sigma)$ be the splitting map. Then the lemma shows that $\mathbf{B}_{n;k}(\Sigma)$ can be generated by the sets

$$X_1 = \{\bar\sigma_1, \dots, \bar\sigma_{n-1}, \bar\mu_1, \dots, \bar\mu_g, \bar\lambda_1, \dots, \bar\lambda_g\},$$
$$X_2 = \{\sigma_1, \dots, \sigma_{k-1}, \zeta_1, \dots, \zeta_n, \mu_1, \dots, \mu_g, \lambda_1, \dots, \lambda_g\},$$

where the generators in X_1 are the images of generators in $\mathbf{B}_{0,n}(\Sigma)$ under the inclusion map $\imath$.

Then the action of $\mathbf{B}_{0,n}(\Sigma)$ on $\mathbf{B}_{n,k}(\Sigma)$ is equivalent to the conjugate action in $\mathbf{B}_{n;k}(\Sigma)$ if we regard these two groups as subgroups of $\mathbf{B}_{n;k}(\Sigma)$. The following easy lemma shows how $\mathbf{B}_{0,n}(\Sigma)$ acts on $\mathbf{B}_{n,k}(\Sigma)$.

Lemma 2.4. *Each generator of $\mathbf{B}_{0,n}(\Sigma)$ acts on $\mathbf{B}_{n,k}(\Sigma)$ as follows.*

(1) *For $1 \le i \le n-1$,*

$$(\bar\sigma_i)_*(\zeta_t) = \begin{cases} \zeta_i\zeta_{i+1}\zeta_i^{-1} & \text{if } t = i, \\ \zeta_i & \text{if } t = i+1. \end{cases}$$

(2) *For $1 \le r \le g$,*

$$(\bar\mu_r)_*(\zeta_1) = \mu_r\zeta_1\mu_r^{-1},$$
$$(\bar\mu_r)_*(\mu_s) = \begin{cases} \mu_r\zeta_1\mu_r\zeta_1^{-1}\mu_r^{-1} & \text{if } s = r, \\ [\mu_r,\zeta_1]\mu_s[\mu_r,\zeta_1]^{-1} & \text{if } r < s, \end{cases}$$
$$(\bar\mu_r)_*(\lambda_s) = \begin{cases} \lambda_r\mu_r\zeta_1^{-1}\mu_r^{-1} & \text{if } s = r, \\ [\mu_r,\zeta_1]\lambda_s[\mu_r,\zeta_1]^{-1} & \text{if } r < s. \end{cases}$$

(3) *For $1 \le r \le g$,*

$$(\bar\lambda_r)_*(\zeta_1) = \lambda_r\zeta_1\lambda_r^{-1},$$
$$(\bar\lambda_r)_*(\mu_s) = \begin{cases} \lambda_r\zeta_1\lambda_r^{-1}\mu_r\zeta_1\lambda_r\zeta_1^{-1}\lambda_r^{-1} & \text{if } s = r, \\ [\lambda_r,\zeta_1]\,\mu_s\,[\lambda_r,\zeta_1]^{-1} & \text{if } r < s, \end{cases}$$
$$(\bar\lambda_r)_*(\lambda_s) = \begin{cases} \lambda_r\zeta_1\lambda_r\zeta_1^{-1}\lambda_r^{-1} & \text{if } s = r, \\ [\lambda_r,\zeta_1]\,\lambda_s\,[\lambda_r,\zeta_1]^{-1} & \text{if } r < s. \end{cases}$$

(4) *All other generators act trivially.*

We can find the presentation for $\mathbf{B}_{n;k}(\Sigma)$ using this lemma as follows.

Lemma 2.5. *The braid group* $\mathbf{B}_{n;k}(\Sigma)$ *admits the presentation in which*

- *the generators are*

$$X_1 = \{\bar{\sigma}_1, \dots, \bar{\sigma}_{n-1}, \bar{\mu}_1, \dots, \bar{\mu}_g, \bar{\lambda}_1, \dots, \bar{\lambda}_g\},$$
$$X_2 = \{\sigma_1, \dots, \sigma_{k-1}, \zeta_1, \dots, \zeta_n, \mu_1, \dots, \mu_g, \lambda_1, \dots, \lambda_g\};$$

- *the relations are*
  (i) (BR_1) *through* (SCR) *among generators in* X_1,
  (ii) (BR_1) *through* (SCR) *among generators in* X_2, *and*
  (iii) $\bar{x}^{-1} y \bar{x} = (\bar{x}_*)(y)$ *for all* $\bar{x} \in X_1$ *and* $y \in X_2$,

where the action by $\bar{x}_*$ *is given in Lemma 2.4.*

Proof. By Lemma 2.3, the intertwining braid group $\mathbf{B}_{n;k}(\Sigma)$ is a semidirect product of the normal subgroup $\mathbf{B}_{n,k}(\Sigma)$ and $\mathbf{B}_{0,n}(\Sigma)$, where $\mathbf{B}_{0,n}(\Sigma)$ acts on $\mathbf{B}_{n,k}(\Sigma)$ by conjugation as shown in Lemma 2.4. Then it is easy to show that the semidirect product $\mathbf{B}_{n;k}(\Sigma)$ admits the desired presentation. □

For surfaces, the condition $(*)$ implies an undesirable consequence:

Lemma 2.6. *Let* $\phi : \mathbf{B}_{n,k}(\Sigma) \to G$ *be an epimorphism satisfying* $\phi = \phi\bar{\beta}_*$ *for any* $\beta \in \mathbf{B}_{n,k}(\Sigma)$. *Then* $\phi(\zeta_i) = 1$ *for* $i = 1, \dots, n$.

Proof. As seen earlier, the hypothesis on ϕ implies that $(\mu_r)_\sharp(y) = y$ for all $y \in G$ and $r = 1, \dots, g$. But by Lemma 2.4(2), we have

$$\begin{aligned}(\mu_r)_\sharp\phi(\lambda_r) &= \phi((\bar{\mu}_r)_*(\lambda_r)) = \phi(\lambda_r\mu_r\zeta_1^{-1}\mu_r^{-1})\\ &= \phi(\lambda_r)\phi((\bar{\mu}_r)_*(\zeta_1^{-1})) = \phi(\lambda_r)(\mu_r)_\sharp(\phi(\zeta_1^{-1})) = \phi(\lambda_r)\phi(\zeta_1^{-1}).\end{aligned}$$

Since $(\mu_r)_\sharp(\phi(\lambda_r)) = \phi(\lambda_r)$ by hypothesis, $\phi(\zeta_1) = 1$ and so $\phi(\zeta_i) = 1$ for all $1 \le i \le n$ by Lemma 2.4(1). □

This lemma says that the condition $(*)$ forces us to set $q = 1$ in the group G_D. Thus $\mathbb{Z}[G_D]$ cannot be a subring of $\mathbb{Z}[G]$, and so a naive attempt to obtain a representation of the braid group $\mathbf{B}_{0,n}(\Sigma)$ using a covering of $B_{n,k}(\Sigma)$ corresponding to any epimorphism $\phi : \mathbf{B}_{n,k}(\Sigma) \to G$ cannot give an extension of any homology linear representation of the classical braid groups.

3. A family of proposed representations

As we have seen in the previous section, we are forced to take a rather small covering of $B_{n,k}(\Sigma)$ in order that the condition $(*)$ be satisfied, that is, the braid action commutes with covering transformations so that it preserves the coefficient. The remedy we propose in this article is to use the same configuration space $B_{n,k}(\Sigma)$ with an extended coefficient ring so that we have some room to adjust coefficients to make the braid action compatible with the coefficients. This remedy is a reasonable thing to do if we hope to construct an extension of homology linear representations of the classical braid groups. Indeed, we successfully obtain an extension that seems the most general among ones obtained from coverings of $B_{n,k}(\Sigma)$.

3A. ***Existence of extensions of homology linear representations.*** We first consider the intertwining braid group $\mathbf{B}_{n;k}(\Sigma)$. Note that $\mathbf{B}_{n;k}(\Sigma)$ is a candidate for group extension of $\mathbf{B}_{n,k}(\Sigma)$ by Lemma 2.3, and $\mathbf{B}_{0,n}(\Sigma)$ acts on $\mathbf{B}_{n;k}(\Sigma)$ by right multiplication and acts on $\mathbf{B}_{n,k}(\Sigma)$ by conjugation because $\mathbf{B}_{n;k}(\Sigma)$ is the semidirect product of $\mathbf{B}_{0,n}(\Sigma)$ and $\mathbf{B}_{n,k}(\Sigma)$.

Let H_Σ be the abstract group depending only on k that admits for $k \geq 2$ the presentation in which

- the generators are $q, t, \overline{m}_1, \dots, \overline{m}_g, \bar{\ell}_1, \dots, \bar{\ell}_g, m_1, \dots, m_g, \ell_1, \dots, \ell_g$;
- the relations are such that all generators commute except that

$$[m_r, \ell_r] = t^2 \quad \text{and} \quad [\overline{m}_r, \ell_r] = [m_r, \bar{\ell}_r] = q.$$

Define ψ_Σ to be the epimorphism from $\mathbf{B}_{n;k}(\Sigma)$ to the group H_Σ such that

$$\begin{aligned} &\psi_\Sigma(\sigma_i) = t, \qquad \psi_\Sigma(\zeta_j) = q, \qquad \psi_\Sigma(\bar{\sigma}_m) = 1, \\ &\psi_\Sigma(\mu_r) = m_r, \quad \psi_\Sigma(\lambda_r) = \ell_r, \quad \psi_\Sigma(\bar{\mu}_r) = \overline{m}_r, \quad \psi_\Sigma(\bar{\lambda}_r) = \bar{\ell}_r, \end{aligned}$$

where $1 \leq i \leq k-1$, $1 \leq j \leq n$, $1 \leq m \leq n-1$ and $1 \leq r \leq g$. If $k = 1$, then we redefine H_Σ to be the quotient of the group above by $t = 1$. Then H_D is isomorphic to G_D defined earlier for all $k \geq 1$, and is a subgroup of H_Σ for any Σ and $k \geq 1$. Even though H_Σ (or H_D) depends on whether $k = 1$ or $k \geq 2$, our notation does not show it for the sake of simplicity.

Let $\phi_\Sigma : \mathbf{B}_{n,k}(\Sigma) \to G_\Sigma$ be the restriction of ψ_Σ to $\mathbf{B}_{n,k}(\Sigma)$ onto G_Σ, the subgroup $\psi_\Sigma(\mathbf{B}_{n,k}(\Sigma))$ of H_Σ. Then G_Σ is generated by

$$\{q, t, m_1, \dots, m_g, \ell_1, \dots, \ell_g\}.$$

Since any two elements of G_Σ commute up to multiplications by central elements q and t, it is a normal subgroup of H_Σ. We can find the covering $p : \tilde{B}_{n,k}(\Sigma) \to B_{n,k}(\Sigma)$ corresponding to $\operatorname{Ker} \phi_\Sigma$. Since the braid group $\mathbf{B}_{0,n}(\Sigma)$ embeds into the mapping class group of punctured surface Σ_n, a braid $\beta \in \mathbf{B}_{0,n}(\Sigma)$ determines a homeomorphism $\bar{\beta}: B_{n,k}(\Sigma) \to B_{n,k}(\Sigma)$. Recall that the induced homomorphism $\bar{\beta}_*$ on $\mathbf{B}_{n,k}(\Sigma)$ is in fact the same as the conjugation by $\imath(\beta)$ where $\imath: \mathbf{B}_{0,n}(\Sigma) \to \mathbf{B}_{n;k}(\Sigma)$ is the splitting map in Lemma 2.3.

Lemma 3.1. *With the notation above, the homeomorphism $\bar{\beta}: B_{n,k}(\Sigma) \to B_{n,k}(\Sigma)$ lifts to a homeomorphism $\tilde{\beta}: \tilde{B}_{n,k}(\Sigma) \to \tilde{B}_{n,k}(\Sigma)$ for any $\beta \in \mathbf{B}_{0,n}(\Sigma)$, and the restriction ϕ_Σ of ψ_Σ satisfies $\beta_\sharp \phi_\Sigma = \phi_\Sigma \bar{\beta}_*$*

Proof. By the lifting criteria, $\bar{\beta}$ lifts to $\tilde{\beta}$ if and only if $\bar{\beta}_*(\mathrm{Ker}\, \phi_\Sigma) \subset \mathrm{Ker}\, \phi_\Sigma$ if and only if there is an induced automorphism $\beta_\sharp$ on G_Σ given by $\beta_\sharp \phi_\Sigma = \phi_\Sigma \bar{\beta}_*$. Thus it suffices to show that $\phi_\Sigma \bar{\beta}_*(W) = 1$ for any $W \in \mathrm{Ker}\, \phi_\Sigma$ and $\beta \in \mathbf{B}_{0,n}(\Sigma)$. Let W be a word in the generators $\{\mu_i, \lambda_i, \sigma_i, \zeta_i\}$ of $\mathbf{B}_{n,k}(\Sigma)$. Since the presentation for H_Σ shows that any two elements are commutative up to multiplications by central elements q and t, we have

$$\phi_\Sigma(W) = W(\mu_i \leftarrow m_i, \lambda_i \leftarrow \ell_i, \sigma_i \leftarrow t, \zeta_i \leftarrow q) = q^c t^d \prod m_i^{a_i} \ell_i^{b_i},$$

where $W(\{x_i \leftarrow y_i\})$ denotes the word obtained from W by replacing the generators x_i by the generators y_i.

Suppose $\phi_\Sigma(W) = 1$. Then $a_i = b_i = 0$ for all $1 \le i \le g$. Thus for generators $\sigma_r, \mu_r, \lambda_r$ for $\mathbf{B}_{0,n}(\Sigma)$, we have

$$\begin{aligned}
\phi_\Sigma((\bar{\sigma}_r)_*(W)) &= \phi_\Sigma(W(\zeta_r \leftarrow \zeta_r\zeta_{r+1}\zeta_r^{-1}, \zeta_{r+1} \leftarrow \zeta_r)) \\
&= W(\mu_i \leftarrow m_i, \lambda_i \leftarrow \ell_i, \sigma_i \leftarrow t, \zeta_i \leftarrow q) = 1, \\
\phi_\Sigma((\bar{\mu}_r)_*(W)) &= \phi_\Sigma(W(\lambda_r \leftarrow \lambda_r\mu_r\zeta_1^{-1}\mu_r^{-1})) \\
&= W(\mu_i \leftarrow m_i, \lambda_i \leftarrow \ell_i, \lambda_r \leftarrow q^{-1}\ell_r, \sigma_i \leftarrow t, \zeta_i \leftarrow q) \\
&= q^{-b_r} W(\mu_i \leftarrow m_i, \lambda_i \leftarrow \ell_i, \sigma_i \leftarrow t, \zeta_i \leftarrow q) = 1, \\
\phi_\Sigma((\bar{\lambda}_r)_*(W)) &= \phi_\Sigma(W(\mu_r \leftarrow \lambda_r\zeta_1\lambda_r^{-1}\mu_r\zeta_1\lambda_r\zeta_1^{-1}\lambda_r^{-1})) \\
&= W(\mu_i \leftarrow m_i, \mu_r \leftarrow qm_r, \lambda_i \leftarrow \ell_i, \sigma_i \leftarrow t, \zeta_i \leftarrow q) \\
&= q^{a_r} W(\mu_i \leftarrow m_i, \lambda_i \leftarrow \ell_i, \sigma_i \leftarrow t, \zeta_i \leftarrow q) = 1.
\end{aligned}$$

Therefore $\phi_\Sigma(\bar{\beta}_*(W)) = 1$ and so $\beta_\sharp(\phi_\Sigma(\alpha)) = \phi_\Sigma(\bar{\beta}_*(\alpha))$ for all $\alpha \in \mathbf{B}_{n,k}(\Sigma)$. □

By the lemma above, we have a $\mathbb{Z}$-module automorphism $\tilde{\beta}_*$ on $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$. Note that $\tilde{\beta}_*$ is not necessarily a $\mathbb{Z}[G_\Sigma]$-module homomorphism since the condition $(*)$ may not hold, that is, the automorphism $\beta_\sharp$ of G_Σ may not be the identity.

On the other hand, $\mathbf{B}_{0,n}(\Sigma)$ acts on $\mathbf{B}_{n;k}(\Sigma)$ by right multiplication and so there is an induced action of β on H_Σ given by $\beta \cdot h = h\psi_\Sigma(\beta)$ for $h \in H_\Sigma$. It is possible to alter the induced action by multiplying with a certain function χ from $\mathbf{B}_{0,n}(\Sigma)$ to the centralizer of G_Σ in H_Σ. We will discuss this possibility in Theorem 4.4.

Using the $\mathbb{Z}$-module automorphism $\tilde{\beta}_*$ and the action on $\mathbf{B}_{n;k}(\Sigma)$ by $\mathbf{B}_{0,n}(\Sigma)$, we construct a $\mathbb{Z}[H_\Sigma]$-module automorphism $\beta\otimes\tilde{\beta}_*$ on $\mathbb{Z}[H_\Sigma]\otimes_{\mathbb{Z}[G_\Sigma]} H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$ by

$$(\beta\otimes\tilde{\beta}_*)(h\otimes c)=(\beta\cdot h)\otimes\tilde{\beta}_*(c)\quad \text{for } h\in H_\Sigma \text{ and } c\in H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma)).$$

Theorem 3.2. *Let Σ be a compact, connected, oriented 2-dimensional manifold with nonempty boundary. Define the group H_Σ (depending on k) and the epimorphism $\psi_\Sigma:\mathbf{B}_{n;k}(\Sigma)\to H_\Sigma$ as above. Let ϕ_Σ be the restriction of ψ_Σ to $\mathbf{B}_{n,k}(\Sigma)$. Set $G_\Sigma=\phi_\Sigma(\mathbf{B}_{n,k}(\Sigma))$. Then there is a homomorphism*

$$\Phi_k:\mathbf{B}_{0,n}(\Sigma)\to \operatorname*{Aut}_{\mathbb{Z}[H_\Sigma]}(\mathbb{Z}[H_\Sigma]\otimes_{\mathbb{Z}[G_\Sigma]} H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))),\quad \beta\mapsto\beta\otimes\tilde{\beta}_*,$$

where the action of β on H_Σ is given by $\beta\cdot h=h\psi(\beta)$ for $h\in H_\Sigma$.

This family Φ_k of representations is an extension of a homology linear representation of the classical braid group $\mathbf{B}_{0,n}(D)$ in the sense of Definition 2.2.

Proof. Clearly Φ_k is a group homomorphism. To see the well-definedness and the $\mathbb{Z}[H_\Sigma]$-linearity of $\Phi_k(\beta)$, we claim that

$$\beta\cdot(hh')=(\beta\cdot h)\beta_\sharp(h')\quad \text{for all } h\in H_\Sigma,\ h'\in G_\Sigma.$$

Then, for $c\in H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$,

$$\begin{aligned}(\beta\otimes\tilde{\beta}_*)(h\otimes h'c)&=(\beta\cdot h)\otimes\tilde{\beta}_*(h'c)=(\beta\cdot h)\otimes\beta_\sharp(h')\tilde{\beta}_*(c)\\&=(\beta\cdot h)\beta_\sharp(h')\otimes\tilde{\beta}_*(c)=(\beta\otimes\tilde{\beta}_*)(hh'\otimes c)\\&=hh'(\beta\otimes\tilde{\beta}_*)(1\otimes c).\end{aligned}$$

Here the last equality is clear by the definition of the action by β.

To show the claim, choose $\alpha\in\mathbf{B}_{n,k}(\Sigma)$ so that $\phi_\Sigma(\alpha)=h'$. By Lemma 3.1, we have

$$\beta_\sharp(\phi_\Sigma(\alpha))=\phi_\Sigma(\bar{\beta}_*(\alpha))=\psi_\Sigma(\bar{\beta}_*(\alpha))=\psi_\Sigma(\beta^{-1}\alpha\beta)=\psi_\Sigma(\beta)^{-1}\phi_\Sigma(\alpha)\psi_\Sigma(\beta).$$

Thus

$$\beta\cdot(hh')=hh'\psi_\Sigma(\beta)=(h\psi_\Sigma(\beta))(\psi_\Sigma(\beta)^{-1}h'\psi_\Sigma(\beta))=(\beta\cdot h)\beta_\sharp(h').$$

To show that Φ_k is an extension of a homology linear representation of the classical braid groups, we regard an n punctured disk D_n as a subspace of Σ_n. Then the configuration space $B_{n,k}(D)$ is a subspace of the configuration space $B_{n,k}(\Sigma)$. For the covering $p:\tilde{B}_{n,k}(\Sigma)\to B_{n,k}(\Sigma)$ corresponding to ϕ_Σ, we denote a connected component of $p^{-1}(B_{n,k})$ by $\tilde{B}_{n,k}(D)$. Since G_D embeds into G_Σ, the restriction $p|_{\tilde{B}_{n,k}(D)}:\tilde{B}_{n,k}(D)\to B_{n,k}(D)$ is the covering over $B_{n,k}(D)$ corresponding to $\psi|_{\mathbf{B}_{n,k}(D)}:\mathbf{B}_{n,k}(D)\to G_D$. In fact, $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))$ is a submodule of

$H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$ as $\mathbb{Z}[G_D]$-module. One can see this more explicitly in the proof of Lemma 3.3. Each braid $\beta \in \mathbf{B}_{0,n}(D)$ gives a $\mathbb{Z}[H_D]$-module automorphism $\beta \otimes \tilde{\beta}_*$ on $\mathbb{Z}[H_D] \otimes_{\mathbb{Z}[G_D]} H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))$. Since $G_D = H_D$, this automorphism is the same as $\tilde{\beta}_*$ on $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))$, which is the homology linear representation of the classical braid group. □

In Section 4, we will show that if one wants to obtain a result similar to the theorem above, the extension $\mathbf{B}_{n;k}(\Sigma)$ of $\mathbf{B}_{n,k}(\Sigma)$ is determined uniquely up to redundant coefficient extension, and the quotient H_Σ is uniquely determined for $k \geq 3$ and is the simplest for $k \geq 1$ in the sense that any proper quotient of H_Σ does not contain G_D properly.

Computation of the proposed representations. We now compute explicit matrix forms of the representations described in Theorem 3.2; these turn out to be extensions of the Burau and Lawrence–Krammer–Bigelow representations of the classical braid groups. The following lemma and its proof show not only that $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$ is a free $\mathbb{Z}[G_\Sigma]$-module but also how to choose a basis. The lemma is an extension of the corresponding lemma on a disk by Bigelow [2004], and we borrow the main idea of his proof.

Lemma 3.3. *The homology group $H_\ell^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$ is the direct sum of*

$$\binom{2g+n+k-2}{k}$$

copies of $\mathbb{Z}[G_\Sigma]$ for $\ell = k$ and is trivial otherwise.

Proof. Let d be a metric on Σ that can be either hyperbolic or Euclidean. Suppose punctures $z_1^0, \dots, z_n^0$ lie on a geodesic. Let γ_j be the geodesic segment joining z_j^0 and z_{j+1}^0 for $1 \leq j \leq n-1$. For $1 \leq i \leq g$, let α_i and β_i be geodesic loops based at z_1^0 that represent the meridian and the longitude of the i-th handle, so that the α_i, β_i, and γ_j are mutually disjoint. Let Γ be the union of all of these arcs, so that Γ_n is a disjoint union of open $2g+n-1$ geodesic segments. Consider

$$B_\Gamma = B_{n,k}(\Gamma) = \{\{z_1, \dots, z_k\} \subset \Gamma_n\}.$$

Then it is not hard to see B_Γ is homeomorphic to a disjoint union of $\binom{2g+n+k-2}{k}$ open k-balls that can be parametrized by $(2g+n-1)$-tuples $(r_1, \dots, r_{2g+n-1})$ of nonnegative integers that add up to k so that the i-th segment of Γ_n contains r_i points from $\{z_1, \dots, z_k\}$.

Let $p : \tilde{B}_{n,k}(\Sigma) \to B_{n,k}(\Sigma)$ be the covering corresponding to the epimorphism $\phi_\Sigma : B_{n,k}(\Sigma) \to G_\Sigma$. We will be done if we can show that the map

$$H_\ell^{\mathrm{BM}}(p^{-1}(B_\Gamma)) \to H_\ell^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$$

induced by the inclusion is an isomorphism, since $H_\ell^{\mathrm{BM}}(B_\Gamma)$ is isomorphic to the direct sum of $\binom{2g+n+k-2}{k}$ copies of $H_\ell(D^k, S^{k-1})$.

Define a family of compact subsets A_ϵ of Σ_n by

$$A_\epsilon = \{\{z_1, \dots, z_k\} \in B_{n,k}(\Sigma) \mid d(z_i, z_j) \geq \epsilon \text{ for } i \neq j,\ d(z_i, z_j^0) \geq \epsilon \text{ for all } i, j\}.$$

Since any compact subset of $B_{n,k}(\Sigma)$ is contained in A_ϵ for sufficiently small $\epsilon > 0$, it suffices to show that

$$H_\ell(p^{-1}(B_\Gamma), p^{-1}(B_\Gamma - A_\epsilon)) \to H_\ell(\tilde{B}_{n,k}(\Sigma), p^{-1}(B_{n,k}(\Sigma) - A_\epsilon))$$

is an isomorphism.

Let $\Sigma_\epsilon \subset \Sigma$ be the closed ϵ-neighborhood of Γ, and let $B_\epsilon = B_{n,k}(\Sigma_\epsilon)$. Then the obvious homotopy collapsing from $B_{n,k}(\Sigma)$ to B_ϵ gives the isomorphism

$$H_\ell(p^{-1}(B_\epsilon), p^{-1}(B_\epsilon - A_\epsilon)) \to H_\ell(\tilde{B}_{n,k}(\Sigma), p^{-1}(B_{n,k}(\Sigma) - A_\epsilon)).$$

Let B be the set of $\{x_1, \dots, x_k\} \in B_\epsilon$ such that for each x_i there exists a unique nearest point in Γ_n. Then B is open and contains $A_\epsilon \cap B_\epsilon$. By excision, the inclusion induces an isomorphism

$$H_\ell(p^{-1}(B), p^{-1}(B - A_\epsilon)) \to H_\ell(p^{-1}(B_\epsilon), p^{-1}(B_\epsilon - A_\epsilon)).$$

Finally, the obvious deformation retract from B to B_Γ gives an isomorphism

$$H_\ell(p^{-1}(B), p^{-1}(B - A_\epsilon)) \to H_\ell(p^{-1}(B_\Gamma), p^{-1}(B_\Gamma - A_\epsilon)). \qquad \square$$

We remark that $B_\epsilon = B_{n,k}(\Sigma_\epsilon)$ and $B_\Gamma = B_{n,k}(\Gamma)$ do not have the same homotopy type even though Γ is a deformation retract of Σ_ϵ. This is because Γ is a 1-dimensional complex and movements of points in Γ avoiding collision are more restricted.

Let $I(n, k, g)$ be the set of $(2g + n - 1)$-tuples of nonnegative integers that add up to k. The proof of Lemma 3.3 shows that a typical basis for $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$ as free $\mathbb{Z}[G_\Sigma]$-module can be indexed by the set $I(n, k, g)$. The proof also shows that a basis for $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(D))$ can be chosen as a subset of a basis for $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$. Thus the homology linear representations for $\mathbf{B}_{0,n}(D)$ appears in matrix forms of proposed representations for $\mathbf{B}_{0,n}(\Sigma)$ as minors.

Krammer [2000; 2002] and Bigelow [2001; 2003; 2004] have shown that there is a natural and useful way of describing a basis geometrically. Recall the loops α_i and β_i and the arcs γ_j from the proof of Lemma 3.3. For $(r_1, \dots, r_{2g+n-1}) \in I(n, k, g)$, choose r_i disjoint duplicates of α_i or β_{i-g} or γ_{i-2g} if $1 \leq i \leq g$ or $g + 1 \leq i \leq 2g$ or $2g + 1 \leq i \leq 2g + n - 1$, respectively. For each i, join these r_i disjoint duplicates to $\partial\Sigma$ by mutually disjoint arcs (that determine a basing). This geometric object is called a *fork*. In fact, a fork uniquely determines a k-cycle in $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$ by lifting the Cartesian product of k curves together with basing

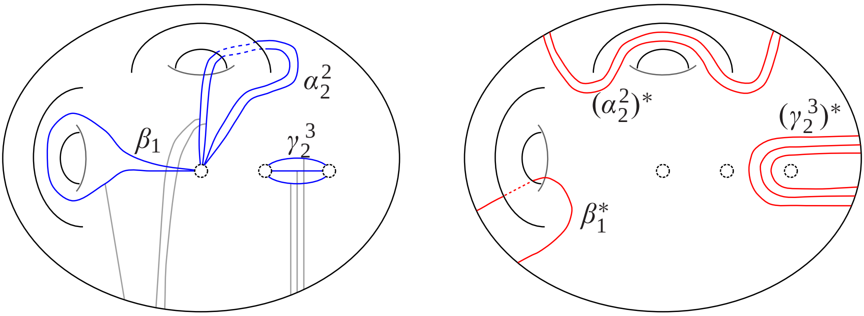


Figure 1. An example of a fork F (left) and its dual noodle N (right).

arcs in the fork. The basing is required to have a unique lift. For example, the fork corresponding to $(0, 2, 1, 0, 0, 3) \in I(3, 6, 2)$ looks like the set of curves on the left of Figure 1.

As Bigelow [2004] showed for the case of the disk, the Poincaré duality, the universal coefficient theorem, and Lemma 3.3 imply that the ordinary relative homology $H_k(\tilde{B}_{n,k}(\Sigma), \partial\tilde{B}_{n,k}(\Sigma))$ is the dual space of the Borel–Moore homology $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$ in the sense that there is a nonsingular sesquilinear pairing

$$\langle\,\cdot\,,\cdot\,\rangle : H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma)) \times H_k(\tilde{B}_{n,k}(\Sigma), \partial\tilde{B}_{n,k}(\Sigma)) \to \mathbb{Z}[S],$$

where S is a skew field containing $\mathbb{Z}[G_\Sigma]$. In fact, the group G_Σ is biordered, and so it can embed into a skew field such as the Mal'cev–Neumann power series ring [Mal'cev 1948; Neumann 1949]. Explicitly, the pairing above is defined by setting, for cycles $F \in H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$ and $N \in H_k(\tilde{B}_{n,k}(\Sigma), \partial\tilde{B}_{n,k}(\Sigma))$ in each homology group,

$$\langle F, N\rangle = \sum_{y \in G_\Sigma} y(F, yN),$$

where $(\,\cdot\,,\cdot\,)$ counts the intersection number.

Let $\alpha_1^*, \dots, \alpha_g^*, \beta_1^*, \dots, \beta_g^*, \gamma_1^*, \dots, \gamma_g^*$ be pairwise disjoint arcs that start and end at $\partial\Sigma$, and suppose α_i^* (or β_i^*, or γ_j^*) intersects only α_i (or β_i, or γ_j) once transversely. We form a basis of $H_k(\tilde{B}_{n,k}(\Sigma), \partial\tilde{B}_{n,k}(\Sigma))$ by duplicating the α_i^*, β_i^* or γ_j^*, depending on a given $(2g+n-1)$-tuple in $I(n, k, g)$. This geometric object is called a *noodle*. In fact, a noodle uniquely determines a relative k-cycle in $H_k(\tilde{B}_{n,k}(\Sigma), \partial\tilde{B}_{n,k}(\Sigma))$ by lifting the Cartesian product of k arcs in the noodle. For example, the noodle corresponding to $(0, 2, 1, 0, 0, 3) \in I(3, 6, 2)$ looks like the set of arcs on the right of Figure 1.

For a k-cycle F determined by a fork and a relative k-cycle N determined by a noodle, the sesquilinear pairing $\langle F, N\rangle$ computes algebraic intersections between them with coefficients in $\mathbb{Z}[G_\Sigma]$. This pairing can easily be computed by recording intersections between the fork and the noodle on Σ. The basis determined by forks and the basis determined by noodles are dual with respect to the pairing.

In the case of a disc, Bigelow [2004] showed that this pairing is invariant under the action by $\mathbf{B}_{0,n}(D)$. However, in the case of a surface Σ of genus ≥ 1, it cannot be invariant under the action by $\mathbf{B}_{0,n}(\Sigma)$. In fact, the pairing cannot be preserved by any braid group action given by a representation Ψ into $\mathrm{Aut}_{\mathbb{Z}[G_\Sigma]}(H_k^{\mathrm{BM}}(\tilde{B}_{n,k}))$. Suppose it is preserved, that is,

$$\langle F, N\rangle = \langle \Psi(\beta)(F), \Psi(\beta)(N)\rangle.$$

for any $\beta \in \mathbf{B}_{0,n}(\Sigma)$, a k-cycle F determined by a fork, and a relative k-cycle N determined by a noodle. Then for any $y \in G_\Sigma$,

$$\begin{aligned} y\langle F, N\rangle &= \langle yF, N\rangle \\ &= \langle \Psi(\beta)(yF), \Psi(\beta)(N)\rangle \\ &= \beta_\sharp(y)\langle \Psi(\beta)(F), \Psi(\beta)(N)\rangle = \beta_\sharp(y)\langle F, N\rangle. \end{aligned}$$

The property $y = \beta_\sharp(y)$ for all $y \in G_\Sigma$ would force us to set $q = 1$ in G_Σ, and so it was abandoned.

Nonetheless, we can extend this pairing to the pairing $\langle\,\cdot\,,\,\cdot\,\rangle_{H_\Sigma}$ from

$$\mathbb{Z}[H_\Sigma] \otimes_{\mathbb{Z}[G_\Sigma]} H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma)) \times \mathbb{Z}[H_\Sigma] \otimes_{\mathbb{Z}[G_\Sigma]} H_k(\tilde{B}_{n,k}(\Sigma), \partial\tilde{B}_{n,k}(\Sigma))$$

to S' defined by

$$\Big\langle \sum_i g_i F_i, \sum_j h_j N_j \Big\rangle_{H_\Sigma} = \sum_{i,j} g_i \langle F_i, N_j\rangle h_j^{-1},$$

where $g_i, h_j \in \mathbb{Z}[H_\Sigma]$ and S' is the skew field containing $\mathbb{Z}[H_\Sigma]$. Note that this extended pairing cannot be invariant under the braid group action given by Φ_k either. However it can be used to compute proposed representations Φ_k explicitly. The following theorem summarizes the above discussion.

Theorem 3.4. *Let the F_i and N_i be k-cycles and relative k-cycles in dual bases determined by forks and noodles. Then $\Phi_k(\beta)$ for each $\beta \in \mathbf{B}_{0,n}(\Sigma)$ is represented by a matrix with respect to the basis $\{F_i \mid 1 \leq i \leq \binom{2g+n+k-2}{k}\}$ whose (i, j)-th entry is given by $\psi_\Sigma(\beta)\langle\tilde{\beta}(F_i), N_j\rangle_{H_\Sigma}$, which is an element of $\mathbb{Z}[H_\Sigma]$ rather than of S'.*

As an example, we will show the matrix form of the representation Φ_1 of the 3-braid group $\mathbf{B}_{0,3}(\Sigma)$ is an extension of the Burau representation when $\Sigma = \Sigma(2, 1)$. Since $k = 1$, the basis of $H_1^{\mathrm{BM}}(\tilde{B}_{3,1}(\Sigma))$ determined by forks can be expressed by $\{\gamma_1, \gamma_2, \alpha_1, \alpha_2, \beta_1, \beta_2\}$ and similarly the dual basis of $H_1(\tilde{B}_{3,1}(\Sigma), \partial\tilde{B}_{3,1}(\Sigma))$ determined by noodles is written by $\{\gamma_1^*, \gamma_2^*, \alpha_1^*, \alpha_2^*, \beta_1^*, \beta_2^*\}$.

Figure 2 shows the action of σ_1 on the fork β_1. It also shows intersection points p_1 and p_2 with γ_1^* and p_3 with β_1^*. In the covering space, the intersection point

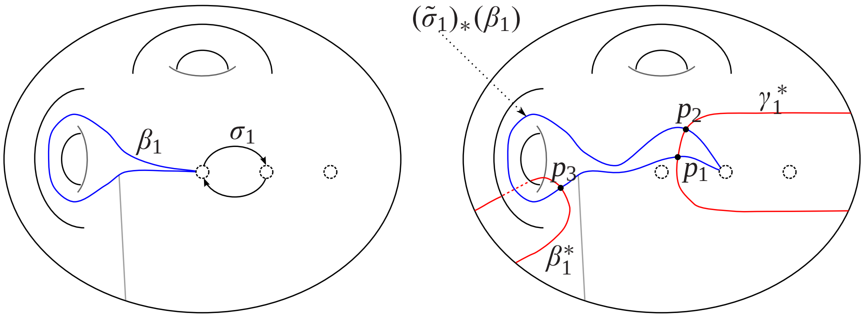


Figure 2. An example of the pairing of a fork and a noodle. At left: fork β_1 and σ_1. At right: fork and noodles.

p_1 lies on the sheet transformed by q since the fork wraps a puncture, and p_2 lies on the sheet transformed by $\ell_1 q$ since the fork contains the longitude of the first handle and wraps a puncture. We have the negative sign for p_2 since the orientation is switched. Finally, p_3 lies on the sheet containing the base point of the covering space. Therefore we have $\Phi_1(\sigma_1)(\beta_1) = \beta_1 + q(1-\ell_1)\gamma_1$. By a similar computation, we can obtain every entry of $\Phi_1(\sigma_1)$:

$$\Phi_1(\sigma_1) = \left(\begin{array}{cc|cccc} -q & 1 & q(1-m_1) & q(1-m_2) & q(1-\ell_1) & q(1-\ell_2) \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline \multicolumn{2}{c|}{\mathbf{0}} & \multicolumn{4}{c}{I_4} \end{array}\right),$$

$$\Phi_1(\sigma_2) = \begin{pmatrix} 1 & 0 \\ q & -q \end{pmatrix} \oplus I_4,$$

$$\Phi_1(\mu_1) = \overline{m}_1 \left(\begin{array}{cc|cc|cc} \multicolumn{2}{c|}{I_2} & \multicolumn{4}{c}{\mathbf{0}} \\ \hline 1 & 0 & m_1 q & q(m_2-1) & \ell_1 - 1 & q(\ell_2-1) \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline \multicolumn{2}{c|}{\mathbf{0}} & \multicolumn{2}{c|}{\mathbf{0}} & \multicolumn{2}{c}{I_2} \end{array}\right),$$

$$\Phi_1(\mu_2) = \overline{m}_2 \left(\begin{array}{cc|cc|cc} \multicolumn{2}{c|}{I_2} & \multicolumn{4}{c}{\mathbf{0}} \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & m_1 - 1 & m_2 q & \ell_1 - 1 & \ell_2 - 1 \\ \hline \multicolumn{2}{c|}{\mathbf{0}} & \multicolumn{2}{c|}{\mathbf{0}} & \multicolumn{2}{c}{I_2} \end{array}\right),$$

$$\Phi_1(\lambda_1) = \bar{\ell}_1 \left(\begin{array}{cc|cc|cc} \multicolumn{2}{c|}{I_2} & \multicolumn{4}{c}{\mathbf{0}} \\ \hline \multicolumn{2}{c|}{\mathbf{0}} & q & 0 & \multicolumn{2}{c}{\mathbf{0}} \\ & & 0 & 1 & & \\ \hline 1 & 0 & q(m_1 q - 1) & q(m_2 - 1) & \ell_1 q & q(\ell_2 - 1) \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right),$$

$$\Phi_1(\lambda_2)=\bar{\ell}_2\left(\begin{array}{cc|cc|cc} \multicolumn{2}{c|}{I_2} & \multicolumn{4}{c}{\mathbf{0}} \\ \hline \multicolumn{2}{c|}{\multirow{2}{*}{\mathbf{0}}} & 1 & 0 & \multicolumn{2}{c}{\multirow{2}{*}{\mathbf{0}}} \\ & & 0 & q & & \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & m_1-1 & q(m_2q-1) & \ell_1-1 & \ell_2 q \end{array}\right).$$

Similarly, we can compute the matrix form for $k=2$ that is the extension of Lawrence–Krammer–Bigelow representation. For $g=1$ and $n=3$, we have by Lemma 3.3 a 10×10 matrix for each generator. Fix a basis for $H_2^{\mathrm{BM}}(\tilde{B}_{3,2}(\Sigma))$ as shown in Figure 1. Let

$$\begin{aligned} &w_{1,1}=(0,0,2,0), \quad w_{1,2}=(0,0,1,1), \quad w_{2,2}=(0,0,0,2),\\ &a_{0,0}=(2,0,0,0), \quad a_{0,1}=(1,0,1,0), \quad a_{0,2}=(1,0,0,1),\\ &b_{0,0}=(0,2,0,0), \quad b_{0,1}=(0,1,1,0), \quad b_{0,2}=(0,1,0,1), \quad z=(1,1,0,0) \end{aligned}$$

in $I(3,2,1)$. Then the action of σ_1 on this basis is as follows:

$$\begin{aligned} \Phi_2(\sigma_1)(w_{1,1}) &= tq^2 w_{1,1},\\ \Phi_2(\sigma_1)(w_{1,2}) &= -tqw_{1,1}-qw_{1,2},\\ \Phi_2(\sigma_1)(w_{2,2}) &= w_{1,1}+(1+t^{-1})w_{1,2}+w_{2,2},\\ \Phi_2(\sigma_1)(a_{0,0}) &= a_{0,0}+q(1+t^{-1})(1-m_1t)a_{0,1}+q^2(m_1^2-(1+t)m_1+1)w_{1,1},\\ \Phi_2(\sigma_1)(a_{0,1}) &= -qa_{0,1}+q^2t(m_1-1)w_{1,1},\\ \Phi_2(\sigma_1)(a_{0,2}) &= a_{0,1}+a_{0,2}+qt(1-m_1)w_{1,1}+q(1-m_1)w_{1,2},\\ \Phi_2(\sigma_1)(b_{0,0}) &= b_{0,0}+q(1+t^{-1})(1-\ell_1t)b_{0,1}+q^2(\ell_1^2-(1+t)\ell_1+1)w_{1,1},\\ \Phi_2(\sigma_1)(b_{0,1}) &= -qb_{0,1}+q^2t(\ell_1-1)w_{1,1},\\ \Phi_2(\sigma_1)(b_{0,2}) &= b_{0,1}+b_{0,2}+qt(1-\ell_1)w_{1,1}+q(1-\ell_1)w_{1,2},\\ \Phi_2(\sigma_1)(z) &= q(t^{-1}-t\ell_1)a_{0,1}+q(1-m_1)b_{0,1}+q^2(1+m_1(\ell_1-1)-t\ell_1)w_{1,1}+z. \end{aligned}$$

The correspondence between the basis $\{v_{j,k}\}$ in [Bigelow 2001] and our basis is

$$\begin{aligned} v_{1,2} &= -tq^{-4}w_{1,1},\\ v_{1,3} &= -tq^{-4}(w_{1,1}+q(1-t^{-1})w_{1,2}+q^2w_{2,2}),\\ v_{2,3} &= -tq^{-2}w_{2,2}. \end{aligned}$$

Then the action of Φ_2 on the basis $\{v_{j,k}\}$ together with substitution $t\mapsto -t$ is exactly that of Lawrence–Krammer–Bigelow representation in [Bigelow 2001].

4. Justification of the proposed representations

To add to the family of representations proposed in the previous section, we will now investigate the possibility that there may be other representations of the surface braid groups that extend the homology linear representations of the classical braid groups. One may try to consider alternatives in the three ways — a group extension of $\mathbf{B}_{n,k}(\Sigma)$ other than $\mathbf{B}_{n;k}(\Sigma)$, a quotient group of $\mathbf{B}_{n;k}(\Sigma)$ other than H_Σ, and an action on H_Σ by $\mathbf{B}_{0,n}(\Sigma)$ other than right multiplication via the quotient map.

***Group extension of* $\mathbf{B}_{n,k}(\Sigma)$.** To make adjustment of coefficients in the most flexible way, we may try to find the largest possible group extension $\mathbf{E}_{n,k}(\Sigma)$ of $\mathbf{B}_{n,k}(\Sigma)$ such that $\mathbf{B}_{0,n}(\Sigma)$ acts on $\mathbf{E}_{n,k}(\Sigma)$. If we regard $\mathbf{B}_{0,n}(\Sigma)$ and $\mathbf{B}_{n,k}(\Sigma)$ as subgroups of some large braid group $\mathbf{B}_{0,n+k+\ell}(\Sigma)$, then $\mathbf{B}_{0,n}(\Sigma)$ acts naturally on $\mathbf{B}_{0,n+k+\ell}(\Sigma)$ as well as on $\mathbf{B}_{n,k}(\Sigma)$ by conjugation. Thus we assume that $\mathbf{B}_{n,k}(\Sigma) \subset \mathbf{E}_{n,k}(\Sigma) \subset \mathbf{B}_{0,n+k+\ell}(\Sigma)$ for some $\ell \geq 0$.

Lemma 4.1. *Let Σ be a surface with nonempty boundary and let Σ' be a collar neighborhood of $\partial\Sigma$. Let $N(\mathbf{B}_{n,k}(\Sigma))$ denote the normalizer of $\mathbf{B}_{n,k}(\Sigma)$ in $\mathbf{B}_{0,n+k+\ell}(\Sigma)$ for some $\ell \geq 0$. Then $N(\mathbf{B}_{n,k}(\Sigma)) \cong \mathbf{B}_{n;k}(\Sigma) \times \mathbf{B}_{0,\ell}(\Sigma')$.*

Proof. We first identify $\mathbf{B}_{n,k}(\Sigma)$ and $\mathbf{B}_{0,n}(\Sigma)$ with the corresponding subgroups of $\mathbf{B}_{0,n+k+\ell}(\Sigma)$ via the embeddings that add trivial ℓ and $k+\ell$ strands, respectively. Then we will show $N(\mathbf{B}_{n,k}(\Sigma)) = \mathbf{B}_{n;k}(\Sigma) \times \mathbf{B}_{0,\ell}(\Sigma')$ as subgroups of $\mathbf{B}_{0,n+k+\ell}(\Sigma)$. It is clear that $\mathbf{B}_{n;k}(\Sigma) \times \mathbf{B}_{0,\ell}(\Sigma') \subset N(\mathbf{B}_{n,k}(\Sigma))$ since $\mathbf{B}_{n,k}(\Sigma)$ is a normal subgroup of $\mathbf{B}_{n;k}(\Sigma)$ from the short exact sequence of Lemma 2.3 and since elements of $\mathbf{B}_{0,\ell}(\Sigma')$ commute with those of $\mathbf{B}_{n,k}(\Sigma)$. Conversely, let $\beta \in N(\mathbf{B}_{n,k}(\Sigma)) \subset \mathbf{B}_{0,n+k+\ell}(\Sigma)$. Any element $\alpha \in \mathbf{B}_{n,k}(\Sigma)$ and its conjugate $\beta^{-1}\alpha\beta \in \mathbf{B}_{n,k}(\Sigma)$ induce permutations that preserve the sets $\{1,\dots,n\}$, $\{n+1,\dots,n+k\}$ and $\{n+k+1,\dots,n+k+\ell\}$. It is easy to see that the induced permutation of β itself must fix these three sets since α can be arbitrary in $\mathbf{B}_{n,k}(\Sigma)$. Thus $\beta \in \mathbf{B}_{n+k;l}(\Sigma)$ and the split exact sequence

$$1 \to \mathbf{B}_{n+k,l}(\Sigma) \longrightarrow \mathbf{B}_{n+k;l}(\Sigma) \xrightarrow{(\pi_{n+k})_*} \mathbf{B}_{0,n+k}(\Sigma) \to 1$$

gives a unique decomposition $\beta = \beta_1\beta_2$ for $\beta_1 \in \mathbf{B}_{0,n+k}(\Sigma)$ and $\beta_2 \in \mathbf{B}_{n+k,\ell}(\Sigma)$. In fact, $\beta_1 = (\pi_{n+k})_*(\beta) \in \mathbf{B}_{n;k}(\Sigma)$ since the epimorphism $(\pi_{n+k})_*$ forgets the last ℓ strands or replaces them by the trivial ℓ-strand braid.

For any $\alpha \in \mathbf{B}_{n,k}(\Sigma) \subset \mathbf{B}_{0,n+k}(\Sigma) \subset \mathbf{B}_{0,n+k+\ell}(\Sigma)$, we have $(\pi_{n+k})_*(\beta_2^{-1}\alpha\beta_2) = \beta_2^{-1}\alpha\beta_2$ since $\beta_2^{-1}\alpha\beta_2 \in \mathbf{B}_{0,n+k}$. On the other hand, $(\pi_{n+k})_*(\beta_2^{-1}\alpha\beta_2) = \alpha$ since $(\pi_{n+k})_*$ replaces the last ℓ strands by the trivial braid. Thus we have $\beta_2^{-1}\alpha\beta_2 = \alpha$. From the presentation of $\mathbf{B}_{0,n+k+\ell}(\Sigma)$ in Section 1A, it is easy to see that β_2 must be a local braid in order for β_2 to commute with every element of $\mathbf{B}_{n,k}(\Sigma)$. Thus we have $\beta_2 \in \mathbf{B}_{0,\ell}(\Sigma')$, where Σ' is an annulus that is a collar neighborhood of $\partial\Sigma$ in Σ. Consequently, we have shown $N(\mathbf{B}_{n,k}(\Sigma)) \subset \mathbf{B}_{n;k}(\Sigma) \times \mathbf{B}_{0,\ell}(\Sigma')$ □

By this lemma, the extension $\mathbf{E}_{n,k}(\Sigma)$ of $\mathbf{B}_{n,k}(\Sigma)$ can be taken as a subgroup of $\mathbf{B}_{n;k}(\Sigma) \times \mathbf{B}_{0,l}(\Sigma')$. We remark that $\mathbf{B}_{n;k}(\Sigma) \times \mathbf{B}_{0,l}(\Sigma')$ is also a subgroup of the intertwining braid group $\mathbf{B}_{n;k+l}(\Sigma)$.

Then we follow the construction given in the discussion before Theorem 3.2 with $\mathbf{E}_{n,k}(\Sigma)$ replacing $\mathbf{B}_{n;k}(\Sigma)$.

Let $\psi : \mathbf{E}_{n,k}(\Sigma) \to H$ be an epimorphism onto a group H. If we choose an action of $\mathbf{B}_{0,n}(\Sigma)$ on the extension $\mathbf{E}_{n,k}(\Sigma)$, then the action is carried over H via ψ and it is convenient to use the convention that $(\beta_1\beta_2)\cdot h = \beta_2 \cdot (\beta_1 \cdot h)$ for $\beta_1, \beta_2 \in \mathbf{B}_{0,n}(\Sigma)$ and $h \in H$. To obtain a $\mathbb{Z}[H]$-module automorphism $\beta \otimes \tilde{\beta}_*$ on $\mathbb{Z}[H] \otimes_{\mathbb{Z}[G]} H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$ that is an extension of a homology linear representation of the classical braid group, this induced action of $\mathbf{B}_{0,n}(\Sigma)$ on H needs to satisfy two conditions.

(i) Lifting criteria: $\beta_\sharp$ exists and $\beta_\sharp(\phi(\alpha)) = \phi(\bar{\beta}_*(\alpha))$ for all $\alpha \in \mathbf{B}_{n,k}(\Sigma)$, where $\phi = \psi|_{\mathbf{B}_{n,k}(\Sigma)}$.

(ii) Linearity and compatibility: $hh'(\beta\cdot 1) = \beta\cdot(hh') = (\beta\cdot h)\beta_\sharp(h')$ for all $h \in H$ and $h' \in G = \phi(\mathbf{B}_{n,k}(\Sigma))$.

As in the proof of Theorem 3.2, we then have

$$(\beta \otimes \tilde{\beta}_*)(h \otimes h'c) = (\beta \otimes \tilde{\beta}_*)(hh' \otimes c) = hh'(\beta \otimes \tilde{\beta}_*)(1 \otimes c)$$

for all $h \in H,\ \ h' \in G$ and $c \in H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$.

Theorem 4.2. *Suppose there are an epimorphism $\psi : \mathbf{E}_{n,k}(\Sigma) \to H$ and an action of $\mathbf{B}_{0,n}(\Sigma)$ on H satisfying the two conditions above. Let Ψ_k be the representation obtained from ψ and the action. Then*

$$\Psi_k = 1_{\mathbb{Z}[H]} \otimes_{\mathbb{Z}[H']} \Psi'_k$$

for a representation Ψ'_k obtained from an epimorphism $\psi' : \mathbf{B}_{n;k}(\Sigma) \to H' \subset H$ and an action $\mathbf{B}_{0,n}(\Sigma)$ on H', where $1_{\mathbb{Z}[H]}$ is the identity map on $\mathbb{Z}[H]$.

Proof. Let $H' = \{\beta\cdot 1 \in H \mid \beta \in \mathbf{B}_{0,n}(\Sigma)\}\phi(\mathbf{B}_{n,k}(\Sigma))$ and $\psi' : \mathbf{B}_{n;k}(\Sigma) \to H'$ be a surjection defined by $\psi'(\beta) = \beta\cdot 1$ for $\beta \in \mathbf{B}_{0,n}(\Sigma)$ and $\psi' = \phi$ on $\mathbf{B}_{n,k}(\Sigma)$. Then since

$$\psi'(\beta_1\beta_2) = (\beta_1\beta_2)\cdot 1 = \beta_2\cdot(\beta_1\cdot 1) = (\beta_1\cdot 1)(\beta_2\cdot 1) = \psi'(\beta_1)\psi'(\beta_2)$$

for all $\beta_1, \beta_2 \in \mathbf{B}_{0,n}(\Sigma)$, the surjection ψ' becomes a homomorphism that preserves the semidirect product structure. Also we have

$$\phi' = \psi'|_{\mathbf{B}_{n,k}(\Sigma)} = \psi|_{\mathbf{B}_{n,k}(\Sigma)} = \phi$$

and so ϕ and ϕ' induce the same homology group $H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$, and two $\mathbb{Z}$-module automorphisms obtained from β coincide.

Consider two representations Ψ_k and Ψ'_k corresponding to ψ and ψ', respectively. Then $\Psi_k(\beta)$ gives a $\mathbb{Z}[H]$-homomorphism on $\mathbb{Z}[H]\otimes_{\mathbb{Z}[G]} H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$ and $\Psi'_k(\beta)$ gives a $\mathbb{Z}[H']$-homomorphism on $\mathbb{Z}[H']\otimes_{\mathbb{Z}[G]} H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$. Since $\mathbb{Z}[H]=\mathbb{Z}[H]\otimes_{\mathbb{Z}[H']}\mathbb{Z}[H']$, the representation $\Psi_k(\beta)$ is a $\mathbb{Z}[H]$-homomorphism on $\mathbb{Z}[H]\otimes_{\mathbb{Z}[H']}\mathbb{Z}[H']\otimes_{\mathbb{Z}[G]} H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$ defined by

$$\Psi_k(\beta)(hh'\otimes c)=hh'(\beta\cdot 1)\otimes\tilde{\beta}_*(c)=h\otimes h'(\beta\cdot 1)\otimes\tilde{\beta}_*(c)$$

for all $h\in\mathbb{Z}[H]$, $h'\in\mathbb{Z}[H']$ and $c\in H_k^{\mathrm{BM}}(\tilde{B}_{n,k}(\Sigma))$. As claimed, this is equal to $1_{\mathbb{Z}[H]}\otimes_{\mathbb{Z}[H']}\Psi'_k(\beta)$. □

This theorem implies that we may assume that $\mathbf{B}_{n;k}(\Sigma)\subset\mathbf{E}_{n,k}(\Sigma)$ without loss of generality. Then by Lemma 4.1, $\mathbf{E}_{n,k}(\Sigma)=\mathbf{B}_{n;k}(\Sigma)\times\mathbf{B}$ for some subgroup $\mathbf{B}$ of $\mathbf{B}_{0,\ell}(\Sigma')$ and the theorem says that any family of representations obtained by using $\mathbf{E}_{n,k}(\Sigma)$ is merely a trivial extension of the family of representations proposed in Section 3.

***Quotient of* $\mathbf{B}_{n;k}(\Sigma)$.** According to the scheme described in Theorem 3.2, it is important to find a good epimorphism $\psi:\mathbf{B}_{n;k}(\Sigma)\to H$ onto some group H.

Since Σ is not a sphere, the inclusion $B_{n,k}(D)\hookrightarrow B_{n,k}(\Sigma)$ induces a monomorphism $\mathbf{B}_{n,k}(D)\hookrightarrow\mathbf{B}_{n,k}(\Sigma)$; see [Birman 1974]. Similarly, $B_{n;k}(D)\hookrightarrow B_{n;k}(\Sigma)$ induces a monomorphism $\mathbf{B}_{n;k}(D)\hookrightarrow\mathbf{B}_{n;k}(\Sigma)$ (to be regarded as an inclusion).

We first determine an epimorphism $\psi_D:\mathbf{B}_{n;k}(D)\to H_D$ to extend the map $\phi_D:\mathbf{B}_{n,k}(D)\to G_D$ for the classical braid groups. Since we want to obtain homology linear representations for the classical braid groups, we should use that $H_D=G_D$, and all of the extra generators $\bar{\sigma}_1,\dots,\bar{\sigma}_{n-1}$ of $\mathbf{B}_{n;k}(D)$ should be sent to the identity by ψ_D, as we have seen earlier in Section 3A. Then $\psi_D|_{\mathbf{B}_{n,k}(D)}=\phi_D$. For some extension H of G_D, let $\psi:\mathbf{B}_{n;k}(\Sigma)\to H$ be an epimorphism. To obtain an extension of homology linear representations of the classical braid groups via ψ, we require the condition

$$(\dagger)\qquad \psi|_{\mathbf{B}_{n;k}(D)}=\psi_D$$

This condition is nothing but a reinterpretation of Definition 2.2 and is necessary to make the diagram

$$\begin{array}{ccccc}
\mathbf{B}_{n,k}(D) & \hookrightarrow & \mathbf{B}_{n;k}(D) & \hookrightarrow & \mathbf{B}_{n;k}(\Sigma)\\
\downarrow{\scriptstyle \phi_D} & & \downarrow{\scriptstyle \psi_D} & & \downarrow{\scriptstyle \psi}\\
G_D & = & H_D & \hookrightarrow & H
\end{array}$$

commutative, so that $\psi|_{\mathbf{B}_{n,k}(D)} = \phi_D$ and we can then apply the construction of Theorem 3.2. We first show that the condition (†) imposes restrictions on the choice of H.

Theorem 4.3. *Let $\psi_\Sigma : \mathbf{B}_{n;k}(\Sigma) \to H_\Sigma$ be the epimorphism defined in Section 3A.*

(1) *Let $h : H_\Sigma \to H$ be an epimorphism such that $h \circ \psi_\Sigma$ satisfies (†). Then h is an isomorphism.*

(2) *Let $\psi : \mathbf{B}_{n;k}(\Sigma) \to H$, an arbitrary epimorphism onto a group H, satisfy (†). Then ψ_Σ factors through ψ, and H is isomorphic to H_Σ for $k \geq 3$.*

Proof. (1) It suffices to show that $h(W) = 1$ implies $W = 1$ for any word W in generators of H_Σ. Assume $k \geq 2$. Using the relations of H_Σ, a given word W can be put into the form

$$W = q^c t^d \prod_{i=1}^{g} W_i, \quad \text{where } W_i = m_i^{a_i} \ell_i^{b_i} \overline{m}_i^{\bar{a}_i} \bar{\ell}_i^{\bar{b}_i}.$$

First consider $[W, \bar{\ell}_r]$. Note that W_r commutes with the other W_i as well as q and t. Since $\bar{\ell}_r$ commutes with all generators except m_r and only W_r contains m_r, we have

$$\begin{aligned}
[W, \bar{\ell}_r] &= \Big(q^c t^d \prod_i W_i\Big) \bar{\ell}_r \Big(q^c t^d \prod_i W_i\Big)^{-1} \bar{\ell}_r^{-1} \\
&= W_r \Big(q^c t^d \prod_{i \neq r} W_i\Big) \bar{\ell}_r \Big(q^c t^d \prod_{i \neq r} W_i\Big)^{-1} W_r^{-1} \bar{\ell}_r^{-1} \\
&= W_r \bar{\ell}_r W_r^{-1} \bar{\ell}_r^{-1} \\
&= \big(m_r^{a_r} \ell_r^{b_r} \overline{m}_r^{\bar{a}_r} \bar{\ell}_r^{\bar{b}_r}\big) \bar{\ell}_r \big(m_r^{a_r} \ell_r^{b_r} \overline{m}_r^{\bar{a}_r} \bar{\ell}_r^{\bar{b}_r}\big)^{-1} \bar{\ell}_r^{-1} \\
&= m_r^{a_r} \big(\ell_r^{b_r} \overline{m}_r^{\bar{a}_r} \bar{\ell}_r^{\bar{b}_r}\big) \bar{\ell}_r \big(\ell_r^{b_r} \overline{m}_r^{\bar{a}_r} \bar{\ell}_r^{\bar{b}_r}\big)^{-1} m_r^{-a_r} \bar{\ell}_r^{-1} \\
&= m_r^{a_r} \bar{\ell}_r m_r^{-a_r} \bar{\ell}_r^{-1} = q^{a_r}.
\end{aligned}$$

The last equality follows from the relation $[m_r, \bar{\ell}_r] = q$. By applying h and using $h(W) = 1$, we have

$$h(q^{a_r}) = h([W, \bar{\ell}_r]) = h(W) h(\bar{\ell}_r) h(W)^{-1} h(\bar{\ell}_r)^{-1} = 1.$$

By (†), h is the identity on G_D that is the subgroup generated by q and t, and q and t are of infinite order. Thus $h(q^{a_r}) = q^{a_r} = 1$ implies $a_r = 0$. Similarly, $b_r = \bar{a}_r = \bar{b}_r = 0$ by considering $[W, \overline{m}_r]$, $[W, \ell_r]$, and $[W, m_r]$. Therefore $W_r = 1$. Since r is arbitrary other than $1 \leq r \leq g$, we now have $W = q^c t^d$. Then $1 = h(W) = q^c t^d$ implies $c = d = 0$. Consequently, $W = 1$.

For the case $k = 1$, the proof is similar but simpler since $t = 1$ in H_Σ.

(2) Consider the commutative diagram

$$\begin{array}{ccc} \mathbf{B}_{n;k}(\Sigma) & \xrightarrow{\psi} & H \cong \mathbf{B}_{n;k}(\Sigma)/\operatorname{Ker}\psi \\ \downarrow \psi_\Sigma & & \downarrow q \\ H_\Sigma \cong \mathbf{B}_{n;k}(\Sigma)/\operatorname{Ker}\psi_\Sigma & \xrightarrow{h} & \mathbf{B}_{n;k}(\Sigma)/(\operatorname{Ker}\psi\cdot\operatorname{Ker}\psi_\Sigma), \end{array}$$

which consists of obvious quotient homomorphisms. Note that the condition (†) is equivalent to $\mathbf{B}_{n;k}(D)/(\operatorname{Ker}\psi\cap\mathbf{B}_{n;k}(D)) \cong G_D$. Thus $\operatorname{Ker}\psi\cap\mathbf{B}_{n;k}(D) = \operatorname{Ker}\psi_\Sigma\cap\mathbf{B}_{n;k}(D)$ since ψ_Σ also satisfies (†). Then

$$\begin{aligned}(\operatorname{Ker}\psi\cdot\operatorname{Ker}\psi_\Sigma)\cap\mathbf{B}_{n;k}(D) &= (\operatorname{Ker}\psi\cap\mathbf{B}_{n;k}(D))\cdot(\operatorname{Ker}\psi_\Sigma\cap\mathbf{B}_{n;k}(D)) \\ &= \operatorname{Ker}\psi_\Sigma\cap\mathbf{B}_{n;k}(D).\end{aligned}$$

Thus $h\circ\psi_\Sigma$ satisfies (†) and h is an isomorphism by part (1). Therefore ψ_Σ factors through ψ via $h^{-1}\circ q$ for $k\geq 1$.

For $k\geq 3$, we will show $\psi:\mathbf{B}_{n;k}(\Sigma)\to H$ factors through ψ_Σ, that is, there is an epimorphism $h: H_\Sigma\to H$ such that $h\psi_\Sigma=\psi$. Then H_Σ is isomorphic to H since ψ_Σ also factors through ψ.

Recall the presentation for $\mathbf{B}_{n;k}(\Sigma)$ in Lemma 2.5. The condition (†) implies $\psi(\sigma_i)=q$, $\psi(\zeta_j)=t$, and $\psi(\bar\sigma_m)=1$ for all $1\leq i\leq k-1$, $1\leq j\leq n$, and $1\leq m\leq n-1$. Since $k\geq 3$, the relation (CR_1) among generators in X_2 is not vacuous and so the relations (CR_1) through (CR_3) for X_2 and the condition (†) imply

$$[\psi(\mu_r),q]=[\psi(\lambda_r),q]=[\psi(\mu_r),t]=[\psi(\lambda_r),t]=[q,t]=1$$

for all $1\leq r\leq g$. Also the relation Lemma 2.5(iii) implies

$$[\psi(\bar\mu_r),q]=[\psi(\bar\lambda_r),q]=[\psi(\bar\mu_r),t]=[\psi(\bar\lambda_r),t]=1 \quad \text{for all } 1\leq r\leq g.$$

Thus q and t lie in the center of H. Using this, all other relations in H_Σ can be shown to hold in H. Therefore ψ induces an epimorphism $h: H_\Sigma\to H$. □

Hence H_Σ is the unique quotient group of $\mathbf{B}_{n;k}(\Sigma)$ satisfying (†) for $k\geq 3$. For $k\leq 2$, the condition (†) does not uniquely determine a quotient group of $\mathbf{B}_{n;k}(\Sigma)$. To take advantage of representations in analyzing the surface braid group $\mathbf{B}_{0,n}(\Sigma)$, one may prefer a simpler coefficient ring as long as the representation carries enough information. For the classical case, there are also several groups satisfying the condition (∗) if we do not assume they are abelian. For the surface braid groups, we cannot obtain any interesting representation if an abelian coefficient ring is used, as discussed in Section 2. Theorem 4.3(1) says that H_Σ is the simplest quotient group satisfying (†) in the sense that any further quotient of H_Σ violates (†).

We now discuss possible actions of $\mathbf{B}_{0,n}(\Sigma)$ on H_Σ induced from ψ_Σ.

Theorem 4.4. *Let $\psi_\Sigma : \mathbf{B}_{n;k}(\Sigma) \to H_\Sigma$ be the epimorphism defined in Section 3A. Let $\beta \cdot h$ denote any action on $h \in H_\Sigma$ by $\beta \in \mathbf{B}_{0,n}(\Sigma)$ that is induced from ψ_Σ and satisfies the two conditions given above Theorem 4.2. Then*

$$\beta \cdot h = h\chi(\beta)\psi_\Sigma(\beta)$$

for some function $\chi : \mathbf{B}_{0,n}(\Sigma) \to C_{H_\Sigma}(G_\Sigma)$ with the property that

$$(\chi, \psi_\Sigma) : \mathbf{B}_{0,n}(\Sigma) \to C_{H_\Sigma}(G_\Sigma) \rtimes H_\Sigma$$

is a homomorphism, where $C_{H_\Sigma}(G_\Sigma)$ denotes the centralizer of G_Σ in H_Σ.

Proof. By the hypotheses of the action, we have

$$h'(\beta \cdot 1) = \beta \cdot (1h') = (\beta \cdot 1)\beta_\sharp(h') \quad \text{and} \quad \beta_\sharp(h') = \psi_\Sigma(\beta)^{-1}h'\psi_\Sigma(\beta)$$

for all $h' \in G_\Sigma$. By combining these two equations, we have

$$\psi_\Sigma(\beta)^{-1}h'\psi_\Sigma(\beta) = (\beta \cdot 1)^{-1}h'(\beta \cdot 1).$$

and so $(\beta \cdot 1)\psi_\Sigma(\beta)^{-1} \in C_{H_\Sigma}(G_\Sigma)$. Hence $(\beta \cdot 1) = \chi(\beta)\psi_\Sigma(\beta)$ for a function $\chi : \mathbf{B}_{0,n}(\Sigma) \to C_{H_\Sigma}(G_\Sigma)$. Since $\chi(\beta_1\beta_2)\psi_\Sigma(\beta_1\beta_2) = (\beta_1\beta_2) \cdot 1 = \beta_2 \cdot (\beta_1 \cdot 1) = (\chi(\beta_1)\psi_\Sigma(\beta_1))\chi(\beta_2)\psi_\Sigma(\beta_2)$, we have

$$\chi(\beta_1\beta_2) = \chi(\beta_1)\psi_\Sigma(\beta_1)\chi(\beta_2)\psi_\Sigma(\beta_1)^{-1}.$$

This implies that

$$\begin{aligned}(\chi(\beta_1\beta_2), \psi_\Sigma(\beta_1\beta_2)) &= (\chi(\beta_1)\psi_\Sigma(\beta_1)\chi(\beta_2)\psi_\Sigma(\beta_1)^{-1}, \psi_\Sigma(\beta_1\beta_2)) \\ &= (\chi(\beta_1), \psi_\Sigma(\beta_1))(\chi(\beta_2), \psi_\Sigma(\beta_2)).\end{aligned}$$

Therefore $(\chi, \psi_\Sigma) : \mathbf{B}_{0,n}(\Sigma) \to C_{H_\Sigma}(G_\Sigma) \rtimes H_\Sigma$ is a homomorphism. □

The function χ in this theorem behaves like a character of $\mathbf{B}_{0,n}(\Sigma)$. In fact, if $k \geq 2$, it can be shown that $C_{H_\Sigma}(G_\Sigma) = Z(H_\Sigma) = \langle q \rangle \oplus \langle t \rangle$. Hence χ can be any homomorphism from $\mathbf{B}_{0,n}(\Sigma)$ to $Z(H_\Sigma)$. In this case, the representations Ψ_k obtained from ψ are given by $\Psi_k = \chi \otimes \Phi_k$ for some character χ, where Φ_k is the proposed representation in Theorem 3.2.

References

[Bardakov 2005] V. G. Bardakov, "Linear representations of the braid groups of some manifolds", *Acta Appl. Math.* **85**:1-3 (2005), 41–48. MR 2006a:20072 Zbl 1106.20030

[Bellingeri 2004] P. Bellingeri, "On presentations of surface braid groups", *J. Algebra* **274**:2 (2004), 543–563. MR 2005a:20052 Zbl 1081.20045

[Bellingeri and Godelle 2007] P. Bellingeri and E. Godelle, "Positive presentations of surface braid groups", *J. Knot Theory Ramifications* **16**:9 (2007), 1219–1233. MR 2008m:20063 Zbl 1149.20031

[Bigelow 1999] S. Bigelow, "The Burau representation is not faithful for $n = 5$", *Geom. Topol.* **3** (1999), 397–404. MR 2001j:20055 Zbl 0942.20017

[Bigelow 2001] S. J. Bigelow, "Braid groups are linear", *J. Amer. Math. Soc.* **14**:2 (2001), 471–486. MR 2002a:20043 Zbl 0988.20021

[Bigelow 2003] S. Bigelow, "The Lawrence–Krammer representation", pp. 51–68 in *Topology and geometry of manifolds* (Athens, GA, 2001), edited by G. Matić and C. McCrory, Proc. Sympos. Pure Math. **71**, Amer. Math. Soc., Providence, RI, 2003. MR 2004j:20077 Zbl 1061.20034

[Bigelow 2004] S. Bigelow, "Homological representations of the Iwahori–Hecke algebra", pp. 493–507 in *Proceedings of the Casson Fest*, edited by C. Gordon and Y. Rieck, Geom. Topol. Monogr. **7**, Geom. Topol. Publ., Coventry, 2004. MR 2006h:20005 Zbl 1080.20003

[Bigelow and Budney 2001] S. J. Bigelow and R. D. Budney, "The mapping class group of a genus two surface is linear", *Algebr. Geom. Topol.* **1** (2001), 699–708. MR 2002k:20065 Zbl 0999.57020

[Birman 1969] J. S. Birman, "On braid groups", *Comm. Pure Appl. Math.* **22** (1969), 41–72. MR 38 #2764 Zbl 0157.30904

[Birman 1974] J. S. Birman, *Braids, links, and mapping class groups*, Annals of Mathematics Studies **82**, Princeton University Press, 1974. MR 51 #11477 Zbl 0305.57013

[Borel and Moore 1960] A. Borel and J. C. Moore, "Homology theory for locally compact spaces", *Michigan Math. J.* **7** (1960), 137–159. MR 24 #A1123 Zbl 0116.40301

[Van Buskirk 1966] J. van Buskirk, "Braid groups of compact 2-manifolds with elements of finite order", *Trans. Amer. Math. Soc.* **122** (1966), 81–97. MR 32 #6440 Zbl 0138.19103

[Fadell and Neuwirth 1962] E. Fadell and L. Neuwirth, "Configuration spaces", *Math. Scand.* **10** (1962), 111–118. MR 25 #4537 Zbl 0136.44104

[Fadell and van Buskirk 1962] E. Fadell and J. van Buskirk, "The braid groups of E^2 and S^2", *Duke Math. J.* **29** (1962), 243–257. MR 25 #4539 Zbl 0122.17804

[Gonçalves and Guaschi 2003] D. L. Gonçalves and J. Guaschi, "On the structure of surface pure braid groups", *J. Pure Appl. Algebra* **182**:1 (2003), 33–64. MR 2004i:20068 Zbl 1053.20035

[González-Meneses 2001] J. González-Meneses, "New presentations of surface braid groups", *J. Knot Theory Ramifications* **10**:3 (2001), 431–451. MR 2002b:20040 Zbl 1030.20024

[Hughes and Ranicki 1996] B. Hughes and A. Ranicki, *Ends of complexes*, Cambridge Tracts in Mathematics **123**, Cambridge University Press, 1996. MR 98f:57039 Zbl 0876.57001

[Jones 1987] V. F. R. Jones, "Hecke algebra representations of braid groups and link polynomials", *Ann. of Math.* (2) **126**:2 (1987), 335–388. MR 89c:46092 Zbl 0631.57005

[Krammer 2000] D. Krammer, "The braid group B_4 is linear", *Invent. Math.* **142**:3 (2000), 451–486. MR 2001k:20078 Zbl 0988.20023

[Krammer 2002] D. Krammer, "Braid groups are linear", *Ann. of Math.* (2) **155**:1 (2002), 131–156. MR 2003c:20040 Zbl 1020.20025

[Lawrence 1990] R. J. Lawrence, "Homological representations of the Hecke algebra", *Comm. Math. Phys.* **135**:1 (1990), 141–191. MR 92d:16020 Zbl 0716.20022

[Mal'cev 1948] A. I. Mal'cev, "On the embedding of group algebras in division algebras", *Doklady Akad. Nauk SSSR* (*N.S.*) **60** (1948), 1499–1501. In Russian. MR 10,8c

[Neumann 1949] B. H. Neumann, "On ordered division rings", *Trans. Amer. Math. Soc.* **66** (1949), 202–252. MR 11,311f Zbl 0035.30401

[Paris and Rolfsen 1999] L. Paris and D. Rolfsen, "Geometric subgroups of surface braid groups", *Ann. Inst. Fourier* (*Grenoble*) **49**:2 (1999), 417–472. MR 2000f:20059 Zbl 0962.20028

[Scott 1970] G. P. Scott, "Braid groups and the group of homeomorphisms of a surface", *Proc. Cambridge Philos. Soc.* **68** (1970), 605–617. MR 42 #3786 Zbl 0203.56302

[Zheng 2005] H. Zheng, "Faithfulness of the Lawrence representation of braid group", preprint, 2005. arXiv math/0509074

[Zinno 2001] M. G. Zinno, "On Krammer's representation of the braid group", *Math. Ann.* **321**:1 (2001), 197–211. MR 2002h:20055 Zbl 1042.20023

Received June 1, 2009. Revised July 9, 2010.

BYUNG HEE AN
DEPARTMENT OF MATHEMATICS
KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY
DAEJEON 305-701
SOUTH KOREA
anbyhee@knot.kaist.ac.kr

KI HYOUNG KO
DEPARTMENT OF MATHEMATICS
KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY
DAEJEON 305-701
SOUTH KOREA
knot@knot.kaist.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

PARAMETRIZATION OF HOLOMORPHIC SEGRE-PRESERVING MAPS

R. BLAIR ANGLE

We investigate holomorphic Segre-preserving maps sending the complexification $\mathcal{M}$ of a generic real analytic submanifold $M \subseteq \mathbb{C}^N$ of finite type at some point p into the complexification $\mathcal{M}'$ of a generic real analytic submanifold $M' \subseteq \mathbb{C}^{N'}$ that is finitely nondegenerate at some point p'. We prove that for a fixed M and M', the germs at $(p, \bar{p})$ of Segre submersive holomorphic Segre-preserving maps sending $(\mathcal{M}, (p, \bar{p}))$ into $(\mathcal{M}', (p', \bar{p}'))$ rationally depend upon their K-jets at $(p, \bar{p})$, for some fixed K depending only on M and M', and these maps are uniquely determined by their K-jets. If, in addition, M and M' are algebraic, we prove that any such map must be algebraic. It follows that the set of germs at $(p, \bar{p})$ of holomorphic Segre-preserving automorphisms of the complexification $\mathcal{M}$ of a generic real analytic submanifold M that is finitely nondegenerate and of finite type at p is an algebraic complex Lie group. We explore the relationship between this automorphism group and the group of automorphisms of M at p.

1. Introduction

Let $M \subseteq \mathbb{C}^N$ be a real analytic submanifold of codimension d, defined locally near $p \in M$ by the real-valued real analytic function $\rho(Z, \bar{Z})$. The *complexification* $\mathcal{M}$ of M is a holomorphic submanifold of $\mathbb{C}^{2N}$ given locally for $(Z, \zeta) \in \mathbb{C}^N \times \mathbb{C}^N$ near $(p, \bar{p})$ by $\mathcal{M} = \{(Z, \zeta) : \rho(Z, \zeta) = 0\}$. Assume M is generic (see Section 2), and let $M' \subseteq \mathbb{C}^{N'}$ be a generic real analytic submanifold of codimension d', let $p' \in M'$, and let $\mathcal{M}'$ denote its complexification. Let $U \subseteq \mathbb{C}^N$ be an open neighborhood of p, and define ${}^*U := \{\bar{Z} : Z \in U\}$. Consider a holomorphic map $\mathcal{H} : (U \times {}^*U, (p, \bar{p})) \to (\mathbb{C}^{2N'}, (p', \bar{p}'))$ of the form

$$\mathcal{H}(Z, \zeta) = (H(Z), \widetilde{H}(\zeta)), \tag{1-1}$$

where $H : U \to \mathbb{C}^{N'}$ and $\widetilde{H} : {}^*U \to \mathbb{C}^{N'}$. Assume that $\mathcal{H}(\mathcal{M} \cap (U \times {}^*U)) \subseteq \mathcal{M}'$. These maps will be the chief object of study in this paper. We will call such a map

MSC2000: 32A10, 32H02, 32H12, 32V35, 32V40.
Keywords: holomorphic Segre-preserving map, complexification, real analytic submanifold, finite jet determination, algebraic submanifold, finitely nondegenerate, finite type.

a *holomorphic Segre-preserving map* (HSPM) since it preserves Segre varieties in a sense that will be made precise in Section 2. With the notation $\bar{\varphi}(z) := \overline{\varphi(\bar{z})}$, we observe that if $\widetilde{H} = \overline{H}$, then H is a holomorphic map defined on U sending $(M \cap U, p)$ into (M', p'). Such maps have been extensively studied. However, HSPMs are relatively new; see [Zhang 2007; Angle 2008a; 2008b] for related recent work. Under certain restrictions, the collection of germs of HSPMs sending $(\mathcal{M}, (p, \bar{p}))$ into $(\mathcal{M}', (p', \bar{p}'))$ is, in a manner to be detailed in subsequent sections, "bigger" than the collection of germs of holomorphic mappings sending (M, p) into (M', p'). We shall see several examples of this in Section 5.

For $p_0 \in \mathbb{C}^m$, let $\mathcal{T}_{p_0}(\mathbb{C}^m)$ denote the holomorphic tangent space of $\mathbb{C}^m$ at p_0. Let $\mathcal{T}^0_{(p,\bar{p})}\mathcal{M} \subseteq \mathcal{T}_{(p,\bar{p})}(\mathbb{C}^{2N})$ denote the set of all vectors of the form

$$\sum_{j=1}^{N} a_j \frac{\partial}{\partial Z_j} + \sum_{j=1}^{N} b_j \frac{\partial}{\partial \zeta_j} \quad \text{such that} \quad \sum_{j=1}^{N} a_j \frac{\partial}{\partial Z_j} \quad \text{and} \quad \sum_{j=1}^{N} b_j \frac{\partial}{\partial \bar{Z}_j}$$

are tangent to M at p. A vector of the form $\sum_{j=1}^{N} a_j(\partial/\partial Z_j)$ tangent to M at p is known as a *holomorphic tangent vector*, and a vector of the form $\sum_{j=1}^{N} b_j(\partial/\partial \bar{Z}_j)$ tangent to M at p is known as an *antiholomorphic tangent vector*. For any germ at $(p, \bar{p})$ of an HSPM $\mathcal{H}$ sending $(\mathcal{M}, (p, \bar{p}))$ into $(\mathcal{M}', (p', \bar{p}'))$, we have

$$\mathcal{D}_{(p,\bar{p})}\mathcal{H}(\mathcal{T}^0_{(p,\bar{p})}\mathcal{M}) \subseteq \mathcal{T}^0_{(p',\bar{p}')}\mathcal{M}',$$

where we define

$$\mathcal{D}_{(p,\bar{p})}\mathcal{H} : \mathcal{T}_{(p,\bar{p})}(\mathbb{C}^{2N}) \to \mathcal{T}_{(p',\bar{p}')}(\mathbb{C}^{2N'}) \quad \text{by} \quad (\mathcal{D}_{(p,\bar{p})}\mathcal{H}(\mathcal{L}))(\varphi) = \mathcal{L}(\varphi \circ \mathcal{H})$$

for any germ at $(p', \bar{p}')$ of a holomorphic function $\varphi : \mathbb{C}^{2N'} \to \mathbb{C}$. We say that $\mathcal{H}$ is *Segre submersive* at $(p, \bar{p})$ if

$$\mathcal{D}_{(p,\bar{p})}\mathcal{H}(\mathcal{T}^0_{(p,\bar{p})}\mathcal{M}) = \mathcal{T}^0_{(p',\bar{p}')}\mathcal{M}'.$$

This definition is independent of choice of coordinates for M and M'.

Given M and M' satisfying certain geometric conditions, our main result is Theorem 1.1, which states that the germs at $(p, \bar{p})$ of HSPMs that are Segre submersive at $(p, \bar{p})$ and send $(\mathcal{M}, (p, \bar{p}))$ into $(\mathcal{M}', (p', \bar{p}'))$ rationally depend upon their K-jets, for some fixed K depending only on M and M'. This result was motivated by, and is a generalization of, results due to Baouendi, Ebenfelt, and Rothschild [1999a] and Baouendi, Rothschild, and Zaitsev [2001]. We also mention a recent paper of Lamel and Mir [2007] for related results. Before stating Theorem 1.1, we present some more notation. Let $J^K(\mathbb{C}^N, \mathbb{C}^{N'})_{(p,p')}$ denote the set of K-jets at p of germs of holomorphic maps from $(\mathbb{C}^N, p)$ into $(\mathbb{C}^{N'}, p')$. (In this paper, we assume that $J^K(\mathbb{C}^N, \mathbb{C}^{N'})_{(p,p')}$ includes only derivatives of *positive* order.) Let j_p^K represent the corresponding *K-jet map* defined on the set of germs at p of holomorphic mappings given by

$$j_p^K \phi = \Big(\frac{\partial^{|\alpha|}\phi}{\partial Z^\alpha}(p)\Big)_{1\le|\alpha|\le K}.$$

Theorem 1.1. *Let $M \subseteq \mathbb{C}^N$ be real analytic, generic, and of finite type at p. Let $M' \subseteq \mathbb{C}^{N'}$ be real analytic, generic, and finitely nondegenerate at p'. Assume that M is of codimension d, that M' is of codimension d', and that $N-d \ge N'-d'$. Then there exist positive integers*

$$K \ge 1 \quad \textit{and} \quad 1 \le r \le \binom{N-d}{N'-d'}^2,$$

depending only on M and M', and $\mathbb{C}^{N'}$-valued holomorphic functions $\Phi_1, \dots, \Phi_r$ defined on an open subset of $\mathbb{C}^N \times J^K(\mathbb{C}^N, \mathbb{C}^{N'})_{(p,p')} \times J^K(\mathbb{C}^N, \mathbb{C}^{N'})_{(\bar p,\bar p')}$ of the form

$$\Phi_l(Z,\Lambda,\Gamma) = \sum_\gamma \frac{P^l_\gamma(\Lambda,\Gamma)}{Q^l_1(\Lambda)^{s^l_\gamma}\, Q^l_2(\Gamma)^{t^l_\gamma}}\,(Z-p)^\gamma\,, \tag{1-2}$$

where s^l_γ and t^l_γ are nonnegative integers, P^l_γ are $\mathbb{C}^{N'}$-valued polynomials, and Q^l_1 and Q^l_2 are $\mathbb{C}$-valued polynomials with real coefficients, such that the following holds. Let $\mathcal{H}(Z,\zeta) = (H(Z), \widetilde H(\zeta))$ be a germ at $(p,\bar p)$ of an HSPM sending $(\mathcal{M},(p,\bar p))$ into $(\mathcal{M}',(p',\bar p'))$ such that $\mathcal{H}$ is Segre submersive at $(p,\bar p)$. Then there exists l, with $1\le l\le r$, such that

$$H(Z) = \Phi_l(Z, j_p^K(H), j_{\bar p}^K(\widetilde H)), \tag{1-3}$$

$$\widetilde H(\zeta) = \overline\Phi_l(\zeta, j_{\bar p}^K(\widetilde H), j_p^K(H)), \tag{1-4}$$

for (Z,ζ) sufficiently close to $(p,\bar p)$. Furthermore, for any

$$(\Lambda_0,\Gamma_0) \in J^K(\mathbb{C}^N, \mathbb{C}^{N'})_{(p,p')} \times J^K(\mathbb{C}^N, \mathbb{C}^{N'})_{(\bar p,\bar p')}$$

such that $Q^l_1(\Lambda_0)\ne 0$ and $Q^l_2(\Gamma_0)\ne 0$, the map Φ_l is holomorphic in a neighborhood of (p,Λ_0,Γ_0).

The appearance of $\overline\Phi_l$ in (1-4) is interesting and will be instrumental in the proof of Corollary 1.4. See Section 2 for precise definitions of *finite type* and *finite nondegeneracy*.

Define

$$\begin{aligned}\mathrm{Aut}(M,p) := \big\{H : (\mathbb{C}^N,p)\to(\mathbb{C}^N,p) \,\big|\, &H \text{ is a germ at } p \text{ of a holomorphic map,}\\ &H \text{ is invertible at } p, \text{ and } H(M)\subseteq M\big\},\\ \mathrm{Aut}_{\mathbb{C}}(\mathcal{M},(p,\bar p)) := \big\{\mathcal{H} : (\mathbb{C}^{2N},(p,\bar p))\to(\mathbb{C}^{2N},(p,\bar p)) \,\big|\, &\mathcal{H} \text{ is a germ at } (p,\bar p)\\ &\text{of an HSPM, } \mathcal{H} \text{ is invertible at } (p,\bar p), \text{ and } \mathcal{H}(\mathcal{M})\subseteq\mathcal{M}\big\}.\end{aligned}$$

We call $\mathrm{Aut}(M,p)$ the *group of automorphisms* of M at p, and $\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},(p,\bar p))$ the *group of holomorphic Segre-preserving automorphisms* of $\mathcal{M}$ at $(p,\bar p)$. Let $J_p^K(\mathbb{C}^N) := J^K(\mathbb{C}^N,\mathbb{C}^N)_{(p,p)}$ be a simplification of notation, define $G_p^K(\mathbb{C}^N)$ to be the set of all elements of $J_p^K(\mathbb{C}^N)$ that correspond to invertible mappings at p, and define a jet map $\eta^K_{(p,\bar p)}$ on the set of germs at $(p,\bar p)$ of HSPMs such that for $\mathcal{H}=(H,\widetilde H)$, $\eta^K_{(p,\bar p)}(\mathcal{H}) := (j_p^K H, j_{\bar p}^K \widetilde H)$. Noting that $M=M'$ implies that $r=1$ in Theorem 1.1, we have the following corollary.

Corollary 1.2. *Let $M\subseteq\mathbb{C}^N$ be a generic real analytic submanifold of finite type at p and finitely nondegenerate at p. Then there exists an integer K depending only on M such that $\eta^K_{(p,\bar p)}$ restricted to* $\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},(p,\bar p))$ *is a homeomorphism onto a closed, holomorphic algebraic submanifold (Lie group) of $G_p^K(\mathbb{C}^N)\times G_{\bar p}^K(\mathbb{C}^N)$.*

Remark 1.3. One consequence of Corollary 1.2 is that j_p^K restricted to $\mathrm{Aut}(M,p)$ is a homeomorphism onto a closed, real algebraic submanifold (Lie group) of $G_p^K(\mathbb{C}^N)$. This fact has already been proven in previous work. For M a hypersurface, Baouendi, Ebenfelt, and Rothschild [1997] showed that $j_p^K(\mathrm{Aut}(M,p))$ is a closed, real analytic submanifold (Lie group) of $G_p^K(\mathbb{C}^N)$. However, they did not show that it is also real algebraic. In [1999a], they proved this fact for M a submanifold of any codimension.

Since $j_p^K(\mathrm{Aut}(M,p))$ is a real algebraic submanifold, it is natural to consider its complexification as a holomorphic submanifold of $G_p^K(\mathbb{C}^N)\times G_{\bar p}^K(\mathbb{C}^N)$. We will denote this complexification $\mathbb{C}\{j_p^K(\mathrm{Aut}(M,p))\}$. What is the relationship between $\mathbb{C}\{j_p^K(\mathrm{Aut}(M,p))\}$ and $\eta^K_{(p,\bar p)}(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},(p,\bar p)))$? The following corollary of Theorem 1.1 says that the former is always contained in the latter, and they are necessarily of the same dimension. In Section 5 we will give examples demonstrating both equality and strict containment.

Corollary 1.4. *Let M and K be as in Corollary 1.2. Let $\mathcal{B}\subseteq G_p^K(\mathbb{C}^N)\times G_{\bar p}^K(\mathbb{C}^N)$ denote the connected component of $\mathbb{C}\{j_p^K(\mathrm{Aut}(M,p))\}$ containing* $(\mathrm{Id},\mathrm{Id}')$, *where* Id (*respectively*, Id') *is the point in $G_p^K(\mathbb{C}^N)$ (respectively, $G_{\bar p}^K(\mathbb{C}^N)$) corresponding to the identity map on $\mathbb{C}^N$. Let $\mathcal{C}\subseteq G_p^K(\mathbb{C}^N)\times G_{\bar p}^K(\mathbb{C}^N)$ denote the connected component of $\eta^K_{(p,\bar p)}$*$(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},(p,\bar p)))$ *that contains* $(\mathrm{Id},\mathrm{Id}')$. *Then*

(i) $\mathbb{C}\{j_p^K(\mathrm{Aut}(M,p))\}\subseteq\eta^K_{(p,\bar p)}(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},(p,\bar p)))$,

(ii) $\mathcal{B} = \mathcal{C}$, *and*

(iii) $\eta^K_{(p,\bar{p})}(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M}, (p, \bar{p})))$ *and* $\mathbb{C}\{j^K_p(\mathrm{Aut}(M, p))\}$ *are made up of finitely many disjoint cosets of* $\mathcal{B}$.

The following application of Theorem 1.1 states that Segre submersive HSPMs sending $(\mathcal{M}, (p, \bar{p}))$ into $(\mathcal{M}', (p', \bar{p}'))$ are uniquely determined by a finite number of derivatives.

Theorem 1.5. *Let M and M' be as defined in Theorem 1.1. Then there exists an integer K, depending only on M and M', such that the following holds. Let $\mathcal{H}_1(Z, \zeta) = (H_1(Z), \widetilde{H}_1(\zeta))$ and $\mathcal{H}_2(Z, \zeta) = (H_2(Z), \widetilde{H}_2(\zeta))$ be germs at $(p, \bar{p})$ of HSPMs sending $(\mathcal{M}, (p, \bar{p}))$ into $(\mathcal{M}', (p', \bar{p}'))$ such that $\mathcal{H}_1$ and $\mathcal{H}_2$ are Segre submersive at $(p, \bar{p})$. If*

$$j^K_p(H_1) = j^K_p(H_2) \quad \textit{and} \quad j^K_{\bar{p}}(\widetilde{H}_1) = j^K_{\bar{p}}(\widetilde{H}_2),$$

then $\mathcal{H}_1(Z, \zeta) \equiv \mathcal{H}_2(Z, \zeta)$.

One of the strengths of Theorem 1.1 lies in the fact that the form of Φ_l leads to Corollaries 1.2 and 1.4. These maps, however, depend upon the jets of both H and $\widetilde{H}$. In Theorem 1.6, we see that it is in fact possible to find functions that express $\mathcal{H}$ entirely in terms of the L-jets of H (or of $\widetilde{H}$) for some L.

Theorem 1.6. *Let M and M' be as in Theorem 1.1. Then there exist positive integers r and L, depending only on M and M', and $\mathbb{C}^{2N'}$-valued holomorphic functions $\Phi^1_1, \ldots, \Phi^1_r$ defined on an open subset of $\mathbb{C}^{2N} \times J^L(\mathbb{C}^N, \mathbb{C}^{N'})_{(p,p')}$ and $\Phi^2_1, \ldots, \Phi^2_r$ defined on an open subset of $\mathbb{C}^{2N} \times J^L(\mathbb{C}^N, \mathbb{C}^{N'})_{(\bar{p},\bar{p}')}$ such that the following holds. Let $\mathcal{H}(Z, \zeta) = (H(Z), \widetilde{H}(\zeta))$ be a germ at $(p, \bar{p})$ of an HSPM sending $(\mathcal{M}, (p, \bar{p}))$ into $(\mathcal{M}', (p', \bar{p}'))$ such that $\mathcal{H}$ is Segre submersive at $(p, \bar{p})$. Then there exist $1 \le l_1, l_2 \le r$ such that*

$$\mathcal{H}(Z, \zeta) = \Phi^1_{l_1}(Z, \zeta, j^L_p(H)),$$
$$\mathcal{H}(Z, \zeta) = \Phi^2_{l_2}(Z, \zeta, j^L_{\bar{p}}(\widetilde{H}))$$

for (Z, ζ) sufficiently close to $(p, \bar{p})$.

The following example shows that Theorem 1.6 does not necessarily hold if M' is finitely degenerate at p'.

Example 1.7. Let $M = M' \subseteq \mathbb{C}^2$ be given by $M = \{\mathrm{Im}\, w = |z|^4\}$ and its complexification by $\mathcal{M} = \{w - \tau = 2iz^2\chi^2\}$, where (z, w) and (χ, τ) are coordinates on $\mathbb{C}^2$. We note that M is of finite type but finitely *degenerate* at 0. Let $H(z, w) = (z, w)$. We can find two *distinct* maps $\widetilde{H}_1(\chi, \tau)$ and $\widetilde{H}_2(\chi, \tau)$ such that $\mathcal{H}_1 = (H, \widetilde{H}_1)$ and $\mathcal{H}_2 = (H, \widetilde{H}_2)$ both send $(\mathcal{M}, 0)$ into $(\mathcal{M}', 0)$ and are Segre submersive at 0. Indeed, let $\widetilde{H}_1(\chi, \tau) = (\chi, \tau)$ and let $\widetilde{H}_2(\chi, \tau) = (-\chi, \tau)$.

Finally, we present a result on algebraicity. Recall that a real analytic mapping is said to be real analytic algebraic if all of its components are real analytic algebraic, and a real analytic submanifold is said to be real algebraic if it can be given by real analytic algebraic defining functions; similarly a holomorphic map is said to be holomorphic algebraic if all of its components are holomorphic algebraic, and a holomorphic submanifold is said to be holomorphic algebraic if it can be given by holomorphic algebraic defining functions.

Theorem 1.8. *Let M and M' be as in Theorem 1.1, and assume that M and M' are real algebraic. Then any germ at $(p', \bar{p}')$ of an HSPM sending $(\mathcal{M}, (p, \bar{p}))$ into $(\mathcal{M}', (p', \bar{p}'))$ that is Segre submersive at $(p, \bar{p})$ is holomorphic algebraic.*

The layout of this paper is as follows. In Section 2, we present some additional background material. Section 3 contains the reformulations and proofs of three of the main results as given in Section 1, while Section 4 is dedicated to proving the main results of Section 1. Section 5 consists of several examples of HSPMs and automorphism groups. In particular, examples demonstrating both equality and nonequality of $\mathbb{C}\{j_p^K(\mathrm{Aut}(M, p))\}$ and $\eta^K_{(p,\bar{p})}(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M}, (p, \bar{p})))$ are provided. See [Angle 2008b] for additional examples.

2. Additional background

Let $M \subseteq \mathbb{C}^N$ be a real analytic submanifold of codimension d. Recall that this means that given any $p \in M$, there exists a real vector-valued real analytic function $\rho = (\rho_1, \ldots, \rho_d)$ defined in a neighborhood of p, satisfying $d\rho_1 \wedge \cdots \wedge d\rho_d \neq 0$ at p, such that M is given locally near p by the vanishing of ρ. We refer to ρ as the local *defining function* for M near p. If, in addition, near any $p \in M$, there exists a local defining function ρ satisfying the stronger condition $\partial_Z \rho_1 \wedge \cdots \wedge \partial_Z \rho_d \neq 0$ at p, then we say that M is *generic*. If M is generic, it can be shown (see for example [Baouendi et al. 1999b]) that there exists a holomorphic change of coordinates $Z = (z, w) \in \mathbb{C}^{N-d} \times \mathbb{C}^d$ vanishing at p and an open neighborhood Ω of 0 such that in these coordinates M is locally given by $\{(z, w) \in \Omega : w = Q(z, \bar{z}, \bar{w})\}$, where $Q(z, \chi, \tau) = (Q^1(z, \chi, \tau), \ldots, Q^d(z, \chi, \tau))$ is a $\mathbb{C}^d$-valued holomorphic function defined near 0 in $\mathbb{C}^{N-d} \times \mathbb{C}^{N-d} \times \mathbb{C}^d$ and satisfying $Q(0, \chi, \tau) \equiv Q(z, 0, \tau) \equiv \tau$. Such coordinates are called *normal coordinates*.

A vector field of the form $\sum_{j=1}^N a_j(Z, \bar{Z})(\partial/\partial \bar{Z}_j)$ tangent to M near p, where the a_j are smooth functions on M, is called a *CR vector field*. We say that M is of *finite type* at p (in the sense of [Kohn 1972] and [Bloom and Graham 1977]) if the CR vector fields, their complex conjugates, and all repeated commutators of these vector fields span the complexified tangent space of M at p. Letting $(\rho_j)_Z := (\partial \rho_j/\partial Z_1, \ldots, \partial \rho_j/\partial Z_N)$ and $L^\alpha := L_1^{\alpha_1} \cdots L_m^{\alpha_m}$, where $\alpha = (\alpha_1, \ldots, \alpha_m)$ and $L_1, \ldots, L_m$ is a basis for the CR vector fields of M near p, we say that M is *finitely nondegenerate* at p if there exists a nonnegative integer K such that

$$\operatorname{span}\{L^{\alpha}(\rho_j)_Z(p) : |\alpha| \le K, 1 \le j \le d\} = \mathbb{C}^N. \tag{2-1}$$

We say that M is *k-nondegenerate* at p if k is the smallest K for which (2-1) holds. It is not difficult to show that if M is given in normal coordinates by $w = Q(z, \bar{z}, \overline{w})$, then M is k-nondegenerate at 0 if and only if the matrix whose rows are $(Q^j_{z_l\chi^\alpha}(0,0,0))_{1\le l\le N-d}$ for $|\alpha| \le K$ and $1 \le j \le d$ has rank $N-d$ for $K \ge k$ and rank less than $N-d$ for $K < k$.

Let $M \subseteq \mathbb{C}^N$ be the generic real analytic submanifold $\{Z \in \Omega : \rho(Z, \bar{Z}) = 0\}$, where Ω is a nonempty open subset of $\mathbb{C}^N$, $\rho(Z,\zeta)$ is holomorphic on $\Omega \times {}^*\Omega$, and $\partial_Z\rho(Z,\zeta) \ne 0$ for all $(Z,\zeta) \in \Omega \times {}^*\Omega$. The complexification of M is then given by $\mathcal{M} = \{(Z,\zeta) \in \Omega \times {}^*\Omega : \rho(Z,\zeta) = 0\}$. Given any $(Z,\zeta) \in \Omega \times {}^*\Omega$, we define the *Segre varieties* of M by

$$\Sigma_Z := \{\zeta \in {}^*\Omega : \rho(Z,\zeta) = 0\} \quad \text{and} \quad \widehat{\Sigma}_\zeta := \{Z \in \Omega : \rho(Z,\zeta) = 0\}.$$

Segre varieties are named for the Italian geometer Beniamino Segre who introduced them [1931]. We note here that $\mathcal{M}$ is sometimes referred to as the *Segre family* associated with M; see [Chern 1975; Faran 1980], for example.

Now let $M' \subseteq \mathbb{C}^{n+1}$ be the generic real analytic submanifold given by $M' = \{(Z' \in \Omega' : \rho'(Z', \bar{Z}') = 0\}$, where Ω' is a nonempty open subset of $\mathbb{C}^{n+1}$, $\rho'(Z',\zeta')$ is holomorphic on $\Omega' \times {}^*\Omega'$, and $\partial_{Z'}\rho'(Z',\zeta') \ne 0$ for all $(Z',\zeta') \in \Omega' \times {}^*\Omega'$. Denote the complexification of M' by $\mathcal{M}'$ and its Segre varieties by $\Sigma'_{Z'}$ and $\widehat{\Sigma}'_{\zeta'}$. Let $p \in M$ and $p' \in M'$, and let $\psi : \Omega \times {}^*\Omega \to \mathbb{C}^{2n+2}$ be a holomorphic map sending $(\mathcal{M}, (p, \bar{p}))$ into $(\mathcal{M}', (p', \bar{p}'))$. Furthermore, we will assume that for any $(Z,\zeta) \in \mathcal{M}$, there exists $(Z',\zeta') \in \mathcal{M}'$ such that

$$\psi(\{Z\} \times \Sigma_Z) \subseteq \{Z'\} \times \Sigma'_{Z'}, \tag{2-2}$$

$$\psi(\widehat{\Sigma}_\zeta \times \{\zeta\}) \subseteq \widehat{\Sigma}'_{\zeta'} \times \{\zeta'\}. \tag{2-3}$$

Proposition 2.1. *The map ψ, when restricted to $\mathcal{M}$, is an HSPM of the form* (1-1).

Faran [1980] proved this for hypersurfaces, but it also holds in higher codimension.

Proof. We show it is true in a neighborhood of $(Z,\zeta) = (Z_0, \zeta_0) \in \mathcal{M}$. Write $\psi(Z,\zeta) = (\phi_1(Z,\zeta), \phi_2(Z,\zeta))$, where ϕ_1 and ϕ_2 are $\mathbb{C}^{N'}$-valued holomorphic functions, and write $\zeta_0 = (\zeta_0', \zeta_0'') \in \mathbb{C}^{N-d} \times \mathbb{C}^d$. As M is generic, the implicit function theorem implies that (after a possible rearrangement of coordinates) there exists a $\mathbb{C}^d$-valued holomorphic function θ, satisfying $\theta(Z_0, \zeta_0') = \zeta_0''$, such that $(Z, \zeta_0', \theta(Z, \zeta_0')) \in \mathcal{M}$ for any Z sufficiently close to Z_0. For any Z near Z_0, define $H(Z) := \phi_1(Z, \zeta_0', \theta(Z, \zeta_0'))$. We claim that $H(Z) = \phi_1(Z,\zeta)$ on $\mathcal{M}$. This is because (2-2) implies that $\phi_1(p_0, \zeta)$ is constant for all $\zeta \in \Sigma_{p_0}$ for any p_0. A similar argument applies to ϕ_2. □

3. Reformulations

In this section, we will assume unless said otherwise that $M \subseteq \mathbb{C}^{m+d}$ and $M' \subseteq \mathbb{C}^{n+e}$ are real analytic generic submanifolds of codimensions d and e, respectively. We will additionally assume that there are open neighborhoods $0 \in U \subseteq \mathbb{C}^{m+d}$ and $0 \in U' \subseteq \mathbb{C}^{n+e}$ such that M is given by

$$M = \{(z, w) \in U : w = Q(z, \bar{z}, \overline{w})\},$$

where $Z = (z, w)$ are normal coordinates, and M' is given by

$$M' = \{(z', w') \in U' : w' = Q'(z', \bar{z}', \overline{w}')\},$$

where $Z' = (z', w')$ are normal coordinates. We also assume that $\rho(z, w, \chi, \tau) := w - Q(z, \chi, \tau)$ is holomorphic on $U \times {}^*U$ and $\partial_Z \rho(z, w, \chi, \tau)$ is nonvanishing on $U \times {}^*U$. Similarly, we assume that $\rho'(z', w', \chi', \tau') := w' - Q'(z', \chi', \tau')$ is holomorphic on $U' \times {}^*U'$ and $\partial_{Z'} \rho'(z', w', \chi', \tau')$ is nonvanishing on $U' \times {}^*U'$. So the complexifications $\mathcal{M}$ and $\mathcal{M}'$ of M and M' are respectively given by

$$\mathcal{M} = \{(z, w, \chi, \tau) \in U \times {}^*U : w = Q(z, \chi, \tau)\}$$

and

$$\mathcal{M}' = \{(z', w', \chi', \tau') \in U' \times {}^*U' : w' = Q'(z', \chi', \tau')\},$$

where $\zeta = (\chi, \tau) \in \mathbb{C}^m \times \mathbb{C}^d$ and $\zeta' = (\chi', \tau') \in \mathbb{C}^n \times \mathbb{C}^e$. Unless otherwise specified, we will assume any (germ of an) HSPM $\mathcal{H}$ sends $(\mathcal{M}, 0)$ into $(\mathcal{M}', 0)$ and is given in the form

$$\mathcal{H}(Z, \zeta) = (H(Z), \widetilde{H}(\zeta)) = (f(Z), g(Z), \tilde{f}(\zeta), \tilde{g}(\zeta)), \tag{3-1}$$

where $f = (f^1, \dots, f^n)$ and $\tilde{f} = (\tilde{f}^1, \dots, \tilde{f}^n)$ are $\mathbb{C}^n$-valued holomorphic functions, $g = (g^1, \dots, g^e)$ and $\tilde{g} = (\tilde{g}^1, \dots, \tilde{g}^e)$ are $\mathbb{C}^e$-valued holomorphic functions, and we write $z = (z_1, \dots, z_m)$, $w = (w_1, \dots, w_d)$, $z' = (z'_1, \dots, z'_n)$, and $w' = (w'_1, \dots, w'_e)$, and similarly for χ, τ, χ', and τ'.

Reformulation of Theorem 1.1. We begin with a technical definition.

Definition 3.1. Let $M \subseteq \mathbb{C}^{m+d}$ be of codimension d and $M' \subseteq \mathbb{C}^{n+e}$ be of codimension e, and assume $m \geq n$. Let $\mathcal{H}$ be an HSPM. Let $\mu = (\mu_1, \dots, \mu_n)$ for some $1 \leq \mu_1 < \dots < \mu_n \leq m$ and $\nu = (\nu_1, \dots, \nu_n)$ for some $1 \leq \nu_1 < \dots < \nu_n \leq m$, and assume that

$$\det\left(\frac{\partial f^k}{\partial z_{\mu_l}}(0)\right)_{1\leq k,l\leq n} \neq 0 \quad \text{and} \quad \det\left(\frac{\partial \tilde{f}^k}{\partial \chi_{\nu_l}}(0)\right)_{1\leq k,l\leq n} \neq 0.$$

Then we say that the map $\mathcal{H}$ satisfies *condition $D_{\mu\nu}$*.

Any given $\mathscr{H}$ may satisfy condition $D_{\mu\nu}$ for several different μ and ν, as the following example illustrates.

Example 3.2. Let $M \subseteq \mathbb{C}^4$ and $M' \subseteq \mathbb{C}^3$ be given by

$$M = \{\operatorname{Im} w = |z_1|^2 + 2\operatorname{Re}(z_3\bar{z}_1 - z_3\bar{z}_2) - |z_2|^2\},$$
$$M' = \{\operatorname{Im} w' = |z_1'|^2 + |z_2'|^2\}.$$

Note that M is of finite type at 0, and M' is finitely nondegenerate at 0. Let $\mathscr{H}$ be given by $\mathscr{H}(z, w, \chi, \tau) = (z_1+z_3, z_1-z_2, w, \chi_1-\chi_2, \chi_2+\chi_3, \tau)$. Then $\mathscr{H}$ satisfies condition $D_{\mu\nu}$ for any permissible μ and ν. That is, μ can be any one of $(1, 2)$, $(1, 3)$, or $(2, 3)$, as can ν.

Our main theorem, from which Theorem 1.1 follows, is Theorem 3.3. Before we present it, we introduce some notation. Given an HSPM $\mathscr{H}$, we can write

$$j_0^K H = ((f_{z_l}^j(0))_{1\le l\le m, 1\le j\le n}, (j_0^K)' H),$$

where $(j_0^K)'H$ represents the remaining derivatives of H at 0. Given any Λ in $J_0^K(\mathbb{C}^{m+d}, \mathbb{C}^{n+e})_{(0,0)}$, we will then write

$$\Lambda = ((\Lambda^{j,l})_{1\le l\le m, 1\le j\le n}, \Lambda'), \tag{3-2}$$

where $(j_0^K H)^{j,l}$ is exactly $f_{z_l}^j(0)$. We define a similar decomposition for $j_0^K \widetilde{H}$.

Theorem 3.3. *Let $M \subseteq \mathbb{C}^{m+d}$ be of codimension d and of finite type at 0. Let $M' \subseteq \mathbb{C}^{n+e}$ be of codimension e and k-nondegenerate at 0. Then there exists a positive integer K depending only on M and M' such that for each $\alpha = (\alpha_1, \dots, \alpha_n)$ with $1 \le \alpha_1 < \cdots < \alpha_n \le m$ and each $\beta = (\beta_1, \dots, \beta_n)$ with $1 \le \beta_1 < \cdots < \beta_n \le m$, there exists a $\mathbb{C}^{n+e}$-valued holomorphic function defined on an open subset of $\mathbb{C}^{m+d} \times J^K(\mathbb{C}^{m+d}, \mathbb{C}^{n+e})_{(0,0)} \times J^K(\mathbb{C}^{m+d}, \mathbb{C}^{n+e})_{(0,0)}$ of the form*

$$\Phi^{\alpha,\beta}(Z, \Lambda, \Gamma) = \sum_{\gamma} \frac{R_\gamma^{\alpha,\beta}(\Lambda, \Gamma)}{(\det(\Lambda^{r,\alpha_j})_{1\le r,j\le n})^{s_{\alpha\beta\gamma}} (\det(\Gamma^{r,\beta_j})_{1\le r,j\le n})^{t_{\alpha\beta\gamma}}} Z^\gamma, \tag{3-3}$$

where $R_\gamma^{\alpha,\beta}$ are $\mathbb{C}^{n+e}$-valued polynomials and $s_{\alpha\beta\gamma}$ and $t_{\alpha\beta\gamma}$ are nonnegative integers, such that if $\mathscr{H}(Z, \zeta) = (H(Z), \widetilde{H}(\zeta))$ is a germ at 0 of an HSPM satisfying condition $D_{\mu\nu}$, then

$$H(Z) = \Phi^{\mu,\nu}(Z, j_0^K(H), j_0^K(\widetilde{H})), \tag{3-4}$$

$$\widetilde{H}(\zeta) = \overline{\Phi^{\nu,\mu}}(\zeta, j_0^K(\widetilde{H}), j_0^K(H)), \tag{3-5}$$

for (Z, ζ) sufficiently close to 0. Furthermore, for any (Λ_0, Γ_0) such that

$$\det(\Lambda_0^{r,\alpha_j})_{1\le r,j\le n} \ne 0 \quad \text{and} \quad \det(\Gamma_0^{r,\beta_j})_{1\le r,j\le n} \ne 0,$$

the function $\Phi^{\alpha,\beta}$ is holomorphic in a neighborhood of $(0, \Lambda_0, \Gamma_0)$.

Remark 3.4. It is implicit in the hypotheses of Theorem 3.3 that $m \geq n$. However, if we assume that $m < n$, even if the matrices $(f_z(0)) := (f^j_{z_l}(0))_{1\leq l\leq m, 1\leq j\leq n}$ and $(\tilde{f}_\chi(0)) := (\tilde{f}^j_{\chi_l}(0))_{1\leq l\leq m, 1\leq j\leq n}$ have maximal rank, the theorem will not hold. Let $M \subseteq \mathbb{C}^4$ be defined by $M = \{\operatorname{Im} w_1 = |z_1|^2, \operatorname{Im} w_2 = |z_2|^2\}$. Let $M' \subseteq \mathbb{C}^4$ be defined by $M' = \{\operatorname{Im} w' = |z'_1|^2 + |z'_2|^2 + |z'_3|^2\}$. Then M is of finite type at 0, and M' is 1-nondegenerate at 0. For any positive integer r, define

$$\mathcal{H}_r(z, w, \chi, \tau) = \big(z_1, z_2, w_1, w_1 + w_2, \chi_1 - 2i\chi_1\tau_1 - 2i\chi_1\tau_1^r, \chi_2, \tau_1^r + \tau_1, \tau_1 + \tau_2 - 2i\tau_1^2 - 2i\tau_1^{r+1}\big).$$

Observe that $\mathcal{H}_r$ is an HSPM sending $(\mathcal{M}, 0)$ into $(\mathcal{M}', 0)$ and is a biholomorphism near 0.

The proof of Theorem 3.3 will be based on arguments from [Baouendi et al. 1999a; 2001]. Before proving the theorem, we first introduce a few lemmas.

Lemma 3.5. *Let $\mathcal{H}(Z, \zeta) = (H(Z), \tilde{H}(\zeta))$ be an HSPM that sends $(\mathcal{M}, 0)$ into $(\mathcal{M}', 0)$. Then $\mathcal{H}'(Z, \zeta) = (\bar{\tilde{H}}(Z), \bar{H}(\zeta))$ is an HSPM sending $(\mathcal{M}, 0)$ into $(\mathcal{M}', 0)$.*

Proof. Let $\rho = (\rho_1, \dots, \rho_d)$ be a defining function for M, and let $\rho' = (\rho'_1, \dots, \rho'_e)$ be a defining function for M'. For $j = 1, \dots, e$ and $k = 1, \dots, d$, there exist holomorphic functions a^j_k such that

$$\rho'_j(H(Z), \tilde{H}(\zeta)) = \sum_{k=1}^d a^j_k(Z, \zeta)\rho_k(Z, \zeta) \quad \text{implies} \tag{3-6}$$

$$\bar{\rho}'_j(\tilde{H}(\zeta), H(Z)) = \sum_{k=1}^d a^j_k(Z, \zeta)\rho_k(Z, \zeta) \quad \text{implies} \tag{3-7}$$

$$\rho'_j(\bar{\tilde{H}}(Z), \bar{H}(\zeta)) = \sum_{k=1}^d \bar{a}^j_k(\zeta, Z)\bar{\rho}_k(\zeta, Z) = \sum_{k=1}^d \bar{a}^j_k(\zeta, Z)\rho_k(Z, \zeta). \tag{3-8}$$

Equations (3-7) and (3-8) follow from the reality of the ρ_j. □

The following notation will be used in Lemmas 3.6, 3.7, and 3.10. Let M, M', and $\mathcal{H}$ be as in Theorem 3.3. We will write $j^K_Z H = ((j^K_Z)'' H, (g_{z^\alpha}(Z))_{|\alpha|\leq K})$, where the remaining derivatives of H at Z are represented by $(j^K_Z)'' H$. Given any $\Lambda \in J^K_Z(\mathbb{C}^{m+d}, \mathbb{C}^{n+e})_{(Z, H(Z))}$, we will also write

$$\Lambda = (\Lambda_1, \Lambda_2), \tag{3-9}$$

where $(j^K_Z H)_2$ is exactly $(g_{z^\alpha}(Z))_{|\alpha|\leq K}$. We do a similar decomposition for $j^K_\zeta \tilde{H}$.

Lemma 3.6. *Let M and M' be as in Theorem 3.3. Then for any $\beta = (\beta_1, \dots, \beta_n)$ and $\alpha = (\alpha_1, \dots, \alpha_n)$, with $1 \leq \alpha_1 < \cdots < \alpha_n \leq m$, there exists a $\mathbb{C}^e$-valued* *holomorphic function ϕ^α_β defined on an open subset of $\mathbb{C}^{K_\beta} \times \mathbb{C}^{m+d} \times \mathbb{C}^{m+d}$, for some integer K_β, of the form*

$$\phi^\alpha_\beta(\Lambda, Z, \zeta) = \sum_{\gamma,\delta,\kappa} \frac{P^{\alpha,\beta}_{\gamma,\delta,\kappa}(\Lambda_1)}{(\det(\Lambda^{j,\alpha_l})_{1\le j,l\le n})^{t_{\alpha\beta\gamma\delta\kappa}}} Z^\gamma \zeta^\delta \Lambda_2^\kappa, \tag{3-10}$$

where $t_{\alpha\beta\gamma\delta\kappa}$ are nonnegative integers, $P^{\alpha,\beta}_{\gamma,\delta,\kappa}$ are $\mathbb{C}^e$-valued polynomials, and ϕ^α_β identically vanishes whenever $\zeta = 0$ and $\Lambda_2 = 0$, such that if $\mathcal{H}$ is a germ at 0 of an HSPM satisfying condition $D_{\mu\nu}$, then for $(Z, \zeta) \in \mathcal{M}$ sufficiently close to 0,

$$Q'_{z'^\beta}(f(Z), \tilde{f}(\zeta), \tilde{g}(\zeta)) = \phi^\mu_\beta(j^{|\beta|}_Z(H), Z, \zeta), \tag{3-11}$$

$$\overline{Q}'_{\chi'^\beta}(\tilde{f}(\zeta), f(Z), g(Z)) = \overline{\phi^\nu_\beta}(j^{|\beta|}_\zeta(\widetilde{H}), \zeta, Z). \tag{3-12}$$

Also ϕ^α_β is holomorphic near $(\Lambda_0, 0, 0)$ for any Λ_0 such that $\det(\Lambda_0^{j,\alpha_l})_{1\le j,l\le n} \neq 0$.

Proof. For $j = 1, \ldots, m$,

$$L_j = \frac{\partial}{\partial z_j} + \sum_{r=1}^{d} Q^r_{z_j}(z, \chi, \tau)\frac{\partial}{\partial w_r} \tag{3-13}$$

are vector fields tangent to $\mathcal{M}$. Let $\hat{z} := (z_{\mu_1}, \ldots, z_{\mu_n})$. Applying $L_{\mu_1}, \ldots, L_{\mu_n}$ to

$$g(z, w) = Q'\big(f(z, w), \widetilde{H}(\chi, \tau)\big) \tag{3-14}$$

we get (in matrix notation)

$$\begin{aligned} &g_{\hat{z}}(z, w) + Q_{\hat{z}}(z, \chi, \tau) g_w(z, w) \\ &\qquad = (f_{\hat{z}}(z, w) + Q_{\hat{z}}(z, \chi, \tau) f_w(z, w)) Q'_{z'}(f(z, w), \widetilde{H}(\chi, \tau)) \end{aligned} \tag{3-15}$$

for all $(z, w, \chi, \tau) \in \mathcal{M}$. By assumption, $(f_{\hat{z}}(0))$ is invertible, so we have near $(z, w, \chi, \tau) = (0, 0, 0, 0)$,

$$\begin{aligned} &Q'_{z'}(f(z, w), \widetilde{H}(\chi, \tau)) \\ &\qquad = (f_{\hat{z}}(z, w) + Q_{\hat{z}}(z, \chi, \tau) f_w(z, w))^{-1} (g_{\hat{z}}(z, w) + Q_{\hat{z}}(z, \chi, \tau) g_w(z, w)). \end{aligned} \tag{3-16}$$

We claim that the right hand side of (3-16) can be written in the form

$$\sum_{\gamma,\delta} \frac{p^\mu_{\gamma,\delta}(j^1_Z(H))}{(\det(f_{\hat{z}}(Z)))^{s_{\mu\gamma\delta}}} Z^\gamma \zeta^\delta, \tag{3-17}$$

where each $p^\mu_{\gamma,\delta}$ is an $n \times e$ polynomial matrix and each $s_{\mu\gamma\delta}$ is a nonnegative integer. This comes from writing the right hand side as

$$(f_{\hat{z}} + Q_{\hat{z}} f_w)^{-1}(g_{\hat{z}} + Q_{\hat{z}} g_w) = (I + f_{\hat{z}}^{-1} f_w Q_{\hat{z}})^{-1}(f_{\hat{z}}^{-1})(g_{\hat{z}} + Q_{\hat{z}} g_w) \tag{3-18}$$

The last factor on the right of (3-18) can clearly be written in the form (3-17), as it is independent of $\det(f_{\hat{z}}(Z))$. The second can be written in the form (3-17) since for any invertible matrix A, we can write A^{-1} as $(1/\det A)(\operatorname{adj} A)$. The first can also be written in the form (3-17). Indeed, since $f_{\hat{z}}^{-1}(0) f_w(0) Q_{\hat{z}}(0) = 0$, then for (z, χ, τ) sufficiently close to 0, we have $(I + B)^{-1} = \sum_{j=0}^{\infty}(-1)^j B^j$, where we define $B := f_{\hat{z}}^{-1} f_w Q_{\hat{z}}$. We then use the aforementioned formula for the inverse of a matrix, and the claim is proved. Defining $j_Z^1 H = ((j_Z^1)' H, (j_Z^1)'' H)$, where $(j_Z^1)'' H = (g_{z_j}(Z))$ and $(j_Z^1)' H$ represents the remaining derivatives at Z, it immediately follows that (3-17) can be written in the form

$$\sum_{\gamma,\delta,\kappa} \frac{p^{\mu}_{\gamma,\delta,\kappa}((j_Z^1)' H)}{(\det(f_{\hat{z}}(Z)))^{s_{\mu\gamma\delta\kappa}}} Z^{\gamma} \zeta^{\delta} ((j_Z^1)'' H)^{\kappa}. \tag{3-19}$$

We get (3-11) from (3-16) and (3-19) by inductively applying the L_j and using the chain rule. To complete the proof of the lemma, we use Lemma 3.5 to see that $(\overline{\tilde{H}}, \overline{H})$ sends $\mathcal{M}$ into $\mathcal{M}'$ and satisfies condition $D_{\nu\mu}$. So as we have seen in this proof,

$$Q'_{z'^{\beta}}(\overline{\tilde{f}}(Z), \bar{f}(\zeta), \bar{g}(\zeta)) = \phi^{\nu}_{\beta}(j_Z^{|\beta|}(\overline{\tilde{H}}), Z, \zeta). \tag{3-20}$$

Taking the complex conjugate of this entire equation gives (3-12).

The fact that $\phi^{\alpha}_{\beta} \equiv 0$ whenever $\zeta = 0$ and $\Lambda_2 = 0$ follows from (3-16), the definition of the L_j given in (3-13), and the fact that $Q_{z^{\beta}}(z, 0, 0) \equiv 0$ for any β. □

Lemma 3.7. *Let M and M' be as in Theorem 3.3. For each $\beta = (\beta_1, \dots, \beta_{n+e})$ and $\alpha = (\alpha_1, \dots, \alpha_n)$ with $1 \le \alpha_1 < \cdots < \alpha_n \le m$, there exists a $\mathbb{C}^{n+e}$-valued function Ψ^{α}_{β}, holomorphic on an open subset of $\mathbb{C}^{m+d} \times \mathbb{C}^{m+d} \times \mathbb{C}^{K_\beta}$ for some integer K_β, of the form*

$$\Psi^{\alpha}_{\beta}(Z, \zeta, \Lambda) = \sum_{\gamma,\delta,\kappa} \frac{P^{\alpha,\beta}_{\gamma,\delta,\kappa}(\Lambda_1)}{(\det(\Lambda^{r,\alpha_l})_{1\le l,r\le n})^{t_{\alpha\beta\gamma\delta\kappa}}} Z^{\gamma} \zeta^{\delta} \Lambda_2^{\kappa}, \tag{3-21}$$

where $P^{\alpha,\beta}_{\gamma,\delta,\kappa}(\Lambda_1)$ are $\mathbb{C}^{n+e}$-valued polynomials and $t_{\alpha\beta\gamma\delta\kappa}$ are nonnegative integers, such that if $\mathcal{H}$ is a germ at 0 of an HSPM satisfying condition $D_{\mu\nu}$, then for $(Z, \zeta) \in \mathcal{M}$ sufficiently close to 0,

$$\partial^{\beta} H(Z) = \Psi^{\nu}_{\beta}(Z, \zeta, j_{\zeta}^{k+|\beta|}(\tilde{H})), \tag{3-22}$$

$$\partial^{\beta} \tilde{H}(\zeta) = \overline{\Psi^{\mu}_{\beta}}(\zeta, Z, j_Z^{k+|\beta|}(H)). \tag{3-23}$$

Also, Ψ^{α}_{β} is holomorphic near $(0, 0, \Lambda_0)$ for any Λ_0 such that $\det(\Lambda_0^{r,\alpha_l})_{1\le l,r\le n} \ne 0$.

Proof. As M' is k-nondegenerate at 0, assume the vectors $\overline{Q}'^{j_1}_{z\chi^{\alpha_1}}(0), \dots, \overline{Q}'^{j_n}_{z\chi^{\alpha_n}}(0)$ span $\mathbb{C}^n$, where each $j_k \in \{1, \dots, e\}$, each $|\alpha_j| \le k$, and $\overline{Q}' = (\overline{Q}'^1, \dots, \overline{Q}'^e)$. From

Lemma 3.6, we have for each $(Z,\zeta)\in\mathcal{M}$

$$\begin{aligned}\overline{Q}'^{j_1}_{\chi'^{\alpha_1}}(\tilde f(\zeta), f(Z), g(Z)) &= \overline{(\phi^{\nu}_{\alpha_1})^{j_1}}(j^{|\alpha_1|}_{\zeta}(\widetilde H),\zeta,Z)\\ &\vdots \\ \overline{Q}'^{j_n}_{\chi'^{\alpha_n}}(\tilde f(\zeta), f(Z), g(Z)) &= \overline{(\phi^{\nu}_{\alpha_n})^{j_n}}(j^{|\alpha_n|}_{\zeta}(\widetilde H),\zeta,Z),\end{aligned} \tag{3-24}$$

where $\phi^{\beta}_{\alpha}=((\phi^{\beta}_{\alpha})^1,\dots,(\phi^{\beta}_{\alpha})^e)$. Using this system of equations, coupled with the fact that normal coordinates for M' imply that $\overline{Q}'(\chi',0,w')\equiv\overline{Q}'(0,z',w')\equiv w'$, we can apply the implicit function theorem to find a map $B^{\nu}:\mathbb{C}^n\times\mathbb{C}^n\to\mathbb{C}^{n+e}$, holomorphic near 0, such that

$$H(Z)=B^{\nu}\big(\tilde f(\zeta),(\overline{(\phi^{\nu}_{\alpha_l})^{j_l}}(j^{|\alpha_l|}_{\zeta}\widetilde H,\zeta,Z))_{1\le l\le n}\big). \tag{3-25}$$

If we Taylor expand, we can write the right hand side of (3-25) as

$$\sum_{\alpha,\beta,\gamma} A^{\nu}_{\alpha\beta\gamma}(\widetilde\Lambda_1)Z^{\alpha}\zeta^{\beta}\widetilde\Lambda_2^{\gamma}, \tag{3-26}$$

where we remind the reader that $\widetilde\Lambda_2$ corresponds to $(\tilde g_{\chi^{\alpha}}(\zeta))$, and $\widetilde\Lambda_1$ corresponds to the remaining derivatives of $\widetilde H$ at ζ. We claim that each $A^{\nu}_{\alpha\beta\gamma}$ is rational. This follows from the fact that $\tilde f(0)=0$, the form of ϕ^{α}_{β} as given in (3-10), and the fact that $\phi^{\alpha}_{\beta}\equiv 0$ whenever $\zeta=0$ and $\Lambda_2=0$ (refer to the statement of Lemma 3.6).

Furthermore, each $A^{\nu}_{\alpha\beta\gamma}(\widetilde\Lambda_1)$ is of the form given in the right hand side of (3-21). This can be seen by Taylor expanding B^{ν} as given in (3-25) and plugging in (3-10). Define

$$\Psi^{\nu}_0(Z,\zeta,\widetilde\Lambda):=\sum_{\alpha,\beta,\gamma} A^{\nu}_{\alpha\beta\gamma}(\widetilde\Lambda_1)Z^{\alpha}\zeta^{\beta}\widetilde\Lambda_2^{\gamma}. \tag{3-27}$$

This proves (3-22) for $|\beta|=0$. For $|\beta|>0$, as every point in $\mathcal{M}$ is of the form $(z,w,\chi,\overline{Q}(\chi,z,w))$, we have

$$H(z,w)\equiv\Psi^{\nu}_0\big(z,w,\chi,\overline{Q}(\chi,z,w),j^k_{(\chi,\overline{Q}(\chi,z,w))}(\widetilde H)\big). \tag{3-28}$$

We inductively differentiate (3-28), applying the chain rule, and (3-22) follows.

To get (3-23), we know from Lemma 3.5 that $(\overline{\widetilde H},\overline H)$ sends $\mathcal{M}$ into $\mathcal{M}'$ and satisfies condition $D_{\nu\mu}$. So

$$\partial^{\beta}\overline{\widetilde H}(Z)=\Psi^{\mu}_{\beta}\big(Z,\zeta,j^{k+|\beta|}_{\zeta}(\overline H)\big). \tag{3-29}$$

Taking the complex conjugate of both sides of this equation, the lemma follows. □

Now we define the *r-th Segre mappings* of M at 0. These mappings were first introduced by Baouendi, Ebenfelt, and Rothschild [1996] and will prove extremely useful in completing the proof of Theorem 3.3. Given a positive integer r, let $t^0, \dots, t^{r-1} \in \mathbb{C}^m$ and define $v^r : \mathbb{C}^{rm} \to \mathbb{C}^{m+d}$ by

$$v^r(t^0, \dots, t^{r-1}) := (t^0, u^r(t^0, \dots, t^{r-1})), \tag{3-30}$$

where $u^r : \mathbb{C}^{rm} \to \mathbb{C}^d$ is given inductively by

$$u^1(t^0) = 0, \quad u^r(t^0, \dots, t^{r-1}) = Q(t^0, t^1, \overline{u^{r-1}}(t^1, \dots, t^{r-1})) \text{ for } r \geq 2. \tag{3-31}$$

Definition 3.8. Let V and W be finite-dimensional complex vector spaces. Let $\mathcal{R}_0(V \times W, V)$ denote the ring of germs of holomorphic functions f at $V \times \{0\}$ in $V \times W$ which can be written in the form $f(\Lambda, \Gamma) = \sum_\alpha p_\alpha(\Lambda)\Gamma^\alpha$, where each $p_\alpha(\Lambda)$ is a polynomial function on V.

The following lemma is proved in [Baouendi et al. 2001]:

Lemma 3.9. *Let $V_0, V_1, \widetilde{V}_0, \widetilde{V}_1$ be finite-dimensional complex vector spaces with fixed bases and $x_0, x_1, \tilde{x}_0, \tilde{x}_1$ be the linear coordinates with respect to these bases. Let $p \in \mathbb{C}[x_0]$ and $\tilde{p} \in \mathbb{C}[\tilde{x}_0]$ be nontrivial polynomial functions on V_0 and $\widetilde{V}_0$ respectively, and let*

$$\phi = (\phi_0, \phi_1) : \mathbb{C} \times V_0 \times V_1 \to \widetilde{V}_0 \times \widetilde{V}_1$$

be a germ of a holomorphic map with components in $\mathcal{R}_0(\mathbb{C} \times V_0 \times V_1, \mathbb{C} \times V_0)$, such that $\phi(\mathbb{C} \times V_0 \times \{0\}) \subseteq \widetilde{V}_0 \times \{0\}$, and satisfying $\tilde{p}(\phi_0(1/p(x_0), x_0, 0)) \not\equiv 0$. Then given any $\tilde{h} \in \mathcal{R}_0(\mathbb{C} \times \widetilde{V}_0 \times \widetilde{V}_1, \mathbb{C} \times \widetilde{V}_0)$, there exists $h \in \mathcal{R}_0(\mathbb{C} \times V_0 \times V_1, \mathbb{C} \times V_0)$ such that

$$\tilde{h}\left(\frac{1}{\tilde{p}(\phi_0(1/p(x_0), x_0, x_1))}, \phi\left(\frac{1}{p(x_0)}, x_0, x_1\right)\right) \equiv h\left(\frac{1}{q(x_0)}, x_0, x_1\right), \tag{3-32}$$

with $q(x_0) := p(x_0)^t \tilde{p}(\phi_0(1/p(x_0), x_0, 0))$ for some positive integer t. Also, h vanishes on $\mathbb{C} \times V_0 \times \{0\}$ if $\tilde{h}$ vanishes on $\mathbb{C} \times \widetilde{V}_0 \times \{0\}$.

Lemma 3.9 will be key in establishing the following lemma.

Lemma 3.10. *Let M and M' be as in Theorem 3.3. Given any $\beta = (\beta_1, \dots, \beta_n)$ and $\alpha = (\alpha_1, \dots, \alpha_n)$, with $1 \leq \beta_1 < \cdots < \beta_n \leq m$ and $1 \leq \alpha_1 < \cdots < \alpha_n \leq m$, and any positive integer s, there exists a $\mathbb{C}^{n+e}$-valued function $\Xi_s^{\alpha,\beta}(x, \Lambda, \Gamma)$ holomorphic on an open subset of $\mathbb{C}^{sm} \times J^{sk}(\mathbb{C}^{m+d}, \mathbb{C}^{n+e})_{(0,0)} \times J^{sk}(\mathbb{C}^{m+d}, \mathbb{C}^{n+e})_{(0,0)}$ of the form*

$$\Xi_s^{\alpha,\beta}(x, \Lambda, \Gamma) = \sum_\gamma \frac{P_\gamma^{\alpha\beta s}(\Lambda, \Gamma)}{Q_\gamma^{\alpha\beta s}(\Lambda, \Gamma)} x^\gamma, \tag{3-33}$$

where each $P_\gamma^{\alpha\beta s}$ is a $\mathbb{C}^{n+e}$-valued polynomial, and

$$Q_\gamma^{\alpha\beta s}(\Lambda, \Gamma) := (\det(\Lambda^{r,\alpha_l})_{1 \leq r,l \leq n})^{u_{\alpha\beta\gamma s}} (\det(\Gamma^{r,\beta_l})_{1 \leq r,l \leq n})^{v_{\alpha\beta\gamma s}}$$

for some nonnegative integers $u_{\alpha\beta\gamma s}$ and $v_{\alpha\beta\gamma s}$, such that if $\mathscr{H}$ is a germ at 0 of an HSPM satisfying condition $D_{\mu\nu}$, then

$$H(v^s(t^0,\dots,t^{s-1})) = \Xi_s^{\mu,\nu}(t^0,\dots,t^{s-1}, j_0^{sk}H, j_0^{sk}\widetilde{H}), \tag{3-34}$$

$$\widetilde{H}(\overline{v^s}(t^0,\dots,t^{s-1})) = \overline{\Xi_s^{\nu,\mu}}(t^0,\dots,t^{s-1}, j_0^{sk}\widetilde{H}, j_0^{sk}H), \tag{3-35}$$

for $t^0,\dots,t^{s-1}$ sufficiently close to 0. Furthermore, for any (Λ_0,Γ_0) such that $\det(\Lambda_0^{r,\alpha_l})_{1\le r,l\le n}\neq 0$ and $\det(\Gamma_0^{r,\beta_l})_{1\le r,l\le n}\neq 0$, the function $\Xi_s^{\alpha,\beta}$ is holomorphic on a neighborhood of $(0,\Lambda_0,\Gamma_0)$.

Proof. We inductively prove something stronger. First, we simplify notation slightly. Define

$$p^\alpha(\Lambda_1(Z)) := \det(\Lambda^{r,\alpha_l}(Z))_{1\le r,l\le n},$$
$$\tilde{p}^\beta(\widetilde{\Lambda}_1(\zeta)) := \det(\widetilde{\Lambda}^{r,\beta_l}(\zeta))_{1\le r,l\le n},$$

where Λ_1 is as defined in (3-9) (and $\widetilde{\Lambda}_1$ is defined similarly). We will show that for any γ and s, there exist nonnegative integers $a^s_{\alpha\beta}$, $b^s_{\alpha\beta}$ and holomorphic maps $\Theta_s^{\alpha,\beta,\gamma}$ with components in

$$\mathscr{R}_0\big(\mathbb{C}\times J^{ks+|\gamma|}(\mathbb{C}^{m+d},\mathbb{C}^{n+e})_{(0,0)}\times J^{ks+|\gamma|}(\mathbb{C}^{m+d},\mathbb{C}^{n+e})_{(0,0)}\times\mathbb{C}^{ms},\ \mathbb{C}\times J^{ks+|\gamma|}(\mathbb{C}^{m+d},\mathbb{C}^{n+e})_{(0,0)}\times J^{ks+|\gamma|}(\mathbb{C}^{m+d},\mathbb{C}^{n+e})_{(0,0)}\big)$$

such that

$$\partial^\gamma H(v^s(t^0,\dots,t^{s-1})) = \Theta_s^{\mu,\nu,\gamma}\left(\frac{1}{p^\mu(\Lambda_1(0))^{a^s_{\mu\nu}}\,\tilde{p}^\nu(\widetilde{\Lambda}_1(0))^{b^s_{\mu\nu}}}, j_0^{ks+|\gamma|}H, j_0^{ks+|\gamma|}\widetilde{H}, t^0,\dots,t^{s-1}\right) \tag{3-36}$$

and

$$\partial^\gamma \widetilde{H}(\overline{v^s}(t^0,\dots,t^{s-1})) = \overline{\Theta_s^{\nu,\mu,\gamma}}\left(\frac{1}{p^\nu(\widetilde{\Lambda}_1(0))^{a^s_{\nu\mu}}\,\tilde{p}^\mu(\Lambda_1(0))^{b^s_{\nu\mu}}}, j_0^{ks+|\gamma|}\widetilde{H}, j_0^{ks+|\gamma|}H, t^0,\dots,t^{s-1}\right). \tag{3-37}$$

First, we reformulate Lemma 3.7 with new notation. Let $j_Z^lH = ((j_Z^l)'H, (j_Z^l)''H)$, where $(j_Z^l)''H = (g_{z^\alpha}(Z))_{|\alpha|\le l}$, and $(j_Z^l)'H$ represents the remaining derivatives at Z. (A similar decomposition applies to $\widetilde{H}$). According to Lemma 3.7, there exist maps θ_γ^α with components in $\mathscr{R}_0(\mathbb{C}\times\mathbb{C}^{l'_\gamma}\times\mathbb{C}^{2m+2d}\times\mathbb{C}^{l''_\gamma}, \mathbb{C}\times\mathbb{C}^{l'_\gamma})$, for some integers l'_γ and l''_γ, such that for $(Z,\zeta)\in\mathcal{M}$,

$$\partial^\gamma H(Z) = \theta_\gamma^\nu\left(\frac{1}{\tilde{p}^\nu((j_\zeta^{k+|\gamma|})'\widetilde{H})}, (j_\zeta^{k+|\gamma|})'\widetilde{H}, Z, \zeta, (j_\zeta^{k+|\gamma|})''\widetilde{H}\right), \tag{3-38}$$

$$\partial^\gamma \widetilde{H}(\zeta) = \overline{\theta_\gamma^\mu}\left(\frac{1}{p^\mu((j_Z^{k+|\gamma|})'H)}, (j_Z^{k+|\gamma|})'H, \zeta, Z, (j_Z^{k+|\gamma|})''H\right). \tag{3-39}$$

It is easy to show that (3-36) and (3-37) hold for $s=1$ by letting $(Z,\zeta)=((z,0),0)$ in (3-38) and $(Z,\zeta)=(0,(\chi,0))$ in (3-39). So now assume (3-36) and (3-37) hold for s replaced by $s-1$, for some $s>1$. We will show they hold for s.

For any s, it is clear from the definition of the Segre mappings that

$$\big(v^s(t^0,\dots,t^{s-1}),\overline{v^{s-1}}(t^1,\dots,t^{s-1})\big)\in\mathcal{M}. \tag{3-40}$$

Using this fact in (3-38), we see that

$$\begin{aligned}\partial^\gamma H(v^s(t^0,\dots,t^{s-1}))\\ =\theta^\nu_\gamma\Bigg(\frac{1}{\tilde p^\nu\big(j^{k+|\gamma|}_{\overline{v^{s-1}}(t^1,\dots,t^{s-1})}\big)'\widetilde H\big)},\big(j^{k+|\gamma|}_{\overline{v^{s-1}}(t^1,\dots,t^{s-1})}\big)'\widetilde H,\\ v^s(t^0,\dots,t^{s-1}),\overline{v^{s-1}}(t^1,\dots,t^{s-1}),\big(j^{k+|\gamma|}_{\overline{v^{s-1}}(t^1,\dots,t^{s-1})}\big)''\widetilde H\Bigg).\end{aligned}$$

But by our induction hypothesis,

$$\begin{aligned}j^{k+|\gamma|}_{\overline{v^{s-1}}(t^1,\dots,t^{s-1})}\widetilde H\\ =\Bigg(\overline{\Theta^{\nu,\mu,\Delta}_{s-1}}\Bigg(\frac{1}{\tilde p^\nu(\widetilde\Lambda_1(0))^{a^{s-1}_{\nu\mu}}p^\mu(\Lambda_1(0))^{b^{s-1}_{\nu\mu}}},j_0^{ks+|\Delta|-k}H,\\ j_0^{ks+|\Delta|-k}\widetilde H,t^1,\dots,t^{s-1}\Bigg)\Bigg)_{|\Delta|\le k+|\gamma|}.\end{aligned}$$

For convenience, we write the tuple on the right hand side of the above as (A,B) where B corresponds to $(\tilde g_{\chi^\alpha}(\overline{v^{s-1}}(t^1,\dots,t^{s-1})))_{|\alpha|\le k+|\gamma|}$, and A corresponds to the remainder. We plug the last equation into the previous to get

$$\partial^\gamma H(v^s(t^0,\dots,t^{s-1}))=\theta^\nu_\gamma\bigg(\frac{1}{\tilde p^\nu(A)},A,v^s(t^0,\dots,t^{s-1}),\overline{v^{s-1}}(t^1,\dots,t^{s-1}),B\bigg).$$

Thus, (3-36) follows from Lemma 3.9.

To finish the proof, we need only show (3-37). Here we apply Lemma 3.5, which tells us that $(\overline{\widetilde H},\overline H)$ sends $\mathcal{M}$ into $\mathcal{M}'$ and satisfies condition $D_{\nu\mu}$. So by (3-36), we see that

$$\begin{aligned}\partial^\gamma\overline{\widetilde H}(v^s(t^0,\dots,t^{s-1}))\\ =\Theta^{\nu,\mu,\gamma}_s\Bigg(\frac{1}{p^\nu(\overline{\widetilde\Lambda}_1(0))^{a^s_{\nu\mu}}\tilde p^\mu(\overline\Lambda_1(0))^{b^s_{\nu\mu}}},j_0^{ks+|\gamma|}\overline{\widetilde H},j_0^{ks+|\gamma|}\overline H,t^0,\dots,t^{s-1}\Bigg).\end{aligned}$$

As p^ν and $\tilde p^\mu$ are polynomials with real coefficients, we take the complex conjugate of both sides of the above to see that (3-37) holds true. □

We are almost ready to complete the proof of Theorem 3.3. First, however, we present three lemmas. Lemma 3.11 can be found (using slightly different language) in [Baouendi et al. 1999a] and is thus presented here without proof. Lemma 3.12

is a generalization of a lemma found in [Baouendi et al. 2001]. Lemma 3.13 can be found in [Baouendi et al. 2001] and is presented here without proof.

Lemma 3.11. *Let M be as in Theorem 3.3. Then there exists an integer r such that the matrix*

$$\left(\frac{\partial v^{2r}}{\partial(t^0, t^{r+1}, t^{r+2}, \dots, t^{2r-1})}(0, x^1, \dots, x^{r-1}, x^r, x^{r-1}, \dots, x^1)\right) \tag{3-41}$$

has rank $m+d$ for all $(x^1, \dots, x^r) \in U \setminus V$, where $U \subseteq \mathbb{C}^{rm}$ is an open neighborhood of the origin, and V is a proper holomorphic subvariety of U. In addition,

$$v^{2r}(0, x^1, \dots, x^{r-1}, x^r, x^{r-1}, \dots, x^1) \equiv 0. \tag{3-42}$$

(*Here, v^{2r} is as defined in* (3-30).)

Lemma 3.12. *Let $V : (\mathbb{C}^{r_1} \times \mathbb{C}^{r_2}, 0) \to (\mathbb{C}^N, 0)$, $r_2 \geq N$, be a holomorphic map defined near 0 satisfying $V(x, \xi)|_{\xi=0} \equiv 0$, with $(x, \xi) \in \mathbb{C}^{r_1} \times \mathbb{C}^{r_2}$, and assume the matrix $((\partial V/\partial \xi)(x, 0))$ has an $N \times N$ minor that is not identically 0. Then there exist holomorphic maps (defined near 0)*

$$\delta : (\mathbb{C}^{r_1}, 0) \to \mathbb{C} \quad \textit{and} \quad \phi : (\mathbb{C}^{r_1} \times \mathbb{C}^N, 0) \to (\mathbb{C}^{r_2}, 0) \tag{3-43}$$

with $\delta(x) \not\equiv 0$, such that

$$V\left(x, \phi\left(x, \frac{Z}{\delta(x)}\right)\right) \equiv Z \tag{3-44}$$

for all $(x, Z) \in \mathbb{C}^{r_1} \times \mathbb{C}^N$ such that $\delta(x) \neq 0$ and both x and $Z/\delta(x)$ are sufficiently small. Furthermore, if V is holomorphic algebraic, then given any sufficiently small x_0 satisfying $\delta(x_0) \neq 0$, the map $\varphi_{x_0}(Z) := \phi(x_0, Z/\delta(x_0))$ is holomorphic algebraic for all Z in a neighborhood of 0.

Proof. Write $\xi = (\xi', \xi'')$, where

$$\xi' = (\xi_1, \dots, \xi_N) \in \mathbb{C}^N \quad \text{and} \quad \xi'' = (\xi_{N+1}, \dots, \xi_{r_2}) \in \mathbb{C}^{r_2 - N}.$$

Assume, without loss of generality, that $\det((\partial V/\partial \xi')(x, 0)) \not\equiv 0$. We wish to solve the equation $Z = V(x, \xi', 0)$ for ξ'. Since $V(x, 0) \equiv 0$, we can write $Z = V(x, \xi', 0) = a(x, \xi')\xi'$, where $a(x, \xi')$ is an $N \times N$ matrix of holomorphic functions defined near 0. By expanding $a(x, \xi')$, we can write

$$Z = V(x, \xi', 0) = a(x, 0)\xi' + ((\xi')^T R_j(x, \xi')\xi')_{1 \leq j \leq N}, \tag{3-45}$$

where each $R_j(x, \xi')$ is an $N \times N$ matrix of holomorphic functions defined near 0. Define $d(x) := \det((\partial V/\partial \xi')(x, 0))$. Using the fact that $(\mathrm{adj}(A))A = \det(A)I$

for any square matrix A, we multiply the far left and far right sides of (3-45) by $b(x) := \operatorname{adj}(a(x,0))$, noting that $a(x,0) = (\partial V/\partial \xi')(x,0)$, to get

$$b(x)Z - d(x)\xi' - b(x)((\xi')^T R_j(x,\xi')\xi')_{1\le j\le N} = 0. \tag{3-46}$$

Divide both sides of (3-46) by $d(x)^2$ and substitute $\tilde{\xi}' = \xi'/d(x)$ and $\widetilde{Z} = Z/d(x)^2$ to get

$$b(x)\widetilde{Z} - \tilde{\xi}' - b(x)((\tilde{\xi}')^T R_j(x, d(x)\tilde{\xi}')\tilde{\xi}')_{1\le j\le N} = 0.$$

By the implicit function theorem, there is a unique holomorphic solution $\tilde{\xi}' = \theta(x,\widetilde{Z})$ defined near 0 such that $\theta(0) = 0$. Thus, the first part of the theorem follows by letting $\delta(x) := d(x)^2$ and

$$\phi(x,y) := (d(x)\theta(x,y), 0) \quad \text{for } (x,y) \in \mathbb{C}^{r_1} \times \mathbb{C}^N.$$

If V is algebraic, the last part of the theorem follows from the *algebraic* implicit function theorem; see [Baouendi et al. 1999b] for example. □

Lemma 3.13. *Let V_0 and V_1 be finite-dimensional vector spaces with fixed linear coordinates x_0 and x_1, respectively. Let $P(x_0,x_1,\lambda) \in \mathcal{R}_0(V_0\times V_1\times\mathbb{C}, V_0)$ with $P(x_0,0,0)\equiv 0$. For a given integer $l \ge 0$, consider the Laurent series expansion*

$$P\left(x_0, \frac{x_1}{\lambda^l}, \lambda\right) = \sum_{\nu\in\mathbb{Z}} c_\nu(x_0,x_1)\lambda^\nu. \tag{3-47}$$

Then $c_0(x_0,0)\equiv 0$, and $c_\nu \in \mathcal{R}_0(V_0\times V_1, V_0)$ for every $\nu\in\mathbb{Z}$.

Proof of Theorem 3.3. Let r be as in Lemma 3.11. Take $x = (x^1,\dots,x^r) \in \mathbb{C}^{rm}$ and $y = (y^0,\dots,y^{r-1}) \in \mathbb{C}^{rm}$. Let $L(x,y) := (y^0, x^1,\dots,x^r, x^{r-1}+y^{r-1},\dots,x^1+y^1)$ and $V(x,y) := v^{2r}(L(x,y))$. In Lemma 3.12, we take $r_1 = r_2 = rm$. From (3-42), we see that $V(x,0)\equiv 0$. Also, from Lemma 3.11, we see that the other hypothesis of Lemma 3.12 holds. Thus we apply Lemma 3.12. Let δ and ϕ be as given in the lemma. We plug these into (3-34) to see that

$$H(Z) \equiv \Xi_{2r}^{\mu,\nu}\left(L\left(x, \phi\left(x, \frac{Z}{\delta(x)}\right)\right), j_0^{2rk}H, j_0^{2rk}\widetilde{H}\right). \tag{3-48}$$

We rewrite the right hand side of (3-48) as

$$H(Z) \equiv \widehat{\Xi}_{2r}^{\mu,\nu}\left(j_0^{2rk}H, j_0^{2rk}\widetilde{H}, \frac{Z}{\delta(x)}, x\right), \tag{3-49}$$

noting that the components of

$$\widehat{\Xi}_{2r}^{\mu,\nu} : J^{2rk}(\mathbb{C}^{m+d},\mathbb{C}^{n+e})_{(0,0)} \times J^{2rk}(\mathbb{C}^{m+d},\mathbb{C}^{n+e})_{(0,0)} \times \mathbb{C}^{m+d}\times\mathbb{C}^{rm} \to \mathbb{C}^{n+e}$$

are holomorphic on an open neighborhood of

$$J^{2rk}(\mathbb{C}^{m+d},\mathbb{C}^{n+e})_{(0,0)} \times J^{2rk}(\mathbb{C}^{m+d},\mathbb{C}^{n+e})_{(0,0)} \times \mathbb{C}^{m+d}\times\mathbb{C}^{rm}.$$

Now choose $x_0 \in \mathbb{C}^{rm}$ such that $\hat{\delta}(t) := \delta(tx_0) \not\equiv 0$, for $t \in \mathbb{C}$. Since $H(Z)$ is independent of x, we can replace $x = tx_0$ in (3-49). There exists a smallest integer l such that $(d^l/dt^l)\hat{\delta}(0) \neq 0$. To make our calculations easier, consider a holomorphic change of variable $\lambda = h(t)$ near the origin in $\mathbb{C}$, where h is determined by $\delta(tx_0) = \lambda^l$. So we now have

$$\begin{aligned}\widehat{\widehat{\Xi}}{}^{\mu,\nu}_{2r}\left(j_0^{2rk}H, j_0^{2rk}\widetilde{H}, \frac{Z}{\lambda^l}, \lambda\right) &:= \widehat{\Xi}^{\mu,\nu}_{2r}\left(j_0^{2rk}H, j_0^{2rk}\widetilde{H}, \frac{Z}{\lambda^l}, x_0 h^{-1}(\lambda)\right) \\ &\equiv H(Z).\end{aligned} \tag{3-50}$$

Observe that the components of $\widehat{\widehat{\Xi}}{}^{\mu,\nu}_{2r}$ are in

$$\begin{aligned}\mathcal{R}_0\big(J^{2rk}(\mathbb{C}^{m+d}, \mathbb{C}^{n+e})_{(0,0)} \times J^{2rk}(\mathbb{C}^{m+d}, \mathbb{C}^{n+e})_{(0,0)} \times \mathbb{C}^{m+d} \times \mathbb{C}, \\ J^{2rk}(\mathbb{C}^{m+d}, \mathbb{C}^{n+e})_{(0,0)} \times J^{2rk}(\mathbb{C}^{m+d}, \mathbb{C}^{n+e})_{(0,0)}\big).\end{aligned}$$

To conclude the proof, we expand the left hand side of (3-50) as a Laurent series in λ. Since $H(Z)$ is independent of λ, we can let $H(Z)$ be the constant term of the Laurent series. By Lemma 3.13 and the form of $\Xi^{\mu,\nu}_{2r}$ given in (3-33), we see that this is exactly of the form (3-3).

Applying Lemma 3.5, we see that $(\overline{\widetilde{H}}, \overline{H})$ sends $\mathcal{M}$ into $\mathcal{M}'$ and satisfies condition $D_{\nu\mu}$. From (3-4), we have

$$\overline{\widetilde{H}}(Z) = \Phi^{\nu,\mu}\left(Z, j_0^K(\overline{\widetilde{H}}), j_0^K(\overline{H})\right). \tag{3-51}$$

Take the complex conjugate of this entire equation, and (3-5) follows. □

Reformulation of Theorem 1.6.

Theorem 3.14. *Let M and M' be as in Theorem 3.3. Then there exists a positive integer L, depending only on M and M', such that for each $\alpha = (\alpha_1, \dots, \alpha_n)$ with $1 \le \alpha_1 < \cdots < \alpha_n \le m$ and each $\beta = (\beta_1, \dots, \beta_n)$ with $1 \le \beta_1 < \cdots < \beta_n \le m$, there exist $\mathbb{C}^{2n+2e}$-valued holomorphic functions $\Phi_1^{\alpha,\beta}$ and $\Phi_2^{\alpha,\beta}$ defined on an open subset of $\mathbb{C}^{m+d} \times \mathbb{C}^{m+d} \times J^L(\mathbb{C}^{m+d}, \mathbb{C}^{n+e})_{(0,0)}$ such that if $\mathcal{H}$ is a germ at 0 of an HSPM satisfying condition $D_{\mu\nu}$, then*

$$\mathcal{H}(Z,\zeta) = (H(Z), \widetilde{H}(\zeta)) = \Phi_1^{\mu,\nu}(Z,\zeta, j_0^L H), \tag{3-52}$$

$$\mathcal{H}(Z,\zeta) = (H(Z), \widetilde{H}(\zeta)) = \Phi_2^{\mu,\nu}(Z,\zeta, j_0^L \widetilde{H}), \tag{3-53}$$

for (Z,ζ) sufficiently close to 0.

Proof. We will prove (3-52); the proof of (3-53) follows similarly. We will show inductively that there exist $\mathbb{C}^{n+e}$-valued holomorphic functions $B_s^{\alpha,\beta,\gamma}$ defined on an open subset of $J^{ks+|\gamma|}(\mathbb{C}^{m+d}, \mathbb{C}^{n+e})_{(0,0)} \times \mathbb{C}^{ms}$, such that

$$\partial^\gamma H(v^s(t^0, \dots, t^{s-1})) = B_s^{\mu,\nu,\gamma}(j_0^{ks+|\gamma|}\mathcal{G}, t^0, \dots, t^{s-1}), \tag{3-54}$$

$$\partial^\gamma \widetilde{H}(\overline{v^s}(t^0, \dots, t^{s-1})) = \overline{B_s^{\nu,\mu,\gamma}}(j_0^{ks+|\gamma|}\mathcal{G}', t^0, \dots, t^{s-1}), \tag{3-55}$$

where $\mathcal{G} = H$ and $\mathcal{G}' = \widetilde{H}$ if s is even, and $\mathcal{G} = \widetilde{H}$ and $\mathcal{G}' = H$ if s is odd.

For $s = 1$, we see that (3-54) and (3-55) hold true by letting $(Z, \zeta) = ((z, 0), 0)$ in (3-22) and $(Z, \zeta) = (0, (\chi, 0))$ in (3-23). For some $s > 1$, assume (3-54) and (3-55) hold for $s - 1$. Assume, without loss of generality, that s is even (a similar proof works for s odd). Since $(v^s(t^0, \dots, v^{s-1}), \overline{v^{s-1}}(t^1, \dots, t^{s-1})) \in \mathcal{M}$, we see from (3-22) that

$$\begin{aligned}\partial^\beta H(v^s(t^0, \dots, t^{s-1}))&\\ \equiv \Psi_\beta^\nu\big(v^s(t^0, \dots, t^{s-1}), \overline{v^{s-1}}(t^1, \dots, t^{s-1}), j^{k+|\beta|}_{\overline{v^{s-1}}(t^1,\dots,t^{s-1})}\widetilde{H}\big).\end{aligned} \tag{3-56}$$

Using (3-55), we see then that

$$\begin{aligned}\partial^\beta H(v^s(t^0, \dots, t^{s-1}))&\\ \equiv \Psi_\beta^\nu\big(v^s(t^0, \dots, t^{s-1}), \overline{v^{s-1}}(t^1, \dots, t^{s-1}),&\\ (\overline{B_{s-1}^{\nu,\mu,\gamma}}(j_0^{k(s-1)+|\gamma|}H, t^1, \dots, t^{s-1}))_{|\gamma| \le k+|\beta|}\big).\end{aligned} \tag{3-57}$$

Now define $B_s^{\mu,\nu,\beta}(\Lambda, t^0, \dots, t^{s-1})$ to be the right hand side of (3-57), with the jets of H replaced by the appropriate corresponding coordinates of Λ.

Using Lemma 3.5, we see that $(\overline{\widetilde{H}}, \overline{H})$ satisfies condition $D_{\nu\mu}$ and sends $\mathcal{M}$ into $\mathcal{M}'$. So, we have from (3-54)

$$\partial^\gamma \overline{\widetilde{H}}(v^s(t^0, \dots, t^{s-1})) = B_s^{\nu,\mu,\gamma}(j_0^{ks+|\gamma|}\overline{\widetilde{H}}, t^0, \dots, t^{s-1}). \tag{3-58}$$

Taking the complex conjugate of both sides gives us (3-55).

Let r be as given in Lemma 3.11. We know from (3-54) and (3-55) that

$$H(v^{2r}(t^0, \dots, t^{2r-1})) = B_{2r}^{\mu,\nu,0}(j_0^{2kr}H, t^0, \dots, t^{2r-1}), \tag{3-59}$$

$$\widetilde{H}(\overline{v^{2r+1}}(t^0, \dots, t^{2r})) = \overline{B_{2r+1}^{\nu,\mu,0}}(j_0^{2kr+k}H, t^0, \dots, t^{2r}). \tag{3-60}$$

Since $v^{l+1}(t^0, \dots, t^{l-1}, 0) = v^l(t^0, \dots, t^{l-1})$ for any positive integer l, we see from Lemma 3.11 that the matrix

$$\left(\frac{\partial v^{2r+1}}{\partial(t^0, t^{r+1}, t^{r+2}, \dots, t^{2r-1})}(0, x^1, \dots, x^{r-1}, x^r, x^{r-1}, \dots, x^1, 0)\right) \tag{3-61}$$

has rank $m + d$ for all $(x_1, \dots, x_r) \in U \setminus V$, for $U \subseteq \mathbb{C}^{rm}$ an open neighborhood of the origin and V a proper holomorphic subvariety of U, and we also see that

$$v^{2r+1}(0, x^1, \dots, x^{r-1}, x^r, x^{r-1}, \dots, x^1, 0) \equiv 0. \tag{3-62}$$

We can now use (3-59) and (3-60) to obtain (3-52) and (3-53) by following exactly the proof of Theorem 3.3. □

Reformulation of Theorem 1.8.

Theorem 3.15. *Let M and M' be as in Theorem 3.3, and assume that M and M' are real algebraic. If $\mathcal{H}$ is a germ at 0 of an HSPM satisfying condition $D_{\mu\nu}$ for some μ and ν, then $\mathcal{H}$ is holomorphic algebraic.*

Proof. An inspection of the proof of Lemma 3.6 shows that the ϕ^{α}_{β} as given in (3-10) are holomorphic algebraic (as M is real algebraic). When solving the system of equations in (3-24), apply the *algebraic* implicit function theorem to see that B^{ν} as given in (3-25) is holomorphic algebraic (as M' is real algebraic). Thus, an inspection of the proof of Lemma 3.7 shows that the Ψ^{α}_{β} as given in (3-21) are holomorphic algebraic. An examination of the proof of Lemma 3.10 then reveals that the $\Xi^{\alpha,\beta}_{s}$ as given in (3-33) are holomorphic algebraic. Finally, in the proof of Theorem 3.3, choose x_0 sufficiently small and satisfying $\delta(x_0) \neq 0$, and substitute $x = x_0$ in (3-48). By Lemma 3.12, we see then that $H(Z)$ is holomorphic algebraic. Similarly, $\tilde{H}(\zeta)$ is holomorphic algebraic. □

4. Proofs of main results

In Section 1, we presented results 1.1, 1.2, 1.4, and 1.5, all of which follow naturally from Theorem 3.3. We also presented Theorem 1.6, which is a direct result of Theorem 3.14, and Theorem 1.8, which is a direct result of Theorem 3.15. In this section, we provide their proofs. First we make the following observations.

Observation 4.1. If $M \subseteq \mathbb{C}^N$ and $M' \subseteq \mathbb{C}^{N'}$ are generic submanifolds of codimensions d and d', respectively, given in normal coordinates by $w = Q(z, \chi, \tau)$ and $w' = Q'(z', \chi', \tau')$, respectively, then a germ at 0 of an HSPM $\mathcal{H} = (f, g, \tilde{f}, \tilde{g})$ sending $(\mathcal{M}, 0)$ into $(\mathcal{M}', 0)$ is Segre submersive at 0 if and only if the matrices $(f_z(0))$ and $(\tilde{f}_\chi(0))$ have rank $N' - d'$. This follows from the fact that a basis for the antiholomorphic vectors tangent to M (respectively, M') at 0 is given by $\{\partial/\partial\bar{z}_j : 1 \le j \le N - d\}$ (respectively, $\{\partial/\partial\bar{z}'_j : 1 \le j \le N' - d'\}$), while a basis for the holomorphic vectors tangent to M (respectively, M') at 0 is given by $\{\partial/\partial z_j : 1 \le j \le N - d\}$ (respectively, $\{\partial/\partial z'_j : 1 \le j \le N' - d'\}$), coupled with the fact that $g_{z_j}(0) = \tilde{g}_{\chi_j}(0) = 0$ for $j = 1, \dots, N - d$.

Observation 4.2. For $p \in \mathbb{C}^N$, let $\phi : (\mathbb{C}^N, 0) \to (\mathbb{C}^N, p)$ be a biholomorphism near 0, and for $p' \in \mathbb{C}^{N'}$, let $\phi' : (\mathbb{C}^{N'}, 0) \to (\mathbb{C}^{N'}, p')$ be a biholomorphism near 0. Then for any nonnegative l, there exist vector-valued polynomial functions F_l and G_l such that if $h : (\mathbb{C}^N, 0) \to (\mathbb{C}^{N'}, 0)$ is any holomorphic map, and $\tilde{h} : (\mathbb{C}^N, p) \to (\mathbb{C}^{N'}, p')$ is given by $\tilde{h} := \phi' \circ h \circ \phi^{-1}$, then $j^l_p \tilde{h} = F_l(j^l_0 h)$ and $j^l_0 h = G_l(j^l_p \tilde{h})$.

Proof of Theorem 1.1. This theorem follows from Theorem 3.3, Observation 4.1, and Observation 4.2. The boundaries on r given in Theorem 1.1 follow from the fact that there are $\binom{m}{n}$ possibilities for μ and $\binom{m}{n}$ possibilities for ν in Theorem 3.3. Therefore, there are $\binom{m}{n}\binom{m}{n}$ possible choices for $\Phi^{\alpha,\beta}$ in (3-3). □

Proof of Corollary 1.2. Without loss of generality, assume $p = 0$. Since $M = M'$, we have $r = 1$ in Theorem 1.1. Define $\Phi := \Phi_1$ as given in (1-2). It then follows from Theorem 1.1 that η_0^K is continuous and injective on $\mathrm{Aut}_{\mathbb{C}}(\mathcal{M}, 0)$. To show that η_0^K is a homeomorphism from $\mathrm{Aut}_{\mathbb{C}}(\mathcal{M}, 0)$ onto $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M}, 0))$, we need to show the continuity of $(\eta_0^K)^{-1}$. Let $\Lambda_j, \widetilde{\Lambda}_j, \Lambda_0, \widetilde{\Lambda}_0$ belong to $G_0^k(\mathbb{C}^N)$ and assume $(\Lambda_j, \widetilde{\Lambda}_j) \in \eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M}, 0))$ converges to $(\Lambda_0, \widetilde{\Lambda}_0) \in \eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M}, 0))$. Then Theorem 1.1 tells us that

$$(\eta_0^K)^{-1}(\Lambda_j, \widetilde{\Lambda}_j) = (\Phi(Z, \Lambda_j, \widetilde{\Lambda}_j), \overline{\Phi}(\zeta, \widetilde{\Lambda}_j, \Lambda_j)),$$
$$(\eta_0^K)^{-1}(\Lambda_0, \widetilde{\Lambda}_0) = (\Phi(Z, \Lambda_0, \widetilde{\Lambda}_0), \overline{\Phi}(\zeta, \widetilde{\Lambda}_0, \Lambda_0)).$$

It follows that $(\eta_0^K)^{-1}(\Lambda_j, \widetilde{\Lambda}_j)$ converges to $(\eta_0^K)^{-1}(\Lambda_0, \widetilde{\Lambda}_0)$.

We now show $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M}, 0))$ is a closed, holomorphic algebraic submanifold of $G_0^K(\mathbb{C}^N) \times G_0^K(\mathbb{C}^N)$. Let $\rho(Z, \bar{Z})$ be a defining function for M near 0. Write $\zeta = (\zeta_1, \zeta_2) \in \mathbb{C}^{N-d} \times \mathbb{C}^d$, where d is the codimension of M. After a possible rearrangement of coordinates, since M is generic, there exists a holomorphic map $\theta : \mathbb{C}^N \times \mathbb{C}^{N-d} \to \mathbb{C}^d$ satisfying $\theta(0) = 0$ such that for all Z and ζ_1 sufficiently close to 0, we have $(Z, \zeta_1, \theta(Z, \zeta_1)) \in \mathcal{M}$. Given $(\Lambda_0, \widetilde{\Lambda}_0) \in G_0^K(\mathbb{C}^N) \times G_0^K(\mathbb{C}^N)$, $(\Lambda_0, \widetilde{\Lambda}_0) \in \eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M}, 0))$ if and only if the following three conditions hold:

$$\Lambda_0 = (S_\gamma(\Lambda_0, \widetilde{\Lambda}_0))_{|\gamma| \le K}, \tag{4-1}$$

$$\widetilde{\Lambda}_0 = (\overline{S_\gamma}(\widetilde{\Lambda}_0, \Lambda_0))_{|\gamma| \le K}, \tag{4-2}$$

$$\rho(\Phi(Z, \Lambda_0, \widetilde{\Lambda}_0), \overline{\Phi}(\zeta_1, \theta(Z, \zeta_1), \widetilde{\Lambda}_0, \Lambda_0)) = 0, \tag{4-3}$$

where S_γ are the rational coefficients in the Taylor expansion given in (1-2). Equations (4-1) and (4-2) can be expressed as a finite set of polynomial equations in Λ_0 and $\widetilde{\Lambda}_0$ as each S_γ is rational. We claim that (4-3) can be expressed as an infinite set of polynomial equations in Λ_0 and $\widetilde{\Lambda}_0$. Indeed, we can write the Taylor expansion as

$$\rho\big(\Phi(Z, \Lambda_0, \widetilde{\Lambda}_0), \overline{\Phi}(\zeta_1, \theta(Z, \zeta_1), \widetilde{\Lambda}_0, \Lambda_0)\big) = \sum_{\alpha,\beta} R_{\alpha\beta}(\Lambda_0, \widetilde{\Lambda}_0) Z^\alpha \zeta_1^\beta.$$

Observe that each $R_{\alpha\beta}$ is rational. This can be seen by noting that $\Phi(0, \Gamma, \widetilde{\Lambda}) \equiv 0$ and $\theta(0) = 0$, and by noting the form of Φ given in Theorem 1.1. The claim follows since the set of polynomial equations comes from setting the numerators of $R_{\alpha\beta}$ equal to 0 for all α and β.

Thus, we see that $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$ is a closed, holomorphic algebraic subvariety of the space $G_0^K(\mathbb{C}^N)\times G_0^K(\mathbb{C}^N)$ since it is given by the vanishing of a set of polynomial equations. To see it is actually a submanifold, note first that it is a subgroup of $G_0^K(\mathbb{C}^N)\times G_0^K(\mathbb{C}^N)$ since multiplication can be defined as follows. Given any $(\Lambda_1,\widetilde{\Lambda}_1)$, $(\Lambda_2,\widetilde{\Lambda}_2)\in\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$, let $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, be the corresponding automorphisms in $\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0)$. Now compose $\mathcal{H}_1$ and $\mathcal{H}_2$, and apply η_0^K to this composition. Under this multiplication, $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$ is a closed subgroup of the Lie group $G_0^K(\mathbb{C}^N)\times G_0^K(\mathbb{C}^N)$, and is thus a Lie subgroup; see [Varadarajan 1974], for example. □

Proof of Corollary 1.4.

Lemma 4.3. *Let $A=(a_{ij})$ be a $d\times d$ invertible matrix, where each $a_{ij}\in\mathbb{C}$. Let $b_1,\dots,b_d\in\mathbb{C}$. Let B_1 be the matrix obtained by replacing row m of A with $(a_{m1}+b_1,\dots,a_{md}+b_d)$ and B_2 be the matrix obtained by replacing row m of A with $(a_{m1}-b_1,\dots,a_{md}-b_d)$. Then at least one of B_1 or B_2 is invertible.*

Proof. Without loss of generality, assume $m=1$. Let $A_n:=(-1)^{n+1}\det M_n$, where M_n is the $(d-1)\times(d-1)$ matrix obtained by deleting the first row and n-th column of A. Assume that $\det B_1=\det B_2=0$. Then expanding along the first row of B_1 gives

$$(a_{11}+b_1)A_1+\cdots+(a_{1d}+b_d)A_d=0, \tag{4-4}$$

and expanding along the first row of B_2 gives

$$(a_{11}-b_1)A_1+\cdots+(a_{1d}-b_d)A_d=0. \tag{4-5}$$

Adding (4-4) and (4-5) gives $2a_{11}A_1+\ldots+2a_{1d}A_d=0$. However, this implies that $\det A=0$, a contradiction. □

We now prove Corollary 1.4. Without loss of generality, assume $p=0$. Let $r(\Lambda,\overline{\Lambda})=(r_1(\Lambda,\overline{\Lambda}),\dots,r_s(\Lambda,\overline{\Lambda}))$ be a defining function for $j_0^K(\mathrm{Aut}(M,0))$ as a real algebraic submanifold of $G_0^K(\mathbb{C}^N)$, where $\Lambda\in G_0^K(\mathbb{C}^N)$; see Remark 1.3. The complexification $\mathbb{C}\{j_0^K(\mathrm{Aut}(M,0))\}$ of this submanifold is thus a complex submanifold of $G_0^K(\mathbb{C}^N)\times G_0^K(\mathbb{C}^N)$ given by the vanishing of $r(\Lambda,\widetilde{\Lambda})$, where $\widetilde{\Lambda}\in G_0^K(\mathbb{C}^N)$. Let $\rho(Z,\bar{Z})$ be a defining function for M near 0. Since $M=M'$, it is clear from the statement of Theorem 3.3 that we can choose $r=1$ in Theorem 1.1. Do so, and define $\Phi:=\Phi_1$ as given in (1-2). From Theorem 1.1, we see that for any $\Lambda\in G_0^K(\mathbb{C}^N)$,

$$\rho(\Phi(Z,\Lambda,\overline{\Lambda}),\overline{\Phi}(\bar{Z},\overline{\Lambda},\Lambda))=A(Z,\Lambda,\bar{Z},\overline{\Lambda})r(\Lambda,\overline{\Lambda})+B(Z,\Lambda,\bar{Z},\overline{\Lambda})\rho(Z,\bar{Z}),$$

where A is a real analytic $d\times s$ matrix, and B is a real analytic $d\times d$ matrix. Complexify to get

$$\rho(\Phi(Z,\Lambda,\widetilde{\Lambda}),\bar{\Phi}(\zeta,\widetilde{\Lambda},\Lambda))=A(Z,\Lambda,\zeta,\widetilde{\Lambda})r(\Lambda,\widetilde{\Lambda})+B(Z,\Lambda,\zeta,\widetilde{\Lambda})\rho(Z,\zeta).$$

Notice that this gives us exactly what we want. This equation says that if $(\Lambda,\widetilde{\Lambda})$ is in $\mathbb{C}\{j_0^K(\mathrm{Aut}(M,0))\}$, then $(\Phi(Z,\Lambda,\widetilde{\Lambda}),\bar{\Phi}(\zeta,\widetilde{\Lambda},\Lambda))$ is in $\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0)$. Now we need only show that

$$\eta_0^K(\Phi(Z,\Lambda,\widetilde{\Lambda}),\bar{\Phi}(\zeta,\widetilde{\Lambda},\Lambda))=(\Lambda,\widetilde{\Lambda}). \tag{4-6}$$

We have the equations

$$\begin{aligned}(\partial_Z^\alpha\Phi(0,\Lambda,\bar{\Lambda}))_{|\alpha|\le K}&=\Lambda+C(\Lambda,\bar{\Lambda})r(\Lambda,\bar{\Lambda}),\\(\partial_{\bar{Z}}^\alpha\bar{\Phi}(0,\bar{\Lambda},\Lambda))_{|\alpha|\le K}&=\bar{\Lambda}+\bar{C}(\bar{\Lambda},\Lambda)r(\Lambda,\bar{\Lambda}),\end{aligned}$$

for C a real analytic matrix. Complexify these to get

$$\begin{aligned}(\partial_Z^\alpha\Phi(0,\Lambda,\widetilde{\Lambda}))_{|\alpha|\le K}&=\Lambda+C(\Lambda,\widetilde{\Lambda})r(\Lambda,\widetilde{\Lambda}),\\(\partial_{\zeta}^\alpha\bar{\Phi}(0,\widetilde{\Lambda},\Lambda))_{|\alpha|\le K}&=\widetilde{\Lambda}+\bar{C}(\widetilde{\Lambda},\Lambda)r(\Lambda,\widetilde{\Lambda}),\end{aligned}$$

and the first part of Corollary 1.4 is proved.

Since we are assuming $p=0$, we take $\mathrm{Id}=\mathrm{Id}'$ in Corollary 1.4. To prove the second part of Corollary 1.4, first we show that $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$ is a complexified submanifold near $(\mathrm{Id},\mathrm{Id})$. In other words, $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))=\mathbb{C}R$, where R is a real submanifold of $G_0^K(\mathbb{C}^N)$ (here $\mathbb{C}R$ denotes the complexification of R). We know from Corollary 1.2 that $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$ is a complex submanifold of $G_0^K(\mathbb{C}^N)\times G_0^K(\mathbb{C}^N)$. Near $(\mathrm{Id},\mathrm{Id})$, let $\hat{s}_1(\Lambda,\widetilde{\Lambda}),\dots,\hat{s}_t(\Lambda,\widetilde{\Lambda})$ be defining functions for $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$. We will assume without loss of generality that these functions are defined on a ball B of sufficiently small radius centered at $(\mathrm{Id},\mathrm{Id})$; this way if (Γ,Λ) is a point in B, then so is (Λ,Γ) and $(\bar{\Lambda},\bar{\Gamma})$. Now we set $s_j(\Lambda,\widetilde{\Lambda})$ to one of the following:

$$\begin{aligned}s_j(\Lambda,\widetilde{\Lambda})&:=\hat{s}_j(\Lambda,\widetilde{\Lambda})+\bar{\hat{s}}_j(\widetilde{\Lambda},\Lambda),\\s_j(\Lambda,\widetilde{\Lambda})&:=i\hat{s}_j(\Lambda,\widetilde{\Lambda})-i\bar{\hat{s}}_j(\widetilde{\Lambda},\Lambda),\end{aligned}$$

We choose between these options as follows. Start with $j=1$. From Lemma 4.3, we can replace $\hat{s}_1$ with one of the above s_1, and in at least one case the differentials of $s_1,\hat{s}_2,\dots,\hat{s}_t$ will be linearly independent near $(\mathrm{Id},\mathrm{Id})$. Choose s_1 so that this is the case. Now do the same thing for $j=2$, then $j=3$, and so forth. Let $\mathcal{R}$ be the submanifold defined by $s_1(\Lambda,\widetilde{\Lambda})=\cdots=s_t(\Lambda,\widetilde{\Lambda})=0$. If $(\Lambda,\widetilde{\Lambda})$ is in $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$, then from Lemma 3.5, we have $(\bar{\widetilde{\Lambda}},\bar{\Lambda})\in\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$. Thus, $\hat{s}_j(\bar{\widetilde{\Lambda}},\bar{\Lambda})=0$, which implies $\bar{\hat{s}}_j(\widetilde{\Lambda},\Lambda)=0$. In other words, near $(\mathrm{Id},\mathrm{Id})$,

$\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0)) \subseteq \mathcal{R}$. But these two submanifolds have equal dimensions. So we see that, in fact, $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0)) = \mathcal{R}$ near (Id, Id).

Now we need only show that $\mathcal{R} = \mathbb{C}R$ for some real submanifold $R \subseteq G_0^K(\mathbb{C}^N)$, and we will have proved our claim. Let

$$R := \{\Lambda : s_1(\Lambda, \overline{\Lambda}) = \cdots = s_t(\Lambda, \overline{\Lambda}) = 0\}. \tag{4-7}$$

Clearly R is a nonempty set as it contains the point $\Lambda = \mathrm{Id}$. Since each s_j is a real function and the differentials of $s_1, \dots, s_t$ are linearly independent, R is a real submanifold.

From Theorem 1.1, we see that if $(H, \widetilde{H}) \in \mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0)$ and $j_0^K(\widetilde{H}) = j_0^K(\overline{H})$, we must have $\widetilde{H} = \overline{H}$. Thus, near (Id, Id),

$$\begin{aligned} \mathbb{C}\{j_0^K(\mathrm{Aut}(M,0))\} \cap \{\widetilde{\Lambda} = \overline{\Lambda}\} &= \eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0)) \cap \{\widetilde{\Lambda} = \overline{\Lambda}\} \\ &= \mathbb{C}R \cap \{\widetilde{\Lambda} = \overline{\Lambda}\}, \end{aligned} \tag{4-8}$$

implying that $j_0^K(\mathrm{Aut}(M,0)) = R$. Thus their complexifications must be equal as well. That is, $\mathbb{C}\{j_0^K(\mathrm{Aut}(M,0))\} = \eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$ near (Id, Id). But both of these are algebraic holomorphic submanifolds. So if they are equal near (Id, Id), then using the notation given in the statement of this corollary, we must have $\mathcal{B} = \mathcal{C}$.

The third part of the corollary comes from the fact that $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$ is a Lie subgroup. Thus, each of its connected components is a coset of $\mathcal{B}$. Since $\mathbb{C}\{j_0^K(\mathrm{Aut}(M,0))\} \subseteq \eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$ and they are both algebraic holomorphic submanifolds, each component of $\mathbb{C}\{j_0^K(\mathrm{Aut}(M,0))\}$ is exactly equal to one of the components of $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$. Algebraicity implies that there are finitely many such components. □

Proof of Theorem 1.5. Assume first that M and M' are given in normal coordinates and that $p = 0$ and $p' = 0$. The equalities of the K-jets imply, in particular, that $\mathcal{H}_1$ and $\mathcal{H}_2$ both satisfy condition $D_{\mu\nu}$ for some μ and ν. Theorem 1.5 now follows from Theorem 3.3 and Observations 4.1 and 4.2. □

Proof of Theorem 1.6. This follows from Theorem 3.14 and Observations 4.1 and 4.2. We leave the details to the reader. □

Proof of Theorem 1.8. As M and M' are real algebraic, they have real analytic algebraic defining functions. When M and M' are expressed in normal coordinates, the new defining functions can also be chosen to be real analytic algebraic. This follows by using the algebraic implicit function theorem in the derivation of the new defining functions; see [Baouendi et al. 1999b] for precise details on deriving normal coordinates and the algebraic implicit function theorem. Also, if $\widetilde{Z} = \varphi(Z)$ is a holomorphic algebraic change of coordinates, then φ^{-1} is a holomorphic algebraic function; this is also a direct consequence of the algebraic implicit function theorem. Theorem 1.8 now follows from Theorem 3.15 and Observation 4.1. □

5. Examples: HSPMs and automorphism groups

For $n > 1$, there exist $M, M' \subseteq \mathbb{C}^{n+1}$ defined near 0 such that there exist no holomorphic maps H satisfying

$$H \text{ is invertible near } 0, \quad H(M) \subseteq M', \quad H(0) = 0, \tag{5-1}$$

yet there exist HSPMs satisfying

$$\mathcal{H} \text{ is invertible near } 0, \quad \mathcal{H}(\mathcal{M}) \subseteq \mathcal{M}', \quad \mathcal{H}(0) = 0. \tag{5-2}$$

Example 5.1. For $n > 1$, let $(z_1, \dots, z_n, w)$ and $(z'_1, \dots, z'_n, w')$ be coordinates on $\mathbb{C}^{n+1}$ and define

$$M = \left\{ \operatorname{Im} w = \sum_{j=1}^{n} \epsilon_j |z_j|^2 \right\} \quad \text{and} \quad M' = \left\{ \operatorname{Im} w' = \sum_{j=1}^{n} \sigma_j |z'_j|^2 \right\},$$

where $\epsilon_j, \sigma_j \in \{-1, 1\}$. Both M and M' are of finite type and finitely nondegenerate at 0. If $|\sum_j \epsilon_j| \neq |\sum_j \sigma_j|$, then there are no holomorphic maps satisfying criteria (5-1). (Indeed, M and M' have different Levi signatures at 0.) However, for $a, c_j \in \mathbb{C} \setminus \{0\}$, the family of maps given by

$$\begin{aligned} &\mathcal{H}(z, w, \chi, \tau) \\ &= \left(\epsilon_1 c_1 z_1, \dots, \epsilon_{n-1} c_{n-1} z_{n-1}, \epsilon_n c_n z_n, a w, \frac{a\sigma_1}{c_1} \chi_1, \dots, \frac{a\sigma_{n-1}}{c_{n-1}} \chi_{n-1}, \frac{a\sigma_n}{c_n} \chi_n, a\tau \right) \end{aligned}$$

satisfy criteria (5-2).

This can also occur in $\mathbb{C}^2$ as the next example illustrates.

Example 5.2. Let $M, M' \subseteq \mathbb{C}^2$ be given by

$$\begin{aligned} M &= \{ \operatorname{Im} w = |z|^2 + 2 \operatorname{Re}(z^4 \bar{z}^2 (1 + i \operatorname{Re} w)) \}, \\ M' &= \{ \operatorname{Im} w' = |z'|^2 + 2 \operatorname{Re}(z'^4 \bar{z}'^2 (1 - i \operatorname{Re} w')) \}. \end{aligned}$$

Notice that M and M' are of finite type and finitely nondegenerate at 0. It can be shown [Chern and Moser 1974] that there are no maps H satisfying criteria (5-1). (Indeed, as M and M' are in Chern–Moser normal form, the fact that the coefficients i and $-i$ are unequal implies that there does not exist a holomorphic map H satisfying criteria (5-1).) However, it is easy to check that the HSPM $\mathcal{H}(z, w, \chi, \tau) = (iz, -w, i\chi, -\tau)$ satisfies (5-2).

Now we will look at some examples of automorphism groups. In Example 5.3, we find that $\mathbb{C}\{j_0^K(\operatorname{Aut}(M, 0))\}$ and $\eta_0^K(\operatorname{Aut}_{\mathbb{C}}(\mathcal{M}, 0))$ are equal.

Example 5.3. Let M be the *Lewy hypersurface* of $\mathbb{C}^2$. It is given by

$$M = \{ \operatorname{Im} w = |z|^2 \}.$$

We note that M is finitely nondegenerate and of finite type at 0. It can be shown (see [Angle 2008b] for the calculations) that every holomorphic Segre-preserving automorphism of $\mathcal{M}$ at 0 is of the form

$$\mathcal{H}(z,w,\chi,\tau)=\left(\frac{\alpha(z+\beta w)}{1-(\gamma+i\beta\tilde{\beta})w-2i\tilde{\beta}z},\frac{\alpha\tilde{\alpha}w}{1-(\gamma+i\beta\tilde{\beta})w-2i\tilde{\beta}z},\right.$$
$$\left.\frac{\tilde{\alpha}(\chi+\tilde{\beta}\tau)}{1-(\gamma-i\beta\tilde{\beta})\tau+2i\beta\chi},\frac{\alpha\tilde{\alpha}\tau}{1-(\gamma-i\beta\tilde{\beta})\tau+2i\beta\chi}\right), \tag{5-3}$$

where $\gamma,\beta,\tilde{\beta}\in\mathbb{C}$ and $\alpha,\tilde{\alpha}\in\mathbb{C}\setminus\{0\}$. Also, every automorphism of M at 0 is of the form

$$H(z,w)=\left(\frac{\alpha(z+\beta w)}{1-(\gamma+i|\beta|^2)w-2i\bar{\beta}z},\frac{|\alpha|^2w}{1-(\gamma+i|\beta|^2)w-2i\bar{\beta}z}\right), \tag{5-4}$$

where $\alpha\in\mathbb{C}\setminus\{0\}$, $\beta\in\mathbb{C}$, and $\gamma\in\mathbb{R}$. The automorphisms in (5-4) follow directly from the automorphisms in (5-3), but those in (5-4) have actually been known for some time [Chern and Moser 1974].

We use (5-3) and (5-4) to show that $\mathbb{C}\{j_0^K(\mathrm{Aut}(M,0))\}=\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$: Let

$$(\Lambda_z^f,\ldots,\Lambda_{ww}^f,\Lambda_z^g,\ldots,\Lambda_{ww}^g,\Lambda_\chi^{\tilde{f}},\ldots,\Lambda_{\tau\tau}^{\tilde{f}},\Lambda_\chi^{\tilde{g}},\ldots,\Lambda_{\tau\tau}^{\tilde{g}})$$

be coordinates on $G_0^2(\mathbb{C}^2)\times G_0^2(\mathbb{C}^2)$, where

$$\Lambda^f_{z^rw^s}\text{ corresponds to }\frac{\partial^{r+s}f}{\partial z^r\partial w^s},\qquad \Lambda^g_{z^rw^s}\text{ corresponds to }\frac{\partial^{r+s}g}{\partial z^r\partial w^s},$$
$$\Lambda^{\tilde{f}}_{\chi^r\tau^s}\text{ corresponds to }\frac{\partial^{r+s}\tilde{f}}{\partial \chi^r\partial \tau^s},\qquad \Lambda^{\tilde{g}}_{\chi^r\tau^s}\text{ corresponds to }\frac{\partial^{r+s}\tilde{g}}{\partial \chi^r\partial \tau^s}.$$

Then (5-4) implies that $\mathbb{C}\{j_0^2(\mathrm{Aut}(M,0))\}$ is given by

$$\begin{aligned}\Big\{\ &\Lambda_w^g=\Lambda_\tau^{\tilde{g}}=\Lambda_z^f\Lambda_\chi^{\tilde{f}},\quad \Lambda_{ww}^g=\Lambda_{\tau\tau}^{\tilde{g}}+2i\Lambda_w^f\Lambda_\tau^{\tilde{f}},\\ &\Lambda_{zw}^g=2i\Lambda_z^f\Lambda_\tau^{\tilde{f}},\qquad \Lambda_{\chi\tau}^{\tilde{g}}=-2i\Lambda_\chi^{\tilde{f}}\Lambda_w^f,\\ &\Lambda_{zw}^f=\frac{\Lambda_{ww}^g}{\Lambda_\chi^{\tilde{f}}},\quad \Lambda_{zz}^f=2i\frac{\Lambda_z^f\Lambda_\tau^{\tilde{f}}}{\Lambda_\chi^{\tilde{f}}},\quad \Lambda_{ww}^f=\frac{\Lambda_{ww}^g\Lambda_w^f}{\Lambda_z^f\Lambda_\chi^{\tilde{f}}},\\ &\Lambda_{\chi\tau}^{\tilde{f}}=\frac{\Lambda_{\tau\tau}^{\tilde{g}}}{\Lambda_z^f},\quad \Lambda_{\chi\chi}^{\tilde{f}}=-2i\frac{\Lambda_\chi^{\tilde{f}}\Lambda_w^f}{\Lambda_z^f},\quad \Lambda_{\tau\tau}^{\tilde{f}}=\frac{\Lambda_{\tau\tau}^{\tilde{g}}\Lambda_\tau^{\tilde{f}}}{\Lambda_\chi^{\tilde{f}}\Lambda_z^f},\\ &\Lambda_z^g=\Lambda_{zz}^g=\Lambda_\chi^{\tilde{g}}=\Lambda_{\chi\chi}^{\tilde{g}}=0\Big\}.\end{aligned} \tag{5-5}$$

It follows from (5-3) that $\eta_0^2(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$ is also given by (5-5).

More interesting, however, are submanifolds for which $\mathbb{C}\{j_0^K(\mathrm{Aut}(M,0))\}$ is not equal to $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$.

Example 5.4. Let $M \subseteq \mathbb{C}^2$ be given by

$$M = \{\mathrm{Im}\, w = |z|^2 + (\mathrm{Re}\, z^2)|z|^2\}.$$

Note that M is finitely nondegenerate and of finite type at 0. Baouendi et al. [1997] showed that there are only two automorphisms of M at 0, namely $H_1(z,w) = (z,w)$ and $H_2(z,w) = (-z,w)$. Thus, $\mathbb{C}\{j_0^K(\mathrm{Aut}(M,0))\}$ also has only two elements. However, the group of holomorphic Segre-preserving automorphisms of $\mathcal{M}$ at 0 (which according to Corollary 1.4 consists of a finite number of elements) contains at least four maps,

$$\begin{aligned} &\mathcal{H}_1(z,w,\chi,\tau) = (z,w,\chi,\tau), && \mathcal{H}_2(z,w,\chi,\tau) = (-z,w,-\chi,\tau),\\ &\mathcal{H}_3(z,w,\chi,\tau) = (-z,-w,\chi,-\tau), && \mathcal{H}_4(z,w,\chi,\tau) = (z,-w,-\chi,-\tau). \end{aligned}$$

The next two examples will compare $\eta_0^K(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$ and $\mathbb{C}\{j_0^K(\mathrm{Aut}(M,0))\}$ for the family $\mathcal{F}$ given by

$$\mathcal{F} = \big\{M = \{\mathrm{Im}\, w = c_1|z|^{2m} + c_2|z|^{2n}\} \;\big|\; 1 < m < n,\, |c_1|^2 + |c_2|^2 \neq 0\big\}.$$

We exclude the *Levi flat* case, $M = \{\mathrm{Im}\, w = 0\}$, since there is no finite jet determination for this M. Notice that each submanifold in $\mathcal{F}$ is of finite type and finitely *degenerate* at 0.

Example 5.5. Assume $c_1 \neq 0$ and $c_2 = 0$. Calculations in [Angle 2008b] show that any holomorphic Segre-preserving automorphism of $\mathcal{M}$ at 0 is given by

$$\begin{aligned} \mathcal{H}(z,w,\chi,\tau) &= (f(z,w), g(z,w), \tilde{f}(\chi,\tau), \tilde{g}(\chi,\tau)) \\ &= \left(\frac{az}{\sqrt[m]{1+\alpha w}}, \frac{a^m\tilde{a}^m w}{1+\alpha w}, \frac{\tilde{a}\chi}{\sqrt[m]{1+\alpha\tau}}, \frac{a^m\tilde{a}^m\tau}{1+\alpha\tau}\right), \end{aligned} \tag{5-6}$$

where $a, \tilde{a} \in \mathbb{C}\setminus\{0\}$, $\alpha \in \mathbb{C}$, and f and $\tilde{f}$ are expressed in terms of any branch of the m-th root.

It immediately follows that any automorphism of M at 0 is of the form

$$H(z,w) = \left(\frac{az}{\sqrt[m]{1+\alpha w}}, \frac{|a|^{2m} w}{1+\alpha w}\right), \tag{5-7}$$

where $a \in \mathbb{C}\setminus\{0\}$, $\alpha \in \mathbb{R}$, and f is expressed in terms of any branch of the m-th root.

In this case, $\eta_0^2(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0)) = \mathbb{C}\{j_0^2(\mathrm{Aut}(M,0))\}$. Indeed, from (5-7), we see that $\mathbb{C}\{j_0^2(\mathrm{Aut}(M,0))\}$ is given by

$$\text{(5-8)}\qquad \big\{\Lambda^g_w = \Lambda^{\tilde g}_\tau = (\Lambda^f_z)^m(\Lambda^{\tilde f}_\chi)^m, \quad \Lambda^g_{ww} = 2m(\Lambda^f_z)^{m-1}(\Lambda^{\tilde f}_\chi)^m\Lambda^f_{zw},$$
$$\Lambda^{\tilde g}_{\tau\tau} = 2m(\Lambda^{\tilde f}_\chi)^{m-1}(\Lambda^f_z)^m\Lambda^{\tilde f}_{\chi\tau},$$
$$\Lambda^f_w = \Lambda^f_{ww} = \Lambda^f_{zz} = \Lambda^g_z = \Lambda^g_{zw} = \Lambda^g_{zz}$$
$$= \Lambda^{\tilde f}_\tau = \Lambda^{\tilde f}_{\tau\tau} = \Lambda^{\tilde f}_{\chi\chi} = \Lambda^{\tilde g}_\chi = \Lambda^{\tilde g}_{\chi\tau} = \Lambda^{\tilde g}_{\chi\chi} = 0\big\}.$$

It follows from (5-6) that $\eta_0^2(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$ is also given by (5-8).

Example 5.6. Assume $c_1, c_2 \neq 0$. Calculations in [Angle 2008b] show that any holomorphic Segre-preserving automorphism of $\mathcal{M}$ at 0 is of one of the $n-m$ forms

$$\text{(5-9)}\qquad \mathcal{H}_c(z,w,\chi,\tau) = (az, c^m w, (c/a)\chi, c^m\tau),$$

where $a \in \mathbb{C}\setminus\{0\}$ and $c \in \{e^{2i\pi r/(n-m)} : r = 0,\dots,n-m-1\}$ (that is, c is a primitive $(n-m)$-th root of unity).

It immediately follows that any automorphism of M at 0 is of the form

$$\text{(5-10)}\qquad H(z,w) = (e^{i\theta}z, w), \quad \text{where } \theta \in \mathbb{R}.$$

Thus, we see from (5-10) that $\mathbb{C}\{j_0^1(\mathrm{Aut}(M,0))\}$ is given by

$$\{\Lambda^g_w = \Lambda^{\tilde g}_\tau = 1,\ \Lambda^f_z\Lambda^{\tilde f}_\chi = 1,\ \Lambda^f_w = \Lambda^{\tilde f}_\tau = \Lambda^g_z = \Lambda^{\tilde g}_\chi = 0\}$$

and thus has positive dimension. For $n = m+1$, (5-9) implies $\mathbb{C}\{j_0^1(\mathrm{Aut}(M,0))\} = \eta_0^1(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$. For $n > m+1$, however, $\mathbb{C}\{j_0^1(\mathrm{Aut}(M,0))\} \subsetneqq \eta_0^1(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$. Indeed, we see from (5-9) that $\eta_0^1(\mathrm{Aut}_{\mathbb{C}}(\mathcal{M},0))$ is equal to the disjoint union of exactly $n-m$ distinct cosets of $\mathbb{C}\{j_0^1(\mathrm{Aut}(M,0))\}$.

References

[Angle 2008a] R. B. Angle, "Geometric properties and related results for holomorphic Segre preserving maps", preprint, 2008. To appear in *Complex Var. Elliptic Equ.* arXiv 0810.2570

[Angle 2008b] R. B. Angle, *Holomorphic Segre preserving maps*, thesis, University of California, San Diego, 2008, available at http://tinyurl.com/3ybdqjb. MR 2626910

[Baouendi et al. 1996] M. S. Baouendi, P. Ebenfelt, and L. P. Rothschild, "Algebraicity of holomorphic mappings between real algebraic sets in $\mathbf{C}^n$", *Acta Math.* **177** (1996), 225–273. MR 99b:32030 Zbl 0890.32005

[Baouendi et al. 1997] M. S. Baouendi, P. Ebenfelt, and L. P. Rothschild, "Parametrization of local biholomorphisms of real analytic hypersurfaces", *Asian J. Math.* **1**:1 (1997), 1–16. MR 99b:32022 Zbl 0943.32021

[Baouendi et al. 1999a] M. S. Baouendi, P. Ebenfelt, and L. P. Rothschild, "Rational dependence of smooth and analytic CR mappings on their jets", *Math. Ann.* **315**:2 (1999), 205–249. MR 2001b:32075 Zbl 0942.32027

[Baouendi et al. 1999b] M. S. Baouendi, P. Ebenfelt, and L. P. Rothschild, *Real submanifolds in complex space and their mappings*, Princeton Math. Series **47**, Princeton University Press, 1999. MR 2000b:32066 Zbl 0944.32040

[Baouendi et al. 2001] M. S. Baouendi, L. P. Rothschild, and D. Zaitsev, "Equivalences of real submanifolds in complex space", *J. Differential Geom.* **59**:2 (2001), 301–351. MR 2003k:32051 Zbl 1037.32030

[Bloom and Graham 1977] T. Bloom and I. Graham, "On "type" conditions for generic real submanifolds of $\mathbf{C}^n$", *Invent. Math.* **40**:3 (1977), 217–243. MR 58 #28644 Zbl 0346.32013

[Chern 1975] S. S. Chern, "On the projective structure of a real hypersurface in C_{n+1}", *Math. Scand.* **36** (1975), 74–82. MR 52 #814 Zbl 0305.53019

[Chern and Moser 1974] S. S. Chern and J. K. Moser, "Real hypersurfaces in complex manifolds", *Acta Math.* **133** (1974), 219–271. MR 54 #13112 Zbl 0302.32015

[Faran 1980] J. J. Faran, "Segre families and real hypersurfaces", *Invent. Math.* **60**:2 (1980), 135–172. MR 81i:32018 Zbl 0464.32011

[Kohn 1972] J. J. Kohn, "Boundary behavior of δ on weakly pseudo-convex manifolds of dimension two", *J. Differential Geometry* **6** (1972), 523–542. MR 48 #727 Zbl 0256.35060

[Lamel and Mir 2007] B. Lamel and N. Mir, "Parametrization of local CR automorphisms by finite jets and applications", *J. Amer. Math. Soc.* **20** (2007), 519–572. MR 2008b:32025 Zbl 1112.32017

[Segre 1931] B. Segre, "Intorno al problema di Poincaré della rappresentazione pseudoconforme", *Rendiconti Accad. d. L. Roma* **13** (1931), 676–683. JFM 57.0404.05

[Varadarajan 1974] V. S. Varadarajan, *Lie groups, Lie algebras, and their representations*, Prentice-Hall, Englewood Cliffs, N.J., 1974. MR 51 #13113 Zbl 0371.22001

[Zhang 2007] Y. Zhang, "Rigidity and holomorphic Segre transversality for holomorphic Segre maps", *Math. Ann.* **337**:2 (2007), 457–478. MR 2008e:32021 Zbl 1117.32012

Received May 21, 2009.

R. BLAIR ANGLE
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA AT SAN DIEGO
LA JOLLA, CA 92093-0112
UNITED STATES

angle@metsci.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

CHERN CLASSES ON DIFFERENTIAL K-THEORY

ULRICH BUNKE

In this note we give a simple, model-independent construction of Chern classes as natural transformations from differential complex K-theory to differential integral cohomology. We verify the expected behavior of these Chern classes with respect to sums and suspension.

1. Statements

Complex K-theory and integral cohomology $H\mathbb{Z}$ are generalized cohomology theories that have unique differential[1] extensions $(\hat{K}, R, I, a, \int)$ and $(\widehat{H\mathbb{Z}}, R, I, a, \int)$ with integration. These extensions are multiplicative in a unique way. We refer to [Bunke and Schick 2010] for a description of the axioms for differential extensions of cohomology theories and a proof of these statements.

The i-th Chern class is a natural transformation of *set-valued* functors

$$c_i : K^0 \to H\mathbb{Z}^{2i}$$

on the category of topological spaces. The product $H\mathbb{Z}^{\mathrm{ev}} := \prod_{i\geq 0} H\mathbb{Z}^{2i}$ is a functor with values in commutative graded rings. We consider the subfunctor

$$H\mathbb{Z}_1^{\mathrm{ev},*} := 1 + \prod_{i\geq 1} H\mathbb{Z}^{2i} \subseteq \prod_{i\geq 0} H\mathbb{Z}^{2i}$$

that takes values in the subgroup of units. The total Chern class

$$c := 1 + c_1 + c_2 + \cdots : K^0 \to H\mathbb{Z}_1^{\mathrm{ev},*}$$

is a natural transformation of *group-valued* functors.

Let $\Omega^*_{\mathrm{cl}}(\dots, K^*) \subseteq \Omega^*(\dots, K^*)$ denote the graded ring valued functors on smooth manifolds of smooth differential forms with coefficients in K^* and its subfunctor of closed forms. We use the powers of the Bott element in K^2 to identify the functors

$$\Omega^0(\dots, K^*) \cong \Omega^{\mathrm{ev}}(\dots) \quad \text{and} \quad \Omega^{-1}(\dots, K^*) \cong \Omega^{\mathrm{odd}}(\dots).$$

MSC2000: 19L10.
Keywords: differential K-theory, Chern classes.

[1] In previous work, we used the term "smooth cohomology" instead of "differential cohomology". We were convinced by D. Freed that the latter is the better name.

We therefore have natural transformations

$$a : \Omega^{\mathrm{odd}} \to \hat{K}^0 \quad \text{and} \quad R : \hat{K}^0 \to \Omega^{\mathrm{ev}}_{\mathrm{cl}},$$

where a only preserves the additive structure, and R is multiplicative.

We consider the symmetric formal power series

$$\widetilde{\mathbf{ch}} := \sum_{i \geq 1} (e^{x_i} - 1) \in \mathbb{Q}[[x_1, x_2, \dots]]$$

in infinitely many variables. We write $\mathbf{ch}_i$ for the homogeneous component of degree i. Then there are polynomials $C_i \in \mathbb{Q}[s_1, s_2, \dots]$ of degree i (where s_i has degree i) such that $C_i(\mathbf{ch}_1, \dots, \mathbf{ch}_i) = \sigma_i$ is the i-th elementary symmetric function in the x_i. The polynomial C_i induces a natural transformation $C_i : \Omega^{\mathrm{ev}} \to \Omega^{2i}$ that maps the even form $\omega = \omega_0 + \omega_2 + \omega_4 + \cdots$, where $\omega_{2k} \in \Omega^{2k}(M)$, to

$$C_i(\omega) := C_i(\omega_2, \dots, \omega_{2i}) \in \Omega^{2i}(M).$$

The following theorem states that the Chern classes have unique lifts to the differential extensions and that these lifts are compatible with the group structures.

Theorem 1.1. (i) *For every $i \geq 1$, there exists a unique natural transformation*

$$\hat{c}_i : \hat{K}^0 \to \widehat{H\mathbb{Z}}^{2i}$$

of set-valued functors on smooth manifolds such that the following diagram commutes:

$$\begin{array}{ccc} \Omega^{\mathrm{ev}} & \xrightarrow{C_i} & \Omega^{2i} \\ \uparrow{\scriptstyle R} & & \uparrow{\scriptstyle R} \\ \hat{K}^0 & \xrightarrow{\hat{c}_i} & \widehat{H\mathbb{Z}}^{2i} \\ \downarrow{\scriptstyle I} & & \downarrow{\scriptstyle I} \\ K^0 & \xrightarrow{c_i} & H\mathbb{Z}^i \end{array} \tag{1}$$

(ii) *The total class $\hat{c} = 1 + \hat{c}_1 + \cdots : \hat{K}^0 \to \widehat{H\mathbb{Z}}_1^{\mathrm{ev},*}$ preserves the group structure.*

Berthomieu [2008] has already constructed lifts of the Chern classes. Our goal is to give a much simpler, model-independent treatment. Further new, but not very deep, points of the theorem above are the assertions about uniqueness and the second statement. Our method of proof is different from Berthomieu's and is in fact a specialization of a general principle used in [Bunke and Schick 2010] and [Bunke 2009a] for the construction of lifts of natural transformations between cohomology functors to their differential refinements.

In the next two paragraphs we connect the differential Chern classes on differential K-theory with previous constructions of differential Chern classes in specific geometric situations.

If $\boldsymbol{V} := (V, h^V, \nabla^V)$ is a hermitian vector bundle with connection over a manifold M, then we have the classes

$$\hat{c}_i^{\mathrm{CS}}(\boldsymbol{V}) \in \widehat{H\mathbb{Z}}^{2i}(M)$$

constructed in [Cheeger and Simons 1985]. In the model of differential K-theory [Bunke and Schick 2009], the geometric bundle is a cycle for a differential K-theory class $[\boldsymbol{V}] \in \hat{K}^0(M)$. We have $\hat{c}_i([\boldsymbol{V}]) = \hat{c}_i^{\mathrm{CS}}(\boldsymbol{V})$.

An even geometric family $\mathcal{E}$ over M (see [Bunke 2009b] for this notion) gives rise to a Bismut superconnection $A(\mathcal{E})$ on an infinite-dimensional Hilbert space bundle $H(\mathcal{E})$ over M. This superconnection

$$A(\mathcal{E}) = D(\mathcal{E}) + \nabla^{H(\mathcal{E})} + \text{higher terms}$$

extends the family of Dirac operators $D(\mathcal{E})$. If the kernel of $D(\mathcal{E})$ is a vector bundle, then it has an induced metric $h^{\ker(D(\mathcal{E}))}$ and connection $\nabla^{\ker(D(\mathcal{E}))}$ obtained from $\nabla^{H(\mathcal{E})}$ by projection. We thus get an induced geometric bundle

$$\mathbf{H}(\mathcal{E}) = (\ker(D(\mathcal{E})), h^{\ker(D(\mathcal{E}))}, \nabla^{\ker(D(\mathcal{E}))})$$

and can define the class $\hat{c}_i^{\mathrm{CS}}(\mathbf{H}(\mathcal{E})) \in \widehat{H\mathbb{Z}}^{2i}(M)$. One goal of [Bunke 2009b], which was not quite achieved there, was to extend this construction to the general case where we do not have a kernel bundle. By assuming that $\mathrm{index}(D(\mathcal{E})) \in K^0(M)$ belongs to the i-th step of the Atiyah–Hirzebruch filtration (that is, that it vanishes after pull-back to any $(i-1)$-dimensional complex), we constructed in that book's 4.1.19 a class $\hat{c}_i(\mathcal{E}) \in \widehat{H\mathbb{Z}}^{2i}(M)$[2] such that $I(\hat{c}_i(\mathcal{E})) = c_i(\mathrm{index}(D(\mathcal{E})))$. On the other hand, the geometric family $\mathcal{E}$ represents a differential K-theory class $[\mathcal{E}, 0] \in \hat{K}^0(M)$ in the model [Bunke and Schick 2009], and we have $I([\mathcal{E}, 0]) = \mathrm{index}(D(\mathcal{E}))$. The class $\hat{c}_i([\mathcal{E}, 0]) \in \widehat{H\mathbb{Z}}^{2i}(M)$ satisfies $I(\hat{c}_i(\mathcal{E})) = c_i(\mathrm{index}(D(\mathcal{E})))$ also and thus gives a second differential refinement of the i-th Chern class of the index of $D(\mathcal{E})$. But in general the class $\hat{c}_i(\mathcal{E})$ differs from $\hat{c}_i([\mathcal{E}, 0])$. This can already be seen on the level of curvatures. Namely, we have

$$R(\hat{c}_i(\mathcal{E})) = R([\mathcal{E}, 0])_{[2i]} \quad \text{and} \quad R(\hat{c}_i([\mathcal{E}, 0])) = C_i(R([\mathcal{E}, 0])),$$

where $\omega_{[2i]}$ denotes the degree-$2i$ component of the form ω. In a sense, this note gives the right answer to the problem considered in [Bunke 2009b].

[2] In [Bunke 2009b] we indexed the Chern classes by their degree, while here we adopt the usual convention.

Finally we discuss odd Chern classes. In topology, the odd Chern classes $c_i^{\mathrm{odd}}: K^{-1} \to H\mathbb{Z}^i$ are related with the even Chern classes by suspension:

$$\begin{array}{ccc} \tilde{K}^0(\Sigma M_+) & \xrightarrow{c_{(i+1)/2}} & \widetilde{H\mathbb{Z}}^{i+1}(\Sigma M_+) \\ \cong\downarrow & & \downarrow\cong \\ K^{-1}(M) & \xrightarrow{c_i^{\mathrm{odd}}} & H\mathbb{Z}^i(M). \end{array}$$

In the smooth context, the suspension isomorphism is replaced by the integration $\int$ along $S^1 \times M \to M$. We have the following odd counterpart of Theorem 1.1.

Theorem 1.2. *For odd $i \in \mathbb{N}$, there are unique natural transformations*

$$\hat{c}_i^{\mathrm{odd}} : \hat{K}^{-1} \to \widehat{H\mathbb{Z}}^i$$

such that

$$\begin{array}{ccc} \hat{K}^0(S^1 \times M) & \xrightarrow{\hat{c}_{(i+1)/2}} & \widehat{H\mathbb{Z}}^{i+1}(S^1 \times M) \\ \int\downarrow & & \downarrow\int \\ \hat{K}^{-1}(M) & \xrightarrow{\hat{c}_i^{\mathrm{odd}}} & \widehat{H\mathbb{Z}}^i(M) \end{array}$$

commutes. The transformation satisfies $I \circ \hat{c}_i^{\mathrm{odd}} = c_i^{\mathrm{odd}} \circ I$.

Let $\pi : W \to B$ be a proper K-oriented map between manifolds. Then we have an Umkehr map $\pi_! : K^*(W) \to K^{*-n}(B)$, where $n = \dim(W) - \dim(B)$. An integral index theorem is an assertion about the Chern classes $c_*(\pi_!(x))$, or $c_*^{\mathrm{odd}}(\pi_!(x))$ for $x \in K^*(W)$, for example, an expression of these classes in terms of the classes $c_*(x)$ or $c_*^{\mathrm{odd}}(x)$, respectively. A prototypical example is given in [Madsen 2009]. The construction of differential lifts of Chern classes makes it possible to ask for geometric refinements of these kinds of results. An example of such a theorem related to the Pfaffian bundle is discussed in [Bunke 2009c].

2. Proofs

Let $\boldsymbol{K}_0 \simeq \mathbb{Z} \times BU$ be a representative of the homotopy type of the classifying space of the functor K^0. By [Bunke and Schick 2010, Proposition 2.1], we may choose a sequence of manifolds $(\mathcal{K}_k)_{k \geq 0}$ together with maps $x_k : \mathcal{K}_k \to \boldsymbol{K}_0$ and $\kappa_k : \mathcal{K}_k \to \mathcal{K}_{k+1}$ such that

(i) $\mathcal{K}_k$ is homotopy equivalent to a k-dimensional CW-complex,

(ii) $\kappa_k : \mathcal{K}_k \to \mathcal{K}_{k+1}$ is an embedding of a closed submanifold,

(iii) $x_k : \mathcal{K}_k \to \boldsymbol{K}_0$ is k-connected, and

(iv) $x_{k+1} \circ \kappa_k = x_k$.

Let $u \in K^0(\boldsymbol{K}_0)$ be the universal class represented by the identity map $\boldsymbol{K}_0 \to \boldsymbol{K}_0$. By [Bunke and Schick 2010, Proposition 2.6] we can further choose a sequence $\hat{u}_k \in \hat{K}^0(\mathcal{K}_k)$ such that $I(\hat{u}_k) = x_k^* u$ and $\kappa_k^* \hat{u}_{k+1} = \hat{u}_k$ for all $k \geq 0$. Then by [ibid., Lemma 3.8] and $2j-1 < k$, we have $H^{2j-1}(\mathcal{K}_k, \mathbb{R}) = 0$. We consider the canonical natural transformation $\iota_{\mathbb{R}} : H\mathbb{Z}^* \to H\mathbb{R}^*$ and the de Rham map $\mathrm{Rham} : \Omega^*_{\mathrm{cl}} \to H\mathbb{R}^*$. Since the latter is multiplicative, we have

$$\iota_{\mathbb{R}}(c_i(I(\hat{u}_k))) = C_i(\mathbf{ch}(I(\hat{u}_k))) = C_i(\mathrm{Rham}(R(\hat{u}_k))) = \mathrm{Rham}(C_i(R(\hat{u}_k))).$$

If we choose $k \geq 2i$, then the diagram

$$\begin{array}{ccc} \widehat{H\mathbb{Z}}^{2i}(\mathcal{K}_k) & \xrightarrow{I} & H\mathbb{Z}^{2i}(\mathcal{K}_k) \\ \downarrow R & & \downarrow \iota_{\mathbb{R}} \\ \Omega^{2i}_{\mathrm{cl}}(\mathcal{K}_k) & \xrightarrow{\mathrm{Rham}} & H\mathbb{R}^{2i}(\mathcal{K}_k) \end{array}$$

is cartesian. Hence for $k \geq 2i$, there exists a unique class $\hat{z}_{i,k} \in \widehat{H\mathbb{Z}}^{2i}(\mathcal{K}_k)$ such that

$$I(\hat{z}_{i,k}) = c_i(I(\hat{u}_k)) \quad \text{and} \quad R(\hat{z}_i) = C_i(R(\hat{u}_k)).$$

Also, we have $\kappa_k^* \hat{z}_{i,k+1} = \hat{z}_{i,k}$. For $k < 2i$, we define $z_{i,k} := (\kappa_k^* \circ \cdots \circ \kappa_{2i-1}^*) z_{i,2i}$.

We now define the natural transformation $\hat{c}_i$. We start with the observation that if $\hat{c}_i$ exists, then it satisfies $\hat{c}_i(\hat{u}_k) = \hat{z}_{i,k}$.

Let $\hat{w} \in \hat{K}^0(M)$. By [ibid., Proposition 2.6] we have $K^0(M) \cong \mathrm{colim}_k[M, \mathcal{K}_k]$, and the underlying class $I(\hat{w}) \in K^0(M)$ can be written as $I(\hat{w}) = f^* x_k^* u$ for some k and $f : M \to \mathcal{K}_k$. We choose a form $\rho \in \Omega^{\mathrm{odd}}(M)$ such that $\hat{w} = f^* \hat{u}_k + a(\rho)$.

We consider a form $\tilde{\rho} \in \Omega^{\mathrm{odd}}([0,1] \times M)$ that restricts to ρ on $\{1\} \times M$ and to 0 on $\{0\} \times M$. We get a class $\tilde{\hat{w}} = \mathrm{pr}_M^* \hat{w} + a(\tilde{\rho}) \in \hat{K}^0([0,1] \times M)$. Note that

$$\tilde{\hat{w}}|_{\{0\} \times M} = f^* \hat{u}_k \quad \text{and} \quad \tilde{\hat{w}}|_{\{1\} \times M} = \hat{w}.$$

If $\hat{c}_i$ exists, then by naturality and the homotopy formula [ibid., (1)], we have

$$\hat{c}_i(\tilde{\hat{w}}|_{\{0\} \times M}) = f^* \hat{z}_{i,k}, \quad \hat{c}_i(\tilde{\hat{w}}|_{\{1\} \times M}) - \hat{c}_i(\tilde{\hat{w}}|_{\{0\} \times M}) = a\Big(\int_{[0,1] \times M/M} R(\hat{c}_i(\tilde{\hat{w}}))\Big).$$

Furthermore, by the commutativity of the upper square in (1), we must require

$$R(\hat{c}_i(\tilde{\hat{w}})) = C_i(R(\tilde{\hat{w}})).$$

Therefore we are forced to define

$$\hat{c}_i(\hat{w}) := f^* \hat{z}_{i,k} + a\Big(\int_{[0,1] \times M/M} C_i(R(\tilde{\hat{w}}))\Big). \tag{2}$$

We see that if $\hat{c}_i$ exists, it is automatically unique.

Lemma 2.1. *The definition of $\hat{c}_i(\hat{\omega})$ by (2) is independent of the choices of $\tilde{\rho}$, ρ and $f : M \to \mathcal{K}_k$.*

Proof. Let us start with a second choice $\tilde{\rho}'$ and write $\tilde{\hat{\omega}}' := \mathrm{pr}_M^* \hat{\omega} + a(\tilde{\rho}')$. Then we can connect $\tilde{\rho}$ with $\tilde{\rho}'$ by a family of such forms, for example, the linear path. This path can be considered as a form $\bar{\rho}$ on $[0, 1] \times [0, 1] \times M$. By construction $\bar{\rho}|_{[0,1]\times\{j\}\times M}$ is constant and has no component in the direction of the first variable for $j = 0, 1$. This implies that

$$R(\tilde{\hat{\omega}}')|_{[0,1]\times\{j\}\times M} = 0. \tag{3}$$

We set $\bar{\hat{\omega}} := \mathrm{pr}_M^* \hat{\omega} + a(\bar{\rho}) \in \hat{K}^0([0, 1] \times [0, 1] \times M)$. By Stokes' theorem we have

$$d \int_{[0,1]\times[0,1]\times M/M} C_i(R((\bar{\hat{\omega}}))) = \int_{[0,1]\times M/M} C_i(R(\tilde{\hat{\omega}}')) - \int_{[0,1]\times M/M} C_i(R(\tilde{\hat{\omega}}))$$

(these are the contributions of the faces $\{j\} \times [0, 1] \times M$) since the integral over the other two faces $[0, 1] \times \{j\} \times M$ vanishes by (3). Since a annihilates exact forms, this implies that

$$a\Big(\int_{[0,1]\times M/M} C_i(R(\tilde{\hat{\omega}}))\Big) = a\Big(\int_{[0,1]\times M/M} C_i(R(\tilde{\hat{\omega}}'))\Big).$$

Assume now that we have chosen a different ρ'. Then $a(\rho' - \rho) = 0$ so that by the exactness axiom [Bunke and Schick 2010, (2)] there exists a class $\hat{\upsilon} \in \hat{K}^1(M)$ with $R(\hat{\upsilon}) = \rho' - \rho$. Let $\hat{e} \in \hat{K}^1(S^1)$ be a lift of the generator of $K^1(S^1) \cong \mathbb{Z}$ with $R(\hat{e}) = dt$. We consider the form $\tilde{\sigma} \in \Omega^{\mathrm{odd}}([0, 1] \times M)$ with no dt component given by

$$\tilde{\sigma}|_{\{t\}\times M} := \int_{[0,t]\times M/M} R(\hat{e} \times \hat{\upsilon}),$$

where we identify $S^1 \cong \mathbb{R}/\mathbb{Z}$ and view the interval $[0, t]$ as a subset of S^1. Then

$$\tilde{\sigma}|_{\{0\}\times M} = 0, \quad \tilde{\sigma}|_{\{1\}\times M} = \rho' - \rho, \quad d\tilde{\sigma} = dt \wedge \mathrm{pr}_M^* R(\hat{\upsilon}) = R(\hat{e} \times \hat{\upsilon}).$$

We now consider

$$\tilde{\hat{\upsilon}} := \mathrm{pr}_M^* \hat{\omega} + \mathrm{pr}_M^* a(\rho) + a(\tilde{\sigma}) \in \hat{K}^0([0, 1] \times M)$$

and calculate modulo the image of d

$$\begin{aligned}\int_{[0,1]\times M/M} C_i(R(\tilde{\hat{\upsilon}})) &\equiv \int_{S^1\times M/M} C_i(R(\mathrm{pr}_M^*(\hat{\omega})) + \mathrm{pr}_M^* d\rho + R(\hat{e} \times \hat{\upsilon}))\\ &\equiv \int_{S^1\times M/M} C_i(R(\mathrm{pr}_M^*(\hat{\omega})) + R(\hat{e} \times \hat{\upsilon}))\\ &\equiv \int_{S^1\times M/M} C_i(R(\mathrm{pr}_M^*(\hat{\omega}) + \hat{e} \times \hat{\upsilon})).\end{aligned}$$

It follows that

$$\begin{aligned}\mathrm{Rham}\Big(\int_{[0,1]\times M/M} C_i(R(\tilde{\hat{v}}))\Big) &= \mathrm{Rham}\Big(\int_{S^1\times M/M} C_i(R(\mathrm{pr}_M^*(\hat{w})+\hat{e}\times\hat{v}))\Big)\\ &= \int_{S^1\times M/M} \mathrm{Rham}(C_i(R(\mathrm{pr}_M^*(\hat{w})+\hat{e}\times\hat{v})))\\ &= \int_{S^1\times M/M} \iota_{\mathbb{R}}(c_i(I(\mathrm{pr}_M^*(\hat{w})+\hat{e}\times\hat{v}))).\end{aligned}$$

In other words, $\mathrm{Rham}(\int_{[0,1]\times M/M} C_i(R(\tilde{\hat{v}})))$ is an integral class, and this implies

$$a\Big(\int_{[0,1]\times M/M} C_i(R(\tilde{\hat{v}}))\Big)=0$$

by [Bunke and Schick 2010, (2)].

If $\tilde{\rho}$ was the path connecting ρ with 0, then we construct the path $\tilde{\rho}'$ from ρ' to 0 by concatenating $\tilde{\rho}$ with $\tilde{\sigma}$ (we may change $\tilde{\rho}$ to ensure a smooth concatenation). Then we get $\tilde{\hat{w}}' := \mathrm{pr}_M^*\,\hat{w}+a(\tilde{\rho}')\in\hat{K}^0([0,1]\times M)$ and

$$\begin{aligned}a\Big(\int_{[0,1]\times M/M} C_i(R(\tilde{\hat{w}}'))\Big) &= a\Big(\int_{[0,1]\times M/M} C_i(R(\tilde{\hat{w}}))\Big)+a\Big(\int_{[0,1]\times M/M} C_i(R(\tilde{\hat{v}}))\Big)\\ &= a\Big(\int_{[0,1]\times M/M} C_i(R(\tilde{\hat{w}}))\Big).\end{aligned}$$

Thus our construction of c_i is independent of the choice of ρ.

Finally we verify that $\hat{c}_i(\hat{w})$ is independent of the choice of $f: M\to\mathcal{K}_k$. If we replace k by $k+1$ and f by $\kappa_k\circ f$, then we obviously get the same result. For two choices $f: M\to\mathcal{K}_k$ and $f': M\to\mathcal{K}_{k'}$, there exists $k''\geq\max\{k,k'\}$ such that $\kappa_k^{k''}\circ f$ and $\kappa_{k'}^{k''}\circ f'$ are homotopic. Here $\kappa_i^j:\mathcal{K}_i\to\mathcal{K}_j$ denotes the composition $\kappa_i^j := \kappa_{j-1}\circ\cdots\circ\kappa_i$ for $j>i$. Therefore it remains to show that a choice $f': M\to\mathcal{K}_k$ homotopic to $f: M\to\mathcal{K}_k$ gives the same result for $\hat{c}_i(\hat{w})$. Let $H:[0,1]\times M\to\mathcal{K}_k$ be a homotopy from f to f'. Then we use H in the construction of $\hat{c}_i(\mathrm{pr}_M^*\,\hat{w})\in\widehat{H\mathbb{Z}}^{2i}([0,1]\times M)$. If we let $\hat{c}_i'(\hat{w})$ denote the result of the construction based on the choice of f' we have by the homotopy formula

$$\hat{c}_i'(\hat{w})-\hat{c}_i'(\hat{w}) = a\Big(\int R(\hat{c}_i(\mathrm{pr}_M^*\,\hat{w}))\Big) = a\Big(\int \mathrm{pr}_M^*\,C_i(\hat{w})\Big)=0. \qquad\square$$

Lemma 2.2. *The construction of $\hat{c}_i$ defines a natural transformation $\hat{c}_i:\hat{K}\to\widehat{H\mathbb{Z}}^{2i}$ of set-valued functors on smooth manifolds.*

Proof. Let $g: N\to M$ be a smooth map between manifolds. Let $\hat{w}\in\hat{K}^0(M)$ and assume that we have constructed $\hat{c}_i(\hat{w})$ using the choices of $f: M\to\mathcal{K}_k$, $\rho\in\Omega^{\mathrm{odd}}(M)$ and $\tilde{\rho}\in\Omega^{\mathrm{odd}}([0,1]\times M)$. Then we construct $\hat{c}_i(g^*\hat{w})$ using the choices $f\circ g : N\to \mathcal{K}_k$, $g^*\rho\in\Omega^{\mathrm{odd}}(N)$ and $(\mathrm{id}\times g)^*\tilde\rho\in\Omega^{\mathrm{odd}}([0,1]\times N)$. With these choices we get $(\mathrm{id}\times g)^*\tilde{\hat{w}} = \widetilde{g^*\hat{w}}\in\hat{K}^0([0,1]\times N)$ and

$$\begin{aligned} g^*\hat{c}_i(\hat{w}) &= g^* f^*\hat{z}_{i,k} + g^* a\Big(\int_{[0,1]\times M/M} C_i(R(\tilde{\hat{w}}))\Big) \\ &= (f\circ g)^*\hat{z}_{i,k} + a\Big(\int_{[0,1]\times M/M} C_i(R((\mathrm{id}\times g)^*\tilde{\hat{w}}))\Big) \\ &= (f\circ g)^*\hat{z}_{i,k} + a\Big(\int_{[0,1]\times M/M} C_i(R(\widetilde{g^*\hat{w}}))\Big) = \hat{c}_i(g^*\hat{w}). \end{aligned}$$

This finishes the proof of Theorem 1.1(i).

To show the part (ii), we consider the natural transformation

$$\hat{B} : \hat{K}^0\times\hat{K}^0\to\widehat{H\mathbb{Z}}^{\mathrm{ev}}$$

given by

$$\hat{B}(\hat{w},\hat{v}) := \hat{c}(\hat{w})\cup\hat{c}(\hat{v}) - \hat{c}(\hat{w}+\hat{v})\in\widehat{H\mathbb{Z}}^{\mathrm{ev}}(M)\quad\text{for } \hat{w},\hat{v}\in\hat{K}^0(M).$$

If we apply I, we get

$$\begin{aligned} I(\hat{B}(\hat{w},\hat{v})) &= I(\hat{c}(\hat{w})\cup\hat{c}(\hat{v})) - I(\hat{c}(\hat{w}+\hat{v})) \\ &= I(\hat{c}(\hat{w}))\cup I(\hat{c}(\hat{v})) - I(\hat{c}(\hat{w}+\hat{v})) \\ &= c(I(\hat{w}))\cup c(I(\hat{v})) - c(I(\hat{w})+I(\hat{v})) = 0. \end{aligned}$$

Let $C = 1 + C_1 + C_2 + \cdots \in\mathbb{Q}[[s_0, s_1,\dots]]$. Then we have the identity

$$C(s_0+s_0', s_1+s_1',\dots) = C(s_0,s_1,\dots)C(s_0',s_1',\dots).$$

Indeed, if $\widetilde{\mathbf{ch}} = \sum_{i\ge1}(e^{x_i}-1)$, then $C(\mathbf{ch}_1,\dots) = \prod_{i\ge1}(1+x_i)$. If we introduce another set of variables x_i' and set $\widetilde{\mathbf{ch}}' = \sum_{i\ge1}(e^{x_i'}-1)$, then

$$\begin{aligned} C(\mathbf{ch}_1+\mathbf{ch}_1', \mathbf{ch}_2+\mathbf{ch}_2',\dots) &= \prod_{i\ge1}(1+x_i)(1+x_i') \\ &= C(\mathbf{ch}_1,\mathbf{ch}_2,\dots)C(\mathbf{ch}_1',\mathbf{ch}_2',\dots). \end{aligned}$$

We now calculate

$$\begin{aligned} R(\hat{B}(\hat{w},\hat{v})) &= R(\hat{c}(\hat{w})\cup\hat{c}(\hat{v})) - R(\hat{c}(\hat{w}+\hat{v})) \\ &= R(\hat{c}(\hat{w}))\cup R(\hat{c}(\hat{v})) - R(\hat{c}(\hat{w}+\hat{v})) \\ &= C(R(\hat{w}))\wedge C(R(\hat{v})) - C(R(\hat{w})+R(\hat{v})) = 0. \end{aligned}$$

Thus $\hat{B}$ factorizes over the subfunctor $H\mathbb{R}^{\mathrm{odd}}/H\mathbb{Z}^{\mathrm{odd}}\subset H\mathbb{R}/\mathbb{Z}^{\mathrm{odd}}\subset\widehat{H\mathbb{Z}}^{\mathrm{ev}}$, where the inclusion is induced by a. Let $\rho\in\Omega^{\mathrm{odd}}(M)$ and $\tilde\rho := t\,\mathrm{pr}_M^*\rho\in\Omega^{\mathrm{odd}}([0,1]\times M)$.

Then we have

$$\hat{B}(\hat{\omega}+a(\rho),\hat{\upsilon})-\hat{B}(\hat{\omega},\hat{\upsilon})=\hat{B}(\mathrm{pr}_M^*\,\hat{\omega}+a(\tilde{\rho}),\hat{\upsilon})|_{\{1\}\times M}-\hat{B}(\mathrm{pr}_M^*\,\hat{\omega}+a(\tilde{\rho}),\hat{\upsilon})|_{\{0\}\times M}.$$

Because $\hat{B}$ takes values in the homotopy invariant subfunctor $H\mathbb{R}^{\mathrm{odd}}/H\mathbb{Z}^{\mathrm{odd}}$, we conclude that $\hat{B}(\hat{\omega}+a(\rho),\hat{\upsilon})=\hat{B}(\hat{\omega},\hat{\upsilon})$. Similarly, we see that $\hat{B}(\hat{\omega},\hat{\upsilon}+a(\rho))=\hat{B}(\hat{\omega},\hat{\upsilon})$. Hence $\hat{B}$ has a factorization over a natural transformation

$$K^0\times K^0\to H\mathbb{R}^{\mathrm{odd}}/H\mathbb{Z}^{\mathrm{odd}}\subset H\mathbb{R}/\mathbb{Z}^{\mathrm{odd}}.$$

Such a natural transformation between homotopy invariant functors on manifolds must be represented by a map of classifying spaces

$$\boldsymbol{K}_0\times\boldsymbol{K}_0\to K(\mathbb{R}/\mathbb{Z},\mathrm{odd}),$$

where $K(\mathbb{R}/\mathbb{Z},\mathrm{odd}):=\bigvee_{i\geq 0}K(\mathbb{R}/\mathbb{Z},2i+1)$ is a wedge of Eilenberg–Mac Lane spaces, that is, by a class in $B\in H^{\mathrm{odd}}(\boldsymbol{K}_0\times\boldsymbol{K}_0;\mathbb{R}/\mathbb{Z})$. Since $\boldsymbol{K}_0$ and therefore $\boldsymbol{K}_0\times\boldsymbol{K}_0$ are even spaces, we know that $H_{\mathrm{odd}}(\boldsymbol{K}_0\times\boldsymbol{K}_0;\mathbb{Z})=0$. Then we have $H^{\mathrm{odd}}(\boldsymbol{K}_0\times\boldsymbol{K}_0;\mathbb{R}/\mathbb{Z})\cong\mathrm{Hom}(H_{\mathrm{odd}}(\boldsymbol{K}_0\times\boldsymbol{K}_0;\mathbb{Z}),\mathbb{R}/\mathbb{Z})=0$ by the universal coefficient formula. We see that $B=0$ and therefore $\hat{B}=0$. This finishes the proof of Theorem 1.1(ii). □

Proof of Theorem 1.2. We let $\hat{e}\in K^1(S^1)$ be, as above, the unique element with $R(\hat{e})=dt$, with $I(\hat{e})=e\in K^1(S^1)$ the canonical generator, and with $\hat{e}|_*=0$ for a basepoint $*\in S^1$. Then we define for odd $i\in\mathbb{N}$ and $\hat{x}\in\hat{K}^{-1}(M)$

$$\hat{c}_i^{\mathrm{odd}}(\hat{x}):=\int\hat{c}_{(i+1)/2}(\hat{e}\times\hat{x}).$$

Note that

$$I\Big(\int\hat{c}_{(i+1)/2}(\hat{e}\times\hat{x})\Big)=\int c_{(i+1)/2}(e\times I(\hat{x})).$$

We have a natural inclusion $\widetilde{H\mathbb{Z}}^*(\Sigma M_+)\subset H\mathbb{Z}^*(S^1\times M)$ since the subspace of classes whose restriction to $\{*\}\times M$ vanishes. Since $e|_*=0$, we see that $e\times I(\hat{x})$ belongs to this subspace. The restriction of $\int$ to this subspace coincides with the suspension isomorphism $\widetilde{H\mathbb{Z}}^{*+1}(\Sigma M_+)\xrightarrow{\sim}H\mathbb{Z}^*(M)$, we have $\int(e\times x)=x$ with inverse $x\mapsto e\times x$. Therefore

$$\int c_{(i+1)/2}(e\times I(\hat{x}))=c_i^{\mathrm{odd}}(I(\hat{x})).$$

In this way we get a natural transformation that has the required property.

Since $\int:\hat{K}^0(S^1\times M)\to\hat{K}^{-1}(M)$ is surjective it is clear that $\hat{c}_i^{\mathrm{odd}}$ is unique. □

References

[Berthomieu 2008] A. Berthomieu, "A version of smooth K-theory adapted to the total Chern class", preprint, 2008. arXiv 0806.4728

[Bunke 2009a] U. Bunke, "Adams operations in smooth K-theory", preprint, 2009. arXiv 0904.4355

[Bunke 2009b] U. Bunke, *Index theory, eta forms, and Deligne cohomology*, Mem. Amer. Math. Soc. **928**, American Mathematical Society, Providence, RI, 2009. MR 2010d:58023

[Bunke 2009c] U. Bunke, "String structures and trivialisations of a pfaffian line bundle", preprint, version 2, 2009. arXiv 0909.0846v2

[Bunke and Schick 2009] U. Bunke and T. Schick, "Smooth K-theory", pp. 45–135 in *Probability to Geometry, II*, Astérisque **328**, Société Mathématique de France, Paris, 2009.

[Bunke and Schick 2010] U. Bunke and T. Schick, "Uniqueness of smooth extensions of generalized cohomology theories", preprint, version 2, 2010. arXiv 0901.4423v2

[Cheeger and Simons 1985] J. Cheeger and J. Simons, "Differential characters and geometric invariants", pp. 50–80 in *Geometry and topology* (College Park, MD, 1983/84), edited by J. Alexander and J. Harer, Lecture Notes in Math. **1167**, Springer, Berlin, 1985. MR 87g:53059 Zbl 0621.57010

[Madsen 2009] I. Madsen, "An integral Riemann–Roch theorem for surface bundles", preprint, 2009. arXiv 0901.4240

Received July 15, 2009.

ULRICH BUNKE
NWF I - MATHEMATIK
UNIVERSITÄT REGENSBURG
93040 REGENSBURG
GERMANY
ulrich.bunke@mathematik.uni-regensburg.de
http://www.mathematik.uni-regensburg.de/Bunke/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

LAPLACIAN SPECTRUM FOR THE NILPOTENT KAC–MOODY LIE ALGEBRAS

DMITRY FUCHS AND CONSTANCE WILMARTH

We prove that a maximal nilpotent subalgebra of a Kac–Moody Lie algebra has an (essentially unique) Euclidean metric whose Laplace operator in the chain complex is scalar on each component of a given degree. Moreover, both the Lie algebra structure and the metric are uniquely determined by this property.

1. Introduction

Let $\mathfrak{g}$ be a real Lie algebra that is either finite-dimensional or has a grading $\mathfrak{g} = \bigoplus_{\boldsymbol{k}\in\mathbb{Z}^n} \mathfrak{g}^{(\boldsymbol{k})}$ such that all the chain spaces

$$C_q^{(\boldsymbol{k})}(\mathfrak{g}) = \bigoplus_{\boldsymbol{k}_1+\cdots+\boldsymbol{k}_q=\boldsymbol{k}} (\mathfrak{g}^{(\boldsymbol{k}_1)} \wedge \cdots \wedge \mathfrak{g}^{(\boldsymbol{k}_q)})$$

are finite-dimensional. (We consider only the case $\mathfrak{g} = \bigoplus_{(k_1,\dots,k_n)\succ(0,\dots,0)} \mathfrak{g}^{(k_1,\dots,k_n)}$ where the notation $(k_1, \dots, k_n) \succ (\ell_1, \dots, \ell_n)$ means that $k_1 \ge \ell_1, \dots, k_n \ge \ell_n$, and $(k_1, \dots, k_n) \ne (\ell_1, \dots, \ell_n)$, and all the spaces $\mathfrak{g}^{(k_1,\dots,k_n)}$ are finite-dimensional.) Suppose that for each value of $\boldsymbol{k}$, some Euclidean structure is fixed for $\mathfrak{g}^{(\boldsymbol{k})}$. Then Euclidean structures arise in all the chain spaces $C_q^{(\boldsymbol{k})}(\mathfrak{g})$, and they give rise to canonical isomorphisms between the chain spaces and the corresponding cochain spaces, $C^q_{(\boldsymbol{k})}(\mathfrak{g}) = (C_q^{(\boldsymbol{k})}(\mathfrak{g}))^*$. Thus, we can regard the boundary and coboundary operators as acting in the same spaces, that is,

$$\partial\colon C_q^{(\boldsymbol{k})}(\mathfrak{g}) \to C_{q-1}^{(\boldsymbol{k})}(\mathfrak{g}) \quad \text{and} \quad \delta\colon C_q^{(\boldsymbol{k})}(\mathfrak{g}) \to C_{q+1}^{(\boldsymbol{k})}(\mathfrak{g}),$$

and to form the *Laplace operators* $\Delta\colon C_q^{(\boldsymbol{k})}(\mathfrak{g}) \to C_q^{(\boldsymbol{k})}(\mathfrak{g})$. Chains (cochains) that are annihilated by Δ are called *harmonic*. The finite-dimensional version of the Hodge–de Rham theory yields the following result.

MSC2000: primary 17B56, 17B67; secondary 81R10.
Keywords: Kac–Moody algebras, Laplace operators.

Proposition 1. *Every harmonic chain (cochain) is a cycle (cocycle), and every homology (cohomology) class of* $\mathfrak{g}$ *with trivial coefficients is represented by a unique harmonic chain (cochain). In particular, there are canonical isomorphisms*

$$\operatorname{Ker}(\Delta\colon C_q^{(k)}(\mathfrak{g}) \to C_q^{(k)}(\mathfrak{g})) = H_q^{(\mathbf{k})}(\mathfrak{g}) = H^q_{(k)}(\mathfrak{g}).$$

For details, see [Fuchs 1986, Section 1.5.3]

Remark. The results discussed below indicate that not only the kernel but also the whole spectrum of the Laplacian must have significance for the (co)homology. We will return to this matter in subsequent publications.

The first computation of the spectrum of the Laplace operator in the cochain complex of an infinite-dimensional Lie algebra was done by I. M. Gel'fand, B. L. Feĭgin, and D. Fuchs (of this paper) [1978], with a subsequent important correction made by F. V. Weinstein [1993], for a maximal nilpotent subalgebra of the Virasoro algebra; this nilpotent Lie algebra is denoted as $L_1(1)$. It has a basis $\{e_i \mid i > 0\}$ with the commutator operation $[e_i, e_j] = (j-i)e_{i+j}$. We introduce in this algebra a $\mathbb{Z}$-grading and a Euclidean structure, letting $\deg e_i = i$ and $\|e_i\| = 1$. For positive integers $i_1, \dots, i_q$ such that $i_r - i_{r-1} \geq 3$ for $r = 2, \dots, q$, let

$$E(i_1, \dots, i_q) = \sum_{s=1}^{q} \binom{i_s}{3} - \sum_{1 \leq \ell < m \leq q} i_\ell i_m,$$

$$\alpha_r(i_1, \dots, i_q) = \begin{cases} 0 & \text{if } r = 1 \text{ and } i_1 < 3, \\ 1 & \text{if } r = 1 \text{ and } i_1 \geq 3, \\ 0 & \text{if } 1 < r \leq q \text{ and } i_r - i_{r-1} = 3, \\ 1 & \text{if } 1 < r \leq q \text{ and } i_r - i_{r-1} > 3, \end{cases}$$

$$\alpha(i_1, \dots, i_q) = \sum_{r=1}^{q} \alpha_r(i_1, \dots, i_q).$$

It is easy to check that $E(1, 4, 7, \dots, 3q-2) = E(2, 5, 8, \dots, 3q-1) = 0$, and all other values of the function E are positive.

Theorem 2 [Gel'fand et al. 1978; Weinstein 1993]. *The set of eigenvalues of the Laplace operator* $\Delta\colon C_*(L_1(1)) \to C_*(L_1(1))$ *coincides with the set of numbers* $E(i_1, \dots, i_q)$. *The multiplicity of the eigenvalue* $E(i_1, \dots, i_q)$ *equals* $2^{\alpha(i_1, \dots, i_q)}$. *(Occasional coincidences* $E(i_1, \dots, i_q) = E(i'_1, \dots, i'_{q'})$ *are possible; in such cases the multiplicities are added.)*

For a sketch of a proof see [Fuchs 1986, Section 2.3.1(B)].

S. Kumar [1984], using ideas from [Kostant 1963], calculated the spectrum of the Laplacian for a maximal nilpotent subalgebra of the Kac–Moody Lie algebra (see also the related works [Feĭgin 1980] and [Lepowsky 1979]). He noted that actually the Laplace operator is scalar in every homogeneous component with respect to the canonical $\mathbb{Z}^r$-grading (where r is the rank of the algebra), and derived a simple formula relating the eigenvalue to the degree.

Here, we develop a different approach to the calculating the Laplace spectrum. In so doing, we will not only obtain a fairly elementary proof of the Kumar formulas, but in addition prove that the relation between the Laplace eigenvalues and degrees uniquely determines both the Lie algebra structure and the Euclidean structure (which may be regarded as an alternative description of the class of Kac–Moody Lie algebras).

We supply below all the necessary definitions; for the general theory of Kac–Moody Lie algebras, see [Kac 1990].

Let $A = \|a_{ij}\|$ be an $n \times n$ matrix with all the diagonal entries equal to 2 and all nondiagonal entries being nonpositive integers. We assume the matrix A is *symmetrizable*, which means that there exists a diagonal matrix D whose diagonal entries $d_1, \dots, d_n$ are positive integers such that the matrix DA is symmetric. We may also assume the matrix A is *irreducible*, which means that there is no partition of $\{1, \dots, n\}$ into nonempty parts I, J such that $a_{ij} = 0$ for all $i \in I$ and $j \in J$. Let $G = G(A)$ be the (real) Kac–Moody Lie algebra with Cartan matrix A, and let $N = N(A)$ be the corresponding nilpotent Lie algebra. In other words, N has a system of generators $e_1, \dots, e_n$ with the defining set of relations $(\operatorname{ad} e_i)^{-a_{ij}+1} e_j = 0$. The algebra N has a natural $\mathbb{Z}^n$-grading $N = \bigoplus_{(k_1,\dots,k_n)\succ(0,\dots,0)} N^{(k_1,\dots,k_n)}$ where $N^{(k_1,\dots,k_n)}$ consists of linear combinations of commutator monomials of the generators involving precisely k_i letters e_i for $i = 1, \dots, n$. The following statement, proved in Section 2, is our main result; compare with [Kumar 1984, Theorem 2.1].

Theorem 3. *There exist unique Euclidean structures in the spaces $N^{(k_1,\dots,k_n)}$ such that $\|e_i\| = 1$ for $i = 1, \dots, n$ and such that the corresponding Laplace operator $\Delta\colon C_*^{(k_1,\dots,k_n)}(N) \to C_*^{(k_1,\dots,k_n)}(N)$ is the multiplication by*

$$E(k_1, \dots, k_n) = \sum_i d_i k_i - \frac{1}{2} \sum_{i,j} d_i a_{ij} k_i k_j.$$

Moreover, if a $\mathbb{Z}^r$-graded Lie algebra is generated by r generators of degrees

$$(1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$$

and the Laplace operator in its cochain complex with respect to some Euclidean structures in the homogeneous components is described by the formulas above, then there exists an isometric isomorphism between this Lie algebra and the nilpotent Kac–Moody Lie algebra N.

Proposition 4. *If $E(k) \neq 0$, then $H_*^{(k)}(N(A)) = 0$.*

This follows from Proposition 1 and Theorem 3.

Proposition 4 is not new: it is essentially contained in [Kac and Kazhdan 1979], which yields a description of a Bernstein–Gel'fand–Gel'fand resolution of the trivial module over a Kac–Moody Lie algebra. This is also a free resolution of the trivial module over $N(A)$.

2. Proof of Theorem 3

2.1. ***The Laplace operator has order 2.*** In the standard calculus, a linear operator $D: C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ is a differential operator of order 1 (that is, $D(f) = af' + bf$ where a and b are functions), if the identity

$$D(fg) = D(f)g + D(g)f - D(1)fg$$

holds for any functions f and g. Similarly, an operator of order 2 is characterized by the identity

$$\begin{aligned} &D(fgh) \\ &\quad = D(fg)h + D(fh)g + D(hg)f - D(f)gh - D(g)fh - D(h)fg + D(1)fgh \end{aligned}$$

(and so on, but we do not need operators of orders greater than 2). It is well known that the commutator of operators of order p and q has order $p + q - 1$.

In the noncommutative (supercommutative) case of chains/cochains of a Lie algebra (with a Euclidean structure), the notion of a differential order looks slightly different. In particular, the operator $\delta\colon C_*(\mathfrak{g}) \to C_*(\mathfrak{g})$ has order 1, meaning

$$\delta(c_1 \wedge c_2) = \delta(c_1) \wedge c_2 + (-1)^{d_1 d_2} \delta(c_2) \wedge c_1 \quad \text{for } c_i \in C_{d_i}(\mathfrak{g}).$$

However, the operator $\partial\colon C_*(\mathfrak{g}) \to C_*(\mathfrak{g})$ has order 2, which means that

$$\begin{aligned} &\partial(c_1 \wedge c_2 \wedge c_3) \\ &\quad = \partial(c_1 \wedge c_2) \wedge c_3 + (-1)^{d_2 d_3} \partial(c_1 \wedge c_3) \wedge c_2 + (-1)^{d_1(d_2+d_3)} \partial(c_2 \wedge c_3) \wedge c_1 \\ &\qquad - \partial(c_1) \wedge c_2 \wedge c_3 - (-1)^{d_1 d_2} \partial(c_2) \wedge c_1 \wedge c_3 - (-1)^{(d_1+d_2)d_3} \partial(c_3) \wedge c_1 \wedge c_2 \end{aligned}$$

for $c_i \in C_{d_i}(\mathfrak{g})$. Since the Laplace operator is a (super)commutator of ∂ and δ, it also has order 2, and we have the following lemma.

Lemma 5. *The Laplace operator* $\Delta\colon C_*(\mathfrak{g}) \to C_*(\mathfrak{g})$ *has order* 2, *that is*,

$$\begin{aligned} &\Delta(c_1 \wedge c_2 \wedge c_3) \\ &\quad = \Delta(c_1 \wedge c_2) \wedge c_3 + (-1)^{d_2 d_3} \Delta(c_1 \wedge c_3) \wedge c_2 + (-1)^{d_1(d_2+d_3)} \Delta(c_2 \wedge c_3) \wedge c_1 \\ &\qquad - \Delta(c_1) \wedge c_2 \wedge c_3 - (-1)^{d_1 d_2} \Delta(c_2) \wedge c_1 \wedge c_3 - (-1)^{(d_1+d_2)d_3} \Delta(c_3) \wedge c_1 \wedge c_2 \end{aligned}$$

for all $c_i \in C_{d_i}(\mathfrak{g})$.

Remark. It is important that Lemma 5 is compatible with Theorem 3 in the sense that if $\mathfrak{g} = N = N(A)$ and $c_i \in C_{d_i}^{(p_i)}(N)$, where $(p_i) = (p_{i1}, \dots, p_{in})$, then every term in the equality of Lemma 5 is $c_1 \wedge c_2 \wedge c_3$ times an approrpiate eigenvalue of Δ, and the equality becomes

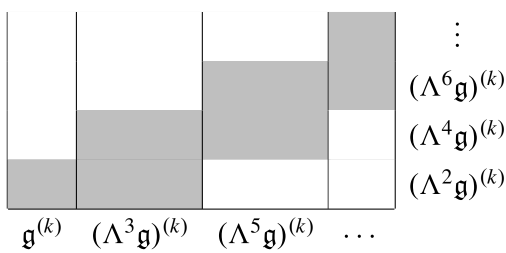
$$E(p_1+p_2+p_3)=E(p_1+p_2)+E(p_1+p_3)+E(p_2+p_3)-E(p_1)-E(p_2)-E(p_3),$$

which is true (because E is a polynomial of degree 2).

2.2. ***Construction of a Lie algebra with a given Laplace operator.*** Let A, a_{ij}, D and d_i be the same as in Section 1. We will now construct a graded Lie algebra $\mathfrak{g} = \bigoplus_{(k_1,\dots,k_n)\succ(0,\dots,0)} \mathfrak{g}^{(k_1,\dots,k_n)}$ with Euclidean structures in the summed-over (finite-dimensional) spaces satisfying the conclusion of Theorem 3 (with N replaced by $\mathfrak{g}$). We will see that $\mathfrak{g}$ is unique up to an isometric isomorphism, provided that $\dim \mathfrak{g}^{(1,0,\dots,0)} = \dim \mathfrak{g}^{(0,1,0,\dots,0)} = \dots = \dim \mathfrak{g}^{(0,\dots,0,1)} = 1$. (Later on, we will see that $\mathfrak{g}$=$N(A)$.)

First consider a given graded Lie algebra $\mathfrak{g}$ with Euclidean structures in $\mathfrak{g}^{(k_1,\dots,k_n)}$. Choose an orthonormal basis in each $\mathfrak{g}^{(k)}$, where $(k) = (k_1, \dots, k_n)$; then wedge products of the elements of the bases in $\mathfrak{g}^{(k)}$ form orthonormal bases in the chain spaces $(\Lambda^q \mathfrak{g})^{(k)}$. For a fixed $(k) \succ (0, \dots, 0)$, consider the matrix with rows (columns) labeled by the elements of our orthonormal bases in $(\Lambda^q \mathfrak{g})^{(k)}$ with q even (odd). Let the shaded blocks represent the boundary/coboundary operators in the chain/cochain complexes of $\mathfrak{g}$, and let the unshaded blocks be zero. To illustrate:

$D_{(k)} =$



Take two rows or two columns of the matrix $D_{(k)}$ corresponding to basis elements $c \in (\Lambda^q \mathfrak{g})^{(k)}$ for $c' \in (\Lambda^{q'} \mathfrak{g})^{(k)}$ (so q and q' have the same parity) and compute their dot product. If $|q'-q| > 2$, then this dot product is obviously zero. If $|q'-q| = 2$, it is also zero because of the relations $\partial \circ \partial = 0$ and $\delta \circ \delta = 0$. Finally, if $q' = q$, then this dot product is the coefficient at c' in $\Delta(c)$ (and the coefficient at c in $\Delta(c')$).

If the Laplace operator $\Delta\colon C_*^{(k)} \to C_*^{(k)}$ is multiplication by a positive number λ, then the dot product of every two different rows, as well as of every two different columns, is equal to zero, and the dot-square of every row or column is equal to λ; in other words, the whole matrix $D_{(k)}$ is an orthogonal matrix times $\sqrt{\lambda}$.

We are ready for the construction announced in the beginning of the section. We use the induction with respect to $|k| = k_1 + \dots + k_n$. We put $\dim \mathfrak{g}^{(0,\dots,0,1,0,\dots,0)} = 1$

and choose (arbitrarily) nonzero vectors $e_1 \in \mathfrak{g}^{(1,0,\dots,0)}, \dots, e_n \in \mathfrak{g}^{(0,\dots,0,1)}$ to have length 1. Take a $(k) = (k_1, \dots, k_n)$ where the k_i are nonnegative integers such that $k_1 + \dots + k_n > 1$. If $E(k) \le 0$, we put $\mathfrak{g}^{(k)} = 0$; let $E(k) > 0$. By the induction hypothesis, the matrix $D_{(k)}$ from above is fully determined, except the bottom left (shaded) block. Away from this block, the dot product of every two distinct rows or columns is zero, and the dot-square of every row or column is equal to $E(k)$. This follows from the identities $\partial \circ \partial = 0$, $\delta \circ \delta = 0$ and also from Lemma 5 and the remark after it, which implies that the Laplace operator $\Delta\colon C_q^{(k)} \to C_q^{(k)}$ with $q \ge 3$ (fully determined) is multiplication by $E(k)$. Thus, the columns of our matrix disjoint from the bottom left box are pairwise orthogonal and have dot-squares $E(k)$. We can construct the missing columns making the whole matrix an orthogonal matrix times $\sqrt{E(k)}$. Since the dot-squares of the rows above the bottom left block are already equal to $E(k)$, the new columns will be confined to this block. Thus, we will have a $\mathfrak{g}^{(k)}$ (with $\dim \mathfrak{g}^{(k)} = \sum_{q \ge 2,\text{ even}} \dim(\Lambda^q \mathfrak{g})^{(k)} - \sum_{q \ge 3,\text{ odd}} \dim(\Lambda^q \mathfrak{g})^{(k)}$) with a ready orthonormal basis, and the new box yields a bracket $[\,\cdot\,,\cdot\,]\colon (\Lambda^2 \mathfrak{g})^{(k)} \to \mathfrak{g}^{(k)}$. Moreover, the orthogonality of the columns of the new box to the columns of the box next to the right means precisely that this bracket satisfies the Jacobi identity. (Notice that it could happen that $\sum_{q \ge 2,\text{ even}} \dim(\Lambda^q \mathfrak{g})^{(k)} = \sum_{q \ge 3,\text{ odd}} \dim(\Lambda^q \mathfrak{g})^{(k)}$; in this case we do not need any new columns and may simply put $\mathfrak{g}^{(k)} = 0$.)

This completes the construction promised in the beginning of the section; the uniqueness is obvious.

End of the proof. It remains to prove that the Lie algebra $\mathfrak{g}$ of Section 2.2 is $N(A)$. This follows from three remarks.

First, it follows from the construction of Section 2.2 that if $(k_1, \dots, k_n) \succ (0, \dots, 0)$ and $k_1 + \dots + k_n \ge 1$, then the bracket mapping $[\,\cdot\,,\cdot\,]\colon (\Lambda^2 \mathfrak{g})^{(k)} \to \mathfrak{g}^{(k)}$ is onto; hence, $\mathfrak{g}$ (like $N(A)$) is generated by $e_1, \dots, e_n$.

Second, the defining relations $(\operatorname{ad} e_i)^{-a_{ij}+1} e_j = 0$ hold. Indeed, the degree (k) of $(\operatorname{ad} e_i)^{-a_{ij}+1} e_j$ is described by the equalities $k_i = -a_{ij} + 1$, $k_j = 1$ and $k_s = 0$ for $s \ne i, j$. Hence,

$$\begin{aligned} E(k) &= \sum d_i k_i - \tfrac{1}{2} \sum a_{ij} k_i k_j \\ &= d_i(-a_{ij}+1) + d_j - d_i(-a_{ij}+1)^2 - d_j - d_i a_{ij}(a_{ij}+1) \\ &= -d_i a_{ij} + d_i + d_j - d_i a_{ij}^2 + 2 d_i a_{ij} - d_i - d_j + d_i a_{ij}^2 - d_i a_{ij} = 0. \end{aligned}$$

By construction, this means that $\mathfrak{g}^{(k)} = 0$; hence $(\operatorname{ad} e_i)^{-a_{ij}+1} e_j = 0$. Thus, there is a graded epimorphism $N(A) \to \mathfrak{g}$.

Third, it is true that $\dim \mathfrak{g}^{(k)} = \dim N(A)^{(k)}$ for all (k). Indeed, for any (k) with $E(k) \ne 0$, the dimensions $\dim \mathfrak{g}^{(k)}$ are determined inductively from the relation $\sum (-1)^q \dim(\Lambda^q \mathfrak{g})^{(k)} = 0$. A similar relation, $\sum (-1)^q \dim(\Lambda^q N(A))^{(k)} = 0$ (for the same values of (k)), follows from Proposition 4 and the Euler–Poincaré lemma.

In addition to that, $\dim \mathfrak{g}^{(k)} = \dim N(A)^{(k)} = 1$ if $(k) = (0, \dots, 0, 1, 0, \dots, 0)$, and $\mathfrak{g}^{(k)} = N(A)^{(k)} = 0$ if $(k) = (k_1, \dots, k_n) \succ (0)$ with $k_1 + \dots + k_n > 1$, and $E(k) \le 0$. Hence, our epimorphism $N(A) \to \mathfrak{g}$ is actually an isomorphism.

3. Conclusion

3.1. ***Canonical basis in $N(A)$.*** The construction of Section 2.2 shows that a maximal nilpotent subalgebra of a Kac–Moody Lie algebra has a canonical Euclidean metric. The metric depends on the choice of generators of length 1, but the commutator relations do not depend on anything. For example, a maximal nilpotent subalgebra of the rank 2 exceptional Lie algebra G_2 has dimension 6. The Cartan matrix is $A = \left[\begin{smallmatrix} 2 & -1 \\ -3 & 2 \end{smallmatrix}\right]$. There is a basis $\{e_{0,1}, e_{1,0}, e_{1,1}, e_{1,2}, e_{1,3}, e_{2,3}\}$ in $N(A)$ with $\deg e_{i,j} = (i, j)$ and the commutation relations

$$[e_{0,1}, e_{1,0}] = \sqrt{3} e_{1,1}, \quad [e_{0,1}, e_{1,1}] = 2e_{1,2}, \quad [e_{0,1}, e_{1,2}] = \sqrt{3} e_{1,3},$$
$$[e_{1,0}, e_{1,3}] = \sqrt{3} e_{2,3}, \quad [e_{1,1}, e_{1,2}] = \sqrt{3} e_{2,3}.$$

If we regard this basis as orthonormal, then the Laplace operator in $C_*^{(p,q)}$ is the multiplication by $3p + q - 3p^2 - q^2 + 3pq$.

A more interesting example is provided by the twisted affine Kac–Moody Lie algebra $A_2^{(2)}$ with Cartan matrix $A = \left[\begin{smallmatrix} 2 & -1 \\ -4 & 2 \end{smallmatrix}\right]$. This Lie algebra (after factoring over the one-dimensional center) is embedded into the current Lie algebra $\mathfrak{sl}(3) \otimes \mathbb{R}[t, t^{-1}]$. It is well known that it has a basis e_i such that $[e_i, e_j] = \alpha_{ij} e_{i+j}$, where the numbers α_{ij} depend only on i, j mod 8 (see [Kac 1990, Exercise 8.16]).

The basis given in [Kac 1990] is not precisely our canonical basis; to get ours, we need to modify it by some coefficients:

$$e_{8s} = \sqrt{2} \begin{bmatrix} t^{2s} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -t^{2s} \end{bmatrix}, \qquad e_{8s+1} = 2 \begin{bmatrix} 0 & t^{2s} & 0 \\ 0 & 0 & t^{2s} \\ 0 & 0 & 0 \end{bmatrix},$$

$$e_{8s+2} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ t^{2s+1} & 0 & 0 \end{bmatrix}, \qquad e_{8s+3} = \begin{bmatrix} 0 & 0 & 0 \\ t^{2s+1} & 0 & 0 \\ 0 & -t^{2s+1} & 0 \end{bmatrix},$$

$$e_{8s+4} = \sqrt{\frac{2}{3}} \begin{bmatrix} t^{2s+1} & 0 & 0 \\ 0 & -2t^{2s+1} & 0 \\ 0 & 0 & t^{2s+1} \end{bmatrix}, \qquad e_{8s+5} = 2 \begin{bmatrix} 0 & t^{2s+1} & 0 \\ 0 & 0 & -t^{2s+1} \\ 0 & 0 & 0 \end{bmatrix},$$

$$e_{8s+6} = 4 \begin{bmatrix} 0 & 0 & t^{2s+1} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad e_{8s+7} = \begin{bmatrix} 0 & 0 & 0 \\ t^{2s+2} & 0 & 0 \\ 0 & t^{2s+2} & 0 \end{bmatrix}.$$

The commutation relations of this basis are given by the formula $[e_{8s+i}, e_{8s'+j}] = \alpha_{ij} e_{8(s+s')+(i+j)}$ for $1 \le i \le 8$ and $1 \le j \le 8$, with the 8×8 matrix $\|\alpha_{ij}\|$ being

$$\begin{bmatrix} 0 & 2 & \sqrt{6} & -\sqrt{6} & -2 & 0 & \sqrt{2} & -\sqrt{2} \\ -2 & 0 & 0 & 0 & 2 & -\sqrt{8} & 0 & \sqrt{8} \\ -\sqrt{6} & 0 & 0 & \sqrt{6} & -\sqrt{2} & 2 & -2 & \sqrt{2} \\ \sqrt{6} & 0 & -\sqrt{6} & 0 & \sqrt{6} & 0 & -\sqrt{6} & 0 \\ 2 & -2 & \sqrt{2} & -\sqrt{6} & 0 & 0 & \sqrt{6} & -\sqrt{2} \\ 0 & \sqrt{8} & -2 & 0 & 0 & 0 & 2 & -\sqrt{8} \\ -\sqrt{2} & 0 & 2 & \sqrt{6} & -\sqrt{6} & -2 & 0 & \sqrt{2} \\ \sqrt{2} & -\sqrt{8} & -\sqrt{2} & 0 & \sqrt{2} & \sqrt{8} & -\sqrt{2} & 0 \end{bmatrix}.$$

The natural grading of the Lie algebra $A_2^{(2)}$ is given by the rule that if $-1 \le s \le 6$, then

$$\deg e_{8n+s} = \begin{cases} (4n+s, 2n) & \text{if } s \le 1, \\ (4n+s-2, 2n+1) & \text{if } s > 1. \end{cases}$$

The Laplace operator $\Delta\colon C_*^{(p,q)} \to C_*^{(p,q)}$ with respect to the metric determined by the basis $\{e_i, i > 0\}$ is the multiplication by $4p+q-4p^2-q^2+4pq$.

3.2. ***Some remarks on the multiplicative structure in $H^*(N(A))$.*** It follows from our results (and actually can be proved directly) that there is a basis in $H^*(N(A))$ represented by uniquely chosen monomial cochains (that is, products of elements of the basis in $C^1(N(A)) = N(A)^*$ dual to our canonical basis). This gives rise to a description of the multiplication in $H^*(N(A))$, which, however, is not very explicit. Let us begin with a couple of simple remarks.

First, it follows from the description above that the multiplication in $H^*(N(A))$ is "square-free": the square of any cohomology class is zero.

Second, every monomial cochain representing a nonzero element of $H^*(N(A))$ should contain at least one factor from $C^1_{(0,\dots,0,1,0,\dots,0)}(N(A))$; this implies that the cohomological length of $H^*(N(A))$ does not exceed the rank of $G(A)$, that is, the size of A.

Third, in the finite-dimensional case, the multiplication in $H^*(N(A))$ satisfies the Poincaré duality: If a nonzero element $\alpha \in H^q(N(A))$ is represented by a monomial cochain $c_{i_1} \dots c_{i_q}$, then the complimentary monomial $c_{j_1} \dots c_{j_r}$, where $q+r = d = \dim N(A)$, also represents a nonzero cohomology class $\beta \in H^r(N(A))$, and $\alpha\beta$ is a nonzero element in $H^d(N(A)) \cong \mathbb{R}$. It follows from the preceding remark that in the finite-dimensional case of rank 2 there are no other nonzero products. (It seems likely that in the infinite-dimensional case of rank 2, the multiplication in $H^*(N(A))$ is trivial.)

Now, let us consider some examples. Let $N(A) = \mathfrak{n}(n)$ be the Lie algebra of (strictly) upper triangular $n \times n$ matrices, associated to the Cartan matrix

$$A = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}.$$

For this Lie algebra, $\dim \mathfrak{n}(n) = \frac{1}{2}n(n-1)$ and $\dim H^*(\mathfrak{n}(n)) = n!$. The basis in $H^*(\mathfrak{n}(n+1))$ is parametrized by the integral points of the ellipsoid $x_1^2 + \cdots + x_n^2 = x_1x_2 + x_2x_3 + \cdots + x_{n-1}x_n + x_1 + \cdots + x_n$, or, still better, by the elements of the Weyl group S_{n+1} whose action on the ellipsoid above is generated by the reflections $s_i(x_1, \ldots, x_n) = (x_1, \ldots, x_{i-1}, -x_i + x_{i-1} + x_{i+1} + 1, x_{i+1}, \ldots, x_n)$ for $i = 1, \ldots, n$ (in this formula, x_0 and x_{n+1} are taken to be zero). If $(p_1, \ldots, p_n) = \sigma(0, \ldots, 0)$ for $\sigma \in S_{n+1}$, then the corresponding cohomology class γ_σ is in $H^\ell(\mathfrak{n}(n+1))$, where ℓ is the length of σ. In this case, $(p_1, \ldots, p_n)$ has a unique presentation as the sum $\sum_{s=1}^q \{i_s, j_s\}$ of different points of the form

$$\{i, j\} = (0, \ldots, 0, \underset{(i)}{1}, \ldots, 1, \underset{(j)}{0}, \ldots, 0) \quad \text{for } 1 \le i < j \le n+1$$

and the class γ_σ is represented by the monomial cochain $c_{i_1,j_1} \cdots c_{i_q,j_q}$, where $c_{i,j}$ takes the value 1 on the one-entry matrix $E_{i,j}$ and takes the value 0 on all other such matrices. Moreover, if the presentations $\sigma(0, \ldots, 0) = \sum\{i_s, j_s\}$, $\sigma'(0, \ldots, 0) = \sum\{i'_t, j'_t\}$ are disjoint and $\sum\{i_s, j_s\} + \sum\{i'_t, j'_t\} = \tau(0, \ldots, 0)$, then $\gamma_\sigma\gamma_{\sigma'} = \gamma_\tau$; in all other cases, $\gamma_\sigma\gamma_{\sigma'} = 0$.

For example, there are 6 permutations in S_3: $\sigma_1 = (1, 2, 3)$, $\sigma_2 = (2, 1, 3)$, $\sigma_3 = (1, 3, 2)$, $\sigma_4 = (2, 3, 1)$, $\sigma_5 = (3, 1, 2)$ and $\sigma_6 = (3, 2, 1)$. Accordingly, there are 6 integral points on the ellipse $x^2 + y^2 - x - y - xy = 0$, given by

$$\begin{aligned} &\sigma_1(0,0) = (0,0), \quad \sigma_4(0,0) = (1,2) = (0,1) + (1,1), \\ &\sigma_2(0,0) = (1,0), \quad \sigma_5(0,0) = (2,1) = (1,0) + (1,1), \\ &\sigma_3(0,0) = (0,1), \quad \sigma_6(0,0) = (2,2) = (1,0) + (0,1) + (1,1), \end{aligned}$$

the cohomology of $\mathfrak{n}(3)$ is spanned by

$$\begin{aligned} \gamma_{\sigma_1} = 1 \in H^0(\mathfrak{n}(3)), &\qquad \gamma_{\sigma_2}, \gamma_{\sigma_3} \in H^1(\mathfrak{n}(3)), \\ \gamma_{\sigma_6} \in H^3(\mathfrak{n}(3)), &\qquad \gamma_{\sigma_4}, \gamma_{\sigma_5} \in H^2(\mathfrak{n}(3)), \end{aligned}$$

$\gamma_{\sigma_2}\gamma_{\sigma_4} = -\gamma_{\sigma_3}\gamma_{\sigma_5} = \gamma_{\sigma_6}$, and all other products of cohomology classes of positive dimensions are zero. Similarly for $\mathfrak{n}(4)$ (we write $\sigma_{(ijkl)}$ for the permutation (i, j, k, l)):

$$\sigma_{(1234)}(0,0,0) = (0,0,0), \quad \sigma_{(2134)}(0,0,0) = (1,0,0), \quad \sigma_{(1324)}(0,0,0) = (0,1,0),$$
$$\sigma_{(1243)}(0,0,0) = (0,0,1), \qquad \sigma_{(3214)}(0,0,0) = (1,0,0)+(0,1,0)+(1,1,0),$$
$$\sigma_{(2314)}(0,0,0) = (0,1,0)+(1,1,0), \quad \sigma_{(2341)}(0,0,0) = (0,0,1)+(0,1,1)+(1,1,1),$$
$$\sigma_{(3124)}(0,0,0) = (1,0,0)+(1,1,0), \quad \sigma_{(3142)}(0,0,0) = (1,0,0)+(0,0,1)+(1,1,1),$$
$$\sigma_{(2143)}(0,0,0) = (1,0,0)+(0,0,1), \quad \sigma_{(2413)}(0,0,0) = (0,1,0)+(1,1,0)+(0,1,1),$$
$$\sigma_{(1342)}(0,0,0) = (0,0,1)+(0,1,1), \quad \sigma_{(4123)}(0,0,0) = (1,0,0)+(1,1,0)+(1,1,1),$$
$$\sigma_{(1423)}(0,0,0) = (0,1,0)+(0,1,1), \quad \sigma_{(1432)}(0,0,0) = (0,1,0)+(0,0,1)+(0,1,1),$$
$$\sigma_{(3241)}(0,0,0) = (1,0,0)+(0,0,1)+(0,1,1)+(1,1,1),$$
$$\sigma_{(2431)}(0,0,0) = (0,1,0)+(0,0,1)+(0,1,1)+(1,1,1),$$
$$\sigma_{(3412)}(0,0,0) = (0,1,0)+(1,1,0)+(0,1,1)+(1,1,1),$$
$$\sigma_{(4213)}(0,0,0) = (1,0,0)+(0,1,0)+(1,1,0)+(1,1,1),$$
$$\sigma_{(4132)}(0,0,0) = (1,0,0)+(0,0,1)+(1,1,0)+(1,1,1),$$
$$\sigma_{(3421)}(0,0,0) = (0,1,0)+(0,0,1)+(1,1,0)+(0,1,1)+(1,1,1),$$
$$\sigma_{(4231)}(0,0,0) = (1,0,0)+(0,0,1)+(1,1,0)+(0,1,1)+(1,1,1),$$
$$\sigma_{(4312)}(0,0,0) = (1,0,0)+(0,1,0)+(1,1,0)+(0,1,1)+(1,1,1),$$
$$\sigma_{(4321)}(0,0,0) = (1,0,0)+(0,1,0)+(0,0,1)+(1,1,0)+(0,1,1)+(1,1,1).$$

The cohomology classes of the corresponding monomial cochains form a basis in the cohomology:

$$\gamma_{(1234)} = 1 \in H^0(\mathfrak{n}(4)),$$
$$\gamma_{(2134)}, \gamma_{(1324)}, \gamma_{(1243)} \in H^1(\mathfrak{n}(4)),$$
$$\gamma_{(2314)}, \gamma_{(3124)}, \gamma_{(2143)}, \gamma_{(1342)}, \gamma_{(1423)} \in H^2(\mathfrak{n}(4)),$$
$$\gamma_{(3214)}, \gamma_{(2341)}, \gamma_{(3142)}, \gamma_{(2413)}, \gamma_{(4123)}, \gamma_{(1432)} \in H^3(\mathfrak{n}(4)),$$
$$\gamma_{(3241)}, \gamma_{(2413)}, \gamma_{(3412)}, \gamma_{(4213)}, \gamma_{(4132)} \in H^4(\mathfrak{n}(4)),$$
$$\gamma_{(3421)}, \gamma_{(4231)}, \gamma_{(4312)} \in H^5(\mathfrak{n}(4)),$$
$$\gamma_{(4321)} \in H^6(\mathfrak{n}(4)).$$

The multiplication is described by the following relations:

$$\gamma_{(2134)}\gamma_{(1243)} = \gamma_{(2143)};$$
$$\gamma_{(2134)}\gamma_{(2314)} = -\gamma_{(1324)}\gamma_{(3124)} = \gamma_{(3214)},$$

$$\begin{aligned}
\gamma_{(1324)}\gamma_{(1342)} &= -\gamma_{(1243)}\gamma_{(1423)} = \gamma_{(1432)};\\
\gamma_{(2134)}\gamma_{(2341)} &= \gamma_{(3241)},\\
\gamma_{(1324)}\gamma_{(2341)} &= \gamma_{(2431)},\\
\gamma_{(1324)}\gamma_{(4123)} &= -\gamma_{(4213)},\\
\gamma_{(1243)}\gamma_{(4123)} &= -\gamma_{(4132)};\\
\gamma_{(1243)}\gamma_{(3412)} &= \gamma_{(2314)}\gamma_{(2341)} = -\gamma_{(3421)},\\
-\gamma_{(3124)}\gamma_{(2341)} &= \gamma_{(1342)}\gamma_{(4123)} = \gamma_{(4231)},\\
\gamma_{(2134)}\gamma_{(3412)} &= \gamma_{(1423)}\gamma_{(4123)} = \gamma_{(4312)};\\
-\gamma_{(2134)}\gamma_{(1243)}\gamma_{(3412)} &= -\gamma_{(2134)}\gamma_{(2314)}\gamma_{(2341)} = \gamma_{(1324)}\gamma_{(3124)}\gamma_{(2341)}\\
&= \gamma_{(1324)}\gamma_{(1342)}\gamma_{(4123)} = -\gamma_{(1243)}\gamma_{(1423)}\gamma_{(4123)}\\
&= \gamma_{(3142)}\gamma_{(2413)} = \gamma_{(4321)}.
\end{aligned}$$

Although the procedure always determines the multiplication in $H^*(N(A))$, it does not give a satisfactory explicit description even of the ring $H^*(\mathfrak{n}(n))$, for reasons unclear to us.

References

[Feĭgin 1980] B. L. Feĭgin, "Cohomology of groups and of algebras of flows", *Uspekhi Mat. Nauk* **35**:2 (1980), 225–226. In Russian; translated in *Russ. Math. Surv.* **35**:2 (1980), 239–240. MR 81h:17018 Zbl 0461.57014

[Fuchs 1986] D. B. Fuks, *Cohomology of infinite-dimensional Lie algebras*, Consultants Bureau, New York, 1986. MR 88b:17001 Zbl 0667.17005

[Gel'fand et al. 1978] I. M. Gel'fand, B. L. Feĭgin, and D. B. Fuks, "Cohomology of infinite-dimensional Lie algebras and Laplace operators", *Funktsional. Anal. i Prilozhen.* **12**:4 (1978), 1–5. In Russian; translated in *Func. Anal. and Appl.* **12** (1978), 243–247. MR 80i:58050 Zbl 0499.57001

[Kac 1990] V. G. Kac, *Infinite-dimensional Lie algebras*, 3rd ed., Cambridge University Press, 1990. MR 92k:17038 Zbl 0716.17022

[Kac and Kazhdan 1979] V. G. Kac and D. A. Kazhdan, "Structure of representations with highest weight of infinite-dimensional Lie algebras", *Adv. in Math.* **34**:1 (1979), 97–108. MR 81d:17004 Zbl 0427.17011

[Kostant 1963] B. Kostant, "Lie algebra cohomology and generalized Schubert cells", *Ann. of Math.* (2) **77** (1963), 72–144. MR 26 #266 Zbl 0134.03503

[Kumar 1984] S. Kumar, "Geometry of Schubert cells and cohomology of Kac–Moody Lie-algebras", *J. Differential Geom.* **20**:2 (1984), 389–431. MR 86j:17020 Zbl 0565.17007

[Lepowsky 1979] J. Lepowsky, "Generalized Verma modules, loop space cohomology and MacDonald-type identities", *Ann. Sci. École Norm. Sup.* (4) **12**:2 (1979), 169–234. MR 81a:17004 Zbl 0414.17007

[Weinstein 1993] F. V. Weinstein, "Filtering bases: a tool to compute cohomologies of abstract subalgebras of the Witt algebra", pp. 155–216 in *Unconventional Lie algebras*, edited by D. Fuchs, Adv. Soviet Math. **17**, Amer. Math. Soc., Providence, RI, 1993. MR 94k:17042 Zbl 0801.17022

Received December 5, 2008.

DMITRY FUCHS
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
ONE SHIELDS AVENUE
DAVIS, CA 95616
UNITED STATES

fuchs@math.ucdavis.edu

CONSTANCE WILMARTH
DEPARTMENT OF MATHEMATICS
NORTHWEST CHRISTIAN UNIVERSITY
828 EAST ELEVENTH AVE.
EUGENE, OR 97401
UNITED STATES

cwilmarth@northwestchristian.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

SIGMA THEORY AND TWISTED CONJUGACY CLASSES

DACIBERG GONÇALVES AND DESSISLAVA HRISTOVA KOCHLOUKOVA

Using Sigma theory we show that for large classes of groups G there is a subgroup H of finite index in $\mathrm{Aut}(G)$ such that for $\varphi \in H$ the Reidemeister number $R(\varphi)$ is infinite. This includes all finitely generated nonpolycyclic groups G that fall into one of the following classes: nilpotent-by-abelian groups of type FP_∞; groups G/G'' of finite Prüfer rank; groups G of type FP_2 without free nonabelian subgroups and with nonpolycyclic maximal metabelian quotient; some direct products of groups; or the pure symmetric automorphism group. Using a different argument we show that the result also holds for 1-ended nonabelian nonsurface limit groups. In some cases, such as with the generalized Thompson's groups $F_{n,0}$ and their finite direct products, $H = \mathrm{Aut}(G)$.

1. Introduction

We study the Reidemeister number $R(\varphi)$ for elements φ of subgroups of finite index in $\mathrm{Aut}(G)$ for large classes of groups G. The Reidemeister number $R(\varphi)$ counts the twisted conjugacy classes $\{ga\varphi(g)^{-1}\}_{g\in G}$ for $a \in G$. The study of $R(\varphi)$ began in the Nielsen–Reidemeister fixed point theory. Following [Taback and Wong 2007], we say a group G has the property R_∞ if $R(\varphi)$ is infinite for every automorphism φ of G. Many classes of groups have the property R_∞: nonelementary Gromov hyperbolic groups [Levitt and Lustig 2000; Fel'shtyn 2001], nonabelian generalized Baumslag–Solitar groups [Levitt 2007], saturated weakly branch groups (including the Grigorchuk group and the Gupta–Sidki group) [Fel'shtyn et al. 2008a], and the Thompson's group F [Bleak et al. 2008].

Fel'shtyn et al. [2008b] showed that for a finitely generated polyfree group G with nonzero Euler characteristic, there is a subgroup H of finite index in $\mathrm{Aut}(G)$ such that the Reidemeister number is infinite for every element of H.

In Sections 4a and 4c, we will show that to the groups in the list above that have the R_∞ property, we can add the generalized Thompson's groups $F_{n,0}$ and any finite direct product of such groups (with possibly different n).

MSC2010: primary 20F65, 20J05, 55M20; secondary 20E45.

Keywords: Reidemeister class, Thompson group, Sigma theory, automorphism of groups, R_∞ property, limit group.

Both authors are partially supported by “bolsa de produtividade de pesquisa” from CNPq, Brazil.

Furthermore we study large classes of groups G where it is possible to show that there is a subgroup H of finite index in $\mathrm{Aut}(G)$ such that $R(\varphi)$ is infinite for every $\varphi \in H$. These are all finitely generated nonpolycyclic groups G such that

(1) G is nilpotent-by-abelian of type FP_∞;

(2) G/G'' is of finite Prüfer rank but not polycyclic;

(3) G is of type FP_2 with no free nonabelian subgroups and its maximal metabelian quotient is not polycyclic;

(4) G is some direct product of groups;

(5) G is the pure symmetric automorphism group;

(6) G is a Houghton group H_n.

The proofs of these results, found in Section 4, rely on the structure of the invariant $\Sigma^1(G)$ for the groups G described above. In the preliminaries we detail the properties of this invariant. It appeared first in the study of finitely presented metabelian groups G [Bieri and Strebel 1980] and was later defined for any finitely generated group G [Bieri et al. 1987]. The geometric invariant $\Sigma^1(G)$ is invariant under the action of the automorphism group of G, which is fundamental for the proofs. In many cases we do not know the full automorphism group $\mathrm{Aut}(G)$, but using Sigma theory we can nevertheless obtain sufficient information on how $\mathrm{Aut}(G)$ can act on the abelianization of G.

In the final section we show that to the list above we can add

(7) G is any 1-ended nonabelian nonsurface limit group.

Orientable surface groups of genus at least 2 and nonorientable surface groups of genus at least 3 are Gromov hyperbolic [Gersten 1999, 3.17.5], and hence as mentioned before have R_∞. Nonorientable surface groups of genus 1 or 2 are not limit groups since in the first case they have torsion and in the second they are virtually abelian but not abelian.

Limits groups appeared in the solution of the Tarski problem by O. Kharlampovich, A. Myasnikov, and independently by Z. Sela. The definition of limit groups was suggested by Z. Sela, and it turned out that the class of limit groups coincides with the class of finitely generated, fully residually free groups studied by G. Baumslag, O. Kharlampovich, A. Myasnikov, and V. Remeslenikov. The class of limit groups includes finitely generated free groups, finite rank free abelian groups, and surface groups of Euler characteristic at most -2. Limit groups are CAT(0) [Alibegović and Bestvina 2006]. Unfortunately we cannot use $\Sigma^1(G)$ to study the Reidemeister numbers for a limit group G because the set $\Sigma^1(G)$ is empty [Kochloukova 2010]. Instead we use the structure theorem of $\mathrm{Aut}(G)$ for 1-ended limit groups G given in [Bumagin et al. 2007].

2. Preliminaries

2a. *Sigma invariants for finitely generated groups.* Let R be an associative ring with 1. All modules and actions in this paper are left ones. For an R-module M, we say that M has type FP_m if there is a projective resolution

$$\cdots \to P_i \to P_{i-1} \to \cdots \to P_0 \to M \to 0$$

such that all projective modules P_i are finitely generated for $i \leq m$. A group G is of type FP_m if the trivial module $\mathbb{Z}$ is of type FP_m as a $\mathbb{Z}G$-module. Every group is of type FP_0, and G is of type FP_1 if and only if G is finitely generated. If G is finitely presented (that is, there is a $K(G, 1)$ with a finite 2-skeleton), then G is of type FP_2. The converse does not hold [Bestvina and Brady 1997].

The homological invariants $\Sigma^m(G, \mathbb{Z})$ were defined in their current form in [Bieri and Renz 1988], but in the case that $m = 1$ and G is metabelian, they can be traced back (under different notation) to [Bieri and Strebel 1980], where they played a crucial part in classifying finitely presented metabelian groups.

For a general finitely generated group G, define the character sphere

$$S(G) = (\mathrm{Hom}_{\mathbb{Z}}(G, \mathbb{R}) \setminus \{0\})/\sim,$$

where $\mathbb{R}$ is considered a group via addition, and $\chi_1 \sim \chi_2$ if there is a positive real number r such that $r\chi_1 = \chi_2$. Write $[\chi]$ for the class of χ in $S(G)$, and define

$$\Sigma^m(G, \mathbb{Z}) = \{[\chi] \in S(G) \mid \mathbb{Z} \text{ is of type } FP_m \text{ as a } \mathbb{Z}G_\chi\text{-module}\},$$

where G_χ is the monoid $\{g \in G \mid \chi(g) \geq 0\}$. By definition,

$$\Sigma^m(G, \mathbb{Z})^c = S(G) \setminus \Sigma^m(G, \mathbb{Z}).$$

It is known that $\Sigma^m(G, \mathbb{Z})$ is always an open subset of $S(G)$ and can be empty (as in the case of a noncyclic free group). If $\Sigma^m(G, \mathbb{Z})$ is not empty, then G is of type FP_m, but the converse does not hold (for example, when G is noncyclic free). The following theorem shows that the invariants $\Sigma^m(G, \mathbb{Z})$ are responsible for the homological type of subgroups of G that contain the commutator.

Theorem 2.1 [Bieri and Renz 1988, Theorem B]. *If N is a subgroup of G that contains the commutator and G is of type FP_m, then N is FP_m if and only if $S(G, N) := \{[\chi] \in S(G) \mid \chi(N) = 0\} \subseteq \Sigma^m(G, \mathbb{Z})$.*

There is a homotopical version $\Sigma^m(G)$ of $\Sigma^m(G, \mathbb{Z})$. In the dimension $m = 1$ the invariants coincide, that is, $\Sigma^1(G, \mathbb{Z}) = \Sigma^1(G)$, and in general $\Sigma^m(G) \subseteq \Sigma^m(G, \mathbb{Z})$, but the inclusion can be strict for $m \geq 2$ [Meier et al. 1998]. As before, the superscript c means complement in $S(G)$.

In general it is hard to calculate all the invariants (both homological and homotopical versions for arbitrary dimension m). In this paper we apply some structural results about $\Sigma^1(G)^c$ to the theory of twisted conjugacy classes.

2b. ***A Sigma invariant for a metabelian group.*** In this section we discuss the original Bieri–Strebel invariant [1980]. This is the form that will be used in the proof of Theorem 4.6.

Let Q be a finitely generated abelian group, and let A be a finitely generated left $\mathbb{Z}Q$-module. If not otherwise stated, all modules in this paper are left ones. Then

$$\Sigma_A(Q) = \{[\chi] \in S(Q) \mid A \text{ is finitely generated as a } \mathbb{Z}Q_\chi\text{-module}\}$$

and

$$\Sigma_A^c(Q) = S(Q) \setminus \Sigma_A(Q).$$

Bieri and Strebel [1981, (1.5)] showed that $\Sigma_A^c(Q)$ is additive, that is, if

$$0 \to A_1 \to A \to A_2 \to 0$$

is a short exact sequence of finitely generated $\mathbb{Z}Q$-modules, then

$$\Sigma_A^c(Q) = \Sigma_{A_1}^c(Q) \cup \Sigma_{A_2}^c(Q).$$

It follows from the definitions that if χ is a real character of a finitely generated metabelian group G such that A is an abelian normal subgroup of G with $Q = G/A$ abelian, and if $\chi(A) \neq 0$, then $[\chi] \in \Sigma^1(G)$. If $\chi(A) = 0$, then $[\chi] \in \Sigma^1(G)^c$ if and only if $[\chi] \in \Sigma_A^c(Q)$ (here we consider χ as a character of Q since $\chi(A) = 0$).

Let I be the annihilator of A in $\mathbb{Z}Q$, that is, $I = \{\lambda \in \mathbb{Z}Q \mid \lambda A = 0\}$, and let $P_1, \dots, P_s$ be the minimal associated primes of A as a $\mathbb{Z}Q$-module. Then we have by [Bieri and Strebel 1981, (1.12)]

$$\Sigma_A^c(Q) = \Sigma_{\mathbb{Z}Q/I}^c(Q) = \bigcup_{1 \le i \le s} \Sigma_{\mathbb{Z}Q/P_i}^c(Q).$$

2c. ***Twisted conjugacy classes.*** Let $\varphi : G \to G$ be a group endomorphism. The set of equivalence classes $\{ga\varphi(g)^{-1}\}_{g \in G}$ for $a \in G$ is denoted by $\mathcal{R}[\varphi]$. The cardinality of $\mathcal{R}[\varphi]$ is denoted by $R(\varphi)$ and called the Reidemeister number of φ.

Let H be a normal subgroup of a group G invariant under an endomorphism φ. Denote by φ' the restriction of φ on H and by $\overline{\varphi}$ the endomorphism induced by φ on $Q = G/H$. Then by [Gonçalves 1998] and [Gonçalves and Wong 2003, (2.2)], there is an exact sequence of sets

(2-1) $$\mathcal{R}[\varphi'] \to \mathcal{R}[\varphi] \to \mathcal{R}[\overline{\varphi}] \to 0.$$

By [Jiang 1983, page 33], if G/H is abelian, nontrivial, torsion-free, and the map $\mathrm{id}_Q - \overline{\varphi}$ is not surjective (that is, $\overline{\varphi}$ is a linear operator with eigenvalue 1), then $R(\overline{\varphi})$ is infinite and consequently $R(\varphi)$ is infinite.

3. General results

A point $[\chi]$ of the character sphere $S(G)$ is called a discrete point (or a rational point) if $[\chi]=[\mu]$ for some character μ such that $\mathrm{Im}(\mu)=\mathbb{Z}$. In this case we call χ discrete or rational.

Lemma 3.1. *Let G be a finitely generated group such that*

$$\Sigma^1(G)^c = \{[\chi_1], \dots, [\chi_m]\}$$

is finite, nonempty, and contains only discrete points, and let φ be an automorphism of G. Let $N=\bigcap_i \mathrm{Ker}(\chi_i)$, let $V=\mathrm{Hom}_{\mathbb{Z}}(G/N,\mathbb{R})$, and let θ be the element of $\mathrm{End}_{\mathbb{Z}}(V)$ *induced by φ.*

Then θ permutes the images $\overline{\chi}_i$ of χ_i in V, where every χ_i is chosen so that the coordinates of $\overline{\chi}_i$ are integers with maximal common divisor 1.

Proof. Since $\Sigma^1(G)^c$ is invariant under automorphisms, we have $\theta(\overline{\chi}_i)=r_i\overline{\chi}_{\pi(i)}$, where π is some element of the symmetric group S_m and the r_i are positive real numbers. By assumption, the coordinates of the entries of $\overline{\chi}_i$ (with respect to some fixed basis of the free abelian group G/N) are coprime integers. Then every r_i is a positive integer.

We aim to show that $r_1=\cdots=r_m=1$. Indeed for $k=m!$, we have $\theta^k(\overline{\chi}_i)=\lambda_i\overline{\chi}_i$ for some positive integer λ_i divisible by r_i for every i. Now φ is invertible and N is a characteristic subgroup of G, so θ is an automorphism of V, and so θ^k is invertible. Furthermore $X=\{\overline{\chi}_i\}_{1\le i\le m}$ spans V. Then by picking up a subset Y of X that is a basis of V, we see that the operator θ^k is diagonalizable with eigenvalues λ_i (note that a fixed $\overline{\chi}_i$ can always be included in some Y). Then $\prod_{\overline{\chi}_i\in Y}\lambda_i=1$, so $\lambda_i=1$ for $\overline{\chi}_i\in Y$. Then $r_i=1$ as required. □

Theorem 3.2. *Let G be a finitely generated group such that*

$$\Sigma^1(G)^c = \{[\chi_1], \dots, [\chi_m]\}$$

is finite, nonempty, and contains only discrete points. Let $N=\bigcap_i \mathrm{Ker}(\chi_i)$, and let $V=\mathrm{Hom}_{\mathbb{Z}}(G/N,\mathbb{R})$. Suppose that the image of $\{\chi_1,\dots,\chi_m\}$ in V is a basis of V as an $\mathbb{R}$-vector space. Then $R(\varphi)$ is infinite for every $\varphi\in\mathrm{Aut}(G)$, that is, G has the property R_∞.

Proof. By the previous lemma, we can fix representatives χ_i (remember we can multiply any χ_i with a positive real number) so that the map $\theta\in\mathrm{End}_{\mathbb{Z}}(V)$, induced by φ, permutes the elements of $\{\overline{\chi}_1, \dots, \overline{\chi}_m\}$, where $\{\overline{\chi}_1, \dots, \overline{\chi}_m\}$ is the image of $\{\chi_1, \dots, \chi_m\}$ in V. Then

$$\theta(\overline{\chi}_1 + \cdots + \overline{\chi}_m) = \overline{\chi}_1 + \cdots + \overline{\chi}_m \quad \text{and} \quad \overline{\chi}_1 + \cdots + \overline{\chi}_m \neq 0$$

in V, that is, θ has eigenvalue 1. Then $R(\overline{\varphi})$ is infinite by the preliminaries from Section 2c; hence $R(\varphi)$ is infinite. □

Theorem 3.3. *Let G be a finitely generated group such that*

$$\Sigma^1(G)^c = \{[\chi_1], \dots, [\chi_m]\}$$

is finite, nonempty, and contains only discrete points. Let $N = \bigcap_i \operatorname{Ker}(\chi_i)$ and $V = \operatorname{Hom}_{\mathbb{Z}}(G/N, \mathbb{R})$. Suppose that the image of $\{\chi_1, \dots, \chi_m\}$ in V lies in an open half subspace. Then $R(\varphi)$ is infinite for every $\varphi \in \operatorname{Aut}(G)$, that is, G has the property R_∞.

Proof. As in the proof of the previous theorem, we can fix representatives χ_i such that $w = \overline{\chi}_1 + \cdots + \overline{\chi}_m$ is fixed by θ, where $\overline{\chi}_i$ is the image of χ_i in V. Since $\{\overline{\chi}_1, \dots, \overline{\chi}_m\}$ lies in an open half subspace of V, we deduce that $w \neq 0$. We continue as in the proof of the previous theorem. □

Corollary 3.4. *Let G be a finitely generated group such that*

$$\Sigma^1(G)^c = \{[\chi_1], \dots, [\chi_m]\}$$

is finite, nonempty, and contains only discrete points. Then there is a subgroup of finite index H of $\operatorname{Aut}(G)$ such that $R(\varphi)$ is infinite for every $\varphi \in H$.

Proof. By Lemma 3.1 the image of $\operatorname{Aut}(G)$ in $\operatorname{End}_{\mathbb{Z}}(V)$ permutes the characters $\overline{\chi}_1, \overline{\chi}_2, \dots, \overline{\chi}_m$, where the coordinates of $\overline{\chi}_i$ are integers with maximal common divisor 1, and $V = \operatorname{Hom}_{\mathbb{Z}}(G/N, \mathbb{R})$ is defined as before. Let H be the kernel of the action of $\operatorname{Aut}(G)$ on $\operatorname{End}_{\mathbb{Z}}(V)$. Then the elements of H fix every $\overline{\chi}_i$; in particular the map induced by φ on the maximal torsion-free abelian quotient of G has eigenvalue 1. We continue as in the proof of the previous theorem. □

Proposition 3.5. *Let G be a finitely generated nilpotent group, and let $\varphi \in \operatorname{Aut}(G)$ be such that $\varphi(g)$ is conjugate to g for some element $g \in G$ of infinite order. Then $R(\varphi)$ is infinite.*

Proof. We induct on the nilpotency length of G. The case when G is abelian is obvious and well known.

Suppose that G is nonabelian. Let $Z(G)$ be the center of G, so $\overline{G} = G/Z(G)$ is of nilpotency length strictly smaller than the nilpotency length of G.

Let $\overline{\varphi}$ be the automorphism of $\overline{G}$ induced by φ. If there is an element of $\overline{G}$ of infinite order that is sent by $\overline{\varphi}$ to a conjugate, then by induction $R(\overline{\varphi}) = \infty$, so $R(\varphi) = \infty$. We can assume that there isn't an element of $\overline{G}$ of infinite order that is sent by $\overline{\varphi}$ to a conjugate. The conditions of [Gonçalves and Wong 2003, Lemma 2.1] then hold, and hence it suffices to show that $R(\varphi') = \infty$ to deduce that $R(\varphi) = \infty$, where φ' is the restriction of φ to N.

Let $g \in G$ be an element of infinite order such that $\varphi(g) \in g^G$. By the preceding paragraph, the image of g in $\overline{G}$ has finite order. If $g \notin Z(G)$, some nontrivial power of g is in $Z(G)$, so we can assume that $g \in Z(G)$. Then $g^G = g$ and $\varphi(g) = g$.

Consider one twisted conjugated class

$$X_h = \{ah\varphi'(a)^{-1}\}_{a \in Z(G)} \quad \text{for some } h \in Z(G).$$

Because there is an element of infinite order in $Z(G)$ that is fixed by φ', we see that $\{a\varphi'(a)^{-1}\}_{a \in Z(G)}$ is a subgroup of infinite index in $Z(G)$; hence X_h is a coset class of a subgroup of infinite index in $Z(G)$. In particular this shows that we have an infinite number of twisted conjugacy classes X_h when $h \in Z(G)$, that is, $R(\varphi') = \infty$. □

4. Applications

4a. ***On the generalized Thompson's group and some soluble groups.*** For a group G of piecewise linear transformations of the unit interval, we say that G is irreducible if G does not fix a point of $(0, 1)$. There are two distinguished real characters ρ and λ of G defined by $\lambda(f) = \log_2 f'(0)$ and $\rho(f) = \log_2 f'(1)$.

Theorem 4.1. *Let G be a finitely generated, irreducible group of piecewise linear transformations of the unit interval; let ρ and λ defined above be nonzero rational nonantipodal characters such that $\lambda(\mathrm{Ker}(\rho)) = \mathrm{Im}(\lambda)$ and $\rho(\mathrm{Ker}(\lambda)) = \mathrm{Im}(\rho)$. Then $R(\varphi)$ is infinite for every $\varphi \in \mathrm{Aut}(G)$, that is, G has the property R_∞.*

Proof. By [Bieri et al. 1987, Theorem 8.1], the invariant $\Sigma^1(G)^c$ is equal to $\{[\rho], [\lambda]\}$, and by hypothesis the conditions of Theorem 3.2 are satisfied. □

The generalized Thompson's group $F_{n,0}$, where $n \geq 2$ (we use the notation of Brin and Guzmán [1998], which differs from the notation in [Brown 1987a]), is defined with the infinite presentation

$$\langle x_0, x_1, x_2, \ldots \mid x_j^{-1} x_i x_j = x_{j+n-1} \text{ for } 0 \leq j < i \rangle;$$

the case $n = 2$ is the classical Thompson's group F. The groups above were first defined and shown to be of type FP_∞ in [Brown 1987a]. Some other groups of PL functions on the interval are known to be of type FP_∞ [Stein 1992]. The automorphism group of $F_{n,0}$ is understood completely only for $n = 2$ [Brin 1996], and only some partial results are known for general $n > 2$ [Brin and Guzmán 1998]. Although we do not know much about the full automorphism group of $F_{n,0}$, we can still deduce the following result from Theorem 4.1. This answers a question asked in [Bleak et al. 2008], where it was shown that the classical Thompson's group F has the property R_∞.

Corollary 4.2. *The generalized Thompson's group $F_{n,0}$ has the property R_∞.*

Proof. The generalized Thompson's group $F_{n,0}$ has several presentations as a group of PL functions of different intervals. In [Brin and Guzmán 1998, Lemma 2.3.1], a description of $F_{n,0}$ is given as the group of PL functions f on the interval $[0, n-1]$, where the set Y_f of break points is finite, the image of Y_f under f is inside $\mathbb{Z}[1/n]$, the slopes of all linear functions are integral powers of n, and the slopes of f at 0 and $n-1$ can be arbitrary integral powers of n. Furthermore $F_{n,0}$ does not fix a point on the open interval $(0, n-1)$. In that paper, the elements of $F_{n,0}$ act on the right of the interval, but by setting $f(x) = (x) f^{-1}$, we get an action on the left. By obvious rescaling (that is, a linear map that gives a bijection between the interval $[0, n-1]$ and $[0, 1]$), we get a description of $F_{n,0}$ as a subgroup of PL functions acting on the left of the interval $[0, 1]$, and then we can apply Theorem 4.1. □

All Σ-invariants are calculated [Bieri et al. 2010] for the original Thompson's group F, but for the generalized ones these invariants are not known except in dimension 1.

Theorem 4.3. *Let G be a nonpolycyclic nilpotent-by-abelian group of type FP_∞. Then $R(\varphi)$ is infinite for every $\varphi \in \mathrm{Aut}(G)$, that is, G has the property R_∞.*

Proof. Soluble groups of type FP_∞ are nilpotent-by-abelian-by-finite, so assuming that G is nilpotent-by-abelian is not a very strong restriction.

By the classification of soluble groups of type FP_∞, G is constructible, that is, built from the trivial group by finite and HNN extensions [Kropholler 1993].

Suppose that the commutator G' is not finitely generated; hence $\Sigma^1(G)^c$ is not empty by Theorem 2.1. By [Bieri and Strebel 1982], $\Sigma^1(G)^c$ is a finite set of discrete points $\{[\chi_1], \dots, [\chi_m]\}$ that lies in an open hemisphere of $S(G)$. Then we can apply Theorem 3.3.

Finally if the commutator G' is finitely generated, G is polycyclic since it is nilpotent, a contradiction. □

A group G is said to be of finite Prüfer rank if there is a number d such that every finitely generated subgroup is generated by at most d elements. In particular if $1 \to A \to G \to Q \to 1$ is a short exact sequence of groups with A and Q abelian, and G finitely generated, then G has finite Prüfer rank if and only if the torsion part of A is finite and $A \otimes_{\mathbb{Z}} \mathbb{Q}$ is finite-dimensional as a $\mathbb{Q}$-vector space. By the link between valuation theory and $\Sigma_A^c(Q)$ [Bieri and Groves 1984, Theorem 8.1], by the existing bijection between $\Sigma_A^c(Q)$ and $\Sigma^1(G)^c$ explained in the penultimate paragraph of Section 2b, and because G has finite Prüfer rank, $\Sigma^1(G)^c$ is a finite set $\{[\chi_1], \dots, [\chi_m]\}$, where $\chi_1, \dots, \chi_m$ are discrete characters.

Lemma 4.4. *Let G be a finitely generated metabelian group of finite Prüfer rank that is not polycyclic. Then there is a subgroup of finite index H of* $\mathrm{Aut}(G)$ *such that $R(\varphi)$ is infinite for every $\varphi \in H$.*

Proof. Since G is not polycyclic, G' is not finitely generated, and so $\Sigma^1(G)^c$ is not empty by Theorem 2.1. On the other hand $\Sigma^1(G)^c$ is a finite set of discrete points. Then we can apply Corollary 3.4. □

Corollary 4.5. *Let G be a finitely generated group such that G/G'' is of finite Prüfer rank and G/G'' is not polycyclic. Then there is a subgroup of finite index H of* $\mathrm{Aut}(G)$ *such that $R(\varphi)$ is infinite for every $\varphi \in H$.*

Proof. By the previous lemma there is a subgroup H_0 of finite index in $\mathrm{Aut}(G/G'')$ such that for every $\varphi_0 \in H_0$ the number $R(\varphi_0)$ is infinite. Then we can define H as the full preimage of H_0 in $\mathrm{Aut}(G)$. □

4b. *More on groups with metabelian quotients of type FP_2.*

Theorem 4.6. *Let G be a finitely generated metabelian group such that G is not polycyclic and $\Sigma^1(G)^c$ is not a finite union of subspheres. Then there is a subgroup of finite index H of* $\mathrm{Aut}(G)$ *such that $R(\varphi)$ is infinite for every $\varphi \in H$.*

Proof. As before since G is not polycyclic (in this case, this is equivalent to G' not being finitely generated) $\Sigma^1(G)^c$ is not empty by Theorem 2.1.

We view $A = G'$ as a left $\mathbb{Z}Q$-module via conjugation, where $Q = G/G'$; we denote this $\mathbb{Z}Q$-action on A by $\circ$, that is, if $q = gA$ and $a \in A$, then $q \circ a = gag^{-1}$.

Let $P_1, \dots, P_s$ be all minimal associated primes of the $\mathbb{Z}Q$-module A; see [Bourbaki 1989]. Let I be the annihilator of $\mathbb{Z}Q$ in A, that is,

$$I = \{\lambda \in \mathbb{Z}Q \mid \lambda \circ A = 0\}.$$

Then $P_1, \dots, P_s$ are the minimal prime ideals above I. Let $\overline{\varphi}$ be the automorphism of $Q = G/G'$ induced by φ. We extend $\overline{\varphi}$ by linearity to a ring endomorphism $\hat{\varphi}$ of $\mathbb{Z}Q$ so that the restriction on $\mathbb{Z}$ is identity, and note that for $\lambda \in I$ we have $0_A = \varphi(0_A) = \varphi(\lambda \circ A) = \hat{\varphi}(\lambda) \circ \varphi(A) = \hat{\varphi}(\lambda) \circ A$, so $\hat{\varphi}(I) = I$. Then $\hat{\varphi}$ permutes the finite set of prime ideals $P_1, \dots, P_s$. Thus by going down to a subgroup of finite index in $\mathrm{Aut}(G)$, we can assume that $\hat{\varphi}(P_i) = P_i$ for every i, and so $\hat{\varphi}$ induces an automorphism φ_i of $R_i = \mathbb{Z}Q/P_i$. By construction, φ_i is the identity on the image of $\mathbb{Z}$ in R_i.

Let $\tilde{\varphi}_i$ be the extension of φ_i to the field of fractions K_i of R_i. Let

$$Q_i = Q/T_i, \quad \text{where } T_i = \{q \in Q \mid q - 1 \in P_i\}.$$

By [Roseblade 1978, Theorem D] or [Bieri and Groves 1986, corollary on p. 426], the group of automorphisms of K_i that fixes the prime subfield k_i and permutes the elements of Q_i induces a finite group on B_i/A_i, where B_i is a minimal subgroup of

Q_i such that $k_i(B_i) \subseteq K_i$ is a purely transcendental extension with rank that equals the rank of the torsion-free group Q_i/B_i, and A_i is the subgroup of elements of Q_i that are algebraic over k_i. The group of automorphisms of K_i induces a finite group on A_i since the Galois group of a finite field extension is finite. Thus the group of automorphisms of K_i that fixes the prime subfield k_i and permutes the elements of Q_i induces a finite group on B_i.

Then by going down to a subgroup of finite index H in $\mathrm{Aut}(G)$, that is, $\varphi \in H$, we can assume that φ_i is trivial on B_i, that is, for $B_i = S_i/T_i$ with $S_i \leq Q$ we have $\varphi(s) \in sT_i$ for every $s \in S_i$. Then if B_i is infinite, the map induced by φ on $G/G' \otimes_{\mathbb{Z}} \mathbb{Q}$ has eigenvalue 1, and we can apply the observations from Section 2c. Thus we can assume henceforth that the B_i are all finite.

We claim that if B_i is finite, then

$$\Sigma^c_{\mathbb{Z}Q/P_i}(Q) = \{[\chi] \in S(Q) \mid \chi(S_i) = 0\} =: S(Q, S_i),$$

where S_i is defined as above, that is, S_i is the subgroup of Q such that $S_i/T_i = B_i$. Indeed by [Bieri and Groves 1984, Theorem 8.1], the class $[\chi]$ is in $\Sigma^c_{\mathbb{Z}Q/P_i}(Q)$ if and only if there is a real valuation of $\mathbb{Z}Q/P_i$ whose restriction on $Q_i = Q/T_i$ is induced by χ. Since K_i is a purely transcendental extension of $k_i(B_i)$ with degree exactly the rank of the torsion-free abelian group Q_i/B_i for every real valuation w of $k_i(B_i)$ and any character $\chi : Q_i \to \mathbb{R}$ such that χ and w coincide on B_i, there is a valuation of K_i that extends both w and χ. Furthermore any real valuation of $k(B_i)$ sends the finite group B_i to zero. Thus any real character of Q_i extends to a real valuation of K_i and so by restriction to a real valuation of $\mathbb{Z}Q/P_i$. Since $T_i - 1$ maps to 0 in $\mathbb{Z}Q/P_i$ and any real valuation of $\mathbb{Z}Q/P_i$ sends 0 to infinity for $[\chi] \in \Sigma^c_{\mathbb{Z}Q/P_i}(Q)$, we always have $\chi(T_i) = \chi(1) = 0$. Since B_i is finite for every real character χ such that $\chi(T_i) = 0$, we have $\chi(S_i) = 0$.

Then as mentioned in Section 2b,

$$\Sigma^1(G)^c = \Sigma^c_A(Q) = \Sigma^c_{\mathbb{Z}Q/I}(Q) = \bigcup_i \Sigma^c_{\mathbb{Z}Q/P_i}(Q).$$

That is, $\Sigma^1(G)^c$ is a finite union of subspheres $S(Q, S_i)$, a contradiction. □

Corollary 4.7. *Let G be a group of type FP_2. Suppose it does not contain a free subgroup of rank 2, and G/G'' is not polycyclic. Then there is a subgroup of finite index H of* $\mathrm{Aut}(G)$ *such that $R(\varphi)$ is infinite for every $\varphi \in H$.*

Proof. By [Bieri and Strebel 1980, Theorem 5.5], the maximal metabelian quotient G/G'' of G has type FP_2, and $\Sigma^1(G/G'')^c$ does not have antipodal points. By Theorem 2.1, since G/G'' is not polycyclic, $\Sigma^1(G/G'')^c$ is not empty. Since any subsphere of $S(G)$ has antipodal points, $\Sigma^1(G/G'')^c$ does not contain a subsphere. Then by Theorem 4.6 there is a subgroup H_0 of finite index in $\mathrm{Aut}(G/G'')$ such that $R(\varphi_0)$ is infinite for every $\varphi_0 \in H_0$. Finally define H to be the full preimage of H_0 in $\mathrm{Aut}(G)$. □

4c. *Direct products of groups.* The direct product of groups is a particular case of a more general construction: A graph product G of groups is constructed from a graph Δ and vertex groups G_v associated with every vertex v in the set V of vertices of Δ. Then G is the free product of all G_v factored by the normal subgroup generated by $[G_v, G_w]$, where v and w are any adjacent vertices in Δ.

Theorem 4.8 [Meinert 1995, Theorem A]. *Let G be a graph product with a finite underlying graph Δ and finitely generated vertex groups $\{G_v\}_{v\in V}$. Let $\chi : G \to \mathbb{R}$ be a nonzero homomorphism. Let Δ' be the full subgraph of Δ spanned by all vertices $V' = \{v \in V \mid \chi(G_v) \neq 0\}$. Then $[\chi] \in \Sigma^1(G)$ if and only if one of the following conditions holds*:

(1) *If $\Delta' = \{v_0\}$, then $[\chi|_{G_{v_0}}] \in \Sigma^1(G_{v_0})$, and each vertex of $V \setminus \{v_0\}$ is adjacent to v_0.*

(2) *If Δ' has at least 2 vertices, then it is connected, and each vertex of $V \setminus V'$ is adjacent to some vertex of Δ'.*

If applied to a finite graph Δ in which every pair of vertices is linked by an edge, then this theorem gives a simple formula for the Σ^1 invariant of a direct product of finitely many groups. There were some attempts to prove analogues for higher-dimensional invariants $\Sigma^m(G, \mathbb{Z})$ and $\Sigma^m(G)$ for G a direct product of groups (such analogues are called *direct product conjectures*). These conjectures turn out to be false in both homological and homotopical versions [Meier et al. 1998; Schütz 2008], but hold if we consider invariants with coefficients in a field [Bieri and Geoghegan 2010].

Corollary 4.9. *Let $H = G_1 \times \cdots \times G_m$, and let $\Sigma^1(G_i)^c$ be finite (possibly empty) and contain only rational points (if any) for all $1 \leq i \leq n$. Suppose further that $\Sigma^1(G_i)^c$ is not empty for at least one i. Then there is a subgroup of finite index H_0 of* $\mathrm{Aut}(H)$ *such that $R(\varphi)$ is infinite for every $\varphi \in H_0$.*

Proof. By Theorem 4.8, $\Sigma^1(G)^c$ is finite, nonempty, and contains only discrete points. Thus we can apply Corollary 3.4. □

Corollary 4.10. *Let $H = G_1 \times \cdots \times G_m$, where G_i are generalized Thompson's groups (not necessarily isomorphic). Then H has the property R_∞.*

Proof. Let ρ_i and λ_i be the characters of H that are zero when restricted to G_j for $j \neq i$, and are the classical ρ and λ from Section 4 when restricted to G_i. By Theorem 4.8, $\Sigma^1(H)^c = \{[\lambda_i], [\rho_i]\}_{1\leq i\leq m}$. Then we can apply Theorem 3.2. □

Examples. Here we exhibit groups G that have abelianization $\mathbb{Z}$ and nonempty $\Sigma^1(G)^c$. Let G be a knot group with commutator G' infinitely generated. For every knot group, we have $G/G' \simeq \mathbb{Z}$. By Theorem 2.1, since G' is infinitely generated, $\Sigma^1(G) \neq S(G)$.

Let D be the double of G, that is, we make the free amalgamated product $G *_{\mathbb{Z}^2} G$, where $\mathbb{Z}^2$ is the torus subgroup of G. Thus D is a 3-manifold group and $D/D' \simeq \mathbb{Z}$. By the results of Thurston [1986] for 3-manifold groups and their translation into algebraic language [Bieri et al. 1987, Corollary F], we have $\Sigma^1(D) = -\Sigma^1(D)$, and $\Sigma^1(D)$ is nonempty if and only if the 3-manifold fibers over S^1; in our case this means that D' is finitely generated. This obviously does not hold as G is a quotient of D, and G' is not finitely generated. Then $\Sigma^1(D)$ is empty. Since $D/D' \simeq \mathbb{Z}$, we see that $\Sigma^1(D)^c = S(D)$ has 2 elements.

4d. ***The pure symmetric automorphism group.*** In this section, G is the pure symmetric automorphism group of the free group F with basis $\{x_1, \dots, x_n\}$. It is generated by $\{\alpha_{i,j}\}_{1 \le i \neq j \le n}$, where

$$\alpha_{i,j}(x_i) = x_j^{-1} x_i x_j \quad \text{and} \quad \alpha_{i,j}(x_k) = x_k \quad \text{for } k \neq i.$$

By [McCool 1986], G is finitely presented with relations

$$[\alpha_{i,j}, \alpha_{k,l}] = 1 \qquad \text{if } i, j, k, l \text{ are different,}$$
$$[\alpha_{i,j}, \alpha_{k,j}] = 1 \quad \text{and} \quad [\alpha_{i,j}\alpha_{k,j}, \alpha_{i,k}] = 1 \qquad \text{if } i, j, k \text{ are different.}$$

By [Brady et al. 2001], G is a duality group of dimension $n-1$. By the main result of [Orlandi-Korner 2000], the class $[\chi]$ is in $\Sigma^1(G)^c$ if and only if exactly one of the following holds:

(1) $\chi(\alpha_{i,j}), \chi(\alpha_{j,i}) \in \mathbb{R}$ for some $1 \le i < j \le n$ and $\chi(\alpha_{r,s}) = 0$ for the remaining indices;

(2) $\chi(\alpha_{j,i}) = -\chi(\alpha_{k,i})$, $\chi(\alpha_{i,k}) = -\chi(\alpha_{j,k})$, and $\chi(\alpha_{k,j}) = -\chi(\alpha_{i,j})$ for some $1 \le i < j < k \le n$ and $\chi(\alpha_{r,s}) = 0$ for the remaining indices.

We write $A_{i,j}$ for the set of characters of first type and $B_{i,j,k}$ for the set of second type (in both cases including the trivial character). Note that $A_{i,j}$ and $B_{i,j,k}$ are $\mathbb{R}$-subspaces of $\operatorname{Hom}_{\mathbb{Z}}(G, \mathbb{R})$ of dimensions 2 and 3, respectively.

Theorem 4.11. *Let G be the pure symmetric automorphism group of a free group of rank n. Then there is a subgroup H of finite index in* $\operatorname{Aut}(G)$ *such that $R(\varphi)$ is infinite for every $\varphi \in H$.*

Proof. Since $\operatorname{Aut}(G)$ permutes

$$W = \Big(\bigcup_{1 \le i < j \le n} A_{i,j}\Big) \cup \Big(\bigcup_{1 \le i < j < k \le n} B_{i,j,k}\Big) \subset \operatorname{Hom}_{\mathbb{Z}}(G, \mathbb{R}),$$

and the only 3-dimensional spaces inside W are the $B_{i,j,k}$, we see that $\mathrm{Aut}(G)$ permutes the spaces $B_{i,j,k}$ and thus also the elements of $M = \bigcup_{1 \le i < j \le n} A_{i,j}$. Also, the only 2-dimensional spaces inside M are the $A_{i,j}$, so $\mathrm{Aut}(G)$ permutes the spaces $A_{i,j}$. By going down to a subgroup of finite index in $\mathrm{Aut}(G)$, we can consider only automorphisms φ of G such that φ sends every $A_{i,j}$ to $A_{i,j}$ and every $B_{i,j,k}$ to $B_{i,j,k}$.

Let $\{\alpha^*_{i,j}\}_{i,j}$ be the basis of $\mathrm{Hom}(G, \mathbb{R})$ dual to $\{\alpha_{i,j}\}_{i,j}$ and φ^* be the automorphism of $\mathrm{Hom}(G, \mathbb{R})$ induced by φ. Note that

$$B_{i,j,k} = \mathbb{R}(\alpha^*_{i,j} - \alpha^*_{k,j}) + \mathbb{R}(\alpha^*_{i,k} - \alpha^*_{j,k}) + \mathbb{R}(\alpha^*_{k,i} - \alpha^*_{j,i})$$

and $B_{i,j,k}$ is invariant under φ^*. Then

$$\varphi^*(\alpha^*_{i,j} - \alpha^*_{k,j}) \in B_{i,j,k}. \tag{4-1}$$

On the other hand, since $A_{i,j}$ and $A_{k,j}$ are invariant under φ^*, we have

$$\varphi^*(\alpha^*_{i,j}) \in \mathbb{R}\alpha^*_{i,j} + \mathbb{R}\alpha^*_{j,i} \quad \text{and} \quad \varphi^*(\alpha^*_{k,j}) \in \mathbb{R}\alpha^*_{k,j} + \mathbb{R}\alpha^*_{j,k}; \tag{4-2}$$

hence

$$\varphi^*(\alpha^*_{i,j} - \alpha^*_{k,j}) \in \mathbb{R}\alpha^*_{i,j} + \mathbb{R}\alpha^*_{j,i} + \mathbb{R}\alpha^*_{k,j} + \mathbb{R}\alpha^*_{j,k}. \tag{4-3}$$

Combining (4-1) and (4-3), we get

$$\varphi^*(\alpha^*_{i,j} - \alpha^*_{k,j}) = a\alpha^*_{i,j} - a\alpha^*_{k,j} \quad \text{for some } a \in \mathbb{R}.$$

Then by (4-2),

$$\varphi^*(\alpha^*_{i,j}) = a\alpha^*_{i,j} \quad \text{and} \quad \varphi^*(\alpha^*_{k,j}) = a\alpha^*_{k,j}.$$

In particular, φ induces a diagonal linear map $\bar{\varphi}$ on the abelianization G/G', and since φ is invertible, the product of eigenvalues is either 1 or -1, and all a are integers. Then a is either 1 or -1, and either $\bar{\varphi}$ has eigenvalue 1 and we are done by the remarks from Section 2c, or $\bar{\varphi}$ is $-\mathrm{id}_{G/G'}$. By going to a subgroup of index 2, we can avoid the last case. □

4e. *Houghton groups.* Let $\mathbb{N} = \{1, 2, \dots\}$ be the set of the natural numbers, and let $M = \{1, 2, \dots, n\} \times \mathbb{N}$ for $n \in \mathbb{N}$. The Houghton group H_n consists of all permutations g of M such that there is an n-tuple $(m_1, \dots, m_n) \in \mathbb{Z}^n$ such that $g(i, r) = (i, r + m_i)$ for all $i \in \{1, \dots, n\}$ and all but finitely many $(i, r) \in M$. These groups were first introduced in [Houghton 1978]. There is an important homomorphism $f : H_n \to \mathbb{Z}^n$ sending each permutation g to the corresponding n-tuple $(m_1, \dots, m_n)$. We always have $\sum_{1 \le i \le n} m_i = 0$.

For $n \ge 2$, the group H_n is finitely generated. More generally K. Brown [1987a] has shown that H_n is of type FP_{n-1} but not of type FP_n. For $n \ge 2$, the group H_n

is generated by $\{t_2, \ldots, t_n\}$, where

$$t_i(j,r) = (j,r) \quad \text{for } i \neq j,\ j \neq 1 \qquad t_i(i,1) = (1,1),$$
$$t_i(i,r) = (i,r-1) \quad \text{for } r > 1, \qquad t_i(1,r) = (1,r+1).$$

Also $H_n/[H_n, H_n] \simeq \mathbb{Z}^{n-1}$; see [Gehrke 1998]. The invariant $\Sigma^1(H_n)^c$ for $n \geq 2$ was calculated in [Brown 1987b] as

$$\Sigma^1(H_n)^c = \{[\alpha_1], \ldots, [\alpha_n]\},$$

where $\alpha_i = -\pi_i f$ and $\pi_i : \mathbb{Z}^n \to \mathbb{Z}$ is the projection to the i-th coordinate. Thus $\Sigma^1(H_n)^c$ is a finite set of rational points of $S(H_n)$, so we can apply Corollary 3.4 and get the following result.

Corollary 4.12. *Let $n \geq 2$ be a natural number, and let $G = H_n$. Then there is a subgroup H of finite index in* Aut(G) *such that $R(\varphi)$ is infinite for every $\varphi \in H$.*

5. Limit groups

In the previous section we saw that sometimes to understand the structure of $\Sigma^1(G)$, it suffices to study the twisted conjugacy classes, although we might not have a complete classification of the whole automorphism group. In this section we study twisted conjugacy classes for 1-ended limit groups G. Kochloukova [2010] proved that $\Sigma^1(G)$ is empty for any limit group G, so we have to find a different approach to study the Reidemeister number $R(\varphi)$. As noted in the introduction, the class of limit groups coincides with the class of finitely generated fully residually free groups.

There are results about the automorphism group of 1-ended limit groups that give a description of a subgroup of finite index of Aut(G) [Bumagin et al. 2007]. In the general case (when the limit group G can be decomposed as a free product), it is known that Out(G) has a subgroup of finite index that has a finite classifying space [Guirardel and Levitt 2007, Theorem 6.5]. Here we do not need the full strength of these results, but use the existence of a canonical abelian JSJ decomposition for a 1-ended limit group G that is neither a surface group nor free abelian of finite rank [Bumagin et al. 2007, Theorem 3.13]. This means that G is the fundamental group of a finite graph of groups with finitely generated abelian edge stabilizers, that is, the group associated to each edge is abelian and there is a canonical graph of groups with these properties. There are limited possibilities for the vertex groups: abelian groups of finite rank, HQ-groups (that is, surface groups), and rigid groups [Bumagin et al. 2007].

The uniqueness of the canonical abelian JSJ decomposition shows that there is a subgroup H_0 of finite index in Aut(G) such that under an automorphism $\varphi \in H_0$ a vertex group goes to a conjugate of a vertex group. Also, up to conjugation and generalized Dehn twists along edges (these twists are defined in [Bumagin et al. 2007, Definition 2.4] and have the property that on a fixed vertex group they act by conjugation with an element of G), an automorphism $\varphi \in H_0$ induces isomorphisms on the vertex groups that permute the adjacent edge groups up to conjugation.

The proof of the following result requires the existence of a canonical JSJ decomposition of G that has bipartite structure as defined in [Bumagin et al. 2007, Theorem 3.13]. This requires that G is a 1-ended limit group that is neither a free abelian nor a surface group.

Theorem 5.1. *Let G be a 1-ended limit group that is neither a free abelian nor a surface group. Then there is a subgroup of finite index H of* $\mathrm{Aut}(G)$ *such that $R(\varphi)$ is infinite for every $\varphi \in H$.*

Proof. Consider the canonical abelian JSJ decomposition of G. Then there is a subgroup of finite index H_0 of $\mathrm{Aut}(G)$ such that for every $\varphi \in H_0$ and every $t \in V(\Gamma) \cup E(\Gamma)$, we have $\varphi(G_t) = G_t^{g_t}$ for some $g_t \in G$. Note that $H_0 = \mathrm{Aut}(G)$ if the underlying graph of groups Γ of the JSJ decomposition of G does not have nontrivial symmetries that permute vertices and edges with isomorphic vertex and edge groups, respectively.

We are particularly interested in edge groups. Since the underlying graph is bipartite, every nonabelian group is linked with an abelian one. Two nonabelian groups or two abelian groups are not directly linked via an edge in the canonical JSJ decomposition. Thus if the connected graph Γ has more than one vertex, there is a nonabelian vertex group, and if Γ has just one vertex, the unique vertex group cannot be abelian as G is nonabelian. The nonabelian groups are divided into two classes: rigid and flexible.

By going down to a subgroup of finite index H_1 in H_0, we can assume that, up to conjugation and generalized Dehn twists along edges, the elements of H_1 that fix any rigid vertex group (not pointwise but as a group) are the identity on this rigid group [Kharlampovich and Myasnikov 2005, Theorem 15.1; Bumagin et al. 2007, Corollary 4.10]. Furthermore the edge groups of flexible vertices (that is, HQ vertices) are infinite cyclic [Bumagin et al. 2007, Definition 3.9], and $\mathrm{Aut}(\mathbb{Z}) = \mathbb{Z}_2$ is finite. Then by further going down to a subgroup H of finite index in H_1, we can consider only automorphisms of G that up to conjugation and generalized Dehn twists along edges send an edge group of a flexible vertex group identically to itself. Thus

$$\varphi(g) \in g^G \text{ for every } \varphi \in H \text{ and every } g \neq 1 \text{ of an edge group.} \tag{5-1}$$

Let $\gamma_i(G)$ be the i-th term of the lower central series of G, and let $t_i(G)$ be the normal subgroup of G containing $\gamma_i(G)$ such that $G/t_i(G)$ is the maximal torsion-free quotient of $G/\gamma_i(G)$. Kochloukova [2010] showed that $\bigcap_{i \geq 1} t_i(G) = 1$, an easy corollary of the fact that G is fully residually free. In other words, for every finite subset T of G there is a homomorphism from G to a free group that is injective on T. Then for the nontrivial element g from (5-1) we can find some i such that $g \notin t_i(G)$. Thus φ induces an automorphism $\overline{\varphi}$ on $Q = G/t_i(G)$ that sends the image $\bar{g}$ of g in Q to some conjugate of $\bar{g}$ in Q. Then $R(\overline{\varphi})$ is infinite by Proposition 3.5; hence $R(\varphi)$ is infinite. □

Acknowledgments

The second author thanks Olga Kharlampovich for conversations during the school of algebra in Rio de Janeiro in 2008 about her results on the automorphism group of 1-ended limit groups [Bumagin et al. 2007].

References

[Alibegović and Bestvina 2006] E. Alibegović and M. Bestvina, "Limit groups are CAT(0)", *J. London Math. Soc.* (2) **74**:1 (2006), 259–272. MR 2007m:57004

[Bestvina and Brady 1997] M. Bestvina and N. Brady, "Morse theory and finiteness properties of groups", *Invent. Math.* **129**:3 (1997), 445–470. MR 98i:20039 Zbl 0888.20021

[Bieri and Geoghegan 2010] R. Bieri and R. Geoghegan, "Sigma invariants of direct products of groups", *Groups Geom. Dyn.* **4**:2 (2010), 251–261. MR 2595091 Zbl 05696864

[Bieri and Groves 1984] R. Bieri and J. R. J. Groves, "The geometry of the set of characters induced by valuations", *J. Reine Angew. Math.* **347** (1984), 168–195. MR 86c:14001 Zbl 0526.13003

[Bieri and Groves 1986] R. Bieri and J. R. J. Groves, "A rigidity property for the set of all characters induced by valuations", *Trans. Amer. Math. Soc.* **294**:2 (1986), 425–434. MR 87i:16015 Zbl 0598.13001

[Bieri and Renz 1988] R. Bieri and B. Renz, "Valuations on free resolutions and higher geometric invariants of groups", *Comment. Math. Helv.* **63**:1 (1988), 464–497. MR 90a:20106 Zbl 0654.20029

[Bieri and Strebel 1980] R. Bieri and R. Strebel, "Valuations and finitely presented metabelian groups", *Proc. London Math. Soc.* (3) **41**:3 (1980), 439–464. MR 81j:20080 Zbl 0448.20029

[Bieri and Strebel 1981] R. Bieri and R. Strebel, "A geometric invariant for modules over an abelian group", *J. Reine Angew. Math.* **322** (1981), 170–189. MR 82f:20017 Zbl 0448.20010

[Bieri and Strebel 1982] R. Bieri and R. Strebel, "A geometric invariant for nilpotent-by-abelian-by-finite groups", *J. Pure Appl. Algebra* **25**:1 (1982), 1–20. MR 83i:20031 Zbl 0485.20026

[Bieri et al. 1987] R. Bieri, W. D. Neumann, and R. Strebel, "A geometric invariant of discrete groups", *Invent. Math.* **90**:3 (1987), 451–477. MR 89b:20108 Zbl 0642.57002

[Bieri et al. 2010] R. Bieri, R. Geoghegan, and D. H. Kochloukova, "The Sigma invariants of Thompson's group F", *Groups Geom. Dyn.* **4**:2 (2010), 263–273. Zbl 05696865

[Bleak et al. 2008] C. Bleak, A. Fel'shtyn, and D. L. Gonçalves, "Twisted conjugacy classes in R. Thompson's group F", *Pacific J. Math.* **238**:1 (2008), 1–6. MR 2009g:20056

[Bourbaki 1989] N. Bourbaki, *Commutative algebra*, Elements of Mathematics **2.2**, Springer, Berlin, 1989. MR 90a:13001 Zbl 0666.13001

[Brady et al. 2001] N. Brady, J. McCammond, J. Meier, and A. Miller, "The pure symmetric automorphisms of a free group form a duality group", *J. Algebra* **246** (2001), 881–896. MR 2002i:20036 Zbl 0995.20010

[Brin 1996] M. G. Brin, "The chameleon groups of Richard J. Thompson: automorphisms and dynamics", *Inst. Hautes Études Sci. Publ. Math.* **84**:1 (1996), 5–33. MR 99e:57003 Zbl 0891.57037

[Brin and Guzmán 1998] M. G. Brin and F. Guzmán, "Automorphisms of generalized Thompson groups", *J. Algebra* **203**:1 (1998), 285–348. MR 99d:20056 Zbl 0930.20039

[Brown 1987a] K. S. Brown, "Finiteness properties of groups", *J. Pure Appl. Algebra* **44**:1–3 (1987), 45–75. MR 88m:20110 Zbl 0613.20033

[Brown 1987b] K. S. Brown, "Trees, valuations, and the Bieri–Neumann–Strebel invariant", *Invent. Math.* **90**:3 (1987), 479–504. MR 89e:20060 Zbl 0663.20033

[Bumagin et al. 2007] I. Bumagin, O. Kharlampovich, and A. Miasnikov, "The isomorphism problem for finitely generated fully residually free groups", *J. Pure Appl. Algebra* **208**:3 (2007), 961–977. MR 2007j:20037 Zbl 1121.20026

[Fel'shtyn 2001] A. L. Fel'shtyn, "The Reidemeister number of any automorphism of a Gromov hyperbolic group is infinite", *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*), *Geom. i Topol.* **279**:6 (2001), 229–240, 250. In Russian; translated in *J. Math. Sci.* **119**:1, (2004), 117–123. MR 2002e:20081

[Fel'shtyn et al. 2008a] A. Fel'shtyn, Y. Leonov, and E. Troitsky, "Twisted conjugacy classes in saturated weakly branch groups", *Geom. Dedicata* **134**:1 (2008), 61–73. MR 2009g:20095

[Fel'shtyn et al. 2008b] A. L. Fel'shtyn, D. L. Gonçalves, and P. Wong, "Twisted conjugacy classes for polyfree groups", preprint, 2008. arXiv 0802.2937

[Gehrke 1998] R. Gehrke, "The higher geometric invariants for groups with sufficient commutativity", *Comm. Algebra* **26**:4 (1998), 1097–1115. MR 99b:20072 Zbl 0905.20024

[Gersten 1999] S. M. Gersten, "Introduction to hyperbolic and automatic groups", pp. 45–70 in *Summer school in group theory* (Banff, 1996), edited by O. Kharlampovich, CRM Proc. Lecture Notes **17**, American Mathematical Society, Providence, RI, 1999. MR 2000a:20093 Zbl 0926.20028

[Gonçalves 1998] D. L. Gonçalves, "Coincidence Reidemeister classes on nilmanifolds and nilpotent fibrations", *Topology Appl.* **83**:3 (1998), 169–186. MR 98m:55003 Zbl 0929.55002

[Gonçalves and Wong 2003] D. Gonçalves and P. Wong, "Twisted conjugacy classes in exponential growth groups", *Bull. London Math. Soc.* **35**:2 (2003), 261–268. MR 2003j:20054 Zbl 1030.20020

[Guirardel and Levitt 2007] V. Guirardel and G. Levitt, "The outer space of a free product", *Proc. Lond. Math. Soc.* (3) **94**:3 (2007), 695–714. MR 2008c:20047 Zbl 1168.20011

[Houghton 1978] C. H. Houghton, "The first cohomology of a group with permutation module coefficients", *Arch. Math.* (*Basel*) **31**:1 (1978), 254–258. MR 80c:20073 Zbl 0377.20044

[Jiang 1983] B. J. Jiang, *Lectures on Nielsen fixed point theory*, Contemporary Mathematics **14**, American Mathematical Society, Providence, RI, 1983. MR 84f:55002 Zbl 0512.55003

[Kharlampovich and Myasnikov 2005] O. Kharlampovich and A. G. Myasnikov, "Effective JSJ decompositions", pp. 87–212 in *Groups, languages, algorithms*, edited by A. V. Borovik, Contemp. Math. **378**, American Mathematical Society, 2005. MR 2006m:20045 Zbl 1093.20019

[Kochloukova 2010] D. H. Kochloukova, "On subdirect products of type FP_m of limit groups", *J. Group Theory* **13**:1 (2010), 1–19. MR 2604842 Zbl 05673300

[Kropholler 1993] P. H. Kropholler, "On groups of type $(\mathrm{FP})_\infty$", *J. Pure Appl. Algebra* **90**:1 (1993), 55–67. MR 94j:20051b Zbl 0816.20042

[Levitt 2007] G. Levitt, "On the automorphism group of generalized Baumslag–Solitar groups", *Geom. Topol.* **11** (2007), 473–515. MR 2008h:20061 Zbl 1143.20014

[Levitt and Lustig 2000] G. Levitt and M. Lustig, "Most automorphisms of a hyperbolic group have very simple dynamics", *Ann. Sci. École Norm. Sup.* (4) **33**:4 (2000), 507–517. MR 2002d:20066 Zbl 0997.20043

[McCool 1986] J. McCool, "On basis-conjugating automorphisms of free groups", *Canad. J. Math.* **38**:6 (1986), 1525–1529. MR 87m:20093 Zbl 0613.20024

[Meier et al. 1998] J. Meier, H. Meinert, and L. VanWyk, "Higher generation subgroup sets and the Σ-invariants of graph groups", *Comment. Math. Helv.* **73**:1 (1998), 22–44. MR 99f:57002 Zbl 0899.57001

[Meinert 1995] H. Meinert, "The Bieri–Neumann–Strebel invariant for graph products of groups", *J. Pure Appl. Algebra* **103**:2 (1995), 205–210. MR 96i:20047 Zbl 0838.20027

[Orlandi-Korner 2000] L. A. Orlandi-Korner, "The Bieri–Neumann–Strebel invariant for basis-conjugating automorphisms of free groups", *Proc. Amer. Math. Soc.* **128**:5 (2000), 1257–1262. MR 2000k:20046 Zbl 0963.20016

[Roseblade 1978] J. E. Roseblade, "Prime ideals in group rings of polycyclic groups", *Proc. London Math. Soc.* (3) **36**:3 (1978), 385–447. MR 58 #10996a Zbl 0391.16008

[Schütz 2008] D. Schütz, "On the direct product conjecture for sigma invariants", *Bull. Lond. Math. Soc.* **40**:4 (2008), 675–684. MR 2009j:20075

[Stein 1992] M. Stein, "Groups of piecewise linear homeomorphisms", *Trans. Amer. Math. Soc.* **332**:2 (1992), 477–514. MR 92k:20075 Zbl 0798.20025

[Taback and Wong 2007] J. Taback and P. Wong, "Twisted conjugacy and quasi-isometry invariance for generalized solvable Baumslag–Solitar groups", *J. Lond. Math. Soc.* (2) **75**:3 (2007), 705–717. MR 2008h:20043 Zbl 1128.20030

[Thurston 1986] W. P. Thurston, "A norm for the homology of 3-manifolds", pp. i–vi and 99–130 Mem. Amer. Math. Soc. **339**, American Mathematical Society, 1986. MR 88h:57014

Received January 22, 2009. Revised December 13, 2009.

DACIBERG GONÇALVES
DEPARTAMENTO DE MATEMÁTICA - IME-USP
UNIVERSIDADE DE SÃO PAULO
CAIXA POSTAL 66281
AGÊNCIA CIDADE DE SÃO PAULO
05314-970 SÃO PAULO, SP
BRAZIL

dlgoncal@ime.usp.br

DESSISLAVA HRISTOVA KOCHLOUKOVA
DEPARTAMENTO DE MATEMÁTICA
UNICAMP
13083-970 CAMPINAS, SP
BRAZIL

desi@ime.unicamp.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

PROPERTIES OF ANNULAR CAPILLARY SURFACES WITH EQUAL CONTACT ANGLES

JAMES GORDON AND DAVID SIEGEL

We consider an annular region $\Omega \subset \mathbb{R}^2$ and analyze the capillary surface $z = u(x, y)$ formed within an annular cylinder $\Omega \times \mathbb{R}$. Assuming identical contact angles γ along the inner and outer boundaries, we determine several qualitative properties of the surface. In particular, we examine the behavior of u in the limiting cases of Ω approaching a disk, a thin ring, and the exterior of a disk.

1. Introduction

The equilibrium liquid-gas interface formed within a capillary tube has been studied extensively over the past two hundred years. The most widely used modern reference is [Finn 1986]. We will consider the related annular geometry in the presence of gravity first examined by Laplace in 1806; see [Laplace 1966, supplements to book X]. Here two concentric circular cylinders define an annular cross section $\Omega \subset \mathbb{R}^2$. If the cylinders are immersed vertically in an infinite reservoir of incompressible fluid, the surface $Z = U(X, Y)$ formed between the tubes will satisfy the boundary value problem

$$\begin{cases} NU = \kappa U & \text{in } \Omega, \\ \hat{\nu} \cdot TU = \cos\gamma & \text{on } \partial\Omega, \end{cases}$$

where $TU = \nabla U / \sqrt{1 + |\nabla U|^2}$, $NU = \nabla \cdot TU$, $\hat{\nu}$ is the exterior unit normal on the boundary $\partial\Omega$ and $\kappa > 0$ is the capillary constant. The contact angle $\gamma \in [0, \pi]$ is defined on the inner and outer boundaries and gives the angle at which the interface meets the bounding wall. For this investigation, γ is assumed to be constant and equal along each cylinder. Such a scenario arises when both tubes are made of the same uniform material.

MSC2000: primary 34B15, 35J25, 76B45; secondary 34E10.
Keywords: annular capillary surface, qualitative, asymptotic.
Gordon is supported by an Ontario Graduate Scholarship. Siegel is supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant.

The axisymmetric nature of such annular solutions allows us to analyze the boundary value problem for an ordinary differential equation:

$$(1)\qquad \begin{cases} \dfrac{1}{R}\left(\dfrac{RU_R}{\sqrt{1+U_R^2}}\right)_R = \kappa U \quad \text{for } R_1 < R < R_2, \\ U_R(R_1^+) = -\cot\gamma, \\ U_R(R_2^-) = \cot\gamma, \end{cases}$$

where U is the surface height, R is the radial variable and $(\cdot)_R$ denotes differentiation with respect to R. System (1) is made dimensionless by introducing the variables

$$u = U/R_2 \quad \text{and} \quad r = R/R_2,$$

which gives

$$(2)\qquad \begin{cases} Nu = \dfrac{(r\sin\psi)_r}{r} = Bu \quad \text{for } a < r < 1 \\ \sin\psi(a) = -\cos\gamma, \\ \sin\psi(1) = \cos\gamma, \end{cases}$$

where B is a positive constant known as the Bond number, and we define $\psi(r)$ as the inclination angle of $u(r)$:

$$\sin\psi(r) = \frac{u_r}{\sqrt{1+u_r^2}}.$$

See Figure 1. The outer radius of the region is now fixed at $r = 1$, while the inner boundary will occur at $r = a$ for $0 < a < 1$. Additionally, we need only consider

$$(3)\qquad 0 \le \gamma < \pi/2$$

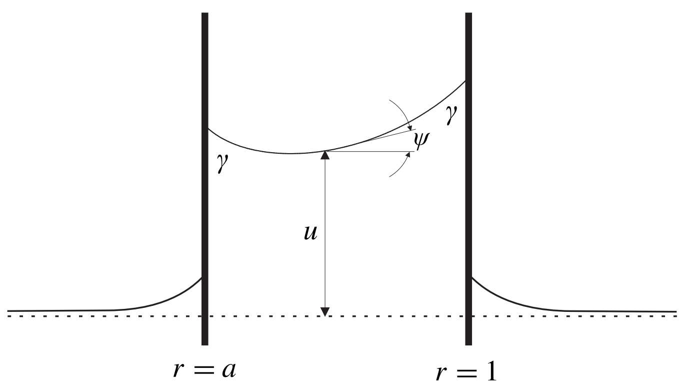


Figure 1. Radial cross section of annular capillary surface.

since the other possibilities are accounted for as follows:

- If $\gamma = \pi/2$, then $u = 0$ is the unique solution.
- For a solution u with $\gamma \in (\pi/2, \pi]$, let $\bar{u} = -u$. We therefore have $N\bar{u} = B\bar{u}$ with $\bar{\gamma} = \pi - \gamma$ or $\bar{\gamma} \in [0, \pi/2)$.

Under (3), the comparison principle [Concus and Finn 1974; Finn 1986] requires u to be positive and bounded for any selection of parameters a and B. Additionally, the volume of u above Ω can be determined by

$$\int_a^1 ru(r)\,dr = \frac{\cos\gamma(1+a)}{B}. \tag{4}$$

Contributions to the annular problem have been made by Elcrat, Kim, and Treinen [2004] and Siegel [2006]; however, this research is still in its fledgling stage. In this paper, the comparison principle is used to provide several qualitative results. We begin in Section 2 by illustrating some general properties of u, the solution to (2); specifically, there exists a unique radius $r = m$ at which u achieves its minimum value, $u(a) < u(1)$, $m \in (a, (1+a)/2)$ and m is monotone increasing with respect to a. Section 3 then explores the behaviour of solutions to the annular problem (1) in the following limiting cases:

- For the dimensionless version of (2), we consider the two cases of $a \to 0$ and $a \to 1$.
- Alternatively, the using the variables

  $$u = U/R_1 \quad \text{and} \quad r = R/R_1$$

  to make (1) dimensionless reformulates it as

  $$\begin{cases} Nu = \dfrac{(r\sin\psi)_r}{r} = Bu & \text{for } 1 < r < b, \\ \sin\psi(1) = -\cos\gamma, \\ \sin\psi(b) = \cos\gamma \end{cases} \tag{5}$$

  The behaviour of u is consequently examined as $b \to \infty$.

2. General properties

In this section, the comparison principle will be used to present a number of qualitative results. We start by confirming the uniqueness of the minimum surface height, which is mentioned under more general conditions in [Elcrat et al. 2004].

Theorem 2.1. *Let u be a solution to the boundary value problem (2). There exists a unique radius $r = m$ at which u achieves its minimum value.*

Proof. Since $\sin\psi$ is continuous with

$$\sin\psi(a) = -\cos\gamma < 0 \quad \text{and} \quad \sin\psi(1) = \cos\gamma > 0,$$

there exists at least one point in $(a, 1)$ where $\sin\psi = 0$, which corresponds to an extremum of u. Define $r = m$ as the first zero of $\sin\psi$. Using the first of (2), we note

$$(\sin\psi)_r = Bu - (\sin\psi)/r \tag{6}$$

$$> 0 \quad \text{for } \sin\psi \le 0 \tag{7}$$

and specifically, $\sin\psi$ is increasing at $r = m$. Suppose there exists more than one point where $\sin\psi = 0$ and let m' be the next zero immediately following m. Because $\sin\psi$ is increasing at m, it must be nonincreasing as it touches the r-axis at m':

$$(\sin\psi)_r|_{r=m'} \le 0.$$

However, this is in contradiction to (7), and m must be the unique extremum point of u. Inequality (7) also implies this is a minimum. □

For the next theorem, we compare boundary heights.

Lemma 2.2. *The function* $\sin\psi$ *is monotone increasing on* $[a, 1]$.

Proof. Given that the zero of $\sin\psi$ is unique, we consider $\sin\psi$ on two subintervals. We have $\sin\psi \le 0$ on $[a, m]$, and (6) ensures that $(\sin\psi)_r > 0$. On $(m, 1]$, $\sin\psi > 0$ and thus u is increasing. In this case, we multiply the first of (2) by r and integrate from m to r to obtain

$$\begin{aligned} \sin\psi(r) &= \frac{B}{r}\int_m^r su(s)\,ds \\ &< \frac{Bu(r)}{r}\left(\frac{r^2-m^2}{2}\right) \\ &< \frac{Bru(r)}{2}. \end{aligned} \tag{8}$$

Therefore $Bu - (\sin\psi)/r > 0$. Equation (6) confirms that $(\sin\psi)_r > 0$. □

Remark. Lemma 2.2 also implies that u is convex.

Theorem 2.3. $u(a) < u(1)$.

Proof. The construction of this proof follows the ideas of [Serrin 1971]. Starting with the annular region Ω, we place as in Figure 2 a line T that separates from Ω a cap Γ. Let Γ' be the reflection of Γ with respect to T, and observe that T is positioned so that Γ' is internally tangent to $\partial\Omega$ at p. Finally, let $\hat{n}$ be the exterior unit normal on $\partial\Gamma'$. With the coordinate system (x, y) oriented so that the y-axis

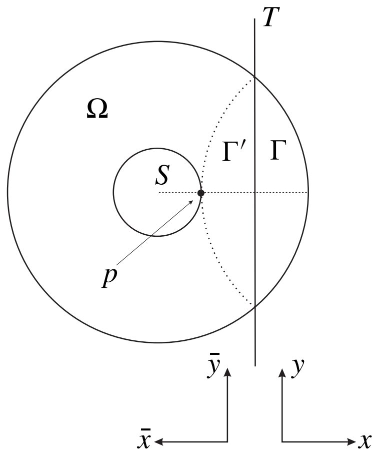


Figure 2. Configuration of reflected region Γ' superimposed onto Ω.

is aligned with T, we define a function $\bar{u}$ on Γ' as

$$\bar{u}(x, y) = u(\bar{x}, \bar{y}) = u(-x, y) \quad \text{for } (x, y) \in \Gamma'.$$

Let $\overline{N}$ be the N operator with respect to the coordinate system $(\bar{x}, \bar{y})$. Clearly, $\overline{N}\bar{u} = B\bar{u}$. However, N is invariant under reflections; thus, $N\bar{u} = \overline{N}\bar{u} = B\bar{u}$ and $\bar{u}$ also satisfies the capillary equation in Γ'. The boundary of Γ' is now decomposed into two pieces, with Σ_α being the portion along T and Σ_β as the remaining curved piece. We subsequently examine how u and $\bar{u}$ compare on each boundary component. It is immediately clear that $u = \bar{u}$ on Σ_α. On Σ_β, note that $\hat{n} \cdot Tu = \sin\psi\ \hat{n} \cdot \hat{r}$, where $\hat{r}$ is the unit vector in the radial direction. Since $\sin\psi$ is increasing, this yields $-\cos\gamma \le \hat{n} \cdot Tu \le \cos\gamma$. Of course, $\hat{n} \cdot T\bar{u} = \cos\gamma$ and hence $\hat{n} \cdot T\bar{u} \ge \hat{n} \cdot Tu$ on Σ_β. As a result, the comparison principle requires

$$\bar{u} \ge u \quad \text{in } \Gamma', \tag{9}$$

which can be extended to the boundary point p by continuity:

$$u(p) \le \bar{u}(p) \quad \text{if and only if} \quad u(a) \le u(1). \tag{10}$$

The possibility of $u(p) = \bar{u}(p)$ is excluded by contradiction. In this case, our attention is restricted to the dashed line S of Figure 2 and both functions are described in terms of the radial variable only. We next assume that $u(a) = u(1)$, which allows the meridional curvature $k_m = (\sin\psi)_r$ of the surface to be compared at $r = a$ and $r = 1$:

$$(\sin\psi)_r|_{r=a} = Bu(a) + (\cos\gamma)/a > Bu(1) - \cos\gamma = (\sin\psi)_r|_{r=1}.$$

Consequently, there exists a $\delta > 0$ such that

$$\min_{r\in[a,a+\delta]}\{(\sin\psi)_r\} > \max_{r\in[1-\delta,1]}\{(\sin\psi)_r\}.$$

We can then integrate $(\sin\psi)_r$ over these regions, giving

$$\sin\psi(a+r) > -\sin\psi(1-r) \quad \text{for all } r\in(0,\delta],$$

and since the function $p/\sqrt{1-p^2}$ is increasing on $(-1, 1)$, we have

$$\frac{\sin\psi(a+r)}{\sqrt{1-\sin^2\psi(a+r)}} > -\frac{\sin\psi(1-r)}{\sqrt{1-\sin^2\psi(1-r)}}.$$

Thus,

$$\begin{aligned} u(a+\delta) &= u(a) + \int_a^{a+\delta} u_s(s)\,ds \\ &= u(a) + \int_a^{a+\delta} \frac{\sin\psi(s)}{\sqrt{1-\sin^2\psi(s)}}\,ds \\ &> u(1) - \int_{1-\delta}^{1} \frac{\sin\psi(s)}{\sqrt{1-\sin^2\psi(s)}}\,ds = u(1-\delta). \end{aligned} \tag{11}$$

This that $u(a+\delta) > \bar{u}(a+\delta)$, which is in contradiction to (9) and the inequality of (10) must be strict. □

Theorem 2.4. *The function u achieves its minimum on $(a, (1+a)/2)$.*

Proof. We refer to Figure 2 and again consider u and $\bar{u}$ along S. The proof will be by contradiction; we assume that the minimum of u occurs at $m \in ((1+a)/2, 1)$. If $\overline{m}$ is defined as the location of the minimum of $\bar{u}$, we then have $\overline{m} \in (a, (1+a)/2)$. However, the convexity of u implies that

$$u(m) < u(\overline{m}) \quad \text{if and only if} \quad \bar{u}(\overline{m}) < u(\overline{m})$$

with $\overline{m} \in \Gamma'$, which is in contradiction to (9). Thus, $m \in (a, (1+a)/2]$. Next, assume $m = (1+a)/2$. Given that $(\sin\psi)_{rr} = u_r - (\sin\psi)_r/r + (\sin\psi)/r^2$, Lemma 2.2 provides $(\sin\psi)_{rr}|_{r=m} < 0$ and continuity requires that there exists a $\delta > 0$ such that $(\sin\psi)_{rr} < 0$ on $[m-\delta, m+\delta]$. With $(\sin\psi)_r$ decreasing on the interval, this gives $-\sin\psi(m-r) > \sin\psi(m+r)$ for all $r \in (0, \delta]$. Finally, an argument similar to (11) yields $u(m-\delta) > u(m+\delta)$, and we conclude $u(m-\delta) > \bar{u}(m-\delta)$. This again contradicts (9); therefore the minimum of u occurs on $(a, (1+a)/2)$. □

Theorem 2.5. *The minimum value m is monotone increasing with respect to a.*

Proof. We proceed by contradiction. First, suppose there exist two inner radii $\bar{a}$ and $\hat{a}$ where m decreases with respect to a. This gives rise to the following configuration as shown in Figure 3:

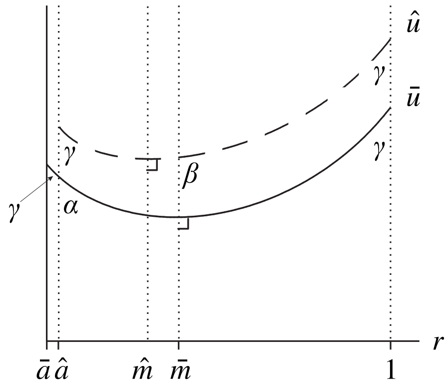


Figure 3. Hypothetical configuration assuming m is decreasing with respect to two values of a.

(i) $\bar{u}$ is the unique solution over $[\bar{a}, 1]$ whose minimum is at $r = \bar{m}$.

(ii) $\hat{u}$ is the unique solution over $[\hat{a}, 1]$ whose minimum is at $r = \hat{m}$.

(iii) $\bar{a} < \hat{a}$.

(iv) $\hat{m} < \bar{m}$.

Consider $\bar{u}$ and $\hat{u}$ on the region $[\hat{a}, 1]$. Here, the contact angle of $\bar{u}$ at $r = \hat{a}$ will be $\alpha > \gamma$, and the comparison principle therefore implies

$$\bar{u} < \hat{u} \quad \text{in } (\hat{a}, 1). \tag{12}$$

Alternatively, we can examine the solutions over $[\bar{m}, 1]$, in which the contact angle of $\hat{u}$ at $r = \bar{m}$ will be $\beta > \pi/2$. Here, the comparison principle would require $\bar{u} > \hat{u}$ in $(\bar{m}, 1)$ which is in disagreement with (12). Consequently, $\bar{m} \le \hat{m}$ for $\bar{a} < \hat{a}$. Now suppose that m is constant for two increasing values of a. Again, $\bar{u}$ and $\hat{u}$ will be configured as before, only with (iv) altered as

(iv)$'$ $\bar{u}$ and $\hat{u}$ share the same minimum at $r = m$.

Figure 4 depicts this possibility. In like manner, we have

$$\bar{u} < \hat{u} \quad \text{in } (\hat{a}, 1). \tag{13}$$

However, on $[m, 1]$, both $\bar{u}$ and $\hat{u}$ have identical contact angles and uniqueness requires $\bar{u} \equiv \hat{u}$, which contradicts (13), and we conclude $\bar{m} < \hat{m}$ for $\bar{a} < \hat{a}$. □

3. Solutions in limiting cases

Preliminary lemmas.

Lemma 3.1. *The function* $(\sin\psi)/r$ *is monotone increasing on* $[a, 1]$.

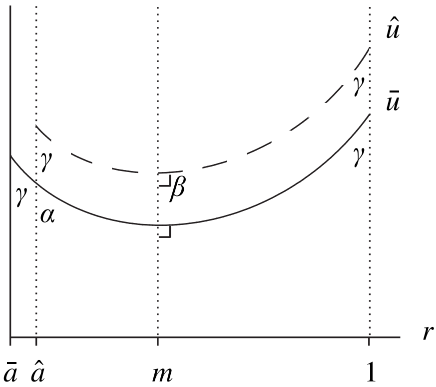


Figure 4. Hypothetical configuration assuming m is constant with respect to two values of a.

Proof. Equation (6) yields $((\sin\psi)/r)_r = (2/r)(Bu/2 - (\sin\psi)/r)$. As we did in Lemma 2.2, we can examine $((\sin\psi)/r)_r$ over two subintervals. On $[a, m]$, $\sin\psi \le 0$ and $((\sin\psi)/r)_r > 0$. On $(m, 1]$, result (8) can be used to claim that $Bu/2 - (\sin\psi)/r > 0$ and thus $((\sin\psi)/r)_r > 0$ for $r \in [a, 1]$. □

Lemma 3.2. *We have* $-(a\cos\gamma)/r < \sin\psi < r\cos\gamma$ *on* $(a, 1)$.

Proof. For the lower bound, we observe that the first of (2) provides the differential inequality $(r\sin\psi)_r = Bru > 0$, and thus $r\sin\psi$ is monotone increasing:

$$r\sin\psi(r) > a\sin\psi(a) = -a\cos\gamma \quad \text{for } r \in (a, 1].$$

For the upper bound, Lemma 3.1 may be used to show that

$$(\sin\psi(r))/r < \sin\psi(1) = \cos\gamma \quad \text{for } r \in [a, 1).$$ □

Approaching a disk. We now consider solutions to (2) as $a \to 0$. As such, reference will be made to the interior solution u_{int}, which solves

$$\begin{cases} (r\sin\psi)_r = Bru_{\text{int}} & \text{for } r \in (0, 1), \\ \sin\psi(0) = 0, \\ \sin\psi(1) = \cos\gamma. \end{cases} \tag{14}$$

See [Finn 1986] for background. Siegel [2006] examined the problem (14), along with the annular problem

$$\begin{cases} (r\sin\psi)_r = Bru & \text{for } r \in (a, 1), \\ \sin\psi(a) = 0, \\ \sin\psi(1) = \cos\gamma. \end{cases} \tag{15}$$

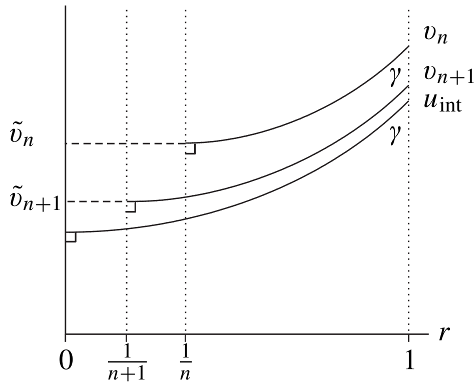


Figure 5. Illustration of $\{v_n\}$ and $\{\tilde{v}_n\}$ compared to u_{int}.

First, it will be shown that the solution of (15) approaches that of (14) as $a \to 0$. Let $\{v_n\}_{n\geq 2}$ be the sequence of functions such that v_n is the unique solution to (15) on the interval $[1/n, 1]$. Thus $\{v_n\}$ is defined on an increasing domain; however, it is desirable to consider also a sequence $\{\tilde{v}_n\}_{n\geq 2}$ of extended functions on $[0, 1]$ by continuing each v_n to $r = 0$ as

$$\tilde{v}_n(r) = \begin{cases} v_n(1/n) & \text{for } r \in [0, 1/n), \\ v_n(r) & \text{for } r \in [1/n, 1]. \end{cases}$$

See Figure 5. Here, $\tilde{v}_n \in C^1[0, 1]$ for all $n \geq 2$. From [Siegel 2006] it can be shown that each function $\tilde{v}_n$, along with the interior solution u_{int}, is increasing and bounded. Siegel also demonstrated that v_n and u_{int} will satisfy the same volume condition:

$$\int_{1/n}^{1} s v_n(s)\, ds = \int_0^1 s u_{\text{int}}(s)\, ds = \frac{\cos\gamma}{B}.$$

Therefore,

$$\int_0^1 s(\tilde{v}_n(s) - u_{\text{int}}(s))\, ds - \int_0^{1/n} s\tilde{v}_n(s)\, ds = 0. \tag{16}$$

Additionally, the comparison principle provides

$$v_{n+1} \leq v_n \quad \text{if and only if} \quad \tilde{v}_{n+1} \leq \tilde{v}_n \text{ for } n \geq 2$$

as well as

$$0 \leq u_{\text{int}} \leq v_n \quad \text{if and only if} \quad 0 \leq u_{\text{int}} \leq \tilde{v}_n \text{ for } n \geq 2.$$

Consequently, we are assured that $\tilde{v}_n \to v$ pointwise on $[0, 1]$ with

$$v \geq u_{\text{int}} \quad \text{on } [0, 1]. \tag{17}$$

Each integral in (16) thus defines a positive decreasing sequence with a defined limit as $n \to \infty$:

$$\lim_{n\to\infty} \int_0^1 s(\tilde{v}_n(s) - u_{\text{int}}(s))\, ds - \lim_{n\to\infty} \int_0^{1/n} s\tilde{v}_n(s)\, ds = 0. \tag{18}$$

The second limit in (18) can be bounded as

$$0 \le \lim_{n\to\infty} \int_0^{1/n} s\tilde{v}_n(s)\, ds \le \tilde{v}_2(1) \cdot \lim_{n\to\infty} \int_0^{1/n} s\, ds = 0,$$

and we conclude $\lim_{n\to\infty} \int_0^{1/n} s\tilde{v}_n(s)\, ds = 0$. The first limit in (18) must now be zero and Lebesgue's dominated convergence theorem can be used to see that

$$0 = \lim_{n\to\infty} \int_0^1 s(\tilde{v}_n(s) - u_{\text{int}}(s))\, ds. = \int_0^1 s(v(s) - u_{\text{int}}(s))\, ds \tag{19}$$

In conjuction with (17), this requires

$$v = u_{\text{int}} \quad \text{almost everywhere.} \tag{20}$$

We further comment that v must be nondecreasing and inequalities that occur in (20) are restricted to jump discontinuities in v. However, suppose such a discontinuity of height $\delta > 0$ occurs at a point $c \in [0, 1)$. Here, there will exist a $d > c$ such that u_{int} is continuous on $[c, d]$ with $v - u_{\text{int}} \ge \delta/2$. This is at odds with (19) being 0 and $v \equiv u_{\text{int}}$ on $[0, 1)$. We can also demonstrate that equality holds at $r = 1$. For $n \ge 2$, we shift u_{int} upward to the position of $\bar{u}_{\text{int}}$ so that $\bar{u}_{\text{int}}(1/n) = v_n(1/n)$. In other words,

$$\bar{u}_{\text{int}} = u_{\text{int}} + v_n(1/n) - u_{\text{int}}(1/n).$$

The comparison principle requires

$$u_{\text{int}}(1) \le v_n(1) \le \bar{u}_{\text{int}}(1). \tag{21}$$

Since $v_n(1/n) = \tilde{v}_n(0)$, we get $\lim_{n\to\infty} v_n(1/n) = \lim_{n\to\infty} \tilde{v}_n(0) = u_{\text{int}}(0)$. This with (21) gives $v(1) = u_{\text{int}}(1)$ and $v \equiv u_{\text{int}}$, as required.

Remark. Dini's theorem can be applied at this point to strengthen the convergence claim on $\{\tilde{v}_n\}$ from pointwise to uniform convergence.

Lemma 3.3. *Define u_a and u_{int} as in Theorem 2.5 and consider m as a function of a. If* $\lim_{a\to 0} m(a) = 0$, *then* $\lim_{a\to 0} u_a(m) = u_{\text{int}}(0)$.

Proof. For a given $m(a)$, select the maximum $n \in \mathbb{N}$ such that $m(a) \le 1/n$, which gives $1/(n+1) < m(a) \le 1/n$. With the sequence of functions $\{v_n\}$, the comparison principle produces the following arrangement, shown in Figure 6:

$$v_{n+1}(1/(n+1)) < u_a(m) \le v_n(1/n) \tag{22}$$

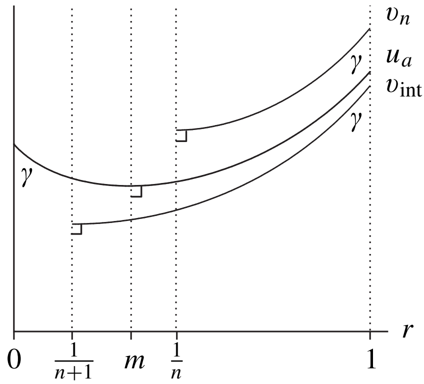


Figure 6. Choosing n so that $v_{n+1}(1/(n+1)) < u_a(m) \leq v_n(1/n)$.

with $\lim_{n\to\infty} v_{n+1}(1/(n+1)) = \lim_{n\to\infty} v_n(1/n) = u_{\text{int}}(0)$. For $\lim_{a\to 0} m(a) = 0$, we have $\lim_{a\to 0} n = \infty$ and (22) requires $\lim_{a\to 0} u_a(m) = u_{\text{int}}(0)$. □

Theorem 3.4. *For* $\gamma \in [0, \pi/2)$, *consider the interior solution* u_{int} *defined on* $[0, 1]$ *together with* u_a, *the solution to* (2) *on* $[a, 1]$. *We have*

$$\lim_{a\to 0} \max_{r\in[a,1]} |u_a(r) - u_{\text{int}}(r)| = 0.$$

Proof. On $[a, 1]$, we compare the contact angles of u_a and u_{int}, noting that the comparison principle requires

$$u_{\text{int}} \leq u_a \quad \text{on } [a, 1]. \tag{23}$$

See Figure 7. Additionally, u_{int} may be shifted upward to the position of $\bar{u}_{\text{int}}$ such that $\bar{u}_{\text{int}}(a) = u_a(a)$. Here again, we use the comparison principle to see that

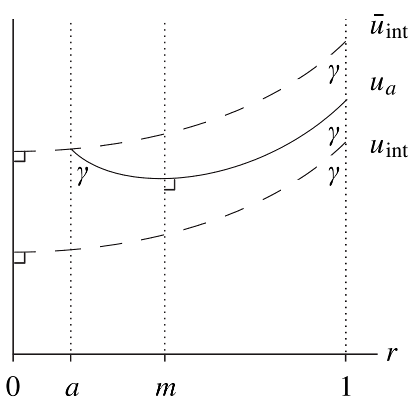


Figure 7. Cross section of comparison surfaces for $a \to 0$.

$u_a \le \bar{u}_{\text{int}}$ on $[a, 1]$. Consequently,

$$\max_{r\in[a,1]} |u_a(r) - u_{\text{int}}(r)| \le [u_a(a) - u_a(m)] + (u_a(m) - u_{\text{int}}(0)), \tag{24}$$

and both bracketed terms of (24) can be bounded. For the first term, we write

$$u_a(a) - u_a(m) = -\int_a^m u_s\, ds = -\int_a^m \frac{\sin\psi}{\sqrt{1-\sin^2\psi}}\, ds,$$

and using Lemma 3.2,

$$u_a(a) - u_a(m) < a\int_a^m \frac{1}{\sqrt{r^2-a^2}}\, ds < a\log(1+\sqrt{1-a^2}) - a\log a$$
$$\to 0 \quad \text{as } a \to 0.$$

For the second term in (24), it is clear that u_a satisifes the boundary value problem (15) on $[m, 1]$. Considering m as a function of a, it is sufficient to show that $\lim_{a\to 0} m(a) = 0$, as Lemma 3.3 would then require $\lim_{a\to 0}(u_a(m) - u_{\text{int}}(0)) = 0$, thus proving the theorem. We proceed by contradiction and assume m does not approach 0. As a result, there exists a $\sigma > 0$ such that

$$m \ge \sigma \quad \text{for all } a \in (0, 1). \tag{25}$$

Suppose that $a < \sigma$. By multiplying the first of (2) by r and integrating from a to m, we have

$$\int_a^m s u_a(s)\, ds = \frac{a\cos\gamma}{B} \quad \text{implies} \quad \lim_{a\to 0}\int_a^m s u_a(s)\, ds = 0.$$

Using (25) and that u_a is decreasing on $[a, m)$, the above integral could also be bounded as $\int_a^m s u_a(s)ds \ge u_a(\sigma)\int_a^\sigma s\, ds$. By (23), $u_a(\sigma) \ge u_{\text{int}}(\sigma)$ so that

$$\lim_{a\to 0}\int_a^m s u_a(s)\, ds \ge u_{\text{int}}(\sigma)\frac{\sigma^2}{2} > 0.$$

This is an impossible situation and m must approach 0 as $a \to 0$. As a result,

$$\max_{r\in[a,1]} |u_a(r) - u_{\text{int}}(r)| \to 0 \quad \text{as } a \to 0. \qquad \square$$

Approaching a thin ring. We next examine solutions to (2) as $a \to 1$. For this, let $u_0 = 2\cos\gamma/(B(1-a))$, the constant function that satisfies the volume condition (4). Also, define the function u_1 by

$$u_1(r) = u_1(a) + \int_a^r \frac{\sin\psi_1(s)}{\sqrt{1-\sin^2\psi_1(s)}}\, ds,$$

with

$$\sin\psi_1(r) = \frac{B}{r}\int_a^r su_0\,ds - \frac{a}{r}\cos\gamma = \frac{\cos\gamma}{1-a}\Big(r - \frac{a}{r}\Big)$$

and

$$u_1(a) = \frac{2\cos\gamma}{B(1-a)} - \frac{1}{1-a^2}\int_a^1 (1-s^2)\frac{\sin\psi_1(s)}{\sqrt{1-\sin^2\psi_1(s)}}.$$

Here, ψ_1 denotes the inclination angle of u_1. Since

$$-(a/r)\cos\gamma \le \sin\psi_1 \le r\cos\gamma, \tag{26}$$

it is easily checked that u_1 is defined and continuous; the choice of $u_1(a)$ ensures that u_1 also satisfies the volume condition. Note that u_1 is a Delaunay surface (that is, a surface of revolution having constant mean curvature) satisfying the differential equation

$$Nu_1 = Bu_0 \quad \text{if and only if} \quad (r\sin\psi_1(r))_r = Bru_0. \tag{27}$$

For $\gamma \neq 0$, it so happens that u_1 will act as a limiting surface as $a \to 1$.

Theorem 3.5. *Define u_a as in Theorem 3.4 and consider the function u_1 described above. For $\gamma \neq 0$, we have $|u_a - u_1| = O((1-a)^3)$ as $a \to 1$.*

Proof. We first bound $|u_a - u_0|$. Using that u_a is convex and $u_a(a) < u_a(1)$, we have

$$|u_a - u_0| \le \max\{u_a(1) - u_0, u_0 - u_a(m)\} < u_a(1) - u_a(m) = \int_m^1 \frac{\sin\psi_a}{\sqrt{1-\sin^2\psi_a}}\,dr,$$

where ψ_a is the inclination angle of u_a. Lemma 3.2 provides that

$$\frac{\sin\psi}{\sqrt{1-\sin^2\psi}} \le \frac{r\cos\gamma}{\sqrt{1-r^2\cos^2\gamma}}$$

and consequently

$$|u - u_0| < \int_m^1 \frac{r\cos\gamma}{\sqrt{1-r^2\cos^2\gamma}}\,dr = \frac{\sqrt{1-m^2\cos^2\gamma} - \sin\gamma}{\cos\gamma} := C(\gamma, m) < C(\gamma, a).$$

Using the first of (2) and (27), we write

$$\sin\psi_a - \sin\psi_1 = \frac{B}{r}\int_a^r s(u_a - u_0)\,ds,$$

or equivalently, $\sin\psi_a - \sin\psi_1 = -(B/r)\int_r^1 s(u_a - u_0)\,ds$. Taken together, these yield

$$|\sin\psi_a - \sin\psi_1| \le \frac{B}{2r}C(\gamma,a)\min\{r^2-a^2, 1-r^2\},$$

and given that $\min\{r^2-a^2, 1-r^2\} \le 2(r^2-a^2)(1-r^2)/(1-a^2)$, we have

$$|\sin\psi_a - \sin\psi_1| \le \frac{B}{r}\,C(\gamma,a)\frac{(r^2-a^2)(1-r^2)}{1-a^2}.$$

Continuing, we bound $|u_a - u_1|$ by first noting that u_a and u_1 have the correct volume; therefore they must intersect at least once in $(a, 1)$, with

$$|u_a - u_1| \le \int_a^1 |(u_a)_r - (u_1)_r|\,dr. \tag{28}$$

To estimate the integrand of (28), we apply the mean value theorem to the function $f(p) = p/\sqrt{1-p^2}$, so that

$$|u_r-(u_{n+1})_r| = \frac{|\sin\psi - \sin\psi_{n+1}|}{(1-\xi^2)^{3/2}} \quad \text{implies} \quad |u_a-u_1| \le \int_a^1 \frac{|\sin\psi_a - \sin\psi_1|}{(1-\xi^2)^{3/2}}\,dr,$$

where ξ lies between $\sin\psi_a$ and $\sin\psi_1$. By Lemma 3.2 and (26), we have $-\cos\gamma < \xi < \cos\gamma$, so that $1-\xi^2 > \sin^2\gamma > 0$ for $\gamma \ne 0$. We may bound $|u_a - u_1|$ further:

$$|u_a - u_1| < \int_a^1 \frac{\dfrac{BC(\gamma,a)}{r}\dfrac{(r^2-a^2)(1-r^2)}{1-a^2}}{\sin^3\gamma}\,dr < \frac{B}{a\sin^3\gamma}C(\gamma,a)(1-a^2)(1-a).$$

Finally, we rewrite $C(\gamma,a)$ as

$$C(\gamma,a) = \frac{\cos\gamma\,(1-a^2)}{\sqrt{1-a^2\cos^2\gamma}+\sin\gamma} < \frac{\cos\gamma\,(1-a^2)}{2\sin\gamma},$$

and thus

$$|u_a - u_1| < \frac{B\cos\gamma}{2a\sin^4\gamma}(1-a^2)^2(1-a) = O((1-a)^3) \quad \text{as } a \to 1. \qquad \square$$

For $\gamma = 0$, the term $(1-\xi^2)$ can no longer be assigned a positive lower bound and the argument above does not yield the asymptotic behaviour of u_a as $a \to 1$. Further work is needed to understand this special case.

Finally, we add to Theorem 3.5 by showing that the limiting surface u_1 will in turn approach the lower portion of a torus as $a \to 1$.

Theorem 3.6. *Consider the function*

$$t(r) = -\sqrt{\left(\frac{1-a}{2}\right)^2\sec^2\gamma - \left(r - \frac{1+a}{2}\right)^2} + b(a,\gamma,B),$$

where

$$b(a,\gamma,B)=\frac{2\cos\gamma}{B(1-a)}+\frac{1-a}{8}\sec^2\gamma(\pi-2\gamma-\sin 2\gamma)+\left(\frac{1-a}{2}\right)\tan\gamma.$$

On $[a,1]$, *the function* $t(r)$ *describes the lower portion of a torus that satisfies the boundary conditions of* (2) *and the volume condition* (4). *For* $\gamma\neq 0$, *we have*

$$|u_1-t|=O((1-a)^2)\quad \text{as } a\to 1.$$

Proof. It can be shown that the inclination angle of $t(r)$ is given as

$$\sin\omega(r)=\frac{\cos\gamma}{1-a}(2r-1-a),$$

with $|\sin\psi_1-\sin\omega|$ being maximized on $[a,1]$ at $r=\sqrt{a}$ such that

$$|\sin\psi_1-\sin\omega|\leq\frac{\cos\gamma}{(1+\sqrt{a})^2}(1-a).$$

We argue analagously to the previous theorem that

$$|u_1-t|\leq\int_a^1\frac{|\sin\psi_1-\sin\omega|}{(1-\xi^2)^{3/2}}\,dr,$$

where $-\cos\gamma\leq\sin\omega<\xi<\sin\psi_1\leq\cos\gamma$. For $\gamma\neq 0$, $|u_1-t|$ is then bounded as

$$|u_1-t|<\int_a^1\frac{\frac{\cos\gamma}{(1+\sqrt{a})^2}(1-a)}{\sin^3\gamma}=O((1-a)^2)\quad\text{as } a\to 1.\qquad\square$$

When considered together, Theorems 3.5 and 3.6 allow us to conclude that for $\gamma\neq 0$, the solution surface u_a approaches the torus portion $t(r)$ as $O((1-a)^2)$:

$$|u_a-t|\leq|u_a-u_1|+|u_1-t|=O((1-a)^2)\quad\text{as } a\to 1.$$

Approaching the exterior of a disk. Finally, consider solutions to (5) where $b\to\infty$. Here, we will make use of the exterior solution u_{ext} that solves

$$\begin{cases}(r\sin\psi)_r=Bru_{\text{ext}} & \text{for } r\in(1,\infty),\\ \sin\psi(1)=-\cos\gamma, \\ \lim\limits_{r\to\infty}u(r)=0.\end{cases}$$

See [Siegel 1980] for background. As well, define a sequence of functions $\{w_n\}_{n\geq 2}$ such that w_n is the solution to the boundary value problem

$$\begin{cases}(r\sin\psi)_r=Brw_n & \text{for } r\in(1,n),\\ \sin\psi(1)=-\cos\gamma, \\ \sin\psi(n)=0.\end{cases}$$

We start by demonstrating that $w_n \to u_{\text{ext}}$ as $n \to \infty$. It can be verified that each function w_n, as well as u_{ext}, is decreasing. Also, the comparison principle requires that $w_{n+1} \le w_n$ and $0 < u_{\text{ext}} \le w_n$ for $n \ge 2$. Furthermore, u_{ext} can be shifted vertically to the position of $\bar{u}_{\text{ext}}$ such that

$$\bar{u}_{\text{ext}} = u_{\text{ext}} + w_n(n) - u_{\text{ext}}(n), \tag{29}$$

and the comparison principle gives $u_{\text{ext}} \le w_n \le \bar{u}_{\text{ext}}$ on $[1, n]$. We consider the limit of (29) as $n \to \infty$. Clearly $\lim_{n\to\infty} u_{\text{ext}}(n) = 0$ and we will prove by contradiction that $\lim_{n\to\infty} w_n(n) = 0$. Assume there exists a $\delta > 0$ such that $w_n(n) \ge \delta$ for all $n \ge 2$. This would imply

$$\int_1^n s w_n \, ds > \delta \int_1^n s \, ds \to \infty \quad \text{as } n \to \infty. \tag{30}$$

However, each w_n obeys the volume condition $\int_1^n s w_n \, ds = (\cos\gamma)/B$, which contradicts (30) and necessarily $\lim_{n\to\infty} w_n(n) = 0$. Therefore, (29) provides that $\lim_{n\to\infty} \bar{u}_{\text{ext}} = u_{\text{ext}}$ and $w_n \to u_{\text{ext}}$ and $n \to \infty$.

The behaviour of u as $b \to \infty$ is divided into the following two theorems, with each considering u on the stated subinterval of $[1, b]$.

Theorem 3.7. *For $\gamma \in [0, \pi/2)$, consider the exterior solution u_{ext} defined on $[1, \infty)$ together with u_b, the solution to (5) on $[1, b]$. Let m be the location of the minimum of u_b. On $[1, m]$, we have*

$$\lim_{b\to\infty} \max_{r\in[1,m]} |u_b(r) - u_{\text{ext}}(r)| = 0.$$

Furthermore, $m = m(b)$ is monotone increasing and $m(b) \to \infty$ as $b \to \infty$.

Proof. We compare the three functions u_{ext}, u_b and $\hat{u}_{\text{ext}}$ on $[1, m]$, where

$$\hat{u}_{\text{ext}} = u_{\text{ext}} + u_b(m) - u_{\text{ext}}(m),$$

with $u_{\text{ext}} \le u_b \le \hat{u}_{\text{ext}}$ on $[1, m]$ by the comparison principle. Similar to Lemma 3.3, we have

$$\lim_{b\to\infty} m(b) = \infty \quad \text{implies} \quad \lim_{b\to\infty} u_b(m) = u_{\text{ext}}(m),$$

and it is sufficient to show that $\lim_{b\to\infty} m(b) = \infty$, since this would require

$$\max_{r\in[1,m]} |u_b(r) - u_{\text{ext}}(r)| \le \hat{u}_{\text{ext}} - u_{\text{ext}} = u_b(m) - u_{\text{ext}}(m)$$
$$\to 0 \quad \text{as } b \to \infty.$$

An argument nearly identical to that proving Theorem 2.5 yields that $m(b)$ is monotone increasing. Furthermore, m increases without bound as $b \to \infty$, which can be shown by contradiction: Assume there exists an $M \in \mathbb{N}$ such that $m(b) \leq M$. The volume condition on u can be used to show that

$$\int_M^b su\,ds \leq \int_m^b su\,ds = \frac{b\cos\gamma}{B}. \tag{31}$$

Additionally, select the function $w_M \in \{w_n\}$ as a lower bound of u on $[1, M]$, so that $w_M(M) \leq w_M \leq u$ on $[1, M]$ by the comparison principle. With (31), this produces

$$w_M(M)(\tfrac{1}{2}(b^2 - M^2)) < \int_M^b su\,ds \leq \frac{b\cos\gamma}{B}. \tag{32}$$

For large enough b, however, (32) cannot hold, and $m \to \infty$ as $b \to \infty$. □

The examination of u on the remaining interval $[m, b]$ will refer to the one-dimensional solution $z(x)$ that solves

$$\begin{cases} \left(\dfrac{z_x}{\sqrt{1+{z_x}^2}}\right)_x = Bz \quad \text{for } x \in (0,\infty), \\ \sin\phi(0) = -\cos\gamma, \\ \lim\limits_{x\to\infty} z(x) = 0, \end{cases} \tag{33}$$

where $\phi(x)$ denotes the inclination angle of $z(x)$. This problem was first considered by Laplace [1966]; a modern treatment is offered by Siegel [1980]. Physically, z represents the height of a capillary surface on one side of an infinite vertical plate.

Theorem 3.8. *Let $\gamma \in [0, \pi/2)$ and define u_b and m as in the previous theorem. Consider the one-dimensional solution z that satisfies (33). On $[m, b]$, we have*

$$\lim_{b\to\infty} \max_{s\in[0,b-m]} |u_b(b-s) - z(s)| = 0.$$

Proof. We employ the functions $z(s)$ and $u_b(b-s)$ for $s \in [0, b-m]$. This amounts to comparing the annular solution with the capillary surface generated by an infinte plate placed tangentially to the outer boundary of Ω. We also introduce the function $\hat{z}$ defined as $\hat{z}(s) = z(s) + u_b(m) - z(b-m)$. Our comparisons will be largely based upon the results of Siegel [1980], where a similar geometry was used to compare the surface z with the interior solution. In our case, the comparison principle requires $z \leq u_b \leq \hat{z}$ and more specifically, $z(s) \leq u_b(b-s) \leq \hat{z}(s)$ for $s \in [0, b-m]$. Thus

$$\max_{s\in[0,b-m]} |u_b(b-s) - z(s)| \leq \hat{z} - z = u_b(m) - z(b-m).$$

From Theorem 3.7, it is clear that $u_b(m) \to u_{\text{ext}}(m) \to 0$ as $b \to \infty$. Additionally, since $m < (1+b)/2$, we have

$$b - m > (b-1)/2 \to \infty \quad \text{as } b \to \infty \quad \text{and} \quad \lim_{b\to\infty} z(b-m) = 0.$$

Therefore, $\max_{s\in[0,b-m]} |u_b(b-s) - z(s)| \to 0$ as $b \to \infty$. □

References

[Concus and Finn 1974] P. Concus and R. Finn, "On capillary free surfaces in the absence of gravity", *Acta Math.* **132** (1974), 177–198. MR 58 #32327a Zbl 0382.76003

[Elcrat et al. 2004] A. Elcrat, T.-E. Kim, and R. Treinen, "Annular capillary surfaces", *Arch. Math. (Basel)* **82**:5 (2004), 449–467. MR 2006e:76055 Zbl 1136.76353

[Finn 1986] R. Finn, *Equilibrium capillary surfaces*, Grundlehren der Mathematischen Wissenschaften **284**, Springer, New York, 1986. MR 88f:49001 Zbl 0583.35002

[Laplace 1966] P. S. Laplace, *Celestial mechanics*, vol. 4, Chelsea, New York, 1966. MR 42 #28 Zbl 1047.01531

[Serrin 1971] J. Serrin, "A symmetry problem in potential theory", *Arch. Rational Mech. Anal.* **43** (1971), 304–318. MR 48 #11545 Zbl 0222.31007

[Siegel 1980] D. Siegel, "Height estimates for capillary surfaces", *Pacific J. Math.* **88**:2 (1980), 471–515. MR 82h:35037 Zbl 0411.35043

[Siegel 2006] D. Siegel, "Approximating symmetric capillary surfaces", *Pacific J. Math.* **224**:2 (2006), 355–365. MR 2007a:76022 Zbl 1116.76015

Received January 8, 2009.

JAMES GORDON
DEPARTMENT OF APPLIED MATHEMATICS
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO N2L 3G1
CANADA

james.gordon@utoronto.ca

DAVID SIEGEL
DEPARTMENT OF APPLIED MATHEMATICS
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO N2L 3G1
CANADA

james.gordon1832@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

APPROXIMATING ANNULAR CAPILLARY SURFACES WITH EQUAL CONTACT ANGLES

JAMES GORDON AND DAVID SIEGEL

Consider an annular region $\Omega \subset \mathbb{R}^2$. We extend the iterative procedure of Siegel to the case of symmetric capillary surfaces $z = u(x, y)$ formed within the annular cylinder $\Omega \times \mathbb{R}$ and having identical contact angles γ along the inner and outer boundaries. We demonstrate convergence under conditions that include $\gamma = 0$, and we recover the interleaving properties noted by Siegel for a particular geometry.

1. Introduction

We continue our examination of annular capillary surfaces of the form described in [Gordon and Siegel 2010]: Two concentric circular cylinders that define an annular cross section $\Omega \subset \mathbb{R}^2$ are immersed vertically in an infinite reservoir of incompressible fluid. Under the influence of gravity, the surface $Z = U(X, Y)$ formed between the tubes will satisfy the boundary value problem

$$\begin{cases} NU = \kappa U & \text{in } \Omega, \\ \hat{\nu} \cdot TU = \cos\gamma & \text{on } \partial\Omega, \end{cases} \tag{1}$$

where $TU = \nabla U / \sqrt{1 + |\nabla U|^2}$, $NU = \nabla \cdot TU$, the exterior unit normal on the boundary $\partial\Omega$ is $\hat{\nu}$, and $\kappa > 0$ is the capillary constant. The contact angle $\gamma \in [0, \pi]$ is defined on the inner and outer boundaries and is the angle at which the interface meets the bounding wall. Again, γ is assumed to be constant and equal along each boundary.

The axisymmetric nature of such annular solutions, along with the change to dimensionless variables

$$u = U/R_2 \quad \text{and} \quad r = R/R_2$$

MSC2000: primary 34B15, 35J25, 76B45; secondary 34E10.
Keywords: annular capillary surface, iterative approximations.
Gordon is supported by an Ontario Graduate Scholarship. Siegel is supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant.

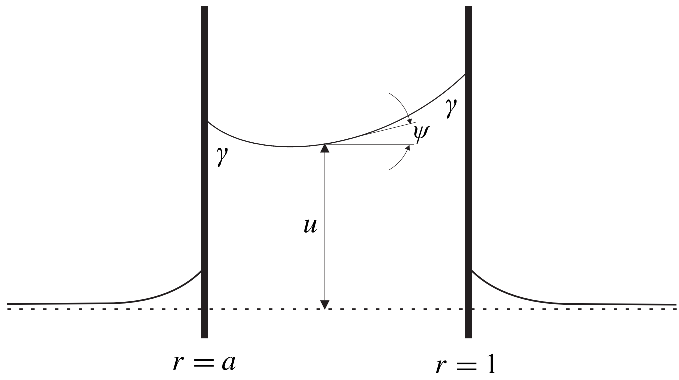


Figure 1. Radial cross section of annular capillary surface.

allows us to convert (1) into a boundary value problem for an ordinary differential equation:

$$\begin{cases} Nu = \dfrac{(r\sin\psi)_r}{r} = Bu & \text{for } a < r < 1, \\ \sin\psi(a) = -\cos\gamma, \\ \sin\psi(1) = \cos\gamma \end{cases} \tag{2}$$

where $B > 0$ is a positive constant known as the Bond number, $(\cdot)_r$ denotes differentiation with respect to the radial variable r, and we define the inclination angle $\psi(r)$ of $u(r)$ as $\sin\psi(r) = u_r/\sqrt{1+u_r^2}$. See Figure 1. Note also that the outer boundary of Ω is now fixed at $r = 1$, while the inner boundary occurs at $r = a$, where $0 < a < 1$. For reasons described in [Gordon and Siegel 2010], we need only consider $0 \le \gamma < \pi/2$. Additionally, u is positive and bounded by the comparison principle [Concus and Finn 1974; Finn 1986], and the volume lifted can be determined from

$$\int_a^1 ru(r)\,dr = \frac{\cos\gamma(1+a)}{B}. \tag{3}$$

In this paper, we apply the iterative procedure introduced by Siegel [2006] to the boundary value problem considered here. Specifically, Section 2 provides conditions under which the approximate functions generated converge to the solution of (2). In [2006], Siegel demonstrated that the iterates display a highly organized interplay, which he called *interleaving properties*. This allowed for bounds to be placed on the surface height at the inner and outer radii. In Section 3, we prove that, under certain conditions, the interleaving properties may be recovered for (2). Section 4 summarizes a new and more complicated behavior that can also occur.

2. Iterative procedure

The iterative scheme of [Siegel 2006] was employed successfully to approximate annular surfaces with the inner contact angle fixed at $\pi/2$. We intend to extend the procedure to the boundary value problem (2). An outline follows: Consider a function u_1 that satisfies the volume condition (3), and suppose there exists a function u_2 such that $Nu_2 = Bu_1$ or equivalently

$$(r \sin \psi_2)_r = Bru_1,$$

where ψ_2 is the inclination angle of u_2. Requiring $\sin \psi_2(a) = -\cos \gamma$, we arrive at an integral equation for ψ_2:

$$\sin \psi_2(r) = \frac{B}{r} \int_a^r s u_1(s)\, ds - \frac{a}{r} \cos \gamma.$$

Since u_1 satisfies the volume requirement, it is easily verified that u_2 also has the correct boundary condition at $r = 1$. Given that $(u_2)_r = \sin \psi_2 / \sqrt{1 - \sin^2 \psi_2}$, we derive an expression for u_2:

$$u_2(r) = u_2(a) + \int_a^r \frac{\sin \psi_2(s)}{\sqrt{1 - \sin^2 \psi_2(s)}}\, ds, \tag{4}$$

with $u_2(a)$ selected so that u_2 has the correct volume, that is,

$$u_2(a) = \frac{2 \cos \gamma}{B(1-a)} - \frac{1}{1-a^2} \int_a^1 (1 - s^2) \frac{\sin \psi_2(s)}{\sqrt{1 - \sin^2 \psi_2(s)}}\, ds. \tag{5}$$

The following theorem solidifies these ideas. Here, a complicating assumption of $\sin \psi_2 \le r \cos \gamma$ arises that did not occur in Siegel's previous analysis.

Theorem 2.1. *Let u_1 be a continuous, positive function defined on $[a, 1]$ that satisfies the volume condition (3). Define*

$$\sin \psi_2(r) = \frac{B}{r} \int_a^r s u_1(s)\, ds - \frac{a}{r} \cos \gamma$$

and assume $\sin \psi_2 \le r \cos \gamma$.

(i) *We have $-(a/r) \cos \gamma \le \sin \psi_2$ on $[a, 1]$.*

(ii) *There exists a function u_2 defined and continuous on $[a, 1]$ given as*

$$u_2(r) = u_2(a) + \int_a^r \frac{\sin \psi_2(s)}{\sqrt{1 - \sin^2 \psi_2(s)}}\, ds$$

with

$$u_2(a) = \frac{2 \cos \gamma}{B(1-a)} - \frac{1}{1-a^2} \int_a^1 (1 - s^2) \frac{\sin \psi_2(s)}{\sqrt{1 - \sin^2 \psi_2(s)}}\, ds.$$

As a result, $Nu_2 = Bu_1$.

(iii) *The function u_2 satisfies both the volume condition and the boundary conditions listed in (2).*

(iv) *There is a unique point $r = m_2$ at which u_2 achieves its minimum value.*

(v) *If*

$$B < \frac{2}{(1-a)(\frac{1}{3}\sqrt{1-a^2}+a\log(1+\sqrt{1-a^2})-a\log a)},$$

then u_2 will also be positive.

Proof. (i) Using that u_1 is positive, we note that $(r\sin\psi_2)_r = Bru_1 > 0$ and the function $r\sin\psi_2$ is monotone increasing. The remainder of the proof mirrors [Gordon and Siegel 2010, Lemma 3.2].

(ii) To show u_2 is defined and continuous, it suffices to show that u_2 is bounded. Since the function $p/\sqrt{1-p^2}$ is increasing on $(-1, 1)$ with

$$-(a/r)\cos\gamma \le \sin\psi_2 \le r\cos\gamma,$$

Equations (4) and (5) give

$$\begin{aligned} u_2(r) &= \frac{2\cos\gamma}{B(1-a)} - \frac{1}{1-a^2}\int_a^1 (1-s^2)\frac{\sin\psi_2}{\sqrt{1-\sin^2\psi_2}}\,ds \\ &\qquad\qquad + \int_a^r \frac{\sin\psi_2}{\sqrt{1-\sin^2\psi_2}}\,ds \\ &\ge \frac{2\cos\gamma}{B(1-a)} - \frac{1}{1-a^2}\int_a^1 (1-s^2)\frac{s\cos\gamma}{\sqrt{1-s^2\cos^2\gamma}}\,ds \\ &\qquad\qquad - \int_a^r \frac{a\cos\gamma}{\sqrt{s^2-a^2\cos^2\gamma}}\,ds \\ &\ge \cos\gamma\Big(\frac{2}{B(1-a)} - \tfrac{1}{3}\sqrt{1-a^2} - a\log(1+\sqrt{1-a^2}) + a\log a\Big). \end{aligned} \tag{6}$$

Given that $a\log a \ge -1/e$ on $(0, 1]$, we can bound u_2 below:

$$u_2(r) > \cos\gamma(2/B - 1/3 - \log 2 - 1/e) > -\infty,$$

Similarly, u_2 can be bounded above:

$$\begin{aligned} u_2(r) &\le \frac{2\cos\gamma}{B(1-a)} + \frac{1}{1-a^2}\int_a^1 (1-s^2)\frac{a\cos\gamma}{\sqrt{s^2-a^2\cos^2\gamma}}ds + \int_a^r \frac{s\cos\gamma}{\sqrt{1-s^2\cos^2\gamma}}\,ds \\ &< \cos\gamma\Big(\frac{2}{B(1-a)} + \log 2 + \frac{1}{e} + 1\Big) < \infty. \end{aligned}$$

Finally, the introductory discussion confirms that $Nu_2 = Bu_1$.

(iii) This is easily proven from the volume condition.

(iv) This argument follows the proof of [Gordon and Siegel 2010, Theorem 2.1].

(v) The lower bound given in (6) is required to be positive:

$$u_2(r) \geq \cos\gamma\left(\frac{2}{B(1-a)} - \tfrac{1}{3}\sqrt{1-a^2} - a\log(1+\sqrt{1-a^2}) + a\log a\right) > 0.$$

Solving for B produces the desired result. □

Theorem 2.1 creates the framework needed to generate a sequence of iterates $\{u_n\}$ defined recursively as

$$Nu_{n+1} = Bu_n \quad \text{for } n \geq 0. \tag{7}$$

We take the initial function u_0 to be the constant function that satisfies the volume condition:

$$u_0 = \frac{2\cos\gamma}{B(1-a)}. \tag{8}$$

It can be shown that for suitable restrictions on B, the sequence $\{u_n\}$ is one in which

- (i) through (v) of Theorem 2.1 are satisfied for all $n \geq 1$, and
- $\{u_n\}$ converges to the solution of the boundary value problem (2).

The following theorem demonstrates these results.

Theorem 2.2 (iterate convergence). *For*

$$B < \frac{2a(1-a^2)\cos\gamma}{2(1+a)(1-a)^2 C(\gamma,m) + a\pi\cos\gamma},$$

the sequence of iterates $\{u_n\}$ generated via (7) and (8) will be continuous and positive. Furthermore, $B\pi/(2(1-a^2)) < 1$ and $\{u_n\}$ converges to u, the solution of (2), with

$$|u-u_n| < C(\gamma,m)\left(B\frac{\pi}{2(1-a^2)}\right)^n.$$

Here, $C(\gamma,m) = (\sqrt{1-m^2\cos^2\gamma} - \sin\gamma)/\cos\gamma$ with $0 < C(\gamma,m) < C(\gamma,a)$, and $r = m$ is the location of the minimum of u.

Proof. We first prove the case for $n = 0$ and proceed inductively. Since u_0 and u satisfy the volume condition, they must intersect at least once in $(a, 1)$. Using that u is convex and $u(a) < u(1)$ [Gordon and Siegel 2010], $|u-u_0|$ is thus bounded as

$$|u-u_0| \leq \max\{u(1)-u_0, u_0-u(m)\} < u(1)-u(m) = \int_m^1 \frac{\sin\psi}{\sqrt{1-\sin^2\psi}}\,dr.$$

Additionally, [Gordon and Siegel 2010, Lemma 3.2] provides that

$$\frac{\sin\psi}{\sqrt{1-\sin^2\psi}} \le \frac{r\cos\gamma}{\sqrt{1-r^2\cos^2\gamma}}$$

and consequently

$$|u-u_0| < \int_m^1 \frac{r\cos\gamma}{\sqrt{1-r^2\cos^2\gamma}}\,dr = \frac{\sqrt{1-m^2\cos^2\gamma}-\sin\gamma}{\cos\gamma} := C(\gamma,m).$$

The case $n=0$ is thus proved. Next, assume u_n is continuous, positive, and satisfies the volume condition. Also, let $|u-u_n| < \beta_n := C(\gamma,m)(B\pi/(2(1-a^2)))^n$. The defining equations for $\{u_n\}$ and u yield

$$\sin\psi - \sin\psi_{n+1} = \frac{B}{r}\int_a^r s(u-u_n)\,ds,$$

or equivalently, $\sin\psi - \sin\psi_{n+1} = -(B/r)\int_r^1 s(u-u_n)\,ds$. When used in tandem, these imply

$$|\sin\psi - \sin\psi_{n+1}| \le \frac{B}{2r}\beta_n \min\{r^2-a^2, 1-r^2\},$$

and since $\min\{r^2-a^2, 1-r^2\} \le 2(r^2-a^2)(1-r^2)/(1-a^2)$, this gives

$$|\sin\psi - \sin\psi_{n+1}| \le \frac{\beta_n B}{r}\frac{(r^2-a^2)(1-r^2)}{1-a^2}. \tag{9}$$

We next bound $\sin\psi_{n+1}$. For $n=0$, $\sin\psi_1$ can be written exactly:

$$\sin\psi_1 = \frac{\cos\gamma}{1-a}\left(r-\frac{a}{r}\right), \tag{10}$$

and it is easily checked that $-(a/r)\cos\gamma \le \sin\psi_1 \le r\cos\gamma$. For $n\ge 1$, we do not have the luxury of an explicit function and we proceed as follows: To show $\sin\psi_{n+1} \le r\cos\gamma$, consider the distance between $\sin\psi_1$ and $\sin\psi_{n+1}$:

$$\begin{aligned}|\sin\psi_1 - \sin\psi_{n+1}| &\le |\sin\psi - \sin\psi_1| + |\sin\psi - \sin\psi_{n+1}| \\ &\le \frac{B}{r}C(\gamma,m)\frac{(r^2-a^2)(1-r^2)}{1-a^2}\left(1+\left(B\frac{\pi}{2(1-a^2)}\right)^n\right).\end{aligned} \tag{11}$$

The last factor in (11) can be bounded by a geometric series:

$$1+\left(B\frac{\pi}{2(1-a^2)}\right)^n < \sum_{k=0}^{\infty}\left(B\frac{\pi}{2(1-a^2)}\right)^k,$$

and since

$$B < \frac{2a(1-a^2)\cos\gamma}{2(1+a)(1-a)^2C(\gamma,m)+a\pi\cos\gamma} < \frac{2(1-a^2)}{\pi},$$

the sum is convergent. We now have

$$|\sin\psi_1 - \sin\psi_{n+1}| \le \frac{B}{r} C(\gamma, m) \frac{2(r^2-a^2)(1-r^2)}{2(1-a^2)-B\pi}. \tag{12}$$

The condition on B can be substituted into (12) to obtain

$$|\sin\psi_1 - \sin\psi_{n+1}| \le \frac{a\cos\gamma}{1-a}\Big(\frac{1}{r}-r\Big).$$

With this in hand, we are able to show that $\sin\psi_{n+1} \le r\cos\gamma$:

$$\begin{aligned} r\cos\gamma - \sin\psi_{n+1} &\ge (r\cos\gamma - \sin\psi_1) - |\sin\psi_1 - \sin\psi_{n+1}| \\ &\ge \bigg(r\cos\gamma - \frac{\cos\gamma}{1-a}\Big(r-\frac{a}{r}\Big)\bigg) - \bigg(\frac{a\cos\gamma}{1-a}\Big(\frac{1}{r}-r\Big)\bigg) = 0, \end{aligned}$$

and conditions (i)–(iv) of Theorem 2.1 apply to u_{n+1}. In addition, the bound on B required here satisfies

$$\begin{aligned} B &< \frac{2a(1-a^2)\cos\gamma}{2(1+a)(1-a)^2 C(\gamma, m) + a\pi\cos\gamma} \\ &< \frac{2}{(1-a)(\frac{1}{3}\sqrt{1-a^2}+a\log(1+\sqrt{1-a^2})-a\log a)}, \end{aligned}$$

and consequently, u_{n+1} also exhibits property (v) of Theorem 2.1. To summarize, u_{n+1} will be continuous, positive and will satisfy the volume condition.

Next, we bound $|u - u_{n+1}|$. Since both u and u_{n+1} have the correct volume, they must intersect at least once in $(a, 1)$. This allows us to state that

$$|u - u_{n+1}| \le \int_a^1 |u_r - (u_{n+1})_r|\, dr. \tag{13}$$

To estimate the integrand of (13), we use the mean value theorem on the function $f(p) = p/\sqrt{1-p^2}$, so that

$$\frac{f(\sin\psi) - f(\sin\psi_{n+1})}{\sin\psi - \sin\psi_{n+1}} = f'(\xi),$$

where ξ lies between $\sin\psi$ and $\sin\psi_{n+1}$. This can be rewritten as

$$|u_r - (u_{n+1})_r| = \frac{|\sin\psi - \sin\psi_{n+1}|}{(1-\xi^2)^{3/2}}. \tag{14}$$

The numerator of (14) has an upper bound given in (9). For the denominator, [Gordon and Siegel 2010, Lemma 3.2] provides bounds on $\sin\psi$ that are identical to those derived above for $\sin\psi_{n+1}$:

$$-\frac{a\cos\gamma}{r} \le \sin\psi \quad \text{and} \quad \sin\psi_{n+1} \le r\cos\gamma,$$

and ξ is bounded as

$$-\frac{a}{r} \le -\frac{a\cos\gamma}{r} < \xi < r\cos\gamma \le r,$$

with $\xi^2 < \max\{a^2/r^2, r^2\}$. The denominator of (14) can thus be estimated using $1-\xi^2 > (1-a^2/r^2)(1-r^2)$, and an upper bound on $|u-u_{n+1}|$ is now possible:

$$\begin{aligned}|u-u_{n+1}| &\le \int_a^1 \frac{|\sin\psi - \sin\psi_{n+1}|}{(1-\xi^2)^{3/2}}\,dr \\ &< \frac{\beta_n B}{1-a^2}\int_a^1 \frac{r^2}{\sqrt{r^2-a^2}\sqrt{1-r^2}}\,dr = \frac{\beta_n B}{1-a^2}\int_0^1 \frac{\sqrt{1-(1-a^2)t^2}}{\sqrt{1-t^2}}\,dt,\end{aligned}$$

where the change of variables $t=\sqrt{(r^2-1)/(a^2-1)}$ is used in the last equality. This integral is always less than $\pi/2$. Hence,

$$|u-u_{n+1}| < \frac{\beta_n B}{1-a^2}\frac{\pi}{2} = \beta_{n+1}$$

and the inductive step is complete. □

3. Single intersection case: interleaving properties

The iterates of [Siegel 2006] were shown there to exhibit the following structure:

(a) $\psi_0 < \psi_2 < \cdots < \psi < \cdots < \psi_3 < \psi_1$ for $r \in (a, 1)$.

(b) $u_1(a) < u_3(a) < \cdots < u(a) < \cdots < u_2(a) < u_0$.

(c) $u_0 < u_2(1) < \cdots < u(1) < \cdots < u_3(1) < u_1(1)$.

These properties were defined collectively by Siegel as the *interleaving properties* of the iterates, with (b) and (c) providing under- and over-estimates for the boundary values of u. Here, the behavior between iterates is more complex, being sensitive to the values of the parameters a, γ and B. However, we are able to recover these interleaving properties under certain conditions. It so happens that it will be necessary to find selections of a, γ and B such that u, u_0 and u_2 will be configured as noted in Figure 2. In other words, there exist unique points b_0 and c_0 in $(a, 1)$ such that

$$\begin{cases} u < u_0 & \text{if } r \in [a, b_0), \\ u > u_0 & \text{if } r \in (b_0, 1], \end{cases} \tag{15}$$

and

$$\begin{cases} u_2 < u_0 & \text{if } r \in [a, c_0), \\ u_2 > u_0 & \text{if } r \in (c_0, 1]. \end{cases} \tag{16}$$

It turns out that if (15) and (16) occur, the interleaving properties are a consequence. We examine the conditions necessary for each configuration below.

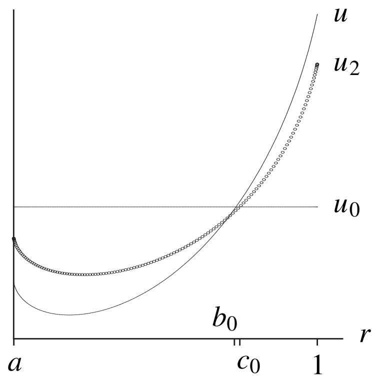


Figure 2. The configuration required for interleaving properties. Whereas u and u_2 must be arranged as shown with respect to u_0, the relative configuration of u and u_2 is not important.

Single intersection of u with u_0. We will show that the configuration of (15) is a result of $u(a) < u_0$, and it is indeed possible to find conditions under which this is true through a comparison with $u_1(a)$. We thus begin by investigating the conditions necessary for $u_1(a) < u_0$. Consider the difference

$$u_0 - u_1(a) = \frac{1}{1-a^2}\int_a^1 (1-s^2)\frac{\sin\psi_1}{\sqrt{1-\sin^2\psi_1}}\,ds. \tag{17}$$

The integrand can be bounded from below by estimating $\sin\psi_1/\sqrt{1-\sin^2\psi_1}$. For $a \le r \le \sqrt{a}$, we use that $\sin\psi_1$ is concave to see $(r-\sqrt{a})\cos\gamma/(\sqrt{a}-a) \le \sin\psi_1$, which gives

$$\frac{\sin\psi_1}{\sqrt{1-\sin^2\psi_1}} \ge \frac{(r-\sqrt{a})\dfrac{\cos\gamma}{\sqrt{a}-a}}{\sqrt{1-\left((r-\sqrt{a})\dfrac{\cos\gamma}{\sqrt{a}-a}\right)^2}},$$

and for $\sqrt{a} \le r < 1$, it is easily seen that $\sin\psi_1/\sqrt{1-\sin^2\psi_1} \ge \sin\psi_1$. Using these bounds in (17), we have

$$u_0 - u_1(a) > \sqrt{\frac{(\sqrt{a}-a)^2}{\cos^2\gamma} - (\sqrt{a}-a)^2} - \frac{\sqrt{a}-a}{\cos\gamma} + \frac{\cos\gamma(1-a^2+2a\log a)}{4(1-a)(1-a^2)},$$

and since $a\log a > (a-1) + \frac{1}{2}(a-1)^2 - \frac{1}{6}(a-1)^3$ on $(0, 1)$, we can arrive at

$$u_0 - u_1(a) > \cos\gamma\left[\frac{1-a}{12(1+a)} - (\sqrt{a}-a)\right].$$

To ensure that $u_1(a) < u_0$, the expression in square brackets must be nonnegative, and after some mechanics, this is shown to be true for $a \in (0, \Lambda^2]$ where

$$\Lambda := \tfrac{1}{6}(9+2\sqrt{353})^{1/3} - \frac{11}{6(9+2\sqrt{353})^{1/3}} \doteq 0.09.$$

To compare $u(a)$ with u_0, we employ the same technique used in the proof of Theorem 2.2, namely, find B so that $u(a)$ lies close enough to $u_1(a)$ and necessarily $u(a) < u_0$.

Theorem 3.1. *For any $\gamma \in [0, \pi/2)$, select $a \le \Lambda^2$ and*

$$B \le \frac{2(1-a^2)\cos\gamma}{\pi}\Big(\frac{1-a}{12(1+a)} - (\sqrt{a}-a)\Big).$$

Under these conditions, there exists a unique $b_0 \in (a, 1)$ such that $u(b_0) = u_0$ with

$$\begin{cases} u < u_0 & \text{if } r \in [a, b_0), \\ u > u_0 & \text{if } r \in (b_0, 1]. \end{cases}$$

Proof. As mentioned, we will require u_1 as a comparison function and it must be confirmed that u_1 is defined and continuous. In fact, since $\sin\psi_1 \le r\cos\gamma$ (see (10)) and since

$$\begin{aligned} B &\le \frac{2(1-a^2)\cos\gamma}{\pi}\Big(\frac{1-a}{12(1+a)} - (\sqrt{a}-a)\Big) \\ &< \frac{2}{(1-a)(\frac{1}{3}\sqrt{1-a^2}+a\log(1+\sqrt{1-a^2})-a\log a)}, \end{aligned}$$

u_1 exhibits all properties of Theorem 2.1. To continue, we write

$$u_0 - u(a) \ge (u_0 - u_1(a)) - |u(a) - u_1(a)|. \tag{18}$$

The selection of a ensures the first term of (18) is positive. For the second term, the proof of Theorem 2.2 for the specific case of u_1 yields

$$|u - u_1| < C(\gamma, m)B\frac{\pi}{2(1-a^2)} < B\frac{\pi}{2(1-a^2)}.$$

with $C(\gamma, m) < 1$. The difference of (18) can now be bounded as

$$u_0 - u(a) > \cos\gamma\Big(\frac{1-a}{12(1+a)} - (\sqrt{a}-a)\Big) - B\frac{\pi}{2(1-a^2)}. \tag{19}$$

Substituting the condition on B produces the desired result,

$$u(a) < u_0. \tag{20}$$

With both u and u_0 having the correct volume, at least one intersection occurs between these functions. The convexity of u, in conjunction with (20), limits this to a unique intersection occurring at a point $b_0 \in (a, 1)$. □

Single intersection of u_2 with u_0. In like manner, we are able to find conditions for $u_2(a) < u_0$ that will result in (16). Before turning to this, it should be verified that under the hypotheses of the previous theorem, u_2 is defined and continuous.

Lemma 3.2 [Siegel 2006]. *Consider two functions v and w defined on $[a, 1]$ with inclination angles given by ψ_v and ψ_w respectively. If $\psi_v < \psi_w$ on $(a, 1)$ and $\int_a^1 rv\,dr = \int_a^1 rw\,dr$, then there exists a unique $b \in (a, 1)$ where $v(b) = w(b)$ and*

$$\begin{cases} w < v & \text{if } r \in [a, b), \\ w > v & \text{if } r \in (b, 1]. \end{cases}$$

To show u_2 is defined and continuous, we first write the difference function

$$r \sin\psi - r \sin\psi_1 = B \int_a^r s(u - u_0)\,ds,$$

which is zero at $r = a$ and $r = 1$. Theorem 3.1 also ensures the function contains a unique extremum at $r = b_0$. Thus, $r \sin\psi - r \sin\psi_1$ must be either positive or negative on $(a, 1)$. It follows from $u(a) < u_0$ that

$$r \sin\psi - r \sin\psi_1 < 0 \quad \text{implies} \quad \psi < \psi_1 \quad \text{for } r \in (a, 1),$$

and Lemma 3.2 requires that there exist a unique $b_1 \in (a, 1)$ where $u_1(b_1) = u(b_1)$ and

$$\begin{cases} u_1 < u & \text{if } r \in [a, b_1), \\ u_1 > u & \text{if } r \in (b_1, 1]. \end{cases}$$

With this in hand, we consider the difference $r \sin\psi - r \sin\psi_2 = B \int_a^r s(u - u_1)\,ds$ and reason accordingly that $\sin\psi_2 < \sin\psi < r \cos\gamma$ for $r \in (a, 1)$. With B bounded as in Theorem 3.1, u_2 will exhibit all properties of Theorem 2.1.

Conditions can now be stated so that $u_2(a) < u_0$. In this case, B will be restricted further than in Theorem 3.1.

Theorem 3.3. *For any $\gamma \in [0, \pi/2)$, select $a \le \Lambda^2$ and*

$$B \le \frac{(1 - a^2)\cos\gamma}{\pi}\left(\frac{1-a}{12(1+a)} - (\sqrt{a} - a)\right).$$

Under these conditions, u_2 satisfies the properties of Theorem 2.1 and there exists a unique $c_0 \in (a, 1)$ such that $u_2(c_0) = u_0$, with

$$\begin{cases} u_2 < u_0 & \text{if } r \in [a, c_0), \\ u_2 > u_0 & \text{if } r \in (c_0, 1]. \end{cases}$$

Proof. Similarly, we write $u_0 - u_2(a) \ge (u_0 - u(a)) - |u(a) - u_2(a)|$. The first term can be bounded as in (19), and an argument identical to the proof of Theorem 2.2

specifically for u_2 estimates the second term. We thus have

$$u_0 - u_2(a) > \cos\gamma\Big(\frac{1-a}{12(1+a)} - (\sqrt{a}-a)\Big) - B\frac{\pi}{2(1-a^2)}\Big(1 + B\frac{\pi}{2(1-a^2)}\Big),$$

and substituting for B gives $u_2(a) < u_0$. As before, the volume condition guarantees an intersection between the two functions. Theorem 2.1(iv) implies that u_2 is monotone decreasing on $[a, m_2)$ and monotone increasing on $(m_2, 1]$. Hence, the intersection is unique and the arrangement between functions easily follows. □

Interleaving properties. We can now prove the interleaving properties for $\{u_n\}$. Here, a and B are restricted so that the single intersection case is guaranteed to occur.

Theorem 3.4. *For any $\gamma \in [0, \pi/2)$, select $a \le \Lambda^2$ and*

$$B \le \frac{(1-a^2)\cos\gamma}{\pi}\Big(\frac{1-a}{12(1+a)} - (\sqrt{a}-a)\Big).$$

Under these conditions, the sequence of iterates $\{u_n\}$ defined by (7) and (8) satisfies (i) through (v) of Theorem 2.1. The iterates exhibit the following properties:

(1) $\psi_2 < \psi_4 < \cdots < \psi < \cdots < \psi_3 < \psi_1$ *for $r \in (a, 1)$.*

(2) $u_1(a) < u_3(a) < \cdots < u(a) < \cdots < u_4(a) < u_2(a)$.

(3) $u_2(1) < u_4(1) < \cdots < u(1) < \cdots < u_3(1) < u_1(1)$.

Proof. We first go through a cycle of recursive arguments and proceed to show that the base case, which is known to be true, sets the cycle in motion. To start, assume that for a certain $k \ge 0$,

(a) u_{2k}, u_{2k+1}, and u_{2k+2} satisfy (i) through (v) of Theorem 2.1 (although u_{2k} does not have to satisfy the boundary conditions);

(b) $\psi < \psi_{2k+1}$ with $\sin\psi_{2k+1} < r\cos\gamma$ for $r \in (a, 1)$.

(c) there exists a unique $c_{2k} \in (a, 1)$ such that

$$\begin{cases} u_{2k+2} < u_{2k} & \text{if } r \in [a, c_{2k}), \\ u_{2k+2} > u_{2k} & \text{if } r \in (c_{2k}, 1]. \end{cases}$$

From (b), Lemma 3.2 requires that there exists a unique $b_{2k+1} \in (a, 1)$ with

$$\begin{cases} u > u_{2k+1} & \text{if } r \in [a, b_{2k+1}), \\ u < u_{2k+1} & \text{if } r \in (b_{2k+1}, 1]. \end{cases}$$

Using the difference function

$$r\sin\psi - r\sin\psi_{2k+2} = B\int_a^r s(u - u_{2k+1})\,ds,$$

we can show that $\sin\psi_{2k+2} < \sin\psi < r\cos\gamma$ for $r \in (a,1)$. This implies that $\psi_{2k+2} < \psi$ on $(a,1)$. Lemma 3.2 can be used again:

$$\begin{cases} u < u_{2k+2} & \text{if } r \in [a, b_{2k+2}), \\ u > u_{2k+2} & \text{if } r \in (b_{2k+2}, 1]. \end{cases}$$

and a new difference function

$$r\sin\psi - r\sin\psi_{2k+3} = B\int_a^r s(u - u_{2k+2})\,ds$$

produces $\psi < \psi_{2k+3}$ on $(a,1)$. It must now be verified that u_{2k+3} is defined. For this, we use

$$r\sin\psi_{2k+1} - r\sin\psi_{2k+3} = B\int_a^r s(u_{2k} - u_{2k+2})\,ds$$

along with (b) and (c) to reason that $\sin\psi_{2k+3} < \sin\psi_{2k+1} < r\cos\gamma$ for $r \in (a,1)$. With B restricted as hypothesized, u_{2k+3} now obeys Theorem 2.1, and Lemma 3.2 ensures

$$\begin{cases} u_{2k+3} > u_{2k+1} & \text{if } r \in [a, c_{2k+1}), \\ u_{2k+3} < u_{2k+1} & \text{if } r \in (c_{2k+1}, 1]. \end{cases}$$

As a final step, we can similarly argue that u_{2k+4} satisfies Theorem 2.1 since

$$\sin\psi_{2k+2} < \sin\psi_{2k+4} < \sin\psi < r\cos\gamma$$

with

$$\begin{cases} u_{2k+4} < u_{2k+2} & \text{if } r \in [a, c_{2k+2}), \\ u_{2k+4} > u_{2k+2} & \text{if } r \in (c_{2k+2}, 1]. \end{cases}$$

The cycle is now complete as (a), (b) and (c) are proved for the next increment of k. In addition, the discussion above provides the summary

$$\psi_{2k+2} < \psi_{2k+4} < \psi < \psi_{2k+3} < \psi_{2k+1} \quad \text{for } r \in (a,1). \tag{21}$$

It remains to verify that (a), (b) and (c) are true for the base case $k = 0$. However, these were shown as a result of Theorems 3.1 and 3.3, making (21) true for all $k \geq 0$:

$$\psi_2 < \psi_4 < \cdots < \psi < \cdots < \psi_3 < \psi_1 \quad \text{for } r \in (a,1). \tag{22}$$

For parts (2) and (3), apply Lemma 3.2 to each adjacent pair of angles in (22) to arrive at

$$u_1(a) < u_3(a) < \cdots < u(a) < \cdots < u_4(a) < u_2(a).$$

and

$$u_2(1) < u_4(1) < \cdots < u(1) < \cdots < u_3(1) < u_1(1). \qquad \square$$

Remark. Examining the proof of Theorem 3.4, we see that the interleaving properties will hold when instead of the restrictions on a and B we assume that the iterates are defined, continuous and positive, $\psi < \psi_1$, and u_2 satisfies (16).

4. Double intersection case

In contrast to the iterates of Siegel (which consistently intersect once with u_0 and result in interleaving properties throughout), here it is also possible to find selections of a, γ and B where u and u_2 intersect twice with u_0. In this case, there exist exactly two points b_{01} and b_{02} in $(a, 1)$ such that $u(b_{01}) = u(b_{02}) = u_0$ with

$$\begin{cases} u > u_0 & \text{if } r \in [a, b_{01}), \\ u < u_0 & \text{if } r \in (b_{01}, b_{02}), \\ u > u_0 & \text{if } r \in (b_{02}, 1]. \end{cases} \tag{23}$$

As well, there exist exactly two points c_{01} and c_{02} in $(a, 1)$ such that $u_2(c_{01}) = u_2(c_{02}) = u_0$ and

$$\begin{cases} u_2 > u_0 & \text{if } r \in [a, c_{01}), \\ u_2 < u_0 & \text{if } r \in (c_{01}, c_{02}), \\ u_2 > u_0 & \text{if } r \in (c_{02}, 1]. \end{cases} \tag{24}$$

Figure 3 demonstrates these configurations. The effect of (23) and (24) on subsequent iterates is far more varied and less understood. For the sake of brevity, we summarize the conditions necessary for (23) and (24) to occur.

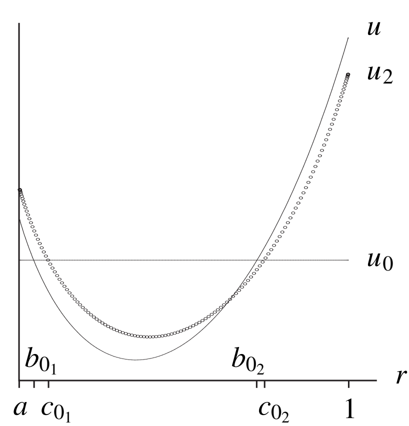


Figure 3. Configuration considered for double intersection case. Whereas u and u_2 must be arranged as shown with respect to u_0, the relative configuration of u and u_2 is not important.

Theorem 4.1. *For a given* $\gamma \in (\arcsin(1/5), \pi/2)$, *select* $a \geq 3/(5\sin\gamma + 2)$ *and*

$$B \leq \frac{\cos\gamma(1-a)^2}{6\pi}\Big(5a - \frac{3-2a}{\sin\gamma}\Big).$$

Under these conditions, there exist exactly two points $b_{01}, b_{02} \in (a, 1)$ *such that* $u(b_{01}) = u(b_{02}) = u_0$ *with*

$$\begin{cases} u > u_0 & \textit{if } r \in [a, b_{01}), \\ u < u_0 & \textit{if } r \in (b_{01}, b_{02}), \\ u > u_0 & \textit{if } r \in (b_{02}, 1]. \end{cases}$$

Theorem 4.2. *For a given* $\gamma \in (\arcsin(1/5), \pi/2)$, *select* $a \geq 3/(5\sin\gamma + 2)$ *and*

$$B \leq \frac{\cos\gamma(1-a)^2}{12\pi}\Big(5a - \frac{3-2a}{\sin\gamma}\Big).$$

Furthermore, assume $\sin\psi_2 < r\cos\gamma$ *on* $(a, 1)$. *Here, there exist exactly two points* $c_{01}, c_{02} \in (a, 1)$ *such that* $u_2(c_{01}) = u_2(c_{02}) = u_0$ *and*

$$\begin{cases} u_2 > u_0 & \textit{if } r \in [a, c_{01}), \\ u_2 < u_0 & \textit{if } r \in (c_{01}, c_{02}), \\ u_2 > u_0 & \textit{if } r \in (c_{02}, 1]. \end{cases}$$

For the most part, the proofs of Theorems 4.1 and 4.2 are analogous to their single intersection counterpart. Note, however, that under the hypotheses of Theorem 4.2, we are unable to prove that $\sin\psi_2 < r\cos\gamma$ on $(a, 1)$ and verify the existence of u_2. This assumption is consequently added to Theorem 4.2.

We briefly outline the resultant behaviors of the double intersection case. Here, it is assumed that all iterates are defined and continuous. Given (23), the behavior of u_1 with respect to u can be examined by considering the difference function

$$r\sin\psi - r\sin\psi_1 = B\int_a^r s(u - u_0)\,ds. \tag{25}$$

In addition to (25) being zero at $r = a$ and $r = 1$, there exist extrema at $r = b_{01}$ and $r = b_{02}$. Since $u(a) > u_0$, ψ and ψ_1 must be arranged as

$$\begin{cases} \psi > \psi_1 & \text{if } r \in (a, \xi_0), \\ \psi < \psi_1 & \text{if } r \in (\xi_0, 1), \end{cases} \tag{26}$$

where $\xi_0 \in (b_{01}, b_{02})$. When (26) is considered in conjunction with the volume condition, three configurations of u with u_1 are possible as illustrated in the first row of Figure 4. Using an analysis similar to the previous section's, we find for configurations A and C (where u and u_1 intersect once) that subsequent iterates will intersect only once with u. Specifically, for configuration A,

$$u_{2n+1}(a) < u(a) < u_{2n+2}(a) \quad \text{and} \quad u_{2n+2}(1) < u(1) < u_{2n+1}(1) \tag{27}$$

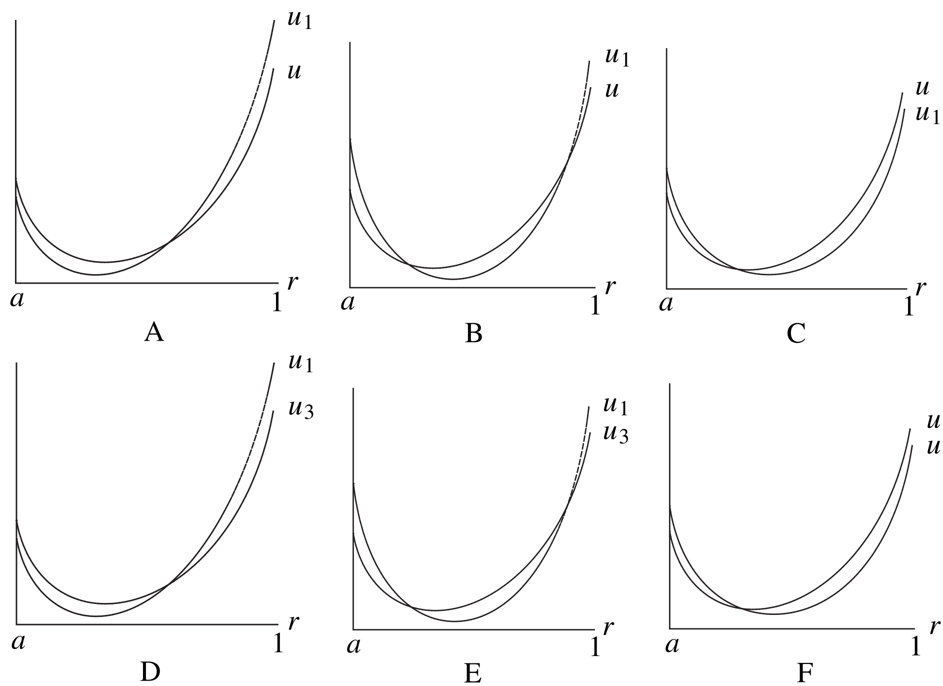


Figure 4. In the first row, potential configurations of u_1 with u_3, assuming (24) holds. In the second, potential configurations of u with u_1, assuming (23) holds.

or for configuration C,

$$u_{2n+2}(a) < u(a) < u_{2n+1}(a) \quad \text{and} \quad u_{2n+1}(1) < u(1) < u_{2n+2}(1)$$

with $n \geq 0$. In configuration B (where u and u_1 intersect twice) the behavior is potentially more diverse. Indeed, three arrangements identical to A, B and C are possible between u_2 and u, and a similar study can be applied here as was done for u versus u_1.

Additionally, (24) can be used to comment on the behavior of u_3 versus u_1. The difference function

$$r \sin \psi_1 - r \sin \psi_3 = B \int_a^r s(u_0 - u_2)\, ds$$

implies three arrangements of u_1 with u_3 are possible as in the second row of Figure 4. Configuration D leads to the predictable behavior

$$\begin{aligned} &\{u_{2n+2}(a)\} \text{ and } \{u_{2n+1}(1)\} \text{ are decreasing for } n \geq 0, \\ &\{u_{2n+1}(a)\} \text{ and } \{u_{2n+2}(1)\} \text{ are increasing for } n \geq 0. \end{aligned} \tag{28}$$

and likewise configuration F gives

$$\{u_{2n+1}(a)\} \text{ and } \{u_{2n+2}(1)\} \text{ are decreasing for } n \geq 0,$$
$$\{u_{2n+2}(a)\} \text{ and } \{u_{2n+1}(1)\} \text{ are increasing for } n \geq 0.$$

Configuration E will itself split into three possible arrangements between u_2 and u_4 that are identical to D, E and F. As one might expect, any configuration of the first row of Figure 4 could conceivably pair with any arrangement of the second row, leading to a far more complex behavior than in the single intersection case. Nevertheless, some pairings will again lead to interleaving iterates. This occurs, for example, when configuration A is paired with configuration D and properties (27) and (28) are matched. The same can be said of pairing configuration C with configuration F. However, if these couples were cross-matched (that is, A–F or C–D), the combined properties would result in diverging iterates. Further research is needed to fully understand the properties of $\{u_n\}$ over the complete parameter space.

References

[Concus and Finn 1974] P. Concus and R. Finn, "On capillary free surfaces in the absence of gravity", *Acta Math.* **132** (1974), 177–198. MR 58 #32327a Zbl 0382.76003

[Finn 1986] R. Finn, *Equilibrium capillary surfaces*, Grundlehren der Mathematischen Wissenschaften **284**, Springer, New York, 1986. MR 88f:49001 Zbl 0583.35002

[Gordon and Siegel 2010] J. Gordon and D. Siegel, "Properties of annular capillary surfaces with equal contact angles", *Pacific J. Math.* **247**:2 (2010), 353–370.

[Siegel 2006] D. Siegel, "Approximating symmetric capillary surfaces", *Pacific J. Math.* **224**:2 (2006), 355–365. MR 2007a:76022 Zbl 1116.76015

Received January 8, 2009.

JAMES GORDON
DEPARTMENT OF APPLIED MATHEMATICS
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO N2L 3G1
CANADA

james.gordon@utoronto.ca

DAVID SIEGEL
DEPARTMENT OF APPLIED MATHEMATICS
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO N2L 3G1
CANADA

james.gordon1832@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

HARMONIC QUASICONFORMAL SELF-MAPPINGS AND MÖBIUS TRANSFORMATIONS OF THE UNIT BALL

DAVID KALAJ AND MIODRAG S. MATELJEVIĆ

We prove in dimension $n > 2$ that a K-quasiconformal harmonic mapping u of the unit ball B^n onto itself is Euclidean bi-Lipschitz if $u(0) = 0$ and $K < 2^{n-1}$. This is an extension of a similar result of Tam and Wan for hyperbolic harmonic mappings with respect to a hyperbolic metric. The proof uses Möbius transformations on the related space and a recent result of the first author, which states that harmonic quasiconformal self-mappings of the unit ball are Lipschitz continuous.

1. Introduction

A twice differentiable function u defined in an open subset Ω of the Euclidean space $\mathbb{R}^n$ is said to be *harmonic* if it satisfies the differential equation

$$\Delta u(x) := D_{11}u(x) + D_{22}u(x) + \cdots + D_{nn}u(x) = 0.$$

Throughout the paper B^n denotes the unit ball in $\mathbb{R}^n$, and S^{n-1} denotes the unit sphere. Also we suppose that $n > 2$ (the case $n = 2$ has already been treated by many authors). Recall that for a vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ with the usual norm $|x| = (\sum_{i=1}^n x_i^2)^{1/2}$ and a matrix $A \in M_{n\times n}$, the matrix norm of A is defined as

$$|A| = \sup\{|Ax| : |x| = 1\}.$$

By $\langle \cdot, \cdot \rangle$ we denote the inner product in $\mathbb{R}^n$. Given $k \in \mathbb{N}$ and a normed space $\mathbb{X}$, the norm of a k-linear mapping from the k-fold Cartesian product of $\mathbb{R}^n$ to $\mathbb{X}$ is defined by

$$|P| = \sup\{|P(v_1, \dots, v_k)| : |v_1| = \cdots = |v_k| = 1\}.$$

For $K \geq q$, a homeomorphism $u : \Omega \to \Omega'$ between two open subsets Ω and Ω' of Euclidean $\mathbb{R}^n$ will be called a *K-quasiconformal* or shortly a quasiconformal mapping if the following two conditions are satisfied.

(i) The homeomorphism u is an absolutely continuous function in almost every segment parallel to some of the coordinate axes, and the partial derivatives

MSC2000: primary 30C65; secondary 31B05.
Keywords: quasiconformal map, harmonic mapping, Lipschitz condition, Möbius transformation, unit ball.

of u exist and are locally L^n integrable functions on Ω. For such a u, we write $u \in \mathrm{ACL}^n$.

(ii) For almost every x in Ω,

$$|Du(x)|^n/K \le J_u(x) \le K l(Du(x))^n,$$

where $l(Du(x)) := \inf\{|Du(x)\zeta| : |\zeta| = 1\}$ and $J_u(x)$ is the Jacobian determinant of u [Rešetnjak 1968].

For a continuous function u, the condition (i) is equivalent to the fact that u belongs to the Sobolev space $W^1_{n,\mathrm{loc}}(\Omega)$.

Let P denote the Poisson kernel, that is,

$$P(x, \eta) = \frac{1 - |x|^2}{|x - \eta|^n} \quad \text{for } x \in B^n \text{ and } \eta \in S^{n-1}.$$

Let $f : S^{n-1} \to \mathbb{R}^n$ be a bounded integrable function on the unit sphere S^{n-1}. The solution of the equation $\Delta u = 0$ in the unit ball $\mathbb{B}^n$ satisfying the boundary condition $u|_{S^{n-1}} = f \in L^1(S^{n-1})$ is given by

$$u(x) = P[f](x) = \int_{S^{n-1}} P(x, \eta) f(\eta) d\sigma(\eta) \quad \text{for } |x| < 1.$$

Here $d\sigma$ is the Lebesgue $(n-1)$-dimensional measure of the sphere S^{n-1} satisfying the condition $P[1](x) \equiv 1$. It is well known that if f is continuous in S^{n-1}, then the mapping $u = P[f]$ has a continuous extension $\tilde{u}$ to the boundary, and $\tilde{u} = f$ on S^{n-1}.

We will consider those harmonic mappings, namely, the solutions of the PDE $\Delta u = 0$, that are also quasiconformal.

Martio [1968] was the first to consider harmonic quasiconformal mappings on the complex plane. Recent papers [Kalaj 2004, 2008; Kalaj and Mateljević 2006; Kalaj and Pavlović 2005, 2009; Manojlović 2009; Pavlović 2002] shed much light on this topic.

Proposition 1.1 [Kalaj 2009]. *Let $u : B^n \to \Omega$ for $n \ge 3$ be a twice differentiable quasiconformal mapping of the unit ball onto the bounded domain Ω in $\mathbb{R}^n$ with a C^2 boundary satisfying the differential inequality*

$$|\Delta u| \le A|Du|^2 + B \quad \textit{for } A, B \ge 0.$$

Then Du (the first derivative of u) is bounded and u is Lipschitz continuous.

Because techniques of complex analysis are not available, the problem in the space $\mathbb{R}^n$ with $n \ge 3$ is much more complicated. For example, any harmonic mapping in a simply connected domain in the plane can be expressed as the sum of an analytic and an antianalytic function. The corresponding representation formula for harmonic mappings in the space is not true. On the other hand, Lewy's theorem and the theorem of Rado, Kneser and Choquet are essentially planar. According to the latter theorem, the harmonic extension (via Poisson integral) of a homeomorphism of the unit circle is always a diffeomorphism of the unit disk. However, in higher dimensions the situation is quite different: Martio [2009] and Melas [1993] constructed a homeomorphism of the unit sphere S^{n-1} for $n \geq 3$ whose harmonic extension fails to be diffeomorphic; see also [Laugesen 1996].

Let $K \in [1, 2^{n-1})$. Our main result, Theorem 3.1, states that the norm of the gradient of any K-quasiconformal harmonic self-mapping u of the unit ball with $u(0) = 0$ is bounded from below by a positive constant c_K that depends only on K. In contrast to the planar case, not all conformal mappings in the space are harmonic; only orthogonal transformations are, while other Möbius transformations are not, at least with respect to the Euclidean metric. However, Möbius transformations will play an important role in this paper. In this regard, Lemma 2.4 is of independent interest. In Section 4 we will give some nontrivial examples of quasiconformal self-mappings of the unit ball, and we will show that our result can be considered as a partial extension of Fefferman's theorem [1974] concerning biholomorphisms between smooth domains in the space.

2. Preliminaries and auxiliary results

Quasiconformal maps are locally well behaved with respect to distance distortion. If $f : \Omega \to \Omega'$ is a K-quasiconformal mapping between domains $\Omega, \Omega' \subset \mathbb{R}^n$, then f is locally Hölder continuous with the exponent $\alpha = K^{1/(1-n)}$, that is,

$$|f(x) - f(y)| \leq M|x - y|^\alpha, \tag{2-1}$$

whenever x and y lie in a fixed compact subset E of Ω; see [Vuorinen 1988, Theorem 10.11]. Here M is a constant depending only on K and E. Such an M can in general tend to infinity as the distance from E to the boundary of Ω tends to zero. However, if the boundary of Ω is regular enough, then an inequality similar to (2-1) holds uniformly in Ω [Gehring and Martio 1985; Koskela et al. 2001].

See also [Finn and Serrin 1958] for related results about the class of (K, K') planar quasiconformal mappings, which generalizes the class of standard quasiconformal mappings.

The following lemma is nothing but a slight reformulation of a corresponding lemma in [Tam and Wan 1998]. For the sake of completeness, we give its proof here and show that the constant is sharp.

Lemma 2.1. *If $u \in C^{1,1}$ is a K-quasiconformal mapping defined in a domain $\Omega \subset \mathbb{R}^n$ for $n \geq 3$, then $J_u(x) > 0$ for $x \in \Omega$ if $K < 2^{n-1}$. The constant 2^{n-1} is sharp.*

Proof. Assume that $J_u(a) = 0$ for some $a \in \Omega$. This yields $Du(a) = 0$. Without loss of generality we can assume that $a = 0$ and $u(0) = 0$. Choose $r > 0$ with $r < \operatorname{dist}(0, \partial\Omega)$, and let $E = \overline{B^n(0, r)} := \{x \in \mathbb{R}^n : |x| \le r\}$. Applying (2-1) to the mapping $f = u^{-1}$ defined in $\Omega' = u(\Omega)$, we obtain

$$|f(y)| \le M_E |y|^{K^{1/(1-n)}} \quad \text{for } y \in u(E),$$

where M_E is a constant depending on E. This implies

$$M_E^{-K^{1/(n-1)}} |x|^{K^{1/(n-1)}} \le |u(x)| \quad \text{for } x \in E. \tag{2-2}$$

Now since u is twice differentiable, with $Du(0) = 0$ and $u(0) = 0$, it follows from Taylor's formula that there exists a positive constant N such that

$$|u(x)| \le N|x|^2 \quad \text{for } x \in E. \tag{2-3}$$

Combining (2-2) and (2-3), we have

$$M_E^{-K^{1/(n-1)}}/N \le |x|^{2-K^{1/(n-1)}} \quad \text{for } x \in E.$$

This is only possible if $2 - K^{1/(n-1)} \le 0$.

Thus $K \ge 2^{n-1}$, which is a contradiction.

To prove sharpness, consider the mapping $u(x) = |x|^\alpha x$ with $\alpha \ge 1$. Then

$$J_u(x) = (1+\alpha)|x|^{n\alpha} \tag{2-4}$$

and

$$|Du(x)| = (\alpha+1)|x|^\alpha. \tag{2-5}$$

By (2-4) and (2-5) it follows that

$$\frac{|Du(x)|^n}{J_u(x)} = (\alpha+1)^{n-1}.$$

Therefore, u is a twice differentiable $(1+\alpha)^{n-1}$-quasiconformal self-mapping of the unit ball with $J_u(0) = 0$, meaning that the constant 2^{n-1} is the best possible. □

Lemma 2.2. *Let u be a harmonic mapping of the unit ball onto itself such that $u(0) = 0$. Then there exists a positive constant C_n such that*

$$\frac{1-|x|^2}{1-|u(x)|^2} \le C_n \quad \textit{for } x \in B^n. \tag{2-6}$$

Proof. Let S^+ denote the northern hemisphere and let S^- be the southern hemisphere. Let $U = P[\chi_{S^+}] - P[\chi_{S^-}]$ be the Poisson integral of the function χ_S that equals 1 on S^+ and -1 on S^-. Then by the Schwarz lemma [Axler et al. 1992],

$$\langle u(x), u(x_0)/|u(x_0)|\rangle \le \big|U(|x|N)\big|$$

for a fixed x_0, where N is the north pole.

It follows that $|u(x_0)|^2 \le \big|U(|x_0|N)\big|^2$. Thus

$$\frac{1-|x|^2}{1-|u(x)|^2} \le \frac{1-|x|^2}{1-U(|x|N)^2} =: g(r) \quad \text{for } r=|x|.$$

We will need Hopf's boundary point lemma:

Lemma 2.3 [Hopf 1952; Protter and Weinberger 1967]. *Let v satisfy $\Delta v \ge 0$ in an open set $D \subset \mathbb{R}^n$ and suppose $v \le M$ in D and $v(P) = M$ for some $P \in \partial D$. Assume that P lies on the boundary of a ball*

$$B^n(a,r) := \{x : |x-a| < r\} \subset D.$$

If v is continuous on $D \cup P$ and if the outward directional derivative $\partial v/\partial n$ exists at P, then $v \equiv M$ or

$$\partial v(P)/\partial n > 0.$$

Applying this lemma to the function $U(x)$ and taking $h(r) = U(rN)$, we obtain

$$h'(1) = \frac{\partial U(N)}{\partial n} > 0.$$

Thus

$$C_n := \sup_{|x|\le 1}\left\{\frac{1-|x|^2}{1-U(|x|N)^2}\right\} < \infty,$$

and the proof of Lemma 2.2 is complete. □

Following the book of Ahlfors [1981], for $a, x \in B^n$ we define

$$[x,a]^2 = 1 + |x|^2|a|^2 - 2\langle x, a\rangle$$

and the inversion x^* of $x \ne 0$ by $x^* = x/|x|^2$. Since

$$\begin{aligned}[x,a]^2 &= |x|^2|x^*-a|^2 = |a|^2|x-a^*|^2 \\ &= \big||x|a - |x|x^*\big|^2 = \big||a|x-|a|a^*\big|^2 = \big||x|a-|x|x^*\big| \cdot \big||a|x-|a|a^*\big|,\end{aligned}$$

we have

$$[x,a]^2 \ge (1-|a|)(1-|x|) \quad \text{and} \quad [x,a]^2 \ge (1-|x|)^2. \tag{2-7}$$

Assume that p is a conformal mapping of the unit ball onto itself. Then it is well known that p is a Möbius transformation of the unit ball onto itself. Under the normalization $p(0) = -a \ne 0$ (or $p(a) = 0$), the mapping p is given (up to some orthogonal transformation of the unit ball) by

$$p(x) = \frac{(1-|a|^2)(x-a) - |x-a|^2 a}{[x,a]^2}. \tag{2-8}$$

Lemma 2.4. *If p is a Möbius transformation of the unit ball onto itself, with $p(0)=a$, then for all $k,l\in\mathbb{N}_0$ with $k>l$, there exists a constant $C_{k,l}$ such that*

$$C_{k,l}\ge k\cdot(k-1)\cdots(l+1),$$

$$\frac{k!\,|a|^{k-1}(1-|a|^2)}{[x,a]^{k+1}}\le|p^{(k)}(x)|\le\frac{C_{k,0}|a|^{k-1}(1-|a|^2)}{[x,a]^{k+1}}\quad \textit{for } x\in B^n, \tag{2-9}$$

$$\frac{\left|p^{(k)}(x)\right|}{\left|p^{(l)}(x)\right|}\le C_{k,l}\frac{1}{(1-|x|)^{k-l}}\quad \textit{for } x\in B^n \quad \textit{and}\quad l>0, \tag{2-10}$$

$$\left|p^{(k)}(x)\right|\le\frac{C_{k,0}}{(1-|p(0)|^2)^{(k-1)/2}}\left(\frac{1-|p(x)|^2}{1-|x|^2}\right)^{(k+1)/2}\textit{for } x\in B^n. \tag{2-11}$$

Proof. It follows from (2-8) that

$$p'(x)=\frac{1-|a|^2}{[x,a]^2}\Delta(x,a),$$

where $\Delta(x,a)=(I-2Q(a))(I-2Q(x-a^*))$, and $Q(y)$ is the matrix whose elements have the form $Q(y)_{i,j}=y_iy_j/|y|^2$. For every $y\in B^n$, we have $K(y):=I-2Q(y)\in O_n$, where O_n is the set of all orthogonal matrices. Thus $\Delta(x,a)$ is an orthogonal matrix as well, and consequently $|\Delta(x,a)|=1$. This means that

$$|p'(x)|=\frac{1-|a|^2}{[x,a]^2}.$$

According to (2-7),

$$|p'(x)|\le\frac{2}{1-|x|}. \tag{2-12}$$

If we put

$$A=\frac{1-|a|^2}{[x,a]^2}\quad\text{and}\quad B=(I-2Q(a))(I-2Q(x-a^*)),$$

then we have $p'=AB$. Therefore, for the $(k+1)$-st derivative of p, we have

$$p^{(k+1)}(x)=\sum_{j=0}^{k}\binom{k}{j}A^{(j)}B^{(k-j)}, \tag{2-13}$$

and $p^{(k+1)}$ can be treated as a k-linear form between the k-fold product of $\mathbb{R}^n$ and and $M_{n\times n}$. We will use the notation

$$Q(y)=\frac{y\otimes y}{|y|^2}, \tag{2-14}$$

where $\otimes$ denotes the tensor product of vectors.

Let us prove that for $k \in \mathbb{N}_0$ there exists a $(2k+2)$-linear form from the $(k+2)$-fold product of $\mathbb{R}^n$ to $M_{n\times n}$ such that

$$Q^{(k)}(y)(h_1, h_2, \dots h_k) = \frac{1}{|y|^{2k+2}} P^k(y, \dots, y, h_1, \dots, h_k). \tag{2-15}$$

We proceed by induction on k. It is evident from (2-14) that (2-15) is true for $k=0$.

Assume that (2-15) is true for some k, and prove it for $k+1$. By (2-15), it follows that

$$\begin{aligned} Q^{(k+1)}&(y)(h_1, h_2, \dots h_k, h_{k+1}) \\ &= \frac{1}{|y|^{2k+2}} \sum_{j=1}^{k+2} P^k(y, \dots, y, \overset{j\downarrow}{h_{k+1}}, y, \dots, y, h_1, \dots, h_k) \\ &\qquad - (k+2) \frac{\langle y, h_{k+1}\rangle}{|y|^{2k+4}} P^k(y, \dots, y, h_1, \dots, h_k), \end{aligned} \tag{2-16}$$

where the j pointing to h_{k+1} denotes that h_{k+1} is in the j-th position. Thus

$$Q^{(k+1)}(y)(h_1, h_2, \dots, h_k, h_{k+1}) = \frac{P^{k+1}(y, \dots, y, h_1, \dots, h_k, h_{k+1})}{|y|^{2(k+1)+2}}, \tag{2-17}$$

where

$$\begin{aligned} P^{k+1}&(e_1, \dots, e_{k+3}, f_1, \dots f_{k+1}) \\ &= \sum_{j=1}^{k+2} \langle e_{k+3}, e_j\rangle P^k(e_1, \dots, \overset{j\downarrow}{f_{k+1}}, \dots, e_{k+2}, f_1, \dots, f_k) \\ &\qquad - 2(k+1)\langle e_{k+3}, f_{k+2}\rangle P^k(e_1, \dots, e_{k+2}, f_1, \dots, f_k). \end{aligned}$$

We first prove the left side of inequality (2-9).

From (2-17) and the induction hypothesis, for $y \neq 0$ we obtain

$$Q^{(k+1)}(y)\left(\frac{y}{|y|}, \dots, \frac{y}{|y|}\right) = 0. \tag{2-18}$$

Thus

$$B^{(k)}(x)\left(\frac{x-a^*}{|x-a^*|}, \dots, \frac{x-a^*}{|x-a^*|}\right) = 0. \tag{2-19}$$

Let us prove the left side of (2-9). First, according to (2-13) and (2-19), we obtain

$$\begin{aligned} |p^{(k+1)}(x)| &= \sup_{|h_1|=\dots=|h_k|=1} |p^{(k+1)}(x)(h_1, \dots, h_k)| \\ &\geq \left| A^{(k)}\left(\frac{x-a^*}{|x-a^*|}, \dots, \frac{x-a^*}{|x-a^*|}\right)\right|. \end{aligned} \tag{2-20}$$

Since P^k is a $(2k+2)$-linear form,

$$|P^k(y,\dots,y,h_1,\dots,h_k)| \le |P^k||y|^{k+2}\prod_{j=1}^{k}|h_j|.$$

Thus $|Q^k(y)| \le |P^k|/|y|^k$, whence we have

$$|Q^k(x-a^*)| \le \frac{|P^k|}{|x-a^*|^k} = \frac{|a|^k|P^k|}{[x,a]^k}.$$

Further, observe that $B(x) = K(a)(I - 2Q(x-a^*))$ for $K(a) \in O_n$. Hence

$$B^{(k)}(x) = -2K(a)Q^{(k)}(x-a^*),$$

and using the identity

$$\frac{1-|p(x)|^2}{1-|x|^2} = \frac{1-|a|^2}{[x,a]^2} = \frac{1-|a|^2}{|a|^2|x-a^*|^2},$$

we obtain

$$|B^k(x)| \le \frac{2|a|^k|P^k|}{[x,a]^k} < \frac{2|a|^k|P^k|(1-|p|^2)^{k/2}}{(1-|a|^2)^{k/2}(1-|x|^2)^{k/2}}. \tag{2-21}$$

To estimate the derivatives of $A(x) = (1-|a|^2)/[x,a]^2$, define

$$H(y) = \frac{1}{|y|^2} = \frac{|a|^2}{1-|a|^2}A(x) \quad \text{for } y = x-a^*.$$

Then

$$H'(y)h_1 = -2\frac{\langle y, h_1\rangle}{|y|^4}.$$

Similarly, it can be proved that for every $k \ge 1$ there exists an $\mathbb{R}$-valued $(2k+2)$-linear form G^{k+1} such that

$$H^{(k+1)}(y)(h_1,h_2,\dots h_{k+1}) = \frac{1}{|y|^{2k+4}}G^{k+1}(y,\dots,y,h_1,\dots,h_{k+1}),$$

where

$$\begin{aligned} G^{k+1}&(e_1,\dots,e_{k+1},f_1,\dots f_{k+1}) \\ &= \sum_{j=1}^{k}\langle e_{k+1},e_j\rangle G^k(e_1,\dots,\overset{j\downarrow}{f_{k+1}},\dots,e_k,f_1,\dots,f_k) \\ &\qquad - 2(k+1)\langle e_{k+1},f_{k+1}\rangle G^k(e_1,\dots,e_k,f_1,\dots,f_k). \end{aligned}$$

Therefore

$$|H^k(y)| \le |G^k|/|y|^{k+2}. \tag{2-22}$$

On the other hand, by the identity $k\cdot(k+1)!-2(k+1)(k+1)!=-(k+2)!$ and the induction hypothesis, we obtain

$$G^k\left(y,\dots,y,\frac{y}{|y|},\dots,\frac{y}{|y|}\right)=\pm\frac{(k+1)!}{|y|^{k+2}}\quad\text{for } k\in\mathbb{N}.$$

Thus

$$\frac{k!}{|y|^{k+2}}=\left|H^k(y)\left(\frac{y}{|y|},\dots,\frac{y}{|y|}\right)\right|\le\left|H^k(y)\right|.$$

In view of the fact

$$\frac{1-|p(x)|^2}{1-|x|^2}=\frac{1-|a|^2}{[x,a]^2},$$

we have

$$\frac{k!|a|^k(1-|a|^2)}{[x,a]^{k+2}}\le\left|A^k(x)\right|\le\frac{2|G^k||a|^k(1-|p|^2)^{1+k/2}}{(1-|a|^2)^{(k-1)/2}(1-|x|^2)^{1+k/2}}.\tag{2-23}$$

Now the left side of (2-9) follows from (2-18), (2-13), (2-20) and (2-23).

Combining (2-13), (2-21), (2-22) and (2-7) for $k\ge 1$, we obtain

$$|p^{(k)}(x)|\le C_{k,0}\frac{|a|^k(1-|a|^2)}{[x,a]^{k+1}}<2C_{k,0}\frac{1}{(1-|x|)^k},$$

where

$$C_{k,0}=|G^{k-1}|+2\sum_{j=1}^{k-1}\binom{k-1}{j}|P^j||G^{(k-1-j)}|.$$

This proves (2-9). Inequalities (2-10) and (2-11) immediately follow from (2-9) and (2-23), and the proof is complete. □

Remark 2.5. For fixed $k,l\in\mathbb{N}_0$, we denote by $C^*_{k,l}$ the infimum taken over all constants $C_{k,l}$ satisfying the previous lemma. In the complex plane $\mathbb{C}$ there holds $C^*_{k,l}=k\cdot(k-1)\cdots(l+1)$, and therefore (2-9) reduces to the equality. According to (2-12), this occurs in the higher dimensions as well for $k=1$. We believe that $C^*_{k,l}=k\cdot(k-1)\cdots(l+1)$ for arbitrary n, k and l.

3. The main result

Theorem 3.1. *Let $K<2^{n-1}$ and let u be a K-quasiconformal harmonic mapping of the unit ball onto itself satisfying the normalization $u(0)=0$. Then there exists a positive constant $c_K>0$ such that*

$$|Du(x)|\ge c_K\quad\textit{for each } x\in B^n.$$

The first step of the proof is similar to that in [Tam and Wan 1998]. However here the problem is more complicated, because Möbius transformations are harmonic with respect to the hyperbolic metric, but not with respect to the Euclidean metric.

Proof. We will prove by contradiction that the function $|Du|$ is uniformly bounded below away from 0. Suppose that there exists a sequence $\{x_i\}$ in B^n such that $Du(x_i) \to 0$ as $i \to \infty$. We will use Proposition 1.1 together with the following lemma.

Lemma 3.2. *Let u be a harmonic Lipschitz mapping of the unit ball onto itself. Let $\{x_i\}$ be a sequence in B^n. For arbitrary $i \in \mathbb{N}$, let p_i and q_i be two Möbius transformations of B^n such that $q_i(0) = x_i$ and $p_i(u(x_i)) = 0$. Take $u_i = p_i \circ u \circ q_i$. Then*

$$|D^{(k)} u_i(x)| \le c_n^k \frac{1}{(1-|x|^2)^k} \quad \text{for } k \in \mathbb{N},$$

where c_n^k is independent of x and i.

Proof. To simplify calculations in this proof, sometimes we will omit the arguments of functions.

Since

$$|p_i'(u)| = \frac{1-|p_i(u)|^2}{1-|u|^2} \tag{3-1}$$

and

$$|q_i'(x)| = \frac{1-|q_i(x)|^2}{1-|x|^2}, \tag{3-2}$$

according to (2-6) it follows that

$$\begin{aligned} |Du_i| &\le |p_i'||Du||q_i'| \\ &\le \frac{1-|p_i(u(q_i(x)))|^2}{1-|u(q_i(x))|^2} \frac{1-|q_i(x)|^2}{1-|x|^2} |Du| \\ &\le C_n |Du|_\infty \frac{1-|p_i(u(q_i(x)))|^2}{1-|x|^2}. \end{aligned}$$

Thus

$$|Du_i| \le C_n |Du|_\infty \frac{1}{1-|x|^2}.$$

For $m \in \mathbb{N}$, we make use of the Cauchy estimates [Axler et al. 1992, pp. 33–35]:

$$|D^m(u)(q_i(x))| \le A_m \frac{|Du|_\infty}{(1-|q_i(x)|)^{m-1}}. \tag{3-3}$$

To estimate the norm of $D^k u_i$ for $k > 1$, we use induction. Obviously, it is complicated to compute $D^k u_i$ for large k. However, it is clear that it can be written as a sum of products:

$$D^k u_i = \sum \Big(p_i^{(\tau)} \prod D^{j_t} u q_i^{(s_{t1})} \cdot \dots \cdot q_i^{(s_{tl_t})} \Big), \tag{3-4}$$

where $\prod$ and $\sum$ denote the corresponding finite product and sum of linear operators. The indices t and τ range at most from 1 to k; the indices j_t, $s_{t1}, \dots, s_{tl_t}$ satisfy similar bounds.

Because of (3-3) we have

$$|p_i^{(\tau)}| \prod |D^{j_t} u| |q_i^{(s_{t1})}| \cdots |q_i^{(s_{tl_t})}| \le \text{const} |p_i^{(\tau)}| \prod \frac{|Du|_\infty}{(1-|q_i(x)|)^{j_t - 1}} |q_i^{(s_{t1})}| \cdots |q_i^{(s_{tl_t})}|. \tag{3-5}$$

Therefore, it is enough to prove that

$$|p_i^{(\tau)}| \prod \frac{|Du|_\infty}{(1-|q_i(x)|)^{j_t - 1}} |q_i^{(s_{t1})}| \dots |q_i^{(s_{tl_t})}| \le \text{const} \frac{1}{(1-|x|)^k}. \tag{3-6}$$

For $k = 1$, inequality (3-6) is satisfied. Assume that (3-6) is true for some k, and therefore (3-5) is true as well. We will prove (3-6) for $k + 1$.

Since $D^{k+1} u_i = D D^k u_i$, the first factor in the corresponding formula (3-4) for $D^{k+1} u_i$, instead of $p_i^{(\tau)} D^{j_t} u q_i^{(s_{t1})} \cdots q_i^{(s_{tl_t})}$, contains the term

$$\big(p_i^{(\tau+1)} D u q_i' D^{j_t} u q_i^{(s_{t1})} \cdots q_i^{(s_{tl_t})} + p_i^{(\tau)} D^{j_t+1} u q_i' \cdot q_i^{(s_{t1})} \cdots q_i^{(s_{tl_t})} + p_i^{(\tau)} D^{j_t} u q_i^{(s_{t1}+1)} \cdots q_i^{(s_{tl_t})} + \dots + p_i^{(\tau)} D^{j_t} u q_i^{(s_{t1})} \dots q_i^{(s_{tl_t}+1)} \big),$$

and consequently the corresponding formulas (3-6) and (3-5), instead of

$$|p_i^{(\tau)}| \frac{|Du|_\infty}{(1-|q_i(x)|)^{j_t - 1}} |q_i^{(s_{t1})}| \dots |q_i^{(s_{tl_t})}|,$$

contain

$$\Bigg(|p_i^{(\tau+1)}| |Du| |q_i'| |D^{j_t} u| |q_i^{(s_{t1})}| \cdots |q_i^{(s_{tl_t})}| + |p_i^{(\tau)}| \frac{|Du|_\infty}{(1-|q_i(x)|)^{j_t}} |q_i'| \cdot |q_i^{(s_{t1})}| \cdots |q_i^{(s_{tl_t})}| + |p_i^{(\tau)}| \frac{|Du|_\infty}{(1-|q_i(x)|)^{j_t - 1}} |q_i^{(s_{t1+1})}| \dots |q_i^{(s_{tl_t})}| + \cdots + |p_i^{(\tau)}| \frac{|Du|_\infty}{(1-|q_i(x)|)^{j_t - 1}} |q_i^{(s_{t1})}| \cdots |q_i^{(s_{tl_t}+1)}| \Bigg).$$

The other factors in (3-4) can be treated similarly.

Applying (3-2), we get

$$\frac{|Du|_\infty}{(1-|q_i(x)|)^{j_t}}|q_i'| = \frac{|Du|_\infty}{(1-|q_i(x)|)^{j_t}}\frac{1-|q_i(x)|^2}{1-|x|^2} \le \frac{|Du|_\infty}{(1-|q_i(x)|)^{j_t-1}}\frac{2}{1-|x|}. \tag{3-7}$$

Next, by applying (2-6) and (2-10), we obtain

$$\left|p_i^{(\tau+1)}Duq_i'\right| \le \frac{\mathrm{const}|p_i^{(\tau)}|}{1-|u(q_i(x))|}\frac{1-|q_i(x)|}{1-|x|} \le \mathrm{const}\frac{|p_i^{(\tau)}|}{1-|x|}. \tag{3-8}$$

On the other hand, according to (2-10) we have

$$|q_i^{(j+1)}| \le \mathrm{const}\frac{|q_i^{(j)}|}{1-|x|}. \tag{3-9}$$

By induction, (3-6) is true for k. The last fact and the estimates (3-7), (3-8) and (3-9) imply that (3-6) is also true for $k+1$. Consequently,

$$|D^{(k)}u_i(x)| \le c_n^k\frac{1}{(1-|x|^2)^k} \quad \text{for } k \in \mathbb{N}. \tag{3-10}$$ □

We are now ready to finish the proof of Theorem 3.1. According to the notations of the previous lemma, $u_i = p_i \circ u \circ q_i$ is a C^∞ K-quasiconformal mapping of the unit ball onto itself, satisfying the condition $u_i(0) = 0$. By (2-6) we have

$$|Du_i(0)| = \frac{1-|x_i|^2}{1-|u(x_i)|^2}|Du(x_i)| \to 0 \quad \text{as } i \to \infty. \tag{3-11}$$

For example, by [Fehlmann and Vuorinen 1988], a subsequence of u_i, also denoted u_i, converges uniformly to a K-quasiconformal map u_0 on the closed unit ball B^n. According to Lemma 3.2 together with Proposition 1.1, u_0 is in $C^\infty(B^n; B^n)$ with $u_0(0) = 0$. The relation (3-11) implies $D(u_0)(0) = 0$. This obviously contradicts the statement of Lemma 2.1. Hence, the proof of Theorem 3.1 is complete. □

Remark 3.3. Let us estimate $Du_i = p_i' Duq_i'$ more precisely. From

$$q_i(x) = \frac{(1-|x_i|^2)(x+x_i)+|x+x_i|^2x_i}{[x,-x_i]^2},$$

according to (3-1), it follows that

$$|q_i'(x)| = \frac{1-|x_i|^2}{|x_i|^2|x+x_i^*|^2}$$

and

$$1-|q_i(x)|^2 = |q_i'(x)|(1-|x|^2).$$

Similarly, since

$$p_i(u) = \frac{(1-|u(x_i)|^2)(u-u(x_i)) + |u-u(x_i)|^2 u(x_i)}{[u, u(x_i)]^2},$$

it follows that

$$|p_i'(u)| = \frac{1-|u(x_i)|^2}{|u(x_i)|^2 |u(q_i(x)) - u(x_i)^*|^2}.$$

Thus

$$|Du_i| \le \frac{1-|u(x_i)|^2}{|u(x_i)|^2 |u(q_i(x)) - u(x_i)^*|^2} |Du|_\infty \frac{1-|x_i|^2}{|x_i|^2 |x + x_i^*|^2}.$$

From

$$|x_i|^2 |x + x_i^*|^2 \ge (1-|x_i|)^2,$$

$$1 - |u(x_i)| \le |Du|_\infty (1-|x_i|),$$

$$|u(x_i)|^2 |u(q_i(x)) - u(x_i)^*|^2 \ge (1-|u(x_i)|)^2,$$

and (2-6) we obtain

$$|Du_i| \le 4 \min\left\{ \frac{|Du|_\infty^2}{[u(q_i(x)), u(x_i)]^2}, \frac{C_n |Du|_\infty}{[x, -x_i]^2} \right\}. \tag{3-12}$$

Assume that $t = \lim_{i\to\infty} x_i \in S^{n-1}$. It follows from (3-12) that

$$|Du_0| = \lim_{i\to\infty} |Du_i|$$

is uniformly bounded in $B^n \setminus B^n(-t, \varepsilon)$ for $\varepsilon > 0$. Is $|Du_0|$ bounded uniformly in B^n?

Theorem 3.4. *Let $K < 2^{n-1}$ and assume that u is a K-quasiconformal harmonic mapping of the unit ball onto B^2 itself satisfying the normalization $u(0) = 0$. Then u is a bi-Lipschitz mapping.*

Proof. According to Proposition 1.1 and Theorem 3.1, there exists a constant $c \ge 1$ such that

$$c^{-1} \le |Du(x)| \le c \quad \text{for } x \in B^n. \tag{3-13}$$

By using (3-13) and the fact that u is quasiconformal, we obtain

$$|D(u^{-1})(u(x))| = \frac{1}{\inf_h |Du(x)h|} = \frac{1}{l(Du(x))} \le \frac{K^{2/n}}{Du(x)} \le cK^{2/n}.$$

Therefore

$$|D(u^{-1})| \le c_1. \tag{3-14}$$

From (3-13) and (3-14) it follows that u is bi-Lipschitz. □

4. Examples of quasiconformal harmonic mappings

We now give examples of nontrivial harmonic quasiconformal self-mappings of the unit ball.

4.1. ***Holomorphic self-mappings of the unit ball.*** Let $B^{2n} \subset \mathbb{C}^n$ and let f be a holomorphic automorphism of the unit ball B^{2n} onto itself. Then f is a quasiconformal harmonic mapping. To prove this fact, observe that $\bar\partial f = 0$ implies $\partial\bar\partial f = 0$. Also f has a holomorphic extension up to the boundary. This means that it is bi-Lipschitz. Therefore f is a quasiconformal harmonic mapping. It is interesting to note that the composition of a harmonic and holomorphic mapping is itself harmonic, because $\partial f \circ h = f_h \partial h + f_{\bar h}\partial \bar h$, and therefore

$$\bar\partial\partial f \circ h = f_{hh}\bar\partial h \partial h + f_{\bar h h}\bar\partial \bar h \partial h + f_{h\bar h}\bar\partial h \partial \bar h + f_{\bar h \bar h}\partial \bar h \bar\partial h.$$

According to Fefferman's theorem [1974], every biholomorphism between two smooth domains has a C^∞ extension to the boundary, which means that these mappings are bi-Lipschitz. Hence, the class of biholomorphic mappings between smooth domains is contained in the class of harmonic quasiconformal mappings. Thus our results can be considered as partial extensions of Fefferman's theorem.

4.2. ***Perturbation of the identity.*** Let us show that small smooth perturbations of the boundary value of a holomorphic automorphism $\phi \in C^2(\overline{B^{2n}})$ of the unit ball onto itself induce harmonic quasiconformal mappings.

Since the composition of a harmonic mapping with a holomorphic automorphism is itself a harmonic mapping, it is enough to perturb the identity map, and after that take the corresponding composition.

Define $I_\delta(x) = x + \delta(x)$, where $x \in B^n$ and $\delta(x) \in B^n$, and take $\phi_\delta = I_\delta/|J_\delta|$, where $|J_\delta|^2 = 1 + 2\langle x, \delta(x)\rangle + |\delta(x)|^2$. Thus

$$|I_\delta(x)| < 1 \quad \text{for } x \in B^n \quad \text{and} \quad |I_\delta(x)| = 1 \quad \text{for } x \in S^{n-1}.$$

We also have

$$|J_\delta(x)|^2 \ge (1 - |\delta(x)|)^2 \quad \text{for } x \in B^n. \tag{4-1}$$

Here $\delta(x)$ is a twice differentiable mapping satisfying

$$|\delta'(x)| < 1 \quad \text{for } x \in S^{n-1}.$$

This condition guarantees the injectivity of $\phi_\delta(x)$ in S^{n-1}. To continue, we use the following result of Gilbarg and Hörmander [1980, Theorem 6.1 and Lemma 2.1].

Proposition 4.1. *The Dirichlet problem $\Delta u = f$ in Ω for $u = u_0$ on $\partial\Omega \in C^1$ has a unique solution $u \in C^{1,\alpha}$ for every $f \in C^{0,\alpha}$ and $u_0 \in C^{1,\alpha}$, and we have*

$$\|u\|_{1,\alpha} \le C(\|u_0\|_{1,\alpha,\partial\Omega} + \|f\|_{0,\alpha}), \tag{4-2}$$

where C is a constant.

To guarantee the injectivity of the harmonic extension $\Phi_\delta(x) = P[\phi_\delta](x)$ of $\phi_\delta(x)$ in the unit ball, we estimate $|D\phi_\delta(x) - \mathrm{Id}|$ and $|D^2\phi_\delta(x)|$, and use (4-2) to conclude that

$$|DP[\phi_\delta](x) - P[\mathrm{Id}]| \le C\big(|D\phi_\delta(x) - \mathrm{Id}| + |D^2\phi_\delta(x)|\big). \tag{4-3}$$

First,

$$D\phi_\delta(x)h = \frac{h+\delta'(x)h}{|J_\delta|} - \frac{I_\delta(x)}{|J_\delta(x)|^3}\big(\langle\delta(x), h\rangle + \langle\delta'(x)h, x+\delta(x)\rangle\big).$$

Therefore, using (4-1) we infer that

$$|D\phi_\delta(x)h - h| \le \left|\frac{2|\delta| + 2|\delta'| + |\delta||\delta'|}{1-|\delta(x)|^2}h\right|,$$

that is,

$$|D\phi_\delta(x) - \mathrm{Id}| \le \frac{2|\delta| + 2|\delta'| + |\delta||\delta'|}{1-|\delta(x)|^2}.$$

Next we find

$$\begin{aligned}
&D^2\phi_\delta(x)(h,k)\\
&= \frac{\delta''(x)(h,k)}{|J_\delta|}\\
&\quad - \frac{I_\delta(x)}{|J_\delta(x)|^3}\big(\langle\delta'(x)k, h\rangle + \langle\delta''(x)(h,k), x+\delta(x)\rangle + \langle\delta'(x)(h), k+\delta'(x)k\rangle\big)\\
&\quad - \frac{k+\delta'(x)k}{|J_\delta(x)|^3}\big(\langle\delta(x), h\rangle + \langle\delta'(x)h, x+\delta(x)\rangle\big)\\
&\quad + 3\frac{I_\delta(x)}{|J_\delta(x)|^5}\big(\langle\delta(x), h\rangle + \langle\delta'(x)h, x+\delta(x)\rangle\big)\big(\langle\delta(x), k\rangle + \langle\delta'(x)k, x+\delta(x)\rangle\big).
\end{aligned}$$

Thus

$$\begin{aligned}
&|D^2\phi_\delta(x)|\\
&\qquad \le \frac{3(|\delta|+|\delta'|+|\delta||\delta'|)^2+(1+|\delta'|)(|\delta|+|\delta'|+|\delta||\delta'|)+2|\delta'|^2+|\delta'|+|\delta''|(2+|\delta|)}{(1-|\delta(x)|^2)^2}.
\end{aligned}$$

Choosing δ such that

$$|D^2\phi_\delta(x)| < \frac{1}{2C} \quad \text{and} \quad |D\phi_\delta(x) - \mathrm{Id}| < \frac{1}{2C},$$

according to (4-3) we obtain

$$|D\Phi_\delta(x) - \mathrm{Id}| < \frac{C}{2C} + \frac{C}{2C} = 1 \quad \text{for } x \in B^n.$$

Thus $|\Phi_\delta(x)-\Phi_\delta(y)+y-x|<|x-y|$, and therefore $0<|\Phi_\delta(x)-\Phi_\delta(y)|$. This implies that Φ_δ is injective. Hence, Φ_δ is a quasiconformal harmonic diffeomorphism of the unit ball onto itself.

In particular, let $I_\varepsilon(x)=(x_1+\varepsilon, x_2, x_3)$ and take $j_\varepsilon=(1+2\varepsilon x_1+\varepsilon^2)^{1/2}$. Define

$$\phi_\varepsilon(x)=I_\varepsilon(x)/j_\varepsilon.$$

Now take $\Phi_\varepsilon=P[\phi_\varepsilon]$. Then for sufficiently small ε, the extension Φ_ε is a diffeomorphism of the unit ball onto itself having a diffeomorphic extension to the boundary. This means that Φ_ε is quasiconformal harmonic mapping.

Direct calculations yield

$$\begin{aligned}\frac{1}{C}\big|D\Phi_\varepsilon(x)-\mathrm{Id}\big| &= \frac{1}{C}\big|P[\phi_\varepsilon-\mathrm{Id}](x)\big| \\ &\le \sup_{|x|=1}\big(|D\phi_\varepsilon(x)-\mathrm{Id}|+|D^2\phi_\varepsilon(x)|\big) \\ &< \sup_{|x|=1}\Bigg\{\bigg(\Big(-1+\frac{\varepsilon x_1+1}{j_\varepsilon^3}\Big)^2+2\Big(-1+\frac{1}{j_\varepsilon}\Big)^2+\frac{\varepsilon^2x_2^2}{j_\varepsilon^6}-\frac{\varepsilon x_3}{j_\varepsilon^3}\bigg)^{1/2} \\ &\qquad +\bigg(2\Big(-1+\frac{1}{j_\varepsilon}\Big)^2+\Big(-1-\frac{\varepsilon(\varepsilon+x_1)}{j_\varepsilon^3}+\frac{1}{j_\varepsilon}\Big)^2+\frac{\varepsilon^2x_2^2}{j_\varepsilon^6}+\frac{\varepsilon^2x_3^2}{j_\varepsilon^6}\bigg)^{1/2}\Bigg\}.\end{aligned}$$

Therefore

$$\lim_{\varepsilon\to 0}|D\Phi_\varepsilon(x)-\mathrm{Id}|=0$$

uniformly on B^n. It follows that there exists $\varepsilon>0$ such that

$$\sup_{|x|\le 1}|D\Phi_\varepsilon(x)-\mathrm{Id}|<1.$$

The inequality $|\Phi_\varepsilon(x)-\Phi_\varepsilon(y)+y-x|<|x-y|$ yields $0<|\Phi_\varepsilon(x)-\Phi_\varepsilon(y)|$. This implies that Φ_ε is injective.

4.3. ***A question.*** Lewy's theorem fails in higher dimensions, as shown by Wood [1991], who constructed the harmonic homeomorphism

$$u(x,y,z)=(x^3-3xz^2+yz,\, y-3xz,\, z),$$

which is not a diffeomorphism. The Jacobian of u is $J_u(x,y,z)=3x^2-3z^2$. This means that u is neither a diffeomorphism nor a quasiconformal mapping. Do there exist quasiconformal harmonic mappings which are not diffeomorphisms? If they exist, then of course $K\ge 2^{n-1}$, as shown above (and in [Tam and Wan 1998]).

Acknowledgment

We thank the referee for providing constructive comments and help in improving this paper.

References

[Ahlfors 1981] L. V. Ahlfors, *Möbius transformations in several dimensions*, University of Minnesota, School of Mathematics, Minneapolis, MN, 1981. MR 84m:30028

[Axler et al. 1992] S. Axler, P. Bourdon, and W. Ramey, *Harmonic function theory*, Graduate Texts in Mathematics **137**, Springer, New York, 1992. MR 93f:31001 Zbl 0765.31001

[Fefferman 1974] C. Fefferman, "The Bergman kernel and biholomorphic mappings of pseudoconvex domains", *Invent. Math.* **26**:1 (1974), 1–65. MR 50 #2562 Zbl 0289.32012

[Fehlmann and Vuorinen 1988] R. Fehlmann and M. Vuorinen, "Mori's theorem for n-dimensional quasiconformal mappings", *Ann. Acad. Sci. Fenn. Ser. A I Math.* **13**:1 (1988), 111–124. MR 90a:30060 Zbl 0615.30018

[Finn and Serrin 1958] R. Finn and J. Serrin, "On the Hölder continuity of quasi-conformal and elliptic mappings", *Trans. Amer. Math. Soc.* **89**:1 (1958), 1–15. MR 20 #4094

[Gehring and Martio 1985] F. W. Gehring and O. Martio, "Lipschitz classes and quasiconformal mappings", *Ann. Acad. Sci. Fenn. Ser. A I Math.* **10** (1985), 203–219. MR 87b:30029 Zbl 0584.30018

[Gilbarg and Hörmander 1980] D. Gilbarg and L. Hörmander, "Intermediate Schauder estimates", *Arch. Rational Mech. Anal.* **74**:4 (1980), 297–318. MR 82a:35038 Zbl 0454.35022

[Hopf 1952] E. Hopf, "A remark on linear elliptic differential equations of second order", *Proc. Amer. Math. Soc.* **3**:5 (1952), 791–793. MR 14,280b Zbl 0048.07802

[Kalaj 2004] D. Kalaj, "Quasiconformal harmonic functions between convex domains", *Publ. Inst. Math.* (*Beograd*) (*N.S.*) **76**:90 (2004), 3–20. MR 2005j:30032

[Kalaj 2008] D. Kalaj, "Quasiconformal and harmonic mappings between Jordan domains", *Math. Z.* **260**:2 (2008), 237–252. MR 2009e:30038 Zbl 1151.30014

[Kalaj 2009] D. Kalaj, "A priori estimate of gradient of a solution to certain differential inequality and quasiconformal mappings", preprint, version 3, 2009. arXiv 0712.3580v3

[Kalaj and Mateljević 2006] D. Kalaj and M. Mateljević, "Inner estimate and quasiconformal harmonic maps between smooth domains", *J. Anal. Math.* **100**:1 (2006), 117–132. MR 2008b:30033 Zbl 1173.30311

[Kalaj and Pavlović 2005] D. Kalaj and M. Pavlović, "Boundary correspondence under quasiconformal harmonic diffeomorphisms of a half-plane", *Ann. Acad. Sci. Fenn. Math.* **30**:1 (2005), 159–165. MR 2006i:30027 Zbl 1071.30016

[Kalaj and Pavlović 2009] D. Kalaj and M. Pavlović, "On quasiconformal self-mappings of the unit disk satisfying Poisson's equation", preprint, 2009, available at http://poincare.matf.bg.ac.rs/~pavlovic/TrAMS.pdf. To appear in *Trans. Amer. Math. Soc.*

[Koskela et al. 2001] P. Koskela, J. Onninen, and J. T. Tyson, "Quasihyperbolic boundary conditions and capacity: Hölder continuity of quasiconformal mappings", *Comment. Math. Helv.* **76**:3 (2001), 416–435. MR 2002j:30026

[Laugesen 1996] R. S. Laugesen, "Injectivity can fail for higher-dimensional harmonic extensions", *Complex Variables Theory Appl.* **28**:4 (1996), 357–369. MR 2001a:31003 Zbl 0871.54020

[Manojlović 2009] V. Manojlović, "Bi-Lipschicity of quasiconformal harmonic mappings in the plane", *Filomat* **23**:1 (2009), 85–89. MR 2568617 Zbl 05649323

[Martio 1968] O. Martio, "On harmonic quasiconformal mappings", *Ann. Acad. Sci. Fenn. Ser. A I No.* **425** (1968), 3–10. MR 38 #4678 Zbl 0162.37902

[Martio 2009] O. Martio, personal communication, 2009.

[Melas 1993] A. D. Melas, "An example of a harmonic map between Euclidean balls", *Proc. Amer. Math. Soc.* **117**:3 (1993), 857–859. MR 93d:58037 Zbl 0836.54007

[Pavlović 2002] M. Pavlović, "Boundary correspondence under harmonic quasiconformal homeomorphisms of the unit disk", *Ann. Acad. Sci. Fenn. Math.* **27**:2 (2002), 365–372. MR 2003g:30030 Zbl 1017.30014

[Protter and Weinberger 1967] M. H. Protter and H. F. Weinberger, *Maximum principles in differential equations*, Prentice-Hall, Englewood Cliffs, NJ, 1967. MR 36 #2935 Zbl 0153.13602

[Rešetnjak 1968] J. G. Rešetnjak, "Generalized derivatives and differentiability almost everywhere", *Mat. Sb.* (*N.S.*) **75(117)**:3 (1968), 323–334. In Russian; translated in *Math. USSR Sb.* **4**:3 (1968), 293–302. MR 37 #754 Zbl 0176.12001

[Tam and Wan 1998] L.-F. Tam and T. Y.-H. Wan, "On quasiconformal harmonic maps", *Pacific J. Math.* **182**:2 (1998), 359–383. MR 99b:58066 Zbl 0892.58017

[Vuorinen 1988] M. Vuorinen, *Conformal geometry and quasiregular mappings*, Lecture Notes in Mathematics **1319**, Springer, Berlin, 1988. MR 89k:30021 Zbl 0646.30025

[Wood 1991] J. C. Wood, "Lewy's theorem fails in higher dimensions", *Math. Scand.* **69**:2 (1991), 166. MR 93a:58024 Zbl 0711.31003

Received June 20, 2009. Revised January 11, 2010.

DAVID KALAJ
UNIVERSITY OF MONTENEGRO
FACULTY OF NATURAL SCIENCES AND MATHEMATICS
DŽORDŽA VAŠINGTONA
81000 PODGORICA
MONTENEGRO

davidk@t-com.me

MIODRAG S. MATELJEVIĆ
UNIVERSITY OF BELGRADE
FACULTY OF MATHEMATICS
STUDENTSKI TRG 16
11000 BELGRADE
SERBIA

miodrag@matf.bg.ac.rs

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

KLEIN BOTTLE AND TOROIDAL DEHN FILLINGS AT DISTANCE 5

SANGYOP LEE

We determine all hyperbolic 3-manifolds M such that $M(\pi)$ contains a Klein bottle, $M(\tau)$ contains an essential torus, and $\Delta(\pi, \tau) = 5$. As a corollary, we prove that if a hyperbolic 3-manifold M has two slopes π and τ on its boundary torus such that $M(\pi)$ is a lens space containing a Klein bottle and $M(\tau)$ is toroidal, then $\Delta(\pi, \tau) \leq 4$.

1. Introduction

Let M be a compact, connected, orientable 3-manifold with a torus boundary component $\partial_0 M$. A *slope* on $\partial_0 M$ is the isotopy class of an essential simple closed curve in $\partial_0 M$. Given a slope γ on $\partial_0 M$, denote by $M(\gamma)$ the 3-manifold obtained by γ*-Dehn filling* on M along $\partial_0 M$, that is, $M(\gamma)$ is obtained from M by gluing a solid torus V_γ along $\partial_0 M$ so that γ bounds a meridional disk of V_γ. For two slopes γ_1, γ_2 on $\partial_0 M$, denote by $\Delta(\gamma_1, \gamma_2)$ the distance between the slopes, which is their geometric intersection number.

We shall say that a 3-manifold M is *hyperbolic* if M with its boundary tori removed admits a complete hyperbolic structure with totally geodesic boundary. A Dehn filling on M is said to be *exceptional* if it produces a nonhyperbolic 3-manifold, which is either reducible, boundary-reducible, annular, toroidal, or a small Seifert fiber space. It is a well-known theorem of Thurston that there are only finitely many exceptional Dehn fillings on each boundary torus of M.

Gordon and Wu [2008] determined all hyperbolic 3-manifolds admitting two toroidal Dehn fillings at distance 4 or 5. In this paper, we determine all hyperbolic 3-manifolds M admitting two Dehn fillings at distance 5, one of which yields a Klein bottle, the other yielding an essential torus.

Following [Martelli and Petronio 2006], we use N to denote the *magic manifold*, the exterior of the chain link with three components in S^3, shown in Figure 1. Using the standard meridian-longitude framing on each boundary component of N, we identify a slope γ with a number in $\mathbb{Q} \cup \{1/0\}$. We denote by $N(r)$ the result of

MSC2000: 57M50.
Keywords: toroidal manifolds, Klein bottles, Dehn fillings.
This research was supported by a Chung-Ang University research grant in 2010.

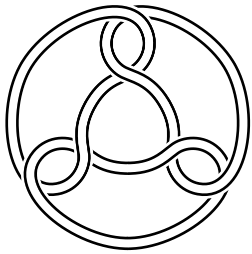


Figure 1. The magic manifold.

Dehn filling on N along a slope corresponding to the number r. Since N admits an automorphism interchanging any two of its boundary components, $N(r)$ is defined independently of the choice of the boundary component of N. Partial Dehn fillings give $N(r, s)$ and $N(r, s, t)$. We also use W to denote the Whitehead link exterior and use $W(r)$ and $W(r, s)$ to denote the corresponding Dehn-filled manifolds. The main result of this paper is the following.

Theorem 1.1. *Let M be a hyperbolic 3-manifold with a torus boundary component $\partial_0 M$. Suppose that there are two slopes π and τ on $\partial_0 M$ such that $M(\pi)$ contains a Klein bottle, $M(\tau)$ contains an essential torus, and $\Delta(\pi, \tau) = 5$. Then $M(\pi)$ is toroidal and either M is equal to either $N(1, -1/3)$, $N(-5/3, -5/3)$, $N(1, 5)$, $N(2, 2)$, or $N(-4, (2n-1)/2)$ for some integer $n \neq 0, -1$.*

We remark that the manifolds in this theorem are identified with some of the manifolds in [Gordon and Wu 2008, Definition 21.3] as follows: $N(1, -1/3) = M_5$, $N(-5/3, -5/3) = M_7$, $N(1, 5) = M_8$, and $N(2, 2) = M_{12}$. Also, $N(-4) = M_3$ is the Whitehead sister link exterior and $N(1, -1/3) = M_5 = W(4/3)$ and $N(1, 5) = M_8 = W(-4)$. See the proofs of Lemmas 2.2, 6.1, 7.3, 7.4, and 8.1.

Corollary 1.2. *Let M be a hyperbolic 3-manifold with ∂M a torus. Suppose that there are two slopes π and τ on ∂M such that $M(\pi)$ is a lens space containing a Klein bottle and $M(\tau)$ contains an essential torus. Then $\Delta(\pi, \tau) \leq 4$.*

Proof. This follows from [Gordon 1999, Theorem 1.1] and Theorem 1.1. □

In an unpublished paper, Teragaito [2000] obtained the same result.

2. Preliminaries

Let M be a hyperbolic 3-manifold with a torus boundary component $\partial_0 M$ such that $M(\pi)$ contains a Klein bottle and $M(\tau)$ contains an essential torus for two slopes π and τ on $\partial_0 M$. Assume $\Delta(\pi, \tau) = 5$.

Lemma 2.1 [Oh 1997; Wu 1998]. *$M(\pi)$ is irreducible.*

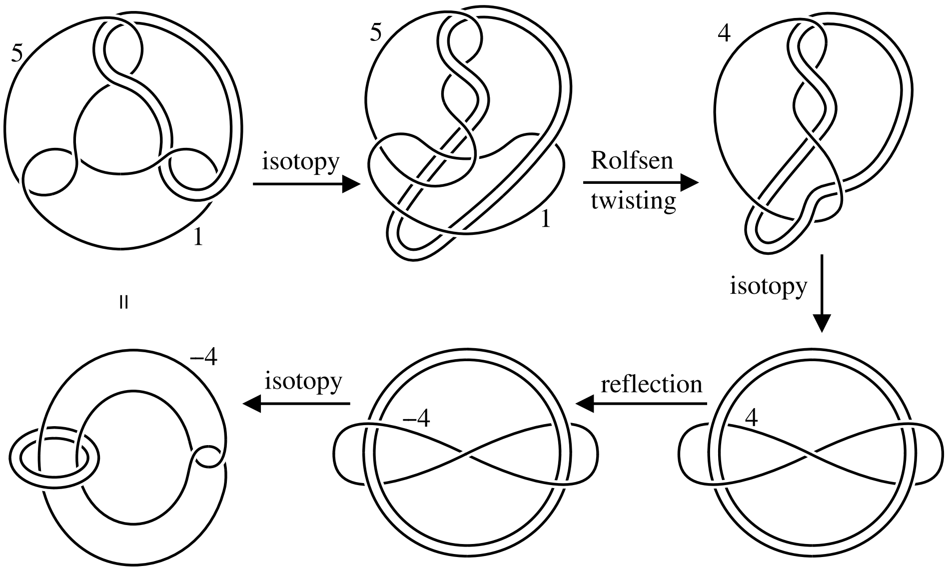


Figure 2. $N(1,5)$ is homeomorphic to $W(-4)$.

Lemma 2.2. *If $M(\tau)$ contains a Klein bottle, then $M(\pi)$ is toroidal and $M = N(1,5) = W(-4)$.*

Proof. Suppose that $M(\tau)$ contains a Klein bottle. Then we have $M = W(-4)$ by [Lee 2007, Theorem 1.4]. Figure 2 shows that $N(1,5) = W(-4)$. In fact, $W(-4)$ is homeomorphic to $M5_1$ in [Martelli and Petronio 2006, Table A.4]. From the table, one sees that M has only one pair of slopes along which Dehn fillings on M give 3-manifolds containing a Klein bottle. Also, the resulting 3-manifolds are toroidal. □

From now on, we assume that $M(\tau)$ does not contain a Klein bottle.

Let $\hat{P}$ be a Klein bottle in $M(\pi)$, chosen so that the core K_π of the attached solid torus V_π intersects $\hat{P}$ transversely and minimally among all Klein bottles in $M(\pi)$. Then $\hat{P} \cap V_\pi$ is a union of meridian disks of V_π, $u_1, \dots, u_p$, numbered successively along V_π. Similarly, we choose an essential torus $\widehat{T}$ in $M(\tau)$ such that $\widehat{T} \cap V_\tau$ is a union of meridian disks of V_τ, $v_1, \dots, v_t$, where t is minimal.

Let $P = \hat{P} \cap M$ and $T = \widehat{T} \cap M$. We may assume that P and T meet transversely. Then $P \cap T$ is a union of circles and arcs. In the usual way, the arc-components of $P \cap T$ define two labeled graphs G_P and G_T on $\hat{P}$ and $\widehat{T}$, respectively. The vertices of G_P and G_T are the meridian disks $u_1, \dots, u_p$ and $v_1, \dots, v_t$, respectively, and the edges are the arc-components of the intersection. A point in $\partial u_x \cap \partial v_y$ is labeled y in G_P and x in G_T. In G_P and G_T, labels $1, \dots, t$ and $1, \dots, p$ respectively appear in order around each vertex, repeating Δ times.

Let q_1, q_2, q_3, q_4, q_5 be the points in $\partial u_x \cap \partial v_y$, which are successively numbered along ∂u_x. Then the points appear in the order of q_d, q_{2d}, q_{3d}, q_{4d}, q_{5d} on ∂v_y in some direction. The number d is called a *jumping number*, and $d = 1$ or 2. See [Gordon and Wu 1999, Lemma 2.10].

Orient the boundary circles of P (respectively T) so that they are mutually homologous on $\partial_0 M$. Every edge of G_P (respectively G_T) has a rectangular neighborhood R whose opposite sides are contained in two (or possibly one) boundary components of P (respectively T). We say the edge is *positive* if some orientation of ∂R is compatible with the orientations of the boundary components. Otherwise, we say it is *negative*. Then we have the *parity rule*: an edge is positive in one graph if and only if it is negative in the other.

Let $G = G_P$ or G_T. We call an edge in G a *level edge* if it has the same label at its endpoints; we call it an x*-edge* if one of its endpoints is labeled x. Let G^+ denote the subgraph of G consisting of all positive edges, and for a label x, let $G^+(x)$ denote the subgraph of G^+ consisting of all x-edges of G^+. A disk face of $G^+(x)$ is called an x*-face*. The boundary of an x-face is called a *Scharlemann cycle* if the x-face is a disk face of G. Note that each edge of a Scharlemann cycle has two consecutive labels, say x and $x+1$, at its endpoints. In this case, the Scharlemann cycle is called an $(x, x+1)$*-Scharlemann cycle*. A Scharlemann cycle of length 2 is called an S*-cycle*. A cycle of positive edges is called an *extended Scharlemann cycle* if it immediately surrounds a Scharlemann cycle. Let $\overline{G}$ denote the *reduced graph* of G, the graph obtained by amalgamating parallel edges of G into a single edge. The *weight* of an edge $\bar{e}$ of $\overline{G}$ is the number of the edges of G in $\bar{e}$.

We call a vertex u_x of G_P a *level vertex* if there exists a positive level x-edge in G_T. Also, we call a vertex v_x of G_T a *Scharlemann vertex* if there exists a Scharlemann cycle with label x in G_P. If e is a positive level x-edge in G_T, then $u_x \cup e$ has a Möbius band neighborhood in $\hat{P}$.

Lemma 2.3. *Suppose that $p \geq 2$. Then G_T satisfies the following.*

(1) *At most two labels of G_T can be labels of positive level edges.*

(2) *G_T cannot contain a Scharlemann cycle.*

(3) *Any family of parallel positive edges in G_T contains at most $p/2+1$ edges. If the family contains $p/2+1$ edges, then the two outermost edges of the family are level.*

(4) *Any family of parallel negative edges in G_T contains at most p edges.*

Proof. See the proof of [Lee and Teragaito 2008, Lemma 6.2]. □

Lemma 2.4. *G_P satisfies the following. Assume $t \geq 3$ in (6) and (7).*

(1) *If G_P contains a Scharlemann cycle, then $\widehat{T}$ is separating in $M(\tau)$.*

(2) *The edges of any Scharlemann cycle of* G_P *cannot be contained in a disk in* $\widehat{T}$.

(3) *If* $t > 2$, *then* G_P *cannot contain an extended Scharlemann cycle.*

(4) *If* G_P *contains two Scharlemann cycles on disjoint label pairs* $\{a, a+1\}$ *and* $\{b, b+1\}$, *then* $a \equiv b \pmod 2$.

(5) G_P *has at most four labels of Scharlemann cycles, that is,* G_T *has at most four Scharlemann vertices.*

(6) *Any family of parallel positive edges in* G_P *contains at most* $t/2+1$ *edges. If* t *is odd, then the family contains less than* $t/2$ *edges.*

(7) *Any family of parallel negative edges in* G_P *contains at most* $t+1$ *edges. If* G_P *contains* $t+1$ *parallel negative edges, then* $G_T^+ = G_T$.

Proof. For (1)–(5), see [Gordon and Wu 2008, Lemma 2.2, parts (4), (5) and (6), and Lemma 2.3, parts (2) and (4)], and for (6) and (7), see [Lee and Teragaito 2008, Lemma 2.5, parts (ii) and (iii)] and [Valdez-Sánchez 2007, Proposition 3.4]. □

Lemma 2.5 [Gordon 1998, Lemma 2.1]. *No two edges are parallel in both graphs* G_P *and* G_T.

Lemma 2.6. G_P *cannot contain two* S*-cycles on disjoint label pairs.*

Proof. If G_P contained two S-cycles on disjoint label pairs, then the construction as in the proof of [Gordon and Luecke 1995, Lemma 3.10] would give a Klein bottle in $M(\tau)$, contradicting our assumption. □

For any submanifold A of a manifold X, we will use $\eta(A)$ to denote a closed regular neighborhood of A in X.

3. Generic case

In this section we will show that the generic case $p \geq 3$ and $t = 3$ or $t \geq 5$ cannot happen. To do this, we first estimate the number of negative (or positive) edge endpoints of the graphs G_P and G_T. Note that the total number of edge endpoints of each graph is $\Delta pt = 5pt$.

Using the argument of [Lee 2007, Section 3], one can prove the following two lemmas and proposition. See [Lee 2007, Lemmas 3.2 and 3.3 and Proposition 3.4].

Lemma 3.1. *Assume* $p \geq 3$. *Let* x *be a label of* G_T *that is not a label of a positive level edge. Then any* x*-face in* G_T *has at least* 4 *sides.*

Lemma 3.2. *Assume* $p \geq 3$. *If* G_T *contains a positive level* x*-edge, then* G_T *cannot contain an* x*-face.*

Proposition 3.3. *Assume* $p \geq 3$.

(1) *Any level vertex of* G_P *has at most* $2t$ *negative edge endpoints.*

(2) *Any nonlevel vertex of G_P has at most $2t-1$ negative edge endpoints.*

In the following lemma, we use an Euler characteristic calculation to give an upper bound for the number of Scharlemann cycles in G_P in terms of the number of negative edge endpoints at a vertex of G_T.

Lemma 3.4. *Suppose that G_T has k ($\geq p$) negative edge endpoints at a vertex v_x. Then G_P contains at least $k-p$ Scharlemann cycles.*

Proof. There is no negative loop edge in G_T, since otherwise $\widehat{T}$ would contain an orientation-reversing curve. Hence G_T has k negative edges incident to v_k and by the parity rule G_P has k positive x-edges.

Let V, E and F be the number of vertices, edges, and disk faces of $G_P^+(x)$, respectively. Then $V=p$, $E=k$, and an Euler characteristic calculation for the graph $G_P^+(x)$ gives $V-E+F\geq\chi(\hat{P})=0$, so $F\geq E-V=k-p$. Recall that each disk face of $G_P^+(x)$ is an x-face in G_P. Hence the number of x-faces in G_P is at least $k-p$, and each contains at least one Scharlemann cycle by [Hayashi and Motegi 1997, Proposition 5.1]. □

Lemma 3.5. *Let v_x be a vertex of G_T. Suppose that any x-face in G_P has at least 3 sides. Then v_x has a most $3p-1$ negative edge endpoints.*

Proof. This lemma is essentially [Lee 2007, Lemma 2.7]. Assume for contradiction that v_x has at least $3p$ negative edge endpoints. Let V, E and F be as in the proof of Lemma 3.4. Then $V=p$, $E\geq 3p$, and $F\geq E-V\geq 2p$. Since any x-face in G_P (and hence any disk face of $G_P^+(x)$) has at least 3 sides, we have $2E\geq 3F\geq 3(E-V)$, which gives $E\leq 3V=3p$. Hence $E=3p$, $F=2p$, and every face of $G_P^+(x)$ is a 3-sided disk face. Since every face of $G_P^+(x)$ is a disk face, we can conclude that $G_P=G_P^+$. So, every x-edge in G_P is positive and hence we have $E=5p$. This is a contradiction. □

Proposition 3.6. *Assume $t\geq 5$.*

(1) *Any Scharlemann vertex of G_T has at most $3p$ negative edge endpoints.*

(2) *Any non-Scharlemann vertex of G_T has at most $3p-1$ negative edge endpoints.*

Proof. If v_x is not a Scharlemann vertex, then any x-face in G_P has at least 3 sides by Lemma 2.4(3), so v_x has at most $3p-1$ negative edge endpoints by Lemma 3.5. Thus we only need to prove the first statement of the proposition. By Lemma 2.4(5), G_T has at most four Scharlemann vertices. We divide our argument into two cases according to the number of Scharlemann vertices of G_T.

First, suppose G_T has at most three Scharlemann vertices. Then any Scharlemann cycle in G_P has label pair $\{a-1,a\}$ or $\{a,a+1\}$ for some label a. Let k be the number of negative edge endpoints of G_T at v_a. Then there are exactly k

positive a-edges in G_P, and by Lemma 3.4 there are at least $k-p$ Scharlemann cycles in G_P. Since $t \geq 5$, no two Scharlemann cycles can share an edge. Since each Scharlemann cycle has at least two edges, the number of positive a-edges in G_P, which is equal to k, is at least $2(k-p)$. So, we have $k \geq 2(k-p)$ and hence $k \leq 2p$. Thus v_a has at most $2p$ negative edge endpoints. If some other vertex v_x of G_T has more than $2p$ negative edge endpoints, then G_P has more than p Scharlemann cycles by Lemma 3.4. This implies that there are more than $2p$ positive a-edges in G_P, which contradicts the fact that v_a has at most $2p$ negative edge endpoints. Hence we conclude that any vertex of G_T has at most $2p$ negative edge endpoints.

Now suppose that G_T has exactly four Scharlemann vertices, say v_1, v_2, v_b and v_{b+1}. Relabeling if necessary, we may assume $b+1 < p$. By Lemma 2.4(4), we have $1 \equiv b \pmod 2$. Since G_P cannot contain two S-cycles on disjoint label pairs, we may assume that any Scharlemann cycle on the label pair $\{b, b+1\}$ has length at least 3. Let m and n be the number of $(1,2)$-Scharlemann cycles and $(b, b+1)$-Scharlemann cycles in G_P, respectively. If $b = 3$, then G_P may contain $(2,3)$-Scharlemann cycles. Let l be the number of $(2,3)$-Scharlemann cycles if $b = 3$, and let $l = 0$ otherwise.

Claim. *Let σ_1, σ_2 be $(b, b+1)$-Scharlemann cycles. Then σ_1 and σ_2 have the same length and there are two families of parallel edges of G_T such that each family contains the same number of edges from σ_1 and σ_2.*

Proof. By the existence of $(1,2)$-Scharlemann cycles and Lemma 2.4(2), there exists an annulus A in $\widehat{T}$ that contains the edges of σ_1 and σ_2. The core of A is an essential curve in $\widehat{T}$. Let $A_{b,b+1}$ be an annulus in ∂V_τ with $A_{b,b+1} \cap \widehat{T} = \partial A_{b,b+1} = \partial v_b \cup \partial v_{b+1}$. Then $F = (A - v_b \cup v_{b+1}) \cup A_{b,b+1}$ is a twice-punctured torus.

Let f_i for $i = 1, 2$ be the disk face of G_P bounded by σ_i. Since f_i is bounded by positive edges, its boundary curve ∂f_i is a nonseparating curve in F. If the two curves ∂f_1 and ∂f_2 are not parallel in F, then we compress F along $f_1 \cup f_2$ to obtain two disks, the boundaries of which are the two boundary curves of F. This implies that either of these disks is a compressing disk for $\widehat{T}$, which gives a contradiction. So, the two curves ∂f_1 and ∂f_2 cobound an annulus in F and the restriction of the annulus onto A $(\subset T)$ is a finite union of bigons, each realizing the parallelism in A between an edge of σ_1 and an edge of σ_2. The two border edges of each bigon are parallel in G_T, or the bigon contains some vertices of G_T in its interior. The second possibility can be ruled out using the argument in the proof of [Lee 2007, Lemma 2.8]. □

Since $(b, b+1)$-Scharlemann cycles have length at least 3, there is a family of parallel edges of G_T containing at least two edges from each such cycle. This family contains at most p edges by Lemma 2.3(4), so we have

$$n \leq p/2.$$

Claim. *G_P contains at most $2p$ Scharlemann cycles.*

Proof. Let k be the number of negative edge endpoints of G_T at v_2. Then there are exactly k positive 2-edges in G_P. Since each Scharlemann cycle in G_P has at least two edges and since any two Scharlemann cycles with label 2 cannot share an edge, the number of 2-edges in Scharlemann cycles in G_P is at least $2m+2l$. Thus we have

$$2l + 2m \leq k.$$

Note that $l+m+n$ is the total number of Scharlemann cycles in G_P. By Lemma 3.4 there are at least $k-p$ Scharlemann cycles in G_P. So, we have

$$k - p \leq l + m + n.$$

Combining the three inequalities above, we obtain

$$l + m \leq k - l - m \leq n + p \leq p/2 + p = 3p/2.$$

Thus we have $l+m+n \leq 3p/2 + p/2 = 2p$. □

By Lemma 3.4 and the previous claim, any vertex of G_T has at most $3p$ negative edge endpoints. □

Lemma 3.7. *$p \leq 2$ or $t \leq 4$.*

Proof. Assume for contradiction that $p \geq 3$ and $t \geq 5$. Let ℓ and s be the number of level vertices of G_P and Scharlemann vertices of G_T, respectively. Then we have $\ell \leq 2$ and $s \leq 4$ by Lemma 2.3(1) and Lemma 2.4(5). Let K be the number of negative edge endpoints of G_P. Then we have $K \leq 2t\ell + (2t-1)(p-\ell)$ by Proposition 3.3. On the other hand, by Proposition 3.6, any Scharlemann vertex of G_T has at least $2p$ positive edge endpoints and any non-Scharlemann vertex of G_T has at least $2p+1$ positive edge endpoints. By the parity rule, K is equal to the number of positive edge endpoints of G_T. So, we have $2ps + (2p+1)(t-s) \leq K$. Combining these inequalities, we obtain

$$2ps + (2p+1)(t-s) \leq K \leq 2t\ell + (2t-1)(p-\ell).$$

This gives $p+t \leq \ell + s \leq 2+4 = 6$, which violates our initial assumption. □

Lemma 3.8. *If $t = 3$, then $p = 2$.*

Proof. Assume $t = 3$. Since the number of edge endpoints of G_T (or G_P) is even, we cannot have $p = 1$.

Let $p \geq 3$. Since $t = 3$, $\widehat{T}$ is a nonseparating torus in $M(\tau)$. So, any vertex of G_T has at most p negative edge endpoints by Lemma 2.4(1) and Lemma 3.4,

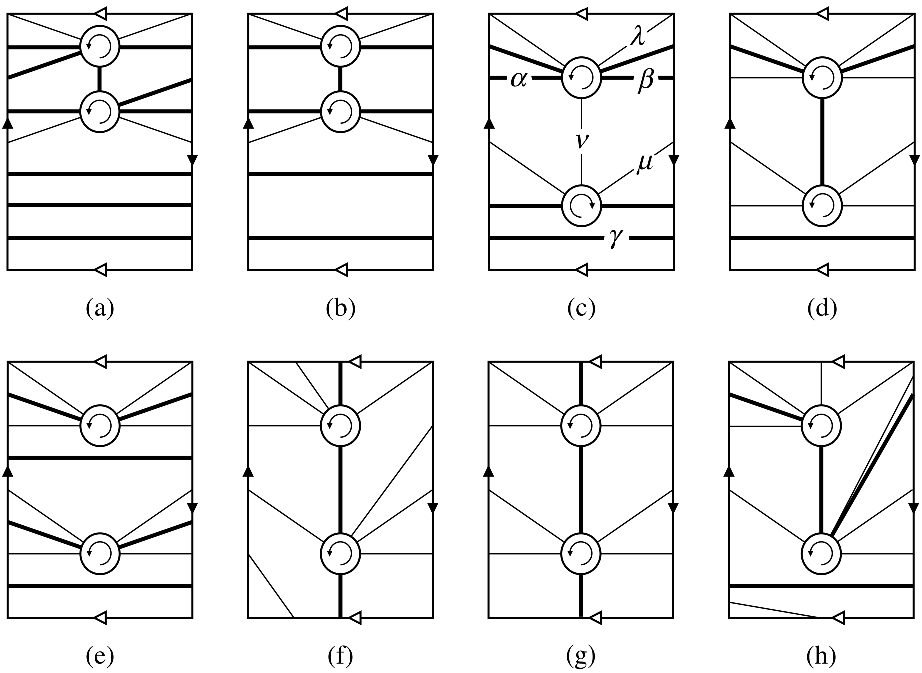


Figure 3. The reduced graph $\overline{G}_P$.

or equivalently it has at least $4p$ positive edge endpoints. Let K be the number of negative edge endpoints of G_P. Then we have $4pt \leq K \leq 2t\ell + (2t-1)(p-\ell)$, which gives $\ell \geq 2pt + p \geq 2\cdot 3\cdot 3 + 3 > 2$. This contradicts that $\ell \leq 2$. □

4. The case $p = 2$

In this case we will show that $t = 1, 2$, or 4. Assume for contradiction that $t = 3$ or $t \geq 5$. If $t \geq 5$, then any vertex of G_T has at most $3p$ negative edge endpoints by Proposition 3.6 and hence G_T has at least $2pt = 4t$ positive edge endpoints. If $t = 3$, we observed in the proof of Lemma 3.8 that any vertex of G_T has at least $4p$ positive edge endpoints, so G_T has at least $4pt = 8t$ positive edge endpoints.

In any case, G_T^+ has at least $2t$ edges. An Euler characteristic calculation shows that G_T^+ has at least t disk faces. By Lemma 2.3(2), each disk face of G_T^+ has at least one level i-edge on its boundary for each $i = 1, 2$. The parity rule implies that each vertex of G_P is a base of a negative loop edge. Then the proof of [Lee 2006, Lemma 5.1] remains valid here to show that $\overline{G}_P$ is a subgraph of one of the graphs in Figure 3, where the thick edges are positive and the thin edges are negative. Note that the number of edges of G_P is $\Delta pt/2 = 5t$.

Lemma 4.1. *G_P contains at least $2t$ positive edges.*

Proof. Assume not. Then G_T^+ has more than $3t$ edges, so an Euler characteristic calculation shows that it has more than $2t$ disk faces. Since each disk face of G_T^+ has at least one level 1-edge and since any such level 1-edge is shared by at most two disk faces of G_T^+, the number of positive level 1-edges in G_T is greater than t. Then, in G_P, the family of parallel negative loop edges based at u_1 contains more than t edges. By Lemma 2.4(7), $G_T^+ = G_T$. Then G_T^+ has $5t$ edges and at least $4t$ disk faces. This also implies that the number of positive level 1-edges is at least $2t$. By Lemma 2.4(4), we have $2t \leq t+1$ and hence $t \leq 1$. This contradicts our assumption that $t = 3$ or $t \geq 5$. □

For the first two graphs in Figure 3, each negative loop edge of $\overline{G}_P$ has weight at most t by Lemma 2.4(7). Hence G_P has at least $3t$ positive edges, which are divided into at most four families of parallel edges. By Lemma 2.4(6), we have $3t \leq 4 \cdot t/2$ if t is odd and $3t \leq 4 \cdot (t/2+1)$ if t is even. Both are impossible, since we assumed $t = 3$ or $t \geq 5$.

For the remaining graphs in the figure, $\overline{G}_P$ has at most three positive edges. By Lemma 4.1, G_P has at least $2t$ positive edges. Hence by Lemma 2.4(6), we have $2t \leq 3 \cdot t/2$ if t is odd and $2t \leq 3 \cdot (t/2+1)$ if t is even. The first inequality is impossible. The latter one is possible only if $t = 6$ and $\overline{G}_P$ is a subgraph of the graph in Figure 3(c). But, using (6) and (7) of Lemma 2.4, one can see that the lower vertex of the graph in Figure 3(c) has less than $5t$ edge endpoints in G_P. This is also impossible.

The following is what we proved in this section.

Lemma 4.2. *If* $p = 2$, *then* $t = 1, 2,$ *or* 4.

5. The case $t = 4$

In this case we will prove $p = 1$. On the contrary we assume $p \geq 2$ throughout this section.

Lemma 5.1. *Let* v_x *be a vertex of* G_T *such that* x *is not a label of an* S*-cycle in* G_P. *Then* v_x *has at most* $3p-1$ *negative edge endpoints, or equivalently it has at least* $2p+1$ *positive edge endpoints.*

Proof. Since x is not a label of an S-cycle of G_P, each x-face in G_P has at least 3 sides by Lemma 2.4(3). Hence the result follows from Lemma 3.5. □

Lemma 5.2. G_P *contains at most* $3p-2$ *Scharlemann cycles.*

Proof. There exists a label of G_P that is not a label of an S-cycle by Lemma 2.6. We may assume that the label is 4. We divide the Scharlemann cycles of G_P into two disjoint families; one family $\mathcal{S}_1$ consists of all Scharlemann cycles having label 2 and the other family $\mathcal{S}_2$ consists of all Scharlemann cycles having label 4. Let s_i (≥ 0) be the number of all Scharlemann cycles in $\mathcal{S}_i$ for $i = 1, 2$. Then $s_1 + s_2$

is the total number of Scharlemann cycles in G_P. Note that no two Scharlemann cycles in G_P can share an edge.

Since each Scharlemann cycle in $\mathcal{S}_2$ has at least 3 edges, there are at least $3s_2$ positive 4-edges in G_P. By the parity rule, G_T has at least $3s_2$ negative edge endpoints at v_4. By Lemma 5.1 we have $3s_2 \leq 3p-1$ and hence $s_2 \leq p-1$.

Now let k be the number of negative edge endpoints of G_T at v_2. Then by Lemma 3.4 we have $k-p \leq s_1+s_2$. On the other hand, G_P has k positive 2-edges. Since any Scharlemann cycle in $\mathcal{S}_1$ has at least two edges, we have $2s_1 \leq k$ and hence $2s_1 - p \leq k-p \leq s_1+s_2$. This gives $s_1 \leq s_2+p \leq (p-1)+p = 2p-1$. We now have $s_1+s_2 \leq (2p-1)+(p-1) \leq 3p-2$. □

Lemma 5.3. *Any vertex of G_T has at most $4p-2$ negative edge endpoints, or equivalently it has at least $p+2$ positive edge endpoints.*

Proof. This follows from Lemmas 3.4 and 5.2. □

Lemma 5.4. *Assume $p \geq 3$. Then $\widehat{T}$ is a separating torus in $M(\tau)$.*

Proof. Suppose that $\widehat{T}$ is nonseparating. Then any vertex of G_T has at least $4p$ positive edge endpoints by Lemma 2.4(1) and Lemma 3.4, so G_T contains at least $4pt/2 (= 8p)$ positive edges. Let n be the number of positive edges of $\overline{G}_T$, and let $q = p/2+1$ if p is even and $q = (p+1)/2$ if p is odd. By Lemma 2.3(3), G_T contains at most nq positive edges. Hence we have $8p \leq nq$, which gives

$$8p/q \leq n.$$

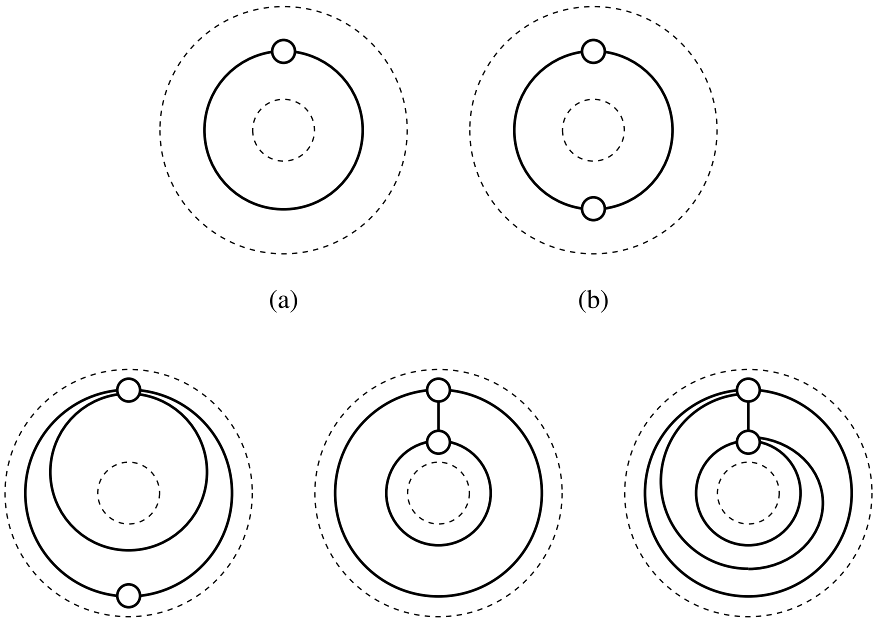
On the other hand, by [Gordon and Wu 2008, Lemma 2.5], $\overline{G}_T$ contains at most $3t$ edges. Hence by (3) and (4) of Lemma 2.3, we have $10p = 5pt/2 \leq nq + (3t-n)p = nq + 12p - np$, which gives

$$n \leq 2p/(p-q).$$

Combining the two inequalities above, we obtain $8p/q \leq n \leq 2p/(p-q)$, which gives $4p \leq 5q$. Solving this inequality, we obtain $3p \leq 10$ if p is even and $3p \leq 5$ if p is odd. Both cases violate the assumption that $p \geq 3$. □

Lemma 5.5. *Assume $p \geq 3$. Then each component $\overline{\Lambda}$ of $\overline{G}_T^+$ is contained in an essential annulus but not in a disk on $\widehat{T}$. There are only five possibilities for $\overline{\Lambda}$, as shown in Figure 4.*

Proof. Since $\widehat{T}$ is separating in G_T, each component of $\overline{G}_T^+$ has one or two vertices. By Lemma 2.3(3) and Lemma 5.3, each vertex of $\overline{G}_T^+$ has valency at least 2. Hence no component of $\overline{G}_T^+$ can be contained in a disk on $\widehat{T}$, so each component is contained in an essential annulus on $\widehat{T}$. By [Teragaito 2006a, Lemma 3.5], there are only five possibilities for $\overline{\Lambda}$. □



(a) (b)

(c) (d) (e)

Figure 4. A component of $\overline{G}_T^+$ with annular support.

Lemma 5.6. $p \leq 2$.

Proof. Assume $p \geq 3$. We may assume that label 4 is not a label of an S-cycle in G_P. Consider the component $\overline{\Lambda}$ of $\overline{G}_T^+$ containing v_4. By Lemma 5.1, v_4 has at least $2p+1$ positive edge endpoints in G_T, so by Lemma 2.3(3) it has valency at least 3 in $\overline{\Lambda}$. Hence $\overline{\Lambda}$ is one of the graphs in Figure 4(c)–(e). In particular, Λ has exactly two vertices v_2 and v_4. Let K be the number of edge endpoints of Λ. Then by Lemmas 5.1 and 5.3 we have

$$(2p+1)+(p+2) \leq K.$$

If x is not a label of a level edge in Λ, then it appears in Λ at most 3 times, since otherwise Λ would contain a 2- or 3-sided x-face, contradicting Lemma 3.1. If x is a label of a level edge in Λ, then it appears in Λ at most 4 times, since otherwise Λ would contain an x-face, contradicting Lemma 3.2. Hence we have

$$K \leq 3(p-\ell)+4\ell = 3p+\ell \leq 3p+2,$$

where ℓ is the number of labels of Λ that are a label of a level edge. Combining the two inequalities above, we obtain $(2p+1)+(p+2) \leq K \leq 3p+2$. This gives a contradiction. □

For the remainder of this section, we assume $p = 2$. Note that the number of edges of G_T is $\Delta pt/2 = 20$.

Lemma 5.7. *Any vertex of G_T has at least four positive edge endpoints.*

Proof. By Lemma 5.2, G_P contains at most 4 Scharlemann cycles. Hence by Lemma 3.4, each vertex of G_T has at most 6 negative edge endpoints, or equivalently has at least 4 positive edge endpoints. □

Using this lemma, one sees that G_T^+ has at least 8 edges. Hence G_T^+ contains at least four disk faces, each of which contains at least one level i-edge for each $i = 1, 2$. This shows that each vertex of $\overline{G}_P$ is a base of a negative loop edge. So, $\overline{G}_P$ is a subgraph of one of the eight graphs in Figure 3.

By Lemma 4.1, G_P has at least 8 positive edges (so, $G_T^+ \neq G_T$). By part (6) of Lemma 2.4, any family of parallel positive edges in G_P contains at most 3 edges. Hence $\overline{G}_P^+$ has at least 3 edges. It follows that $\overline{G}_P$ is a subgraph of one of the first three graphs in Figure 3.

Assume that $\overline{G}_P$ is a subgraph of the graph in Figure 3(a) or (b). Then $\overline{G}_P$ has exactly two negative edges, each of which containing at most 4 edges of G_P by Lemma 2.4(7). Hence G_P has at least 12 positive edges. In fact, by Lemma 2.4(6), G_P has exactly 12 positive edges and $\overline{G}_P$ is the graph in Figure 3(a). Also, each positive edge of $\overline{G}_P$ contains exactly three edges of G_P. Examining the labels of G_P, one sees that G_P must contain two S-cycles on disjoint label pairs. This contradicts Lemma 2.6.

Hence $\overline{G}_P$ is a subgraph of the graph in Figure 3(c). Label the edges of $\overline{G}_P$ as in the figure, and let $|\cdot|$ denote the weight of the corresponding reduced edge. Then $|\alpha|, |\beta|, |\gamma| \leq 3$ and $|\lambda|, |\mu|, |\nu| \leq 4$. But, the number of edge endpoints of G_P at the lower vertex of the graph in Figure 3(c) is $|\alpha| + |\beta| + 2|\mu| + |\nu| \leq 18$. This is impossible since each vertex of G_P has $\Delta t = 20$ edge endpoints. Hence we conclude that $p = 2$ is impossible.

Summarizing the results obtained in this section, we have the following.

Lemma 5.8. *If $t = 4$, then $p = 1$.*

6. The case $t = 1$

In this case, the reduced graph $\overline{G}_T$ has at most 3 edges. See Figure 5. The number of edges of G_T is $\Delta pt/2 = 5p/2$ (so, p is even). By Lemma 2.3(3) we have $5p/2 \leq 3(p/2 + 1)$ and hence $p \leq 3$. Since p is even, we have $p = 2$ and we can determine the graph pair G_T, G_P as shown in Figure 6. One can see that the jumping number for the graph pair is 1, so the edge correspondence between the two graphs is as shown in the figure.

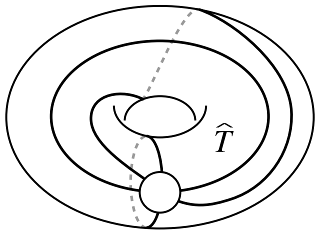


Figure 5. The reduced graph $\overline{G}_T$.

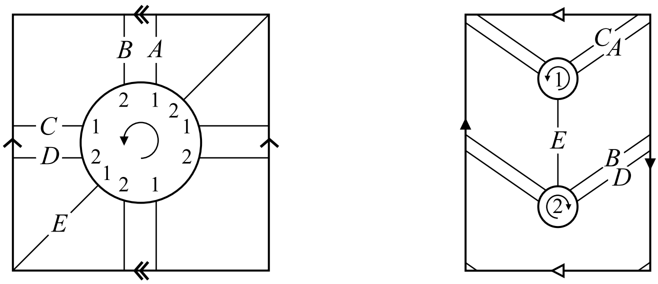


Figure 6. The graph pair G_T, G_P.

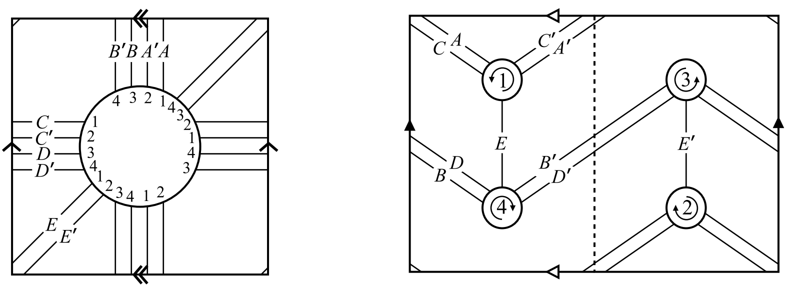


Figure 7. The graph pair G'_T, G_S.

A thin neighborhood $\eta(\hat{P})$ of $\hat{P}$ is a twisted I-bundle over the Klein bottle $\hat{P}$. Its boundary, $\hat{S} = \partial\eta(\hat{P})$, is a torus. Let $S = \hat{S} \cap M$. As done in Section 2, we construct two labeled graphs G_S and G'_T from the intersection of S and T, where G'_T is obtained by doubling the edges of G_T and G_S double-covers G_P. See Figure 7 for the graphs G'_T and G_S and the edge correspondence between them. The graph G_S is homeomorphic to the graph shown in Figure 8(a).

Let $Z = M(\pi) - \mathrm{Int}(\eta(\hat{P}))$. Then $M(\pi) = \eta(\hat{P}) \cup_{\hat{S}} Z$ and $Z \cap V_\pi$ is a union of two 1-handles V_{41} and V_{23}, where $V_{i,i+1}$ is the part of V_π between two vertices of G_S labeled i and $i+1$. Let f, g, and h be the faces of G'_T bounded by the edges $A' \cup B$, $A \cup D' \cup E$, and $B' \cup C \cup E'$, respectively. By compressing the genus 3 surface $\partial(\eta(\hat{P}) \cup V_\pi)$ along the three disks f, g, and h, one obtains a 2-sphere

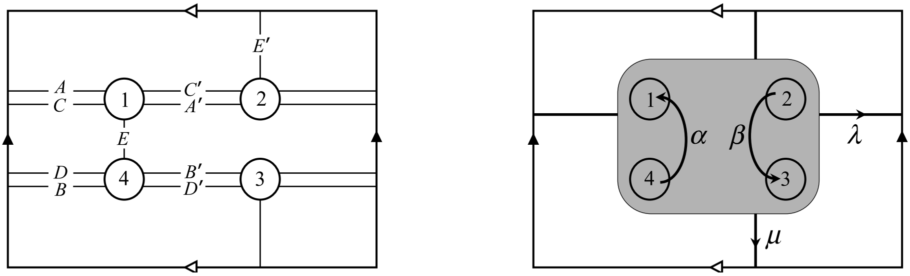


Figure 8. Generators for $\pi_1(Z)$.

in $M(\pi)$, which bounds a 3-ball by the irreducibility of $M(\pi)$. This implies that $\eta(\hat{S} \cup V_{41} \cup V_{23} \cup f \cup g \cup h)$ is Z minus a 3-ball. Thus $\eta(\hat{S} \cup V_{41} \cup V_{23} \cup f \cup g \cup h)$ and Z have the same fundamental group.

To calculate $\pi_1(Z)$, we follow an argument in [Teragaito 2000]. As a base point of Z, we take a disk containing the vertices of G_S as shown in Figure 8(b). The group $\pi_1(Z)$ has four generators α, β, λ, and μ as shown in the figure, where α and β are represented by the cores of V_{41} and V_{23}, respectively. The two generators λ and μ give a relation $\lambda\mu = \mu\lambda$ and the three disks f, g, and h give three relations $\lambda\alpha\beta = 1$, $\lambda\alpha^{-2}\beta^{-1} = 1$, and $\mu\beta\alpha\lambda^{-1}\beta = 1$, respectively. Hence $\pi_1(Z)$ has the presentation

$$\langle \alpha, \beta, \lambda, \mu : \lambda\mu = \mu\lambda, \lambda\alpha\beta = 1, \lambda\alpha^{-2}\beta^{-1} = 1, \mu\beta\alpha\lambda^{-1}\beta = 1 \rangle.$$

Using the last two relations, one can the eliminate two generators λ and μ to obtain $\pi_1(Z) = \langle \alpha, \beta : \alpha^3\beta^2 = 1 \rangle$. This group is isomorphic to the fundamental group of the trefoil knot exterior, so Z is not a solid torus. This implies that $\hat{S}$ is an essential torus in $M(\pi)$.

The graph pair in Figure 7 is homeomorphic to that in [Gordon and Wu 2008, Figure 11.10]. Hence M is homeomorphic to M_5 in the notation of [ibid., Definition 21.3].

Lemma 6.1. *If $t = 1$, then $M(\pi)$ is toroidal and $M = N(1, -1/3) = W(4/3)$.*

Proof. We only need to show that $M = N(1, -1/3) = W(4/3)$. By applying similar moves as in Figure 2, one can see that $N(1, -1/3)$ is homeomorphic to $W(4/3)$. From [Martelli and Petronio 2006, Table A.4], one sees that $N(1, -1/3, -4)$ contains a Klein bottle and $N(1, -1/3, 1)$ is toroidal. Here, $\Delta(-4, 1) = 5$.

We already saw that if $t = 1$, then M is uniquely determined ($M = M_5$). Hence we only need to show $N(1, -1/3)$ contains a properly embedded once-punctured torus with boundary slope 1. By a Rolfsen twisting (see Figure 2), slope 1 on $\partial N(1, -1/3)$ is changed into slope 0 on the boundary torus of $W(4/3)$. It is easy to see that slope 0 is a boundary slope of a once-punctured torus in $W(4/3)$. □

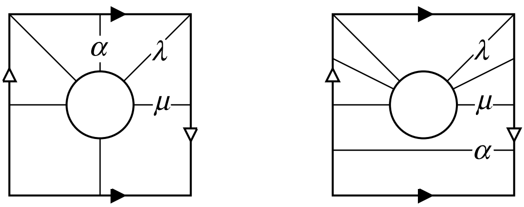


Figure 9. The reduced graph $\overline{G}_P$.

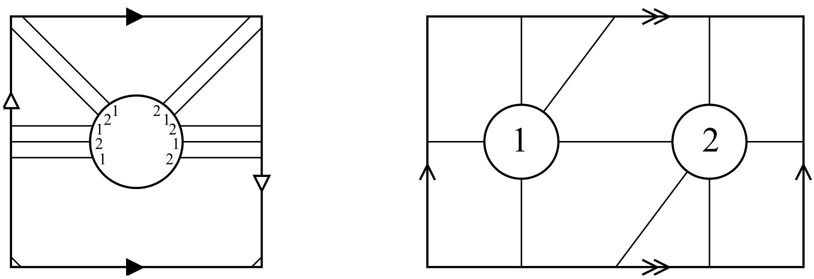


Figure 10. The graph pair G_P, G_T.

7. The case $p = 1$

In this case the reduced graph $\overline{G}_P$ has one of the forms in Figure 9. Label the edges of $\overline{G}_P$ as in the figure. Note that α is positive while λ and μ are negative. We write $G_P = \Gamma_1(|\alpha|, |\lambda|, |\mu|)$ or $\Gamma_2(|\alpha|, |\lambda|, |\mu|)$ according to whether $\overline{G}_P$ is the first or second graph in Figure 9. Up to homeomorphism of $\hat{P}$, we have $\Gamma_i(a, b, c) \cong \Gamma_i(a, c, b)$ for each $i = 1, 2$.

Lemma 7.1. $|\alpha| > 0$.

Proof. Assume for contradiction that $|\alpha| = 0$. We have $|\lambda|, |\mu| \leq t + 1$ by Lemma 2.4(7). The number of edges of G_P is $\Delta pt/2 = 5t/2$ (so, t must be even) and hence $5t/2 = |\lambda| + |\mu| \leq 2t + 2$, giving $t \leq 4$.

If $t = 4$, then $|\lambda| = |\mu| = 5$; this is impossible by [Teragaito 2006b, Lemma 8.6]. If $t = 2$, then $(|\lambda|, |\mu|) = (2, 3)$ or $(3, 2)$. We may assume $(|\lambda|, |\mu|) = (2, 3)$. Then using Lemma 2.5, we can determine the graph pair G_P, G_T as in Figure 10. But a jumping number argument as in the first paragraph of the proof of [Goda and Teragaito 2005, Proposition 8.7] rules out this possibility. □

Lemma 7.2. $t \geq 4$ *is impossible.*

Proof. Assume $t \geq 4$. Since $|\alpha| > 0$, $G_T^+ \neq G_T$ and hence we have $|\lambda|, |\mu| \leq t$ by Lemma 2.4(7). The total number of edges of G_P is $\Delta pt/2 = 5t/2$, so $|\alpha| \geq t/2$.

Hence $|\alpha| = t/2$ or $t/2 + 1$ by Lemma 2.4(6). But $\alpha = t/2 + 1$ is impossible by [Teragaito 2006b, Lemma 8.12]. Thus $G_P = \Gamma_1(t/2, t, t)$ or $\Gamma_2(t/2, t, t)$. The latter is impossible by [Teragaito 2006b, Lemma 8.11], the former is possible only if $t = 4$

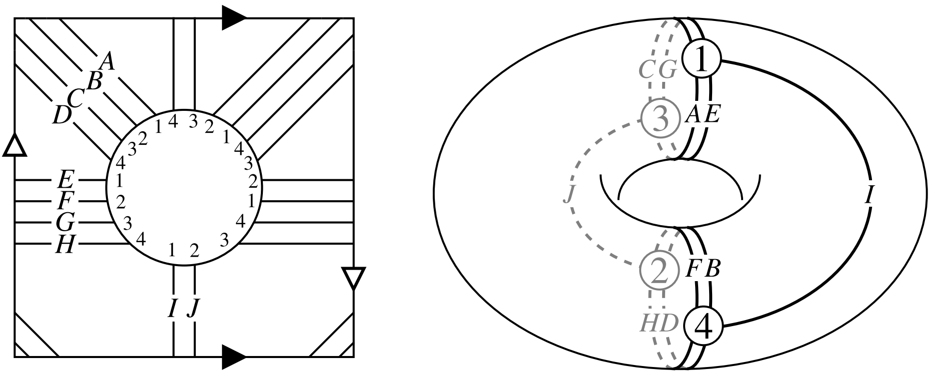


Figure 11. The graph pair G_P, G_T.

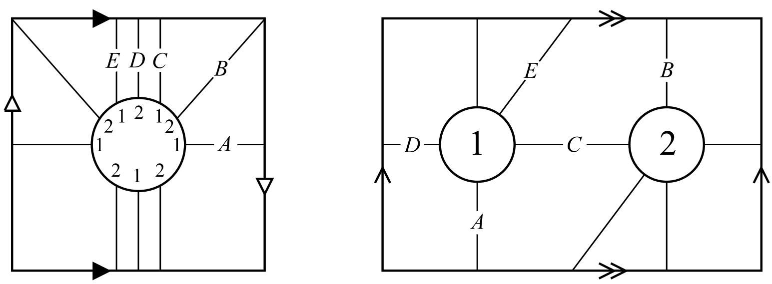


Figure 12. The graph pair G_P, G_T.

by [Teragaito 2006b, Lemma 8.10], and the graph pair G_P, G_T is determined as in Figure 11.

Let b_1, b_2 and b_3 be the bigon faces of G_P bounded by the edges $A \cup B$, $C \cup D$, and $I \cup J$, respectively. Let $V_{i,i+1}$ be the part of V_τ between v_i and v_{i+1} for each $i \in \{1, 2, 3, 4\}$. Then for each $j = 1, 2$, shrinking V_{12} and V_{34} to their cores in $b_j \cup b_3 \cup V_{12} \cup V_{34}$ gives a Möbius band B_j in $M(\tau)$ such that $\partial B_j \subset \widehat{T}$. Isotope B_1 so that $B_1 \cap B_2 = \partial B_1 = \partial B_2$. Then $B_1 \cup B_2$ is a Klein bottle in $M(\tau)$, contradicting our assumption that $M(\tau)$ does not contain a Klein bottle. □

Hence $t = 2$. The proof of [Goda and Teragaito 2005, Proposition 8.7] shows that the only two possibilities for G_P are $G_P \cong \Gamma_1(3, 1, 1)$ or $\Gamma_2(3, 2, 0)$.

Lemma 7.3. *If $G_P \cong \Gamma_1(3, 1, 1)$, then $M(\pi)$ is toroidal and $M = N(2, 2)$.*

Proof. Using Lemma 2.5, we can determine the graph pair G_P, G_T as in Figure 12. The jumping number is 1 and the edge correspondence is as shown in the figure. The graph pair G_S, G'_T obtained from G_P and G_T as in Section 6 is shown in Figure 13.

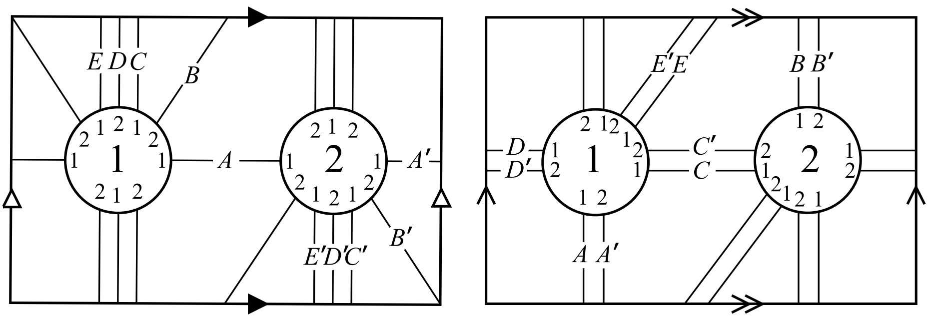


Figure 13. The graph pair G_S, G'_T.

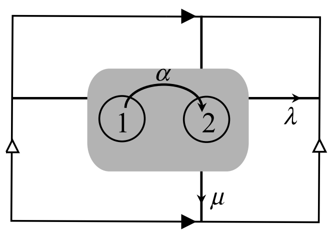


Figure 14. Generators for $\pi_1(Z)$.

Let $Z = M(\pi) - \mathrm{Int}(\eta(\hat{P}))$ and $V_{12} = V_\pi \cap Z$, and let f and g be the faces of G'_T bounded by the edges $A' \cup C \cup E'$ and $B \cup C' \cup E$, respectively. Compressing the genus 2 surface $\partial(\eta(\hat{P}) \cup V_\pi)$ along the disks f and g gives a 2-sphere in $M(\pi)$, which bounds a 3-ball. Hence $\pi_1(\eta(\hat{S} \cup V_{12} \cup f \cup g)) \cong \pi_1(Z)$.

As a base point, we take a disk containing the two vertices of G_S as shown in Figure 14. The group $\pi_1(Z)$ has three generators α, λ, and μ as in the figure, where α is represented by the core of V_{12}. The torus $\hat{S}$ gives a relation $\lambda\mu = \mu\lambda$ and the disks f and g give two relations $\lambda\alpha\mu\alpha^{-1}\mu\alpha = 1$ and $\mu\alpha\mu\alpha^{-1}\mu^{-1}\alpha^{-1} = 1$. Hence $\pi_1(Z)$ has the presentation

$$\langle \alpha, \lambda, \mu : \lambda\mu = \mu\lambda,\ \lambda\alpha\mu\alpha^{-1}\mu\alpha = 1,\ \mu\alpha\mu\alpha^{-1}\mu^{-1}\alpha^{-1} = 1 \rangle.$$

One sees that $\pi_1(Z) = \langle \alpha, \mu : \alpha\mu\alpha = \mu\alpha\mu \rangle$, which is isomorphic to the fundamental group of the trefoil knot exterior. This implies that Z is not a solid torus. Hence $\hat{S}$ is an essential torus in $M(\pi)$.

The graph pair G_S, G'_T is shown in [Gordon and Wu 2008, Figure 20.4] with the order reversed. By [ibid., Theorem 21.4], M is homeomorphic to M_{12} in the notation of that paper. They proved in [ibid., Lemma 22.2] that M_{12} is the double branched cover of the tangle in [ibid., Figure 22.12(b)], which is the tangle on the top left in Figure 15. Using isotopies and the Montesinos trick, one can see that $M = M_{12}$ is homeomorphic to the 3-manifold in the top right of Figure 15, where a

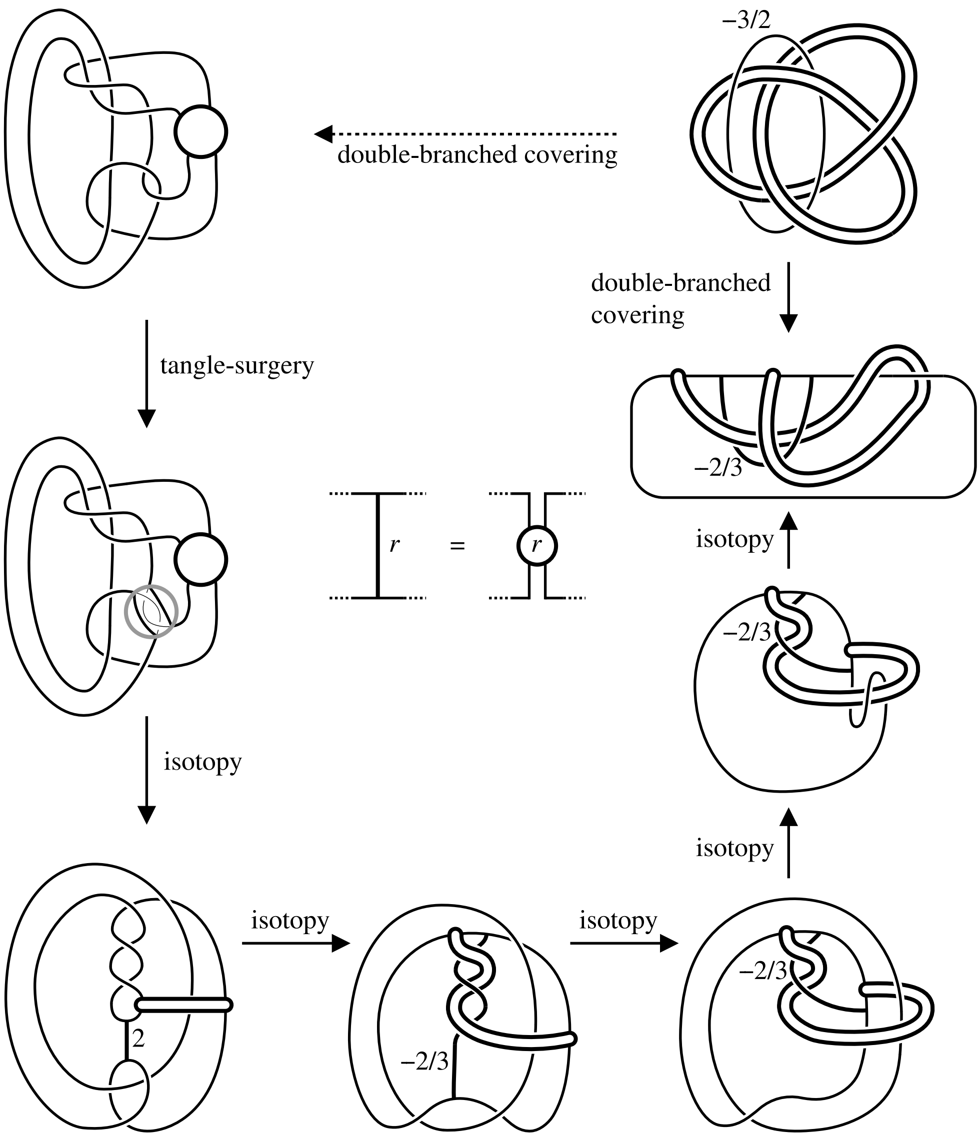


Figure 15. A surgery description for M_{12}.

thick arc with a rational number r represents a rational tangle of slope r as shown in the center of the figure. (See [Eudave-Muñoz 2002, Section 2] for the definition of rational tangles.) Figure 16 shows that $M = M_{12}$ is homeomorphic to $N(2, 2)$. See [Gompf and Stipsicz 1999, Chapter 5]. □

Let V be a solid torus and K a knot on ∂V that wraps around V in the longitudinal direction l times and in the meridional direction m times. Push K into

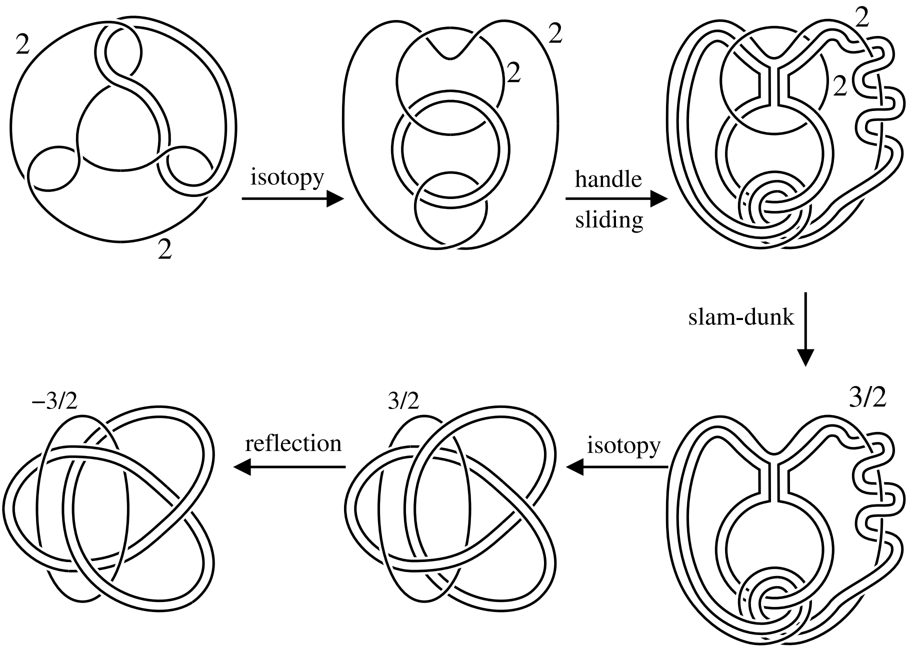


Figure 16. $N(2, 2)$ is homeomorphic to M_{12}.

the interior of V and remove its open regular neighborhood from V. The resulting manifold will be denoted by $C(l, m)$ and called a *cable space* of type (l, m).

Lemma 7.4. *If $G_P \cong \Gamma_2(3, 2, 0)$, then $M(\pi)$ is toroidal and $M = N(-4, \frac{1}{2}(2n-1))$ for some integer $n \neq 0, -1$.*

Proof. Note that $\Gamma_2(3, 2, 0) \cong \Gamma_2(3, 0, 2)$. Assume $G_P \cong \Gamma_2(3, 0, 2)$. Using Lemma 2.5, we can determine G_T as in Figure 12. Each face of G_T is a disk, so there is no circle component of $P \cap T$. There exists a nondisk face in G_P; this face is homeomorphic to a Möbius band. Let k be an orientation-reversing curve on the nondisk face.

Let $X = M - \text{Int}(\eta(k))$, $\hat{B} = \hat{P} - \text{Int}(\eta(k))$, $B' = \hat{P} \cap \eta(k)$ and $B = X \cap \hat{B}$. Then $\hat{B}$ and B' are Möbius bands and B is a once-punctured Möbius band. Since M is bounded by a single torus (see [Lee 2007, Theorem 1.3]), X is bounded by two tori ∂M and $\partial\eta(k)$. Let $T_0 = \partial M$, $T_1 = \partial\eta(k)$, and $\partial_i B = \partial B \cap T_i (i = 0, 1)$. Let $X(\pi) = X \cup V_\pi$ and $X(\tau) = X \cup V_\tau$.

Note that $\widehat{T}$ is essential in $X(\tau)$; it is incompressible in $X(\tau)$, since otherwise it would be compressible in $M(\tau)$, and it is not boundary-parallel in $X(\tau)$, since otherwise it would bound a solid torus in $M(\tau)$. Since G_P contains Scharlemann cycles, by Lemma 2.4(1), $\widehat{T}$ is separating in $M(\tau)$ (and hence in $X(\tau)$).

Claim. *X is hyperbolic.*

Proof. We first show that X is irreducible. On the contrary, suppose that X contains an essential sphere Q. Since M is irreducible, Q is separating in X. In particular, Q separates the two boundary components of X. By an isotopy of Q, we may assume that Q meets each of B and T transversely. We may also assume that Q meets each of B and T minimally among all essential spheres in X. Then Q and T are disjoint, since otherwise T would be compressible. But Q and B cannot be disjoint because B has one boundary component on each of T_0 and T_1. Since Q and T are disjoint, each component of $Q \cap B$ is parallel to $\partial_1 B$ in B. Compressing $\hat{P}$ along a disk component of $Q - B$ gives a projective plane in $M(\pi)$. This contradicts [Jin et al. 2003, Theorem 1.1]. Hence X is irreducible.

Each T_i for $i = 0, 1$ is incompressible in X, since otherwise after compression it would become a sphere bounding a 3-ball by the irreducibility of X, implying that X is a solid torus. Thus X is boundary-irreducible.

The manifold M, which is obtained from X by Dehn filling, is hyperbolic. Hence X cannot be Seifert fibered.

We only need to prove that X is atoroidal. Suppose that X contains an essential torus U. Since M is atoroidal and irreducible, U separates X into two components. Let X_0 and X_1 be the two components, where $T_i \subset X_i$ for $i = 0, 1$ and $X_1 \cup \eta(k)$ is a solid torus. We may assume that U was chosen so that X_0 contains no essential torus in its interior. We also assume that U intersects each of B and T transversely and minimally. Since M is orientable, each component of $U \cap B$ is parallel in B to either $\partial_0 B$ or $\partial_1 B$.

Suppose some components of $U \cap B$ are parallel to $\partial_0 B$. Let A ($\subset B$) be the annulus cut off by the outermost such component. Then A is contained in X_0 and intersects the tori T_0 and U. The boundary circle of A on U is essential, since otherwise X would be boundary-reducible. The frontier of $\eta(T_0 \cup A \cup U)$ is a torus in X_0. Since X_0 is irreducible and atoroidal, the torus bounds a solid torus in X_0. This implies that X_0 is a cable space, which contradicts the hyperbolicity of M.

Hence all the components of $U \cap B$ are parallel to $\partial_1 B$. The outermost component cuts off an annulus A' ($\subset B$), which lies in X_1. One boundary component of A' is $\partial_1 B$ and the other is an essential curve in U. Since $A' \cup B'$ is a Möbius band with boundary on U, $\eta(U \cup A' \cup B') = \eta(U \cup A') \cup \eta(k)$ is homeomorphic to the cable space $C(2, 1)$. One boundary component of $\eta(U \cup A' \cup B')$ is parallel to U and the other bounds a solid torus J in X_1 since otherwise either that component would be essential in M, contradicting the hyperbolicity of M, or it would compress into an essential sphere in X_1, contradicting the irreducibility of X. The core of A', which is a Seifert fiber of $\eta(U \cup A' \cup B')$, is homotopic to the core of J, since otherwise U would be an essential torus in M, contradicting the hyperbolicity of M again. This implies that $X_1 \cong \eta(U \cup A') \cup J \cong U \times I$, showing that U is boundary-parallel in X. This contradicts the choice of U. □

Neither $X(\pi)$ nor $X(\tau)$ is hyperbolic (the former contains a Möbius band $\hat{B}$ and the latter contains an essential torus $\widehat{T}$) and $\Delta(\pi, \tau) = 5$, so it follows from [Lee 2007, Theorem 1.1] that X is the exterior of the Whitehead sister link. Hence M is the result of a Dehn filling on the link exterior.

The results of exceptional Dehn fillings on the Whitehead sister link exterior are shown in [Martelli and Petronio 2006, Table A.1]. From the table, one sees that each of $X(\pi)$ and $X(\tau)$ contains a unique essential torus cutting it into the trefoil knot exterior and the cable space $C(2, 1)$. Let E and C denote the knot exterior and the cable space, respectively. Let $V = \eta(k)$ and let T_2 be the common boundary torus of E and C in $X(\pi)$. Then $X(\pi) = E \cup_{T_2} C$, $M(\pi) = E \cup_{T_2} (C \cup_{T_1} V)$, and $\partial C = T_1 \cup T_2$.

Claim. *T_2 is an essential torus in $M(\pi)$.*

Proof. Suppose that $C \cup V$ is a solid torus. (Otherwise, $T_2\,(= \partial(C \cup V))$ is an essential torus in $M(\pi)$.) Consider the curves on T_2. By an (r, s)-*curve*, we mean a curve on T_2 that wraps around the solid torus $C \cup V$ in the longitudinal direction r times and in the meridional direction s times. Then $C \cup V$ is a fibered solid torus whose regular fibers on T_2 are $(2, 1)$-curves.

Note that E and C are Seifert fiber spaces whose fibers intersect exactly once in their common boundary T_2. See [Martelli and Petronio 2006, Table A.1]. Suppose that an (a, b)-curve is a regular fiber of E. Then we have $a - 2b = 1$ and hence $a = 2b + 1$. This implies that $M(\pi)$ is a Seifert fiber space over the 2-sphere with three exceptional fibers of indices 2, 3, and $|2b + 1|$. (Note that E is a Seifert fiber space over the disk with two exceptional fibers of indices 2 and 3.) Such a Seifert fiber space does not contain a Klein bottle, which contradicts the assumption that $M(\pi)$ contains a Klein bottle. □

Since the Whitehead sister link exterior X has a self-homeomorphism interchanging its two boundary tori, we may assume that T_0 is the knotted boundary torus of X. Let $X(r_0, r_1)$ denote the 3-manifold obtained from X by performing a Dehn filling on T_i along slope r_i for each $i = 0, 1$. Partial Dehn fillings give $X(r_0) = X(r_0, \cdot)$ and $X(\cdot, r_1)$. Recall that M is obtained from X by performing a Dehn filling along the torus T_1, so $M = X(\cdot, r)$ for some $r \in \mathbb{Q} \cup \{1/0\}$.

By [ibid., Proposition 1.5] we have

$$N(-\tfrac{3}{2}, \alpha, \beta) = N\Big(-4, -\frac{\alpha+1}{\alpha+2}, -\beta - 3\Big) = N\Big(-\tfrac{3}{2}, -\frac{2\alpha+5}{\alpha+2}, -\frac{2\beta+5}{\beta+2}\Big).$$

Using this, one sees

$$N(-4, r, s) = N\Big(-4, -\frac{r+2}{r+1}, -\frac{s+2}{s+1}\Big).$$

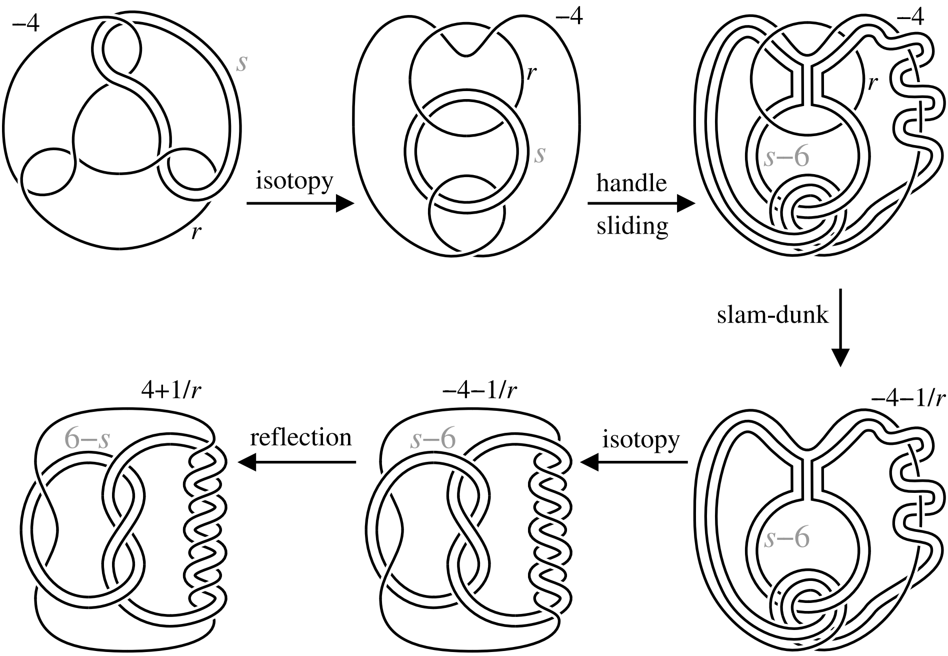


Figure 17. $N(-4)$ is homeomorphic to the Whitehead sister link exterior.

In particular, we have

$$N(-4, r) = N\Big(-4, -\frac{r+2}{r+1}\Big).$$

Figure 17 shows that

$$N(-4, r) = X(\,\cdot\,, 4+1/r) \quad \text{and} \quad N(-4, r, s) = X(6-s, 4+1/r).$$

It is known that the Whitehead sister link exterior X has exactly 5 exceptional slopes on any boundary component (see [ibid., Table A.1]). One can see that the set $\mathscr{E}$ of exceptional slopes of X on T_0 is $\mathscr{E} = \{1/0, 6, 7, 8, 9, 13/2\}$. Here, $\pi, \tau \in \mathscr{E}$ and $\{\pi, \tau\} = \{9, 13/2\}$. (Note that $\Delta(9, 13/2) = 5$.)

Assume $\pi = 13/2$. Then $\tau = 9$. Let $M = X(\,\cdot\,, 4+1/r) = N(-4, r)$ for some $r \in \mathbb{Q} \cup \{1/0\}$. Then

$$M(\pi) = X(\pi, r) = X(\tfrac{13}{2}, 4+1/r) = N(-4, r, -\tfrac{1}{2}) = N(-4, -\tfrac{1}{2}, r).$$

Since $M(\pi)$ contains a Klein bottle, $r = \frac{1}{2}(2n-1)$ for some integer $n \neq 0$. See the last row of [ibid., Table 3]. If $n = -1$, then

$$\begin{aligned} M(\tau) = X(\tau, 4+1/r) = X(9, 4+1/r) &= N(-4, r, -3) \\ &= N(-3, -4, r) = N(-3, -4, -\tfrac{3}{2}) \end{aligned}$$

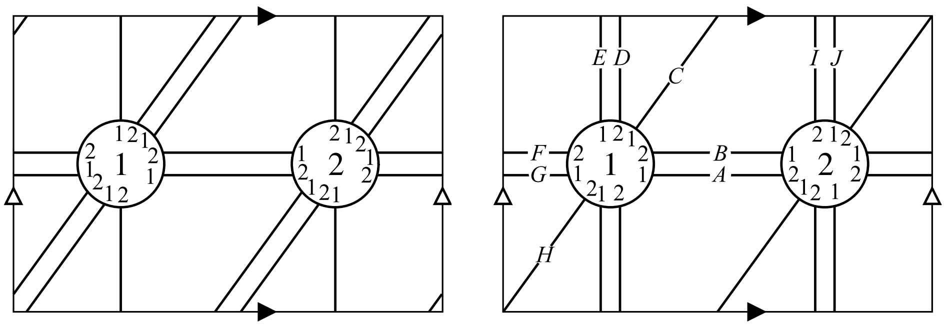


Figure 18. The graph G_T.

is not toroidal. See the last row for slope -3 in [ibid., Table 2]. We conclude that $M = N(-4, \frac{1}{2}(2n-1))$ for some integer $n \neq 0, -1$.

Assume $\pi = 9$. Then $\tau = 13/2$. Let $M = X(\,\cdot\,, 4+1/r) = N(-4, r)$ for some $r \in \mathbb{Q} \cup \{1/0\}$. Then

$$\begin{aligned} M(\pi) = X(\pi, 4+1/r) = X(9, 4+1/r) &= N(-4, r, -3) \\ &= N(-4, -3, r) = N(-4, -\tfrac{1}{2}, -\tfrac{r+2}{r+1}). \end{aligned}$$

Since $M(\pi)$ contains a Klein bottle, $-\frac{r+2}{r+1} = \frac{1}{2}(2n-1)$ for some integer $n \neq 0$. See the last row of [ibid., Table 3]. If $n = -1$, then

$$\begin{aligned} M(\tau) &= X(\tau, 4+1/r) = X(\tfrac{13}{2}, 4+1/r) = N(-4, r, -\tfrac{1}{2}) = N(-4, -\tfrac{1}{2}, r) \\ &= N(-4, -3, -\tfrac{r+2}{r+1}) = N(-4, -3, -\tfrac{3}{2}) = N(-3, -4, -\tfrac{3}{2}) \end{aligned}$$

is not toroidal. See the last row for slope -3 in [ibid., Table 2]. Hence $M = N(-4, r) = N(-4, -\frac{r+2}{r+1}) = N(-4, \frac{1}{2}(2n-1))$ for some integer $n \neq 0, -1$. □

8. The case $p \geq 2$ and $t = 2$

In this case, the argument in [Goda and Teragaito 2005, Section 9] shows the following.

- $p = 2$;
- nonloop edges of G_T are negative; and
- G_T is one of the graphs in Figure 18.

For the first graph in Figure 18, the argument in the second paragraph of the proof of [Teragaito 2006b, Lemma 7.4] shows that $M(\tau)$ contains a Klein bottle, contradicting our assumption. Hence G_T is the second graph in Figure 18. Then the graph G_P is uniquely determined as shown in Figure 19. See the third paragraph of the proof of [ibid., Lemma 7.4]. We obtain a graph pair G_S, G'_T from G_P and G_T

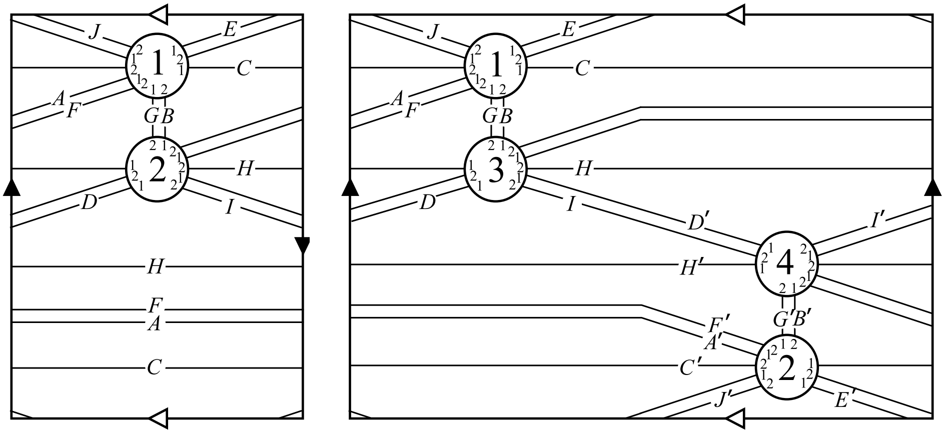


Figure 19. The graphs G_P and G_S.

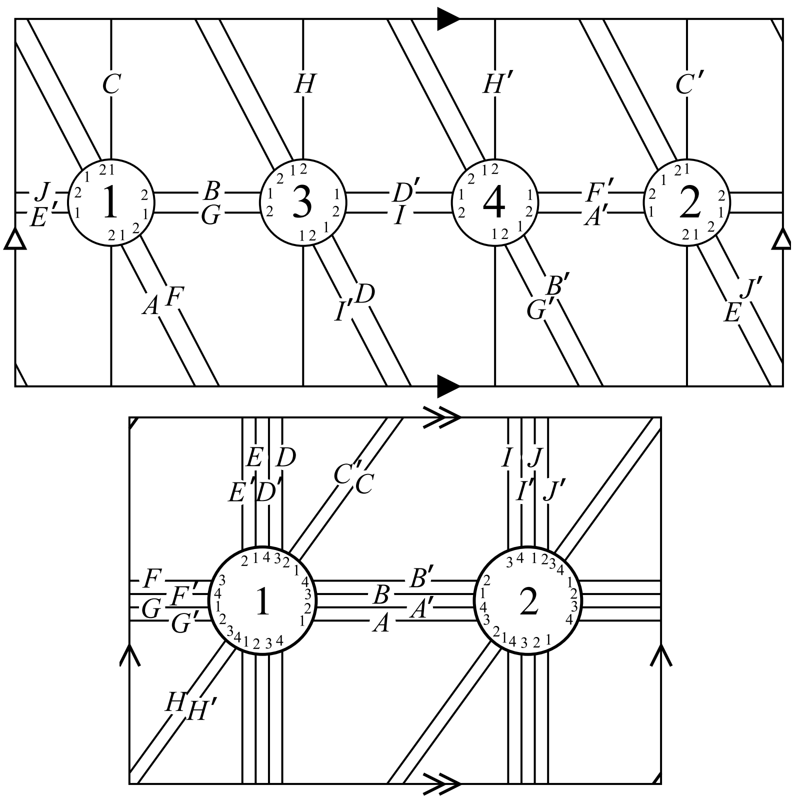


Figure 20. The graph pair G_S, G'_T.

as in Section 6. See Figure 19 for G_S. See Figure 20 for the edge correspondence between the graphs G_S and G'_T.

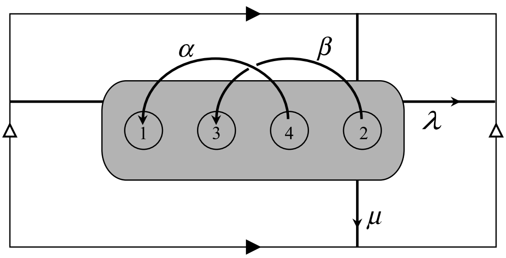


Figure 21. Generators for $\pi_1(Z)$.

Let $Z = M(\pi) - \operatorname{Int}\eta(\hat{P})$, and let f, g, and h be the faces of G'_T bounded by the edges $A' \cup B$, $D' \cup E$ and $A \cup C' \cup D$, respectively. The group $\pi_1(Z)$ has four generators $\alpha, \beta, \lambda, \mu$ as shown in Figure 21, where α and β are represented by the cores of the two 1-handles V_{41}, V_{23} in $V_\pi \cap Z$. The three disks f, g, and h give three relations $\alpha\beta^{-1} = 1$, $\mu\lambda\beta\alpha = 1$, and $\alpha\mu\beta^{-1}\mu\beta\mu = 1$. Hence $\pi_1(Z)$ has the presentation

$$\langle \alpha, \beta, \lambda, \mu : \lambda\mu = \mu\lambda, \alpha\beta^{-1} = 1, \mu\lambda\beta\alpha = 1, \alpha\mu\beta^{-1}\mu\beta\mu = 1\rangle.$$

Since $\alpha = \beta$ and $\lambda = \mu^{-1}\alpha^{-1}\beta^{-1} = \mu^{-1}\alpha^{-2}$, we have

$$\pi_1(Z) = \langle \alpha, \mu : \alpha^2\mu = \mu\alpha^2, \alpha\mu\alpha^{-1}\mu\alpha\mu = 1\rangle.$$

Letting $\gamma = \mu\alpha$, one sees that $\pi_1(Z) = \langle \alpha, \gamma : \alpha^2 = \gamma^3\rangle$, which implies that Z is not a solid torus. Hence $\hat{S}$ is an essential torus and $M(\pi)$ is toroidal.

The graph pair G_S, G'_T is shown in [Gordon and Wu 2008, Figure 16.6]. By [ibid., Theorem 21.4], M is homeomorphic to M_7 in the notation of that paper. The double branched cover of the tangle on the top left in Figure 22 is M_7 (see [ibid., Lemma 22.2]). The figure shows that $M_7 = N(-5/3, -5/3)$.

Summarizing the results in this section, we obtain the following.

Lemma 8.1. *If $p \geq 2$ and $t = 2$, then $M(\pi)$ is toroidal and $M = N(-5/3, -5/3)$.*

References

[Eudave-Muñoz 2002] M. Eudave-Muñoz, "On hyperbolic knots with Seifert fibered Dehn surgeries", *Topology Appl.* **121**:1-2 (2002), 119–141. MR 2003c:57005 Zbl 1009.57010

[Goda and Teragaito 2005] H. Goda and M. Teragaito, "On hyperbolic 3-manifolds realizing the maximal distance between toroidal Dehn fillings", *Algebr. Geom. Topol.* **5** (2005), 463–507. MR 2006i:57009 Zbl 1082.57011

[Gompf and Stipsicz 1999] R. E. Gompf and A. I. Stipsicz, 4-*manifolds and Kirby calculus*, Graduate Studies in Mathematics **20**, American Mathematical Society, Providence, RI, 1999. MR 2000h:57038 Zbl 0933.57020

[Gordon 1998] C. M. Gordon, "Boundary slopes of punctured tori in 3-manifolds", *Trans. Amer. Math. Soc.* **350**:5 (1998), 1713–1790. MR 98h:57032 Zbl 0896.57011

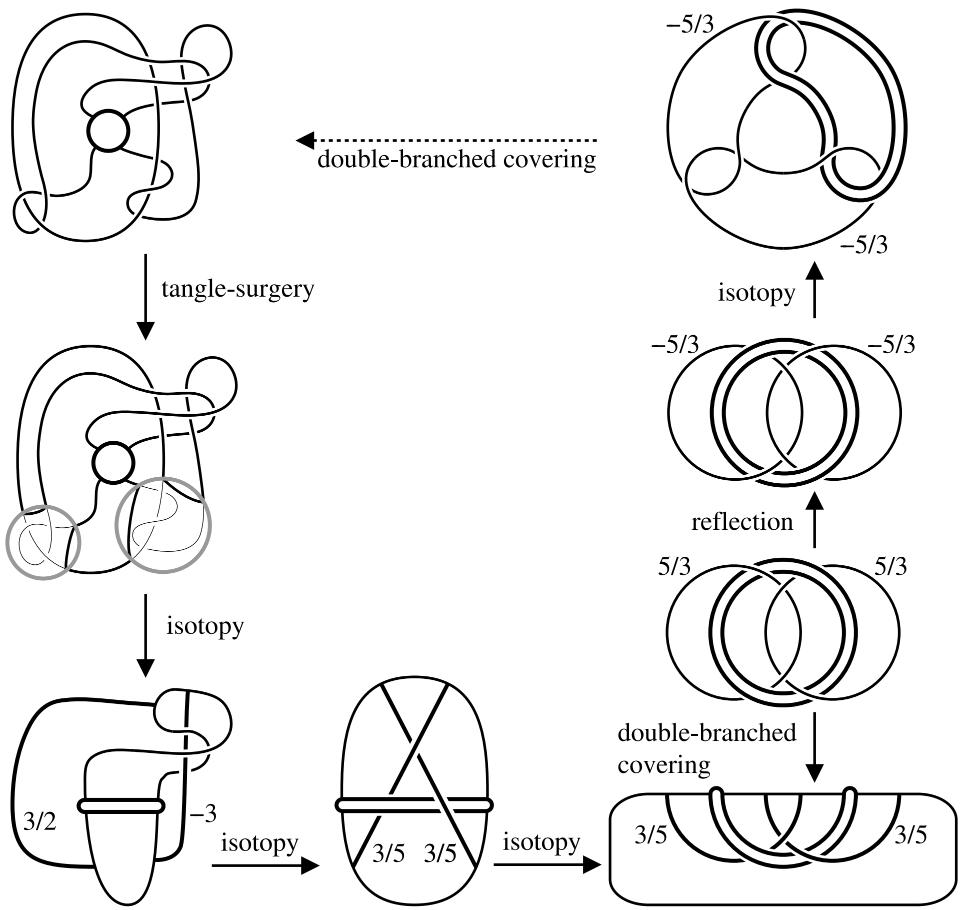


Figure 22. $N(-5/3, -5/3)$ is homeomorphic to M_7.

[Gordon 1999] C. M. Gordon, "Toroidal Dehn surgeries on knots in lens spaces", *Math. Proc. Cambridge Philos. Soc.* **125**:3 (1999), 433–440. MR 99h:57006 Zbl 0926.57014

[Gordon and Luecke 1995] C. M. Gordon and J. Luecke, "Dehn surgeries on knots creating essential tori, I", *Comm. Anal. Geom.* **3**:3-4 (1995), 597–644. MR 96k:57003 Zbl 0865.57015

[Gordon and Wu 1999] C. M. Gordon and Y.-Q. Wu, "Toroidal and annular Dehn fillings", *Proc. London Math. Soc.* (3) **78**:3 (1999), 662–700. MR 2000b:57029 Zbl 1024.57020

[Gordon and Wu 2008] C. M. Gordon and Y.-Q. Wu, *Toroidal Dehn fillings on hyperbolic 3-manifolds*, Mem. Amer. Math. Soc. **909**, American Mathematical Society, Providence, RI, 2008. MR 2009c:57036 Zbl 1166.57014

[Hayashi and Motegi 1997] C. Hayashi and K. Motegi, "Only single twists on unknots can produce composite knots", *Trans. Amer. Math. Soc.* **349**:11 (1997), 4465–4479. MR 98b:57010b Zbl 0883.57005

[Jin et al. 2003] G. T. Jin, S. Lee, S. Oh, and M. Teragaito, "P^2-reducing and toroidal Dehn fillings", *Math. Proc. Cambridge Philos. Soc.* **134**:2 (2003), 271–288. MR 2004c:57035 Zbl 1028.57011

[Lee 2006] S. Lee, "Dehn fillings yielding Klein bottles", *Int. Math. Res. Not.* **2006** (2006), Art. ID 24253. MR 2007b:57038 Zbl 1106.57016

[Lee 2007] S. Lee, "Exceptional Dehn fillings on hyperbolic 3-manifolds with at least two boundary components", *Topology* **46**:5 (2007), 437–468. MR 2008h:57032 Zbl 1123.57011

[Lee and Teragaito 2008] S. Lee and M. Teragaito, "Boundary structure of hyperbolic 3-manifolds admitting annular and toroidal fillings at large distance", *Canad. J. Math.* **60**:1 (2008), 164–188. MR 2009a:57029 Zbl 1136.57007

[Martelli and Petronio 2006] B. Martelli and C. Petronio, "Dehn filling of the "magic" 3-manifold", *Comm. Anal. Geom.* **14**:5 (2006), 969–1026. MR 2007k:57042 Zbl 1118.57018

[Oh 1997] S. Oh, "Reducible and toroidal 3-manifolds obtained by Dehn fillings", *Topology Appl.* **75**:1 (1997), 93–104. MR 98a:57027 Zbl 0870.57008

[Teragaito 2000] M. Teragaito, "Toroidal Dehn fillings and lens spaces containing Klein bottles", manuscript, 2000, available at http://home.hiroshima-u.ac.jp/teragai/kb-tor.ps.gz.

[Teragaito 2006a] M. Teragaito, "Distance between toroidal surgeries on hyperbolic knots in the 3-sphere", *Trans. Amer. Math. Soc.* **358**:3 (2006), 1051–1075. MR 2006h:57005 Zbl 1089.57007

[Teragaito 2006b] M. Teragaito, "Toroidal Dehn fillings on large hyperbolic 3-manifolds", *Comm. Anal. Geom.* **14**:3 (2006), 565–601. MR 2007f:57045

[Valdez-Sánchez 2007] L. G. Valdez-Sánchez, "Toroidal and Klein bottle boundary slopes", *Topology Appl.* **154**:3 (2007), 584–603. MR 2007j:57025 Zbl 1113.57002

[Wu 1998] Y.-Q. Wu, "Dehn fillings producing reducible manifolds and toroidal manifolds", *Topology* **37**:1 (1998), 95–108. MR 98j:57033 Zbl 0886.57012

Received May 11, 2009.

SANGYOP LEE
DEPARTMENT OF MATHEMATICS
CHUNG-ANG UNIVERSITY
221 HEUKSEOK-DONG, DONGJAK-GU
SEOUL 156-756
SOUTH KOREA
sylee@cau.ac.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

REPRESENTATIONS OF THE TWO-FOLD CENTRAL EXTENSION OF $SL_2(\mathbb{Q}_2)$

HUNG YEAN LOKE AND GORDAN SAVIN

We define a notion of pseudospherical type for smooth representations of the nontrivial two fold central extension of $SL_2(\mathbb{Q}_2)$. We describe completely the irreducible representations that contain the pseudospherical type. We relate our results to Kohnen's plus and minus spaces of classical modular forms of half integral weight.

1. Introduction

Let $\mathbb{Q}$ be the field of rational numbers. For every place v of $\mathbb{Q}$, let $\mathbb{Q}_v$ denote the corresponding local field. Then $\mathbb{Q}_v = \mathbb{R}$ or $\mathbb{Q}_p$ for a prime p. The group $SL_2(\mathbb{Q}_v)$ has a nontrivial two-fold central extension

$$1 \to \mu_2 \to G(\mathbb{Q}_v) \to SL_2(\mathbb{Q}_v) \to 1, \tag{1}$$

where $\mu_2 = \{\pm 1\}$. Recall that an irreducible representation of $G(\mathbb{Q}_v)$ is called genuine if the central subgroup μ_2 acts faithfully on it. Gelbart's book [1976] contains a basic theory of genuine representations of $G(\mathbb{R})$ and $G(\mathbb{Q}_p)$ for $p \neq 2$. Our intent is to develop a theory in the case of $G(\mathbb{Q}_2)$. The main difference between $G(\mathbb{Q}_2)$ and $G(\mathbb{Q}_p)$ for $p \neq 2$ lies in the fact that the central extension splits over $SL_2(\mathbb{Z}_p)$ when $p \neq 2$. In particular, we have a subgroup $K_p \subseteq G(\mathbb{Q}_p)$ isomorphic to $SL_2(\mathbb{Z}_p)$ under the natural projection from $G(\mathbb{Q}_p)$ to $SL_2(\mathbb{Q}_p)$ for every $p \neq 2$. A genuine representation π of $G(\mathbb{Q}_p)$ is called *unramified* if it contains a nonzero K_p-fixed vector.

Assume now that $p = 2$. Let K denote the full inverse image of $SL_2(\mathbb{Z}_2)$ in $G(\mathbb{Q}_2)$. In this case the central extension splits over a smaller subgroup. More precisely, we have a subgroup $K_1(4) \subseteq K$ isomorphic to the subgroup of $SL_2(\mathbb{Z}_2)$ given by the congruence

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod 4$$

MSC2000: 11F70, 22E50, 22E55.
Keywords: modular forms, Shimura correspondence, Hecke algebra, pseudospherical representations.
Savin is supported by NSF grants DMS-0551846 and DMS-0852429.

In this paper we completely describe genuine irreducible representations of $G(\mathbb{Q}_2)$ containing nonzero $K_1(4)$-fixed vectors. More precisely, in Section 3, we describe a Hecke algebra $H(\gamma)$ that captures the structure of all representations generated by $K_1(4)$-fixed vectors and with a fixed central character γ. In Section 4, we show that $H(\gamma)$ is isomorphic to the Iwahori–Matsumoto Hecke algebra for $\mathrm{PGL}_2(\mathbb{Q}_2)$. In this way we get a correspondence between (some) representations of $G(\mathbb{Q}_2)$ and representations of $\mathrm{PGL}_2(\mathbb{Q}_2)$. We call this correspondence a local Shimura correspondence.

In Section 5, we show that the compact group K has exactly two irreducible genuine representations, with the fixed central character γ, containing nonzero $K_1(4)$-fixed vectors. These representations are denoted by $V(2)$ and $V(-1)$ and have dimensions 2 and 4, respectively. We show that a representation π of $G(\mathbb{Q}_2)$ has $V(2)$ as a K-type if and only if it corresponds to an unramified representation of $\mathrm{PGL}_2(\mathbb{Q}_2)$, by the local Shimura correspondence. Thus, it is natural to define unramified representations of $G(\mathbb{Q}_2)$ to be those that contain $V(2)$ as a K-type, and we call $V(2)$ a pseudospherical type.

We should point out that the center of $G(\mathbb{Q}_2)$ is a cyclic group of order 4. Thus, we have two different genuine central characters γ and two classes of unramified representations. This is analogous to the case of the real group $G(\mathbb{R})$, where the weights $-1/2$ and $1/2$ are called pseudospherical types.

We apply our local results in a global setting in Section 8. Let $\mathbb{A}$ be the ring of adeles, and let $G(\mathbb{A})$ be the two-fold cover of $\mathrm{SL}_2(\mathbb{A})$. Let $r > 1$ be an odd integer. Let $\pi = \otimes\pi_v$ be a genuine cuspidal automorphic representation such that

- π_∞ is a holomorphic discrete series representation with the lowest weight $r/2$,
- π_p is unramified for all $p \neq 2$, and
- π_2 contains a nonzero $K_1(4)$-fixed vector.

Every such π corresponds to a Hecke eigenspace in $S_{r/2}(\Gamma_0(4))$, the space of cuspidal modular forms of weight $r/2$. Roughly speaking, a function $f = \bigotimes f_v$ in π gives naturally a modular form in $S_{r/2}(\Gamma_0(4))$ if f_∞ is a lowest weight vector in π_∞, f_p is K_p-fixed and f_2 is $K_1(4)$-fixed. Since the space of $K_1(4)$-fixed vectors in π_2 is two-dimensional, unless π_2 is a Steinberg representation, the cuspidal automorphic representation π gives rise to a two-dimensional Hecke eigenspace in $S_{r/2}(\Gamma_0(4))$. We can pick a line in this subspace by taking f_2 to be in the K-type isomorphic to $V(2)$. In this way we get a representation-theoretic description of Kohnen's plus space $S^+_{r/2}(\Gamma_0(4))$ [1980]. We also obtain that "new forms" in Kohnen's minus space $S^-_{r/2}(\Gamma_0(4))$ correspond to automorphic representations π, where the local component π_2 is a Steinberg representation. Finally, we show that the global Shimura correspondence is compatible with our local Shimura correspondence at the place $p = 2$.

Representations of $G(\mathbb{Q}_2)$ have been studied in great detail by Waldspurger [1980; 1981; 1991]. However, his approach does not involve the Hecke algebra $H(\gamma)$. Furthermore, a representation-theoretic description of Kohnen's plus space has already been given in [Baruch and Mao 2007]. That approach relies heavily on the mentioned results of Waldspurger, where needed local results are hard to extract.

2. Double cover of $\mathrm{SL}_2(\mathbb{Q}_v)$

We now describe the double cover $G(\mathbb{Q}_v)$ in (1). A section $\boldsymbol{s} : \mathrm{SL}_2(\mathbb{Q}_v) \to G(\mathbb{Q}_v)$ allows us to identify $G(\mathbb{Q}_v)$ with the set $\mathrm{SL}_2(\mathbb{Q}_v) \times \mu_2$, with group law

$$(g_1, \epsilon_1)(g_2, \epsilon_2) = (g_1 g_2, \epsilon_1 \epsilon_2 \sigma_v(g_1, g_2)),$$

where $\sigma_v(g_1, g_2)$ is a cocycle that depends on $\boldsymbol{s}$. Following [Gelbart 1976], we make the following choice of the cocycle σ_v. Let $(\cdot\,,\cdot)_v$ be the Hilbert symbol over $\mathbb{Q}_v$. For $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Q}_v)$, we define

$$x(g) = \begin{cases} c & \text{if } c \neq 0, \\ d & \text{if } c = 0 \end{cases} \quad \text{and} \quad s(g) = \begin{cases} (c, d)_v & \text{if } v \text{ is a finite prime, } cd \neq 0 \\ & \qquad\qquad \text{and } \mathrm{ord}(c) \text{ is odd,} \\ 1 & \text{otherwise.} \end{cases}$$

Then $\sigma_v(g_1, g_2) = (x(g_1 g_2) x(g_1), x(g_1 g_2) x(g_2))_v s(g_1) s(g_2) s(g_1 g_2)$.

An advantage of this particular section is that $K_p = \boldsymbol{s}(\mathrm{SL}_2(\mathbb{Z}_p))$ is a subgroup in $G(\mathbb{Q}_p)$ if $p \neq 2$. If $p = 2$, we define

$$K_1(4) = \left\{ \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, 1 \right) \in \mathrm{SL}_2(\mathbb{Z}_2) \times \{\pm 1\} : a \in 1 + 4\mathbb{Z}_2,\, c \in 4\mathbb{Z}_2 \right\}.$$

By [Gelbart 1976, Proposition 2.14], $K_1(4)$ is a compact subgroup of $G(\mathbb{Q}_2)$.

A smooth representation of $G(\mathbb{Q}_v)$ is called *genuine* if μ_2 acts nontrivially. If p is an odd prime number, a smooth genuine representation of $G(\mathbb{Q}_p)$ is called *unramified* if it contains a vector fixed by K_p. A vector fixed by K_p is called a *spherical* vector.

If $p = 2$, a smooth genuine representation is called *tamely ramified* if it contains a vector fixed by $K_1(4)$. Unfortunately $\mathrm{SL}_2(\mathbb{Z}_2)$ does not split in $G(\mathbb{Q}_2)$, so we cannot define spherical vectors in the same manner as those for odd primes. The objective of this paper is to motivate and define spherical vectors of genuine representations of $G(\mathbb{Q}_2)$.

We set up some notation for later. For $u \in \mathbb{Q}_v$ and $t \in \mathbb{Q}_v^\times$, we define these elements in $\mathrm{SL}_2(\mathbb{Q}_v)$:

$$\underline{x}(u) = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}, \qquad \underline{y}(u) = \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix}, \qquad \underline{w}(t) = \begin{pmatrix} 0 & t \\ -t^{-1} & 0 \end{pmatrix}, \qquad \underline{h}(t) = \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix}.$$

Let $x(u) = \boldsymbol{s}(\underline{x}(u))$, $y(u) = \boldsymbol{s}(\underline{y}(u))$, $w(t) = \boldsymbol{s}(\underline{w}(t))$ and $h(t) = \boldsymbol{s}(\underline{h}(t))$ in $G(\mathbb{Q}_v)$. Note that $h(t)h(s) = h(ts)(t, s)_v$. Let $N = \{x(u) : u \in \mathbb{Q}_v\}$, $\overline{N} = \{y(u) : u \in \mathbb{Q}_v\}$ and T be the subgroup of G generated by elements $h(t)$.

3. Hecke algebra at $p = 2$

We fix $p = 2$ from Sections 3 through 6. We will denote $G(\mathbb{Q}_2)$ by G and $K_1(4)$ by K_1. The objective of these sections is to classify genuine representations of G containing a nonzero vector fixed by K_1.

Let M be the center of G. It is a cyclic group of order 4 generated by $h(-1)$. (Note that $h(-1)h(-1) = (-1, -1)_2 = -1 \in \mu_2$.) Thus, a genuine central character γ is determined by its value on $h(-1)$; this value is a fourth root of 1. Let K and K_0 be the open compact subgroups in G equal to the inverse images of $\mathrm{SL}_2(\mathbb{Z}_2)$ and

$$\left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}_2) : c \in 4\mathbb{Z}_2\right\}$$

respectively. Let $K(4) \subset K_1$ denote the principal congruence subgroup. It is the image under the section $\boldsymbol{s}$ of the subgroup of $\mathrm{SL}_2(\mathbb{Z}_2)$ consisting of matrices congruent to 1 modulo 4. We have $K \supset K_0 \supset K_1 \supset K(4)$ and $K_0 = M \times K_1$. We extend the central character γ to K_0, so that it is trivial on K_1. Given a smooth representation (π, V) of G, we let

$$V^{K_0,\gamma} := \{v \in V : \pi(k_0)v = \gamma(k_0)v \text{ for all } k_0 \in K_0\}.$$

Let $\mathcal{R}(G, \gamma)$ denote the category of admissible smooth (necessarily genuine) representations V of G such that $V^{K_0,\gamma}$ generates V as a G-module.

Next we define the corresponding Hecke algebra. Let $C_c(G)$ denote the set of locally constant, compactly supported functions on G. Let

$$H(\gamma) = \{f : C_c(G) : f(k_0 g k_0') = \overline{\gamma}(k_0) f(g) \overline{\gamma}(k_0') \text{ for all } k_0, k_0' \in K_0\}.$$

For $f_1, f_2 \in H(\gamma)$, we define

$$f_1 \cdot f_2(g_0) = \int_G f_1(g) f_2(g^{-1} g_0) dg = \int_G f_1(g_0 g) f_2(g^{-1}) dg,$$

where dg is the Haar measure on G such that the measure of K_0 is 1. Then $H(\gamma)$ is a $\mathbb{C}$-algebra. For $f \in H(\gamma)$ and $v \in V$, we have

$$\pi(f)v = \int_G f(g)\pi(g)v \, dg \in V^{K_0,\gamma}.$$

In this way $V^{K_0,\gamma}$ is a left $H(\gamma)$-module. Let $\mathcal{R}(H(\gamma))$ denote the category of finite-dimensional left $H(\gamma)$-modules. We have a functor $A : \mathcal{R}(G, \gamma) \to \mathcal{R}(H(\gamma))$

given by $V \mapsto V^{K_0,\gamma}$. Since the group K_0 has a triangular decomposition

$$K_0 = (K_0 \cap \overline{N})(K_0 \cap T)(K_0 \cap N),$$

the functor A is an equivalence of categories. This follows, in essence, from [Casselman 1995, Corollary 3.3.6]; see also [Borel 1976] and [Bernšteĭn and Zelevinskiĭ 1976, Theorem 4.2].

Our immediate goal is to understand the structure of $H(\gamma)$. The character γ of the center M extends to a character γ of T that is trivial on $K_1 \cap T$, and $\gamma(h(2^n)) = 1$ for all $n \in \mathbb{Z}$. Let us abbreviate $\gamma(t) = \gamma(h(t))$. We define $\zeta = (1 + \gamma(-1))/\sqrt{2}$. Note that ζ is a primitive 8-th root of 1. The character γ of T is invariant under conjugation by $w = w(1)$. We can now extend the character γ from T to the normalizer $N_G(T)$ by defining $\gamma(w) = \zeta$.

We define some functions in $H(\gamma)$. For g in $N_G(T)$, we set X_g to be the function supported on $K_0 g K_0$ such that

$$X_g(k_0 g k_0') = \overline{\gamma}(k_0)\overline{\gamma}(g)\overline{\gamma}(k_0') \quad \text{for all } k_0, k_0' \in K_0.$$

Note that this definition depends only on the image of g in the affine Weyl group $W_a := N_G(T)/(T \cap K_0)$.

Proposition 1. *Functions X_g for g in W_a form a basis of $H(\gamma)$.*

Proof. We need first to determine the K_0-double cosets in G. This can be easily determined in $SL_2(\mathbb{Q}_2)$ using row-column reduction. In addition to $h(2^n)$ and $w(2^{-n})$ the double coset representatives are

$$y(2),\quad h(2^n)y(2),\quad y(2)h(2^{-n}),\quad y(2)w(2^{-n}),\quad w(2^{-n})y(2),\quad y(2)w(2^{-n})y(2),$$

where $n \geq 1$ in all cases. We claim that the Hecke algebra is not supported on these cosets.

Lemma 2. *The commutator of $x(2)$ and $y(2)$ modulo the principal congruence subgroup $K(4)$ is equal to $-1 \in \mu_2$.*

Proof. This can be easily checked using the multiplication rule. It also follows from applying [Stein 1973, Corollary 2.9] to the ring $A = \mathbb{Z}/4\mathbb{Z}$, □

Now we can easily finish the proof of proposition. Indeed if f is in $H(\gamma)$, then

$$f(y(2)) = f(y(2)x(2)) = -f(x(2)y(2)) = -f(y(2))$$

by the lemma above. This implies that f must vanish on $y(2)$. Other cases are dealt with in the same manner. □

Let $\ell : N_G(T) \to \mathbb{Z}$ be defined by $\ell(g) = \log_2(n)$, where n is the number of left (or right) K_0-cosets in the double coset $K_0 g K_0$. In other words, the volume of $K_0 g K_0$ is $2^{\ell(g)}$. For example, $w(2^{-1})$ normalizes K_0, so $\ell(w(2^{-1})) = 1$.

Proposition 3. *For every integer n, we have*

$$\ell(h(2^n)) = 2|n| \quad \textit{and} \quad \ell(w(2^{-n})) = 2|1-n|.$$

More precisely, we have the following decompositions of double cosets:

(i) *If* $n \geq 0$,

$$K_0 h(2^n) K_0 = \bigcup_{u \in \mathbb{Z}/2^{2n}\mathbb{Z}} x(u) h(2^n) K_0 = \bigcup_{u \in \mathbb{Z}/2^{2n}\mathbb{Z}} K_0 h(2^n) y(4u).$$

(ii) *If* $n \geq 1$,

$$K_0 h(2^{-n}) K_0 = \bigcup_{u \in \mathbb{Z}/2^{2n}\mathbb{Z}} y(4u) h(2^{-n}) K_0 = \bigcup_{u \in \mathbb{Z}/2^{2n}\mathbb{Z}} K_0 h(2^{-n}) x(u).$$

(iii) *If* $n \geq 0$,

$$K_0 w(2^n) K_0 = \bigcup_{u \in \mathbb{Z}/2^{2n+2}\mathbb{Z}} x(u) w(2^n) K_0 = \bigcup_{u \in \mathbb{Z}/2^{2n+2}\mathbb{Z}} K_0 w(2^n) x(u).$$

(iv) *If* $n \geq 1$,

$$K_0 w(2^{-n}) K_0 = \bigcup_{u \in \mathbb{Z}/2^{2n-2}\mathbb{Z}} y(4u) w(2^{-n}) K_0 = \bigcup_{u \in \mathbb{Z}/2^{2n-2}\mathbb{Z}} K_0 w(2^{-n}) y(4u).$$

Proof. This follows easily from the decomposition $K_0 = (K_0 \cap \overline{N})(K_0 \cap T)(K_0 \cap N)$. Details are left to the reader. □

We record the following tautological lemma:

Lemma 4. *Let g_1 and g_2 be two elements in $N_G(T)$. If $\ell(g_1 g_2) = \ell(g_1) + \ell(g_2)$ then $X_{g_1} \cdot X_{g_2} = X_{g_1 g_2}$.*

Let $T_n = X_{h(2^n)}$ and $U_n = X_{w(2^{-n})}$.

Proposition 5. *Let $T_w = \sqrt{1/2}\, U_0$. We have the following identities, where m, n are any integers unless further specified.*

(i) $(T_w + 1)(T_w - 2) = 0$.

(ii) $U_1 \cdot U_1 = 1$.

(iii) *If $m, n \geq 0$ or $m, n \leq 0$, then $T_m \cdot T_n = T_{m+n}$.*

(iv) $U_1 \cdot T_n = U_{n+1}$ *and* $T_n \cdot U_1 = U_{1-n}$.

(v) $U_1 \cdot U_n = T_{n-1}$ *and* $U_n \cdot U_1 = T_{1-n}$.

Proof. All statements except for (i) follow from Lemma 4. For (i), we need to show $T_w^2 = T_w \cdot T_w = T_w + 2$. Since T_w^2 is supported in K, this is equivalent to $T_w^2(1) = 2$

and $T_w^2(w(1)) = T_w(w(1))$. Suppose $f_1, f_2 \in H(\gamma)$, where f_1 is supported on $K_0 r K_0 = \bigsqcup_{i=1}^{s} r_i K_0$ (disjoint union). Then

$$f_1 \cdot f_2(g) = \sum_{i=1}^{s} f_1(r_i) f_2(r_i^{-1} g).$$

We can apply this observation to $f_1 = f_2 = T_w$. Proposition 3(iii) with $n = 0$ gives a decomposition of $K_0 w(1) K_0$ into single cosets. Hence

$$T_w^2(g) = \sum_{u \bmod 4} T_w(x(u)w(1)) \cdot T_w(w(-1)x(-u)g).$$

If $g = 1$, this gives $T_w^2(1) = 4T_w(w(1)) \cdot T_w(w(-1))$. Since $T_w(w(1)) = 2^{-1/2}\bar{\zeta}$ and $T_w(w(-1)) = 2^{-1/2}\zeta$, we obtain that $T_w^2(1) = 2$. If $g = w(1)$, then

$$T_w^2(w(1)) = T_w(w(1)) \sum_{u \bmod 4} T_w(y(u)).$$

If $u = 0$ or 2, then $y(u)$ is not in $K_0 w(1) K_0$ and $T_w(y(u)) = 0$. If $u = \pm 1$, then $y(u) = x(u)w(-u)x(u)$, and we can rewrite

$$T_w^2(w(1)) = T_w(w(1))[T_w(w(1)) + T_w(w(-1))] = T_w(w(1)). \qquad \square$$

Here is the main result of this section.

Theorem 6. *The Hecke algebra $H(\gamma)$ is generated by T_w and U_1 as an abstract $\mathbb{C}$-algebra modulo the relations*

(a) $(T_w - 2)(T_w + 1) = 0$ *and*

(b) $U_1^2 = 1$.

Proof. Suppose H is the abstract algebra generated by $U_0 = \sqrt{2}T_w$ and U_1 modulo the relations (a) and (b). We have a natural homomorphism $B : H \to H(\gamma)$ of $\mathbb{C}$-algebras. By Proposition 1, $H(\gamma)$ is spanned by T_n and U_n and by Proposition 5, these elements are generated by U_0 and U_1. This shows that B is surjective. To show that it is injective, suppose $h \in H$ is in the kernel of B. Since U_0 and U_1 satisfy quadratic relations, $h = \sum_i c_i u_i$, where $c_i \in \mathbb{C}$ and $u_i \in H$ is of the form $U_1 U_0 U_1 U_0 \cdots$ or $U_0 U_1 U_0 U_1 \cdots$. Because $U_0 U_1 = T_1$, $B(u_i)$ is either T_n, $T_n U_1 = U_{1-n}$, $U_1 T_n = U_{n+1}$, or $U_1 T_n U_1 = T_{-n}$. These elements have disjoint support as functions in $H(\gamma)$. Therefore $B(h) = \sum_i c_i B(u_i) = 0$ implies that $c_i = 0$ and $h = 0$. This proves that B is an injection and Theorem 6. $\square$

We now give two consequences of Theorem 6:

Proposition 7. *The element $Z := T_1/2 + (T_1/2)^{-1}$ belongs to the center of $H(\gamma)$.*

Proof. By Proposition 5, T_1 and U_1 generate $H(\gamma)$. Clearly Z commutes with T_1. It suffices to show that Z commutes with U_1. Since $T_1 = U_0U_1$, we can use quadratic relations satisfied by U_0 and U_1 to write

$$2Z = U_0U_1 + U_1U_0 - 2^{1/2}U_1. \tag{2}$$

Hence Z commutes with U_1. □

Proposition 8. *For $n \geq 0$, T_n is an invertible element in the algebra $H(\gamma)$.*

Proof. The quadratic relations satisfied by U_0 and U_1 imply that U_0 and U_1 are invertible, and so is T_1, since $T_1 = U_0U_1$. Hence $T_n = T_1^n$ is invertible. □

Suppose (π, V) is a representation in $\mathcal{R}(G, \gamma)$. Let $V(N)$ denote the span of $\pi(n)v - v$ for all $v \in V$ and $n \in N$, and let $V_N = V/V(N)$ be the Jacquet module. Let

$$(V_N)^{K_0 \cap T, \gamma} = \{v \in V_N : \pi_{V_N}(t)v = \gamma(t)v \text{ for all } t \in K_0 \cap T\}.$$

The invertibility of T_n implies the following; see [Borel 1976, Lemma 4.7].

Corollary 9. *Suppose (π, V) is a representation in $\mathcal{R}(G, \gamma)$. Then the canonical map $V^{K_0,\gamma} \to (V_N)^{K_0 \cap T, \gamma}$ is a bijection. In particular V_N is nonzero, and V cannot be a supercuspidal representation.* □

4. Local Shimura correspondence

Let $G' = \mathrm{PGL}_2(\mathbb{Q}_2)$. Let I be its Iwahori subgroup and H' be its Iwahori–Hecke algebra. Let T'_w and U'_1 denote the characteristic functions of

$$I\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}I \quad \text{and} \quad I\begin{pmatrix} 0 & 1 \\ p & 0 \end{pmatrix}I$$

respectively. Then H' is the abstract $\mathbb{C}$-algebra generated by T'_w and U'_1 satisfying the same relations as (a) and (b) of Theorem 6; see [Matsumoto 1977]. This gives the next corollary.

Corollary 10. *The Hecke algebras $H(\gamma)$ and H' are isomorphic $\mathbb{C}$-algebras.*

Let $\mathcal{R}(H')$ denote the category of finite-dimensional representations of H'. Let $\mathcal{R}(G', I)$ denote the category of admissible smooth representations V of G' such that V^I generates V as a G'-module. By [Borel 1976; Bernšteĭn and Zelevinskiĭ 1976], the functor $V \mapsto V^I$ is an equivalence of categories from $\mathcal{R}(G', I)$ to $\mathcal{R}(H)$. The isomorphism in Corollary 10 establishes an equivalence of categories between $\mathcal{R}(H(\gamma))$ and $\mathcal{R}(H')$. Hence the following four categories are equivalent:

$$\mathcal{R}(G, \gamma) \simeq \mathcal{R}(H(\gamma)) \simeq \mathcal{R}(H') \simeq \mathcal{R}(G', I).$$

If V is a representation in $\mathcal{R}(G, \gamma)$, then we call the corresponding representation in $\mathcal{R}(G', I)$ the *local Shimura lift* of V. We denote it by $\mathrm{Sh}(V)$.

Proposition 11. *Let V be a representation in $\mathcal{R}(G,\gamma)$. Then the following are equivalent.*

(i) *The local Shimura lift* $\mathrm{Sh}(V)$ *is a spherical representation of G'.*

(ii) *The action of T'_w on* $\mathrm{Sh}(V)^I$ *has an eigenvalue* 2.

(iii) *The action of T_w on $V^{K_0,\gamma}$ has an eigenvalue* 2.

Proof. The projection map to $G'(\mathbb{Z}_2)$-fixed vectors in $\mathrm{Sh}(V)$ is given by $\frac{1}{3}(T'_w+1)$, since T'_w+1 is the characteristic function of $G'(\mathbb{Z}_2)$ and the volume of $G'(\mathbb{Z}_2)$ is 3. It follows that a $G'(\mathbb{Z}_2)$-fixed vector is an eigenvector of T'_w with eigenvalue 2. This proves the equivalence of (i) and (ii). The equivalence of (ii) and (iii) follows from Corollary 10. □

This proposition motivates the following definition.

Definition. Let V be a smooth representation of G. An eigenvector of T_w in $V^{K_0,\gamma}$ with an eigenvalue 2 is called a γ-*spherical* vector. The representation is called a γ-*unramified* or γ-*spherical* representation if it contains a γ-spherical vector.

5. Pseudospherical representation of K at $p = 2$

We retain the notations in Sections 3 and 4 where $p=2$. In the previous section we defined a representation V of G to be unramified if $V^{K_0,\gamma}\neq 0$ and T_w has an eigenvalue 2 on $V^{K_0,\gamma}$. In this section, we will reinterpret this condition in terms of representations of K, and see that K has only two irreducible representations E such that $E^{K_0,\gamma}\neq 0$. For both representations, $E^{K_0,\gamma}$ is one-dimensional and they are distinguished by the action of T_w on $E^{K_0,\gamma}$. That eigenvalue can be either 2 or -1, so we use the eigenvalue to denote the representations by $V(2)$ and $V(-1)$. Their dimensions are 2 and 4, respectively. Thus, a representation of G is unramified if and only if it contains the two-dimensional K-type $V(2)$, which we may call a pseudospherical type.

If $E^{K_0,\gamma}\neq 0$, then, by Frobenius reciprocity, the K-type E is a summand of a six-dimensional induced representation

$$I_K(\gamma) := \mathrm{Ind}_{K_0}^{K}\gamma = \{\phi : K\to\mathbb{C} : \phi(k_0k)=\gamma(k_0)\phi(k) \text{ for all } k\in K,\ k_0\in K_0\}.$$

Here the group K acts on it by right translation, denoted π_R. Let $H_K(\gamma)$ denote the subalgebra of $H(\gamma)$ consisting of functions supported on K. We have the action of $H_K(\gamma)$ on $I_K(\gamma)^{K_0,\gamma}$, also denoted by π_R. By Proposition 1, $H_K(\gamma)=\mathbb{C}1\oplus\mathbb{C}T_w$ and it is a commutative subalgebra. The algebra $H_K(\gamma)$ is antiisomorphic to the algebra $H_K(\bar{\gamma})$ via the map $f\mapsto\hat{f}$, where $\hat{f}(g)=f(g^{-1})$.

For $f\in H_K(\bar{\gamma})$ and $\phi\in I_K(\gamma)$, we set

$$(\pi_L(f)\phi)(g)=\int_K f(k)\phi(k^{-1}g)dk \quad \text{for all } g\in K.$$

This action commutes with the right action π_R of K on $I_K(\gamma)$ and

$$H_K(\bar{\gamma}) = \mathrm{End}_K(I_K(\gamma)).$$

Note that $I_K(\gamma)^{K_0,\gamma} = H(\bar{\gamma})$. The actions π_L and π_R of $H(\bar{\gamma})$ and $H(\gamma)$ on $I_K(\gamma)^{K_0,\gamma} = H(\bar{\gamma})$ are related by $\pi_L(\hat{f}) = \pi_R(f)$.

We define the functions $F_{-1} := \frac{1}{3}(2 - T_w)$ and $F_2 := \frac{1}{3}(T_w + 1)$ in $H_K(\gamma)$. Then $\{F_{-1}, F_2\}$ is a basis of idempotents of $H_K(\gamma)$.

For $j = -1, 2$, let $V(j) = \pi_L(\hat{F}_j)I_K(\gamma)$. In other words $V(j)$ is the eigenspace of $\pi_L(\hat{T}_w)$ on $I_K(\gamma)$ corresponding to the eigenvalue j. Note that $\hat{F}_j \in V(j)$ and $\pi_R(T_w)\hat{F}_j = j\hat{F}_j$. In particular $\hat{F}_2$ is a γ-spherical vector.

Proposition 12. (i) *We have* $I_K(\gamma) = V(-1) \oplus V(2)$, *where each summand is an irreducible representation of* K.

(ii) *We have* $\dim V(-1) = 4$ *and* $\dim V(2) = 2$.

(iii) *The* K*-submodule* $V(2)$ *contains a* γ*-spherical vector* $\hat{F}_2$. *The space of* γ*-spherical vectors is one-dimensional.*

(iv) *The* K*-submodule* $V(-1)$ *does not have any* γ*-spherical vector.*

Proof. Since $\dim \mathrm{End}(I_K(\gamma)) = 2$, both $V(-1)$ and $V(2)$ are irreducible K-modules. This proves (i).

To compute the dimensions of $V(-1)$ and $V(2)$ we need a lemma.

Lemma 13. *The operator* $\pi_L(\hat{T}_w)$ *as an element in* $\mathrm{End}_K(I_K(\gamma))$ *has trace* 0.

Proof. For $g \in K$, let ϕ_g be an element of $I_K(\gamma)$ such that ϕ_g is supported on $K_0 g$ and $\phi_g(k_0 g) = \gamma(k_0)$. Let S be a set of representatives of $K_0\backslash K$. Then $\{\phi_g : g \in S\}$ is a basis of $I_K(\gamma)$. To prove the lemma, it suffices to show that $(\pi_L(\hat{T}_w)\phi_g)(g) = 0$. Indeed, this shows that the matrix of $\pi_L(\hat{T}_w)$ in the basis ϕ_g has vanishing diagonal entries. Note that $\pi_L(T_w)\phi_g$ is supported on $K_0 w(1) K_0 g$. If $(\pi_L(\hat{T}_w)\phi_g)(g) \neq 0$, then $g \in K_0 w(1) K_0 g$ and $1 \in K_0 w(1) K_0$. This is a contradiction since K_0 is not equal to $K_0 w(1) K_0$. □

We have $\dim V(2) + \dim V(-1) = \dim I_K(\gamma) = [K : K_0] = 6$. By the lemma, $2\dim V(2) - \dim V(-1) = 0$. This implies $\dim V(-1) = 4$ and $\dim V(2) = 2$ and proves Proposition 12(ii). We have $I_K(\gamma)^{K_0,\gamma} = H_K(\bar{\gamma})$ and $\pi_R(F_j)I_K(\gamma) = \mathbb{C}\hat{F}_j$ for $j = -1, 2$. The vector $\hat{F}_2$ is γ-spherical while $\hat{F}_{-1}$ is not. This proves parts (iii) and (iv). □

Theorem 14. *A smooth representation* V *of* G *with central character* γ *is* γ*-unramified if and only if there is a nontrivial* K*-module homomorphism* l *from* $V(2)$ *to* V. *A vector in* V *proportional to* $l(\hat{F}_2)$ *is a* γ*-spherical vector of* V.

Proof. A γ-spherical vector in V generates a representation of K in which every irreducible K-submodule is isomorphic to an irreducible submodule of $I_K(\gamma)$. Now the theorem follows from Proposition 12. □

6. Unramified principal series representations at $p = 2$

In this section, we continue to assume $p = 2$ and use notation of Sections 3 to 5. We will show that γ-unramified representations appear as submodules of principal series representations.

We recall the character γ of T in Section 3. Let $(\pi_s, I(\gamma, s))$ be the normalized induced principal series representation, where $I(\gamma, s)$ is the set of smooth functions $\phi : G \to \mathbb{C}$ satisfying

$$\phi(\epsilon x(u)h(t)g) = \epsilon\gamma(t)|t|^{s+1}\phi(g) \quad \text{for all } \epsilon \in \mu_2,\, u \in \mathbb{Q}_2 \text{ and } t \in \mathbb{Q}_2^\times.$$

The group G acts by left translation: $(\pi_s(g)\phi)(g') = \phi(g'g)$.

Proposition 15. *An irreducible γ-unramified representation V is isomorphic to a submodule of some $I(\gamma, s)$.*

Proof. By Corollary 9, $(V_N)^{K_0\cap T,\gamma}$ is nonzero. Hence there is a nontrivial T-homomorphism $V_N \to \gamma\nu^{s+1}$ for some $s \in \mathbb{C}$. Here ν is the character $\nu(\underline{h}(t)) = |t|$. By Frobenius reciprocity, there is a nontrivial map $V \to I(\gamma, s)$ that is an injection because V is irreducible. □

We recall that $K(4)$ is the principal congruence subgroup in K_1. Restricting functions ϕ in $I(\gamma, s)$ to K gives a natural isomorphism $l : I_K(\gamma) \to I(\gamma, s)^{K(4)}$ of K-modules.

Theorem 16. *The K-types $V(2)$ and $V(-1)$ are of multiplicity one in $I(\gamma, s)$. The space $I(\gamma, s)^{K_0,\gamma}$ is 2-dimensional and is spanned by $l(\hat{F}_2)$ and $l(\hat{F}_{-1})$.*

Waldspurger [1981, Chapter VI] describes an explicit basis of $I(\gamma, s)^{K_0,\gamma}$ and calculates the action of the operator T_1.

We will describe a scalar multiple ϕ_j of $l(\hat{F}_j) \in V_j$ that is more convenient for later calculations. Let $d_2 = 1$ and $d_{-1} = -2$, and define ϕ_j to be the unique vector in $I(\gamma, s)$ whose restriction to K is given by

$$\phi_j(k) = \begin{cases} d_j\gamma(k) & \text{if } k \in K_0, \\ 2^{-1/2}\zeta\gamma(k_0k_0') & \text{if } k = k_0w(1)k_0' \in K_0w(1)K_0, \\ 0 & \text{otherwise.} \end{cases}$$

We define an intertwining map $M(s) : I(\gamma, s) \to I(\gamma, -s)$ by

$$(M(s)\phi)(g) = \int_{\mathbb{Q}_2} \phi(w(1)x(u)g)\,du,$$

where $g \in G$ and du is the Haar measure on $\mathbb{Q}_2$ such that the measure of $\mathbb{Z}_2$ is 1.

Proposition 17. *We have*

$$M(s)\phi_2 = \frac{\zeta}{\sqrt{2}}\left(\frac{1-\frac{1}{2}(2^{-2s})}{1-2^{-2s}}\right)\phi_2 \quad \textit{and} \quad M(s)\phi_{-1} = -\frac{\zeta}{2\sqrt{2}}\left(\frac{1-2(2^{-2s})}{1-2^{-2s}}\right)\phi_{-1}.$$

Proof. Since the vector ϕ_j is unique up to a scalar in $I_K(\gamma)$, we have $M(s)\phi_j = c\phi_j$ for some $c \in \mathbb{C}$. It remains to determine $c = d_j^{-1} M(s)\phi_j(1)$.

If $u \notin \mathbb{Z}_2$, we have $w(1)x(u) = (-1,u)_2 \cdot x(-u^{-1})h(u^{-1})y(u^{-1})$. We write $u^{-1} = 2^m v$, where $v \in \mathbb{Z}_2^\times$ and $m \geq 1$. Recall that $\gamma(t) = \gamma(h(t))$. Then

$$\begin{aligned} &M(s)\phi_j(1) \\ &\quad = \int_{\mathbb{Z}_2} \phi_j(w(1)x(u))du + \sum_{m=1}^{\infty} 2^{m-1} \int_{\mathbb{Z}_2^\times} (-1, 2^m v)_2 \phi_j(h(2^m v)y(2^m v))d^\times v \\ &\quad = 2^{-1/2}\zeta + \sum_{m=1}^{\infty} 2^{-ms-1} \int_{\mathbb{Z}_2^\times} (-1, v)_2 \gamma(2^m v)\phi_j(y(2^m v))d^\times v, \end{aligned}$$

where $d^\times v$ is the Haar measure of $\mathbb{Z}_2^\times$ with total measure 1. Now $\phi_j(y(2^m v)) = 0$ if $m = 1$ and it is equal to 1 if $m \geq 2$. Since $\gamma(2^m v) = \gamma(2^m)\gamma(v)(2^m, v)_2$ and $\gamma(2^m) = 1$, we can rewrite

$$\begin{aligned} M(s)\phi_j(1) &= 2^{-1/2}\zeta + d_j \sum_{m=2}^{\infty} 2^{-ms-1} \int_{\mathbb{Z}_2^\times} (2, v)_2^m (-1, v)_2 \gamma(v) d^\times v \\ &= 2^{-1/2}\zeta + d_j \sum_{m=2}^{\infty} 2^{-ms-1} \frac{1}{4} \sum_{v \in (\mathbb{Z}/8\mathbb{Z})^\times} (2, v)_2^m (-1, v)_2 \gamma(v). \end{aligned}$$

The sum $\sum_{v\in(\mathbb{Z}/8\mathbb{Z})^\times}$ on the right is zero if m is odd, and equals $\sqrt{2}\zeta$ if m is even. Finally adding up all the terms gives the constant c and the lemma. □

Let $s_0 = 1/2$ or $1/2 + i\pi/\log 2$. From Proposition 17, ϕ_{-1} lies in the kernel of $M(s_0)$, so $I(\gamma, s_0)$ is reducible. Indeed $I(\gamma, s_0)$ has a unique irreducible quotient that is an even Weil representation.

Definition. Let $s_0 = 1/2$ or $1/2 + i\pi/\log 2$. The kernel of $M(s_0)$ is called the *Steinberg* representation of $G(\mathbb{Q}_2)$. We shall denote this representation by $St(\epsilon)$, where $\epsilon = \pm 1$ such that $2^{s_0} = \epsilon\sqrt{2}$.

We claim that $St(\epsilon)$ is an irreducible representation of $G(\mathbb{Q}_2)$. Indeed by [Loke and Savin 2010, Section 6], we have

(3) $$I(\gamma, s)_N^{\mathrm{ss}} \cong \gamma|\cdot|^{s+1} \oplus \gamma|\cdot|^{-s+1} \quad \text{for every } s \in \mathbb{C},$$

where $I(\gamma, s)_N^{\mathrm{ss}}$ is the semisimplification of $I(\gamma, s)_N$ as a T-module. Hence $I(\gamma, s)$ has at most length 2. The claim now follows because $St(\epsilon)$ is a proper submodule of $I(\gamma, s_0)$. Also see [Savin 2004, Section 7].

Corollary 18. *The even Weil representation contains the irreducible K-module $V(2)$. It is a γ-unramified representation. The Steinberg representation contains the irreducible K-module $V(-1)$.* □

Proposition 19. *Let $Z = T_1/2 + (T_1/2)^{-1}$ be the central element in the Hecke algebra $H(\gamma)$ as in Proposition 7. Then $\pi_s(Z)$ acts on $I(\gamma, s)^{K_0,\gamma}$ as the scalar $2^s + 2^{-s}$.*

Proof. By Corollary 9, the natural projection of $I(\gamma, s)$ onto $I(\gamma, s)_N$ gives an isomorphism of $I(\gamma, s)^{K_0,\gamma}$ and $I(\gamma, s)_N$. From Proposition 3(i)'s decomposition of $K_0 h(2) K_0$ into single K_0-cosets, it follows that the action of T_1 on $I(\gamma, s)^{K_0,\gamma}$ corresponds to the action of $4 \cdot \pi_{s,N}(h(2))$ on $I(\gamma, s)_N$. By (3), the eigenvalues of $T_1/2$ are 2^s and 2^{-s}. □

Corollary 20. *An irreducible γ-unramified representation is uniquely determined by the eigenvalue of the action of Z on its γ-spherical vector.*

Proof. Suppose the irreducible γ-unramified representation is a subquotient of both $I(\gamma, s)$ and $I(\gamma, s')$. Then by Proposition 19, $2^s + 2^{-s} = 2^{s'} + 2^{-s'}$, which implies $2^s = 2^{s'}$ or $2^s = 2^{-s'}$. By Proposition 17 both $I(\gamma, s)$ and $I(\gamma, -s)$ have the same irreducible γ-unramified subquotient. □

Corollary 21. *The Steinberg representation $St(\epsilon)$ corresponds to the one-dimensional representation of $H(\gamma)$ given by $T_w = -1$ and $U_1 = -\epsilon$.*

Proof. We know that $T_w = -1$ on $St(\epsilon)^{K_0,\gamma}$. It remains to compute the action of U_1. Since $St(\epsilon)$ is a subquotient of $I(\gamma, s_0)$, where $2^{s_0} = \epsilon\sqrt{2}$, the central element Z acts on $St(\epsilon)$ by the scalar $\epsilon(2^{1/2} + 2^{-1/2})$. By (2) we have $2^{1/2} Z = T_w U_1 + U_1 T_w - U_1$. Hence $U_1 = -\epsilon$ as claimed. □

Let V be an irreducible γ-unramified representation. By Proposition 15, we may assume that V is the unique γ-unramified subquotient of $I(\gamma, s)$ for some $s \in \mathbb{C}$. By Proposition 11, its local Shimura lift $V' = \mathrm{Sh}(V)$ is an unramified irreducible representation of $G' = \mathrm{PGL}_2(\mathbb{Q}_2)$. Let B' be the Borel subgroup of G'. We may realize V' as the unramified irreducible subquotient of the normalized induced principal series representation $(\pi'_s, I'(t))$ with trivial central character. Here $I'(t) = \mathrm{Ind}_{B'}^{G'} \omega^t$ (normalized induction), where ω is the character

$$\omega\begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} = |a_1/a_2|.$$

The next theorem is similar to [Waldspurger 1991, Proposition 4]. There the local correspondence is defined by restricting the oscillator representation to the dual pair $G(\mathbb{Q}_2) \times \mathrm{PGL}_2(\mathbb{Q}_2)$, while the local Shimura lift used here is defined by the Hecke algebra isomorphism.

Theorem 22. *If V is the unique γ-unramified irreducible subquotient of $I(\gamma, s)$, then its local Shimura lift* $\mathrm{Sh}(V)$ *is the unique unramified irreducible subquotient of $I'(s)$.*

Proof. Assume that $\mathrm{Sh}(V)$ is a subquotient of $I'(t)$. By Proposition 19 the central operator Z in $H(\gamma)$ acts on $I(\gamma, s)^{K_0,\gamma}$ by the scalar $2^s + 2^{-s}$. The corresponding operator Z' in the algebra H' acts on $I'(t)$ by $2^t + 2^{-t}$. Thus, $2^s + 2^{-s} = 2^t + 2^{-t}$. Solving the equation gives $2^s = 2^t$ or $2^s = 2^{-t}$. Both $I'(t)$ and $I'(-t)$ have the same irreducible subquotients, so we may set $s = t$. □

We deduce a corollary that is a part of [Waldspurger 1980, Propositions 1 and 2].

Corollary 23. *The principal series representation $I(\gamma, s)$ is reducible if and only if $s = 1/2$ or $1/2 + i\pi/\log 2$.*

Proof. Let V be the γ-unramified irreducible subquotient of $I(\gamma, s)$. Let W be the unramified irreducible subquotient of $I'(s)$. Then $V = I(\gamma, s)$ if and only if $\dim V^{K_0,\gamma} = 2$. By Theorem 22, $\dim V^{K_0,\gamma} = \dim W^I$. Now $\dim W^I = 2$ if and only if $I'(s)$ is irreducible. Finally $I'(s)$ is irreducible if and only if $s \neq 1/2$ and $1/2 + i\pi/\log 2$. □

7. Automorphic forms

In this section we review a connection between automorphic forms and classical modular forms of half integral weight. This is mostly well known material that can be found in [Gelbart 1976, Chapters 2 and 3] and in [Waldspurger 1981]. We then transfer the action of the Hecke algebra $H(\gamma)$ to the setting of classical modular forms.

Let $\mathbb{A} = \prod_v \mathbb{Q}_v$ be the ring of adeles over $\mathbb{Q}$. We recall K_p, $s(g)$ and the cocycle σ_v defined in Section 2. Let $G(\mathbb{A}) = \mathrm{SL}_2(\mathbb{A}) \times \{\pm 1\}$ as a set. For $g_1 = (g_{1,v})$, $g_2 = (g_{2,v}) \in \mathrm{SL}_2(\mathbb{A})$ and $\epsilon_1, \epsilon_2 \in \{\pm 1\}$, the group law on $G(\mathbb{A})$ is given by

$$(g_1, \epsilon_1)(g_2, \epsilon_2) = (g_1 g_2, \epsilon_1 \epsilon_2 \sigma(g_1, g_2)),$$

where $\sigma(g_1, g_2) = \prod_v \sigma_v(g_{1,v}, g_{2,v})$. Then $\mathrm{pr} : G(\mathbb{A}) \to \mathrm{SL}_2(\mathbb{A})$, $(g, \epsilon) \mapsto g$ is a twofold cover that splits over the subgroup $\mathrm{SL}_2(\mathbb{Q})$. Since $\mathrm{SL}_2(\mathbb{Q})$ is perfect, this splitting is unique and given by

$$s_{\mathbb{Q}} : \mathrm{SL}_2(\mathbb{Q}) \to G(\mathbb{A}), \quad g \mapsto (g, s_{\mathbb{A}}(g)), \quad \text{where } s_{\mathbb{A}}(g) = \prod_v s(g_v).$$

We also need a description of a maximal compact subgroup in $G(\mathbb{R})$. Let

$$\underline{k}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) \quad \text{for } -\pi < \theta \leq \pi.$$

Then $\underline{K}_\infty := \{\underline{k}(\theta) : -\pi < \theta \le \pi\}$ is a maximal compact subgroup in $\mathrm{SL}_2(\mathbb{R})$. Let $K_\infty = \{k(\theta) : -2\pi < \theta \le 2\pi\}$, where

$$k(\theta) = \begin{cases} (\underline{k}(\theta), 1) & \text{if } -\pi < \theta \le \pi, \\ (\underline{k}(\theta), -1) & \text{if } -2\pi < \theta \le -\pi \text{ or } \pi < \theta \le 2\pi. \end{cases}$$

Then K_∞ is a maximal compact subgroup of $G(\mathbb{R})$ and $\mathrm{pr}(K_\infty) = \underline{K}_\infty$. If r is an odd integer, then $k(\theta) \mapsto e^{ir\theta/2}$ defines a genuine character of K_∞

Let $A_{r/2}(4)$ denote the set of functions φ in $L^2(\mathrm{SL}_2(\mathbb{Q})\backslash G(\mathbb{A}))$ satisfying the following properties:

(1) $\varphi(gk_1) = \varphi(g)$ for all $k_1 \in K_1(4) \prod_{p \ne 2,\infty} K_p$;

(2) $\varphi(gk_0) = \gamma(k_0)\varphi(g)$ for all $k_0 \in K_0$ in $G(\mathbb{Q}_2)$, where $\gamma(-1) = -i^r$;

(3) $\varphi(gk(\theta)) = e^{i\frac{r}{2}\theta}\varphi(g)$;

(4) φ is smooth as a function on $G(\mathbb{R})$ and satisfies $\triangle\varphi = -\frac{1}{4}r(\frac{1}{4}r - 1)\varphi$, where $\triangle$ is the Casimir operator; and

(5) φ is cuspidal, that is, $\int_{N(\mathbb{Q})\backslash N(\mathbb{A})} \varphi(x(u)g)du = 0$ for all $g \in G(\mathbb{A})$.

A basis of $A_{r/2}(4)$ arises from cuspidal automorphic representations $\pi = \bigotimes_v \pi_v$ of $G(\mathbb{A})$ such that π_∞ is a holomorphic discrete series representation with the lowest weight $r/2$, π_p is unramified for all $p \ne 2$, and π_2 contains a $K_1(4)$-fixed vectors. In particular, $\pi_2^{K_0,\gamma} \ne 0$ for some central character γ. Note that γ is determined by r. Indeed, since the local components of $s_{\mathbb{Q}}(h(-1))$ for $v \ne \infty, 2$ are contained in K_p, we have $\varphi(1) = \varphi(s_{\mathbb{Q}}(h(-1))) = \gamma(-1)e^{i\pi r/2}\varphi(1)$, and therefore $\gamma(-1) = -i^r$.

Let $\mathcal{H}$ be the complex upper half plane. For elements $\underline{g} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$, $g = (\underline{g}, \epsilon) \in G(\mathbb{R})$ and $z \in \mathcal{H}$, we define

$$gz = \underline{g}z = \frac{az+b}{cz+d}.$$

We define a holomorphic function on $\mathcal{H}$ by

$$J(g, z) = J((\underline{g}, \epsilon), z) := \epsilon\,(cz+d)^{1/2}.$$

Here we choose $w^{1/2}$ so that $-\pi/2 < \arg(w^{1/2}) \le \pi/2$. We call $J(g, z)$ a *factor of automorphy*. By [Gelbart 1976, Lemma 3.3], it has $J(gg', z) = J(g, g'z)J(g', z)$ for any two g and g' in $G(\mathbb{R})$. Define a congruence subgroup $\Gamma_0(4)$ by

$$\Gamma_0(4) := G(\mathbb{R}) \cap \Big(s_{\mathbb{Q}}(\mathrm{SL}_2(\mathbb{Q})) \cdot K_0(4) \cdot \prod_{p \ne 2} K_p\Big).$$

Similarly, define $\Gamma_1(4) \subseteq \Gamma_0(4)$ by replacing $K_0(4)$ with $K_1(4)$. Let $S_{r/2}(\Gamma_0(4))$ and $S_{r/2}(\Gamma_1(4))$ be the spaces of classical modular forms of weight $r/2$. By

[Koblitz 1984, page 183], we have $S_{r/2}(\Gamma_0(4)) = S_{r/2}(\Gamma_1(4))$. We will denote this space by $S_{r/2}(4)$.

By [Gelbart 1976, Proposition 3.1], there is a bijection $Q : A_{r/2}(4) \to S_{r/2}(4)$, which we recall: If $\varphi \in A_{r/2}(4)$, then

$$(Q\varphi)(z) = \varphi(g_\infty) J(g_\infty, i)^r, \quad \text{where } z = g_\infty i \in \mathcal{H}.$$

Conversely, given $f \in S_{r/2}(4)$, let $g \in G(\mathbb{A})$. By [Gelbart 1976, Lemma 3.2], $g = g_{\mathbb{Q}} g_\infty k$ for some $g_{\mathbb{Q}} \in \boldsymbol{s}_{\mathbb{Q}}(\mathrm{SL}_2(\mathbb{Q}))$, $g_\infty \in G(\mathbb{R})$ and $k \in K_1(4) \prod_{p \neq 2,\infty} K_p$. Then $(Q^{-1} f)(g) = f(g_\infty(i)) J(g_\infty, i)^{-r}$.

Using the bijection Q, we define another bijection between the spaces of operators by

$$\boldsymbol{q} : \mathrm{End}_{\mathbb{C}}(A_{r/2}(4)) \to \mathrm{End}_{\mathbb{C}}(S_{r/2}(4)), \quad L \mapsto QLQ^{-1}.$$

Since the Hecke algebra $H(\gamma)$ defined in Section 3 acts on $A_{r/2}(4)$, it is of interest to reinterpret this action in terms of classical modular forms.

Proposition 24. *Let U_1 and T_1 be the operators in the local Hecke algebra $H(\gamma)$, where $\gamma(-1) = -i^r$. Recall that $\zeta = (1 - i^r)/\sqrt{2}$. For $f(z) \in S_{r/2}(4)$, we have*

(i) $(\boldsymbol{q}(U_1) f)(z) = \bar{\zeta}(2z)^{-r/2} f(-1/(4z))$ *and*

(ii) $(\boldsymbol{q}(T_1) f)(z) = 2^{-r/2} \sum_{u=0}^{3} f((z+u)/4)$.

Proof. (i) Suppose $\varphi = Q^{-1}(f) \in A_{r/2}(4)$. For every place v, let $w_v = w(2^{-1})$ be the element in $G(\mathbb{Q}_v)$ defined in Section 2. By Proposition 3(iv),

$$(U_1\varphi)(g_\infty) = \int_{K_0 w_2 K_0} U_1(k)\varphi(g_\infty k)dk = U_1(w_2)\varphi(g_\infty w_2) = \bar{\zeta}\varphi(g_\infty w_2).$$

Next, consider $\underline{w}(2^{-1})$ in $\mathrm{SL}_2(\mathbb{Q})$. By [Gelbart 1976, (2.30)], $s_{\mathbb{Q}}(\underline{w}(2^{-1})) = \prod w_v$. Since φ is left $\mathrm{SL}_2(\mathbb{Q})$-invariant, and right K_p-invariant for $p \neq 2$,

$$\bar{\zeta}\varphi(g_\infty w_2) = \bar{\zeta}\varphi(s_{\mathbb{Q}}(\underline{w}(2^{-1}))^{-1} g_\infty w_2) = \bar{\zeta}\varphi\Big(\Big(\prod_{v\neq 2} w_v^{-1}\Big) g_\infty\Big) = \bar{\zeta}\varphi(w_\infty^{-1} g_\infty).$$

Applying Q to this equation gives (i). Part (ii) is proved analogously. □

8. Kohnen's plus space

Hecke eigenforms in $S_{r/2}(4)$ correspond to cuspidal automorphic representations π such that π_∞ is a discrete series representation of lowest weight $r/2$, π_p is unramified for all $p \neq 2$, and π_2 has $K_1(4)$-fixed vectors. In particular, $\pi_2^{K_0,\gamma} \neq 0$ for the central character $\gamma(-1) = -i^r$. If π_2 is a principal series representation, then $\pi_2^{K_0,\gamma}$ is 2-dimensional and therefore the corresponding Hecke eigenspace in $S_{r/2}(4)$ is also 2-dimensional. Kohnen's plus space is introduced to resolve this ambiguity. In terms of the space of automorphic functions $A_{r/2}(4)$, it is clear what to do. Decompose $A_{r/2}(4) = A^+_{r/2}(4) \oplus A^-_{r/2}(4)$, where $A^+_{r/2}(4)$ is the eigenspace of the local Hecke operator T_w with eigenvalue 2, while $A^-_{r/2}(4)$ is the eigenspace with eigenvalue -1. Since the presence of the eigenvalue 2 for T_w acting on π_2 eliminates a possibility that π_2 is a Steinberg representation, we see that there is a one-to-one correspondence between Hecke eigenforms in $A^+_{r/2}(4)$ and cuspidal automorphic representations π (as above) such that π_2 is a γ-unramified representation. The classical Kohnen plus space is (essentially) $Q(A^+_{r/2}(4))$, as will be explained in a moment. Niwa [1977] defines two operators $\boldsymbol{T}_4$ and $\boldsymbol{W}_4$ on $S_{r/2}(4)$ by

$$(\boldsymbol{T}_4 f)(z) = \frac{1}{4}\sum_{u=0}^{3} f((z+u)/4) \quad \text{and} \quad (\boldsymbol{W}_4 f)(z) = (-2iz)^{-r/2} f(-1/(4z)).$$

Note that $\boldsymbol{W}_4^2 = 1$. Let $\kappa = (r-1)/2$. Niwa shows that the operator

$$\boldsymbol{W} = (-1)^{(r^2-1)/8} 2^{1-\kappa} \boldsymbol{W}_4 \boldsymbol{T}_4$$

on $S_{r/2}(4)$ satisfies[1] the quadratic relation $(\boldsymbol{W}+1)(\boldsymbol{W}-2) = 0$. Kohnen defines $S^+_{r/2}(4)$ and $S^-_{r/2}(4)$ to be the eigenspaces of $\boldsymbol{W}$ on $S_{r/2}(4)$ of eigenvalues 2 and -1, respectively [Kohnen 1980]. Proposition 24 says that

$$\boldsymbol{q}(U_1) = (-1)^{r^2-1/8} \boldsymbol{W}_4 \quad \text{and} \quad \boldsymbol{q}(T_1) = 2^{3/2-\kappa} \boldsymbol{T}_4,$$

where the sign $(-1)^{r^2-1/8}$ is the quotient of

$$\bar{\zeta} = (1+i^r)/\sqrt{2} \quad \text{and} \quad i^{r/2} = ((1+i)/\sqrt{2})^r.$$

Since $T_w = \sqrt{2}^{-1} T_1 U_1$, it follows that $\boldsymbol{q}(T_w)$ and $\boldsymbol{W}$ are conjugates of each other by $\boldsymbol{W}_4$. Thus Kohnen's plus space is simply a conjugate of our space:

$$Q(A^+_{r/2}(4)) = \boldsymbol{W}_4(S^+_{r/2}(4)).$$

Because $\boldsymbol{W}_4$ commutes with the classical Hecke operators $\boldsymbol{T}_{p^2}$ whenever $p \neq 2$, $Q(A^+_{r/2}(4))$ and $S^+_{r/2}(4)$ are isomorphic as $\mathbb{C}[\boldsymbol{T}_{3^2}, \boldsymbol{T}_{5^2}, \dots]$-modules.

There is another description of $S^+_{r/2}(4)$ in terms of Fourier coefficients. It consists of the cusp forms whose n-th Fourier coefficient vanishes whenever $(-1)^\kappa n \equiv 2, 3 \pmod 4$. Kohnen defines a Hecke operator $\boldsymbol{T}_4^+$ that preserves $S^+_{r/2}(4)$ in the following way: For $f(z) = \sum_n a_n q^n \in S^+_{r/2}(4)$, set $(\boldsymbol{T}_4^+ f)(z) = \sum_n b_n q^n$ where the sum is taken over integers $n > 0$ and $(-1)^\kappa n \equiv 0, 1 \pmod 4$, and

$$b_n = a_{4n} + \left(\frac{(-1)^\kappa n}{2}\right) 2^{\kappa-1} a_n + 2^{r-2} a_{n/4}.$$

[1]In [Kohnen 1980], the operator is $\boldsymbol{T}_4 \boldsymbol{W}_4$ acting on the right, that is, $\boldsymbol{T}_4$ acts first and $\boldsymbol{W}_4$ follows.

Here $a_{n/4} = 0$ if n is not a multiple of 4. The large parentheses denote the Legendre symbol.

We can now formulate and prove our main global results.

Theorem 25. *There is a one-to-one correspondence between Hecke eigenforms f in $S^+_{r/2}(4)$ and irreducible cuspidal automorphic representations $\pi = \bigotimes_v \pi_v$ in $L^2(\mathrm{SL}_2(\mathbb{Q})\backslash G(\mathbb{A}))$ such that*

(i) *π_∞ is the discrete series representation of $G(\mathbb{R})$ with lowest weight $r/2$;*

(ii) *π_p is unramified for all odd primes p;*

(iii) *π_2 is γ-unramified, where $\gamma(-1) = -i^r$; and*

(iv) *if $\boldsymbol{T}^+_4 f = \lambda_2 f$, then a γ-spherical vector in π_2 is an eigenvector for $Z = T_1/2 + (T_1/2)^{-1}$ with eigenvalue $2^{1-r/2}\lambda_2$.*

Note that λ_2 determines the eigenvalue of Z on a γ-spherical vector, which in turn determines π_2 uniquely by Corollary 20.

Proof. The first three statements are clear, since $Q^{-1}(\boldsymbol{W}_4 f)$ is a Hecke eigenform in $A^+_{r/2}(4)$ that is contained in a cuspidal automorphic representation π with these properties. It remains to show (iv).

Lemma 26. *Let f be in $S^+_{r/2}(4)$. Then $\boldsymbol{T}^+_4 f = 2^{r/2-1}\boldsymbol{q}(Z) f$.*

Proof. Recall that T_1 is invertible by Proposition 8. Hence, it suffices to show that $2^{2-r/2}\boldsymbol{q}(T_1)\boldsymbol{T}^+_4 = \boldsymbol{q}(T_1^2 + 4)$. If $f(z) = \sum_{n=1}^\infty a_n q^n \in S_{r/2}(4)$, then $(\boldsymbol{q}(T_1) f)(z) = 2^{2-r/2}\sum_{n=0}^\infty a_{4n} q^n$ by Proposition 24. Thus, if $f(z) \in S^+_{r/2}(4)$, then one computes

$$2^{2-r/2}(\boldsymbol{q}(T_1)\boldsymbol{T}^+_4 f)(z) = (\boldsymbol{q}(T_1^2+4) f)(z) = \sum_n (2^{4-r} a_{16n} + 4a_n) q^n. \qquad \square$$

Now we can finish the proof of Theorem 25. If $\boldsymbol{T}^+_4 f = \lambda_p f$, then Lemma 26 implies that $Q^{-1}(f)$ is an eigenform for Z with eigenvalue $2^{1-r/2}\lambda_2$. Since $\boldsymbol{W}_4 = (-1)^{(r^2-1)/8}\boldsymbol{q}(U_1)$ and Z commutes with U_1, we see that $Q^{-1}(\boldsymbol{W}_4 f)$ is also an eigenform for Z with the same eigenvalue. $\square$

If f is a Hecke eigenform in $S^+_{r/2}(4)$, then by [Kohnen 1980, Theorem 1(ii)] the corresponding Shimura lift $f' = \mathrm{Sh}(f)$ is a Hecke eigenform in $S_{r-1}(\mathrm{SL}_2(\mathbb{Z}))$. Recall that $G' = \mathrm{PGL}_2$. There is a bijection between Hecke eigenforms f' in $S_{r-1}(\mathrm{SL}_2(\mathbb{Z}))$ and irreducible cuspidal automorphic representations $\pi' = \bigotimes_v \pi'_v$ in $L^2(G'(\mathbb{Q})\backslash G'(\mathbb{A}))$ such that π'_∞ is a discrete series representation with lowest weight $r-1$ and π'_p is unramified for all primes p; see [Gelbart 1975, Proposition 3.1]. Recall the local Shimura lift $\mathrm{Sh}(\pi_2)$ in Proposition 11 of a γ-unramified representation π_2 of $G(\mathbb{Q}_2)$. The following corollary gives a precise representation-theoretic description of the Shimura correspondence at the place $p = 2$.

Corollary 27. *Let f be a Hecke eigenform in $S^+_{r/2}(4)$. Let $\pi = \bigotimes_v \pi_v$ be the cuspidal automorphic representation corresponding to f in Theorem 25. Let $\pi' = \bigotimes_v \pi'_v$ be the cuspidal automorphic representations of $L^2(G'(\mathbb{Q})\backslash G'(\mathbb{A}))$ corresponding to the Hecke eigenform $f' = \mathrm{Sh}(f)$ in $S_{r-1}(\mathrm{SL}_2(\mathbb{Z}))$. Then $\mathrm{Sh}(\pi_2) = \pi'_2$.*

Proof. If $\boldsymbol{T}^+_4 f = \lambda_2 f$, then $\boldsymbol{T}_2 f' = \lambda_2 f'$ by [Kohnen 1980, Theorem 1(ii)], where $\boldsymbol{T}_2$ is the classical Hecke operator action on $S_{r-1}(\mathrm{SL}_2(\mathbb{Z}))$. By [Gelbart 1975, Proposition 5.2.1], one checks that π'_2 is indeed isomorphic to $\mathrm{Sh}(\pi_2)$. □

Let π be a cuspidal automorphic representation of $G(\mathbb{A})$ as in Theorem 25 and π' be the corresponding cuspidal automorphic representation of $G'(\mathbb{A})$ as in Corollary 27. By the Ramanujan conjecture, proved by Deligne, $\pi'_2 = \mathrm{Sh}(\pi_2)$ is a tempered irreducible unramified representation, so $\pi'_2 = I'(s)$ for some $s \in i\mathbb{R}$. This implies that $\pi_2 = I(\gamma, s)$ by Theorem 22 and Corollary 23. Thus $\pi_2^{K_0,\gamma}$ is an irreducible $H(\gamma)$-module of dimension 2. It corresponds under Q to a two-dimensional subspace of $S_{r/2}(4)$ spanned by a line in $S^+_{r/2}(4)$ and a line in $S^-_{r/2}(4)$.

On the other hand, if $\pi_2 = St(\epsilon)$ is a Steinberg representation of $G(\mathbb{Q}_2)$ (see the definition before Corollary 18), then π corresponds under Q to an Hecke eigenform in $S^-_{r/2}(4)$. More precisely:

Theorem 28. *There is a one-to-one correspondence between Hecke eigenforms f in $S^-_{r/2}(4)$ such that $\mathbf{W}_4 f = -\epsilon(-1)^{(r^2-1)/8} f$ for some $\epsilon = \pm 1$ and irreducible cuspidal automorphic representations $\pi = \bigotimes_v \pi_v$ in $L^2(\mathrm{SL}_2(\mathbb{Q})\backslash G(\mathbb{A}))$ such that*

(i) *π_∞ is the discrete series representation of $G(\mathbb{R})$ with lowest weight $r/2$,*

(ii) *π_p is unramified for all odd primes p, and*

(iii) *π_2 is the Steinberg representation $St(\epsilon)$.*

Proof. Recall by Corollary 21 that T_w and U_1 act on the one-dimensional space $St(\epsilon)^{K_0,\gamma}$ by -1 and $-\epsilon$. The theorem now follows from Proposition 24 and the definition of $S^-_{r/2}(4)$. □

Acknowledgment

Loke thanks the University of Utah for its hospitality during work on this paper.

References

[Baruch and Mao 2007] E. M. Baruch and Z. Mao, "Central value of automorphic L-functions", *Geom. Funct. Anal.* **17**:2 (2007), 333–384. MR 2008g:11075 Zbl 05166590

[Bernšteĭn and Zelevinskiĭ 1976] I. N. Bernšteĭn and A. V. Zelevinskiĭ, "Representations of the group $\mathrm{GL}(n, F)$, where F is a local non-Archimedean field", *Uspehi Mat. Nauk* **31**:3 (1976), 5–70. In Russian; translated in *Russian Math. Surveys* **31**:3 (1976), 1–68. MR 54 #12988

[Borel 1976] A. Borel, "Admissible representations of a semi-simple group over a local field with vectors fixed under an Iwahori subgroup", *Invent. Math.* **35** (1976), 233–259. MR 56 #3196 Zbl 0334.22012

[Casselman 1995] W. Casselman, "Introduction to the theory of admissible representations of p-adic reductive groups", draft of workshop notes from Séminaire Paul Sally, 1995, available at http://www.math.ubc.ca/~cass/research/pdf/p-adic-book.pdf.

[Gelbart 1975] S. S. Gelbart, *Automorphic forms on adèle groups*, Annals of Mathematics Studies **83**, Princeton University Press, 1975. MR 52 #280 Zbl 0329.10018

[Gelbart 1976] S. S. Gelbart, *Weil's representation and the spectrum of the metaplectic group*, Lecture Notes in Math. **530**, Springer, Berlin, 1976. MR 54 #12654 Zbl 0365.22017

[Koblitz 1984] N. Koblitz, *Introduction to elliptic curves and modular forms*, Graduate Texts in Mathematics **97**, Springer, New York, 1984. MR 86c:11040 Zbl 0553.10019

[Kohnen 1980] W. Kohnen, "Modular forms of half-integral weight on $\Gamma_0(4)$", *Math. Ann.* **248**:3 (1980), 249–266. MR 81j:10030 Zbl 0416.10023

[Loke and Savin 2010] H. Y. Loke and G. Savin, "Modular forms on nonlinear double covers of algebraic groups", *Trans. Amer. Math. Soc.* **362** (2010), 4901–4920.

[Matsumoto 1977] H. Matsumoto, *Analyse harmonique dans les systèmes de Tits bornologiques de type affine*, Lecture Notes in Math. **590**, Springer, Berlin, 1977. MR 58 #28315 Zbl 0366.22001

[Niwa 1977] S. Niwa, "On Shimura's trace formula", *Nagoya Math. J.* **66** (1977), 183–202. MR 58 #27781 Zbl 0351.10018

[Savin 2004] G. Savin, "Lectures on representations of p-adic groups", pp. 19–46 in *Representations of real and p-adic groups*, edited by E.-C. Tan and C.-B. Zhu, Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap. **2**, Singapore Univ. Press, 2004. MR 2005g:22021 Zbl 1056.22011

[Stein 1973] M. R. Stein, "Surjective stability in dimension 0 for K_2 and related functors", *Trans. Amer. Math. Soc.* **178** (1973), 165–191. MR 48 #6267 Zbl 0267.18015

[Waldspurger 1980] J.-L. Waldspurger, "Correspondance de Shimura", *J. Math. Pures Appl.* (9) **59**:1 (1980), 1–132. MR 83f:10029

[Waldspurger 1981] J.-L. Waldspurger, "Sur les coefficients de Fourier des formes modulaires de poids demi-entier", *J. Math. Pures Appl.* (9) **60**:4 (1981), 375–484. MR 83h:10061 Zbl 0431.10015

[Waldspurger 1991] J.-L. Waldspurger, "Correspondances de Shimura et quaternions", *Forum Math.* **3**:3 (1991), 219–307. MR 92g:11054 Zbl 0724.11026

Received July 20, 2009.

HUNG YEAN LOKE
NATIONAL UNIVERSITY OF SINGAPORE
10, LOWER KENT RIDGE ROAD
SINGAPORE 119076
SINGAPORE

matlhy@nus.edu.sg
http://www.math.nus.edu.sg/~matlhy/

GORDAN SAVIN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH
SALT LAKE CITY, UT 84112
UNITED STATES

savin@math.utah.edu
http://www.math.utah.edu/~savin/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

LARGE QUANTUM CORRECTIONS IN MIRROR SYMMETRY FOR A 2-DIMENSIONAL LAGRANGIAN SUBMANIFOLD WITH AN ELLIPTIC UMBILIC

GIOVANNI MARELLI

Given the Lagrangian fibration $\mathbb{R}^4 \to \mathbb{R}^2$ and a Lagrangian submanifold, exhibiting an elliptic umbilic and supporting a flat line bundle, we study in the context of mirror symmetry the corrections necessary to solve the monodromy of the holomorphic structure of the mirror bundle on the dual fibration. This is a preliminary step towards the understanding of quantum corrections in this specific case.

1. Introduction

The first steps in the study of mirror symmetry, assuming the existence of dual torus fibrations X and $\widehat{X}$, has been undertaken in papers such as [Fukaya 2005; Arinkin and Polishchuk 2001; Leung et al. 2001; Bruzzo et al. 2001; 2002]: Under certain hypotheses, a transform is provided, defined on some subcategory of the Fukaya category of X, that maps pairs formed by a Lagrangian submanifold L and a $U(1)$-flat connection ∇ to holomorphic bundles $\widehat{E}$ over $\widehat{X}$. The caustic K of L is always assumed to be empty. The purpose of this paper is to start understanding how to remove this hypothesis.

We focus our attention on the Lagrangian fibration $\mathbb{R}^4 \to \mathbb{R}^2$ and consider a Lagrangian map $f : L \hookrightarrow \mathbb{R}^4 \to \mathbb{R}^2$. Generically, f exhibits only folds and cusps, which are singularities of codimension 1 and 2 respectively. If we restrict the fibrations and L to the subset $\mathbb{R}^2 \setminus K$, then the Lagrangian map f has no singular points, and so we can try to apply the constructions contained in the papers mentioned before. We can hope to get a holomorphic bundle $\widehat{E}$ on the dual fibration restricted to $\mathbb{R}^2 \setminus K$, and whose holomorphic structure can be extended to the whole fibration over $\mathbb{R}^2$. However this hope is in general vain (we consider the elliptic umbilic in Section 4, but see also the same example described in [Fukaya 2005, Section 5.4]): What may happen, as in the case we are going to study, is that K is a compact curve, and in the noncompact subset of $\mathbb{R}^2$ determined by K, the holomorphic structure of $\widehat{E}$ presents a monodromy when going around the caustic K, and this

MSC2000: primary 14J32, 58E05, 58E07; secondary 37G25, 53D12.
Keywords: mirror symmetry, quantum corrections, elliptic umbilic, Morse homology.

hinders us from extending the mirror bundle to K and gluing it to the mirror bundle constructed inside K, and so from producing a holomorphic bundle $\widehat{E}$ on the whole dual fibration. Some kinds of *quantum corrections* are thus required to obtain a holomorphic bundle defined on the whole dual fibration $\widehat{\mathbb{R}^4} \to \mathbb{R}^2$.

One program to perform quantum corrections is outlined in [Fukaya 2005]. The idea is that there is a "classical" or, as it is called there, "semiflat" situation, which is the one where X is a Lagrangian fibration with only smooth fibers. In this case a complex structure is easily constructed on the mirror manifold $\hat{X}$; however as soon singular fibers are allowed, the complex structure constructed on the smooth part does not extend to the whole $\hat{X}$. So there is a need for some corrections to extend the complex structure.

Computations from physics suggest that pseudoholomorphic discs provide these corrections. When a Lagrangian submanifold L is given on X, if it is not a ramified cover a holomorphic mirror bundle is obtained on $\hat{X}$. When the caustic is allowed, the situation is similar to the case of singular fibers, and the classical holomorphic structure, obtained from L away from the caustic on the mirror bundle, must be deformed to extend it to the whole mirror bundle. Again, the idea outlined in [Fukaya 2005] is that quantum corrections are provided by the instanton effect, that is, by counting pseudoholomorphic strips in $\mathbb{R}^4$ that bound L and the fiber F_x of the fibration.

In a more general framework, quantum corrections are related to the obstruction theory of well-definedness of Floer homology. As proposed in [Fukaya 2000] as a general idea and holding beyond the specific case considered there, the fiber over $x \in \mathbb{R}^2 \setminus K$ of the mirror bundle $\widehat{E}$ on $\widehat{\mathbb{R}^4}$ is constructed as the Lagrangian intersection Floer homology of L and of the Lagrangian fiber of $\mathbb{R}^4$ over x. Assume that K contains just one singular point and that this point is an elliptic umbilic. We know that in dimension 2 this singularity is neither stable nor generic; however, from [Marelli 2006a; 2006b], we know how the caustic K and the bifurcation locus B change when f is slightly perturbed to $\tilde{f}$.

According to a conjecture proposed by Fukaya [2005, Section 3.5], near K, Lagrangian intersection Floer homology is equivalent to Morse homology defined by means of the generating function f of L, which is a Morse function far from K and B. This conjecture allows us to switch from Floer to Morse homology. More precisely, Fukaya conjectured that near the caustic, the moduli space of pseudoholomorphic discs bounded by a fiber and by the Lagrangian submanifold L is isotopic to the moduli space of gradient lines of the generating function of L between the points of intersections of the fiber with L. This conjecture has been proved in [Fukaya and Oh 1997] for the case of the cotangent bundle and in some of the examples considered in [Fukaya 2005]. Its purpose is just to provide a way to simplify the computations involved in working with pseudoholomorphic discs.

In our case, we apply this for a small perturbation of the elliptic umbilic. In view of this, that is, considering gradient lines rather than pseudoholomorphic discs, our result can be only a first step in the study of quantum corrections in mirror symmetry for the example considered.

Another important fact to be remarked is that, put simply, the homological mirror symmetry conjecture establishes an equivalence between the Fukaya category on one side and the derived category of coherent sheaves on the mirror side; however here our concern is only about objects and not also about morphisms, so again this work should be considered only as a preliminary step in understanding mirror symmetry.

With all the limitations outlined above, quantum corrections (that is, rules to glue the holomorphic Morse homology bundle $\hat{\hat{E}}$, relative in our case to $\tilde{f}$) are then defined across folds that are not limit points of bifurcation lines, and across bifurcation lines far from their intersections. Fukaya [2005, Section 5.4] explained that the cancellation of monodromy, which may look accidental, is in fact related to the phenomenon of wall crossing of Floer homology. In our case, which is approximated in the sense that we use Morse homology instead of Floer homology, the caustic and bifurcation locus represent the walls in an analogous phenomenon for Morse homology. We check that in this way the holomorphic structure of $\hat{\hat{E}}$ can be extended to the codimension 2 subset of $\mathbb{R}^2$ containing the remaining points of $\tilde{K}$ and $\tilde{B}$, that is, the intersection points of bifurcation lines, folds that are limit points of bifurcation lines, and cusps. We realize however that these corrections are not enough to extend the holomorphic structure of $\hat{\hat{E}}$ to cusps. A correction of different kind is thus required, which is related to the possibility of defining a spin structure on $\tilde{L}$ or, better, a relative spin structure. This has to do with the orientation problem in Floer homology theory; see [Fukaya et al. 2000]. In this way, the monodromy around the caustic is also canceled, and so the mirror bundle $\hat{\hat{E}}$ can be endowed with a holomorphic structure defined on the whole dual fibration.

2. Preliminaries

Throughout this paper we will use results from [Marelli 2006a; 2006b], whose contents we now summarize.

- We introduced Lagrangian bundles $\pi : X \to B$, Lagrangian maps $g : L \to B$ and their generating functions, and defined the caustic K of L as the set of critical values of $\pi \circ g$;
- We recalled the classification of Lagrangian singularities. We noted that in dimension 2 only folds and cusps are stable and generic; the elliptic umbilic, which is the case considered in this paper, is stable and generic starting from dimension 3. However it can appear as unstable singularity (that is, it breaks

in folds and cusps under a small perturbation) in dimension 2; it is given by the generating function

$$f(y_1, y_2) = \tfrac{1}{3}y_1^3 - y_1 y_2^2.$$

- We showed that the elliptic umbilic in dimension 2 becomes after a small perturbation a Lagrangian map whose caustic is a tricuspoid, a curve with three edges, whose points are folds, and three cusps at vertexes.
- For f the generating function of L, we defined the family $f_x : \mathbb{R}^n \to \mathbb{R}$ of functions, where x is a point in the base of the fibration, as

$$f_x(y) = f(y) - x \cdot y$$

  and considered the gradient system

$$\nabla f_x(y) = \frac{dy}{dt} \tag{1}$$

  whose solutions we called gradient lines. We noted that the caustic K of L is the subset of points x where the gradient field ∇f_x exhibits a degenerate critical point. We defined B, the bifurcation locus of L, as the subset of points x where f_x is a Morse function but ∇f_x is not Morse–Smale, that is, where the phase portrait of ∇f_x features a saddle-to-saddle separatrix.
- We studied how the bifurcation locus of the elliptic umbilic, represented by three straight half-lines with a common vertex at the origin, is modified after a small perturbation. Far from the caustic, a tricuspoid, the bifurcation locus looks as that of the unperturbed elliptic umbilic with three bifurcation lines; in a neighborhood of the caustic, these are half-lines that generically have vertex at a fold point of the caustic. As for the mutual positions of the bifurcation lines and their possible intersections, see [Marelli 2006b, Theorem 4.14] and the pictures there representing the diagrams that can be expected.

3. The mirror bundle

We recall how the mirror bundle should be constructed for the trivial fibration $\mathbb{R}^4 \to \mathbb{R}^2$ and its dual (but more generally also for $\mathbb{R}^{2n} \to \mathbb{R}^n$). In [Leung et al. 2001; Bruzzo et al. 2001; 2002], it is defined by a kind of Fourier–Mukai transform that associates to each pair formed by a Lagrangian submanifold L, in the given Lagrangian fibration, and a local system ∇ on it, a vector bundle $\hat{E}$ on the dual fibration, endowed with a connection $\hat{\nabla}$. Its curvature $\hat{F}$ satisfies $\hat{F}^{0,2} = 0$ and so induces a holomorphic structure on $\hat{E}$. This is achieved under certain hypotheses, among which that L has no caustic. On the other hand, in [Fukaya 2000], the fiber of the mirror bundle $\hat{E}$ over the point (x, w) of the dual fibration (x is a coordinate on the base and w on the fiber) is defined as the Lagrangian intersection Floer homology of L and F_x:

$$\hat{E}_{(x,w)} = HF((L,\nabla),(F_x,w)),$$

where $w \in \hat{F}_x$ defines a flat connection on F_x. For the affine Lagrangian submanifolds considered in [Fukaya 2000], HF^k is nonvanishing only when k equals the dimension of the fiber. A holomorphic frame is then defined on $\hat{E}$. These two constructions are equivalent in the cases considered in the papers above, that is, when assuming at least that the fibration has no singular fibers and that L has no caustic.

Here, we will follow mainly the second construction (though sometimes we use also the Fourier–Mukai construction), since this approach seems to be more suitable as quantum corrections are provided by pseudoholomorphic discs. However, as explained in the introduction, using a conjecture by Fukaya [2005, Section 3.5], near the caustic we switch from Floer homology and pseudoholomorphic discs to Morse homology and gradient lines. So that we don't introduce notation that we don't use, we state the conjecture informally and refer to [Fukaya 2005] for the precise formulation.

Conjecture 3.1. *The moduli space of gradient lines is isotopic to the moduli space of pseudoholomorphic discs in a neighborhood of a point of the caustic.*

Partial progress towards a proof of this conjecture has been made by Floer[1988] and by Fukaya and Oh [1997], the latter in the case of the cotangent bundle.

The transfer to Morse homology is then performed as follows. Consider the trivial Lagrangian fibration $\mathbb{R}^{2n} \to \mathbb{R}^n$. To L is associated a (local) generating function $f : \mathbb{R}^n \to \mathbb{R}$. We consider as in Section 2 the family $f_x : \mathbb{R}^n \to \mathbb{R}$ and the gradient system (1). Let K and B be the caustic and the bifurcation locus of L. If $x \notin K \cup B$, with some further hypotheses on f (see [Schwarz 1993]), the Morse complex is defined over x. The space of k-chains is the free $\mathbb{C}$-module generated by critical points of Morse index k and the differential is defined counting gradient lines, that is, the solutions of the gradient system (1) joining two critical points whose Morse indexes differ by 1. The fiber of the mirror bundle is defined as the Morse homology $\hat{E}_{(x,w)} = HM(f_x)$ of the Morse complex over x, and a holomorphic frame is constructed similarly to that proposed in [Fukaya 2000; 2005]. Namely, by writing $\nabla = d + A$, a section $e(x)$ of $\hat{E}$ turns out to be holomorphic and descends on the torus fibers when multiplied by the weight

$$\exp\Bigl(2\pi\Bigl(\frac{h(x)}{2} - \frac{A(x)}{4\pi} + i\frac{\partial h}{\partial x}\cdot w\Bigr)\Bigr),$$

where h is a multivalued function on the base such that each sheet of L is locally the graph of dh. In other words, h is a set of local generating functions defined in the coordinates of the base, one for each sheet of L. The problem is to glue this bundle along the caustic K and the bifurcation locus B.

4. The monodromy of the elliptic umbilic

Consider the trivial Lagrangian torus fibration $\mathbb{R}^4 \to \mathbb{R}^2$ and a Lagrangian submanifold L whose caustic K contains an elliptic umbilic q. In a neighborhood of q, we can choose symplectic coordinates (y_1, y_2, x_1, x_2), with coordinates y_1 and y_2 on the fibers and x_1 and x_2 on the base of the fibration, such that L is given by the generating function

$$f : \mathbb{R}^2 \to \mathbb{R}, \quad (y_1, y_2) \mapsto \tfrac{1}{3} y_1^3 - y_1 y_2^2 \tag{2}$$

Since we will be working in a neighborhood of q, we can use the local coordinates just introduced. This means we consider the Lagrangian fibration $\mathbb{R}^4 \to \mathbb{R}^2$ and the Lagrangian submanifold L defined by the generating function f. Associated to f, we have the caustic K and the bifurcation locus B. By hypothesis, $K = \{(0, 0)\}$, while [Marelli 2006b] shows that B is given by three half-lines from $(0, 0)$, defined by $t \to te^{i\alpha}$ for $\alpha = 0, 2\pi/3, 4\pi/3$ and $t > 0$.

Consider a line bundle E over L with a flat $U(1)$-connection ∇. The pair (L, ∇) defines an object in the Fukaya category of the symplectic manifold $\mathbb{R}^4$. On $\mathbb{R}^2 \setminus K$ the function f has no critical points, so we can apply results of [Bruzzo et al. 2002] or [Fukaya 2000], thus producing a bundle $\widehat{E}$ of rank 2 over the total space of the dual fibration restricted to $\mathbb{R}^2 \setminus K$. A hermitian connection $\widehat{\nabla}$ can be defined on $\widehat{E}$, thus inducing a holomorphic structure on $\widehat{E}$. Note that L is a 2-sheeted cover of $\mathbb{R}^2 \setminus K$. Thus for $x \in \mathbb{R}^2 \setminus K$ if $p_1(x)$ and $p_2(x)$ denote the elements of $L \cap F_x$, where F_x is the fiber of the Lagrangian fibration $\mathbb{R}^4 \to \mathbb{R}^2$ over x, and if z_1 and z_2 are coordinates along the fibers of the dual fibration, then the connection $\widehat{\nabla}$ can be written as $d + \widehat{A}$, with

$$\widehat{A}(x) = i(p_1(x)dz_1 + p_2(x)dz_2). \tag{3}$$

However, let $\Gamma \in \pi_1(\mathbb{R}^2 \setminus K)$, $\Gamma : [0, 1] \to \mathbb{R}^2$, and consider the continuous maps

$$M^i_\Gamma : [0, 1] \to \mathbb{R}^4, \quad t \mapsto p_i(\Gamma(t)) \quad \text{for } i = 1, 2. \tag{4}$$

Let $M^i_\Gamma(t)_F$ be the projection onto $F_{\Gamma(t)} \cong \mathbb{R}^2$ of $M^i_\Gamma(t)$.

Definition 4.1. The monodromy of the holomorphic structure of $\widehat{E}$ is the map

$$\mathcal{M} : \pi_1(\mathbb{R}^2 \setminus K) \to \mathrm{End}(\mathbb{R}^2), \quad \mathcal{M}(\Gamma)(M^i_\Gamma(0)_F) = M^i_\Gamma(1)_F. \tag{5}$$

Since $\Gamma(0) = \Gamma(1)$, $M^i_\Gamma(0)$ and $M^i_\Gamma(1)$ belong to the same fiber. Also, the endomorphism $\mathcal{M}(\Gamma)$ is well-defined, since $\{M^i_\Gamma(t)\}$ for $i = 1, 2$ is a basis of $F_{\Gamma(t)}$.

Lemma 4.2. *If* Γ *is a nontrivial simple loop in* $\pi_1(\mathbb{R}^2 \setminus K)$, *then the monodromy* $\mathcal{M}$ *of the holomorphic structure* $\widehat{E}$ *on* Γ *can be represented by the matrix*

$$\mathcal{M}(\Gamma) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Proof. This follows because the points $p_1(x)$ and $p_2(x)$ exchange when going around the origin. In fact, since L can be written as $x_1 = y_1^2 - y_2^2$ and $x_2 = -2y_1y_2$, it becomes $z = \overline{w}^2$ by writing $z = x_1 + ix_2$ and $w = y_1 + iy_2$. □

This lemma shows that $\widehat{E}$ cannot be extended to a holomorphic bundle on the whole dual fibration over $\mathbb{R}^2$. For this, some “quantum correction” must be added; see also [Fukaya 2005, Section 5.4].

5. Perturbations of the elliptic umbilic

Consider now a small perturbation $\tilde{f}$ of f. The caustic $\widetilde{K}$ and the bifurcation locus $\widetilde{B}$ of $\tilde{f}$ were studied in [Marelli 2006a] and [Marelli 2006b], respectively. There $\widetilde{K}$ was shown to be diffeomorphic to a tricuspoid, and $\widetilde{B}$, outside a disc containing $\widetilde{K}$, looks as the bifurcation locus of the unperturbed f, while inside this disc its structure can be highly complicated and bifurcation lines can intersect. See [Marelli 2006b] for pictures of the several admissible diagrams representing the reciprocal positions of $\widetilde{K}$ and $\widetilde{B}$ inside the disc. At first we restrict our attention to the subset $\mathbb{R}^2 \setminus \widetilde{K}$. Given a flat connection $\widetilde{\nabla}$ on the Lagrangian submanifold $\widetilde{L}$ defined by $\tilde{f}$, we construct a holomorphic bundle $\widehat{\widetilde{E}}$ on each of the two connected components of $\mathbb{R}^2 \setminus \widetilde{K}$, as explained in [Bruzzo et al. 2002] or in [Fukaya 2000]. As done in Section 4 for $\widehat{E}$, we can define the monodromy $\widetilde{\mathcal{M}}$ of the holomorphic structure of $\widehat{\widetilde{E}}$ and prove the following lemma:

Lemma 5.1. *If* Γ *is a nontrivial simple loop in* $\pi_1(\mathbb{R}^2 \setminus K)$, *then the monodromy* $\widetilde{\mathcal{M}}$ *of the holomorphic structure of* $\widehat{\widetilde{E}}$ *on* Γ *can be represented by the matrix*

$$\widetilde{\mathcal{M}}(\Gamma) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Proof. Since f is perturbed on a compact subset D containing the origin, it follows that $\tilde{f}$ coincides with f outside D and that $\widetilde{K} \subset D$. So $\widetilde{\mathcal{M}}(\Gamma) = \mathcal{M}(\Gamma)$. □

Therefore, outside the caustic, the holomorphic structure of $\widehat{\widetilde{E}}$ also exhibits a monodromy.

6. Quantum corrections to perturbations of the elliptic umbilic

The problem is to solve the monodromy and extend the holomorphic structure of $\widehat{\widetilde{E}}$ across the caustic $\widetilde{K}$, gluing it with the holomorphic structure inside $\widetilde{K}$. The way to achieve this is to construct $\widehat{\widehat{E}}$ with its holomorphic structure on $\mathbb{R}^2 \setminus (\widetilde{K} \cup \widetilde{B})$, define morphism gluing this structure across $\widetilde{K}$ and $\widetilde{B}$ and check if the monodromy is solved. This is what we mean by quantum corrections. We are going to define quantum corrections on $\widehat{\widehat{E}}$. Then, since a holomorphic section is obtained by multiplying a section of $\widehat{\widehat{E}}$ by a suitable weight, we will obtain quantum corrections for holomorphic sections of $\widehat{\widehat{E}}$. If a section can be extended to $\widetilde{K} \cup \widetilde{B}$, the same will hold for a holomorphic section. The features of the set $\mathbb{R}^2 \setminus (\widetilde{K} \cup \widetilde{B})$, namely, the mutual positions of $\widetilde{K}$ and $\widetilde{B}$, are described in [Marelli 2006b, Theorem 4.14].

We explain now how to construct the mirror bundle (of Section 3) far from $\widetilde{K} \cup \widetilde{B}$ in this case. The function $\tilde{f}_x$ defined by $\tilde{f}_x(y) = \tilde{f}(y) - x \cdot y$ is a Morse function for every $x \in \mathbb{R}^2 \setminus (\widetilde{K} \cup \widetilde{B})$. As computed in [Marelli 2006b], if x lies inside the caustic, $\tilde{f}_x$ has four critical points: three saddles $s_i(x)$, the points with Morse index 1, and an unstable node $n(x)$, the point with Morse index 2. Thus the Morse complex is

$$0 \leftarrow 0 \leftarrow \bigoplus_{i=1}^{3} \mathbb{C}[s_i(x)] \xleftarrow{\partial_x} \mathbb{C}[n(x)] \leftarrow 0 \leftarrow \cdots, \tag{6}$$

where $\mathbb{C}[s_i(x)]$ and $\mathbb{C}[n(x)]$ denote the free modules over $\mathbb{C}$ generated by $s_i(x)$ and $n(x)$, respectively. The differential ∂ can be defined after an orientation is chosen on the moduli space of gradient lines from n to s_i (see [Schwarz 1993] or [McDuff and Salamon 2004] for a more detailed construction of Morse homology). In our case, ∂_x can be defined as $\partial_x n(x) = s_1(x) + s_2(x) + s_3(x)$ (anyway, the Morse complex has only two nontrivial terms, so ∂ automatically satisfies $\partial^2 = 0$); we fix this choice of orientation of gradient lines.

If x lies outside the caustic, $\tilde{f}_x$ has two saddles as critical points, so the Morse complex is simply given by

$$0 \leftarrow 0 \leftarrow \mathbb{C}[s_i(x)] \oplus \mathbb{C}[s_j(x)] \leftarrow 0 \leftarrow \cdots. \tag{7}$$

Definition 6.1. The fiber $\widehat{\widehat{E}}_x$ of $\widehat{\widehat{E}}$ over $x \in \mathbb{R}^2 \setminus (\widetilde{K} \cup \widetilde{B})$ is defined to be the Morse homology of the Morse complex (6) or (7) if x lies respectively inside or outside the caustic.

In our case, Morse homology has only one nontrivial term, so for x inside the caustic

$$\widehat{\widehat{E}}_x = \frac{\bigoplus_{i=1}^{3} \mathbb{C}[s_i(x)]}{\partial_x(\mathbb{C}[n(x)])},$$

while for x outside the caustic, $\widehat{\widehat{E}}_x = \mathbb{C}[s_i(x)] \oplus \mathbb{C}[s_j(x)]$.

Definition 6.2. On each connected component of $\mathbb{R}^2 \setminus (\widetilde{K} \cup \widetilde{B})$, we define $\widehat{\widehat{E}}$ as the trivial bundle whose fiber at $x \in U_i$ is given by Definition 6.1.

We define now morphisms gluing the holomorphic bundle $\widehat{\widehat{E}}$ along $\widetilde{K}$ and $\widetilde{B}$. We start by considering the subset $\widetilde{K}_F$ of $\widetilde{K}$ consisting of folds that are not limit points of bifurcation lines. It is a codimension 1 subset of $\mathbb{R}^2$. Suppose U and

V are two connected components of $\mathbb{R}^2 \setminus (\widetilde{K} \cup \widetilde{B})$, lying respectively outside and inside the caustic, such that $\partial U \cap \partial V \neq \varnothing$, and let $\widetilde{K}_i \subset \partial U \cap \partial V \cap \widetilde{K}_F$ be a connected component of $\widetilde{K}_F$. For simplicity, suppose that V is inside the caustic and U is outside, so that along $\widetilde{K}_i$ the node n and the saddle s_i in V glue together and disappear in U. (The pair (n, s_i) is also called a birth/death pair.)

Definition 6.3. The isomorphism $\widehat{\widehat{E}}(U) \cong \widehat{\widehat{E}}(V)$ gluing $\widehat{\widehat{E}}$ along $\widetilde{K}_i$ is defined as the one induced in homology by the inclusion

$$\mathbb{C}[s_j] \oplus \mathbb{C}[s_k] \hookrightarrow \bigoplus_{l=1}^{3} \mathbb{C}[s_l(x)] \quad \text{for } j, k \neq i.$$

It is a good definition since the inclusion preserves kernel and image of the differential of the Morse complex.

The second group of definitions is concerned instead with gluing along the subset $\widetilde{B}_1$ of $\widetilde{B}$ consisting of points that are not intersection of bifurcation lines. It is a codimension 1 subset of $\mathbb{R}^2$.

Definition 6.4. For each $x \in \mathbb{R}^2 \setminus (\widetilde{K} \cup \widetilde{B})$ lying inside the caustic, we define the incidence matrix $I(x) = (I(x)_i) \in \mathrm{Mat}(3, 1)$ such that $I(x)_i = 0$ if there is no gradient line from $n(x)$ to $s_i(x)$, and $I(x)_i = 1$ otherwise.

Remark 6.5. Similar definitions in a different setting appear in [Igusa 2002; 1993; Igusa and Klein 1993], highlighting the relation between Morse theory and algebraic K-theory. The definition of incidence matrix also resembles that of transition matrix given in [Kokubu 2000].

The incidence matrix at x gives information about the phase portrait of the gradient vector field $\nabla \tilde{f}_x$ and is related to the Morse differential simply as

$$\partial_x n(x) = I(x)_1 s_1(x) + I(x)_2 s_2(x) + I(x)_3 s_3(x).$$

The incidence matrix is constant on each connected component of $\mathbb{R}^2 \setminus (\widetilde{K} \cup \widetilde{B})$: Indeed, the gradient vector fields $\nabla \tilde{f}_x$ are orbit equivalent for all x in the same connected component, and so the Morse complexes are isomorphic. Let U and V be two such components lying inside the caustic such that $\partial U \cap \partial V \neq \varnothing$, with incidence matrix $I(U)$ and $I(V)$, respectively. For $\tau \in \{1, 0, -1\}$, let $E_{ij}(\tau) \in \mathrm{Mat}(3, 3)$ be the triangular matrix whose (k, l)-entry is 1 if $k = l$, is τ if $k = i$ and $l = j$, and is 0 otherwise. By results in [Marelli 2006b], crossing a bifurcation line can change at most only one of the entries of the incidence matrix. Therefore either

- $I(U) \neq I(V)$ and so there is only one $k \in \{1, 2, 3\}$ such that $I(U)_k \neq I(V)_k$, or
- $I(U) = I(V)$.

Definition 6.6. Consider the transformation matrix from U to V associated to points in $\partial U \cap \partial V \cap \widetilde{B}_1$ of a bifurcation line of $\widetilde{B}$. When this bifurcation line is characterized by the appearance of a nongeneric gradient line from s_i to s_j, the transformation matrix is of the form $E_{ij}(\tau)$, with $E_{ij}(\tau)I(U) = I(V)$.

When $I(U) \neq I(V)$ it follows that $\tau = 1$ if $I(U)_j = 0$, and $\tau = -1$ if $I(U)_j = 1$. When instead $I(U) = I(V)$, there is an ambiguity in the choice of τ, which will be discussed in Example 6.8.

We give two examples to clarify the previous definition.

Example 6.7. Suppose the phase portrait of $\nabla \tilde{f}_x$ for $x \in U$ and for $x \in V$ is represented by the incidence matrices $I(U) = (1, 1, 1)$ and $I(V) = (1, 1, 0)$, respectively. There are two possible bifurcations from U to V (see [Marelli 2006a; 2006b] for further explanations and some pictures): Either the nongeneric gradient line $\gamma_{s_1 s_3}$ or the nongeneric gradient line $\gamma_{s_2 s_3}$ appears in the phase portrait of $\nabla \tilde{f}_x$ when x is the bifurcation point. The first bifurcation corresponds to the transformation matrix $E_{31}(-1)$, while the second corresponds to $E_{32}(-1)$. Instead, if crossing from V to U, the same bifurcations give the transformation matrices $E_{31}(1)$ and $E_{32}(1)$, respectively.

Example 6.8. $I(U) = I(V)$ occurs only in case (c) of the proof of Proposition 6.11 and shown in Figure 1 (which contains the notation), along the bifurcation line between δ and ϵ. The phase portraits in δ and ϵ, which are represented respectively in [Marelli 2006b, Figures 4.20 and 4.19], can be summarized here as follows. The separatrixes that connect s_1 and s_3 to n in α (the phase portrait over α is shown in [Marelli 2006b, Figure 4.17]) can form a saddle-to-saddle separatrix in ϵ, but this can not occur in δ. This can provide a criterion for the choice of τ (which cannot be justified further here) considering only the special example of the perturbed elliptic umbilic. The matrix $M(\omega_3)$ in the proof of Proposition 6.11 is the transformation matrix from ϵ to δ. There the choice of τ is the one that solves the monodromy.

Suppose now that U and V lie outside the caustic $\widetilde{K}$ and $\partial U \cap \partial V \cap \widetilde{B}_1$ is a subset of $\widetilde{B}_j$, one of the three bifurcation lines forming the bifurcation diagram $\widetilde{B}$, and assume $\widetilde{B}_j$ enters into $\widetilde{K}$ at a point p through the side l_j of $\widetilde{K}$, where n and s_j form a birth/death pair. Since we are working in a neighborhood of $\widetilde{K}$, we can assume that $p \in \partial U \cap \partial V$. Inside the caustic and in a neighborhood of p, we can associate a transformation matrix $E_{ik}(\tau)$ to $\widetilde{B}_j$ using Definition 6.6.

Definition 6.9. If U and V lie outside $\widetilde{K}$ and are as above, the transformation matrix from U to V associated to points in $\partial U \cap \partial V \cap \widetilde{B}_1$ of the bifurcation line $\widetilde{B}_j$ is the matrix $E_{ik}(\tau) \in \mathrm{Mat}(2, 2)$ obtained from $E_{ik}(\tau) \in \mathrm{Mat}(3, 3)$ above by deleting the j-th row and the j-th column.

The transformation matrix we associate to a bifurcation line $\widetilde{B}_j$ from U to V defines a morphism between the Morse complexes of U and V.

Definition 6.10. The isomorphism $\widehat{\widehat{E}}(U) \cong \widehat{\widehat{E}}(V)$ gluing $\widehat{\widehat{E}}$ along $\widetilde{B}_j$ is the one induced by the transformation matrix of Definition 6.6 or Definition 6.9 associated to the bifurcation line $\widetilde{B}_j$.

We have now to check that we can extend $\widehat{\widehat{E}}$ through the codimension 2 subset given by intersection points of bifurcation lines, limit points of bifurcation lines on the caustic, and the three cusps.

We start by considering intersection points of bifurcation lines. In [Marelli 2006b] we analyzed the conditions under which two bifurcation lines can intersect themselves.

Proposition 6.11. *The holomorphic bundle $\widehat{\widehat{E}}$ can be extended through intersection points of bifurcation lines.*

Proof. We check that, for all possible cases of intersection of bifurcation lines described in [Marelli 2006b] and for a chosen loop Γ around the intersection point p, the composition of the transformation matrices of bifurcation lines at intersection points with Γ is the identity.

From [Marelli 2006b], we know there are the three cases (a), (b), and (c) of Figure 1. The phase portraits in the subsets determined by bifurcation lines and of bifurcations in cases are represented in [ibid., Figures 4.7, 4.8 and 4.9 for (a), 4.11, 4.12, 4.13 and 4.14 for (b), and 4.17, 4.18, 4.19, 4.20, 4.21 and 4.22 for (c)].

In case (a), we know that the two bifurcation lines are characterized by the appearance of the same saddle-to-saddle separatrix, obtained by gluing the same pair of separatrices. So, choosing a simple loop Γ around p intersecting for simplicity the bifurcation lines at four points w_i for $i = 1, \dots, 4$, and associating to each w_i a transition matrix $M(w_i)$ according to Definition 6.6, we have

$$M(w_1) = M(w_3) = M(w_2)^{-1} = M(w_4)^{-1},$$

and thus

$$M(w_4)M(w_3)M(w_2)M(w_1) = \mathrm{Id}\,.$$

This implies that there is no monodromy around p and so the holomorphic bundle $\widehat{\widehat{E}}$ can be extended across p.

In case (b), we chose again a simple loop Γ around p intersecting the bifurcation lines at four points w_i for $i = 1, \dots, 4$. Suppose w_1 belongs to the bifurcation line from α to β, w_2 to the bifurcation line from β to δ, w_3 to the bifurcation line from δ to γ, and w_4 to the bifurcation line from γ to α. The transformation matrices

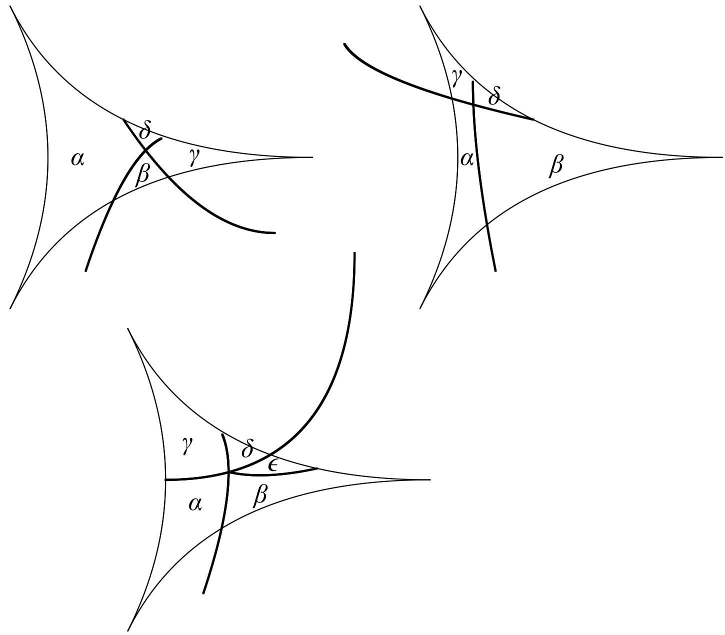


Figure 1. Clockwise from top left: Intersection of bifurcations lines in cases (a), (b), and (c)

associated by Definition 6.6 to the bifurcation lines at each w_i are given by

$$M(w_1)=\begin{pmatrix}1&-1&0\\0&1&0\\0&0&1\end{pmatrix},\qquad M(w_2)=\begin{pmatrix}1&0&0\\0&1&0\\0&-1&1\end{pmatrix},$$

$$M(w_3)=\begin{pmatrix}1&1&0\\0&1&0\\0&0&1\end{pmatrix},\qquad M(w_4)=\begin{pmatrix}1&0&0\\0&1&0\\0&1&1\end{pmatrix}.$$

Then $M(w_4)M(w_3)M(w_2)M(w_1)=\mathrm{Id}$, and so the holomorphic bundle $\widehat{\widehat{E}}$ can be extended across p.

In case (c) choose a simple loop Γ around p that intersects the bifurcation lines at five points w_i for $i=1,\dots,5$, starting from the bifurcation line from α to β and then proceeding anticlockwise. The transformation matrices are then

$$M(w_1)=\begin{pmatrix}1&-1&0\\0&1&0\\0&0&1\end{pmatrix},\quad M(w_2)=\begin{pmatrix}1&0&0\\0&1&0\\0&-1&1\end{pmatrix},\quad M(w_3)=\begin{pmatrix}1&0&0\\0&1&0\\-1&0&1\end{pmatrix},$$

$$M(w_4) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad M(w_5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Then $M(w_5)M(w_4)M(w_3)M(w_2)M(w_1) = \mathrm{Id}$, and so the holomorphic bundle $\widehat{\widetilde{E}}$ can be extended across p. □

We analyze now the behavior of $\widehat{\widetilde{E}}$ around limit points of bifurcation lines belonging to the caustic.

Proposition 6.12. *The holomorphic bundle $\widehat{\widetilde{E}}$ can be extended through limit points of bifurcation lines belonging to the caustic, when they are not cusps.*

Proof. From [Marelli 2006b] we know there are two cases: generically, either (i) the bifurcation line $\widetilde{B}$ enters into the caustic $\widetilde{K}$ at a fold or (ii) it is a half-line with origin at a fold (and the bifurcation line $\widetilde{B}$, near its origin, lies inside $\widetilde{K}$). See Figure 2. In both cases, let us denote this fold by p.

In case (i), p is not a cusp. So at each point of the caustic $\widetilde{K}$ near p, the node n glues with a saddle, which we suppose is s_1. Suppose also that the half-line $\widetilde{B}$ has its endpoint on the side of the caustic where n glues with s_2. With α the region marked in Figure 2, suppose that the phase portrait of $\nabla \tilde{f}_x$ for $x \in \alpha$ contains all the gradient lines γ_{ns_i}. Choose a simple loop Γ around p intersecting $\widetilde{B}$ at two points w_1 and w_3, and $\widetilde{K}$ at two points w_2 and w_4. Suppose w_1 lies inside the caustic and w_4 outside. The transition matrices at w_1 and w_3, according respectively to Definitions 6.6 and 6.9, are

$$M(w_1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \quad \text{and} \quad M(w_3) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Consider an element $h \in \widehat{\widetilde{E}}_x$ for $x \in \alpha$. Since $\widehat{\widetilde{E}}_x = (\bigoplus_{i=1}^3 \mathbb{C}[s_i(x)])/\partial_x(\mathbb{C}[n])$, we write h as an equivalence class $[(h_1, h_2, h_3)]$ in the basis (s_1, s_2, s_3) of $\mathbb{C}[s_i(x)]$,

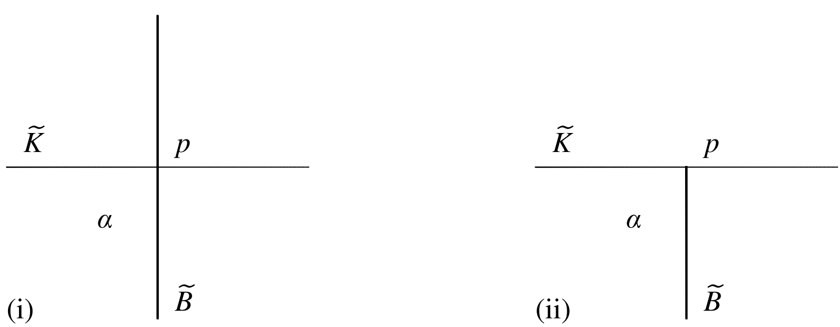


Figure 2. Mutual positions of bifurcation lines and caustic.

where $(h_1, h_2, h_3) \sim (h_1 + c, h_2 + c, h_3 + c)$ for every $c \in \mathbb{C}$. Moving along Γ from α into β, crossing $\widetilde{B}$ at w_1, we transform h by $M(w_1)$. At β, we have $(h_1, h_2, h_3) \sim (h_1 + c, h_2 + c, h_3)$ for every $c \in \mathbb{C}$, so we can write

$$[M(w_1)h] = [(h_1, h_2, -h_2 + h_3)] = [(0, h_2 - h_1, -h_2 + h_3)].$$

According to Definition 6.3, when crossing $\widetilde{K}$ at w_2 we have the gluing isomorphism

$$[(0, h_2 - h_1, -h_2 + h_3)] \cong (h_2 - h_1, -h_2 + h_3).$$

Crossing now $\widetilde{B}$ along Γ at w_3, we have

$$[M(w_3)(h_2 - h_1, -h_2 + h_3)^t] = (h_2 - h_1, h_3 - h_1).$$

Crossing $\widetilde{K}$ at w_4 and using the gluing isomorphism of Definition 6.3 we obtain

$$(h_2 - h_1, h_3 - h_1) \cong [(0, h_2 - h_1, h_3 - h_1)] = [(h_1, h_2, h_3)].$$

This shows that there is no monodromy and so $\widehat{\widehat{E}}$ can be extended through p.

In case (ii), suppose for simplicity that at p the node n and the saddle s_1 form the birth/death pair; that $\widetilde{B}$ intersects $\widetilde{K}$ at another point where n and s_2 form the birth/death pair; and that for $x \in \alpha$ the phase portrait of $\nabla \tilde{f}_x$ contains all the gradient lines γ_{ns_i}. Choose a simple loop Γ around p intersecting $\widetilde{B}$ at the point w_1, and $\widetilde{K}$ at two points w_2 and w_3. We know w_1 lies inside the caustic. The transformation matrix of Definition 6.6 at w_1 is

$$M(w_1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

Consider an element $h \in \widehat{\widehat{E}}_x$ for $x \in \alpha$, and write it again as in case (i) as an equivalence class $[(h_1, h_2, h_3)]$ in the basis (s_1, s_2, s_3) of $\mathbb{C}[s_i(x)]$. Going along Γ into β, crossing $\widetilde{B}$ in w_1, we transform h by $M(w_1)$. In β we have $(h_1, h_2, h_3) \sim (h_1 + c, h_2 + c, h_3)$ for every $c \in \mathbb{C}$, so we can write

$$[M(w_1)h] = [(h_1, h_2, -h_1 + h_3)] = [(0, h_2 - h_1, -h_1 + h_3)].$$

Now, crossing $\widetilde{K}$ at w_2 and using the gluing isomorphism of Definition 6.3, we have

$$[(0, h_2 - h_1, -h_1 + h_3)] \cong (h_2 - h_1, -h_1 + h_3).$$

Finally, entering into $\widetilde{K}$ through w_3 and using again the gluing isomorphism, we obtain in (α)

$$(h_2 - h_1, -h_1 + h_3) \cong [(0, h_2 - h_1, -h_1 + h_3)] = [(h_1, h_2, h_3)].$$

This shows that there is no monodromy and so $\widehat{\widehat{E}}$ can be extended through p. $\square$

Now we check if $\widehat{\widehat{E}}$ can be extended to cusps. To start suppose that at a cusp c the node n glues with the saddles s_2 and s_3. According to [Marelli 2006b] there are two cases: either (I) for x in a neighborhood of c, inside the caustic, the phase portrait of $\nabla \tilde{f}_x$ contains all the gradient lines γ_{ns_i}, or (II) it contains only γ_{ns_2} and γ_{ns_3}. In both cases a monodromy appears around the cusp.

Lemma 6.13. *In case (I), if Γ is a nontrivial simple loop around c, the monodromy of the holomorphic structure of $\widehat{\widehat{E}}$ along Γ is represented by the matrix*

$$M = \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix}. \tag{8}$$

Proof. For x outside the caustic, since $\widehat{\widehat{E}}_x = \mathbb{C}[s_1] \oplus \mathbb{C}[s_j]$, we may write an element $h \in \widehat{\widehat{E}}_x$ as (h_1, h_j). Suppose for $k \in \{2, 3\}$ that l_k is the branch of the caustic where n glues with s_k. Then on l_k the gluing isomorphism of Definition 6.3 identifies s_j with the saddle different from s_k and s_1. So, entering into the caustic through l_2 we have

$$(h_1, h_j) \cong [(h_1, h_j, 0)] = [(h_1 - h_j, 0, -h_j)].$$

Now exiting from the caustic through l_3, we have

$$[(h_1 - h_j, 0, -h_j)] \cong (h_1 - h_j, -h_j),$$

which gives the expected monodromy. □

Lemma 6.14. *In case (II), if Γ is a nontrivial simple loop around c, the monodromy of the holomorphic structure of $\widehat{\widehat{E}}$ along Γ is represented by the matrix*

$$M = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{9}$$

Proof. Using the notation in the proof of the previous lemma, we have, entering into the caustic through l_2

$$(h_1, h_j) \cong [(h_1, h_j, 0)] = [(h_1, 0, -h_j)].$$

Exiting the caustic through l_3, we have

$$[(h_1, 0, -h_j)] \cong (h_1, -h_j),$$

which gives the expected monodromy. □

In both cases, the matrix M is invertible. This means that the same monodromy is associated to Γ and to its opposite Γ^{-1} in $\pi_1(L \setminus \{c\})$.

If now at c the node n glues with the saddles s_1 and s_2 we have a similar result:

Lemma 6.15. *If* Γ *is a nontrivial simple loop around* c, *the monodromy of the holomorphic structure of* $\widehat{\widehat{E}}$ *along* Γ *is represented in case* (I) *by the matrix*

$$M = \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix} \tag{10}$$

and in case (II) *by the matrix*

$$M = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{11}$$

Proof. The proof is analogous to that of Lemmas 6.13 and 6.14. □

Again observe that the matrix M is invertible, meaning that Γ and Γ^{-1} provide the same monodromy.

Finally, if the node n glues at c with the saddles s_1 and s_3, we obtain:

Lemma 6.16. *If* Γ *is a nontrivial simple loop around* c, *the monodromy of the holomorphic structure of* $\widehat{\widehat{E}}$ *along* Γ *is represented in case* (I) *by the matrix*

$$M = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} \quad \textit{or by its inverse} \quad M^{-1} = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} \tag{12}$$

and in case (II) *by the matrix*

$$M = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \textit{or by its inverse} \quad M^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{13}$$

Proof. The proof is similar to that of Lemma 6.13 and 6.14. □

In both cases, if Γ is associated to M (say), then Γ^{-1} is associated to M^{-1}.

To solve the monodromy around the cusps it is necessary to add a new kind of correction. It is related to the possibility of defining a spin structure on $\widetilde{L}$ and to the problem of orientation in Lagrangian intersection Floer homology. (In fact from [Fukaya et al. 2000] we know that the existence of a relative spin structure on $\widetilde{L}$ is a condition for the orientability of the moduli space of pseudoholomorphic discs.) This is suggested intuitively by what follows: Consider the composition $\pi \circ i : \widetilde{L} \hookrightarrow \mathbb{R}^4 \to \mathbb{R}^2$, where π is the projection of the fibration and i is the Lagrangian immersion, and note that a spin structure can be induced at least on the subset of $\widetilde{L}$ where $d\pi$ is invertible, that is, on $\widetilde{L} \setminus \pi^{-1}(\widetilde{K})$. This means that the caustic or a subset of it represents an obstruction to the existence of a spin structure on $\widetilde{L}$.

The following result shows that the set of cusps is actually the obstruction to the existence of a spin structure on a Lagrangian submanifold L with generating function f. It proves, in fact, that the second Stiefel–Whitney class $w_2(L) \in H^2(L, \mathbb{Z}_2)$ of L, which represents the obstruction to the existence of spin structures on L, has the set of cusps as Poincaré dual in $H_0(L, \mathbb{Z}_2)$.

Lemma 6.17. *We have* $\mathrm{PD}(w_2(L)) = A_3(f)$ *where* $A_3(f)$ *is the set of singular points of* f *of type* A_3, *that is*, *the set of cusps.*

Proof. The equality is mainly proved by using the Thom polynomials of Lagrangian singularities. The proof is essentially given in [Kazarian 2003], where it follows from other major results given there. Kazarian first demonstrates that the cohomology class $\mathrm{PD}(\Omega(f))$, the Poincaré dual to the locus $\Omega(f)$ of singularities of f of class Ω, is equal to the Thom polynomial P_Ω associated to Ω. Kazarian computes Thom polynomials (see also [Vassilyev 1988]), and in particular, when $\Omega = A_3$, shows that $P_\Omega = w_2(T^*L) = w_2(TL)$. □

Let A be an immersed 1-dimensional submanifold of $\mathbb{R}^2$ with three nonintersecting connected components, each of which is a half-line with vertex at one of the three cusps of the caustic. To solve the monodromy around the cusps it is enough to glue (for example along A) the holomorphic structure so as to cancel the monodromy. The problem is to justify this procedure, which for the moment is just an ad hoc correction. As said, the idea, coming from the orientation problem of Lagrangian intersection Floer homology and confirmed by Lemma 6.17, is that the ability to define a spin structure on some flat bundle on $\widetilde{L}$ should provide such a correction. We make the following natural definition:

Definition 6.18. Along each half-line forming the submanifold A, depending on which cusp the half-line has as vertex, we glue the holomorphic bundle $\widehat{\widehat{E}}$ using the inverse of morphism (8), (10), or (12) in case (I) and (9), (11) or (13) in case (II).

This correction is called *orientation twist* in [Fukaya 2005].

Proposition 6.19. *If* $\widehat{\widehat{E}}$ *is glued along* A *according to Definition 6.18*, *then its holomorphic structure can be extended across the cusp.*

Proof. The proof is a direct consequence of Lemma 6.13, 6.14, 6.15 and 6.16, since the corrections applied are just the inverses of what we want to cancel. □

We explain now how to justify Definition 6.18, generalizing the idea outlined in [Fukaya 2005, Section 5.4]. Before considering the case of a perturbed elliptic umbilic, let us examine for simplicity a Lagrangian submanifold L exhibiting a cusp c. In this case, A is a half-line with vertex in c. Consider a ball U containing c. Since U is contractible, L owns a spin structure over U. On the other hand, $d\pi$ is invertible over the complement of U and so induces a spin structure on L. Since by Lemma 6.17, $w_2(L)$ does not vanish because of c, it follows that the nonexistence of a spin structure on L comes from the gluing of TL along the boundary of U. The purpose now is to show how A can provide both a correction to TL, by defining a new bundle carrying a spin structure, and a correction to the flat $U(1)$-line bundle $\mathscr{L}$ on L, yielding the gluing that cancels the monodromy. Consider representations

$$\rho : \pi_1(\mathbb{R}^2 \setminus \{c\}) \to \{1, -1\} = O(1) \subset U(1)$$

defining two representations $\rho^{O(1)} := \rho_1$ and $\rho^{U(1)} := \rho_2$. According to this choice we have, respectively, a flat $O(1)$-bundle $\mathscr{L}_{\rho_1}$ or a flat $U(1)$-bundle $\mathscr{L}_{\rho_2}$ on $\mathbb{R}^2 \setminus \{c\}$. There are two possibilities for ρ. It is either the trivial or the nontrivial group homomorphism $\mathbb{Z} \to \{1, -1\}$. When ρ is the nontrivial representation, its values on a path $\Gamma \in \pi_1(\mathbb{R}^2 \setminus \{c\})$ are given by the intersection number of Γ and A. $\mathscr{L}_{\rho_1}$ is the trivial bundle when ρ is trivial but is a Möbius strip when ρ is nontrivial. The same holds for $\mathscr{L}_{\rho_2}$; the bundle $\mathscr{L}_{\rho_2}$ restricted to a generator $\Gamma \cong S^1$ of $\pi_1(\mathbb{R}^2 \setminus \{c\})$ is the flat line bundle on the torus $T^1 = S^1$ with factor of automorphy equal to either 1 or -1 according to whether ρ is trivial or not. In other words, we may think of a section of $\mathscr{L}_{\rho_2}$ over Γ as multiplied by respectively 1 or -1 at $\Gamma \cap A$ (the factor of automorphy for $U(1)$-line bundles on tori and the induced connection on the mirror bundle are explained [Bruzzo et al. 2001; 2002]). The projection of the fibration $\pi : \mathbb{R}^4 \to \mathbb{R}^2$ and the composition $\pi \circ i$, where $i : L \hookrightarrow \mathbb{R}^4$ is the Lagrangian immersion, define respectively bundles $\mathscr{L}_\rho^{\mathbb{R}^4} = \pi^* \mathscr{L}_\rho$ on $\mathbb{R}^4$ and $\mathscr{L}_\rho^L = (\pi \circ i)^* \mathscr{L}_\rho$ on L, away from $\pi^{-1}(c)$ and $(\pi \circ i)^{-1}(c)$, respectively, where ρ can be either ρ_1 or ρ_2.

If ρ_1 is the nontrivial representation, then, since a Möbius strip has $w_1 = 1$, by setting $M = \mathscr{L}_{\rho_1}^{\mathbb{R}^4} \oplus \mathscr{L}_{\rho_1}^{\mathbb{R}^4}$, we have

$$w_1(M) = 2w_1(\mathscr{L}_{\rho_1}^{\mathbb{R}^4}) = 0 \quad \text{and} \quad w_2(M) = 2w_2(\mathscr{L}_{\rho_1}^{\mathbb{R}^4}) + w_1(\mathscr{L}_{\rho_1}^{\mathbb{R}^4}) w_1(\mathscr{L}_{\rho_2}^{\mathbb{R}^4}) = 1.$$

This implies that the bundle $TL \oplus M|_L$ over L carries a spin structure. In fact, since $\mathscr{L}_\rho^L = i^*(\pi^* \mathscr{L}_\rho) = i^*(\mathscr{L}_\rho^{\mathbb{R}^4})$ and so $w_2(\mathscr{L}_\rho^L) = i^* w_2(\mathscr{L}_\rho^{\mathbb{R}^4})$, we have

$$w_2(TL \oplus M|_L) = w_2(TL) + w_1(TL) w_1(M|_L) + w_2(M|_L) = 0$$

in $H^2(L; \mathbb{Z}_2)$. This together with the facts that L has dimension 2 and that M is a real orientable vector bundle on $\mathbb{R}^4$ implies by definition that L has a relative spin structure.

Now consider the flat line bundle $\mathscr{L} \otimes \mathscr{L}_{\rho_2}^L$ over L with connection $\nabla_\rho = \nabla \otimes \nabla_{\rho_2}^L$, where $(\mathscr{L}, \nabla)$ is the given flat line bundle over L and $\nabla_{\rho_2}^L$ is the flat connection of $\mathscr{L}_{\rho_2}^L$ defined by ρ_2, and consider the effect of the connection $\hat{\nabla}_\rho$ on the transformed bundle $\hat{E}$. It induces a nontrivial gluing along A, given by multiplication by -1, which cancels the monodromy along c, given also by multiplication by -1. In fact, if s_1 and s_2 are the saddles and l_1 and l_2 are the sides of the caustic where the node n glues together with s_1 and s_2, respectively, we have according to Definition 6.3 that $(h) \cong [(h, 0)]$ along l_1. In Morse homology we have the equality $[(h, 0)] = [(0, -h)]$ along l_2. According to Definition 6.3, we have $[(0, -h)] \cong (-h)$. Finally, along A, the connection $\hat{\nabla}_\rho$ gives the gluing $(-h) \cong (h)$.

Consider now our case of a perturbed elliptic umbilic. Take a suitable ball U containing a cusp c of $\widetilde{L}$ such that $\widetilde{L} \cap \pi^{-1}(U)$ has two connected components.

For simplicity, suppose that c is the cusp of the caustic where n, s_2 and s_3 glue together. Identify the critical points of the gradient system over x and the points of $\widetilde{L}$ over x. Then of the two components of $\widetilde{L}\cap\pi^{-1}(U)$, one contains s_1 and the other s_2 and s_3. Note that $T\widetilde{L}$ carries a spin structure over the first component but not over the second, where we find the same situation described above for the cusp. So choose ρ so that $\mathscr{L}^{\widetilde{L}}_{\rho_1}$ and $\mathscr{L}^{\widetilde{L}}_{\rho_2}$ are the trivial flat line bundles over the component containing s_1 and the nontrivial one over the component containing s_2 and s_3. As above, set $M=\mathscr{L}^{\widetilde{L}}_{\rho_1}\oplus\mathscr{L}^{\widetilde{L}}_{\rho_1}$. Then $T\widetilde{L}\oplus M|_L$ carries a spin structure on both the components. Moreover, the connection $\widehat{\widehat{\nabla}}_\rho$ on the mirror bundle $\widehat{\widehat{E}}$ induced by the connection $\widetilde{\nabla}_\rho=\widetilde{\nabla}\oplus\widetilde{\nabla}^{\widetilde{L}}_{\rho_2}$ cancels the monodromy of Lemmas 6.13 and 6.14 as we will now explain.

Consider first the case (II) described by Lemma 6.14. Since no gradient line exists from n to s_1, it can be treated as done above for the cusp. We get that the flat connection gives a gluing along A that is multiplication by 1 on chains generated by s_1 and multiplication by -1 on chains generated by s_2 or s_3; this cancels in homology the monodromy of Lemma 6.14.

Consider now case (I) described by Lemma 6.13. The gluing provided by $\widehat{\widehat{\nabla}}_\rho$ must commute with the equivalence among cycles in Morse homology in order to define a gluing in homology, and this is not automatic as in case (II) because of the gradient line from n to s_1. In fact, the connection $\widehat{\widehat{\nabla}}_\rho$ induces a connection on $\partial(\langle n\rangle)=\langle s_1+s_2+s_3\rangle$ characterized by a gluing that is multiplication by -1. On the other hand, the connection on $\sum_{i=1}^3\mathbb{C}[s_i]$ has factor of automorphy -1 on the chains s_2 and s_3 and 1 on s_1. This means that it does not commute with the action on cycles determined by the differential ∂. Thus, to induce a connection in homology, that is, on the quotient $\sum_{i=1}^3\mathbb{C}[s_i]/\partial(\langle n\rangle)$, the connection at the chain level, that is, on $\sum_{i=1}^3\mathbb{C}[s_i]$, must be split into two parts. One of these, commuting with that on $\partial(\langle n\rangle)$, will induce a connection in homology. The problem is the choice of a splitting of the connection at the chains level. This is performed as follows. The gluing $(h_1,h_2,h_3)\cong(h_1,-h_2,-h_3)$ is split as $(h_1,-h_2,-h_3)=(h_1-h_2-h_3,-h_2,-h_3)+(h_2+h_3,0,0)$ and on the quotient the gluing given by $[(h_1,-h_2,-h_3)]=[(h_1-h_2-h_3,-h_2,-h_3)]$ is induced. Indeed, it commutes with the Morse differential:

$$\begin{aligned}(h_1,h_2,h_3)&\cong(h_1+g,h_2+g,h_3+g)\\&\cong(h_1+g-h_2-g-h_3-g,-h_2-g,-h_3-g)\\&=(h_1-h_2-h_3-g,-h_2-g,-h_3-g),\end{aligned}$$

where the first equivalence is that among cycles in Morse homology and the second is the gluing, and

$$(h_1,h_2,h_3)\cong(h_1-h_2-h_3,-h_2,-h_3)\cong(h_1-h_2-h_3-g,-h_2-g,-h_3-g)$$

where now the first equivalence is the gluing and the second is that among cycles in Morse homology. The splitting we chose corresponds to a gluing, at the chain level, given by multiplication by 1 on the generator s_1, and, on the generators s_2 and s_3, by multiplication by -1, followed by a projection parallel to s_1 onto the lines generated by s_3 and s_2, respectively. A better justification for this choice requires, perhaps, considering a more general situation than that of a perturbed elliptic umbilic. Anyway, this solves the monodromy. Indeed, as in the proof of Lemma 6.13, we have along l_2 $(h_1, h_j) \cong [(h_1, h_j, 0)] = [(h_1 - h_j, 0, -h_j)]$, the connection ∇_ρ gives the gluing

$$[(h_1 - h_j, 0, -h_j)] \cong [(h_1 - h_j + h_j, 0, h_j)] = [(h_1, 0, h_j)],$$

and along l_3 we have $[(h_1, 0, h_j)] \cong (h_1, h_j)$.

It remains to check that there is no monodromy in the holomorphic structure of $\widehat{\widehat{E}}$ when going along a loop Γ such that the caustic lies in the compact region of $\mathbb{R}^2$ determined by Γ, as described in Lemma 5.1.

Theorem 6.20. *The monodromy of Lemma 5.1 is solved by the following corrections*: $\widehat{\widehat{E}}$ *is glued by means of the morphisms of Definition 6.3 along the caustic* $\widetilde{K}$, *of Definition 6.10 along the bifurcation locus* $\widetilde{B}$, *and of Definition 6.18 along the relative cycle* A.

Proof. The theorem follows from Propositions 6.11, 6.12 and 6.19. □

As an example, we write the transformation matrices associated to bifurcation lines and to half-lines forming the relative cycles A, which are met by a loop Γ as described above, and show that their composition is the identity, implying that the expected monodromy is canceled. Consider, for instance, the bifurcation diagram of Figure 3.

Assume for simplicity that Γ is directed counterclockwise. Set $a_i = A_i \cap \Gamma$ and $b_i = \widetilde{B}_i \cap \Gamma$, where A_i are the half-lines forming the relative cycle A, and $\widetilde{B}_i$ are the bifurcation lines for $i = 1, 2, 3$. Then the matrices corresponding to gluing morphisms at points a_i and b_i are

$$M(b_1) = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \quad M(b_2) = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, \quad M(b_3) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$
$$M(a_1) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad M(a_2) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad M(a_3) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Observe now that $M(b_3)M(a_3)M(a_2)M(b_2)M(a_1)M(b_1) = \mathrm{Id}$, which implies that the monodromy is solved.

With such corrections, the mirror bundle $\widehat{\widehat{E}}$ is endowed with a holomorphic structure that can be extended along the caustic and the bifurcation locus.

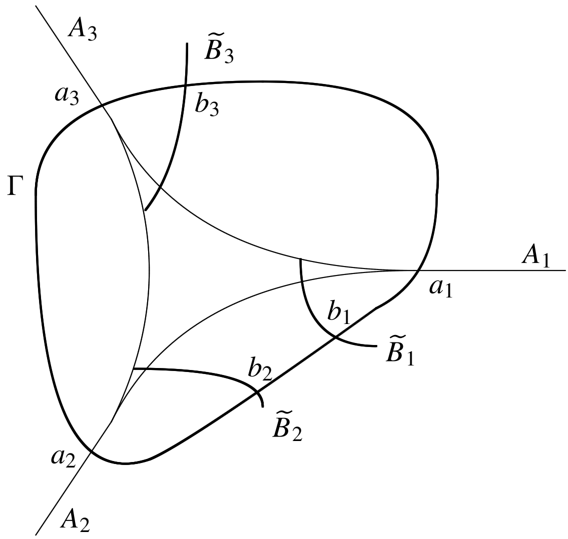


Figure 3. An allowed bifurcation diagram together with the half-cycle A and the loop Γ.

Acknowledgements

I wish to thank K. Fukaya, whose suggestions and help were decisive for achieving the results here.

References

[Arinkin and Polishchuk 2001] D. Arinkin and A. Polishchuk, "Fukaya category and Fourier transform", pp. 261–274 in *Winter School on Mirror Symmetry, Vector Bundles and Lagrangian Submanifolds* (Cambridge, MA, 1999), edited by C. Vafa and S.-T. Yau, AMS/IP Stud. Adv. Math. **23**, Amer. Math. Soc., Providence, RI, 2001. MR 2002k:32040 Zbl 1082.14522

[Bruzzo et al. 2001] U. Bruzzo, G. Marelli, and F. Pioli, "A Fourier transform for sheaves on real tori: I, The equivalence Sky$(T) \simeq$ Loc$(\hat{T})$", *J. Geom. Phys.* **39**:2 (2001), 174–182. MR 2002g:32024 Zbl 1058.14505

[Bruzzo et al. 2002] U. Bruzzo, G. Marelli, and F. Pioli, "A Fourier transform for sheaves on real tori, II: Relative theory", *J. Geom. Phys.* **41**:4 (2002), 312–329. MR 2003a:32035 Zbl 1071.14515

[Floer 1988] A. Floer, "Morse theory for Lagrangian intersections", *J. Differential Geom.* **28**:3 (1988), 513–547. MR 90f:58058 Zbl 0674.57027

[Fukaya 2000] K. Fukaya, "Mirror symmetry of Abelian varieties and multi-theta functions", preprint, 2000, Available at http://www.math.kyoto-u.ac.jp/~fukaya/abelrev.pdf.

[Fukaya 2005] K. Fukaya, "Multivalued Morse theory, asymptotic analysis and mirror symmetry", pp. 205–278 in *Graphs and patterns in mathematics and theoretical physics*, edited by M. Lyubich and L. Takhtajan, Proc. Sympos. Pure Math. **73**, Amer. Math. Soc., Providence, RI, 2005. MR 2006a:53100 Zbl 1085.53080

[Fukaya and Oh 1997] K. Fukaya and Y.-G. Oh, "Zero-loop open strings in the cotangent bundle and Morse homotopy", *Asian J. Math.* **1**:1 (1997), 96–180. MR 99e:58038 Zbl 0938.32009

[Fukaya et al. 2000] K. Fukaya, Y. Oh, H. Ohta, and K. Ono, "Lagrangian intersection Floer Theory: Anomaly and obstruction", preprint, 2000, Available at http://www.math.kyoto-u.ac.jp/~fukaya/fooo.dvi.

[Igusa 1993] K. Igusa, "The Borel regulator map on pictures, I: A dilogarithm formula", *K-Theory* **7**:3 (1993), 201–224. MR 95c:19005a Zbl 0793.19001

[Igusa 2002] K. Igusa, *Higher Franz–Reidemeister torsion*, AMS/IP Studies in Advanced Mathematics **31**, American Mathematical Society, Providence, RI, 2002. MR 2004f:19003 Zbl 1016.19001

[Igusa and Klein 1993] K. Igusa and J. Klein, "The Borel regulator map on pictures, II: An example from Morse theory", *K-Theory* **7**:3 (1993), 225–267. MR 95c:19005b Zbl 0793.19002

[Kazarian 2003] M. Kazarian, "Thom polynomials for Lagrange, Legendre, and critical point function singularities", *Proc. London Math. Soc.* (3) **86**:3 (2003), 707–734. MR 2005a:32025 Zbl 1034.32019

[Kokubu 2000] H. Kokubu, "On transition matrices", pp. 219–224 in *International Conference on Differential Equations* (Berlin, 1999), vol. 1, edited by B. Fiedler et al., World Scientific, River Edge, NJ, 2000. MR 1870126 Zbl 0966.37002

[Leung et al. 2001] N. C. Leung, S.-T. Yau, and E. Zaslow, "From special Lagrangian to Hermitian–Yang–Mills via Fourier-Mukai transform", pp. 209–225 in *Winter School on Mirror Symmetry, Vector Bundles and Lagrangian Submanifolds* (Cambridge, MA, 1999), edited by C. Vafa and S.-T. Yau, AMS/IP Stud. Adv. Math. **23**, Amer. Math. Soc., Providence, RI, 2001. MR 2002k:53102 Zbl 1081.53516

[Marelli 2006a] G. Marelli, "Two-dimensional Lagrangian singularities and bifurcations of gradient lines, I", *J. Geom. Phys.* **56**:9 (2006), 1688–1708. MR 2007f:53099 Zbl 1108.53040

[Marelli 2006b] G. Marelli, "Two-dimensional Lagrangian singularities and bifurcations of gradient lines, II", *J. Geom. Phys.* **56**:9 (2006), 1875–1892. MR 2007f:53100 Zbl 1108.53041

[McDuff and Salamon 2004] D. McDuff and D. Salamon, *J-holomorphic curves and symplectic topology*, American Mathematical Society Colloquium Publications **52**, American Mathematical Society, Providence, RI, 2004. MR 2004m:53154 Zbl 1064.53051

[Schwarz 1993] M. Schwarz, *Morse homology*, Progress in Mathematics **111**, Birkhäuser, Basel, 1993. MR 95a:58022 Zbl 0806.57020

[Vassilyev 1988] V. A. Vassilyev, *Lagrange and Legendre characteristic classes*, Advanced Studies in Contemporary Mathematics **3**, Gordon and Breach, New York, 1988. MR 91k:57034 Zbl 0715.53001

Received December 21, 2008. Revised May 31, 2010.

GIOVANNI MARELLI
KOREA INSTITUTE FOR ADVANCED STUDY
HOEGIRO 87
DONGDAEMUN-GU
SEOUL 130-722
SOUTH KOREA
marelli@kias.re.kr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

CROSSED POINTED CATEGORIES AND THEIR EQUIVARIANTIZATIONS

DEEPAK NAIDU

We propose a notion, quasiabelian third cohomology of crossed modules, which generalizes Eilenberg and Mac Lane's abelian and Ospel's quasiabelian cohomology. We classify crossed pointed categories in terms of it. We apply the process of equivariantization to the latter to obtain braided fusion categories, which may be viewed as generalizations of the categories of modules over twisted Drinfeld doubles of finite groups. As a consequence, we obtain a description of all braided group-theoretical categories. We give a criterion for these categories to be modular. We describe the quasitriangular quasi-Hopf algebras underlying these categories.

1. Introduction

Turaev's notion [2000; 2008] of a crossed category (short for braided group-crossed category) has attracted much attention recently [Drinfeld et al. 2010; Kirillov 2001a; 2001b; Müger 2004; 2005]. Roughly, a crossed category consists of a group G, a G-graded tensor category $\mathscr{C}$, an action $g \mapsto T_g$ of G on $\mathscr{C}$ by tensor autoequivalences, and G-braidings $c(X, Y) : X \otimes Y \xrightarrow{\sim} T_g(Y) \otimes X$ for $X, Y \in \mathscr{C}$, satisfying certain compatibility conditions. Crossed categories are known to arise in various contexts; for instance, Müger [2004] showed that Galois extensions of braided tensor categories have a natural structure of crossed categories. In [2005], Müger established a connection between 1-dimensional quantum field theories and crossed categories. Kirillov [2001b] showed that crossed categories arise in the theory of vertex operator algebras.

A fusion category is said to be *pointed* if all its simple objects are invertible. One of the goals of this paper is to classify all crossed pointed categories. From [Joyal and Street 1993], it is known that braided pointed categories are classified by Eilenberg and Mac Lane's abelian cohomology $\mathrm{H}^3_{\mathrm{ab}}(A, \mathbb{K}^\times)$, where A is a finite abelian group. On the other hand, certain crossed pointed categories in which the group action is strict were described by Turaev [2000; 2008] in terms of Ospel's quasiabelian cohomology $\mathrm{H}^3_{\mathrm{qa}}(G, \mathbb{K}^\times)$, where G is a (not necessarily abelian) finite

MSC2000: primary 18D10; secondary 16W30.
Keywords: braided crossed G-categories, equivariantization, group-theoretical fusion categories.

group. As remarked in [Müger 2005, Section 4.9], to obtain a complete classification of crossed pointed categories one must allow for nonstrict group actions. To this end, Definition 3.4 generalizes Ospel's quasiabelian cohomology to the notion of quasiabelian third cohomology $\mathrm{H}^3_{\mathrm{qa}}(\mathcal{X}, \mathbb{K}^\times)$ of a crossed module $\mathcal{X}$. To any given $\xi \in Z^3_{\mathrm{qa}}(\mathcal{X}, \mathbb{K}^\times)$, we associate a crossed pointed category $\mathcal{C}(\xi)$ and show that all crossed pointed categories are of this form.

Another idea that has been studied extensively recently is that of a modular category. Examples of modular categories arise in quantum group theory, three-dimensional topology, vertex operator algebras and rational conformal field theory. Let G be a finite group. Perhaps the most accessible construction of a modular category is that of the category of modules over the Drinfeld double $D(G)$ of G. Let ω be a 3-cocycle on G. In [1990; 1992], Dijkgraaf, Pasquier, and Roche introduced a quasitriangular quasi-Hopf algebra $D^\omega(G)$, generalizing the Drinfeld double $D(G)$. It is well known that the category $D^\omega(G)$-Mod of modules over $D^\omega(G)$ is a modular category. Modular categories resembling $D^\omega(G)$-Mod arise naturally from crossed pointed categories. An important feature of a general crossed fusion category is that the application of the equivariantization process (which is analogous to taking the invariants under a group action) yields a braided fusion category. We apply the equivariantization process to the crossed pointed category $\mathcal{C}(\xi)$ and study the resulting braided fusion category, which resembles the category $D^\omega(G)$-Mod. As a consequence, we obtain a description of *all* braided group-theoretical categories. In Proposition 5.6, we show that $\mathcal{C}(\xi)$ is modular if and only if ξ is nondegenerate in the sense of Definition 3.10 and a certain homomorphism is surjective.

By a general result, the equivariantization of the category $\mathcal{C}(\xi)$ is equivalent as a braided fusion category to the category of modules over some finite-dimensional quasitriangular quasi-Hopf algebra H. In the sequel we describe such an H. Namely, given $\xi \in Z^3_{\mathrm{qa}}(\mathcal{X}, \mathbb{K}^\times)$, we construct a finite-dimensional quasitriangular quasi-Hopf algebra $H(\xi)$, generalizing $D^\omega(G)$, and show that $\mathcal{C}(\xi) \cong H(\xi)$-Mod, as braided fusion categories.

Outline. Section 2 recalls essential definitions and results about nondegenerate fusion categories, equivariantization, and crossed categories. Section 3 proposes the notion of quasiabelian third cohomology of crossed modules. In Section 4, we construct crossed pointed categories from quasiabelian 3-cocycles and classify the former. In Section 5, we apply the process of equivariantization to the categories obtained in Section 4 and study the resulting braided fusion categories. In Section 6, we construct finite-dimensional quasitriangular quasi-Hopf algebras from quasiabelian 3-cocycles and show that these underlie the braided fusion categories obtained in Section 5.

2. Preliminaries

We will freely use the language and basic theory of fusion categories and modular categories [Bakalov and Kirillov 2001; Ostrik 2003; Etingof et al. 2005].

2a. ***Conventions.*** Let $\mathbb{K}$ be an algebraically closed field of characteristic 0. The multiplicative group of nonzero elements of $\mathbb{K}$ will be denoted by $\mathbb{K}^\times$. Unless otherwise stated, all cocycles will have coefficients in the trivial module $\mathbb{K}^\times$. All functors will be assumed to be additive and $\mathbb{K}$-linear on the morphism spaces. The unit object of a tensor category will be denoted by $\mathbb{1}$. The identity element of a group will be denoted by e.

2b. ***Morita equivalence.*** Following [Müger 2003a], we say that two fusion categories $\mathcal{C}$ and $\mathcal{D}$ are Morita equivalent if $\mathcal{D}$ is equivalent to the dual fusion category $\mathcal{C}^*_{\mathcal{M}}$ for some indecomposable right $\mathcal{C}$-module category $\mathcal{M}$; see also [Etingof et al. 2005; Ospel 1999]. This is known to be an equivalence relation on the class of fusion categories. A fusion category is said to be *pointed* if all its simple objects are invertible. A fusion category is *group-theoretical* if it is Morita equivalent to a pointed category.

2c. ***Nondegenerate fusion categories.*** Let $\mathcal{C}$ be a braided fusion category with braiding c. Following [Müger 2003b], we say two objects X and Y of $\mathcal{C}$ *centralize* each other if $c(Y, X) \circ c(X, Y) = \mathrm{id}_{X\otimes Y}$.

The *centralizer* of a fusion subcategory $\mathcal{D} \subseteq \mathcal{C}$ is the full fusion subcategory $\mathcal{D}'$ of $\mathcal{C}$ consisting of all objects $X \in \mathcal{C}$ that centralize every object in $\mathcal{D}$. The category $\mathcal{C}$ is said to be *nondegenerate* if $\mathcal{C}' = \mathrm{Vec}$ (the fusion category generated by the unit object). If $\mathcal{C}$ is a premodular category, that is, if it has a twist, then it is nondegenerate if and only if it is modular [Beliakova and Blanchet 2001; Müger 2003b; Drinfeld et al. 2010].

Proposition 2.1. *Let $\mathcal{C}$ be a nondegenerate fusion category. Suppose $\mathcal{C}$ admits a twist. Then the set of twists on $\mathcal{C}$ is in bijection with the set of invertible self-dual objects of $\mathcal{C}$.*

Proof. Let $\mathrm{Aut}_\otimes(\mathrm{id}_\mathcal{C})$ denote the group of tensor automorphisms of the identity tensor functor $\mathrm{id}_\mathcal{C}$. Let $\mathrm{Aut}^*_\otimes(\mathrm{id}_\mathcal{C}) := \{\varphi \in \mathrm{Aut}_\otimes(\mathrm{id}_\mathcal{C}) \mid \varphi_{X^*} = (\varphi_X)^* \text{ for all } X \in \mathcal{C}\}$. Let θ be a fixed twist on $\mathcal{C}$. The map $\varphi \mapsto \theta_\varphi$ defined by $(\theta_\varphi)_X := \theta_X \circ \varphi_X$ for all $X \in \mathcal{C}$ is a bijection from $\mathrm{Aut}^*_\otimes(\mathrm{id}_\mathcal{C})$ to the set of all twists on $\mathcal{C}$.

Let $X_1, X_2, \dots$ denote the simple objects of $\mathcal{C}$, and let $G(\mathcal{C})$ denote the group of invertible objects of $\mathcal{C}$. Let S denote the S-matrix of $\mathcal{C}$ with respect to θ. It was shown in [Gelaki and Nikshych 2008] that the map

$$G(\mathcal{C}) \to \mathrm{Aut}_\otimes(\mathrm{id}_\mathcal{C}), \quad X_j \mapsto \varphi_j, \quad \text{where } (\varphi_j)_{X_i} := \frac{S_{ij}}{d(X_i)d(X_j)}\,\mathrm{id}_{X_i},$$

is an isomorphism. It is easy to check that this map restricts to a bijection between the set of invertible self-dual objects of $\mathcal{C}$ and the set $\mathrm{Aut}^*_{\otimes}(\mathrm{id}_{\mathcal{C}})$. □

2d. *Equivariantization.* Recall that a tensor functor between two tensor categories $\mathcal{C}$ and $\mathcal{D}$ is a triple (F, φ, φ_0), where $F : \mathcal{C} \to \mathcal{D}$ is a functor, φ is a natural isomorphism $F \circ \otimes_{\mathcal{C}} \xrightarrow{\sim} \otimes_{\mathcal{D}} \circ (F \times F)$, and φ_0 is an isomorphism $F(\mathbb{1}_{\mathcal{C}}) \xrightarrow{\sim} \mathbb{1}_{\mathcal{D}}$ satisfying certain compatibility conditions; see [Kassel 1995]. We will call φ the *tensor structure* on F and φ_0 the *unit-preserving structure* on F. For a group G, we will denote by $\underline{G}$ the tensor category whose objects are elements of G, whose morphisms are the identities, and whose tensor product is given by the group operation in G.

Let $\mathcal{C}$ be a fusion category with an action of a finite group G given by a tensor functor $T : \underline{G} \to \mathrm{Aut}_{\otimes}(\mathcal{C})$, $g \mapsto T_g$. Let γ be the tensor structure on the functor T. In this situation one can define a *G-equivariant object* in $\mathcal{C}$ to be a pair $(X, \{u_g\}_{g\in G})$ in which X is an object of $\mathcal{C}$ and

$$u_g : T_g(X) \xrightarrow{\sim} X \quad \text{for } g \in G, \tag{1}$$

is a family of isomorphisms, called the *equivariant structure on* X, such that

$$u_{gh} = u_g \circ T_g(u_h) \circ \gamma_{g,h}(X) \quad \text{for all } g, h \in G. \tag{2}$$

One defines morphisms between equivariant objects to be morphisms in $\mathcal{C}$ that commute with the equivariant structures. The *equivariantization* $\mathcal{C}^G$ *of* $\mathcal{C}$ is the category of G-equivariant objects of $\mathcal{C}$ [Kirillov 2001a; Arkhipov and Gaitsgory 2003; Gaitsgory 2005; Tambara 2001]. The equivariantization category $\mathcal{C}^G$ is a fusion category with tensor product defined by

$$(X, \{u_g\}_{g\in G}) \otimes (X', \{u'_g\}_{g\in G}) := (X \otimes X', \{\tilde{u}_g\}_{g\in G})$$

for $(X, \{u_g\}_{g\in G}), (X', \{u'_g\}_{g\in G}) \in \mathcal{C}^G$, where

$$\tilde{u}_g := (u_g \otimes u'_g) \circ \mu_g(X, X') \tag{3}$$

for all $g \in G$. Here μ_g is the tensor structure on the functor T_g, $g \in G$.

Remark 2.2. We have $\mathrm{FPdim}(\mathcal{C}^G) = |G|\, \mathrm{FPdim}(\mathcal{C})$.

2e. *Crossed categories.* Recall that a *grading* of a fusion category $\mathcal{C}$ by a finite group G is a decomposition $\mathcal{C} = \bigoplus_{g\in G} \mathcal{C}_g$ of $\mathcal{C}$ into a direct sum of full abelian subcategories such that $\otimes$ maps $\mathcal{C}_g \times \mathcal{C}_h$ to $\mathcal{C}_{gh}$ and $*$ maps $\mathcal{C}_g$ to $\mathcal{C}_{g^{-1}}$ for all $g, h \in G$. Note that $\mathcal{C}_e$, called the *trivial component*, is a fusion subcategory of $\mathcal{C}$. A grading is said to *faithful* if $\mathcal{C}_g \neq 0$ for all $g \in G$.

Below we recall the notion of a *crossed category* (short for *braided group-crossed category*), introduced by Turaev [2000; 2008] in a more general form; see also [Drinfeld et al. 2010; Müger 2004; 2005].

Definition 2.3. A *crossed fusion category* is an octuple $(\mathscr{C}, G, T, \gamma, \iota, \mu, \nu, c)$ in which

- G is a finite group;
- $\mathscr{C}$ is a fusion category with (not necessarily faithful) G-grading $\mathscr{C} = \bigoplus_{g\in G} \mathscr{C}_g$;
- $T : \underline{G} \to \mathrm{Aut}_\otimes(\mathscr{C})$, $g \mapsto T_g$ is a tensor functor satisfying $T_g(\mathscr{C}_h) \subset \mathscr{C}_{ghg^{-1}}$, with tensor structure γ and unit-preserving structure ι;
- μ is a family $\{\mu_g\}_{g\in G}$, where μ_g is a tensor structure on T_g;
- ν is a family $\{\nu_g\}_{g\in G}$, where ν_g is a unit-preserving structure on T_g;
- $c(X, Y) : X \otimes Y \xrightarrow{\sim} T_g(Y) \otimes X$ for $X \in \mathscr{C}_g$ and $Y \in \mathscr{C}$ is a family of natural isomorphisms, called *G-braiding*;

and the following compatibility conditions are satisfed:

(i) $$(\gamma_{g,h}(Y) \otimes \mathrm{id}_{T_g(X)}) \circ (\gamma^{-1}_{ghg^{-1},g}(Y) \otimes \mathrm{id}_{T_g(X)}) \circ c(T_g(X), T_g(Y)) \circ \mu_g(X, Y) = \mu_g(T_h(Y), X) \circ T_g(c(X, Y))$$

for all $g, h \in G$ and objects $X \in \mathscr{C}_h$ and $Y \in \mathscr{C}$.

(ii) $$\alpha^{-1}_{T_g(T_h(Z)),X,Y} \circ (\gamma_{g,h}(Z) \otimes \mathrm{id}_{X\otimes Y}) \circ c(X \otimes Y, Z) \circ \alpha^{-1}_{X,Y,Z} = (c(X, T_h(Z)) \otimes \mathrm{id}_Y) \circ \alpha^{-1}_{X,T_h(Z),Y} \circ (\mathrm{id}_X \otimes c(Y, Z)),$$

for all $g, h \in G$ and objects $X \in \mathscr{C}_g$, $Y \in \mathscr{C}_h$ and $Z \in \mathscr{C}$.

(iii) $$\alpha_{T_g(Y),T_g(Z),X} \circ (\mu_g(Y, Z) \otimes \mathrm{id}_X) \circ c(X, Y \otimes Z) \circ \alpha_{X,Y,Z} = (\mathrm{id}_{T_g(Y)} \otimes c(X, Z)) \circ \alpha_{T_g(Y),X,Z} \circ (c(X, Y) \otimes \mathrm{id}_Z),$$

for all $g \in G$ and objects $X \in \mathscr{C}_g$ and $Y, Z \in \mathscr{C}$.

(Here α denotes the associativity constraint of $\mathscr{C}$.)

Remark 2.4. The trivial component of a crossed fusion category is a braided fusion category.

Now let $\mathscr{C} := (\mathscr{C}, G, T, \gamma, \iota, \mu, \nu, c)$ be a crossed fusion category. Kirillov [2001a] and Müger [2004] explain that the equivariantization category $\mathscr{C}^G$ admits a braiding, that is, $\mathscr{C}^G$ is a braided fusion category. The braiding $\tilde{c}$ on $\mathscr{C}^G$ is defined as follows. Let $(X, \{u_g\}_{g\in G})$ and $(X', \{u'_g\}_{g\in G})$ be objects of $\mathscr{C}^G$. Let $X = \bigoplus_{g\in G} X_g$ be a decomposition of X with respect to the G-grading of $\mathscr{C}$. Then $\tilde{c}_{X,X'}$ is given by the composition

(4) $$X \otimes X' = \bigoplus_{g\in G} X_g \otimes X' \xrightarrow{\bigoplus c_{X_g,X'}} \bigoplus_{g\in G} T_g(X') \otimes X_g \xrightarrow{\bigoplus u'_g \otimes \mathrm{id}_{X_g}} \bigoplus_{g\in G} X' \otimes X_g = X' \otimes X.$$

Remark 2.5. It is shown in [Drinfeld et al. 2010] that the equivariantization category $\mathcal{C}^G$ is nondegenerate if and only if the G-grading is faithful and the trivial component $\mathcal{C}_e$ is nondegenerate.

Definition 2.6. Consider two crossed fusion categories $\mathcal{C} = (\mathcal{C}, G, T, \gamma, \iota, \mu, \nu, c)$ and $\mathcal{C}' = (\mathcal{C}', G', T', \gamma', \iota', \mu', \nu', c')$. A *crossed tensor functor* from $\mathcal{C}$ to $\mathcal{C}'$ is a quintuple $(f, F, \eta, \eta_0, \beta)$ in which

- $f : G \to G'$ is a group homomorphism,
- $F : \mathcal{C} \to \mathcal{C}'$ is a tensor functor with tensor structure η and unit-preserving structure η_0, and
- β is a family $\{\beta_g\}_{g\in G}$, where $\beta_g : F \circ T_g \xrightarrow{\sim} T'_{f(g)} \circ F$ is an isomorphism of tensor functors,

and the following compatibility conditions are satisfed:

(i) $F(\mathcal{C}_g) \subseteq \mathcal{C}'_{f(g)}$ for all $g \in G$.

(ii) $(\beta_g(Y) \otimes \mathrm{id}_{F(X)}) \circ \eta(T_g(Y), X) \circ F(c(X, Y)) = c'(F(X), F(Y)) \circ \eta(X, Y)$ for all $g \in G$ and objects $X \in \mathcal{C}_g$ and $Y \in \mathcal{C}$.

(iii) $T'_{f(g)}(\beta_h(X)) \circ \beta_g(T_h(X)) \circ F(\gamma_{g,h}(X)) = \gamma'_{f(g),f(h)}(F(X)) \circ \beta_{gh}(X)$ for all $g, h \in G$ and objects $X \in \mathcal{C}$.

We say that $(f, F, \eta, \eta_0, \beta)$ is an equivalence if f is an isomorphism and F is an equivalence.

2f. *Pointed categories.* A fusion category is said to be pointed if all its simple objects are invertible.

Let X be a finite group and ω be a 3-cocycle on X. We associate to the pair (X, ω) a pointed category Vec_X^ω whose objects are X-graded finite-dimensional vector spaces over $\mathbb{K}$, whose morphisms are linear transformations that respect the grading, and whose unit object is the ground field $\mathbb{K}$ supported on $\{e\}$. The tensor product $V \otimes W$ of homogeneous objects $V, W \in \mathrm{Vec}_X^\omega$ of degrees $x, y \in X$, respectively, is defined to be the homogeneous object $V \otimes_{\mathbb{K}} W$ of degree xy.

The associativity constraint α is defined by

$$\alpha_{U,V,W} : (U \otimes V) \otimes W \xrightarrow{\sim} U \otimes (V \otimes W), \quad (u \otimes v) \otimes w \mapsto \omega(x, y, z) u \otimes (v \otimes w),$$

where $U, V, W \in \mathrm{Vec}_G^\omega$ and $u \in U$, $v \in V$, $w \in W$ are homogeneous elements of degrees $x, y, z \in X$, respectively.

The left and right unit constraints λ and ρ, respectively, are defined by

$$\lambda_V := \mathbb{K} \otimes V \xrightarrow{\sim} V, \quad 1 \otimes v \mapsto \omega(e, e, x)^{-1} v,$$
$$\rho_V := V \otimes \mathbb{K} \xrightarrow{\sim} V, \quad v \otimes 1 \mapsto \omega(x, e, e) v,$$

where $V \in \mathrm{Vec}_G^\omega$ and $v \in V$ is a homogeneous element of degree $x \in X$.

Every pointed category is equivalent to some Vec_X^{ω}.

2g. *Crossed modules.* Recall that a (*finite*) *crossed module* is a triple (G, X, ∂), where G and X are (finite) groups with G acting on X as automorphisms, denoted $(g, x) \mapsto {}^{g}x$, and $\partial : X \to G$ is a homomorphism satisfying

$$\begin{aligned} {}^{\partial(x)}x' &= xx'x^{-1} && \text{for } x, x' \in X, \\ \partial({}^{g}x) &= g\partial(x)g^{-1} && \text{for } g \in G \text{ and } x \in X. \end{aligned}$$

Note that $\operatorname{Ker}\partial$ is a central subgroup of X.

A homomorphism of crossed modules $(G, X, \partial) \to (G', X', \partial')$ is a pair of group homomorphisms $(f : G \to G', F : X \to X')$ such that $\partial' \circ F = f \circ \partial$ and $F({}^{g}x) = {}^{f(g)}F(x)$ for $g \in G$. We say that (f, F) is an *isomorphism* if both f and F are isomorphisms.

3. Quasiabelian third cohomology of crossed modules

Let A be an abelian group. Eilenberg and Mac Lane [Eilenberg and Mac Lane 1953; 1954; Mac Lane 1952] argue that the cohomology groups $\mathrm{H}^n(A, \mathbb{K}^\times)$ are inappropriate since they do not take into account the abelianness of A, and so should be replaced by groups $\mathrm{H}^n_{\mathrm{ab}}(A, \mathbb{K}^\times)$. (For the cohomology theory for crossed modules, see [Whitehead 1949].) Below we recall the definition of $\mathrm{H}^3_{\mathrm{ab}}(A, \mathbb{K}^\times)$.

An *abelian* 3-*cocycle* on A is a pair (ω, c), where ω is a normalized 3-cocycle on A, that is, for all $w, x, y, z \in A$,

$$\begin{aligned} &\omega(x, y, z) = 1 \quad \text{if } x, y \text{ or } z \text{ is the identity}, \\ &\omega(x, y, z)\omega(w, xy, z)\omega(w, x, y) = \omega(w, x, yz)\omega(wx, y, z) \end{aligned}$$

and c is a 2-cochain on A (that is, $c \in C^2(A, \mathbb{K}^\times)$) satisfying the equations

$$\begin{aligned} c(xy, z) &= \frac{\omega(x, y, z)\omega(z, x, y)}{\omega(x, z, y)} c(x, z)c(y, z), \\ c(x, yz) &= \frac{\omega(y, x, z)}{\omega(x, y, z)\omega(y, z, x)} c(x, y)c(x, z) \end{aligned}$$

for all $x, y, z \in A$.

Abelian 3-cocycles on A form an abelian group, denoted by $Z^3_{\mathrm{ab}}(A, \mathbb{K}^\times)$, under pointwise multiplication. The group of coboundaries is defined by

$$B^3_{\mathrm{ab}}(A, \mathbb{K}^\times) := \left\{ \left(d\eta, (x, y) \mapsto \frac{\eta(y, x)}{\eta(x, y)} \right) \,\middle|\, \text{normalized } \eta \in C^2(G, \mathbb{K}^\times) \right\},$$

which is a subgroup of $Z^3_{\mathrm{ab}}(A, \mathbb{K}^\times)$. The quotient $Z^3_{\mathrm{ab}}(A, \mathbb{K}^\times)/B^3_{\mathrm{ab}}(A, \mathbb{K}^\times)$ is the *abelian third cohomology* of A, denoted $\mathrm{H}^3_{\mathrm{ab}}(A, \mathbb{K}^\times)$.

Remark 3.1. The group $\mathrm{H}^3_{\mathrm{ab}}(A, \mathbb{K}^\times)$ is isomorphic to the group of quadratic forms on A; see [Mac Lane 1952].

Definition 3.2. We say an abelian 3-cocycle (ω, c) on A is *nondegenerate* if the symmetric bicharacter $A \times A \to \mathbb{K}^\times$, $(x, y) \mapsto c(y, x)c(x, y)$ is nondegenerate.

In [1999], C. Ospel generalized the notion of abelian third cohomology in the following way. Let G be a (not necessarily abelian) group. A *quasiabelian* 3-*cocycle* on G is a pair (ω, c), where ω is a 3-cocycle on G and c is a 2-cochain on G (that is, $c \in C^2(G, \mathbb{K}^\times)$) satisfying for all $g, x, y, z \in G$ the equations

$$\begin{aligned}
\omega(gxg^{-1}, gyg^{-1}, gzg^{-1}) &= \omega(x, y, z),\\
c(gxg^{-1}, gyg^{-1}) &= c(x, y),\\
c(xy, z) &= \frac{\omega(x, y, z)\omega(xyz(xy)^{-1}, x, y)}{\omega(x, yzy^{-1}, y)} c(x, yzy^{-1})c(y, z),\\
c(x, yz) &= \frac{\omega(xyx^{-1}, x, z)}{\omega(x, y, z)\omega(y, z, x)} c(x, y)c(x, z),
\end{aligned}$$

Note 3.3. The third equation above appeared in a slightly different but equivalent form in [Ospel 1999].

Quasiabelian 3-cocycles on G form an abelian group, denoted by $Z^3_{\mathrm{qa}}(G, \mathbb{K}^\times)$, under pointwise multiplication. The group of coboundaries is defined by

$$B^3_{\mathrm{qa}}(G, \mathbb{K}^\times) := \left\{ \left(d(\eta), (x, y) \mapsto \frac{\eta(y, x)}{\eta(x, y)} \right) \,\middle|\, \text{conjugation-invariant } \eta \in C^2(G, \mathbb{K}^\times) \right\},$$

which is a subgroup of $Z^3_{\mathrm{qa}}(G, \mathbb{K}^\times)$. The quotient $Z^3_{\mathrm{qa}}(G, \mathbb{K}^\times)/B^3_{\mathrm{qa}}(G, \mathbb{K}^\times)$ is the *quasiabelian third cohomology* of G, denoted $\mathrm{H}^3_{\mathrm{qa}}(G, \mathbb{K}^\times)$. If G is abelian, quasiabelian cohomology reduces to abelian cohomology: $\mathrm{H}^3_{\mathrm{ab}}(G, \mathbb{K}^\times) = \mathrm{H}^3_{\mathrm{qa}}(G, \mathbb{K}^\times)$.

We can extend Ospel's quasiabelian cohomology for groups to cover crossed modules: We allow G to act on an arbitrary group X (not just $X = G$). The first condition, $\omega^g = \omega$, in Ospel's definition is replaced by the condition that ω^g is cohomologous to ω via μ_g. The second condition, $c^g = c$, is extended similarly, as are the others. This results in the following definition, whose main motivation is the classification of crossed pointed categories (see Section 4).

Definition 3.4. A *quasiabelian* 3-*cocycle* on a crossed module $\mathscr{X} = (G, X, \partial)$ is a quadruple (ω, γ, μ, c), where

(a) $\omega \in Z^3(X, \mathbb{K}^\times)$,

(b) $\gamma \in Z^2(G, C^1(X, \mathbb{K}^\times))$,

(c) $\mu \in C^1(G, C^2(X, \mathbb{K}^\times))$ and satisfies $d(\mu_g) = \omega^g/\omega$ for $g \in G$, that is,

$$\frac{\mu_g(y,z)\mu_g(x,yz)}{\mu_g(xy,z)\mu_g(x,y)} = \frac{\omega^g(x,y,z)}{\omega(x,y,z)} \quad \text{for } g \in G \text{ and } x, y, z \in X,$$

(d) $d(\gamma_{g,h}) = (d\mu)_{g,h}$ for $g, h \in G$, that is,

$$\frac{\gamma_{g,h}(x)\gamma_{g,h}(y)}{\gamma_{g,h}(xy)} = \frac{\mu_g({}^h x, {}^h y)\mu_h(x,y)}{\mu_{gh}(x,y)} \quad \text{for } g, h \in G \text{ and } x, y \in X,$$

(e) $c \in C^2(X, \mathbb{K}^\times)$ and satisfies

$$(\mathrm{e}_1) \quad \frac{c^g(x,y)}{c(x,y)} = \frac{\mu_g(xyx^{-1},x)}{\mu_g(x,y)} \frac{\gamma_{g\partial(x)g^{-1},g}(y)}{\gamma_{g,\partial(x)}(y)} \qquad \text{for } g \in G, x, y \in X,$$

$$(\mathrm{e}_2) \quad c(xy,z) = \frac{\omega(x,y,z)\omega((xy)z(xy)^{-1},x,y)}{\omega(x,yzy^{-1},y)\gamma_{\partial(x),\partial(y)}(z)} c(x,yzy^{-1})c(y,z) \qquad \text{for } x, y, z \in X,$$

$$(\mathrm{e}_3) \quad c(x,yz) = \frac{\omega(xyx^{-1},x,z)}{\omega(x,y,z)\omega(xyx^{-1},xzx^{-1},x)\mu_{\partial(x)}(y,z)} c(x,y)c(x,z) \qquad \text{for } x, y, z \in X.$$

Note 3.5. Here C^n denotes the space of n-cochains, Z^n denotes the space of n-cocycles, and d is the usual *differential operator* [Brown 1982]. (The definition of d depends on whether the module under consideration is left or right.) The action $(g, x) \mapsto {}^g x$ of G on X induces a right action of G on $C^n(X, \mathbb{K}^\times)$ by translations. The map $c^g \in C^2(X, \mathbb{K}^\times)$ is defined by $c^g(x, y) := c({}^g x, {}^g y)$ and the map ω^g is defined similarly.

Quasiabelian 3-cocycles on a crossed module $\mathscr{X} = (G, X, \partial)$ form an abelian group $Z^3_{\mathrm{qa}}(\mathscr{X}, \mathbb{K}^\times)$ under pointwise multiplication.

We define the group of coboundaries by

$$B^3_{\mathrm{qa}}(\mathscr{X}, \mathbb{K}^\times) := \left\{ \left(d\eta, d\beta, g \mapsto d(\beta_g)\frac{\eta^g}{\eta}, (x,y) \mapsto \beta_{\partial(x)}(y)\frac{\eta(xyx^{-1},x)}{\eta(x,y)} \right) \;\middle|\; \begin{array}{l} \eta \in C^2(X, \mathbb{K}^\times), \\ \beta \in C^1(G, C^1(X, \mathbb{K}^\times)) \end{array} \right\}.$$

A direct computation shows that $B^3_{\mathrm{qa}}(\mathscr{X}, \mathbb{K}^\times) \subseteq Z^3_{\mathrm{qa}}(\mathscr{X}, \mathbb{K}^\times)$.

Definition 3.6. The *quasiabelian third cohomology* of a crossed module $\mathscr{X}$ is the quotient of $Z^3_{\mathrm{qa}}(\mathscr{X}, \mathbb{K}^\times)$ by $B^3_{\mathrm{qa}}(\mathscr{X}, \mathbb{K}^\times)$. We denote it by $\mathrm{H}^3_{\mathrm{qa}}(\mathscr{X}, \mathbb{K}^\times)$.

Remark 3.7. Let G be a group. Consider the crossed module $\mathscr{G} = (G, G, \mathrm{id}_G)$, where G acts on itself by conjugation.

(i) There is a homomorphism $\mathrm{H}^3_{\mathrm{qa}}(G, \mathbb{K}^\times) \to \mathrm{H}^3_{\mathrm{qa}}(\mathscr{G}, \mathbb{K}^\times)$ induced from

$$Z^3_{\mathrm{qa}}(G, \mathbb{K}^\times) \to Z^3_{\mathrm{qa}}(\mathscr{G}, \mathbb{K}^\times), \quad (\omega, c) \mapsto (\omega, 1, 1, c).$$

(ii) There exists a homomorphism $\mathrm{H}^3(G, \mathbb{K}^\times) \to \mathrm{H}^3_{\mathrm{qa}}(\mathscr{G}, \mathbb{K}^\times)$; see Lemma 6.3.

Definition 3.8. A quasiabelian 3-cocycle (ω, γ, μ, c) is normalized if

$$\omega(x, y, z) = 1 \text{ if } x, y \text{ or } z \text{ is the identity}, \quad \gamma_{g,h}(x) = 1 \text{ if } g, h \text{ or } x \text{ is the identity},$$
$$\mu_g(x, y) = 1 \text{ if } x, y \text{ or } g \text{ is the identity}, \quad c(x, y) = 1 \text{ if } x \text{ or } y \text{ is the identity}.$$

Note 3.9. Every quasiabelian 3-cocycle is cohomologous to a normalized one.

Let (ω, γ, μ, c) be a normalized quasiabelian 3-cocycle on a crossed module (G, X, ∂). Then $(\omega|_{\mathrm{Ker}\,\partial}, c|_{\mathrm{Ker}\,\partial})$ is an abelian 3-cocycle on the (abelian) group $\mathrm{Ker}\,\partial$.

Definition 3.10. A normalized quasiabelian 3-cocycle (ω, γ, μ, c) on a crossed module (G, X, ∂) is *nondegenerate* if the abelian 3-cocycle $(\omega|_{\mathrm{Ker}\,\partial}, c|_{\mathrm{Ker}\,\partial})$ on the (abelian) group $\mathrm{Ker}\,\partial$ is nondegenerate.

Any homomorphism $(f, F) : (G', X', \partial') = \mathscr{X}' \to \mathscr{X} = (G, X, \partial)$ of crossed modules induces a homomorphism

$$Z^3_{\mathrm{qa}}(\mathscr{X}, \mathbb{K}^\times) \to Z^3_{\mathrm{qa}}(\mathscr{X}', \mathbb{K}^\times), \quad (\omega, \gamma, \mu, c) \mapsto (\omega, \gamma, \mu, c)^{(f,F)},$$

where

$$(\omega, \gamma,\ \mu, c)^{(f,F)} = (\omega \circ F^{\times 3},\ (g, h) \mapsto \gamma_{f(g), f(h)} \circ F,\ g \mapsto \mu_{f(g)} \circ F^{\times 2}, c \circ F^{\times 2}).$$

It is straightforward to check that this homomorphism preserves coboundaries and thereby provides a homomorphism $\mathrm{H}^3_{\mathrm{qa}}(\mathscr{X}, \mathbb{K}^\times) \to \mathrm{H}^3_{\mathrm{qa}}(\mathscr{X}', \mathbb{K}^\times)$. Consequently, for any crossed module $\mathscr{X}$ there is a natural action of the group of automorphisms $\mathrm{Aut}(\mathscr{X})$ of $\mathscr{X}$ on $\mathrm{H}^3_{\mathrm{qa}}(\mathscr{X}, \mathbb{K}^\times)$.

4. Classification of crossed pointed categories

In this section we classify crossed pointed categories in terms of quasiabelian third cohomology of crossed modules.

4a. ***Construction of a crossed pointed category from a quasiabelian 3-cocyle on a crossed module.*** To a quasiabelian 3-cocycle (ω, γ, μ, c) on a finite crossed module (G, X, ∂), we associate a crossed pointed category $(\mathscr{C}, G, T, \tilde{\gamma}, \iota, \tilde{\mu}, \nu, \tilde{c})$ as follows. As a fusion category, $\mathscr{C} = \mathrm{Vec}^\omega_X$. For each $g \in G$, let $\mathscr{C}_g$ denote the full abelian subcategory consisting of objects of Vec^ω_X supported on $\partial^{-1}(g) \subset X$, that is, objects of $\mathscr{C}_g$ are defined to be finite-dimensional $\partial^{-1}(g)$-graded vector spaces (we set $\mathscr{C}_g := \{0\}$ if $\partial^{-1}(g)$ is empty). This defines a G-grading $\mathscr{C} = \bigoplus_{g \in G} \mathscr{C}_g$ of $\mathscr{C}$.

Next we define a functor $T : \underline{G} \to \mathrm{Aut}_{\otimes}(\mathscr{C})$, $g \mapsto T_g$ as follows. Let $V \in \mathrm{Vec}_X^{\omega}$ be a homogeneous object of degree $x \in X$. The functor $T_g : \mathrm{Vec}_X^{\omega} \xrightarrow{\sim} \mathrm{Vec}_X^{\omega}$ is defined by $T_g(V) := V$ (as a vector space) and the degree of $T_g(V)$ is defined to be ${}^g x$. The T_g are extended to nonhomogeneous objects and morphisms in the obvious way.

The tensor structure $\tilde{\gamma}$ on the functor $T : \underline{G} \to \mathrm{Aut}_{\otimes}(\mathscr{C})$ is defined by

$$\gamma_{g,h}(x)\,\mathrm{id}_V =: \tilde{\gamma}_{g,h}(V) : T_{gh}(V) \xrightarrow{\sim} (T_g \circ T_h)(V)$$

for all homogeneous objects $V \in \mathrm{Vec}_X^{\omega}$ of degree $x \in X$, and $g, h \in G$.

The unit-preserving structure $\iota : T_e \xrightarrow{\sim} \mathrm{id}_{\mathscr{C}}$ on the functor $T : \underline{G} \to \mathrm{Aut}_{\otimes}(\mathscr{C})$ is defined by

$$\gamma_{e,e}^{-1}(x)\,\mathrm{id}_V =: \iota(V) : T_e(V) \xrightarrow{\sim} \mathrm{id}_{\mathscr{C}}(V)$$

for all homogeneous objects $V \in \mathrm{Vec}_X^{\omega}$ of degree $x \in X$.

The tensor structure $\tilde{\mu}_g$ on the functor $T_g : \mathrm{Vec}_X^{\omega} \xrightarrow{\sim} \mathrm{Vec}_X^{\omega}$ for $g \in G$ is defined by

$$\mu_g(x, y)\,\mathrm{id}_{V \otimes_{\mathbb{K}} W} =: \tilde{\mu}_g(V, W) : T_g(V \otimes W) \xrightarrow{\sim} T_g(V) \otimes T_g(W)$$

for all homogeneous objects $V, W \in \mathrm{Vec}_X^{\omega}$ of degrees $x, y \in X$, respectively.

The unit-preserving structure ν_g on the functor $T_g : \mathrm{Vec}_X^{\omega} \xrightarrow{\sim} \mathrm{Vec}_X^{\omega}$ for $g \in G$ is defined by

$$\mu_g^{-1}(e, e)\,\mathrm{id}_{\mathbb{K}} =: \nu_g : T_g(\mathbb{K}) \xrightarrow{\sim} \mathbb{K}.$$

For $V, W \in \mathrm{Vec}$, let $\tau_{V,W}$ denote the flip operator $V \otimes_{\mathbb{K}} W \xrightarrow{\sim} W \otimes_{\mathbb{K}} V$ that takes $v \otimes_{\mathbb{K}} w \mapsto w \otimes_{\mathbb{K}} v$. The G-braiding $\tilde{c}$ is defined by

$$c(x, y)\tau_{V,W} =: \tilde{c}(V, W) : V \otimes W \xrightarrow{\sim} T_g(W) \otimes V$$

for all homogeneous objects $V, W \in \mathrm{Vec}_X^{\omega}$ of degrees $x, y \in X$. Here $g = \partial(x)$.

The crossed module axioms of (G, X, ∂) and the quasiabelian 3-cocycle axioms of (ω, γ, μ, c) together ensure that the necessary axioms of a crossed category are satisfied. Specifically, Definition 3.4(c) ensures that $\tilde{\mu_g}$ is a tensor structure on the functor T_g defined above. Definition 3.4(d) of ensures that $\tilde{\gamma}$ is a tensor structure on the functor T. The conditions of Definition 3.4(e_1)–(e_3) correspond to the axioms of Definition 3.4(i)–(iii), respectively.

We will denote the crossed pointed category constructed above by $\mathscr{C}(\omega, \gamma, \mu, c)$.

Remark 4.1. The trivial component $\mathscr{C}(\omega, \gamma, \mu, c)_e$ of $\mathscr{C}(\omega, \gamma, \mu, c)$ (under the G-grading) is a braided fusion category. As a fusion category, $\mathscr{C}(\omega, \gamma, \mu, c)_e = \mathrm{Vec}_{\mathrm{Ker}\,\partial}^{\omega|_{\mathrm{Ker}\,\partial}}$. Suppose that the quasiabelian 3-cocycle (ω, γ, μ, c) is normalized. Then the braiding on the trivial component is given by

$$V \otimes W \to W \otimes V, \quad v \otimes w \mapsto c(x, y) w \otimes v$$

for all homogeneous objects $V, W \in \mathrm{Vec}_{\mathrm{Ker}\,\partial}^{\omega|_{\mathrm{Ker}\,\partial}}$ of degrees $x, y \in \mathrm{Ker}\,\partial$. Clearly, the braided fusion category $\mathscr{C}(\omega, \gamma, \mu, c)_e$ is nondegenerate if and only if the quasiabelian 3-cocycle (ω, γ, μ, c) is nondegenerate in the sense of Definition 3.10.

4b. *Classification.*

Proposition 4.2. *Let $\mathscr{C}(\omega, \gamma, \mu, c)$ and $\mathscr{C}(\omega', \gamma', \mu', c')$ be crossed pointed categories as constructed in Section 4a. Then $\mathscr{C}(\omega, \gamma, \mu, c) \cong \mathscr{C}(\omega', \gamma', \mu', c')$ as crossed categories if and only if there is an isomorphism (f, F) of the underlying (finite) crossed modules such that the quasiabelian 3-cocycles $(\omega', \gamma', \mu', c')^{(f,F)}$ and (ω, γ, μ, c) are cohomologous.*

Proof. Suppose (G, X, ∂) and (G', X', ∂') are the underlying (finite) crossed modules of $\mathscr{C}(\omega, \gamma, \mu, c)$ and $\mathscr{C}(\omega', \gamma', \mu', c')$, respectively. Let (f, F) be an isomorphism from (G, X, ∂) to (G', X', ∂') such that the quasiabelian 3-cocycles $(\omega', \gamma', \mu', c')^{(f,F)}$ and (ω, γ, μ, c) are cohomologous via (η, β); see Section 3. In what follows we will construct an equivalence $(f, \tilde{F}, \tilde{\eta}, \eta_0, \tilde{\beta})$ of crossed categories (see Definition 2.6) from $\mathscr{C}(\omega, \gamma, \mu, c)$ to $\mathscr{C}(\omega', \gamma', \mu', c')$.

Recall that $\mathscr{C}(\omega, \gamma, \mu, c) = \mathrm{Vec}_X^{\omega}$ and $\mathscr{C}(\omega', \gamma', \mu', c') = \mathrm{Vec}_{X'}^{\omega'}$ as fusion categories. Let $V \in \mathrm{Vec}_X^{\omega}$ be a homogeneous object of degree $x \in X$. Define a functor $\tilde{F} : \mathrm{Vec}_X^{\omega} \to \mathrm{Vec}_{X'}^{\omega'}$ by $\tilde{F}(V) := V$ (as a vector space), and define the degree of $\tilde{F}(V)$ to be $F(x)$. The functor $\tilde{F}$ extends to nonhomogeneous objects and morphisms in the obvious way.

The tensor structure $\tilde{\eta}$ on the functor $\tilde{F}$ is defined by

$$\eta(x, y)\,\mathrm{id}_{V \otimes_{\mathbb{K}} W} =: \tilde{\eta}(V, W) : \tilde{F}(V \otimes W) \xrightarrow{\sim} \tilde{F}(V) \otimes \tilde{F}(W)$$

for all homogeneous objects $V, W \in \mathrm{Vec}_X^{\omega}$ of degrees $x, y \in X$, respectively.

The definition of the unit-preserving structure η_0 on $\tilde{F}$ is obvious. It is easy to verify that $(\tilde{F}, \tilde{\eta}, \eta_0)$ is an equivalence of tensor categories.

Next we define isomorphisms $\tilde{\beta}_g : F \circ T_g \xrightarrow{\sim} T'_{f(g)} \circ F$ for $g \in G$ of tensor functors by

$$\beta_g(x)\,\mathrm{id}_V =: \tilde{\beta}_g(V) : (\tilde{F} \circ T_g)(V) \xrightarrow{\sim} (T'_{f(g)} \circ \tilde{F})(V)$$

for all homogeneous objects $V \in \mathrm{Vec}_X^{\omega}$ of degree $x \in X$.

It is easy to verify that axioms (i)–(iii) of Definition 2.6 are satisfied. This shows that $\mathscr{C}(\omega, \gamma, \mu, c) \cong \mathscr{C}(\omega', \gamma', \mu', c')$ as crossed categories.

The converse is clear from the construction above. □

Remark 4.3. Proposition 4.2 shows that if the quasiabelian 3-cocycles (ω, γ, μ, c) and $(\omega', \gamma', \mu', c')$ are cohomologous (on the same crossed module (G, X, ∂)), then the corresponding crossed pointed categories $\mathscr{C}(\omega, \gamma, \mu, c)$ and $\mathscr{C}(\omega', \gamma', \mu', c')$ are equivalent.

Recall from Section 3 that for any crossed module $\mathscr{X}$ there is a natural action of $\mathrm{Aut}(\mathscr{X})$ on the quasiabelian third cohomology $\mathrm{H}^3_{\mathrm{qa}}(\mathscr{X}, \mathbb{K}^\times)$ of $\mathscr{X}$.

Theorem 4.4. *Crossed pointed categories are classified, up to equivalence, by the orbits of the quasiabelian third cohomology* $\mathrm{H}^3_{\mathrm{qa}}(\mathscr{X}, \mathbb{K}^\times)$ *(of a finite crossed module* $\mathscr{X}$*) under the action of* $\mathrm{Aut}(\mathscr{X})$.

Proof. Every crossed pointed category is equivalent to some $\mathscr{C}(\omega, \gamma, \mu, c)$ with underlying (finite) crossed module $\mathscr{X}$. Now apply Proposition 4.2. □

5. Equivariantization of $\mathscr{C}(\omega, \gamma, \mu, c)$

Throughout this section, let (ω, γ, μ, c) be a normalized quasiabelian 3-cocycle on a finite crossed module (G, X, ∂). In Section 4a, we associated to (ω, γ, μ, c) a crossed pointed category $\mathscr{C}(\omega, \gamma, \mu, c)$. Our goal in this section is to apply the equivariantization process to $\mathscr{C}(\omega, \gamma, \mu, c)$ and study the resulting braided fusion category.

5a. *Description.* Recall that $\mathscr{C}(\omega, \gamma, \mu, c) = \mathrm{Vec}_X^\omega$ as a fusion category.

Proposition 5.1. *An object of the equivariantization category* $\mathscr{C}(\omega, \gamma, \mu, c)^G$ *is an X-graded vector space V together with a twisted action $\triangleright$ of G on V that is compatible with the grading in the sense that*

$$gh \triangleright v = \gamma_{g,h}(x)(g \triangleright (h \triangleright v)), \qquad e \triangleright v = v, \quad \mathrm{degree}(g \triangleright v) = {}^g x \tag{4}$$

for all $v \in V$ homogeneous of degree $x \in X$ and $g, h \in G$. Morphisms in the category are linear maps preserving the twisted action and grading. The twisted action of G on the tensor product is given by

$$g \triangleright (v \otimes w) = \mu_g(x, y)(g \triangleright v \otimes g \triangleright w) \tag{5}$$

for homogeneous v, w of degrees $x, y \in X$, respectively. The associativity constraint on the category is given by

$$(u \otimes v) \otimes w \mapsto \omega(x, y, z) u \otimes (v \otimes w)$$

for all homogeneous u, v, w of degrees $x, y, z \in X$. The braiding on the category is given by

$$v \otimes w \mapsto c(x, y)(\partial(x) \triangleright w \otimes v) \tag{6}$$

for all homogeneous v, w of degrees $x, y \in X$.

Proof. The action $\triangleright$ corresponds to equivariant structure (1). Equation (4) corresponds to (2). The definition of the tensor product (5) comes from (3) and the definition of the braiding (6) comes from (4). □

Remark 5.2. For a simple special case of the description above, take the quasi-abelian 3-cocycle (ω, γ, μ, c) (on the finite crossed module (G, X, ∂)) to be trivial. Then the corresponding equivariantization category $\mathcal{C}(1, 1, 1, 1)^G$ admits a simple description in that objects of this category are G-equivariant vector bundles on X. This braided fusion category was considered in [Bantay 2005]. This category is not nondegenerate in general: By Proposition 5.6 it is nondegenerate if and only if ∂ is an isomorphism. In this case, the category is equivalent to $D(G)$-Mod, as a braided fusion category.

Theorem 5.3. *Every braided group-theoretical category is equivalent to* $\mathcal{C}(\xi)^G$ *for some normalized quasiabelian* 3*-cocycle* ξ *on a finite crossed module* (G, X, ∂).

Proof. This follows from [Naidu et al. 2009], where it was shown that every braided group-theoretical category is the equivariantization of a pointed category. □

Lemma 5.4. *For any* $x \in X$*, let* $\mathrm{Stab}_G(x)$ *denote the stabilizer of* x *in* G*, that is,* $\mathrm{Stab}_G(x) = \{g \in G \mid {}^g x = x\}$. *Define* $\phi_x : \mathrm{Stab}_G(x) \times \mathrm{Stab}_G(x) \to \mathbb{K}^\times$ *by*

$$\phi_x(g, h) := \gamma_{g,h}(x) \quad \text{for } g, h \in \mathrm{Stab}_G(x).$$

Then ϕ_x *is a* 2*-cocycle on* $\mathrm{Stab}_G(x)$.

Proof. The condition of Definition 3.4(b) on γ means that

$$\gamma_{h,k}(x)\gamma_{g,hk}(x) = \gamma_{gh,k}(x)\gamma_{g,h}({}^k x)$$

for all $g, h, k \in G$, $x \in X$. Restricting to $\mathrm{Stab}_G(x)$ we get the stated assertion. □

Let R be a complete set of representatives of orbits of X under the action of G.

Proposition 5.5. *There is a bijection between the set of isomorphism classes of simple objects of* $\mathcal{C}(\omega, \gamma, \mu, c)^G$ *and the isomorphism classes of the set*

$$\Gamma := \{(a, V) \mid a \in R,\ V \text{ is an irreducible module over } \mathbb{K}_{\phi_a}[\mathrm{Stab}_G(a)]\}, \tag{7}$$

where ϕ_a *is the* 2*-cocycle defined in Lemma 5.4.*

Proof. Let $\mathrm{Irr}(\mathcal{C}(\omega, \gamma, \mu, c)^G)$ denote the set of simple objects of $\mathcal{C}^G$. We will define a map

$$\Gamma \to \mathrm{Irr}(\mathcal{C}(\omega, \gamma, \mu, c)^G) \tag{8}$$

and show that it induces a bijection between the isomorphism classes of the source and target sets. Let $g_1, g_2, \ldots$ be coset representatives of $\mathrm{Stab}_G(a)$ in G. Pick any $(a, V) \in \Gamma$. We define the map (8) by

$$(a, V) \mapsto \tilde{V} = \bigoplus_{g_i} V_{{}^{g_i}a}, \tag{9}$$

where $V_{{}^{g_i}a} = V$ as a vector space and $\mathrm{degree}(V_{{}^{g_i}a}) = {}^{g_i}a$. The twisted action of G on $\tilde{V}$ is given by

$$h \triangleright v := \frac{\gamma_{g_j,t}(a)}{\gamma_{h,g_i}(a)}(t \triangleright v) \tag{10}$$

for all $v \in \tilde{V}$ homogeneous of degree ${}^{g_i}a$ with $t \in \mathrm{Stab}_G(a)$ uniquely determined by the equation $hg_i = g_j t$. The degree of $h \triangleright v$ is defined to be ${}^{g_j}a$.

To prove that the map (8) defined via (9) and (10) is well defined, we need to show that the action defined in (10) satisfies (4). This amounts to verifying that the scalars

$$\frac{\gamma_{g_k,st}(a)\gamma_{s,t}(a)}{\gamma_{gh,g_i}(a)} \quad \text{and} \quad \frac{\gamma_{g,h}({}^{g_i}a)\gamma_{g_j,t}(a)\gamma_{g_k,s}(a)}{\gamma_{h,g_i}(a)\gamma_{g,g_j}(a)}$$

are equal for all $g, h \in G$ and $s, t \in \mathrm{Stab}_G(a)$ with $hg_i = g_j t$ and $gg_j = g_k s$. But this follows from applying the condition (b) on γ of Definition 3.4 successively to the quadruples (g, h, g_i, a), (g, g_j, t, a) and (g_k, s, t, a).

We now show that the map (8) induces a bijection between isomorphism classes of the source and target sets. It is clear that the map (8) preserves isomorphic objects. Furthermore, the object in $\mathrm{Irr}(\mathscr{C}(\omega, \gamma, \mu, c)^G)$ corresponding to $(a, V) \in \Gamma$ has FP-dimension equal to $|{}^G a| \dim_{\mathbb{K}} V$, where ${}^G a$ denotes the orbit containing a. The sum of squares of FP-dimensions of isomorphism classes of objects in the image of (8) is

$$\begin{aligned}\sum_{a\in R} \sum_{V \in \mathrm{Irr}(\mathbb{K}_{\phi_a}[\mathrm{Stab}_G(a)])} |{}^G a|^2 (\dim V)^2 &= \sum_{a\in R} |{}^G a|^2 |\mathrm{Stab}_G(a)| \\ &= \sum_{a\in R} |{}^G a|\,|G| \\ &= |G|\,|X| = \mathrm{FPdim}(\mathscr{C}(\omega, \gamma, \mu, c)^G). \qquad \square\end{aligned}$$

5b. ***Twist and S-matrix.*** As before, R denotes a complete set of representatives of orbits of X under the action of G. By Proposition 5.5, the simple objects of $\mathscr{C}(\omega, \gamma, \mu, c)^G$ correspond to pairs (a, χ), where $a \in R$ and χ is an irreducible ϕ_a-character of $\mathrm{Stab}_G(a)$. Note that $\mathscr{C}(\omega, \gamma, \mu, c)^G$ admits a canonical twist θ with respect to which categorical dimensions coincide with FP-dimensions. The values of θ on simple objects are given by $\theta_{(a,\chi)} = c(a, a)\chi(\partial(a))/\deg \chi$. A direct calculation shows that the S-matrix S is given by

$$S_{(a,\chi),(b,\chi')} = \sum_{\substack{x\in({}^G a)\\ y\in({}^G b)\cap C_X(x)}} c(x, y)c(y, x)\frac{\gamma_{g,\partial({}^{g^{-1}}y)}(a)\gamma_{h,\partial({}^{h^{-1}}x)}(b)}{\gamma_{\partial(y),g}(a)\gamma_{\partial(x),h}(b)}\chi({}^{g^{-1}}y)\chi'({}^{h^{-1}}x),$$

where in each summand g and h are defined by ${}^g a = x$ and ${}^h b = y$. Note that the choice of g and h does not affect the sum.

5c. *Modularity.* As before, (ω, γ, μ, c) is a normalized quasiabelian 3-cocycle on a finite crossed module (G, X, ∂).

Proposition 5.6. *The braided fusion category $\mathscr{C}(\omega, \gamma, \mu, c)^G$ is nondegenerate if and only if the homomorphism ∂ is surjective and (ω, γ, μ, c) is nondegenerate in the sense of Definition 3.10.*

Proof. This follows immediately by combining Remark 2.5 and Remark 4.1. □

Assume ∂ is surjective and (ω, γ, μ, c) is nondegenerate. Then $\mathscr{C}(\omega, \gamma, \mu, c)^G$ together with the canonical twist given in Section 5b is a modular category, that is, the S-matrix described in Section 5b is invertible. In this situation, using the orthogonality relations for projective characters we obtain that the Gauss sum and central charge of $\mathscr{C}(\omega, \gamma, \mu, c)^G$ are given respectively by

$$\tau(\mathscr{C}(\omega, \gamma, \mu, c)^G) = |G| \sum_{a \in \operatorname{Ker} \partial} c(a, a),$$

$$\zeta(\mathscr{C}(\omega, \gamma, \mu, c)^G) = \frac{1}{\sqrt{|\operatorname{Ker} \partial|}} \sum_{a \in \operatorname{Ker} \partial} c(a, a).$$

Note 5.7. The sum $\sum_{a \in \operatorname{Ker} \partial} c(a, a)$ is the classical Gauss sum for the quadratic form $a \mapsto c(a, a)$ on the abelian group $\operatorname{Ker} \partial$.

Remark 5.8. The category $\mathscr{C}(\omega, \gamma, \mu, c)^G$ may admit other twists besides the canonical one. In view of Theorem 5.3 and Proposition 5.6, a description of all twists on $\mathscr{C}(\omega, \gamma, \mu, c)^G$ will imply a description of *all* modular group-theoretical categories. The former is easily obtained using Proposition 2.1.

6. Quasitriangular quasi-Hopf algebra arising from quasiabelian 3-cocycles on crossed modules

Suppose (ω, γ, μ, c) is a normalized quasiabelian 3-cocycle on a finite crossed module (G, X, ∂). In the previous section we described the braided fusion category $\mathscr{C}(\omega, \gamma, \mu, c)^G$. This category is integral (that is, the FP-dimensions of objects are integers), so there exists a finite-dimensional quasitriangular quasi-Hopf algebra H such that $\mathscr{C}(\omega, \gamma, \mu, c)^G \cong H$-Mod, as braided fusion categories; see [Etingof et al. 2005, Theorem 8.33] and [Kassel 1995, Section XV.2]. Our goal in this section is to describe such an H.

In what follows we associate to (ω, γ, μ, c) a finite-dimensional quasitriangular quasi-Hopf algebra $H(\omega, \gamma, \mu, c)$, which may be viewed as a generalization of the twisted Drinfeld double of a finite group. Let $H(\omega, \gamma, \mu, c)$ be a finite-dimensional vector space with a basis $\{t_x g\}_{(x,g)\in X\times G}$ indexed by the set $X\times G$. Define a product on $H(\omega,\gamma,\mu,c)$ by

(11) $$(t_x g)(t_y h) := \delta_{x,{}^h y}\gamma_{g,h}(y)^{-1} t_y(gh).$$

This product admits a unit

(12) $$1 = \sum_{x\in X} t_x e.$$

Define a coproduct $\Delta : H(\omega,\gamma,\mu,c) \to H(\omega,\gamma,\mu,c)\otimes H(\omega,\gamma,\mu,c)$ and counit $\varepsilon : H(\omega,\gamma,\mu,c) \to \mathbb{K}$ by

(13) $$\Delta(t_x g) := \sum_{\substack{a,b\in X\\ ab=x}} \mu_g(a,b) t_a g \otimes t_b g,$$

(14) $$\varepsilon(t_x g) := \delta_{x,e}.$$

Also, set

(15) $$\Phi := \sum_{x,y,z\in X} \omega(x,y,z) t_x e \otimes t_y e \otimes t_z e,$$

(16) $$R := \sum_{x,y\in X} c(x,y) t_x e \otimes t_y \partial(x),$$

(17) $$\alpha := 1, \quad \beta := \sum_{x\in X} \omega(x^{-1},x,x^{-1}) t_x e.$$

Finally, define a linear map $S : H(\omega,\gamma,\mu,c) \to H(\omega,\gamma,\mu,c)$ by

(18) $$S(t_x g) := \frac{\gamma_{g^{-1},g}(x^{-1})}{\mu_g(x,x^{-1})} t_{{}^g x^{-1}} g^{-1}.$$

Proposition 6.1. *The product unit, coproduct Δ, counit ε, Drinfeld associator Φ and antiautomorphism S of (11)–(14), (15) and (18) make $H(\omega,\gamma,\mu,c)$ a quasitriangular quasi-Hopf algebra with universal R-matrix R (16) in the sense of* [Kassel 1995, *Definitions 1.1, 2.1, and 5.1*].

Proof. The proof is completely similar to the one for the twisted Drinfeld double of a finite group: Associativity of the product is equivalent to the equality

$$\gamma_{h,k}(x)\gamma_{g,hk}(x) = \gamma_{gh,k}(x)\gamma_{g,h}({}^k x) \quad \text{for } g,h,k\in G, x\in X,$$

which holds by axiom (b) in Definition 3.4. Quasicoassociativity of the coproduct is equivalent to the equality

$$\frac{\mu_g(y,z)\mu_g(x,yz)}{\mu_g(xy,z)\mu_g(x,y)} = \frac{\omega({}^g x,{}^g y,{}^g z)}{\omega(x,y,z)} \quad \text{for } g\in G, x,y,z\in X,$$

which holds by axiom (c) in Definition 3.4. That the coproduct is a morphism of algebras is equivalent to the equality

$$\frac{\gamma_{g,h}(x)\gamma_{g,h}(y)}{\gamma_{g,h}(xy)} = \frac{\mu_g({}^h x, {}^h y)\mu_h(x, y)}{\mu_{gh}(x, y)} \quad \text{for } g, h \in G \text{ and } x, y \in X,$$

which holds by axiom (d) in Definition 3.4.

We note that the inverse of the R-matrix R is

$$R^{-1} = \sum_{x,y\in X} c(x, x^{-1}yx)^{-1}\gamma_{\partial(x),\partial(x^{-1})}(y)^{-1} t_x e \otimes t_y \partial(x^{-1}).$$

The R-matrix axioms on R hold due to axioms (e_1)–(e_3) in Definition 3.4.

Finally, axioms (a)–(d) in Definition 3.4 ensure that S is indeed an antiautomorphism that satisfies the required axioms. □

Proposition 6.2. *Let (ω, γ, μ, c) be a normalized quasiabelian 3-cocycle on a finite crossed module (G, X, ∂). The categories $\mathcal{C}(\omega, \gamma, \mu, c)^G$ (see Section 5) and $H(\omega, \gamma, \mu, c)$-Mod are equivalent as braided fusion categories.*

Proof. Let V be a (left) module over $H(\omega, \gamma, \mu, c)$, with action denoted by $\cdot$. Note that V admits an X-grading: $V = \bigoplus_{x\in X} V_x$, where $V_x = (t_x e)\cdot V$. Define a twisted action of G on V by $g \triangleright v := (t_x g)\cdot v$ for all $v \in V$ homogeneous of degree $x \in X$. Observe that the degree of $g \triangleright v$ is ${}^g x$, since $(t_x g)(t_x e) = (t_{{}^g x} e)(t_x g)$. The aforementioned action is twisted in that $gh \triangleright v = \gamma_{g,h}(x)(g \triangleright (h \triangleright v))$. Note that the twisted action of G completely determines the action of $H(\omega, \gamma, \mu, c)$ on the module V. The associativity constraint on the category $H(\omega, \gamma, \mu, c)$-Mod (which is defined using the Drinfeld associator Φ of (15)) is given by

$$(u \otimes v) \otimes w \mapsto \omega(x, y, z) u \otimes (v \otimes w)$$

for all homogeneous u, v, w of degrees $x, y, z \in X$. The braiding on the category $H(\omega, \gamma, \mu, c)$-Mod (which is defined using the R-matrix R of (16)) is given by

$$v \otimes w \mapsto c(x, y)(\partial(x) \triangleright w \otimes v)$$

for homogeneous v, w of degrees $x, y \in X$. Now compare with Proposition 5.1. □

We next explain the relation between the quasitriangular quasi-Hopf algebras constructed above and the twisted Drinfeld double of a finite group. Let ω be a normalized 3-cocycle on a finite group G.

For all $g, h, x, y \in G$, define

$$\gamma_{g,h}(x) := \frac{\omega(g, h, x)\omega(ghxh^{-1}g^{-1}, g, h)}{\omega(g, hxh^{-1}, h)}, \tag{19}$$

$$\mu_g(x, y) := \frac{\omega(gxg^{-1}, g, y)}{\omega(gxg^{-1}, gyg^{-1}, g)\omega(g, x, y)}. \tag{20}$$

A direct computation establishes the following.

Lemma 6.3. *The quadruple* $(\omega, \gamma, \mu, 1)$, *where* γ *and* μ *are respectively defined by* (19) *and* (20), *is a quasiabelian* 3*-cocycle on the crossed module* (G, G, id_G) (*where* G *acts on itself by conjugation*) *in the sense of Definition 3.4.*

Let $(\omega, \gamma, \mu, 1)$ be the quasiabelian 3-cocycle on (G, G, id_G) constructed in Lemma 6.3. Then, evidently, $H(\omega^{-1}, \gamma^{-1}, \mu^{-1}, 1) \cong D^{\omega}(G)$ as quasitriangular quasi-Hopf algebras. In particular, $\mathscr{C}(\omega^{-1}, \gamma^{-1}, \mu^{-1}, 1)^G \cong D^{\omega}(G)$-Mod as braided fusion categories.

Acknowledgments

I thank D. Nikshych and S. Witherspoon for useful discussions.

References

[Arkhipov and Gaitsgory 2003] S. Arkhipov and D. Gaitsgory, "Another realization of the category of modules over the small quantum group", *Adv. Math.* **173**:1 (2003), 114–143. MR 2004e:17010 Zbl 1025.17004

[Bakalov and Kirillov 2001] B. Bakalov and A. Kirillov, Jr., *Lectures on tensor categories and modular functors*, University Lecture Series **21**, American Mathematical Society, Providence, RI, 2001. MR 2002d:18003 Zbl 0965.18002

[Bantay 2005] P. Bantay, "Characters of crossed modules and premodular categories", preprint, 2005. arXiv math/0512542

[Beliakova and Blanchet 2001] A. Beliakova and C. Blanchet, "Modular categories of types B, C and D", *Comment. Math. Helv.* **76**:3 (2001), 467–500. MR 2003e:57050 Zbl 0996.57003

[Brown 1982] K. S. Brown, *Cohomology of groups*, Graduate Texts in Mathematics **87**, Springer, New York, 1982. MR 83k:20002 Zbl 0584.20036

[Dijkgraaf et al. 1990] R. Dijkgraaf, V. Pasquier, and P. Roche, "Quasi Hopf algebras, group cohomology and orbifold models", *Nuclear Phys. B Proc. Suppl.* **18B** (1990), 60–72. MR 92m:81238 Zbl 0957.81670

[Dijkgraaf et al. 1992] R. Dijkgraaf, V. Pasquier, and P. Roche, "Quasi-Hopf algebras, group cohomology and orbifold models", pp. 75–98 in *Integrable systems and quantum groups* (Pavia, 1990), edited by P. Binétruy et al., World Sci. Publ., River Edge, NJ, 1992. MR 1178543 Zbl 0925.16024

[Drinfeld et al. 2010] V. Drinfeld, S. Gelaki, D. Nikshych, and V. Ostrik, "On braided fusion categories, I", *Selecta Math.* (*N.S.*) **16**:1 (2010), 1–119. MR 2609644 Zbl 05716902

[Eilenberg and Mac Lane 1953] S. Eilenberg and S. Mac Lane, "On the groups of $H(\Pi, n)$, I", *Ann. of Math.* (2) **58** (1953), 55–106. MR 15,54b Zbl 0050.39304

[Eilenberg and Mac Lane 1954] S. Eilenberg and S. Mac Lane, "On the groups $H(\Pi, n)$, II: Methods of computation", *Ann. of Math.* (2) **60** (1954), 49–139. MR 16,391a Zbl 0055.41704

[Etingof et al. 2005] P. Etingof, D. Nikshych, and V. Ostrik, "On fusion categories", *Ann. of Math.* (2) **162**:2 (2005), 581–642. MR 2006m:16051 Zbl 1125.16025

[Gaitsgory 2005] D. Gaitsgory, "The notion of category over an algebraic stack", preprint, 2005. arXiv math.AG/0507192

[Gelaki and Nikshych 2008] S. Gelaki and D. Nikshych, "Nilpotent fusion categories", *Adv. Math.* **217**:3 (2008), 1053–1071. MR 2009b:18015 Zbl 1168.18004

[Joyal and Street 1993] A. Joyal and R. Street, "Braided tensor categories", *Adv. Math.* **102**:1 (1993), 20–78. MR 94m:18008 Zbl 0817.18007

[Kassel 1995] C. Kassel, *Quantum groups*, Graduate Texts in Mathematics **155**, Springer, New York, 1995. MR 96e:17041 Zbl 0808.17003

[Kirillov 2001a] A. Kirillov, Jr., "Modular categories and orbifold models, II", preprint, 2001. arXiv math/0110221

[Kirillov 2001b] A. Kirillov, Jr., "On G-equivariant modular categories", preprint, 2001. arXiv math/0401119

[Mac Lane 1952] S. MacLane, "Cohomology theory of Abelian groups", pp. 8–14 in *Proceedings of the International Congress of Mathematicians* (Cambridge, MA, 1950), vol. 2, Amer. Math. Soc., Providence, RI, 1952. MR 13,529d Zbl 0049.01401

[Müger 2003a] M. Müger, "From subfactors to categories and topology, I: Frobenius algebras in and Morita equivalence of tensor categories", *J. Pure Appl. Algebra* **180**:1-2 (2003), 81–157. MR 2004f:18013 Zbl 1033.18002

[Müger 2003b] M. Müger, "On the structure of modular categories", *Proc. London Math. Soc.* (3) **87**:2 (2003), 291–308. MR 2004g:18009 Zbl 1037.18005

[Müger 2004] M. Müger, "Galois extensions of braided tensor categories and braided crossed G-categories", *J. Algebra* **277**:1 (2004), 256–281. MR 2005b:18011

[Müger 2005] M. Müger, "Conformal orbifold theories and braided crossed G-categories", *Comm. Math. Phys.* **260**:3 (2005), 727–762. MR 2007b:81228 Zbl 1160.81454

[Naidu et al. 2009] D. Naidu, D. Nikshych, and S. Witherspoon, "Fusion subcategories of representation categories of twisted quantum doubles of finite groups", *Int. Math. Res. Not.* **2009**:22 (2009), 4183–4219. MR 2552301 Zbl 05640426

[Ospel 1999] C. Ospel, *Tressages et théories cohomologiques pour les algèbres de Hopf: Application aux invariants des 3-variétés*, thesis, Univ. Strasbourg, 1999. MR 2001i:16071

[Ostrik 2003] V. Ostrik, "Module categories, weak Hopf algebras and modular invariants", *Transform. Groups* **8**:2 (2003), 177–206. MR 2004h:18006 Zbl 1044.18004

[Tambara 2001] D. Tambara, "Invariants and semi-direct products for finite group actions on tensor categories", *J. Math. Soc. Japan* **53**:2 (2001), 429–456. MR 2002e:18010 Zbl 0980.18003

[Turaev 2000] V. Turaev, "Homotopy field theory in dimension 3 and crossed group-categories", preprint, 2000. arXiv math/0005291

[Turaev 2008] V. Turaev, "Crossed group-categories", *Arab. J. Sci. Eng. Sect. C Theme Issues* **33**:2 (2008), 483–503. MR 2500054 Zbl 1185.18009

[Whitehead 1949] J. H. C. Whitehead, "Combinatorial homotopy, I", *Bull. Amer. Math. Soc.* **55** (1949), 213–245. MR 11,48b Zbl 0040.38704

Received June 1, 2009. Revised May 17, 2010.

DEEPAK NAIDU
DEPARTMENT OF MATHEMATICS
TEXAS A&M UNIVERSITY
COLLEGE STATION, TX 77843-3368
UNITED STATES
dnaidu@math.tamu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 247, No. 2, 2010

PAINLEVÉ ANALYSIS OF GENERALIZED ZAKHAROV EQUATIONS

HASSAN A. ZEDAN AND SALMA M. AL-TUWAIRQI

We consider the invariance and integrability properties of the generalized Zakharov equations and obtain an exact invariant solution. We derive the Painlevé property of these equations and obtain some exact solutions using the truncated Painlevé expansion.

1. Introduction

We consider the generalized Zakharov equations (GZEs) for the complex envelope $E(x,t)$ of the high-frequency wave and the real low-frequency field $\eta(x,t)$ in the form

$$iE_t + E_{xx} - 2\beta|E|^2E + 2E\eta = 0, \tag{1}$$

$$\eta_{tt} - \eta_{xx} = -(|E|^2)_{xx}, \tag{2}$$

where the cubic term in Equation (1) describes nonlinear self-interaction in the high-frequency subsystem; such a term corresponds to a self-focusing effect in plasma physics. The coefficient β is a real constant that can be positive or negative. The sound velocity and the coupling constant in Equation (2) have been normalized to unity for simplicity. The GZEs are a universal model of interaction between high- and low-frequency waves in one dimension. The collisions between solitary waves of GZEs have been simulated in detail in [Hadouaj et al. 1991]. When $\beta = 0$, this system is reduced to the classical Zakharov equations for plasma physics.

Several aspects of the GZEs have been studied. Malomed, Anderson, Lisak, Quiroga-Teixeiro and Stenflo [1997] analyzed them using a variational approach. Wang and Li [2005] introduced periodic wave solutions using the extended F-expansion method. By considering the modified Adomian decomposition method, Wang, Dai, Wu, Lei and Zhang [2007] calculated exact and numerical solutions. Zhang [2007] constructed exact traveling wave solutions by a direct algebraic method. Li, Li and Lin [2008] used the exp-function method to obtain exact solutions. Javidi and Golbabi [2008] obtained exact and numerical solutions by the

MSC2000: primary 35Q51; secondary 35K05.
Keywords: generalized Zakharov equations, Lie symmetries, Painlevé expansion.

variational iteration method. Abbasbandy, Babolian and Ashtiani [2009] applied homotopy analysis method to obtain a solution.

The objective of this work is to investigate the symmetry and conduct Painlevé analysis of GZEs. It will be organized as follows. In Section 2, we obtain the Lie point symmetry group of GZEs and its similarity reductions. Moreover, we also construct explicit analytic invariant solutions. In Section 3, we inspect the singularity structure of GZEs by means of the Weiss–Tabor–Carnevale procedure. In Section 4, we obtain some exact solutions using the truncated Painlevé expansion.

2. Symmetry analysis

Symmetry is one of the most important concepts in the area of partial differential equations [Bluman and Kumei 1989]. To find the Lie symmetries of the GZEs, we express the complex envelope as $E(x,t) = u(x,t) + iv(x,t)$, with real high-frequency waves $u(x,t)$ and $v(x,t)$. Substituting it into Equations (1) and (2) and separating the imaginary and real parts we obtain the equations

$$
\begin{aligned}
u_{xx} - v_t - 2\beta(u^2+v^2)u + 2\eta u &= 0,\\
u_t + v_{xx} - 2\beta(u^2+v^2)v + 2\eta v &= 0,\\
\eta_{tt} - \eta_{xx} &= -(u^2+v^2)_{xx}.
\end{aligned} \tag{3}
$$

Consider a one-parameter Lie group of infinitesimal transformations of the form

$$
\begin{aligned}
&u \to U = u + \varepsilon\phi_1(x,t,u,v,\eta),\\
&v \to V = v + \varepsilon\phi_2(x,t,u,v,\eta),\\
&\eta \to \Lambda = \eta + \varepsilon\phi_3(x,t,u,v,\eta),\\
&x \to X = x + \varepsilon\xi_1(x,t,u,v,\eta),\\
&t \to T = t + \varepsilon\xi_2(x,t,u,v,\eta) \quad \text{for } \varepsilon \ll 1,
\end{aligned} \tag{4}
$$

with infinitesimal generator

$$
\widetilde{X} = \xi_1\frac{\partial}{\partial x} + \xi_2\frac{\partial}{\partial t} + \phi_1\frac{\partial}{\partial u} + \phi_2\frac{\partial}{\partial v} + \phi_3\frac{\partial}{\partial \eta}. \tag{5}
$$

Require that (3) are invariant under (4) by direct substitution [Bluman and Kumei 1989]. Eliminate u_t, v_t, η_{tt} using (3), and set to zero all coefficients of the independent terms of the polynomials of u, v and η and their partial derivatives. We then obtain the determining equations for the infinitesimals. By solving these equations we have

$$\phi_1 = -(k_3 + k_4 t + k_5 t^2)v,$$
$$\phi_2 = (k_3 + k_4 t + k_5 t^2)u,$$
$$\phi_3 = \tfrac{1}{2}k_4 + k_5 t,$$
$$\xi_1 = k_1, \quad \xi_2 = k_2,$$

where k_1, k_2, k_3, k_4 and k_5 are arbitrary parameters. The infinitesimal generator of the Lie algebra associated with each parameter k_i is obtained from the generator (5):

$$L_1 = \frac{\partial}{\partial x}, \quad L_2 = \frac{\partial}{\partial t}, \quad L_4 = tu\frac{\partial}{\partial v} - tv\frac{\partial}{\partial u} + \frac{1}{2}\frac{\partial}{\partial \eta},$$
$$L_3 = u\frac{\partial}{\partial v} - v\frac{\partial}{\partial u}, \qquad L_5 = t^2 u\frac{\partial}{\partial v} - t^2 v\frac{\partial}{\partial u} + t\frac{\partial}{\partial \eta}.$$

The similarity solution can be obtained from the invariant surface equation

$$\frac{dt}{k_2} = \frac{dx}{k_1} = \frac{du}{-(k_3+k_4t+k_5t^2)v} = \frac{dv}{(k_3+k_4t+k_5t^2)u} = \frac{d\eta}{k_4/2+k_5t}.$$

For the most general generator X, we obtain the similarity solution

$$u(x,t) = c\cos\left(\frac{k_5}{3k_2}t^3 + \frac{k_4}{2k_2}t^2 + \frac{k_3}{k_2}t + F_1(z)\right),$$
$$v(x,t) = c\sin\left(\frac{k_5}{3k_2}t^3 + \frac{k_4}{2k_2}t^2 + \frac{k_3}{k_2}t + F_1(z)\right),$$
$$\eta(x,t) = \frac{k_5}{2k_2}t^2 + \frac{k_4}{2k_2}t + F_2(z),$$

where F_1 and F_2 are similarity functions of the similarity variable $z = k_1 t - k_2 x$, satisfying the similarity equations

$$\begin{aligned} 0 &= -k_2^3(F_1'' - (F_1')^2) + k_1k_2F_1' - 2k_2F_2 + k_3 + 2c^2\beta k_2, \\ 0 &= k_2^3(F_1'' + (F_1')^2) + k_1k_2F_1' - 2k_2F_2 + k_3 + 2c^2\beta k_2, \\ 0 &= F_2'' + k_5/(k_2(k_1^2 - k_2^2)), \end{aligned} \tag{6}$$

where prime denotes differentiation with respect to z. Solving equations (6), we obtain the solution

$$F_1 = \frac{1}{2k_2^3}\left(-k_1k_2 + \sqrt{(k_1k_2)^2 - 4k_2^3(k_3 - 2k_2(a_2 - c^2\beta))}\right)z + a_0,$$
$$F_2 = a_1,$$

where a_0 and a_1 are arbitrary constants. Hence the exact invariant solution to

GZEs (1) and (2) can be written as

$$E(x,t)=c\exp\biggl(\frac{i}{2k_2^3}\bigl(k_2^2\bigl(\tfrac{2}{3}k_5t^3+k_4t^2+2k_3t\bigr)+(k_1t-k_2x)\sqrt{(k_1k_2)^2-4k_2^3(k_3-2k_2(a_1-c^2\beta))}-k_1k_2+2k_2^3a_0\bigr)\biggr),$$

$$\eta(x,t)=\frac{k_5}{2k_2}t^2+\frac{k_4}{2k_2}t+a_1.$$

3. Painlevé analysis

The Painlevé test is one of the impressive ways to test whether partial differential equations are integrable or nonintegrable. Myrzakulov [1999] has confirmed the Painlevé nature of the (2+1)-dimensional Zakharov equation. In this paper we will show that GZEs possess the Painlevé property. Various approaches can be applied to investigate the Painlevé integrability. Here we will use WTC method [Weiss et al. 1983]. Consider Laurent expansion of the solutions of (3) in a local neighborhood of a movable singular manifold $\phi(x,t)=0$:

$$u(x,t)=\sum_{j=0}^{\infty}u_j\phi^{j+\alpha_1},\quad v(x,t)=\sum_{j=0}^{\infty}v_j\phi^{j+\alpha_2},\quad \eta(x,t)=\sum_{j=0}^{\infty}\eta_j\phi^{j+\alpha_3},$$

where $u_j(x,t)$, $v_j(x,t)$ and $\eta_j(x,t)$ are analytic functions and the three α_i are integers to be determined. It is sufficient to substitute

$$u(x,t)=u_0\phi^{\alpha_1},\quad v(x,t)=v_0\phi^{\alpha_2},\quad \eta(x,t)=\eta_0\phi^{\alpha_3},$$

into equations (3) to find the dominant behavior and leading exponents α_i. Balancing the dominant terms, we get $\alpha_1=\alpha_2=-1$ and $\alpha_3=-2$, and the equations

$$u_0^2+v_0^2=\phi_x^2-\frac{\phi_x^4}{\beta^2\phi_t^2+\beta(1-\beta)\phi_x^2}\quad\text{and}\quad \eta_0=-\frac{\phi_x^4}{\beta^2\phi_t^2+(1-\beta)\phi_x^2}.$$

The first shows that either u_0 or v_0 is arbitrary. Substituting the series expansion

$$u=u_0\phi^{-1}+u_r\phi^{r-1},\quad v=v_0\phi^{-1}+v_r\phi^{r-1},\quad \eta=\eta_0\phi^{-2}+\eta_r\phi^{r-2},$$

into equations (3) and collecting the coefficients of u_r, v_r and η_r for leading terms, the resonance are found to have the values $r=-1,0,2,3,3,4$. As usual, the resonance $r=-1$ represents the arbitrariness of the singularity manifold $\phi(x,t)=0$, and the remaining resonance values indicate the arbitrariness of five functions (one of $\{u_0, v_0, \eta_0\}$, one of $\{u_2, v_2, \eta_2\}$, two of $\{u_3, v_3, \eta_3\}$ and one of $\{u_4, v_4, \eta_4\}$).

4. Some solutions to the generalized Zakharov equation

When an equation has the Painlevé property, one can find some kinds of exact solutions by truncating the Painlevé expansion of the solution about the movable singularity manifold at the constant level term, leading to the so-called Backlund transformation of the equation. For GZEs, the transformation has the form

$$u = u_0\phi^{-1} + u_1, \quad v = v_0\phi^{-1} + v_1, \quad \eta = \eta_0\phi^{-2} + \eta_1\phi^{-1} + \eta_2. \tag{7}$$

Our approach to finding an exact solution for u, v and η is similar to that of [Choudhury 2006]. Substituting equations (7) into equations (3) and equating the coefficients of ϕ^{-i} for $i = 0, 1, 2, 3, 4$ to zero, we obtain this system of PDEs:

$$0 = -2\beta u_1^2 v_1 - 2\beta v_1^3 + 2v_1\eta_2 + u_{1,t} + v_{1,xx}, \tag{8}$$

$$0 = -2\beta u_1^2 v_0 - 4\beta u_0 u_1 v_1 - 6\beta v_0 v_1^2 + 2v_1\eta_1 + 2v_0\eta_2 + u_{0,t} + v_{0,xx}, \tag{9}$$

$$0 = -4\beta u_0 u_1 v_0 - 2\beta u_0^2 v_1 - 6\beta v_0^2 v_1 + 2v_1\eta_0 + 2v_0\eta_1 - u_0\phi_t - 2v_{0,x}\phi_x - v_0\phi_{xx}, \tag{10}$$

$$0 = -2\beta u_0^2 v_0 - 2\beta v_0^3 + 2v_0\eta_0 + 2v_0\phi_x^2, \tag{11}$$

$$0 = -2\beta u_1^3 - 2\beta u_1 v_1^2 + 2u_1\eta_2 - v_{1,t} + u_{1,xx}, \tag{12}$$

$$0 = -6\beta u_0 u_1^2 - 4\beta u_1 v_0 v_1 - 2\beta u_0 v_1^2 + 2u_1\eta_1 + 2u_0\eta_2 - v_{0,t} + u_{0,xx}, \tag{13}$$

$$0 = -6\beta u_0^2 u_1 - 2\beta u_1 v_0^2 - 4\beta u_0 v_0 v_1 + 2u_1\eta_0 + 2u_0\eta_1 + v_0\phi_t - 2u_{0,x}\phi_x - u_0\phi_{xx}, \tag{14}$$

$$0 = -2\beta u_0^3 - 2\beta u_0 v_0^2 + 2u_0\eta_0 + 2u_0\phi_x^2, \tag{15}$$

$$0 = \eta_{2,tt} + 2u_{1,x}^2 + 2v_{1,x}^2 + 2u_1 u_{1,xx} + 2v_1 v_{1,xx} - \eta_{2,xx}, \tag{16}$$

$$0 = \eta_{1,tt} + 4u_{0,x}u_{1,x} + 4v_{0,x}v_{1,x} + 2u_1 u_{0,xx} + 2u_0 u_{1,xx} + 2v_1 v_{0,xx} + 2v_0 v_{1,xx} - \eta_{1,xx}, \tag{17}$$

$$\begin{aligned} 0 = {} & -2\eta_{1,t}\phi_t - \eta_1\phi_{tt} + \eta_{0,tt} - 4u_1\phi_x u_{0,x} + 2u_{0,x}^2 - 4u_0\phi_x u_{1,x} \\ & - 4v_1\phi_x v_{0,x} + 2v_{0,x}^2 - 4v_0\phi_x v_{1,x} + 2\phi_x\eta_{1,x} - 2u_0 u_1\phi_{xx} \\ & - 2v_0 v_1\phi_{xx} + \eta_1\phi_{xx} + 2u_0 u_{0,xx} + 2v_0 v_{0,xx} - \eta_{0,xx}, \end{aligned} \tag{18}$$

$$\begin{aligned} 0 = {} & 2\eta_1\phi_t^2 - 4\phi_t\eta_{0,t} - 2\eta_0\phi_{tt} + 4u_0 u_1\phi_x^2 + 4v_0 v_1\phi_x^2 - 2\eta_1\phi_x^2 \\ & - 8u_0\phi_x u_{0,x} - 8v_0\phi_x v_{0,x} + 4\phi_x\eta_{0,x} - 2u_0^2\phi_{xx} - 2v_0^2\phi_{xx} + 2\eta_0\phi_{xx}, \end{aligned} \tag{19}$$

$$0 = 6\eta_0\phi_t^2 + 6u_0^2\phi_x^2 + 6v_0^2\phi_x^2 - 6\eta_0\phi_x^2. \tag{20}$$

Substituting a trial solution $\phi(x,t) = 1 + \exp(iQ(x,t))$ into system (8)–(20), we consider two cases:

Case 1. Let $v_0(x,t) = \exp(iQ(x,t))$ be the arbitrary function; then equation (11), (15) and (10) become

$$0 = -2e^{iQ}(e^{2iQ}(\beta + Q_x^2) + \beta u_0^2 - \eta_0), \tag{21}$$

$$0 = -2u_0(e^{2iQ}(\beta + Q_x^2) + \beta u_0^2 - \eta_0), \tag{22}$$

$$\begin{aligned} 0 = {} & -2\beta u_0^2 v_1 + v_1(-6\beta e^{2iQ} + 2\eta_0) + 2e^{iQ}\eta_1 \\ & - e^{iQ}u_0(4\beta u_1 + iQ_t) + 3e^{2iQ}Q_x^2 - ie^{2iQ}Q_{xx}. \end{aligned} \tag{23}$$

It is convenient to separate (23) into two equations:

$$0=-e^{iQ}u_0(iQ_t)-ie^{2iQ}Q_{xx},$$
$$0=-2\beta u_0^2 v_1+v_1(-6\beta e^{2iQ}+2\eta_0)+2e^{iQ}\eta_1-4\beta e^{iQ}u_0u_1+3e^{2iQ}Q_x^2.$$

Solving the first for u_0 we get $u_0(x,t)=-Q_{xx}e^{iQ}/Q_t$. Substituting this into (21) we obtain

$$\eta_0(x,t)=(\beta Q_t^2+Q_t^2Q_x^2+\beta Q_{xx}^2)e^{2iQ}/Q_t^2.$$

Letting $Q(x,t)=f_1(t)x+f_2(t)$ and substituting this into system (8)–(20), we find that (20) has the form

$$0=6e^{4i(f_1x+f_2)}(f_1^4-\beta(xf_1'+f_2')^2+f_1^2(-1+\beta-x^2f_1'^2-2xf_1'f_2'-f_2'^2)),$$
$$0=f_1^4-\beta(xf_1'+f_2')^2+f_1^2(-1+\beta-x^2f_1'^2-2xf_1'f_2'-f_2'^2).$$

Equating the coefficients of x^2, x^1 and x^0 in the second equation to zero, we obtain

$$0=-f_1^2+\beta f_1^2+f_1^4-\beta f_2'^2-f_1^2f_2'^2,$$
$$0=-2\beta f_1'f_2'-2f_1^2f_1'f_2',$$
$$0=-f_1'^2(\beta+f_1^2).$$

The first two are satisfied if $f_1(t)=c_1$. Substituting this into the last and solving for f_2, we get

$$f_2(t)=\sqrt{\frac{-c_1^2(c_1^2+\beta-1)}{-\beta-c_1^2}}\,t+c_2.$$

Substituting all the above results into equations (10), (13) and (14) and solving for v_1, u_1 and η_1 we find $v_1(x,t)=-1/2$,

$$u_1(x,t)=\frac{i\sqrt{(\beta+c_1^2)(\beta+c_1^2-1)}}{2c_1(\beta+c_1^2)}\quad\text{and}\quad \eta_1(x,t)=-(\beta+c_1^2)\exp(ih(x,t)),$$

where $h(x,t)=c_1x+c_2+\sqrt{c_1^2(1-\beta-c_1^2)}t/\sqrt{-\beta-c_1^2}$. Solving (8) or (9) or (12) for η_2, we get

$$\eta_2(x,t)=\frac{\beta(1-c_1^2+c_1^4)-\beta^2(1-c_1^2)}{4c_1^2(\beta+c_1^2)},$$

Substituting all the results into the truncated expansion (7), we obtain the corresponding exact solution of GZEs (3):

$$u(x,t)=-\frac{i\sqrt{1-\beta-c_1^2}}{2c_1\sqrt{-\beta-c_1^2}},\qquad v(x,t)=-\frac{1}{2}+\frac{\exp(ih(x,t))}{1+\exp(ih(x,t))},$$
$$\eta(x,t)=\frac{\beta(1-c_1^2+c_1^4)-\beta^2(1-c_1^2)}{4c_1^2(\beta+c_1^2)}+\frac{(\beta+c_1^2)\exp(ih(x,t))}{1+\exp(ih(x,t))}\Big(\frac{\exp(ih(x,t))}{1+\exp(ih(x,t))}-1\Big).$$

Figures 1, 2, 3, 4 and 5 illustrate the solution for certain β, c_1 and c_2.

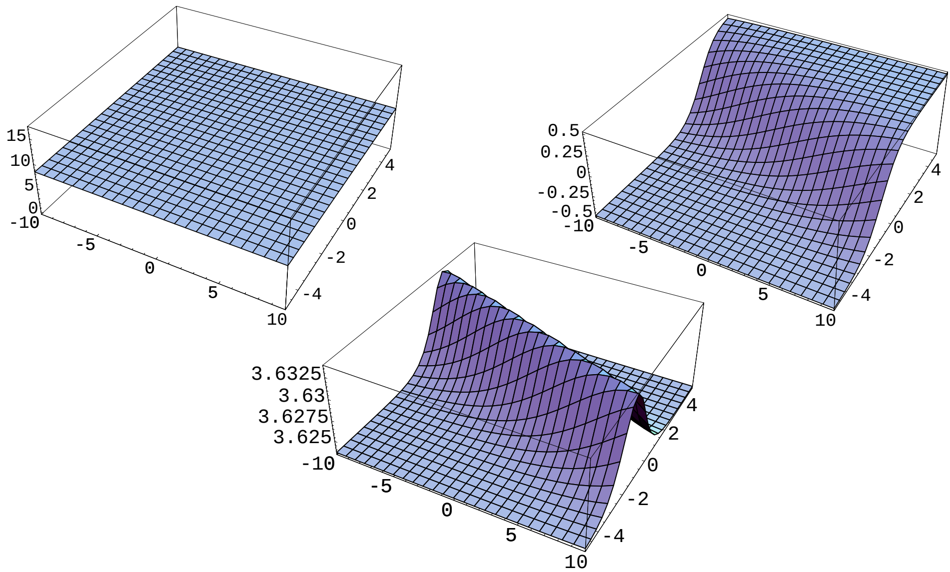


Figure 1. Clockwise from top left: Graphs of $u(x,t)$, $v(x,t)$ and $\eta(x,t)$ for $\beta = 0.05$, $c_1 = -0.3i$ and $c_2 = i$.

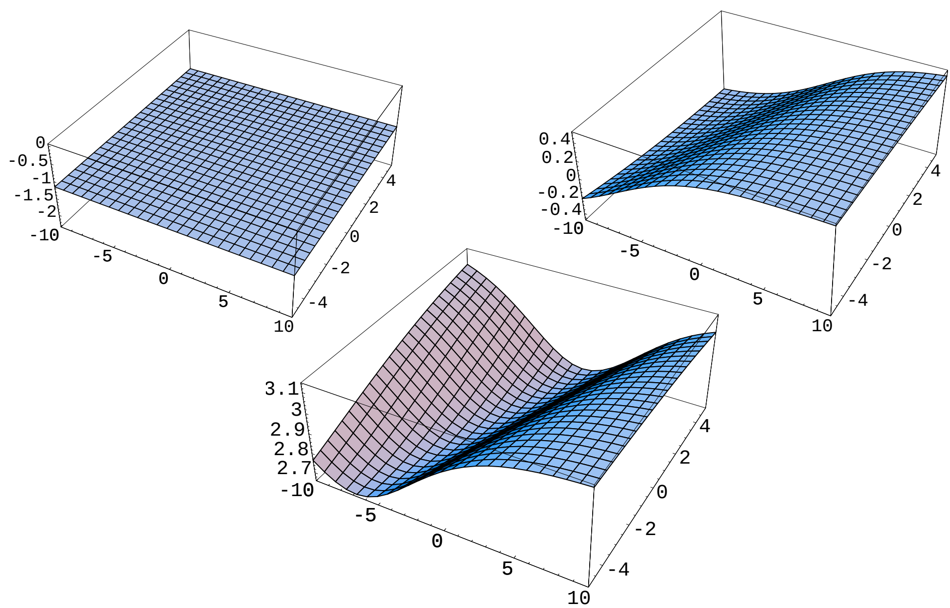


Figure 2. Clockwise from top left: Graphs of $u(x,t)$, $v(x,t)$ and $\eta(x,t)$ for $\beta = 2$, $c_1 = -0.3i$ and $c_2 = -i$.

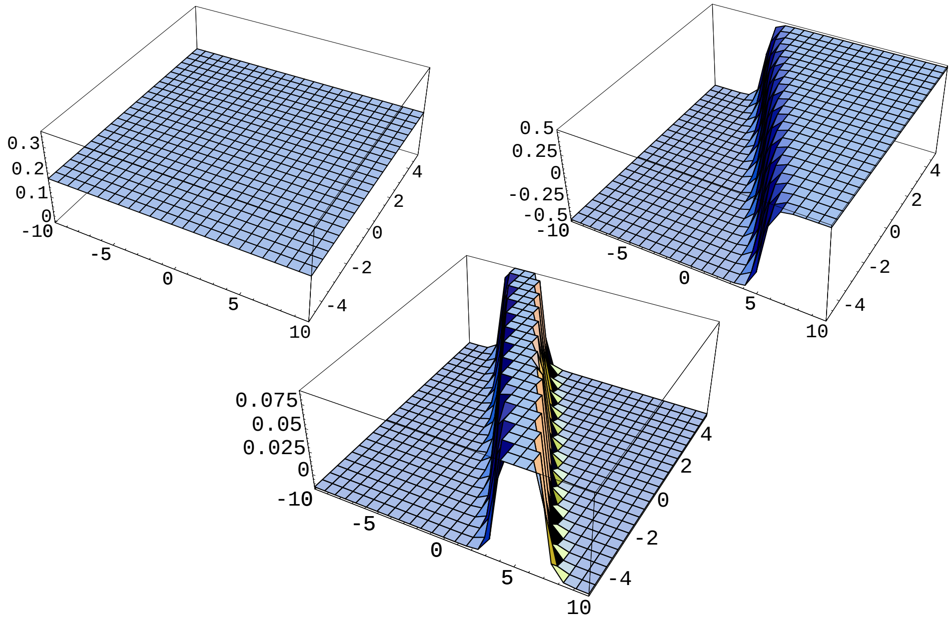


Figure 3. Clockwise from top left: Graphs of $u(x,t)$, $v(x,t)$ and $\eta(x,t)$ for $\beta=-0.05$, $c_1=-3i$ and $c_2=0.4i$.

Case 2: Let

$$u_0(x,t)=\exp(iQ(x,t))$$

be the arbitrary function. Then separate (14) into two equations:

$$0=iv_0Q_t-ie^{iQ}Q_{xx},$$
$$0=-2u_1(3\beta e^{2iQ}+\beta v_0^2-\eta_0)+e^{iQ}(2\eta_1+v_0(-4\beta v_1+iQ_t)+3e^{iQ}Q_x^2).$$

From the first we obtain $v_0(x,t)=Q_{xx}e^{iQ}/Q_t$. Solving (11) for η_0, we get

$$\eta_0(x,t)=(\beta Q_t^2+Q_t^2Q_x^2+\beta Q_{xx}^2)e^{2iQ}/Q_t^2.$$

Letting $Q(x,t)=f_1(t)x+f_2(t)$ and substituting this into system (8)–(20), we find that (20) assumes the form

$$0=6e^{4i(f_1x+f_2)}(f_1^4-\beta(xf_1'+f_2')^2+f_1^2(-1+\beta-x^2f_1'^2-2xf_1'f_2'-f_2'^2)),$$
$$0=f_1^4-\beta(xf_1'+f_2')^2+f_1^2(-1+\beta-x^2f_1'^2-2xf_1'f_2'-f_2'^2).$$

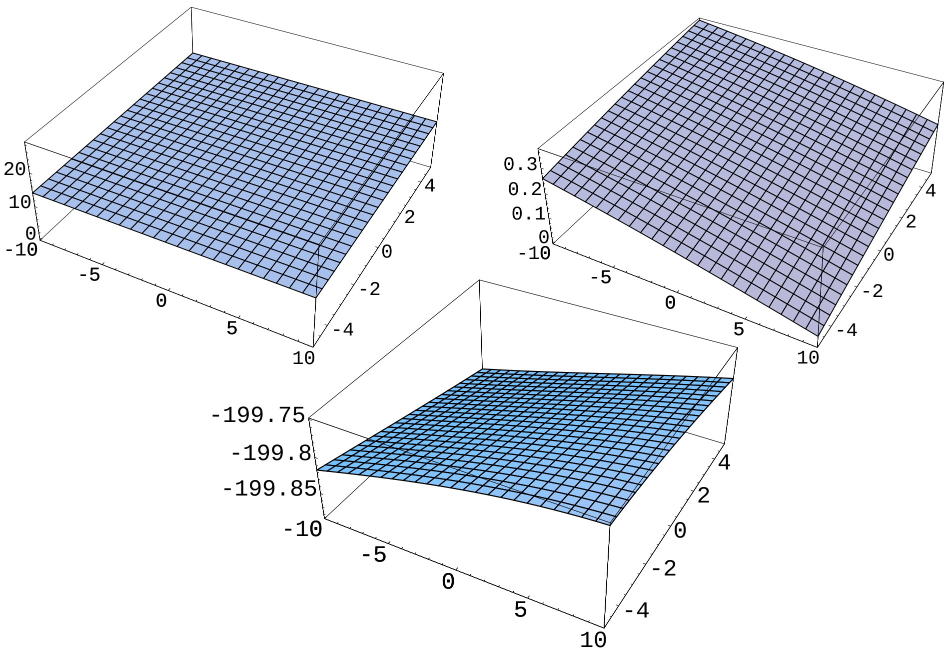


Figure 4. Clockwise from top left: Graphs of $u(x,t)$, $v(x,t)$ and $\eta(x,t)$ for $\beta=-1,\ c_1=0.05i$ and $c_2=-i$.

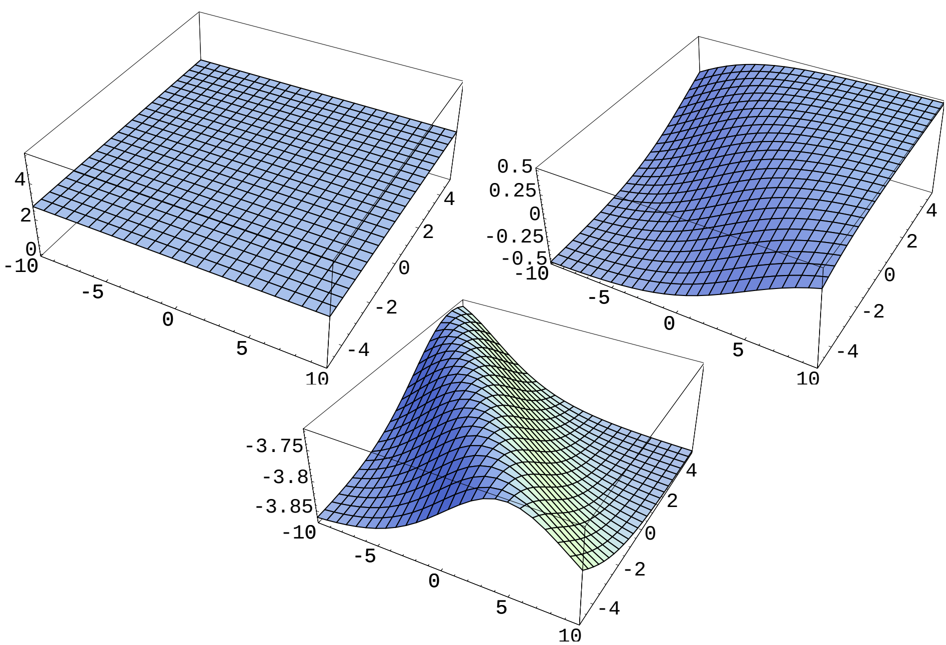


Figure 5. Clockwise from top left: Graphs of $u(x,t)$, $v(x,t)$ and $\eta(x,t)$ for $\beta=-0.5$, $c_1=-0.3$ and $c_2=-i$.

Equating the coefficients of x^2, x^1 and x^0 in the second of these to zero, we obtain

$$\begin{aligned}
0 &= -f_1^2 + \beta f_1^2 + f_1^4 - \beta f_2'^2 - f_1^2 f_2'^2,\\
0 &= -2\beta f_1' f_2' - 2 f_1^2 f_1' f_2',\\
0 &= -f_1'^2(\beta + f_1^2).
\end{aligned}$$

The latter two are satisfied if $f_1(t) = c_1$. Substituting this into the first equation above and solving for f_2, we get

$$f_2(t) = t\sqrt{\frac{-c_1^2(c_1^2+\beta-1)}{-\beta-c_1^2}} + c_2.$$

Substituting all the above results into Equations (10), (14) and (19) and solving for υ_1, u_1 and η_1 we find

$$\begin{aligned}
u_1(x,t) &= -1/2,\\
\upsilon_1(x,t) &= \frac{i\sqrt{(-\beta-c_1^2+1)}}{2c_1(\beta+c_1^2)},\\
\eta_1(x,t) &= -(\beta+c_1^2)\exp(ih(x,t)),
\end{aligned}$$

where $h(x,t) = c_1 x + c_2 + t\sqrt{c_1^2(1-\beta-c_1^2)}/\sqrt{-\beta-c_1^2}$. Solving (8) or (12) or (13) for η_2, we get

$$\eta_2(x,t) = \frac{\beta(1-c_1^2+c_1^4) - \beta^2(1-c_1^2)}{4c_1^2(\beta+c_1^2)},$$

Substituting all the results into the truncated expansion (7), we obtain the corresponding exact solution of GZEs (3):

$$\begin{aligned}
\upsilon(x,t) &= \frac{i\sqrt{1-\beta-c_1^2}}{2c_1\sqrt{-\beta-c_1^2}}\\
u(x,t) &= -\frac{1}{2} + \frac{\exp(ih(x,t))}{1+\exp(ih(x,t))}\\
\eta(x,t) &= \frac{\beta(1-c_1^2+c_1^4) - \beta^2(1-c_1^2)}{4c_1^2(\beta+c_1^2)}\\
&\qquad + \frac{(\beta+c_1^2)\exp(ih(x,t))}{1+\exp(ih(x,t))}\left(\frac{\exp(ih(x,t))}{1+\exp(ih(x,t))} - 1\right).
\end{aligned}$$

Figures 6, 7, 8, 9 and 10 illustrate the solution for certain β, c_1 and c_2.

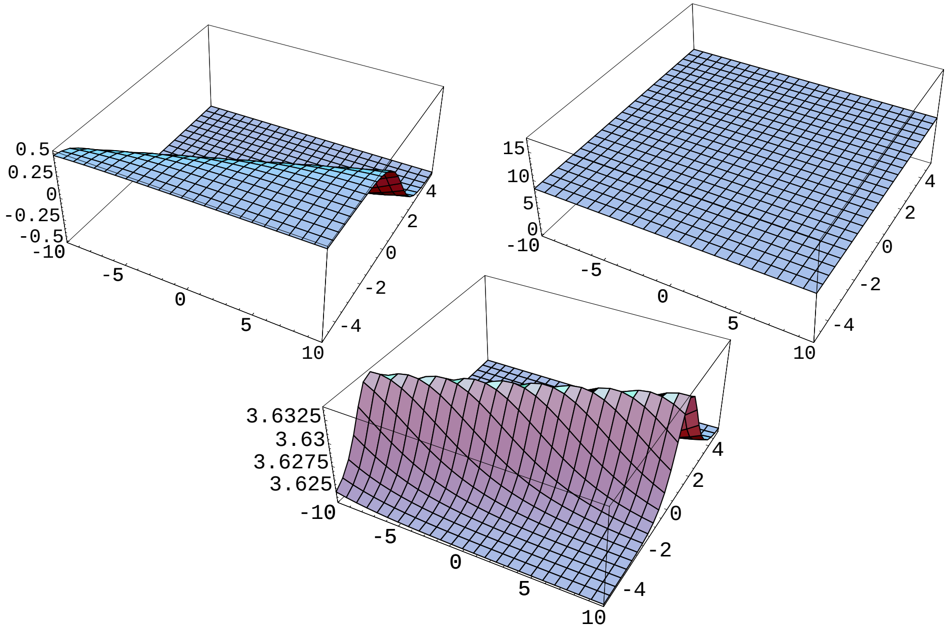


Figure 6. Clockwise from top left: Graphs of $u(x,t)$, $v(x,t)$ and $\eta(x,t)$ for $\beta = 0.05$, $c_1 = -0.3i$ and $c_2 = i$.

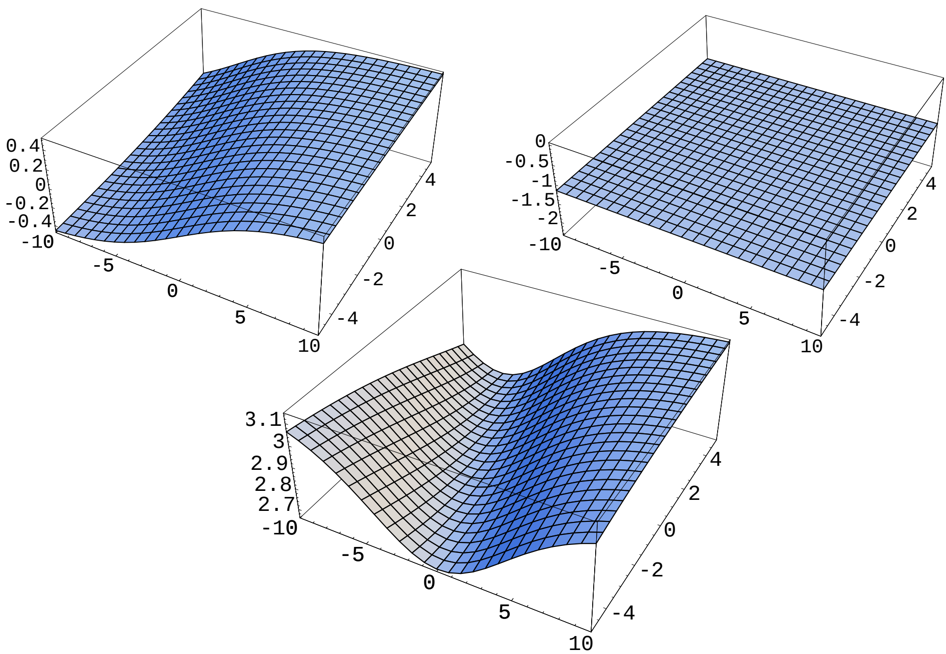


Figure 7. Clockwise from top left: Graphs of $u(x,t)$, $v(x,t)$ and $\eta(x,t)$ for $\beta = 2$, $c_1 = -0.3i$ and $c_2 = -i$.

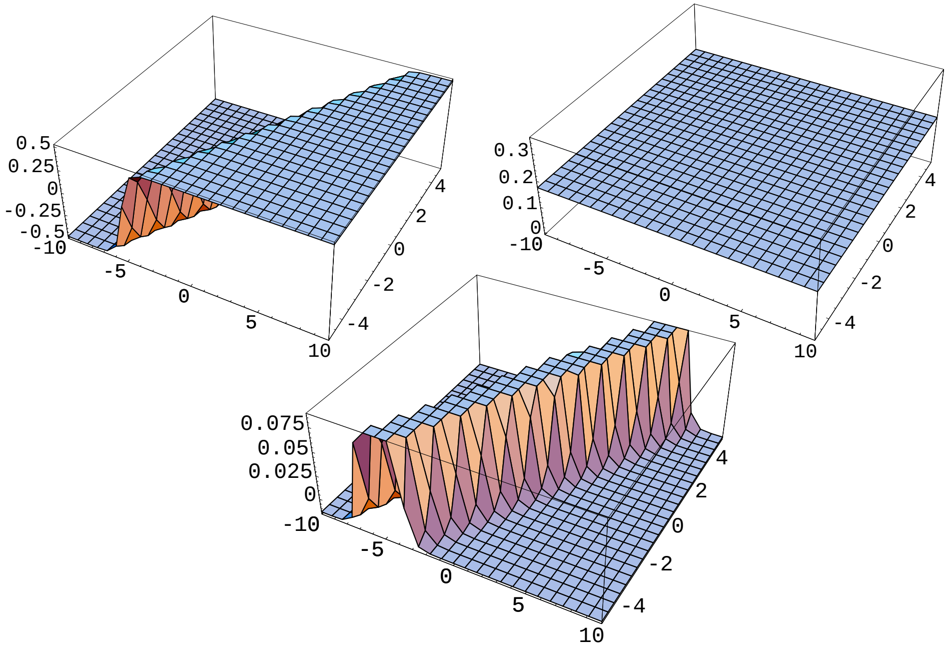


Figure 8. Clockwise from top left: Graphs of $u(x,t)$, $v(x,t)$ and $\eta(x,t)$ for $\beta = -0.05$, $c_1 = -3i$ and $c_2 = 0.4i$.

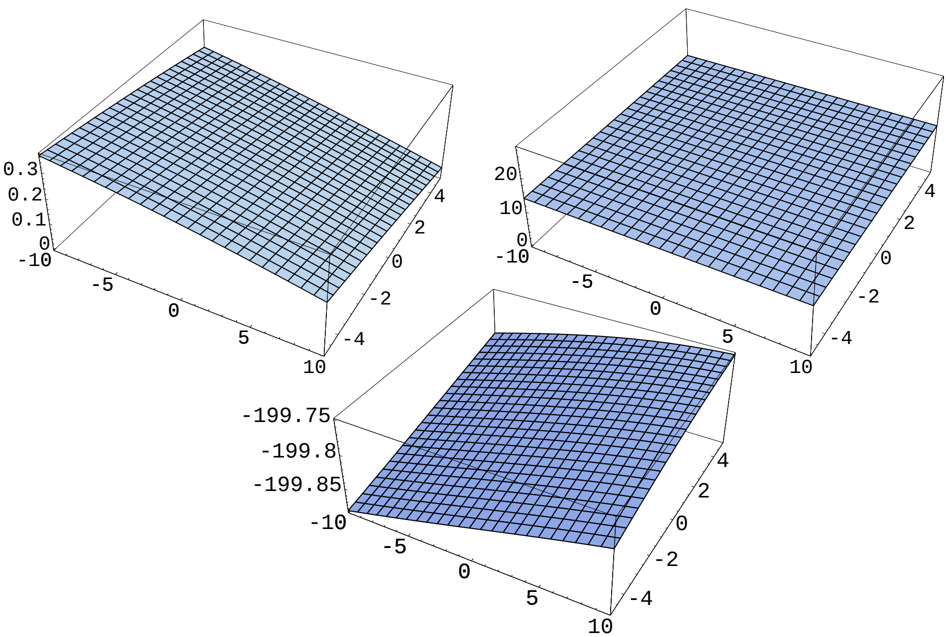


Figure 9. Clockwise from top left: Graphs of $u(x,t)$, $v(x,t)$ and $\eta(x,t)$ for $\beta = -1$, $c_1 = 0.05$ and $c_2 = -i$.

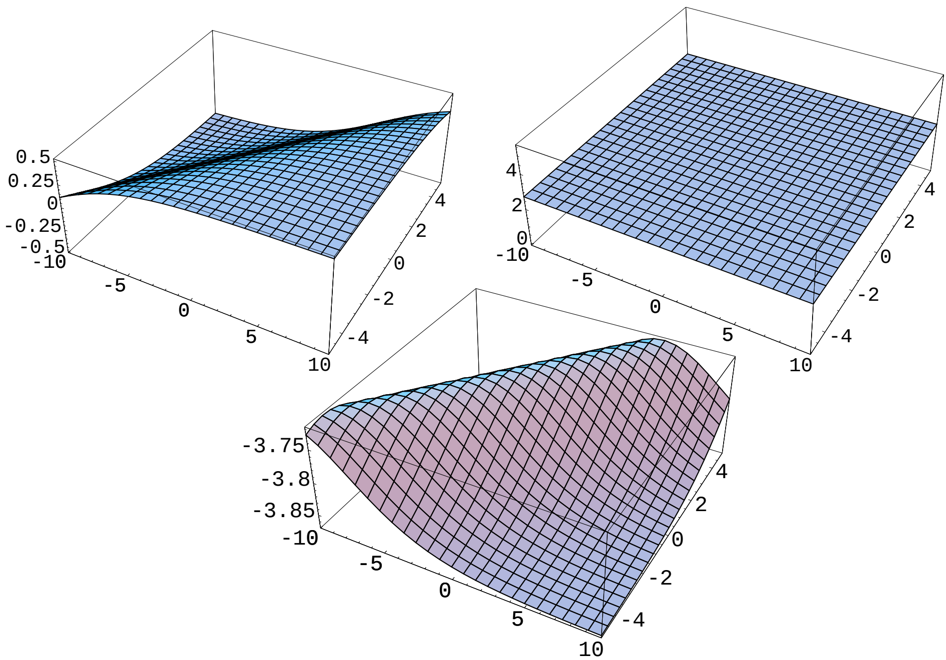


Figure 10. Clockwise from top left: Graphs of $u(x,t)$, $v(x,t)$ and $\eta(x,t)$ for $\beta = -0.5$, $c_1 = -0.3$ and $c_2 = -i$.

References

[Abbasbandy et al. 2009] S. Abbasbandy, E. Babolian, and M. Ashtiani, "Numerical solution of the generalized Zakharov equation by homotopy analysis method", *Commun. Nonlinear Sci. Numer. Simul.* **14**:12 (2009), 4114–4121. MR 2537572 Zbl 1159.65348

[Bluman and Kumei 1989] G. W. Bluman and S. Kumei, *Symmetries and differential equations*, Applied Mathematical Sciences **81**, Springer, New York, 1989. MR 91b:58284 Zbl 0698.35001

[Choudhury 2006] S. R. Choudhury, "Painlevé analysis of nonlinear evolution equations — an algorithmic method", *Chaos Solitons Fractals* **27**:1 (2006), 139–152. MR 2006b:37132 Zbl 1088.35553

[Hadouaj et al. 1991] H. Hadouaj, B. A. Malomed, and G. A. Maugin, "Dynamics of a soliton in a generalized Zakharov system with dissipation", *Phys. Rev. A* (3) **44**:6 (1991), 3925–3931. MR 92g:35190

[Javidi and Golbabi 2008] M. Javidi and A. Golbabi, "Exact and numerical solitary wave solutions of generalized Zakharov equation by the variational iteration method", *Chaos, Solitons and Fractals* **36** (2008), 309–313. Zbl 05235317

[Li et al. 2008] Y.-Z. Li, K.-M. Li, and C. Lin, "Exp-function method for solving the generalized Zakharov equations", *Appl. Math. Comput.* **205**:1 (2008), 197–201. MR 2466623 Zbl 1160.35523

[Malomed et al. 1997] B. Malomed, D. Anderson, M. Lisak, M. Quiroga-Teixeiro, and L. Stenflo, "Dynamics of solitary waves in the Zakharov model equations", *Phys. Rev. E* **55** (1997), 962–968.

[Myrzakulov 1999] R. Myrzakulov, "Singularity structure analysis, integrability solitons and dromions in $(2+1)$-dimensional Zakharov equations", preprint, 1999. arXiv solv-int/9903008

[Wang and Li 2005] M. Wang and X. Li, "Extended F-expansion method and periodic wave solutions for the generalized Zakharov equations", *Phys. Lett. A* **343** (2005), 48–54. MR 2005m:35266 Zbl 1181.35255

[Wang et al. 2007] Y.-Y. Wang, C.-Q. Dai, L. Wu, and J.-F. Zhang, "Exact and numerical solitary wave solutions of generalized Zakharov equation by the Adomian decomposition method", *Chaos Solitons Fractals* **32**:3 (2007), 1208–1214. MR 2007h:35322 Zbl 1130.35120

[Weiss et al. 1983] J. Weiss, M. Tabor, and G. Carnevale, "The Painlevé property for partial differential equations", *J. Math. Phys.* **24**:3 (1983), 522–526. MR 84c:35101 Zbl 0514.35083

[Zhang 2007] H. Zhang, "New exact travelling wave solutions of the generalized Zakharov equations", *Rep. Math. Phys.* **60**:1 (2007), 97–106. MR 2008i:35233 Zbl 1170.35524

Received June 23, 2009. Revised January 10, 2010.

HASSAN A. ZEDAN
MATHEMATICS DEPARTMENT
KAFR EL-SHEIKH UNIVERSITY
KAFR EL-SHEIKH
EGYPT
Current address:
MATHEMATICS DEPARTMENT
KING ABDUL AZIZ UNIVERSITY
PO BOX 80203
JEDDAH 21589
SAUDI ARABIA
hassanzedan2003@yahoo.com

SALMA M. AL-TUWAIRQI
MATHEMATICS DEPARTMENT
KING ABDULAZIZ UNIVERSITY
JEDDAH 21589
SAUDI ARABIA
saltuwairqi@gmail.com

CONTENTS

Volume 247, no. 1 and no. 2

Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/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 247 No. 2 October 2010