Pacific Journal of Mathematics

Volume 248 No. 1 November 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. 248, No. 1, 2010

AN EXISTENCE THEOREM OF CONFORMAL SCALAR-FLAT METRICS ON MANIFOLDS WITH BOUNDARY

SÉRGIO DE MOURA ALMARAZ

Let (M, g) be a compact Riemannian manifold with boundary. We address the Yamabe-type problem of finding a conformal scalar-flat metric on M whose boundary is a constant mean curvature hypersurface. When the boundary is umbilic, we prove an existence theorem that finishes some of the remaining cases of this problem.

1. Introduction

J. Escobar [1992a] has studied the following Yamabe-type problem for manifolds with boundary:

Yamabe problem. Let (M^n, g) be a compact Riemannian manifold of dimension $n \geq 3$ with boundary ∂M. Is there a scalar-flat metric on M that is conformal to g and has ∂M as a constant mean curvature hypersurface?

In dimension two, the classical Riemann mapping theorem says that any simply connected, proper domain of the plane is conformally diffeomorphic to a disk. This theorem is false in higher dimensions since the only bounded open subsets of $\mathbb{R}^n$ for $n \geq 3$ that are conformally diffeomorphic to Euclidean balls are the Euclidean balls themselves. The Yamabe-type problem proposed by Escobar can be viewed as an extension of the Riemann mapping theorem for higher dimensions.

In analytical terms, this problem corresponds to finding a positive solution to

$$\begin{cases} L_g u = 0 & \text{in } M, \\ B_g u + K u^{n/(n-2)} = 0 & \text{on } \partial M \end{cases} \tag{1-1}$$

for some constant K, where $L_g = \Delta_g - \frac{1}{4}(n-2)/(n-1) R_g$ is the conformal Laplacian and $B_g = \partial/\partial\eta - \frac{1}{2}(n-2)h_g$. Here Δ_g is the Laplace–Beltrami operator, R_g is the scalar curvature, h_g is the mean curvature of ∂M and η is the inward unit normal vector to ∂M.

MSC2000: primary 53C25; secondary 35J65.
Keywords: Yamabe problem, manifold with boundary, scalar curvature, mean curvature, Weyl tensor.
The author was supported at IMPA by CNPq-Brazil.

The solutions of the equations (1-1) are the critical points of the functional

$$Q(u) = \frac{\displaystyle\int_M |\nabla_g u|^2 + \frac{n-2}{4(n-1)} R_g u^2 dv_g + \frac{n-2}{2}\int_{\partial M} h_g u^2 d\sigma_g}{\left(\displaystyle\int_{\partial M} u^2(n-1)/(n-2) d\sigma_g\right)^{(n-2)/(n-1)}},$$

where dv_g and $d\sigma_g$ denote the volume forms of M and ∂M, respectively. Escobar introduced the conformally invariant Sobolev quotient

$$Q(M, \partial M) = \inf\{Q(u) : u \in C^1(M), u \not\equiv 0 \text{ on } \partial M\}$$

and proved that it satisfies $Q(M, \partial M) \leq Q(B^n, \partial B)$. Here, B^n denotes the unit ball in $\mathbb{R}^n$ endowed with the Euclidean metric.

Under the hypothesis that $Q(M, \partial M)$ is finite (which is the case when $R_g \geq 0$), he also showed that the strict inequality

$$Q(M, \partial M) < Q(B^n, \partial B) \tag{1-2}$$

implies the existence of a minimizing solution of the equations (1-1).

Notation. We denote by (M^n, g) a compact Riemannian manifold of dimension $n \geq 3$ with boundary ∂M and finite Sobolev quotient $Q(M, \partial M)$.

Theorem 1.1 [Escobar 1992a]. *Assume that one of the following conditions holds*:

(1) $n \geq 6$ *and* M *has a nonumbilic point on* ∂M;

(2) $n \geq 6$, M *is locally conformally flat and* ∂M *is umbilic*;

(3) $n = 4$ *or* 5 *and* ∂M *is umbilic*;

(4) $n = 3$.

Then $Q(M, \partial M) < Q(B^n, \partial B)$ *and there is a minimizing solution to* (1-1).

The proof for $n = 6$ under condition (1) appeared later, in [Escobar 1996b].

Further existence results were obtained by F. Marques in [Marques 2005] and [Marques 2007]. Together, these results can be stated as follows:

Theorem 1.2 [Marques 2005; Marques 2007]. *Assume that one of the following conditions holds*:

(1) $n \geq 8$, $\overline{W}(x) \neq 0$ *for some* $x \in \partial M$ *and* ∂M *is umbilic*;

(2) $n \geq 9$, $W(x) \neq 0$ *for some* $x \in \partial M$ *and* ∂M *is umbilic*;

(3) $n = 4$ *or* 5 *and* ∂M *is not umbilic.*

Then $Q(M, \partial M) < Q(B^n, \partial B)$ *and there is a minimizing solution to* (1-1).

Here, W denotes the Weyl tensor of M and $\overline{W}$ the Weyl tensor of ∂M.

Our main result deals with the remaining dimensions $n = 6, 7$ and 8 when the boundary is umbilic and $W \neq 0$ at some boundary point:

Theorem 1.3. *Suppose that $n = 6, 7$ or 8, that ∂M is umbilic and that $W(x) \neq 0$ for some $x \in \partial M$. Then $Q(M, \partial M) < Q(B^n, \partial B)$ and there is a minimizing solution to the equations* (1-1).

These cases are similar to the case of dimensions 4 and 5 when the boundary is not umbilic, studied in [Marques 2007].

Other works concerning conformal deformation on manifolds with boundary include [Ahmedou 2003; Ambrosetti et al. 2002; Ben Ayed et al. 2005; Brendle 2002; Djadli et al. 2003; 2004; Escobar 1992b; 1996a; Escobar and Garcia 2004; Felli and Ould Ahmedou 2003; 2005; Han and Li 1999; 2000]

We will now discuss the strategy in the proof of Theorem 1.3. We assume that ∂M is umbilic and choose $x_0 \in \partial M$ so that $W(x_0) \neq 0$. Our proof is explicitly based on constructing a test function ψ such that

$$Q(\psi) < Q(B^n, \partial B). \tag{1-3}$$

The function ψ has support in a small half-ball around the point x_0. The usual strategy in this kind of problem (which goes back to [Aubin 1976]) is to define the function ψ in the small half-ball as one of the standard entire solutions to the corresponding Euclidean equations. In our context those are

$$U_\epsilon(x) = \left(\frac{\epsilon}{x_1^2 + \cdots + x_{n-1}^2 + (\epsilon + x_n)^2}\right)^{(n-2)/2}, \tag{1-4}$$

where $x = (x_1, \ldots, x_n)$ and $x_n \geq 0$.

The next step would be to expand the quotient of ψ in powers of ϵ and, by exploiting the local geometry around x_0, show that the inequality (1-3) holds if ϵ is small. To simplify the asymptotic analysis, we use conformal Fermi coordinates centered at x_0. This concept, introduced in [Marques 2005], plays the same role that conformal normal coordinates (see [Lee and Parker 1987]) did in the case of manifolds without boundary.

When $n \geq 9$, the strict inequality (1-3) was proved in [Marques 2005]. The difficulty arises because, when $3 \leq n \leq 8$, the first correction term in the expansion does not have the right sign. When $3 \leq n \leq 5$, Escobar proved the strict inequality by applying the positive mass theorem, a global construction originally due to Schoen [1984]. This argument does not work when $6 \leq n \leq 8$ because the metric is not sufficiently flat around the point x_0.

As we mentioned before, the situation under the hypothesis of Theorem 1.3 is quite similar to the cases of dimensions 4 and 5 when the boundary is not umbilic, cases solved by Marques [2007]. As he pointed out, the test functions U_ϵ are not optimal in these cases but the problem is still local. This phenomenon does not appear in the classical solution of the Yamabe problem for manifolds without boundary. However, perturbed test functions have been used in the works of Hebey and Vaugon [1993], Brendle [2007] and Khuri, Marques and Schoen [2009].

To prove the inequality (1-3), inspired by the ideas of Marques, we introduce

$$\phi_\epsilon(x) = \epsilon^{n-2/2} R_{ninj}(x_0) x_i x_j x_n^2 (x_1^2 + \cdots + x_{n-1}^2 + (\epsilon + x_n)^2)^{-n/2}.$$

Our test function ψ is defined as $\psi = U_\epsilon + \phi_\epsilon$ around $x_0 \in \partial M$.

In Section 2 we expand the metric g in Fermi coordinates and discuss the conformal Fermi coordinates. In Section 3 we prove Theorem 1.3 by estimating $Q(\psi)$.

Notation. Throughout, we use the index notation for tensors, with commas denoting covariant differentiation. We adopt the summation convention whenever confusion does not result. When dealing with Fermi coordinates, we will use indices $1 \le i, j, k, l, m, p, r, s \le n-1$ and $1 \le a, b, c, d \le n$. Lines over an object mean that the metric is being restricted to the boundary.

We set $\det g = \det g_{ab}$. We will denote by ∇_g or ∇ the covariant derivative and by Δ_g or Δ the Laplace–Beltrami operator. The full curvature tensor will be denoted by R_{abcd}, the Ricci tensor by R_{ab}, and the scalar curvature by R_g or R. The second fundamental form of the boundary will be denoted by h_{ij} and the mean curvature, $(n-1)^{-1}\,\mathrm{tr}(h_{ij})$, by h_g or h. We will denote the Weyl tensor by W_{abcd}.

We let $\mathbb{R}^n_+$ denote the half-space $\{x = (x_1, \dots, x_n) \in \mathbb{R}^n;\ x_n \ge 0\}$. If $x \in \mathbb{R}^n_+$ we set $\bar{x} = (x_1, \dots, x_{n-1}) \in \mathbb{R}^{n-1} \cong \partial\mathbb{R}^n$. We will denote by $B^+_\delta(0)$ (or B^+_δ for short) the half-ball $B_\delta(0) \cap \mathbb{R}^n_+$, where $B_\delta(0)$ is the Euclidean open ball of radius $\delta > 0$ centered at the origin of $\mathbb{R}^n$. Given a subset $\mathscr{C} \subset \mathbb{R}^n_+$, we set $\partial^+\mathscr{C} = \partial\mathscr{C} \cap (\mathbb{R}^n_+ \backslash \partial\mathbb{R}^n_+)$ and $\partial'\mathscr{C} = \mathscr{C} \cap \partial\mathbb{R}^n_+$.

We denote the volume forms of M and ∂M denoted by dv_g and $d\sigma_g$, respectively. The n-dimensional sphere of radius r in $\mathbb{R}^{n+1}$ will be denoted by S^n_r. We denote the volume of the n-dimensional unit sphere S^n_1 by σ_n.

For $\mathscr{C} \subset M$, we define the energy of a function u in $\mathscr{C}$ by

$$E_{\mathscr{C}}(u) = \int_{\mathscr{C}} \Big(|\nabla_g u|^2 + \frac{n-2}{4(n-1)} R_g u^2 \Big) dv_g + \frac{n-2}{2} \int_{\partial'\mathscr{C}} h_g u^2 d\sigma_g.$$

2. Coordinate expansions for the metric

In this section we will write expansions for the metric g in Fermi coordinates. We will also discuss the concept of conformal Fermi coordinates. The results of this section are basically proved on [Marques 2005, pages 1602–1609 and 1618].

Definition 2.1. Let $x_0 \in \partial M$. Choose geodesic normal coordinates $(x_1, \dots, x_{n-1})$ on the boundary, centered at x_0. We say that $(x_1, \dots, x_n)$ for small $x_n \ge 0$ are the Fermi coordinates (centered at x_0) of the point $\exp_x(x_n\eta(x)) \in M$. Here, we denote by $\eta(x)$ the inward unit vector normal to ∂M at $x \in \partial M$.

It is easy to see that in these coordinates, $g_{nn} \equiv 1$ and $g_{jn} \equiv 0$ for $j = 1, \dots, n-1$.

Fix $x_0 \in \partial M$. The existence of conformal Fermi coordinates is stated as follows:

Proposition 2.2. *For any given integer $N \geq 1$ there is a metric $\tilde{g}$, conformal to g, such that* $\det \tilde{g}(x) = 1 + O(|x|^N)$ *in $\tilde{g}$-Fermi coordinates centered at x_0. Moreover, $h_{\tilde{g}}(x) = O(|x|^{N-1})$.*

The first statement of Proposition 2.2 is [Marques 2005, Proposition 3.1]. The second follows from the equation

$$h_g = \frac{-1}{2(n-1)} g^{ij} g_{ij,n} = \frac{-1}{2(n-1)} (\log \det g)_{,n}.$$

The next three lemmas will also be used in the computations of the next section.

Lemma 2.3. *Suppose that ∂M is umbilic. Then, in conformal Fermi coordinates centered at x_0, we have $h_{ij}(x) = O(|x|^N)$, where N can be taken arbirarily large, and*

$$\begin{aligned} g^{ij}(x) = \delta_{ij} &+ \tfrac{1}{3}\overline{R}_{ikjl}x_k x_l + R_{ninj}x_n^2 + \tfrac{1}{6}\overline{R}_{ikjl;m}x_k x_l x_m + R_{ninj;k}x_n^2 x_k + \tfrac{1}{3}R_{ninj;n}x_n^3 \\ &+ (\tfrac{1}{20}\overline{R}_{ikjl;mp} + \tfrac{1}{15}\overline{R}_{iksl}\overline{R}_{jmsp})x_k x_l x_m x_p \\ &+ (\tfrac{1}{2}R_{ninj;kl} + \tfrac{1}{3}\operatorname{Sym}_{ij}(\overline{R}_{iksl}R_{nsnj}))x_n^2 x_k x_l \\ &+ \tfrac{1}{3}R_{ninj;nk}x_n^3 x_k + (\tfrac{1}{12}R_{ninj;nn} + \tfrac{2}{3}R_{nins}R_{nsnj})x_n^4 + O(|x|^5). \end{aligned}$$

Here, every coefficient is computed at x_0.

Lemma 2.4. *Suppose that ∂M is umbilic. Then, in conformal Fermi coordinates centered at x_0, we have these equalities at x_0:*

(i) $\overline{R}_{kl} = \operatorname{Sym}_{klm}(\overline{R}_{kl;m}) = 0$,

(ii) $R_{nn} = R_{nn;k} = \operatorname{Sym}_{kl}(R_{nn;kl}) = 0$,

(iii) $R_{nn;n} = 0$,

(iv) $\operatorname{Sym}_{klmp}(\frac{1}{2}\overline{R}_{kl;mp} + \frac{1}{9}\overline{R}_{ikjl}\overline{R}_{imjp}) = 0$,

(v) $R_{nn;nk} = 0$,

(vi) $R_{nn;nn} + 2(R_{ninj})^2 = 0$,

(vii) $R_{ij} = R_{ninj}$,

(viii) $R_{ijkn} = R_{ijkn;j} = 0$,

(ix) $R = R_{,j} = R_{,n} = 0$,

(x) $R_{,ii} = -\frac{1}{6}(\overline{W}_{ijkl})^2$,

(xi) $R_{ninj;ij} = -\frac{1}{2}R_{;nn} - (R_{ninj})^2$.

The idea for proving items (i)–(vi) of Lemma 2.4 is to express g_{ij} as the exponential of a matrix A_{ij}. Then we just observe that trace$(A_{ij}) = O(|x|^N)$ for any arbitrarily large integer N. Items (vii)–(xi) are applications of the Gauss and Codazzi equations and the Bianchi identity. Item (x) uses the fact that Fermi coordinates are normal on the boundary.

Lemma 2.5. *Suppose ∂M is umbilic. Then $W_{abcd}(x_0) = 0$ in conformal Fermi coordinates centered at $x_0 \in \partial M$ if and only if $R_{ninj}(x_0) = \overline{W}_{ijkl}(x_0) = 0$.*

Proof of Lemma 2.5. Recall that the Weyl tensor is defined by

$$W_{abcd} = R_{abcd} - \frac{1}{n-2}(R_{ac}g_{bd} - R_{ad}g_{bc} + R_{bd}g_{ac} - R_{bc}g_{ad}) + \frac{R}{(n-2)(n-1)}(g_{ac}g_{bd} - g_{ad}g_{bc}). \tag{2-1}$$

By the symmetries of the Weyl tensor, $W_{nnnn} = W_{nnni} = W_{nnij} = 0$. By the identity (2-1) and Lemma 2.4(viii), we have $W_{nijk}(x_0) = 0$. From the identity (2-1) again and from parts (ii), (vii) and (ix) of Lemma 2.4, we have

$$W_{ninj} = \frac{n-3}{n-2}R_{ninj}$$

and

$$W_{ijkl} = \overline{W}_{ijkl} - \frac{1}{n-2}(R_{nink}g_{jl} - R_{ninl}g_{jk} + R_{njnl}g_{ik} - R_{njnk}g_{il})$$

at x_0. In the last equation we also used the Gauss equation. The result follows. □

3. Estimating the Sobolev quotient

In this section, we will prove Theorem 1.3 by constructing a function ψ such that

$$Q(\psi) < Q(B^n, \partial B).$$

We first recall that the positive number $Q(B^n, \partial B)$ also appears as the best constant in the Sobolev-trace inequality

$$\left(\int_{\partial\mathbb{R}^n_+} |u|^{2(n-1)/(n-2)} d\bar{x}\right)^{(n-2)/n-1} \leq \frac{1}{Q(B^n, \partial B)} \int_{\mathbb{R}^n_+} |\nabla u|^2 dx$$

for every $u \in H^1(\mathbb{R}^n_+)$. Escobar [1988] and independently Beckner [1993] proved that the equality is achieved by the functions U_ϵ defined in (1-4). They are solutions to the boundary value problem

$$\begin{cases} \Delta U_\epsilon = 0 & \text{in } \mathbb{R}^n_+, \\ \dfrac{\partial U_\epsilon}{\partial y_n} + (n-2)U_\epsilon^{n/(n-2)} = 0 & \text{on } \partial\mathbb{R}^n_+. \end{cases} \tag{3-1}$$

One can check, using integration by parts, that

$$\int_{\mathbb{R}^n_+} |\nabla U_\epsilon|^2 dx = (n-2) \int_{\partial\mathbb{R}^n_+} U_\epsilon^{2(n-1)/(n-2)} dx$$

and also that

$$Q(B^n, \partial B) = (n-2)\Big(\int_{\partial\mathbb{R}^n_+} U_\epsilon^{2(n-1)/(n-2)} dx\Big)^{1/(n-1)}. \tag{3-2}$$

Assumption. In what remains, we will assume that ∂M is umbilic and there is a point $x_0 \in \partial M$ such that $W(x_0) \neq 0$.

Since the Sobolev quotient $Q(M, \partial M)$ is a conformal invariant, we can use conformal Fermi coordinates centered at x_0.

Convention. Henceforth, all the curvature terms are evaluated at x_0. We fix conformal Fermi coordinates centered at x_0 and work in a half-ball $B^+_{2\delta} = B^+_{2\delta}(0) \subset \mathbb{R}^n_+$.

In particular, for any arbitrarily large N, we can write the volume element dv_g as

$$dv_g = (1 + O(|x|^N))dx. \tag{3-3}$$

Often we will use that, for any homogeneous polynomial p_k of degree k,

$$\int_{S^{n-2}_r} p_k = \frac{r^2}{k(k+n-3)} \int_{S^{n-2}_r} \Delta p_k. \tag{3-4}$$

We will now construct the test function ψ. Set

$$\phi_\epsilon(x) = \epsilon^{(n-2)/2} A R_{ninj} x_i x_j x_n^2 ((\epsilon + x_n)^2 + |\bar{x}|^2)^{-n/2}, \tag{3-5}$$

for $A \in \mathbb{R}$ to be fixed later, and

$$\phi(y) = A R_{ninj} y_i y_j y_n^2 ((1 + y_n)^2 + |\bar{y}|^2)^{-n/2}. \tag{3-6}$$

Thus, $\phi_\epsilon(x) = \epsilon^{2-(n-2)/2} \phi(\epsilon^{-1}x)$. Set $U = U_1$. Thus, $U_\epsilon(x) = \epsilon^{-(n-2)/2} U(\epsilon^{-1}x)$. Note that $U_\epsilon(x) + \phi_\epsilon(x) = (1 + O(|x|^2))U_\epsilon(x)$. Hence, if δ is sufficiently small,

$$\tfrac{1}{2} U_\epsilon \leq U_\epsilon + \phi_\epsilon \leq 2U_\epsilon \quad \text{in } B^+_{2\delta}.$$

Let $r \mapsto \chi(r)$ be a smooth cut-off function satisfying $\chi(r) = 1$ for $0 \leq r \leq \delta$, $\chi(r) = 0$ for $r \geq 2\delta$ and $0 \leq \chi \leq 1$ and $|\chi'(r)| \leq C\delta^{-1}$ if $\delta \leq r \leq 2\delta$. Our test function is defined by

$$\psi(x) = \chi(|x|)(U_\epsilon(x) + \phi_\epsilon(x)).$$

3.1. ***Estimating the energy of ψ.*** The energy of ψ is given by

$$\begin{aligned} E_M(\psi) &= \int_M \Big(|\nabla_g\psi|^2 + \frac{n-2}{4(n-1)}R_g\psi^2\Big)dv_g + \frac{n-2}{2}\int_{\partial M} h_g\psi^2 d\sigma_g \\ &= E_{B^+_\delta}(\psi) + E_{B^+_{2\delta}\backslash B^+_\delta}(\psi). \end{aligned} \tag{3-7}$$

Observe that

$$|\nabla_g\psi|^2 \le C|\nabla\psi|^2 \le C|\nabla\chi|^2(U_\epsilon+\phi_\epsilon)^2 + C\chi^2|\nabla(U_\epsilon+\phi_\epsilon)|^2.$$

Hence,

$$\begin{aligned} E_{B^+_{2\delta}\backslash B^+_\delta}(\psi) \le C\int_{B^+_{2\delta}\backslash B^+_\delta} |\nabla\chi|^2U_\epsilon^2dx + C\int_{B^+_{2\delta}\backslash B^+_\delta} \chi^2|\nabla U_\epsilon|^2dx \\ + C\int_{B^+_{2\delta}\backslash B^+_\delta} R_gU_\epsilon^2dx + C\int_{\partial' B^+_{2\delta}\backslash \partial' B^+_\delta} h_gU_\epsilon^2d\bar{x}, \end{aligned}$$

Thus,

$$E_{B^+_{2\delta}\backslash B^+_\delta}(\psi) \le C\epsilon^{n-2}\delta^{2-n}. \tag{3-8}$$

The first term in the right hand side of (3-7) is

$$\begin{aligned} E_{B^+_\delta}(\psi) &= E_{B^+_\delta}(U_\epsilon+\phi_\epsilon) \\ &= \int_{B^+_\delta}\Big(|\nabla_g(U_\epsilon+\phi_\epsilon)|^2 + \frac{n-2}{4(n-1)}R_g(U_\epsilon+\phi_\epsilon)^2\Big)dv_g \\ &\qquad + \frac{n-2}{2}\int_{\partial' B^+_\delta} h_g(U_\epsilon+\phi_\epsilon)^2d\sigma_g \\ &= \int_{B^+_\delta}|\nabla(U_\epsilon+\phi_\epsilon)|^2dx \\ &\qquad + \int_{B^+_\delta}(g^{ij}-\delta^{ij})\partial_i(U_\epsilon+\phi_\epsilon)\partial_j(U_\epsilon+\phi_\epsilon)dx \\ &\qquad + \frac{n-2}{4(n-1)}\int_{B^+_\delta}R_g(U_\epsilon+\phi_\epsilon)^2dx + C\epsilon^{n-2}\delta. \end{aligned} \tag{3-9}$$

Here, we used the identity (3-3) for the volume term and Proposition 2.2 for the integral involving h_g.

We will treat the three integral terms in the right hand side of (3-9) in the next three lemmas.

Lemma 3.1. *We have*

$$
\begin{aligned}
\int_{B_\delta^+} &|\nabla(U_\epsilon+\phi_\epsilon)|^2 dx \\
&\le Q(B^n,\partial B^n)\Big(\int_{\partial M}\psi^{2(n-1)/(n-2)}dx\Big)^{(n-2)/(n-1)}+C\epsilon^{n-2}\delta^{2-n}\\
&\quad -\frac{4}{(n+1)(n-1)}\epsilon^4A^2(R_{ninj})^2\int_{B_{\delta\epsilon^{-1}}}\frac{y_n^2|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^n}dy\\
&\quad +\frac{8n}{(n+1)(n-1)}\epsilon^4A^2(R_{ninj})^2\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^3|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}}dy\\
&\quad +\frac{12n}{(n+1)(n-1)}\epsilon^4A^2(R_{ninj})^2\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^4|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}}dy.
\end{aligned}
$$

Proof. Since $R_{nn}=0$ by Lemma 2.4(ii), we have $\int_{S_r^{n-2}} R_{ninj}y_iy_j d\sigma_r(y)=0$. Thus we see that

$$
\int_{B_\delta^+}|\nabla(U_\epsilon+\phi_\epsilon)|^2dx=\int_{B_\delta^+}|\nabla U_\epsilon|^2dx+\int_{B_\delta^+}|\nabla\phi_\epsilon|^2dx. \tag{3-10}
$$

Integrating by parts equations (3-1) and using the identity (3-2) we obtain

$$
\begin{aligned}
\int_{B_\delta^+}|\nabla U_\epsilon|^2dx &\le Q(B^n,\partial B^n)\Big(\int_{\partial' B_\delta^+}U_\epsilon^{2(n-1)/(n-2)}dx\Big)^{(n-2)/(n-1)}\\
&\le Q(B^n,\partial B^n)\Big(\int_{\partial M}\psi^{2(n-1)/(n-2)}dx\Big)^{(n-2)/(n-1)}.
\end{aligned}
$$

In the first inequality above we used that $\partial U_\epsilon/\partial\eta>0$ on $\partial^+B_\delta^+$, where η denotes the inward unit normal vector. In the second we used that $\phi_\epsilon=0$ on ∂M.

For the second term in the right hand side of (3-10), an integration by parts plus a change of variables gives

$$
\int_{B_\delta^+}|\nabla\phi_\epsilon|^2dx\le-\epsilon^4\int_{B^+_{\epsilon^{-1}\delta}}(\Delta\phi)\phi dy+C\epsilon^{n-2}\delta^{2-n},
$$

Here we have used that $\int_{\partial' B_\delta^+}(\partial\phi_\epsilon/\partial x_n)\phi_\epsilon d\bar{x}=0$; the term $C\epsilon^{n-2}\delta^{2-n}$ comes from the integral over $\partial^+B_\delta^+$.

Claim. *The function ϕ satisfies*

$$
\begin{aligned}
\Delta\phi(y)=2AR_{ninj}y_iy_j((1+y_n)^2+|\bar{y}|^2)^{-n/2}&\\
-4nAR_{ninj}y_iy_jy_n((1+y_n)^2+|\bar{y}|^2)^{-(n+2)/2}&\\
-6nAR_{ninj}y_iy_jy_n^2((1+y_n)^2+|\bar{y}|^2)^{-(n+2)/2}&.
\end{aligned}
$$

Proof. We set $Z(y)=((1+y_n)^2+|\bar{y}|^2)$. Since $R_{nn}=0$,

$$\begin{aligned}
&\Delta(R_{ninj}y_iy_jy_n^2Z^{-n/2})\\
&\quad=\Delta(R_{ninj}y_iy_jy_n^2)Z^{-n/2}+R_{ninj}y_iy_jy_n^2\Delta(Z^{-n/2})\\
&\qquad+2\partial_k(R_{ninj}y_iy_jy_n^2)\partial_k(Z^{-n/2})+2\partial_n(R_{ninj}y_iy_jy_n^2)\partial_n(Z^{-n/2})\\
&\quad=2R_{ninj}y_iy_jZ^{-n/2}+2nR_{ninj}y_iy_jy_n^2Z^{-(n+2)/2}\\
&\qquad-4nR_{ninj}y_iy_jy_n^2Z^{-(n+2)/2}-4nR_{ninj}y_iy_jy_n(y_n+1)Z^{-(n+2)/2}\\
&\quad=2R_{ninj}y_iy_jZ^{-n/2}-6nR_{ninj}y_iy_jy_n^2Z^{-(n+2)/2}\\
&\qquad\qquad-4nR_{ninj}y_iy_jy_nZ^{-(n+2)/2}. \qquad\square
\end{aligned}$$

Using this claim, we get

$$\begin{aligned}
\int_{B^+_{\delta\epsilon^{-1}}}(\Delta\phi)\phi dy&=2A^2\int_{B^+_{\delta\epsilon^{-1}}}((1+y_n)^2+|\bar{y}|^2)^{-n}R_{ninj}R_{nknl}y_iy_jy_ky_ly_n^2dy\\
&\quad-4nA^2\int_{B^+_{\delta\epsilon^{-1}}}((1+y_n)^2+|\bar{y}|^2)^{-n-1}R_{ninj}R_{nknl}y_iy_jy_ky_ly_n^3dy\\
&\quad-6nA^2\int_{B^+_{\delta\epsilon^{-1}}}((1+y_n)^2+|\bar{y}|^2)^{-n-1}R_{ninj}R_{nknl}y_iy_jy_ky_ly_n^4dy.
\end{aligned}$$

Since $\Delta^2(R_{ninj}R_{nknl}y_iy_jy_ky_l)=16(R_{ninj})^2$,

$$\int_{S_r^{n-2}}R_{ninj}R_{nknl}y_iy_jy_ky_ld\sigma_r=\frac{2\sigma_{n-2}}{(n+1)(n-1)}r^{n+2}(R_{ninj})^2.$$

Thus

$$\begin{aligned}
\int_{B^+_{\delta\epsilon^{-1}}}(\Delta\phi)\phi dy&=\frac{4}{(n+1)(n-1)}A^2(R_{ninj})^2\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^2|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^n}dy\\
&\quad-\frac{8n}{(n+1)(n-1)}A^2(R_{ninj})^2\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^3|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}}dy\\
&\quad-\frac{12n}{(n+1)(n-1)}A^2(R_{ninj})^2\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^4|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}}dy.
\end{aligned}$$

Hence,

$$\begin{aligned}
\int_{B^+_{\delta}}|\nabla\phi_\epsilon|^2dx&\le-\frac{4}{(n+1)(n-1)}\epsilon^4A^2(R_{ninj})^2\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^2|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^n}dy\\
&\quad+\frac{8n}{(n+1)(n-1)}\epsilon^4A^2(R_{ninj})^2\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^3|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}}dy
\end{aligned}$$

$$+\frac{12n}{(n+1)(n-1)}\epsilon^4 A^2 (R_{ninj})^2 \int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^4|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}}dy + C\epsilon^{n-2}\delta^{2-n}. \quad \square$$

Lemma 3.2. *We have*

$$\begin{aligned}\int_{B^+_\delta}&(g^{ij}-\delta^{ij})\partial_i(U_\epsilon+\phi_\epsilon)\partial_j(U_\epsilon+\phi_\epsilon)dx\\ &=\frac{(n-2)^2}{(n+1)(n-1)}\epsilon^4 R_{ninj;ij}\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^2|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^n}dy\\ &\quad+\frac{(n-2)^2}{2(n-1)}\epsilon^4(R_{ninj})^2\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^4|\bar{y}|^2}{((1+y_n)^2+|\bar{y}|^2)^n}dy\\ &\quad-\frac{4n(n-2)}{(n+1)(n-1)}\epsilon^4 A(R_{ninj})^2\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^4|\bar{y}|^2}{((1+y_n)^2+|\bar{y}|^2)^{n+1}}dy+E_1,\end{aligned}$$

where

$$E_1=\begin{cases}O(\epsilon^4\delta^{-4}) & \text{if } n=6,\\ O(\epsilon^5\log(\delta\epsilon^{-1})) & \text{if } n=7,\\ O(\epsilon^5) & \text{if } n\ge 8.\end{cases}$$

Proof. Observe that

$$\begin{aligned}\int_{B^+_\delta}&(g^{ij}-\delta^{ij})\partial_i(U_\epsilon+\phi_\epsilon)\partial_j(U_\epsilon+\phi_\epsilon)dx\\ &=\int_{B^+_\delta}(g^{ij}-\delta^{ij})\partial_i U_\epsilon\partial_j U_\epsilon dx+2\int_{B^+_\delta}(g^{ij}-\delta^{ij})\partial_i U_\epsilon\partial_j\phi_\epsilon dx\\ &\quad+\int_{B^+_\delta}(g^{ij}-\delta^{ij})\partial_i\phi_\epsilon\partial_j\phi_\epsilon dx.\end{aligned}\tag{3-11}$$

We will handle separately the three terms in the right side of this. The first is

$$\begin{aligned}\int_{B^+_\delta}&(g^{ij}-\delta^{ij})(x)\partial_i U_\epsilon(x)\partial_j U_\epsilon(x)dx\\ &=\int_{B^+_{\delta\epsilon^{-1}}}(g^{ij}-\delta^{ij})(\epsilon y)\partial_i U(y)\partial_j U(y)dy\\ &=(n-2)^2\int_{B^+_{\delta\epsilon^{-1}}}((1+y_n)^2+|\bar{y}|^2)^{-n}(g^{ij}-\delta^{ij})(\epsilon y)y_i y_j dy.\end{aligned}$$

Hence, using Lemma A.1 we obtain

$$\begin{aligned}\int_{B^+_\delta}&(g^{ij}-\delta^{ij})(x)\partial_i U_\epsilon(x)\partial_j U_\epsilon(x)dx\\ &=\frac{(n-2)^2}{(n+1)(n-1)}\epsilon^4 R_{ninj;ij}\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^2|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^n}dy\\ &\quad+\frac{(n-2)^2}{2(n-1)}\epsilon^4(R_{ninj})^2\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^4|\bar{y}|^2}{((1+y_n)^2+|\bar{y}|^2)^n}dy+E_1',\end{aligned}$$

where

$$E_1'=\begin{cases}O(\epsilon^4\delta) & \text{if } n=6,\\ O(\epsilon^5\log(\delta\epsilon^{-1})) & \text{if } n=7,\\ O(\epsilon^5) & \text{if } n\geq 8.\end{cases}$$

The second term is

$$\begin{aligned}2\int_{B^+_\delta}&(g^{ij}-\delta^{ij})(x)\partial_i U_\epsilon(x)\partial_j\phi_\epsilon(x)dx\\ &=-2\int_{B^+_\delta}(g^{ij}-\delta^{ij})(x)\partial_i\partial_j U_\epsilon(x)\phi_\epsilon(x)dx\\ &\quad-2\int_{B^+_\delta}(\partial_i g^{ij})(x)\partial_j U_\epsilon(x)\phi_\epsilon(x)dx+O(\epsilon^{n-2}\delta^{2-n})\\ &=-2\epsilon^2\int_{B^+_{\delta\epsilon^{-1}}}(g^{ij}-\delta^{ij})(\epsilon y)\partial_i\partial_j U(y)\phi(y)dy\\ &\quad-2\epsilon^3\int_{B^+_{\delta\epsilon^{-1}}}(\partial_i g^{ij})(\epsilon y)\partial_j U(y)\phi(y)dy+O(\epsilon^{n-2}\delta^{2-n}).\end{aligned}\tag{3-12}$$

But

$$\begin{aligned}-2\epsilon^2&\int_{B^+_{\delta\epsilon^{-1}}}(g^{ij}-\delta^{ij})(\epsilon y)\partial_i\partial_j U(y)\phi(y)dy\\ &=-2(n-2)\epsilon^2 A\int_{B^+_{\delta\epsilon^{-1}}}((1+y_n)^2+|\bar{y}|^2)^{-n-1}(g^{ij}-\delta^{ij})(\epsilon y)\\ &\qquad\qquad\cdot\big(ny_iy_j-((1+y_n)^2+|\bar{y}|^2)\delta_{ij}\big)R_{nknl}y_ky_ly_n^2dy\\ &=-\frac{4n(n-2)}{(n+1)(n-1)}\epsilon^4A(R_{ninj})^2\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^4|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}}dy+E_2',\end{aligned}$$

where

$$E_2'=\begin{cases}O(\epsilon^4\delta) & \text{if } n=6,\\ O(\epsilon^5\log(\delta\epsilon^{-1})) & \text{if } n=7,\\ O(\epsilon^5) & \text{if } n\geq 8\end{cases}$$

and in the last equality, we used Lemma A.2 and the fact that Lemma 2.3, together with parts (i), (ii) and (iii) of Lemma 2.4, implies

$$\int_{S_r^{n-2}} (g^{ij}-\delta^{ij})(\epsilon y)\delta_{ij} R_{nknl} y_k y_l d\sigma_r(y) = \int_{S_r^{n-2}} O(\epsilon^4|y|^4) R_{nknl} y_k y_l d\sigma_r(y).$$

We also have, by Lemma 2.3 and Lemma 2.4(i),

$$-2\epsilon^3 \int_{B^+_{\delta\epsilon^{-1}}} (\partial_i g^{ij})(\epsilon y)\partial_j U(y)\phi(y)dy = E_3' = \begin{cases} O(\epsilon^4\delta) & \text{if } n=6,\\ O(\epsilon^5\log(\delta\epsilon^{-1})) & \text{if } n=7,\\ O(\epsilon^5) & \text{if } n\geq 8.\end{cases}$$

Hence

$$\begin{aligned}2\int_{B^+_\delta} &(g^{ij}-\delta^{ij})(x)\partial_i U_\epsilon(x)\partial_j\phi_\epsilon(x)dx\\ &= E_2' + E_3' - \frac{4n(n-2)}{(n+1)(n-1)}\epsilon^4 A(R_{ninj})^2 \int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^4|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}}dy.\end{aligned}$$

Finally, the third term in the right hand side of (3-11) is written as

$$\begin{aligned}\int_{B^+_\delta} (g^{ij}-\delta^{ij})(x)\partial_i\phi_\epsilon(x)\partial_j\phi_\epsilon(x)dx &= \epsilon^4\int_{B^+_{\delta\epsilon^{-1}}} (g^{ij}-\delta^{ij})(\epsilon y)\partial_i\phi(y)\partial_j\phi(y)dy\\ &= \begin{cases} O(\epsilon^4\delta) & \text{if } n=6,\\ O(\epsilon^5\log(\delta\epsilon^{-1})) & \text{if } n=7,\\ O(\epsilon^5) & \text{if } n\geq 8.\end{cases}\end{aligned}$$

The result now follows if we choose ϵ small enough that $\log(\delta\epsilon^{-1}) > \delta^{2-n}$. □

Lemma 3.3. *We have*

$$\begin{aligned}\frac{n-2}{4(n-1)}\int_{B^+_\delta} R_g(U_\epsilon+\phi_\epsilon)^2dx = &\\ &\frac{n-2}{8(n-1)}\epsilon^4 R_{;nn}\int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^2}{((1+y_n)^2+|\bar{y}|^2)^{n-2}}dy\\ &- \frac{n-2}{24(n-1)^2}\epsilon^4(\overline{W}_{ijkl})^2\int_{B^+_{\delta\epsilon^{-1}}} \frac{|\bar{y}|^2}{((1+y_n)^2+|\bar{y}|^2)^{n-2}}dy + E_2,\end{aligned}$$

where

$$E_2 = \begin{cases} O(\epsilon^4\delta) & \text{if } n=6,\\ O(\epsilon^5\log(\delta\epsilon^{-1})) & \text{if } n=7,\\ O(\epsilon^5) & \text{if } n\geq 8.\end{cases}$$

Proof. We first observe that

$$\int_{B^+_\delta} R_g(U_\epsilon+\phi_\epsilon)^2 dx = \int_{B^+_\delta} R_g U_\epsilon^2 dx + 2\int_{B^+_\delta} R_g U_\epsilon \phi_\epsilon dx + \int_{B^+_\delta} R_g \phi_\epsilon^2 dx. \tag{3-13}$$

We will handle each term in the right hand side of (3-13) separately. Using Lemma A.3, we see that the first is

$$\begin{aligned}\int_{B^+_\delta} R_g(x)U_\epsilon(x)^2 dx &= \epsilon^2 \int_{B^+_{\delta\epsilon^{-1}}} R_g(\epsilon y)U^2(y)dy \\ &= \tfrac{1}{2}\epsilon^4 R_{;nn}\int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^2}{((1+y_n)^2+|\bar{y}|^2)^{n-2}}dy + E_4' \qquad (3\text{-}14)\\ &\quad - \frac{1}{12(n-1)}\epsilon^4(\overline{W}_{ijkl})^2\int_{B^+_{\delta\epsilon^{-1}}} \frac{|\bar{y}|^2}{((1+y_n)^2+|\bar{y}|^2)^{n-2}}dy,\end{aligned}$$

where

$$E_4' = \begin{cases} O(\epsilon^4\delta) & \text{if } n=6,\\ O(\epsilon^5\log(\delta\epsilon^{-1})) & \text{if } n=7,\\ O(\epsilon^5) & \text{if } n\geq 8.\end{cases}$$

By Lemma 2.4(ix), the second term is

$$\begin{aligned}2\int_{B^+_\delta} R_g(x)U_\epsilon(x)\phi_\epsilon(x)dx &= 2\epsilon^4\int_{B^+_{\delta\epsilon^{-1}}} R_g(\epsilon y)U(y)\phi(y)dy \\ &= \begin{cases} O(\epsilon^4\delta) & \text{if } n=6,\\ O(\epsilon^5\log(\delta\epsilon^{-1})) & \text{if } n=7,\\ O(\epsilon^5) & \text{if } n\geq 8\end{cases}\end{aligned}$$

and the last term is

$$\int_{B^+_\delta} R_g\phi_\epsilon^2 dx = \begin{cases} O(\epsilon^4\delta) & \text{if } n=6,\\ O(\epsilon^5\log(\delta\epsilon^{-1})) & \text{if } n=7,\\ O(\epsilon^5) & \text{if } n\geq 8.\end{cases}$$

□

Proof of Theorem 1.3. It follows from Lemmas 3.1, 3.2 and 3.3 and the identities (3-7), (3-8) and (3-9) that

$$\begin{aligned}E_M(\psi) &\leq Q(B^n,\partial B^n)\Big(\int_{\partial M}\psi^{2(n-1)/(n-2)}\Big)^{(n-2)/(n-1)} + E \qquad (3\text{-}15)\\ &\quad -\epsilon^4\frac{4A^2}{(n+1)(n-1)}(R_{ninj})^2\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^2|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^n}dy \\ &\quad +\epsilon^4\frac{(n-2)^2}{(n+1)(n-1)}R_{ninj;ij}\int_{B^+_{\delta\epsilon^{-1}}}\frac{y_n^2|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^n}dy\end{aligned}$$

$$+\epsilon^4 \frac{8nA^2}{(n+1)(n-1)}(R_{ninj})^2 \int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^3|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}} dy$$
$$+\epsilon^4 \frac{12nA^2}{(n+1)(n-1)}(R_{ninj})^2 \int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^4|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}} dy$$
$$-\epsilon^4 \frac{4n(n-2)A}{(n+1)(n-1)}(R_{ninj})^2 \int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^4|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}} dy$$
$$+\epsilon^4 \frac{(n-2)^2}{2(n-1)}(R_{ninj})^2 \int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^4|\bar{y}|^2}{((1+y_n)^2+|\bar{y}|^2)^{n}} dy$$
$$+\epsilon^4 \frac{n-2}{8(n-1)} R_{;nn} \int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^2}{((1+y_n)^2+|\bar{y}|^2)^{n-2}} dy$$
$$-\epsilon^4 \frac{n-2}{48(n-1)^2}(\overline{W}_{ijkl})^2 \int_{B^+_{\delta\epsilon^{-1}}} \frac{|\bar{y}|^2}{((1+y_n)^2+|\bar{y}|^2)^{n-2}} dy.$$

where

$$E = \begin{cases} O(\epsilon^4\delta^{-4}) & \text{if } n=6, \\ O(\epsilon^5 \log(\delta\epsilon^{-1})) & \text{if } n=7, \\ O(\epsilon^5) & \text{if } n \geq 8. \end{cases}$$

We divide the rest of the proof into two cases.

The case $n = 7, 8$. Set $I = \int_0^\infty r^n/(r^2+1)^n dr$. We will apply the change of variables $\bar{z} = (1+y_n)^{-1}\bar{y}$ and Lemmas B.1 and B.2 to compare the different integrals in the expansion (3-15).

These integrals are

$$I_1 = \int_{\mathbb{R}^n_+} \frac{y_n^2|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^n} dy_n d\bar{y} = \int_0^\infty y_n^2(1+y_n)^{3-n} dy_n \int_{\mathbb{R}^{n-1}} \frac{|\bar{z}|^4}{(1+|\bar{z}|^2)^n} d\bar{z}$$
$$= \frac{2(n+1)\sigma_{n-2} I}{(n-3)(n-4)(n-5)(n-6)},$$
$$I_2 = \int_{\mathbb{R}^n_+} \frac{y_n^3|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}} dy_n d\bar{y} = \int_0^\infty y_n^3(1+y_n)^{1-n} dy_n \int_{\mathbb{R}^{n-1}} \frac{|\bar{z}|^4}{(1+|\bar{z}|^2)^{n+1}} d\bar{z}$$
$$= \frac{3(n+1)\sigma_{n-2} I}{n(n-2)(n-3)(n-4)(n-5)},$$
$$I_3 = \int_{\mathbb{R}^n_+} \frac{y_n^4|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}} dy_n d\bar{y} = \int_0^\infty y_n^4(1+y_n)^{1-n} dy_n \int_{\mathbb{R}^{n-1}} \frac{|\bar{z}|^4}{(1+|\bar{z}|^2)^{n+1}} d\bar{z}$$
$$= \frac{12(n+1)\sigma_{n-2} I}{n(n-2)(n-3)(n-4)(n-5)(n-6)},$$

$$I_4=\int_{\mathbb{R}^n_+}\frac{y_n^4|\bar{y}|^2}{((1+y_n)^2+|\bar{y}|^2)^n}dy_n d\bar{y}=\int_0^\infty y_n^4(1+y_n)^{1-n}dy_n\int_{\mathbb{R}^{n-1}}\frac{|\bar{z}|^2}{(1+|\bar{z}|^2)^n}d\bar{z}$$
$$=\frac{24\sigma_{n-2}I}{(n-2)(n-3)(n-4)(n-5)(n-6)},$$
$$I_5=\int_{\mathbb{R}^n_+}\frac{y_n^2}{((1+y_n)^2+|\bar{y}|^2)^{n-2}}dy_n d\bar{y}=\int_0^\infty y_n^2(1+y_n)^{3-n}dy_n\int_{\mathbb{R}^{n-1}}\frac{1}{(1+|\bar{z}|^2)^{n-2}}d\bar{z}$$
$$=\frac{8(n-2)\sigma_{n-2}I}{(n-3)(n-4)(n-5)(n-6)}.$$

Thus,

$$\begin{aligned}E_M(\psi)\le{}& Q(B^n,\partial B^n)\Big(\int_{\partial M}\psi^{2(n-1)/(n-2)}\Big)^{(n-2)/(n-1)}+E'\\&+\epsilon^4\Big(-\frac{4A^2}{(n+1)(n-1)}I_1+\frac{8nA^2}{(n+1)(n-1)}I_2+\frac{(n-2)^2}{2(n-1)}I_4\Big)(R_{ninj})^2\\&+\epsilon^4\Big(\frac{12nA^2}{(n+1)(n-1)}-\frac{4n(n-2)A}{(n+1)(n-1)}\Big)I_3\cdot(R_{ninj})^2\\&+\epsilon^4\frac{(n-2)^2}{(n+1)(n-1)}I_1\cdot R_{ninj;ij}+\epsilon^4\frac{n-2}{8(n-1)}I_5\cdot R_{;nn}\\&-\epsilon^4\frac{n-2}{48(n-1)^2}(\overline{W}_{ijkl})^2\int_{\mathbb{R}^n_+}\frac{|\bar{y}|^2}{((1+y_n)^2+|\bar{y}|^2)^{n-2}}dy.\end{aligned}\tag{3-16}$$

where

$$E'=\begin{cases}O(\epsilon^5\log(\delta\epsilon^{-1})) & \text{if } n=7,\\ O(\epsilon^5) & \text{if } n=8.\end{cases}$$

Using Lemma 2.4(xi) and substituting the expressions obtained for $I_1,\dots,I_5$ in the expansion (3-16), the coefficients of $R_{ninj;ij}$ and $R_{;nn}$ cancel out and we obtain

$$\begin{aligned}E_M(\psi)&\\ \le{}& Q(B^n,\partial B^n)\Big(\int_{\partial M}\psi^{2(n-1)/(n-2)}\Big)^{(n-2)/(n-1)}+E'\\&+\epsilon^4\sigma_{n-2}I\cdot\gamma\big(16(n+1)A^2-48(n-2)A+2(8-n)(n-2)^2\big)(R_{ninj})^2\\&-\epsilon^4\frac{n-2}{48(n-1)^2}(\overline{W}_{ijkl})^2\int_{\mathbb{R}^n_+}\frac{|\bar{y}|^2}{((1+y_n)^2+|\bar{y}|^2)^{n-2}}dy,\end{aligned}\tag{3-17}$$

where

$$1/\gamma=(n-1)(n-2)(n-3)(n-4)(n-5)(n-6).$$

Choosing $A=1$, the term $16(n+1)A^2-48(n-2)A+2(8-n)(n-2)^2$ in the expansion (3-17) is -62 for $n=7$ and -144 for $n=8$. Thus, for small ϵ, since $W_{abcd}(x_0) \neq 0$, the expansion (3-17) together with Lemma 2.5 implies that, in dimensions 7 and 8,

$$E_M(\psi) < Q(B^n, \partial B^n)\Big(\int_{\partial M} \psi^{2(n-1)/(n-2)}\Big)^{(n-2)/(n-1)}.$$

The case $n = 6$. We will again apply the change of variables $\bar{z} = (1+y_n)^{-1}\bar{y}$ and Lemma B.1 to compare the different integrals in the expansion (3-15). In the next estimates we are always assuming $n = 6$.

In this case, the first integral is

$$\begin{aligned} I_{1,\delta/\epsilon} &= \int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^2|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^n} dy_n d\bar{y} \\ &= \int_{B^+_{\delta\epsilon^{-1}} \cap \{y_n \leq \delta/2\epsilon\}} \frac{y_n^2|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^n} dy_n d\bar{y} + O(1) \\ &= \int_{\mathbb{R}^n_+ \cap \{y_n \leq \delta/2\epsilon\}} \frac{y_n^2|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^n} dy_n d\bar{y} + O(1). \end{aligned}$$

Hence

$$\begin{aligned} I_{1,\delta/\epsilon} &= \int_0^{\delta/2\epsilon} y_n^2(1+y_n)^{3-n} dy_n \int_{\mathbb{R}^{n-1}} \frac{|\bar{z}|^4}{(1+|\bar{z}|^2)^n} d\bar{z} + O(1) \\ &= \log(\delta\epsilon^{-1})\frac{n+1}{n-3}\sigma_{n-2} I + O(1). \end{aligned}$$

The second integral is

$$I_{2,\delta/\epsilon} = \int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^3|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}} dy_n d\bar{y} = O(1).$$

The others are similar to $I_{1,\delta/\epsilon}$:

$$\begin{aligned} I_{3,\delta/\epsilon} &= \int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^4|\bar{y}|^4}{((1+y_n)^2+|\bar{y}|^2)^{n+1}} dy_n d\bar{y} \\ &= \int_0^{\delta/2\epsilon} y_n^4(1+y_n)^{1-n} dy_n \int_{\mathbb{R}^{n-1}} \frac{|\bar{z}|^4}{(1+|\bar{z}|^2)^{n+1}} d\bar{z} + O(1) \\ &= \log(\delta\epsilon^{-1})\frac{n+1}{2n}\sigma_{n-2} I + O(1), \\ I_{4,\delta/\epsilon} &= \int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^4|\bar{y}|^2}{((1+y_n)^2+|\bar{y}|^2)^n} dy_n d\bar{y} \\ &= \int_0^{\delta/2\epsilon} y_n^4(1+y_n)^{1-n} dy_n \int_{\mathbb{R}^{n-1}} \frac{|\bar{z}|^2}{(1+|\bar{z}|^2)^n} d\bar{z} \\ &= \log(\delta\epsilon^{-1})\sigma_{n-2} I + O(1), \end{aligned}$$

$$
\begin{aligned}
I_{5,\delta/\epsilon} &= \int_{B^+_{\delta\epsilon^{-1}}} \frac{y_n^2}{((1+y_n)^2+|\bar{y}|^2)^{n-2}} dy_n d\bar{y} \\
&= \int_0^{\delta/2\epsilon} y_n^2(1+y_n)^{3-n} dy_n \int_{\mathbb{R}^{n-1}} \frac{1}{(1+|\bar{z}|^2)^{n-2}} d\bar{z} + O(1) \\
&= \log(\delta\epsilon^{-1}) \frac{4(n-2)}{n-3} \sigma_{n-2} I + O(1), \\
I_{6,\delta/\epsilon} &= \int_{B^+_{\delta\epsilon^{-1}}} \frac{|\bar{y}|^2}{((1+y_n)^2+|\bar{y}|^2)^{n-2}} dy_n d\bar{y} \\
&= \int_0^{\delta/2\epsilon} (1+y_n)^{5-n} dy_n \int_{\mathbb{R}^{n-1}} \frac{|\bar{z}|^2}{(1+|\bar{z}|^2)^{n-2}} d\bar{z} + O(1) \\
&= \log(\delta\epsilon^{-1}) \frac{4(n-1)(n-2)}{(n-3)(n-5)} \sigma_{n-2} I + O(1).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
E_M(\psi) \le\ & Q(B^n, \partial B^n)\Big(\int_{\partial M} \psi^{2(n-1)/(n-2)}\Big)^{(n-2)/(n-1)} + O(\epsilon^4\delta^{-4}) \\
& + \epsilon^4\Big(-\frac{4A^2}{(n+1)(n-1)} I_{1,\delta/\epsilon} + \frac{(n-2)^2}{2(n-1)} I_{4,\delta/\epsilon}\Big)(R_{ninj})^2 \\
& + \epsilon^4\Big(\frac{12nA^2}{(n+1)(n-1)} - \frac{4n(n-2)A}{(n+1)(n-1)}\Big) I_{3,\delta/\epsilon} \cdot (R_{ninj})^2 \\
& + \epsilon^4 \frac{(n-2)^2}{(n+1)(n-1)} I_{1,\delta/\epsilon} \cdot R_{ninj;ij} + \epsilon^4 \frac{n-2}{8(n-1)} I_{5,\delta/\epsilon} \cdot R_{;nn} \\
& - \epsilon^4 \frac{n-2}{48(n-1)^2} I_{6,\delta/\epsilon} \cdot (\overline{W}_{ijkl})^2.
\end{aligned} \tag{3-18}
$$

Using Lemma 2.4(xi) and substituting the expressions obtained for $I_{1,\delta/\epsilon}$ through $I_{6,\delta/\epsilon}$ in expansion (3-18), we find the coefficients of $R_{ninj;ij}$ and $R_{;nn}$ cancel out and obtain

$$
\begin{aligned}
E_M(\psi) \le\ & Q(B^n, \partial B^n)\Big(\int_{\partial M} \psi^{2(n-1)/(n-2)}\Big)^{(n-2)/(n-1)} + O(\epsilon^4\delta^{-4}) \\
& + \epsilon^4 \log(\delta\epsilon^{-1})\sigma_{n-2} I \\
& \cdot \Big(\frac{6(n-3)-4}{(n-1)(n-3)} A^2 - \frac{2(n-2)}{n-1} A + \frac{(n-2)^2(n-5)}{2(n-1)(n-3)}\Big)(R_{ninj})^2 \\
& - \epsilon^4 \log(\delta\epsilon^{-1})\sigma_{n-2} I \frac{(n-2)^2}{12(n-1)(n-3)(n-5)} (\overline{W}_{ijkl})^2.
\end{aligned} \tag{3-19}
$$

Choosing $A = 1$, the term to the left of $(R_{ninj})^2$ in the expansion (3-19) is $-2/15$ for $n = 6$. Thus, for small ϵ, since $W_{abcd}(x_0) \neq 0$, the expansion (3-19) together with Lemma 2.5 implies that, in dimension $n = 6$

$$E_M(\psi) < Q(B^n, \partial B^n)\Big(\int_{\partial M} \psi^{2(n-1)/(n-2)}\Big)^{(n-2)/(n-1)}. \qquad \square$$

Appendix A.

In this section, we will use the results of Section 2 to calculate some integrals used in the computations of Section 3. As before, all curvature coefficients are evaluated at $x_0 \in \partial M$, around which we center conformal Fermi coordinates.

Lemma A.1. *We have*

$$\int_{S_r^{n-2}} (g^{ij} - \delta^{ij})(\epsilon y) y_i y_j d\sigma_r(y) = \sigma_{n-2}\epsilon^4 \frac{y_n^2 r^{n+2}}{(n+1)(n-1)} R_{ninj;ij} + \sigma_{n-2}\epsilon^4 \frac{y_n^4 r^n}{2(n-1)} (R_{ninj})^2 + O(\epsilon^5 |(r, y_n)|^{n+5}).$$

Proof. By Lemma 2.3,

$$\int_{S_r^{n-2}} (g^{ij} - \delta^{ij})(\epsilon y) y_i y_j d\sigma_r(y) = \epsilon^4 \int_{S_r^{n-2}} \tfrac{1}{2} R_{ninj;kl} y_i y_j y_k y_l d\sigma_r(y) + O(\epsilon^5 |(r, y_n)|^{n+5}) + \epsilon^4 y_n^2 \int_{S_r^{n-2}} (\tfrac{1}{12} R_{ninj;nn} + \tfrac{2}{3} R_{nins} R_{nsnj}) y_i y_j d\sigma_r(y).$$

Then we just use the identity (3-4), Lemma 2.4 and the fact that

$$\Delta^2(R_{ninj;kl} y_i y_j y_k y_l) = 16 R_{ninj;ij}. \qquad \square$$

Lemma A.2. *We have*

$$\int_{S_r^{n-2}} (g^{ij} - \delta^{ij})(\epsilon y) R_{nknl} y_i y_j y_k y_l d\sigma_r(y) = \frac{2}{(n+1)(n-1)} \sigma_{n-2}\epsilon^2 y_n^2 r^{n+2} (R_{ninj})^2 + O(\epsilon^5 |(r, y_n)|^{n+5})$$

Proof. As in Lemma A.1, the result follows from

$$\int_{S_r^{n-2}} (g^{ij} - \delta^{ij})(\epsilon y) R_{nknl} y_i y_j y_k y_l d\sigma_r(y) = \epsilon^2 y_n^2 \int_{S_r^{n-2}} R_{ninj} R_{nknl} y_i y_j y_k y_l d\sigma_r(y) + O(\epsilon^5 |(r, y_n)|^{n+5}),$$

the fact that $\Delta^2(R_{ninj} R_{nknl} y_i y_j y_k y_l) = 16 (R_{ninj})^2$, and the identity (3-4). $\square$

Lemma A.3. *We have*

$$\int_{S_r^{n-2}} R_g(\epsilon y)d\sigma_r(y) = \sigma_{n-2}\epsilon^2\Big(\tfrac{1}{2}y_n^2 r^{n-2}R_{;nn} - \frac{1}{12(n-1)}r^n(\overline{W}_{ijkl})^2\Big) + O(\epsilon^3|(r,y_n)|^{n+1}).$$

Proof. As in Lemma A.1, the result follows from Lemma 2.4(x), (3-4), and

$$\int_{S_r^{n-2}} R_g(\epsilon y)d\sigma_r(y) = \epsilon^2 y_n^2\int_{S_r^{n-2}}\tfrac{1}{2}R_{;nn}d\sigma_r(y) + \epsilon^2\int_{S_r^{n-2}}\tfrac{1}{2}R_{;ij}y_iy_jd\sigma_r(y) + O(\epsilon^3|(r,y_n)|^{n+1}). \qquad \square$$

Appendix B.

Finally, we prove results used in the computations of Section 3.

Lemma B.1. *We have*

(a) $\displaystyle\int_0^\infty \frac{s^\alpha ds}{(1+s^2)^m} = \frac{2m}{\alpha+1}\int_0^\infty \frac{s^{\alpha+2}ds}{(1+s^2)^{m+1}}$ *for* $\alpha+1<2m$;

(b) $\displaystyle\int_0^\infty \frac{s^\alpha ds}{(1+s^2)^m} = \frac{2m}{2m-\alpha-1}\int_0^\infty \frac{s^{\alpha}ds}{(1+s^2)^{m+1}}$ *for* $\alpha+1<2m$;

(c) $\displaystyle\int_0^\infty \frac{s^\alpha ds}{(1+s^2)^m} = \frac{2m-\alpha-3}{\alpha+1}\int_0^\infty \frac{s^{\alpha+2}ds}{(1+s^2)^{m}}$ *for* $\alpha+3<2m$.

Proof. Integrating by parts, we get

$$\int_0^\infty \frac{s^{\alpha+2}ds}{(1+s^2)^{m+1}} = \int_0^\infty s^{\alpha+1}\frac{s\,ds}{(1+s^2)^{m+1}} = \frac{\alpha+1}{2m}\int_0^\infty \frac{s^\alpha ds}{(1+s^2)^m}$$

for $\alpha+1<2m$, which proves item (a).

Item (b) follows from (a) and from

$$\int_0^\infty \frac{s^\alpha ds}{(1+s^2)^m} = \int_0^\infty \frac{s^\alpha(1+s^2)}{(1+s^2)^{m+1}}ds = \int_0^\infty \frac{s^\alpha ds}{(1+s^2)^{m+1}} + \int_0^\infty \frac{s^{\alpha+2}ds}{(1+s^2)^{m+1}}.$$

To prove (c), observe that by (a),

$$\int_0^\infty \frac{s^\alpha ds}{(1+s^2)^{m-1}} = \frac{2(m-1)}{\alpha+1}\int_0^\infty \frac{s^{\alpha+2}ds}{(1+s^2)^m},$$

for $\alpha+3<2m$. But by (b), we have

$$\int_0^\infty \frac{s^\alpha ds}{(1+s^2)^{m-1}} = \frac{2(m-1)}{2(m-1)-\alpha-1}\int_0^\infty \frac{s^\alpha ds}{(1+s^2)^m}. \qquad \square$$

Lemma B.2. $\displaystyle\int_0^\infty \frac{t^k}{(1+t)^m}dt = \frac{k!}{(m-1)(m-2)\cdots(m-1-k)}$ *for* $m>k+1$.

Proof. The proof follows straightforwardly from treating the integral

$$\int_0^\infty \frac{t^{k-1}}{(1+t)^{m-1}} dt$$

two ways: first, by integrating it by parts, and second by borrowing a factor of $(1+t)$ from its denominator and dividing that integral into two. □

Acknowledgments

This paper is taken from part of my doctoral thesis [Almaraz 2009] at IMPA. I would like to express my gratitude to my advisor Professor Fernando C. Marques for numerous mathematical conversations and constant encouragement.

References

[Ahmedou 2003] M. O. Ahmedou, "A Riemann mapping type theorem in higher dimensions, I: The conformally flat case with umbilic boundary", pp. 1–18 in *Nonlinear equations: Methods, models and applications* (Bergamo, 2001), edited by D. Lupo et al., Progr. Nonlinear Differential Equations Appl. **54**, Birkhäuser, Basel, 2003. MR 2004m:58036 Zbl 1041.58010

[Almaraz 2009] S. Almaraz, *Existence and compactness theorems for the Yamabe problem on manifolds with boundary*, thesis, IMPA, Brazil, 2009.

[Ambrosetti et al. 2002] A. Ambrosetti, Y. Li, and A. Malchiodi, "On the Yamabe problem and the scalar curvature problems under boundary conditions", *Math. Ann.* **322**:4 (2002), 667–699. MR 2003e:53038 Zbl 1005.53034

[Aubin 1976] T. Aubin, "Équations différentielles non linéaires et problème de Yamabe concernant la courbure scalaire", *J. Math. Pures Appl.* (9) **55**:3 (1976), 269–296. MR 55 #4288 Zbl 0336.53033

[Beckner 1993] W. Beckner, "Sharp Sobolev inequalities on the sphere and the Moser–Trudinger inequality", *Ann. of Math.* (2) **138**:1 (1993), 213–242. MR 94m:58232 Zbl 0826.58042

[Ben Ayed et al. 2005] M. Ben Ayed, K. El Mehdi, and M. Ould Ahmedou, "The scalar curvature problem on the four dimensional half sphere", *Calc. Var. Partial Differential Equations* **22**:4 (2005), 465–482. MR 2006i:35079 Zbl 1130.35048

[Brendle 2002] S. Brendle, "A generalization of the Yamabe flow for manifolds with boundary", *Asian J. Math.* **6**:4 (2002), 625–644. MR 2003m:53052 Zbl 1039.53035

[Brendle 2007] S. Brendle, "Convergence of the Yamabe flow in dimension 6 and higher", *Invent. Math.* **170**:3 (2007), 541–576. MR 2008k:53136 Zbl 1130.53044

[Djadli et al. 2003] Z. Djadli, A. Malchiodi, and M. Ould Ahmedou, "Prescribing scalar and boundary mean curvature on the three dimensional half sphere", *J. Geom. Anal.* **13**:2 (2003), 255–289. MR 2004d:53032 Zbl 1092.53028

[Djadli et al. 2004] Z. Djadli, A. Malchiodi, and M. O. Ahmedou, "The prescribed boundary mean curvature problem on $\mathbb{B}^4$", *J. Differential Equations* **206**:2 (2004), 373–398. MR 2005h:35117 Zbl 1108.35070

[Escobar 1988] J. F. Escobar, "Sharp constant in a Sobolev trace inequality", *Indiana Univ. Math. J.* **37**:3 (1988), 687–698. MR 90a:46071 Zbl 0666.35014

[Escobar 1992a] J. F. Escobar, "Conformal deformation of a Riemannian metric to a scalar flat metric with constant mean curvature on the boundary", *Ann. of Math.* (2) **136**:1 (1992), 1–50. MR 93e:53046 Zbl 0766.53033

[Escobar 1992b] J. F. Escobar, "The Yamabe problem on manifolds with boundary", *J. Differential Geom.* **35**:1 (1992), 21–84. MR 93b:53030 Zbl 0771.53017

[Escobar 1996a] J. F. Escobar, "Conformal deformation of a Riemannian metric to a constant scalar curvature metric with constant mean curvature on the boundary", *Indiana Univ. Math. J.* **45**:4 (1996), 917–943. MR 98d:53051 Zbl 0881.53037

[Escobar 1996b] J. F. Escobar, "Conformal metrics with prescribed mean curvature on the boundary", *Calc. Var. Partial Differential Equations* **4**:6 (1996), 559–592. MR 97h:53040 Zbl 0867.53034

[Escobar and Garcia 2004] J. F. Escobar and G. Garcia, "Conformal metrics on the ball with zero scalar curvature and prescribed mean curvature on the boundary", *J. Funct. Anal.* **211**:1 (2004), 71–152. MR 2005f:53050 Zbl 1056.53026

[Felli and Ould Ahmedou 2003] V. Felli and M. Ould Ahmedou, "Compactness results in conformal deformations of Riemannian metrics on manifolds with boundaries", *Math. Z.* **244**:1 (2003), 175–210. MR 2004d:53034 Zbl 1076.53047

[Felli and Ould Ahmedou 2005] V. Felli and M. Ould Ahmedou, "A geometric equation with critical nonlinearity on the boundary", *Pacific J. Math.* **218**:1 (2005), 75–99. MR 2007b:53079 Zbl 1137.53328

[Han and Li 1999] Z.-C. Han and Y. Li, "The Yamabe problem on manifolds with boundary: Existence and compactness results", *Duke Math. J.* **99**:3 (1999), 489–542. MR 2000j:53045 ZBL 0945.53023

[Han and Li 2000] Z.-C. Han and Y. Li, "The existence of conformal metrics with constant scalar curvature and constant boundary mean curvature", *Comm. Analysis Geom.* **8**:4 (2000), 809–869. MR 2001m:53062 Zbl 0990.53033

[Hebey and Vaugon 1993] E. Hebey and M. Vaugon, "Le problème de Yamabe équivariant", *Bull. Sci. Math.* **117**:2 (1993), 241–286. MR 94k:53056 Zbl 0786.53024

[Khuri et al. 2009] M. A. Khuri, F. C. Marques, and R. M. Schoen, "A compactness theorem for the Yamabe problem", *J. Differential Geom.* **81**:1 (2009), 143–196. MR 2010e:53065 Zbl 1162.53029

[Lee and Parker 1987] J. M. Lee and T. H. Parker, "The Yamabe problem", *Bull. Amer. Math. Soc. (N.S.)* **17**:1 (1987), 37–91. MR 88f:53001 Zbl 0633.53062

[Marques 2005] F. C. Marques, "Existence results for the Yamabe problem on manifolds with boundary", *Indiana Univ. Math. J.* **54**:6 (2005), 1599–1620. MR 2006j:53047 Zbl 1090.53043

[Marques 2007] F. C. Marques, "Conformal deformations to scalar-flat metrics with constant mean curvature on the boundary", *Comm. Analysis Geom.* **15**:2 (2007), 381–405. MR 2008i:53046 Zbl 1132.53021

[Schoen 1984] R. Schoen, "Conformal deformation of a Riemannian metric to constant scalar curvature", *J. Differential Geom.* **20**:2 (1984), 479–495. MR 86i:58137 Zbl 0576.53028

Received July 4, 2009.

SÉRGIO DE MOURA ALMARAZ
INSTITUTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL FLUMINENSE (UFF)
RUA MÁRIO SANTOS BRAGA S/N
24020-140 NITERÓI, RJ
BRAZIL

almaraz@impa.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

PARASURFACE GROUPS

KHALID BOU-RABEE

A residually nilpotent group is *k-parafree* if all of its lower central series quotients match those of a free group of rank k. Magnus proved that k-parafree groups of rank k are themselves free. We mimic this theory with surface groups playing the role of free groups. Our main result shows that the analog of Magnus' theorem is false in this setting.

Introduction

This article is motivated by three stories. The first story concerns a theorem of Wilhelm Magnus. Recall that the *lower central series* of a group G is defined to be

$$\gamma_1(G) := G \text{ and } \gamma_k(G) := [G, \gamma_{k-1}(G)] \quad \text{for } k \geq 2,$$

where $[A, B]$ denotes the group generated by commutators of elements of A with elements of B. The *rank* of G is the size of a minimal generating set of G. In [1939], Magnus gave a beautiful characterization of free groups in terms of their lower central series.

Theorem (Magnus' theorem on parafree groups). *Let F_k be a nonabelian free group of rank k and G a group of rank k. If $G/\gamma_i(G) \cong F_k/\gamma_i(F_k)$ for all i, then $G \cong F_k$.*

Following this result, Hanna Neumann inquired whether it was possible for two residually nilpotent groups G and G' to have $G/\gamma_i(G) \cong G'/\gamma_i(G')$ for all i without having $G \cong G'$; see [Liriano 2007]. Gilbert Baumslag [1967] gave a positive answer to this question by constructing what are now known as parafree groups that are not themselves free. A group G is *parafree* if

(1) G is residually nilpotent, and

(2) there exists a finitely generated free group F such that $G/\gamma_i(G) \cong F/\gamma_i(F)$ for all i.

MSC2000: 20F22.

Keywords: Magnus' theorem, Magnus' conjugacy theorem, parasurface, parafree, almost surface groups, automorphism-nilpotent separable.

By Magnus' theorem, Baumslag's examples necessarily have rank different from the corresponding free group. In this paper we give new examples addressing Neumann's question. Specifically, we construct residually nilpotent groups G that share the same lower central series quotients with a surface group but are not themselves surface groups. These examples are analogous to Baumslag's parafree groups, where the role of free groups is replaced by surface groups. Consequently, we call such groups *parasurface groups*. Since the parasurface examples constructed in this paper have the same rank as their corresponding surface groups, the analog of Magnus' theorem for parasurface groups is false.

Theorem 1. *Let Γ_g be the genus g surface group. There exists a rank $2g$ residually nilpotent group G such that $G/\gamma_i(G) \cong \Gamma_g/\gamma_i(\Gamma_g)$ for all i.*

We now turn towards the second story, which concerns residual properties in free groups. Our second story requires some notation. We say two elements $g, h \in G$ are *nilpotent-conjugacy equivalent* if the images of g and h in all nilpotent quotients of G are conjugate. We say a group G is *conjugacy-nilpotent separable* if any pair of nilpotent-conjugacy equivalent elements must be conjugate. Free groups are known to be conjugacy-nilpotent separable [Lyndon and Schupp 2001; Paris 2009]. A natural question to ask is whether the role of inner automorphisms may be played by automorphisms. In this vein, we say two elements $g, h \in G$ are *automorphism equivalent* if there exists some $\phi \in \operatorname{Aut} G$ with $\phi(g) = h$. Further, we say g is *nilpotent-automorphism equivalent* to h if the images of g and h in all nilpotent quotients of G are automorphism equivalent. A group G is *automorphism-nilpotent separable* if all pairs of nilpotent-automorphism equivalent elements must be automorphism equivalent. Free groups are conjugacy-nilpotent separable; however:

Theorem 2. *Nonabelian free groups of even rank are not automorphism-nilpotent separable.*

In the same flavor, Orin Chein [1969] showed that there exist automorphisms of nilpotent quotients of F_3 that do not lift to automorphisms of F_3.

Our third and final story concerns Magnus' conjugacy theorem for groups with one defining relator [Magnus et al. 1976, Theorem 4.11, page 261], which we state below. Let $\langle\langle w \rangle\rangle_H$ be the group generated by the H-conjugates of w for a subgroup H of G.

Theorem (Magnus' conjugacy theorem for groups with one defining relator). *Let s and t be elements in F_k such that $\langle\langle s \rangle\rangle_{F_k} = \langle\langle t \rangle\rangle_{F_k}$. Then s is conjugate to t^ϵ for $\epsilon = 1$ or -1.*

Our next result demonstrates that this theorem does *not* generalize to one-relator nilpotent groups. Let $F_{k,i} = F_k/\gamma_i(F_k)$ be the *free rank k, i-step nilpotent quotient.* Let $\phi_{k,i}$ be the projection $F_k \to F_{k,i}$.

Theorem 3. *Let $F_4 = \langle a, b, c, d\rangle$ be the free group of rank* 4. *Let $w = [a, b][c, d]$. Then*

$$\langle\langle \phi_{4,i}(w[w, bwb^{-1}])\rangle\rangle_{F_{4,i}} = \langle\langle \phi_{4,i}\,(w)\rangle\rangle_{F_{4,i}} \quad \textit{for all } i.$$

However, for large i, the element w is not conjugate to $w[w, bwb^{-1}]$ in $F_{4,i}$.

1. Almost surface groups

Let Γ_g be the fundamental group of a closed hyperbolic surface of genus g. A group G is a *weakly g-parasurface group* if $G/\gamma_k(G) \cong \Gamma_g/\gamma_k(\Gamma_g)$ for all $k \geq 1$. If G is weakly g-parasurface and residually nilpotent, we say that G is *g-parasurface*. Let $G = F/N$ be a weakly g-parasurface group where F is a free group of rank $2g$ on generators $a_1, a_2, \ldots, a_{2g-1}, a_{2g}$. Set $w = [a_1, a_2]\cdots[a_{2g-1}, a_{2g}]$. Recall that $\langle\langle w\rangle\rangle_F$ is the normal closure of w in F. Then we have the following trichotomy (see Theorem 4 below) for such groups G:

(Type I) There exists an isomorphism $\phi : F \to F$ such that $\phi(N) \geq \langle\langle w\rangle\rangle_F$.

(Type II) There exists an isomorphism $\phi : F \to F$ such that $\phi(N) < \langle\langle w\rangle\rangle_F$.

(Type III) G is not of Type I or II.

The following theorem demonstrates that only examples of Type I or III may be residually nilpotent.

Theorem 4. *Groups of Type I must be surface groups. Further, groups of Type II are never parasurface.*

Our next two theorems show that although surface groups are residually nilpotent, there exist examples of Type II and III. That is, weakly parasurface groups that are not parasurface groups exist. Further, parasurface groups that are not surface groups exist.

Theorem 5. *Let $k > 2$ be even, and let $F_k = \langle a_1, a_2, \ldots, a_k\rangle$. Suppose $w = [a_1, a_2]\cdots[a_{k-1}, a_k]$, and let γ be an element in F_k. If $w[w, \gamma w\gamma^{-1}]$ is cyclically reduced and of different word length than w in F_k, then the group*

$$G = \langle a_1, a_2, \ldots, a_k : w[w, \gamma w\gamma^{-1}]\rangle$$

is weakly $k/2$-parasurface of Type II.

In Theorem 5, one can take $\gamma = a_2$, for example.

Theorem 6. *Let $k > 2$ be even, let $F_k = \langle a_1, a_2, \ldots, a_k\rangle$, and let δ be in the commutator subgroup of $F(a_1, a_2)$. If $[a_1\delta, a_2]$ is cyclically reduced and is of different word length than $[a_1, a_2]$ in $F(a_1, a_2)$, then the group*

$$G = \langle a_1, a_2, \ldots, a_k : [a_1\delta, a_2][a_3, a_4]\cdots[a_{k-1}, a_k]\rangle$$

is $k/2$-parasurface of Type III.

In Theorem 6 one can take $\delta = [[a_1, a_2], a_1]$, for example.

2. Proofs of the main results

2.1. *Preliminaries.* We first list a couple of results needed in the proofs of our main theorems. The first is from Magnus, Korass and Solitar [1976, Lemma 5.9, page 350].

Lemma 7. *The Frattini subgroup of a nilpotent group contains the derived subgroup.*

For the following theorem, see [Azarov 1998, Theorem 1].

Theorem 8 (Azarov's theorem). *Let A and B be free groups, and let α and β be nonidentity elements of the groups A and B, respectively. Let $G = (A * B; \alpha = \beta)$. Let n be the largest positive integer such that $y^n = \beta$ has a solution in B. If $n = 1$, then G is a residually finite p-group.*

2.2. *The proofs.* Before proving Theorems 1, 2, and 3 from the introduction, we prove Theorems 4, 5, and 6.

Proof of Theorem 4. We first show that groups of Type I must be surface groups. For the sake of a contradiction, suppose that G is weakly g-parasurface of Type I and is not isomorphic to Γ_g. Let F and $K = \langle\langle w \rangle\rangle_F$ be as in the definition of Type I groups. By assumption, there exists an isomorphism $\phi : F \to F$ such that $\phi(N) \geq K$. The isomorphism ϕ^{-1} induces a homomorphism $\rho_i : \Gamma_g/\gamma_i(\Gamma_g) \to G/\gamma_i(G)$ that is surjective for all i. Since finitely generated nilpotent groups are Hopfian (see for example [de la Harpe 2000, Section III.A.19]), the maps ρ_i must be isomorphisms for all i. On the other hand, since G is not isomorphic to Γ_g, we must have some $\gamma \in \phi(N) - K$. Further, $F/K = \Gamma_g$ is residually nilpotent, so there exists some i such that $\gamma \neq 1$ in $\Gamma_g/\gamma_i(\Gamma_g)$. Since $\gamma \in \ker \rho_i$, we have a contradiction.

We now show that groups of Type II are never residually nilpotent. For the sake of a contradiction, suppose that G is a residually nilpotent group of Type II. Let F and $K = \langle\langle w \rangle\rangle_F$ be as in the definition of Type II groups. By assumption, the map $\phi : F \to F$ induces a map $\psi : G \to \Gamma_g$ that is onto with nontrivial kernel. Let $\gamma \in \ker \psi$. Since G is residually nilpotent, there exists i such that $g \notin \gamma_i(G)$. Hence, the induced map $\rho_i : G/\gamma_i(G) \to \Gamma_g/\gamma_i(\Gamma_g)$ is onto but not bijective, which is impossible since finitely generated nilpotent groups are Hopfian. □

Proof of Theorem 5. Let G, w, and γ be as in the statement of Theorem 5. We first show that G is weakly parasurface:

Claim 9. *We have $w = 1$ in all quotients $G/\gamma_i(G)$.*

Proof of claim. Let $H_1 = \langle\langle \phi_{k,i}(w[w, \gamma w \gamma^{-1}]) \rangle\rangle$ and $H_2 = \langle\langle \phi_{k,i}(w) \rangle\rangle$ in $F_k/\gamma_i(F_k)$. Clearly $H_1 \leq H_2$. Further, the image of H_1 in $H_2/[H_2, H_2]$ generates as

$$\phi_{k,i}(w[w, \gamma w \gamma^{-1}]) = \phi_{k,i}(w) \mod [H_2, H_2].$$

Hence, since H_2 is nilpotent, Lemma 7 implies that $H_1 = H_2$, and so the claim follows. □

If $w \neq 1$ in G, then the claim also shows that G is not residually nilpotent. Suppose, for the sake of a contradiction, that $w = 1$ in G. Then by Magnus' conjugacy theorem for groups with one defining relator, w^ϵ and $w[w, \gamma w \gamma^{-1}]$ are conjugate for $\epsilon = 1$ or -1, but this is impossible since they are both cyclically reduced words of different word lengths [Magnus et al. 1976, Theorem 1.3, page 36]. Hence G is not residually nilpotent and G is not parasurface. Further, as $w = 1$ in all nilpotent quotients, G is weakly parasurface. The proof of Theorem 5 is complete. □

Proof of Theorem 6. Let G and δ be as in the statement of Theorem 6. That G is residually nilpotent follows from Theorem 8 applied to

$$A = F_{k-2} = \langle a_1, a_2, \ldots, a_{k-2} \rangle \quad \text{and} \quad B = F_2 = \langle a_{k-1}, a_k \rangle,$$

with $\alpha = [a_1\delta, a_2][a_3, a_4] \cdots [a_{k-3}, a_{k-2}]$ and $\beta = ([a_{k-1}, a_k])^{-1}$ and the following claim:

Claim 10. *If $\beta = y^n$ in B, where $y \in B$, then $n = 1$.*

Proof of claim. If $\beta = y^n$ for some $n > 1$, then since $B/\langle\langle \beta \rangle\rangle$ is torsion-free, $y \in \langle\langle \beta \rangle\rangle$. Hence, $\langle\langle y \rangle\rangle = \langle\langle \beta \rangle\rangle$, and so by Magnus' conjugacy theorem for groups with one defining relator, y is conjugate to β or β^{-1}. But then β would have the same cyclically reduced form word length as β^n, contradicting [Magnus et al. 1976, Theorem 1.3, page 36]. □

We now show that G is not a surface group. Let $H = \langle a_1\delta, a_2, \ldots, a_{k-1}, a_k \rangle \leq G$. The next two claims imply that H is a nonfree proper subgroup of G of rank k. However, if G were a surface group, it would have to be the surface group $\Gamma_{k/2}$, of genus $k/2$. Any rank k proper subgroup of $\Gamma_{k/2}$ must be free, so H must be free, a contradiction.

Claim 11. *H is not a free group and is of rank k.*

Proof of claim. H is rank k, since $\{a_1\delta, a_2, \ldots, a_k\}$ generate $G/[G, G]$ and since $G/[G, G] = \mathbb{Z}^k$. If H were free, it would be free of rank k. Let $x_1, x_2, \ldots, x_k$ be a free basis for H. The map $H \to H$ given by $x_1 \mapsto a_1\delta$ and $x_k \mapsto a_k$ for $k > 1$ is an isomorphism because H is Hopfian (being a free group). So H is freely generated by $\{a_1\delta, a_2, \ldots, a_k\}$, but this is impossible since $[a_1\delta, a_2] \cdots [a_{k-1}, a_k] = 1$. □

Claim 12. *H is a proper subgroup of G.*

Proof of claim. Indeed, the element a_1 cannot be in H. For suppose $a_1 \in H$, and let N be the normal subgroup generated by a_k for $k > 2$. Then we have $G/N = \langle a_1, a_2 : [a_1\delta, a_2]\rangle$. Since $a_1 \in H$, G/N is abelian, and so has presentation $G/N = \langle a_1, a_2 : [a_1, a_2]\rangle$. But then the normal subgroup generated by $[a_1, a_2]$ and the normal subgroup generated by $[a_1\delta, a_2]$ are equal in $F_2 = \langle a_1, a_2\rangle$. By Magnus' conjugacy theorem for groups with one defining relator, $[a_1, a_2]^\epsilon$ must be conjugate to $[a_1\delta, a_2]$ for $\epsilon = 1$ or -1 in $F_2 = \langle a_1, a_2\rangle$, which is impossible since $[a_1, a_2]$ and $[a_1\delta, a_2]$ are cyclically reduced and have different word lengths [Magnus et al. 1976, Theorem 1.3, page 36]. Hence $a_1 \notin H$. □

We finish the proof of Theorem 6 by showing that all of the lower central series quotients of G match those of a surface group of genus $k/2$.

Claim 13. *G is weakly $k/2$-parasurface.*

Proof of claim. Let ψ be the map defined by $a_1 \mapsto a_1\delta$ and $a_k \mapsto a_k$ for $k > 1$. This gives a well-defined map $F_{k,i} \to F_{k,i}$, where $F_{k,i} := F_k/\gamma_i(F_k)$. This is an epimorphism by Lemma 7. Since finitely generated nilpotent groups are Hopfian, $\psi : F_{k,i} \to F_{k,i}$ must be an isomorphism. Therefore, the induced map on $\Gamma_{k/2}/\gamma_i(\Gamma_{k/2}) \to G/\gamma_i(G)$ is an isomorphism, as claimed. □

The proof of Theorem 6 is now complete. □

We are now ready to quickly prove all of our theorems stated in the introduction.

Proof of Theorem 1. Theorem 6 gives the desired parasurface groups. □

Proof of Theorem 2. The proof of Claim 13 with $\delta = [[a_1, a_2], a_1]$ shows that

$$[a_1, a_2]\cdots[a_{k-1}, a_k] \quad \text{and} \quad [a_1\delta, a_2]\cdots[a_{k-1}, a_k]$$

are nilpotent-automorphism equivalent. However, $[a_1, a_2]\cdots[a_{k-1}, a_k]$ is not automorphism equivalent to $[a_1\delta, a_2]\cdots[a_{k-1}, a_k]$ in F_k by Theorem 6. □

Proof of Theorem 3. The element w is not conjugate to $w[w, bwb^{-1}]$ in F_4. Hence, since F_4 is conjugacy-nilpotent separable, there exists some large $N > 0$ such that $\phi_{4,i}(w)$ is not conjugate to $\phi_{4,i}(w[w, bwb^{-1}])$ in $F_{4,i}$ for all $i > N$. Moreover, the equality

$$\langle\langle \phi_{4,i}(w[w, bwb^{-1}])\rangle\rangle_{F_{4,i}} = \langle\langle \phi_{4,i}(w)\rangle\rangle_{F_{4,i}} \quad \text{for all } i$$

is an immediate consequence of Claim 9. □

3. Final remarks

We have shown that there exist groups that are almost surface groups in the sense that they share all their lower central series quotients with a surface group but are not themselves surface groups. In light of our examples, we pose the following question.

Question 14. What properties do parasurface groups share with surface groups?

As a small step in answering this question, we present this:

Theorem 15. *Any finite-index subgroup of a parasurface group is not free.*

Proof. Let G be a parasurface group. Note that G is not free, for if it were, Magnus' theorem would imply that the fundamental group of some compact surface is free. Further, G is torsion-free, for otherwise by residual nilpotence, there would exist torsion elements in $\Gamma_g/\gamma_k(\Gamma_g)$ for some g and k, but this is impossible by [Labute 1970].

Let $\mathrm{cd}(G)$ denote the cohomological dimension of G. If $\Gamma \leq G$ is a free group of finite index, then by [Brown 1994, Theorem 3.1, page 190] and the fact that G is torsion-free, $\mathrm{cd}\, G = \mathrm{cd}\, \Gamma$. Hence, $\mathrm{cd}\, G = 1$, but then by [Stallings 1968] and [Swan 1969], G must itself be free, a contradiction. □

Acknowledgements. I am extremely grateful to my advisor, Benson Farb, for his teaching and inspiration, and to my coadvisor, Ben McReynolds, for all his time and help. Further, I thank Justin Malestein for some useful discussions and careful reading of this paper. Finally, I would like to thank Alex Wright, Tom Church, Blair Davey, and Thomas Zamojski for reading early drafts of this paper.

References

[Azarov 1998] D. N. Azarov, "On the residual nilpotence of free products of free groups with cyclic amalgamation", *Mat. Zametki* **64**:1 (1998), 3–8. In Russian; translated in *Math. Notes* **64**:1 (1998), 3–7. MR 2000g:20038 Zbl 0922.20030

[Baumslag 1967] G. Baumslag, "Groups with the same lower central sequence as a relatively free group, I: The groups", *Trans. Amer. Math. Soc.* **129** (1967), 308–321. MR 36 #248 Zbl 0153.35002

[Brown 1994] K. S. Brown, *Cohomology of groups*, Graduate Texts in Mathematics **87**, Springer, New York, 1994. MR 96a:20072 Zbl 0823.20049

[Chein 1969] O. Chein, "Subgroups of IA automorphisms of a free group", *Acta Math.* **123** (1969), 1–12. MR 41 #307 Zbl 0184.04101

[de la Harpe 2000] P. de la Harpe, *Topics in geometric group theory*, University of Chicago Press, Chicago, IL, 2000. MR 2001i:20081 Zbl 0965.20025

[Labute 1970] J. P. Labute, "On the descending central series of groups with a single defining relation", *J. Algebra* **14** (1970), 16–23. MR 40 #4342 Zbl 0198.34601

[Liriano 2007] S. Liriano, "Algebraic geometric invariants of parafree groups", *Internat. J. Algebra Comput.* **17**:1 (2007), 155–169. MR 2008a:20056 Zbl 1131.20022

[Lyndon and Schupp 2001] R. C. Lyndon and P. E. Schupp, *Combinatorial group theory*, Springer, Berlin, 2001. MR 2001i:20064 Zbl 0997.20037

[Magnus 1939] W. Magnus, "Über freie Faktorgruppen und freie Untergruppen gegebener Gruppen", *Monatsh. Math. Phys.* **47**:1 (1939), 307–313. MR 1550819 Zbl 0021.30004

[Magnus et al. 1976] W. Magnus, A. Karrass, and D. Solitar, *Combinatorial group theory: Presentations of groups in terms of generators and relations*, 2nd ed., Dover, New York, 1976. MR 54 #10423 Zbl 0362.20023

[Paris 2009] L. Paris, “Residual p properties of mapping class groups and surface groups”, *Trans. Amer. Math. Soc.* **361**:5 (2009), 2487–2507. MR 2010f:20037 Zbl 05551113

[Stallings 1968] J. R. Stallings, “On torsion-free groups with infinitely many ends”, *Ann. of Math.* (2) **88** (1968), 312–334. MR 37 #4153 Zbl 0238.20036

[Swan 1969] R. G. Swan, “Groups of cohomological dimension one”, *J. Algebra* **12** (1969), 585–610. MR 39 #1531 Zbl 0188.07001

Received August 12, 2009. Revised December 6, 2009.

KHALID BOU-RABEE
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CHICAGO
CHICAGO, IL 60637
UNITED STATES

khalid@math.uchicago.edu
http://www.math.uchicago.edu/∼khalid

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

EXPRESSIONS FOR CATALAN KRONECKER PRODUCTS

ANDREW A. H. BROWN,
STEPHANIE VAN WILLIGENBURG AND MIKE ZABROCKI

We give some elementary manifestly positive formulas for the Kronecker products $s_{(d,d)} * s_{(d+k,d-k)}$. These formulas demonstrate some fundamental properties of the Kronecker coefficients, and we use them to deduce a number of enumerative and combinatorial results.

1. Introduction

A classic open problem in algebraic combinatorics is to explain in a manifestly positive combinatorial formula the Kronecker product (or internal product) of two Schur functions. This product is the Frobenius image of the internal tensor product of two irreducible symmetric group modules, or it is alternatively the characters of the induced tensor product of general linear group modules. Although for representation theoretic reasons this expression clearly has nonnegative coefficients when expanded in terms of Schur functions, it remains an open problem to provide a satisfying positive combinatorial or algebraic formula for the Kronecker product of two Schur functions.

Many attempts have been made to capture some aspect of these coefficients, for example, special cases [Bessenrodt and Behns 2004; Bessenrodt and Kleshchev 1999; Remmel and Whitehead 1994; Rosas 2001], asymptotics [Ballantine and Orellana 2005; 2005], stability [Vallejo 1999], the complexity of calculating them [Bürgisser and Ikenmeyer 2008], and conditions under which they are nonzero [Dvir 1993]. Given that the Littlewood–Richardson rule and many successors have so compactly and cleanly been able to describe the external product of two Schur functions, it seems as though some new ideas for capturing the combinatorics of Kronecker coefficients are needed.

The results in this paper were inspired by the symmetric function identity of [Garsia et al. 2009, Theorem I.1] for the Kronecker product $s_{(d,d)} * s_{(d,d)}$ of two

MSC2000: primary 05E05; secondary 05E10, 20C30, 20C33.
Keywords: Catalan number, Kronecker coefficient, Kronecker product, Schur function, symmetric function.
The authors were supported in part by the National Sciences and Engineering Research Council of Canada.

Schur functions. More precisely, for a subset of partitions X of $2d$, if we set $[X] = \sum_{\lambda \in X} s_\lambda$, called a *rug*, then

$$s_{(d,d)} * s_{(d,d)} = [\text{4 parts all even or all odd}]. \tag{1-1}$$

This identity differs significantly from most published results on the Kronecker product: instead of giving a combinatorial interpretation or algorithm, it clearly states exactly which partitions have nonzero coefficients and that all of the coefficients are 0 or 1.

This computation arose in the solution to a mathematical physics problem related to resolving the interference of 4 qubits [Wallach 2005] because the sum of these coefficients is equal to the dimensions of polynomial invariants of four copies of $SL(2, \mathbb{C})$ acting on $\mathbb{C}^8$. Understanding the Kronecker product of $s_{(d,d)}$ with s_λ for partitions λ with 4 parts that are all even or all odd would be useful for calculating the dimensions of invariants of six copies of $SL(2, \mathbb{C})$ acting on $\mathbb{C}^{12}$, which is a measure of entanglement of 6 qubits. Ultimately we would like to be able to compute

$$CT_{a_1,a_2,\dots,a_k}\left(\frac{\prod_{i=1}^{k}(1-a_i^2)}{\prod_{S\subseteq\{1,2,\dots k\}}(1-q\prod_{i\in S}a_i/\prod_{j\notin S}a_j)}\right) = \sum_{d\geq 0}\langle s_{(d,d)}^{*k}, s_{(2d)}\rangle q^{2d}$$

(see [Garsia et al. 2010, formulas I.4 and I.5] and [Luque and Thibon 2006] for a discussion), where CT represents the operation of taking the constant term and equations of this type are a motivation for understanding the Kronecker product with $s_{(d,d)}$ as completely as possible.

Using $s_{(d,d)} * s_{(d,d)}$ as our inspiration, we show in Corollary 3.6 that

$$s_{(d,d)} * s_{(d+1,d-1)} = [\text{2 even parts and 2 odd parts}]$$

and with a similar computation we also derive that

$$s_{(d,d)} * s_{(d+2,d-2)} = [\text{4 parts, all even or all odd, but not 3 the same}] + [\text{4 distinct parts}].$$

Interestingly, this formula says that all of the coefficients in the Schur expansion of $s_{(d,d)} * s_{(d+2,d-2)}$ are either 0, 1 or 2, and that the coefficient is 2 for those Schur functions indexed by partitions with 4 distinct parts that are all even or all odd.

These and further examples suggest that the Schur function expansion of $s_{(d,d)} * s_{(d+k,d-k)}$ has the pattern of a boolean lattice of subsets, in that it can be written as the sum of $\lfloor k/2 \rfloor + 1$ intersecting sums of Schur functions each with coefficient 1.

The main result of this article is Theorem 3.1, which states that

$$s_{(d,d)} * s_{(d+k,d-k)} = \sum_{i=0}^{k} [(k+i, k, i)P] + \sum_{i=1}^{k} [(k+i+1, k+1, i)P], \tag{1-2}$$

where we have used the notation γP to represent the set of partitions λ of $2d$ of length less than or equal to 4 such that $\lambda - \gamma$ (representing a vector difference) is a partition with 4 even parts or 4 odd parts. The disjoint sets of this sum can be grouped so that the sum is of only $\lfloor k/2 \rfloor + 1$ terms, which shows that the coefficients always lie in the range 0 through $\lfloor k/2 \rfloor + 1$. The most interesting aspect of this formula is that we see the lattice of subsets arising in a natural and unexpected way in a representation-theoretic setting. This is potentially part of a more general result and the hope is that this particular model will shed light on a general formula for the Kronecker product of two Schur functions, but our main motivation for computing these is to develop computational tools.

There are yet further motivations for restricting our attention to the Kronecker product of $s_{(d,d)}$ with another Schur function. The Schur functions indexed by the partition (d, d) are a special family for several combinatorial reasons, and so there is reason to believe that their behavior will be more accessible than the general case of the Kronecker product of two Schur functions. More precisely, Schur functions indexed by partitions with two parts are notable because they are the difference of two homogeneous symmetric functions, for which a combinatorial formula for the Kronecker product is known. In addition, a partition (d, d) is rectangular and hence falls under a second category of Schur functions that are often combinatorially more straightforward to manipulate than the general case.

From the hook length formula it follows that the number of standard tableau of shape (d, d) is equal to the Catalan number

$$C_d = \frac{1}{d+1}\binom{2d}{d}.$$

Therefore, from the perspective of S_{2d} representations, we may, by taking the Kronecker product with the Schur function $s_{(d,d)}$ and the Frobenius image of a module, explain how the tensor of a representation with a particular irreducible module of dimension C_d decomposes.

The paper is structured as follows. In Section 2, we review pertinent background information including necessary symmetric function notation and lemmas needed for later computation. In Section 3, we consider a generalization of formula (1-1) to an expression for $s_{(d,d)} * s_{(d+k,d-k)}$. Finally, Section 4 is devoted to combinatorial and symmetric function consequences of our results. In particular, we are able to give generating functions for the partitions that have a particular coefficient in the expression $s_{(d,d)} * s_{(d+k,d-k)}$.

2. Background

Partitions. A *partition* λ of an integer n, denoted $\lambda \vdash n$, is a finite sequence of nonnegative integers $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\ell)$ whose values sum to n. The *height* or *length* of the partition, denoted $\ell(\lambda)$, is the maximum index for which $\lambda_{\ell(\lambda)} > 0$. We call the λ_i *parts* or *rows* of the partition, and if λ_i appears n_i times we abbreviate this subsequence to $\lambda_i^{n_i}$. With this in mind if $\lambda = (k^{n_k}, (k-1)^{n_{k-1}}, \ldots, 1^{n_1})$, then we define $z_\lambda = 1^{n_1} 1!\, 2^{n_2} 2! \cdots k^{n_k} k!$. The 0 parts of the partition are optional and we will assume that $\lambda_i = 0$ for $i > \ell(\lambda)$.

Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ be a partition of n. To form the *diagram* associated with λ, place a cell at each point (j, i) in matrix notation, where $1 \leq i \leq \lambda_j$ and $1 \leq j \leq \ell$. We say λ has *transpose* λ' if the diagram for λ' is given by the points (i, j) for which $1 \leq i \leq \lambda_j$ and $1 \leq j \leq \ell$.

Symmetric functions and the Kronecker product. The ring of symmetric functions is the graded subring of $\mathbb{Q}[x_1, x_2, \ldots]$ given by $\Lambda := \mathbb{Q}[p_1, p_2, \ldots]$, where $p_i = x_1^i + x_2^i + \cdots$ are the elementary power sum symmetric functions. For $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_\ell)$, we define $p_\lambda := p_{\lambda_1} p_{\lambda_2} \cdots p_{\lambda_\ell}$. The interested reader should consult a reference such as [Macdonald 1995] for more details of the structure of this ring. It is straightforward to see that $\{p_\lambda\}_{\lambda \vdash n \geq 0}$ forms a basis for Λ. This basis is orthogonal to itself with respect to the scalar product on Λ:

$$\langle p_\lambda, p_\mu \rangle = z_\lambda \delta_{\lambda\mu}.$$

However, our focus for this paper will be the basis $\{s_\lambda\}_{\lambda \vdash n \geq 0}$ of Λ known as the basis of *Schur functions*, which is the orthonormal basis under the scalar product:

$$\langle s_\lambda, s_\mu \rangle = \delta_{\lambda\mu}.$$

The *Kronecker product* is the operation

$$\frac{p_\lambda}{z_\lambda} * \frac{p_\mu}{z_\mu} = \delta_{\lambda\mu} \frac{p_\lambda}{z_\lambda} \tag{2-1}$$

on symmetric functions that in terms of the Schur functions becomes

$$s_\mu * s_\nu = \sum_{\lambda \vdash |\mu|} C_{\mu\nu\lambda} s_\lambda.$$

The *Kronecker coefficients* $C_{\mu\nu\lambda}$ encode the inner tensor product of symmetric group representations. That is, if we denote the irreducible S_n module indexed by a partition λ by M^λ and if $M^\mu \otimes M^\nu$ represents the tensor of two modules with the diagonal action, then the module decomposes as

$$M^\mu \otimes M^\nu \simeq \bigoplus_\lambda (M^\lambda)^{\oplus C_{\mu\nu\lambda}}.$$

The Kronecker coefficients also encode the decomposition of GL_{nm} polynomial representations to $\mathrm{GL}_n \otimes \mathrm{GL}_m$ representations

$$\mathrm{Res}^{\mathrm{GL}_{mn}}_{\mathrm{GL}_n \otimes \mathrm{GL}_m}(V^\lambda) \simeq \bigoplus_{\mu,\nu} (V^\mu \otimes V^\nu)^{\oplus C_{\mu\nu\lambda}}.$$

It easily follows from (2-1) and the linearity of the product that these coefficients satisfy the symmetries

$$C_{\mu\nu\lambda} = C_{\nu\mu\lambda} = C_{\mu\lambda\nu} = C_{\mu'\nu'\lambda} \quad \text{and} \quad C_{\lambda\mu(n)} = C_{\lambda\mu'(1^n)} = \begin{cases} 1 & \text{if } \lambda = \mu, \\ 0 & \text{otherwise,} \end{cases}$$

which we will use extensively in what follows.

We will need some symmetric function identities. Recall that for a symmetric function f, the symmetric function operator $f^\perp$ (read "eff perp") is defined to be the operator dual to multiplication with respect to the scalar product. That is,

$$\langle f^\perp g, h\rangle = \langle g, f \cdot h\rangle. \tag{2-2}$$

The perp operator can also be defined linearly by $s_\lambda^\perp s_\mu = s_{\mu/\lambda} = \sum_{\nu \vdash |\mu|-|\lambda|} c^\mu_{\lambda\nu} s_\nu$, where the $c^\mu_{\lambda\nu}$ are the Littlewood–Richardson coefficients. We will make use of the following well-known relation, which connects the internal and external products:

$$\langle s_\lambda f, g * h\rangle = \sum_{\mu,\nu \vdash |\lambda|} C_{\lambda\mu\nu} \langle f, (s_\mu^\perp g) * (s_\nu^\perp h)\rangle. \tag{2-3}$$

This formula follows because of the relationship between the internal coproduct and the scalar product. It may also be seen to hold on the power sum basis since

$$\left\langle p_\lambda p_\mu, \frac{p_\gamma}{z_\gamma} * \frac{p_\nu}{z_\nu}\right\rangle = \left\langle p_\mu, \left(p_\lambda^\perp \frac{p_\gamma}{z_\gamma}\right) * \left(p_\lambda^\perp \frac{p_\nu}{z_\nu}\right)\right\rangle, \tag{2-4}$$

which holds because both the sides are 1 if and only if $\gamma = \nu$ and are both equal to the union of the parts of λ and μ. This given, (2-3) follows by linearity.

From these two identities and the Littlewood–Richardson rule, we derive this:

Lemma 2.1. *If $\ell(\lambda) > 4$, then $C_{(d,d)(a,b)\lambda} = 0$. Otherwise it satisfies the following recurrences. If $\ell(\lambda) = 4$, then*

$$\langle s_{(d,d)} * s_{(d+k,d-k)}, s_\lambda\rangle = \langle s_{(d-2,d-2)} * s_{(d+k-2,d-k-2)}, s_{\lambda-(1^4)}\rangle. \tag{2-5}$$

If $\ell(\lambda) = 3$, then

$$\begin{aligned} \langle s_{(d,d)} * s_{(d+k,d-k)}, s_\lambda\rangle = {} & \langle s_{(d-1,d-1)} * s_{(d+k-1,d-k-1)}, s_{(1)} s_{\lambda-(1^3)}\rangle \\ & - \langle s_{(d-2,d-2)} * s_{(d+k-2,d-k-2)}, s_{(1)}^\perp s_{\lambda-(1^3)}\rangle. \end{aligned} \tag{2-6}$$

If $\ell(\lambda)=2$, *then*

$$\langle s_{(d,d)} * s_{(d+k,d-k)}, s_\lambda\rangle = \begin{cases} 1 & \text{if } k \equiv \lambda_2 \ (\mathrm{mod}\ 2) \text{ and } \lambda_2 \geq k, \\ 0 & \text{otherwise.} \end{cases} \tag{2-7}$$

Proof. For (2-7), see [Remmel and Whitehead 1994, Theorem 3.3] and [Rosas 2001, Corollary 1]. The result also appears in [Garsia et al. 2010, Theorem 2.2].

Regev [1980] proves that the maximum height of an indexing partition of the terms of $s_{(d,d)} * s_{(d+k,d-k)}$ will be at most 4; hence we can conclude that if $\ell(\lambda)>4$, then $C_{(d,d)(a,b)\lambda}=0$.

Assume that $\ell(\lambda)=4$. By the Pieri rule, $s_{(1^4)}s_{\lambda-(1^4)}$ is equal to s_λ plus terms of the form s_γ, where $\ell(\gamma)>4$. As a consequence,

$$\begin{aligned}\langle s_{(d,d)} * s_{(d+k,d-k)}, s_\lambda\rangle &= \langle s_{(d,d)} * s_{(d+k,d-k)}, s_{(1^4)}s_{\lambda-(1^4)}\rangle \\ &= \sum_{\mu\vdash 4} \langle (s_\mu^\perp s_{(d,d)}) * (s_{\mu'}^\perp s_{(d+k,d-k)}), s_{\lambda-(1^4)}\rangle.\end{aligned}$$

Every term in this sum is 0 unless both μ and μ' have length no more than 2. The only term for which this is true is $\mu=(2,2)$, and $s_{(2,2)}^\perp(s_{(a,b)}) = s_{(a-2,b-2)}$; hence (2-5) holds.

Assume that $\ell(\lambda)=3$. Although there are cases to check, it follows again from the Pieri rule that $s_{(1^3)}s_{\lambda-(1^3)} - s_{(1^4)}s_{(1)}^\perp s_{\lambda-(1^3)}$ is equal to s_λ plus terms involving s_γ, where $\ell(\gamma)>4$. Therefore,

$$\begin{aligned}\langle s_{(d,d)} * s_{(d+k,d-k)}, s_\lambda\rangle &= \langle s_{(d,d)} * s_{(d+k,d-k)}, s_{(1^3)}s_{\lambda-(1^3)} - s_{(1^4)}s_{(1)}^\perp s_{\lambda-(1^3)}\rangle \\ &= \sum_{\mu\vdash 3} \langle (s_\mu^\perp s_{(d,d)}) * (s_{\mu'}^\perp s_{(d+k,d-k)}), s_{\lambda-(1^3)}\rangle \\ &\qquad - \langle s_{(d-2,d-2)} * s_{(d+k-2,d-k-2)}, s_{(1)}^\perp s_{\lambda-(1^3)}\rangle.\end{aligned}$$

Again, in the sum the only terms that are not equal to 0 are those such that the length of both μ and μ' less than or equal to 2, and in this case the only such partition is $\mu=(2,1)$. By the Littlewood–Richardson rule, $s_{(2,1)}^\perp(s_{(a,b)}) = s_{(1)}^\perp(s_{(a-1,b-1)})$; hence this last expression is equal to

$$\begin{aligned}&\langle (s_{(1)}^\perp s_{(d-1,d-1)}) * (s_{(1)}^\perp s_{(d+k-1,d-k-1)}), s_{\lambda-(1^3)}\rangle \\ &\qquad - \langle s_{(d-2,d-2)} * s_{(d+k-2,d-k-2)}, s_{(1)}^\perp s_{\lambda-(1^3)}\rangle \\ &\qquad\qquad = \langle s_{(d-1,d-1)} * s_{(d+k-1,d-k-1)}), s_{(1)}s_{\lambda-(1^3)}\rangle \\ &\qquad\qquad\qquad - \langle s_{(d-2,d-2)} * s_{(d+k-2,d-k-2)}, s_{(1)}^\perp s_{\lambda-(1^3)}\rangle. \quad \square\end{aligned}$$

To express our main results we will use the characteristic of a boolean-valued proposition. If R is a proposition, then we denote the propositional characteristic (or indicator) function of R by

$$(\!(R)\!) = \begin{cases} 1 & \text{if proposition } R \text{ is true,} \\ 0 & \text{otherwise.} \end{cases}$$

3. The Kronecker product $s_{(d,d)} * s_{(d+k,d-k)}$

Let P indicate the set of partitions with four even parts or four odd parts, and let $\overline{P}$ be the set of partitions with 2 even parts and 2 odd parts (that is, the complement of P in the set of partitions of even size with at most 4 parts).

We let γP represent the set of partitions λ of $2d$ (the value of d will be implicit in the left hand side of the expression) such that $\lambda - \gamma \in P$. We also let $(\gamma \uplus \alpha) P = \gamma P \cup \alpha P$. In the cases we consider, the partitions in γP and αP are disjoint.

Theorem 3.1. *For λ a partition of $2d$,*

$$\langle s_{(d,d)} * s_{(d+k,d-k)}, s_\lambda \rangle = \sum_{i=0}^{k} (\!(\lambda \in (k+i, k, i) P)\!) + \sum_{i=1}^{k} (\!(\lambda \in (k+i+1, k+1, i) P)\!). \tag{3-1}$$

In the notation we have introduced, Theorem 3.1 can easily be restated:

Corollary 3.2. *For $k \geq 0$, if k is odd, then*

$$\begin{aligned} s_{(d,d)} * s_{(d+k,d-k)} &= [((k, k) \uplus (k+1, k, 1) \uplus (k+2, k+1, 1)) P] \\ &\quad + \sum_{i=1}^{(k-1)/2} [((k+2i, k, 2i) \uplus (k+2i+1, k+1, 2i) \\ &\qquad \uplus (k+2i+1, k, 2i+1) \uplus (k+2i+2, k+1, 2i+1)) P], \end{aligned}$$

and if k is even, then

$$\begin{aligned} s_{(d,d)} * s_{(d+k,d-k)} = [(k, k) P] + \sum_{i=1}^{k/2} [((k+2i-1, k, 2i-1) \uplus (k+2i, k+1, 2i-1) \\ \uplus (k+2i, k, 2i) \uplus (k+2i+1, k+1, 2i)) P]. \end{aligned}$$

As a consequence, the coefficients of $s_{(d,d)} * s_{(d+k,d-k)}$ will all be less than or equal to $\lfloor k/2 \rfloor + 1$.

Remark 3.3. The upper bound on the coefficients that appear in the expressions $s_{(d,d)} * s_{(d+k,d-k)}$ are sharp in that for sufficiently large d, there is a coefficient that will be equal to $\lfloor k/2 \rfloor + 1$.

Remark 3.4. This regrouping of the rugs is not unique but is useful because there are partitions that will fall in the intersection of each of these sets. This set of rugs is also not unique in that it is possible to describe other collections of sets of partitions (for example, see (3-4)).

Remark 3.5. This case has been considered in more generality by Remmel and Whitehead [1994] and Rosas [Rosas 2001] in their study of the Kronecker product of two Schur functions indexed by two-row shapes. In relation to these results, our first (lengthy) derivation of a similar result started with the formula of Rosas, but was later replaced by the current simpler identity. Meanwhile, R. Orellana, in a personal communication, has informed us that the journal version of the earlier paper has an error in it. Our computer implementation of [Remmel and Whitehead 1994, Theorem 2.1] does not agree, for example, with direct computation for $(h, k) = (4, 1)$ and $(l, m) = (3, 2)$ and $\nu = (3, 1, 1)$. Consequently, we wanted an independent proof of Theorem 3.1, and as a result our derivation is an elementary proof that uses induction involving symmetric function identities.

We note that for $k = 2$, the expression stated in the introduction does not exactly follow this decomposition, but it does follow from some manipulation.

Corollary 3.6. *For $d \geq 1$,*

$$s_{(d,d)} * s_{(d,d)} = [P], \tag{3-2}$$

$$s_{(d,d)} * s_{(d+1,d-1)} = [\overline{P}] \tag{3-3}$$

and for $d \geq 2$,

$$s_{(d,d)} * s_{(d+2,d-2)} = [P \cap \textit{at most two equal parts}] + [\textit{distinct partitions}]. \tag{3-4}$$

Proof. Note that (3-2) is just a restatement of Corollary 3.2 in the case that $k = 0$ and is [Garsia et al. 2009, Theorem I.1].

First, by Theorem 3.1 we note that

$$C_{(d,d)(d+1,d-1)\lambda} = (\!(\lambda \in (1, 1)P)\!) + (\!(\lambda \in (2, 1, 1)P)\!) + (\!(\lambda \in (3, 2, 1)P)\!).$$

If $\lambda \in P$, then $\lambda - (1, 1)$, $\lambda - (2, 1, 1)$, $\lambda - (3, 2, 1)$ are all not in P, so each of the terms in that expression are 0. If λ is a partition with two even parts and two odd parts (that is, $\lambda \in \overline{P}$), then either $\lambda_1 \equiv \lambda_2$ and $\lambda_3 \equiv \lambda_4 \pmod 2$ or $\lambda_1 \equiv \lambda_3$ and $\lambda_2 \equiv \lambda_4 \pmod 2$ or $\lambda_2 \equiv \lambda_3$ and $\lambda_1 \equiv \lambda_4 \pmod 2$. In each of these three cases, exactly one of the expressions $(\!(\lambda \in (1, 1)P)\!)$, $(\!(\lambda \in (2, 1, 1)P)\!)$ or $(\!(\lambda \in (3, 2, 1)P)\!)$ will be 1 and the other two will be zero. Therefore,

$$\sum_{\lambda \vdash 2d} C_{(d,d)(d+1,d-1)\lambda} s_\lambda = [\overline{P}].$$

We also have by Theorem 3.1

$$\begin{aligned} C_{(d,d)(d+2,d-2)\lambda} = (\!(\lambda \in (2,2)P)\!) + (\!(\lambda \in (4,2,2)P)\!) - (\!(\lambda \in (6,4,2)P)\!) \\ + (\!(\lambda \in (3,2,1)P)\!) + (\!(\lambda \in (4,3,1)P)\!) \\ + (\!(\lambda \in (5,3,2)P)\!) + (\!(\lambda \in (6,4,2)P)\!). \end{aligned}$$

Any distinct partition in P is also in $(6, 4, 2)P$. Every distinct partition in $\overline{P}$ will have two odd parts and two even parts and will be in one of $(3, 2, 1)P$, $(4, 3, 1)P$ or $(5, 3, 2)P$, depending on which of λ_2, λ_1 or λ_3 is equal to λ_4 (mod 2), respectively. Therefore, we have

$$\text{(3-5)} \qquad [\text{distinct partitions}] = [((3,2,1) \uplus (4,3,1) \uplus (5,3,2) \uplus (6,4,2))P].$$

If $\lambda \in (2, 2)P \cap (4, 2, 2)P$, then $\lambda_2 \geq \lambda_3 + 2$ because $\lambda \in (2, 2)P$, and $\lambda_1 \geq \lambda_2 + 2$ and $\lambda_3 \geq \lambda_4 + 2$ because $\lambda \in (4, 2, 2)P$, so $\lambda \in (6, 4, 2)P$. Conversely, one verifies that in fact $(2, 2)P \cap (4, 2, 2)P = (6, 4, 2)P$; hence

$$[(2,2)P \cup (4,2,2)P] = [(2,2)P] + [(4,2,2)P] - [(6,4,2)P].$$

If $\lambda \in P$ does not have three equal parts, then either $\lambda_2 > \lambda_3$, or $\lambda_1 > \lambda_2$ and $\lambda_3 > \lambda_4$. Therefore, $\lambda \in (2, 2)P \cup (4, 2, 2)P$ and hence $(2, 2)P \cup (4, 2, 2)P = P \cap (\text{at most two equal parts})$. □

Proof of Theorem 3.1. Our proof proceeds by induction on the value of d and uses the Lemma 2.1. We will consider two base cases because (2-6) and (2-5) give recurrences for two smaller values of d. The exception for this is of course that λ is a partition of length 2 since it is easily verified that the two sides of (3-1) agree: the only term on the right hand side of the equation that can be nonzero is $(\!(\lambda \in (k, k)P)\!)$.

When $d = k$, the left hand side of (3-1) is $\langle s_{(k,k)} * s_{(2k)}, s_\lambda \rangle$, which is 1 if $\lambda = (k, k)$ and 0 otherwise. On the right hand side of (3-1) we have $(\!(\lambda \in (k, k)P)\!)$ is 1 if and only if $\lambda = (k, k)$ and all other terms are 0, and hence the two expressions agree.

If $d = k + 1$, then $s_{(k+1,k+1)} * s_{(2k-1,1)} = s_{(k+1,k,1)} + s_{(k+2,k)}$. The only partitions λ of $2k + 2$ such that the indicator functions on the right hand side of (3-1) can be satisfied are $(\!(\lambda \in (k, k)P)\!)$ when $\lambda = (k+2, k)$ and $(\!(\lambda \in (k+1, k, 1)P)\!)$ when $\lambda = (k+1, k, 1)$. All others must be 0 because the partitions that are subtracted off are larger than $2k+2$.

Now assume that (3-1) holds for all values strictly smaller than d. If $\ell(\lambda) = 4$, then $\lambda - \gamma \in P$ if and only if $\lambda - \gamma - (1^4) \in P$ for all partitions γ of length less than or equal to 3, so

$$\begin{aligned}\langle s_{(d,d)} * s_{(d+k,d-k)}, s_\lambda\rangle &= \langle s_{(d-2,d-2)} * s_{(d+k-2,d-k-2)}, s_{\lambda-(1^4)}\rangle \\ &= \sum_{i=0}^{k} (\!(\lambda-(1^4) \in (k+i, k, i)P)\!) + \sum_{i=1}^{k} (\!(\lambda-(1^4) \in (k+i+1, k+1, i)P)\!) \\ &= \sum_{i=0}^{k} (\!(\lambda \in (k+i, k, i)P)\!) + \sum_{i=1}^{k} (\!(\lambda \in (k+i+1, k+1, i)P)\!).\end{aligned}$$

So we can now assume that $\ell(\lambda) = 3$. By (2-6) we need to consider the coefficients of the form $\langle s_{(d,d)} * s_{(d+k,d-k)}, s_\mu\rangle$, where s_μ appears in the expansion of $s_{(1)} s_{\lambda-(1^3)}$ or $s^\perp_{(1)} s_{\lambda-(1^3)}$. If λ has three distinct parts and $\lambda_3 \geq 2$ then $\mu = \lambda - \delta$, where

$$\delta \in \{(1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 1, 1, -1), (2, 1, 1), (1, 2, 1), (1, 1, 2)\}$$

and we can assume by induction that these expand into terms having the form $\pm(\!(\lambda - \delta - \gamma \in P)\!)$, where γ is a partition. However, if λ is not distinct or $\lambda_3 = 1$, then for some δ in the set, $\lambda - \delta$ will not be a partition and $(\!(\lambda - \delta \in \gamma P)\!)$ will be 0, and we can add these terms to our formulas so that we can treat the argument uniformly and not have to consider different possible λ.

One obvious reduction we can make to treat the expressions more uniformly is to note that $(\!(\lambda - (1, 1, 1, -1) \in \gamma P)\!) = (\!(\lambda - (2, 2, 2) \in \gamma P)\!)$.

Let

$$\begin{aligned}C_1 &= \{(1, 1, 0), (1, 0, 1), (0, 1, 1), (2, 2, 2)\},\\ C_2 &= \{(2, 1, 1), (1, 2, 1), (1, 1, 2)\}.\end{aligned}$$

By the induction hypothesis and (2-6), we have

$$\begin{aligned}&\langle s_{(d,d)} * s_{(d+k,d-k)}, s_\lambda\rangle \\ &\quad= \sum_{\delta\in C_1}\Big(\sum_{i=0}^{k} (\!(\lambda-\delta \in (k+i, k, i)P)\!) + \sum_{i=1}^{k} (\!(\lambda-\delta \in (k+i+1, k+1, i)P)\!)\Big) \\ &\qquad - \sum_{\delta\in C_2}\Big(\sum_{i=0}^{k} (\!(\lambda-\delta \in (k+i, k, i)P)\!) + \sum_{i=1}^{k} (\!(\lambda-\delta \in (k+i+1, k+1, i)P)\!)\Big).\end{aligned}$$

We notice that

$$\begin{aligned}\lambda - (2, 2, 2) - (k+i, k, i) &= \lambda - (1, 1, 2) - (k+i+1, k+1, i),\\ \lambda - (2, 1, 1) - (k+i, k, i) &= \lambda - (1, 0, 1) - (k+i+1, k+1, i),\\ \lambda - (1, 2, 1) - (k+i, k, i) &= \lambda - (0, 1, 1) - (k+i+1, k+1, i),\end{aligned}$$

so the corresponding terms always cancel. With this reduction, we are left with the terms

$$\sum_{\delta\in C_3}\sum_{i=0}^{k}(\!(\lambda-\delta\in(k+i,k,i)P)\!)+\sum_{\delta\in C_4}\sum_{i=1}^{k}(\!(\lambda-\delta\in(k+i+1,k+1,i)P)\!)$$
$$-\sum_{i=0}^{k}(\!(\lambda-(1,1,2)\in(k+i,k,i)P)\!)-\sum_{\delta\in C_5}\sum_{i=1}^{k}(\!(\lambda-\delta\in(k+i+1,k+1,i)P)\!)$$
$$-(\!(\lambda-(1,2,1)\in(k,k)P)\!)-(\!(\lambda-(2,1,1)\in(k,k)P)\!)+(\!(\lambda-(2,2,2)\in(k,k)P)\!),$$

where

$$C_3=\{(0,1,1),(1,0,1),(1,1,0)\},$$
$$C_4=\{(1,1,0),(2,2,2)\},$$
$$C_5=\{(2,1,1),(1,2,1)\}.$$

Next we notice that

$$\lambda-(1,1,2)-(k+i,k,i)=\lambda-(0,1,1)-(k+i+1,k,i+1),$$
$$\lambda-(2,1,1)-(k+i+1,k+1,i)=\lambda-(1,1,0)-(k+i+2,k+1,i+1),$$
$$\lambda-(2,2,2)-(k+i+1,k+1,i)=\lambda-(1,2,1)-(k+i+2,k+1,i+1).$$

Then by canceling these terms and joining the compositions that are being subtracted off in the sum, these sums reduce to the expression

$$\sum_{i=0}^{k}(\!(\lambda\in(k+i+1,k,i+1)P)\!)+\sum_{i=0}^{k}(\!(\lambda\in(k+i+1,k+1,i)P)\!)$$
$$+(\!(\lambda\in(k,k+1,1)P)\!)+(\!(\lambda\in(k+3,k+2,1)P)\!)$$
$$+(\!(\lambda\in(2k+3,k+3,k+2)P)\!)+(\!(\lambda\in(k+2,k+2,2)P)\!)$$
$$-(\!(\lambda\in(2k+1,k+1,k+2)P)\!)-(\!(\lambda\in(2k+3,k+2,k+1)P)\!)$$
$$-(\!(\lambda\in(k+3,k+3,2)P)\!)-(\!(\lambda\in(k+1,k+2,1)P)\!)$$
$$-(\!(\lambda\in(k+2,k+1,1)P)\!).$$

Since $\ell(\lambda)=3$, if $\lambda-(a,b)\in P$, then $\lambda_1-a\geq\lambda_2-b\geq\lambda_3\geq 2$, which is true if and only if $\lambda_1-a-2\geq\lambda_2-b-2\geq\lambda_3-2\geq 0$. In particular,

$$(\!(\lambda\in(k+2,k+2,2)P)\!)=(\!(\lambda\in(k,k)P)\!),$$
$$(\!(\lambda\in(k+3,k+3,2)P)\!)=(\!(\lambda\in(k+1,k+1)P)\!).$$

By verifying a few conditions it is easy to check that $\lambda\in(r,s,s+1)P$ if and only if $\lambda\in(r+2,s+2,s+1)P$, and similarly $\lambda\in(s,s+1,r)P$ if and only if $\lambda\in(s+2,s+1,r)P$. With this relationship, we have these equivalences between the terms appearing in the expression above:

$$\begin{aligned}
(\!(\lambda \in (k, k+1, 1)P)\!) &= (\!(\lambda \in (k+2, k+1, 1)P)\!),\\
(\!(\lambda \in (k+1, k+2, 1)P)\!) &= (\!(\lambda \in (k+3, k+2, 1)P)\!),\\
(\!(\lambda \in (2k+1, k+1, k+2)P)\!) &= (\!(\lambda \in (2k+3, k+3, k+2)P)\!),\\
(\!(\lambda \in (2k+1, k, k+1)P)\!) &= (\!(\lambda \in (2k+3, k+2, k+1)P)\!).
\end{aligned}$$

After we cancel these terms the expression reduces to

$$\sum_{i=0}^{k-1}(\!(\lambda \in (k+i+1, k, i+1)P)\!) + \sum_{i=1}^{k}(\!(\lambda \in (k+i+1, k+1, i)P)\!) + (\!(\lambda \in (k, k)P)\!).$$

This concludes the proof by induction on d since we know the identity holds for each partition λ of length 2, 3 or 4. □

4. Combinatorial and symmetric function consequences

4.1. ***Tableaux of height less than or equal to* 4.** Since every partition of even size and of length less than or equal to 4 lies in either P or $\overline{P}$, Corollary 3.6 has the following corollary.

Corollary 4.1. *For d a positive integer*,

$$\sum_{\lambda \vdash 2d, \ell(\lambda)\le 4} s_\lambda = s_{(d,d)} * (s_{(d,d)} + s_{(d+1,d-1)}),$$

$$\sum_{\lambda \vdash 2d-1, \ell(\lambda)\le 4} s_\lambda = s_{(d,d-1)} * s_{(d,d-1)}.$$

Proof. For the sum over partitions of $2d$, (1-1) (or (3-1)) says that $s_{(d,d)} * s_{(d,d)}$ is the sum over all s_λ with $\lambda \vdash 2d$ having four even parts or four odd parts, and Corollary 3.6 says that $s_{(d,d)}*s_{(d+1,d-1)}$ is the sum over s_λ with $\lambda \vdash 2d$ where λ does not have four odd parts or four even parts. Hence, $s_{(d,d)} * s_{(d,d)} + s_{(d,d)} * s_{(d+1,d-1)}$ is the sum over s_λ where λ runs over all partitions with less than or equal to 4 parts.

For the other identity, we use (2-3) to derive

$$\langle s_{(d,d-1)} * s_{(d,d-1)}, s_\lambda\rangle = \langle s_{(d,d)} * s_{(d,d)}, s_{(1)}s_\lambda\rangle.$$

If λ is a partition of $2d-1$, then the expression is 0 if $\ell(\lambda) > 4$; if $\ell(\lambda) \le 4$ then $s_{(1)}s_\lambda$ is a sum of at most 5 terms, $s_{(\lambda_1+1,\lambda_2,\lambda_3,\lambda_4)}$, $s_{(\lambda_1,\lambda_2+1,\lambda_3,\lambda_4)}$, $s_{(\lambda_1,\lambda_2,\lambda_3+1,\lambda_4)}$, $s_{(\lambda_1,\lambda_2,\lambda_3,\lambda_4+1)}$ and $s_{(\lambda_1,\lambda_2,\lambda_3,\lambda_4,1)}$. Because λ has exactly 3 or 1 terms that are odd, exactly one of these will have 4 even parts or 4 odd parts. □

Regev [1981], Gouyou-Beauchamps [1989], Gessel [1990] and later Bergeron, Krob, Favreau, and Gascon [Bergeron et al. 1995; Bergeron and Gascon 2000] studied tableaux of bounded height. For $y_k(n)$ equal to the number of standard tableaux of height less than or equal to k, Gessel [1990] remarks that expressions for $y_k(n)$ exist for $k = 2, 3, 4, 5$ that are simpler than the k-fold sum that one would expect to see. This is perhaps because all four of these cases have more general formulas in terms of characters.

In the case $k = 4$, Corollary 4.1 is a statement about characters indexed by partitions of bounded height. In particular, if those characters are evaluated at the identity we see a previously known result:

Corollary 4.2. $y_4(n) = C_{\lfloor (n+1)/2 \rfloor} C_{\lceil (n+1)/2 \rceil}$.

This follows from the hook length formula that says the number of standard tableaux of shape (d, d) is C_d, the number of standard tableaux of $(d, d-1)$ is C_d, and the number of standard tableaux of shape that are either of shape (d, d) or $(d+1, d-1)$ is C_{d+1}.

Interestingly, some known expressions for $y_2(n)$, $y_3(n)$ and $y_5(n)$ can also be explained in terms of symmetric function identities using the Pieri rule.

4.2. *Generating functions for partitions with coefficient r in* $s_{(d,d)} * s_{(d+k,d-k)}$. An easy consequence of Theorem 3.1 is a generating function formula for the sum of the coefficients of the expressions $s_{(d,d)} * s_{(d+k,d-k)}$.

Corollary 4.3. *For a fixed* $k \geq 1$,

$$G_k(q) := \sum_{d \geq k} \Big(\sum_{\lambda \vdash 2d} \langle s_{(d,d)} * s_{(d+k,d-k)}, s_\lambda \rangle \Big) q^d = \frac{q^k + q^{k+1} + q^{2k+1} + \sum_{r=k+2}^{2k} 2q^r}{(1-q)(1-q^2)^2(1-q^3)}.$$

Remark 4.4. Corollary 4.3 only holds for $k > 0$. In the case that $k = 0$, the numerator of the expression above is different and we have from Corollary 3.6

$$\begin{aligned} G_0(q) = \sum_{d \geq 0} \Big(\sum_{\lambda \vdash 2d} \langle s_{(d,d)} * s_{(d,d)}, s_\lambda \rangle \Big) q^d &= \sum_{d \geq 0} \langle s_{(d,d)} * s_{(d,d)}, s_{(d,d)} * s_{(d,d)} \rangle q^d \\ &= \sum_{d \geq 0} \langle [P], [P] \rangle q^d = \sum_{d \geq 0} |P| q^d \\ &= \frac{1}{(1-q)(1-q^2)^2(1-q^3)}. \end{aligned}$$

This last equality is the formula given in [Garsia et al. 2010, Corollary 1.2] and it follows because the generating function for partitions with even parts and length less than or equal to 4 is $1/((1-q)(1-q^2)(1-q^3)(1-q^4))$, and the generating function for the partitions of $2d$ with odd parts of length less than or equal to 4 is $q^2/((1-q)(1-q^2)(1-q^3)(1-q^4))$. The sum of these two generating functions is equal to a generating function for the number of nonzero coefficients of $s_{(d,d)} * s_{(d,d)}$.

Proof. Theorem 3.1 gives a formula for $s_{(d,d)} * s_{(d+k,d-k)}$ in terms of rugs of the form $[\gamma P]$. We can calculate that for each of the rugs that appears in this expression

$$\begin{aligned}\sum_{d\ge k}\Big(\sum_{\lambda\vdash 2d}\langle[\gamma P], s_\lambda\rangle\Big)q^d &= \sum_{d\ge k}\Big(\sum_{\lambda\vdash 2d}\langle[P], s_{\lambda-\gamma}\rangle\Big)q^d\\ &= \sum_{d\ge k}\Big(\sum_{\mu\vdash(2d-|\gamma|)}\langle[P], s_\mu\rangle\Big)q^{d+|\gamma|/2}\\ &= \frac{q^{|\gamma|/2}}{(1-q)(1-q^2)^2(1-q^3)}.\end{aligned}$$

Now $s_{(d,d)} * s_{(d+k,d-k)}$ is the sum of rugs of the form $[\gamma P]$ with γ equal to (k,k), $(k+1,k,1)$ and $(2k+1,k+1,k)$ each contribute a term to the numerator of the form q^k, q^{k+1} and q^{2k+1}, respectively. The rugs in which γ is equal to $(k+i+1,k+1,i)$ and $(k+i+1,k,i+1)$ for $1\le i\le k-1$ each contribute a term $2q^{k+i+1}$ to the numerator. □

In order to compute other generating functions of Kronecker products, we need the following very surprising theorem. It says that the partitions such that the $C_{(d,d)(d+k,d-k)\lambda}$ are of coefficient $r>1$ are exactly those partitions $\gamma+(6,4,2)$ for which $C_{(d-6,d-6)(d-6+(k-2),d-6-(k-2))\gamma}$ is equal to $r-1$.

Theorem 4.5. *For $k\ge 2$, assume that $C_{(d,d)(d+k,d-k)\lambda}>0$. Then*

$$C_{(d+6,d+6)(d+k+8,d-k+4)(\lambda+(6,4,2))} = C_{(d,d)(d+k,d-k)\lambda}+1.$$

Lemma 4.6. *For γ a partition with $\ell(\gamma)\le 4$, $\lambda\in\gamma P$ if and only if $\lambda+(6,4,2)$ is in both $(\gamma_1+2,\gamma_2+2,\gamma_3,\gamma_4)P$ and $(\gamma_1+4,\gamma_2+2,\gamma_3+2,\gamma_4)P$.*

Proof. If $\lambda\in\gamma P$, then $\lambda-\gamma$ is a partition with four even parts or four odd parts. Hence, both $\lambda-\gamma+(2,2)=(\lambda+(6,4,2))-(\gamma+(4,2,2))$ and $\lambda-\gamma+(4,2,2)=(\lambda+(6,4,2))-(\gamma+(2,2))$ are elements of P.

Conversely, assume that $\lambda+(6,4,2)$ is as stated. Then $\lambda_1+6-(\gamma_1+4)\ge\lambda_2+4-(\gamma_2+2)$, $\lambda_2+4-(\gamma_2+2)\ge\lambda_3+2-\gamma_3$, and $\lambda_3+2-(\gamma_3+2)\ge\lambda_4-\gamma_4\ge 0$. This implies that $\lambda-\gamma$ is a partition and since $\lambda-\gamma+(2,2)$ has four even or four odd parts; then so does $\lambda-\gamma$ and hence $\lambda\in\gamma P$. □

Proof of Theorem 4.5. Consider the case where λ is a partition of $2d$ with $\lambda_2-k\equiv\lambda_4 \pmod 2$ since the case where $\lambda_2-k\not\equiv\lambda_4 \pmod 2$ is analogous and just uses different nonzero terms in the sum below. From Theorem 3.1, we have

$$C_{(d,d)(d+k,d-k)\lambda} = \sum_{i=0}^{k}(\!(\lambda\in(k+i,k,i)P)\!), \tag{4-1}$$

since the other terms are clearly zero in this case. If $\lambda_2-k\ge\lambda_3$, then the terms in this sum will be nonzero as long as $0\le i\le\lambda_3-\lambda_4$ and $0\le k+i\le\lambda_1-\lambda_2$

and $\lambda_3 - i \equiv \lambda_4 \pmod 2$. Consider the case where $\lambda_3 \equiv \lambda_4 \pmod 2$; then (4-1) is equal to $a+1$, where $a = \lfloor \min(\lambda_3 - \lambda_4, \lambda_1 - \lambda_2 - k, k)/2 \rfloor$ since the terms that are nonzero in this sum are $(\!(\lambda \in (k+2j, k, 2j)P)\!)$, where $0 \le j \le a$. By Lemma 4.6, these terms are true if and only if $(\!(\lambda + (6,4,2) \in (k+2+2j, k+2, 2j)P)\!)$ are true for all $0 \le j \le a+1$. But again by Theorem 3.1, in this case we also have

$$C_{(d+6,d+6)(d+k+8,d-k-4)(\lambda+(6,4,2))} = a+2 = C_{(d,d)(d+k,d-k)\lambda} + 1.$$

The case in which $\lambda_3 \equiv \lambda_4 + 1 \pmod 2$ is similar, but the terms of the form $(\!(\lambda \in (k+2j+1, k, 2j+1)P)\!)$ in (4-1) are nonzero if and only if the terms $(\!(\lambda + (6,4,2) \in (k+2+2j+1, k+2, 2j+1)P)\!)$ contribute to the expression for $C_{(d+6,d+6)(d+k+8,d-k-4)(\lambda+(6,4,2))}$ and there is exactly one more nonzero term; hence $C_{(d+6,d+6)(d+k+8,d-k-4)(\lambda+(6,4,2))} = C_{(d,d)(d+k,d-k)\lambda} + 1$. □

Now for computations it is useful to have a way of determining exactly the number of partitions of $2d$ that have a given coefficient. For integers $d, k, r > 0$, we let $L_{d,k,r}$ be the number of partitions λ of $2d$ with $\langle s_{(d,d)} * s_{(d+k,d-k)}, s_\lambda \rangle = r$. Theorem 3.1 has shown that $L_{d,k,r} = 0$ for $r > \lfloor k/2 \rfloor + 1$, and Theorem 4.5 says $L_{d,k,r} = L_{d-6,k-2,r-1} + 1$ for $r > 1$. These recurrences will allow us to completely determine the generating functions for the coefficients $L_{d,k,r}$. We set

$$L_{k,r}(q) = \sum_{d \ge 0} L_{d,k,r} q^d.$$

Corollary 4.7. *With the convention that $G_k(q) = 0$ for $k < 0$, we have $L_{k,r}(q) = 0$ for $r > \lfloor k/2 \rfloor + 1$, and*

$$L_{k,1}(q) = G_k(q) - 2q^6 G_{k-2}(q) + q^{12} G_{k-4}(q), \tag{4-2}$$

$$L_{k,r}(q) = q^{6r-6} L_{k-2r+2,1}(q). \tag{4-3}$$

Proof. Theorem 4.5 explains (4-3) because

$$\begin{aligned} L_{k,r}(q) &= \sum_{d \ge 0} \#\{\lambda : C_{(d,d)(d+k,d-k)\lambda} = r\} q^d \\ &= \sum_{d \ge 0} \#\{\lambda : C_{(d-6,d-6)(d+k-8,d-k+4)(\lambda+(6,4,2))} = r-1\} q^d \\ &= \sum_{d \ge 0} \#\{\lambda : C_{(d-6r+6,d-6r+6)(d+k-8r+8,d-k+4r-4)(\lambda+(6r-6,4r-4,2r-2))} = 1\} q^d \\ &= q^{6r-6} \sum_{d \ge 0} \#\{\lambda : C_{(d-6r+6,d-6r+6)(d+k-8r+8,d-k+4r-4)(\lambda+(6r-6,4r-4,2r-2))} = 1\} \\ &\qquad \cdot q^{d-6r+6} \\ &= q^{6r-6} L_{k-2r+2,1}(q). \end{aligned}$$

Now we also have by definition and Theorem 3.1 that

$$G_k(q) = \sum_{r=1}^{\lfloor k/2 \rfloor + 1} r L_{k,r}(q). \tag{4-4}$$

Hence, we can use this formula and (4-3) to define $L_{k,r}(q)$ recursively. It remains to show that the formula for $L_{k,1}(q)$ stated in (4-2) satisfies this formula, which we do by induction. Note that $L_{0,1}(q) = G_0(q)$ and $L_{1,1}(q) = G_1(q)$ and $L_{k,1}(q) = 0$ for $k < 0$. Then assuming that the formula holds for values smaller than $k > 1$, we have from (4-4)

$$\begin{aligned}
L_{k,1}(q) &= G_k(q) - \sum_{r=2}^{\lfloor k/2 \rfloor + 1} r L_{k,r}(q) \\
&= G_k(q) - \sum_{r \geq 2} r q^{6r-6} L_{k-2r+2,1}(q) \\
&= G_k(q) - \sum_{r \geq 2} r q^{6r-6} (G_{k-2r+2}(q) - 2q^6 G_{k-2r}(q) + q^{12} G_{k-2r-2}(q)) \\
&= G_k(q) - \sum_{r \geq 1} (r+1) q^{6r} G_{k-2r}(q) + \sum_{r \geq 2} 2r q^{6r} G_{k-2r}(q) \\
&\qquad\qquad - \sum_{r \geq 3} (r-1) q^{6r} G_{k-2r}(q) \\
&= G_k(q) - 2q^6 G_{k-2}(q) + q^{12} G_{k-4}(q).
\end{aligned}$$

Therefore, by induction (4-2) holds for all $k > 0$. □

References

[Ballantine and Orellana 2005] C. M. Ballantine and R. C. Orellana, "On the Kronecker product $s_{(n-p,p)} * s_\lambda$", *Electron. J. Combin.* **12** (2005), Research Paper 28, 26 pp. MR 2006g:05227 Zbl 1076.05082

[Ballentine and Orellana 2005] C. Ballentine and R. Orellana, "A combinatorial interpretation for the coefficients in the Kronecker product $s_{(n,n-p)} * s_\lambda$", *Sém. Lothar. Combin.* **54A** (2005), Art. B54Af.

[Bergeron and Gascon 2000] F. Bergeron and F. Gascon, "Counting Young tableaux of bounded height", *J. Integer Seq.* **3**:1 (2000), Art. 00.1.7. MR 2001g:05102 Zbl 0953.05079

[Bergeron et al. 1995] F. Bergeron, L. Favreau, and D. Krob, "Conjectures on the enumeration of tableaux of bounded height", *Discrete Math.* **139** (1995), 463–468. MR 96f:05188 Zbl 0833.05079

[Bessenrodt and Behns 2004] C. Bessenrodt and C. Behns, "On the Durfee size of Kronecker products of characters of the symmetric group and its double covers", *J. Algebra* **280**:1 (2004), 132–144. MR 2005d:20020 Zbl 1073.20004

[Bessenrodt and Kleshchev 1999] C. Bessenrodt and A. Kleshchev, "On Kronecker products of complex representations of the symmetric and alternating groups", *Pacific J. Math.* **190**:2 (1999), 201–223. MR 2000i:20017 Zbl 1009.20013

[Bürgisser and Ikenmeyer 2008] P. Bürgisser and C. Ikenmeyer, "The complexity of computing Kronecker coefficients", pp. 357–368 in *20th Annual International Conference on Formal Power Series and Algebraic Combinatorics* (Valparaiso-Viña del Mar, Chile, 2008), Discrete Mathematics & Theoretical Computer Science, 2008.

[Dvir 1993] Y. Dvir, "On the Kronecker product of S_n characters", *J. Algebra* **154**:1 (1993), 125–140. MR 94a:20023 Zbl 0848.20006

[Garsia et al. 2009] A. Garsia, N. Wallach, G. Xin, and M. Zabrocki, "Hilbert series of invariants, constant terms and Kostka–Foulkes polynomials", *Discrete Math.* **309**:16 (2009), 5206–5230. MR 2548921 Zbl 05648061

[Garsia et al. 2010] A. Garsia, N. Wallach, G. Xin, and M. Zabrocki, "Kronecker coefficients via symmetric functions and constant term identities", preprint, 2010, available at http://garsia.math.yorku.ca/~zabrocki/papers/KronViaSym.pdf.

[Gessel 1990] I. M. Gessel, "Symmetric functions and P-recursiveness", *J. Combin. Theory Ser. A* **53**:2 (1990), 257–285. MR 91c:05190 Zbl 0704.05001

[Gouyou-Beauchamps 1989] D. Gouyou-Beauchamps, "Standard Young tableaux of height 4 and 5", *European J. Combin.* **10**:1 (1989), 69–82. MR 90b:05013 Zbl 0672.05012

[Luque and Thibon 2006] J.-G. Luque and J.-Y. Thibon, "Algebraic invariants of five qubits", *J. Phys. A* **39**:2 (2006), 371–377. MR 2006h:81048 Zbl 1084.81020

[Macdonald 1995] I. G. Macdonald, *Symmetric functions and Hall polynomials*, 2nd ed., Oxford University Press, New York, 1995. MR 96h:05207 Zbl 0824.05059

[Regev 1980] A. Regev, "The Kronecker product of S_n-characters and an $A \otimes B$ theorem for Capelli identities", *J. Algebra* **66**:2 (1980), 505–510. MR 82f:16015 Zbl 0445.16012

[Regev 1981] A. Regev, "Asymptotic values for degrees associated with strips of Young diagrams", *Adv. in Math.* **41**:2 (1981), 115–136. MR 82h:20015 Zbl 0509.20009

[Remmel and Whitehead 1994] J. B. Remmel and T. Whitehead, "On the Kronecker product of Schur functions of two row shapes", *Bull. Belg. Math. Soc. Simon Stevin* **1**:5 (1994), 649–683. MR 96h:05211 Zbl 0827.20017

[Rosas 2001] M. H. Rosas, "The Kronecker product of Schur functions indexed by two-row shapes or hook shapes", *J. Algebraic Combin.* **14**:2 (2001), 153–173. MR 2003b:05154 Zbl 0999.05098

[Vallejo 1999] E. Vallejo, "Stability of Kronecker products of irreducible characters of the symmetric group", *Electron. J. Combin.* **6** (1999), R39. MR 2001b:20022 Zbl 0944.05099

[Wallach 2005] N. R. Wallach, "The Hilbert series of measures of entanglement for 4 qubits", *Acta Appl. Math.* **86**:1-2 (2005), 203–220. MR 2006c:81020 Zbl 1097.81023

Received August 18, 2009.

ANDREW A. H. BROWN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BC V6T 1Z2
CANADA
Current address:
DEPARTMENT OF COMBINATORICS AND OPTIMIZATION
UNIVERSITY OF WATERLOO
WATERLOO, ON N2L 3G1
CANADA

a29brown@services107.math.uwaterloo.ca

Stephanie van Willigenburg
Department of Mathematics
University of British Columbia
Vancouver, BC V6T 1Z2
Canada
steph@math.ubc.ca

Mike Zabrocki
Department of Mathematics and Statistics
York University
Toronto, ON M3J 1P3
Canada
zabrocki@mathstat.yorku.ca
http://garsia.math.yorku.ca/~zabrocki/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

METRIC PROPERTIES OF HIGHER-DIMENSIONAL THOMPSON'S GROUPS

JOSÉ BURILLO AND SEAN CLEARY

Higher-dimensional Thompson's groups nV are finitely presented groups that generalize dyadic self-maps of the unit interval to dyadic self-maps of n-dimensional unit cubes. We describe some of the metric properties of these groups. We describe their elements based upon tree-pair diagrams and give upper and lower bounds for word length in terms of the size of the diagrams. Though these bounds are somewhat separated, we show that there are elements realizing the lower bounds and that the fraction of elements that are close to the upper bound converges to 1, showing that the bounds are optimal and that the upper bound is generically achieved.

Introduction

Thompson's groups provide a wide range of interesting examples of unusual group-theoretic behavior. The family of Thompson's groups includes the original groups described by Thompson, commonly denoted F, T and V, as well as generalizations in many different directions. The bulk of these generalizations includes groups that can be regarded as self-maps of the unit interval. Brin [2004; 2005] generalized V to higher-dimensional groups nV, which are described naturally in terms of dyadic self-maps of n-dimensional cubes. Little is known about these groups aside from their simplicity. Brin describes their elements geometrically in terms of dyadic interpolations of collections of dyadic blocks, and gives presentations for $2V$.

We describe elements in higher-dimensional Thompson's groups as being given by tree-pair diagrams. Though these diagrams usually take advantage of the natural left-to-right ordering of subintervals of the unit interval, by using several types of carets, Brin describes elements of nV via tree-pair diagrams. A natural question is how the size of these tree-pair diagrams corresponds to the word length of elements with respect to finite generating sets.

MSC2000: 20F65.
Keywords: Thompson's groups.
The authors are grateful for the hospitality of the Centre de Recerca Matemàtica. Burillo acknowledges support from MEC grant MTM2006-13544-C02-02 and is grateful for the hospitality of The City College of New York. This work was supported in part by NSF grant 0811002 and grants from The City University of New York PSC-CUNY Research Award Program.

We give upper and lower bounds for the word lengths of elements of higher-dimensional Thompson's groups nV with respect to finite generating sets, in terms of the size of tree-pair descriptions of elements. An element with a minimal tree-pair description of size N has word length between $\log N$ and $N \log N$ (up to the standard affine equivalences) with respect to the standard finite generating set. This of course also thus holds for any finite generating set.

1. Background on higher-dimensional Thompson's groups

Brin [2004] describes higher-dimensional Thompson's groups nV, giving presentations and showing that the group $2V$ is simple. Brin [2004; 2005] defines nV as a subgroup of the homeomorphism group of the n-fold product of the Cantor set with itself, and then shows there are a number of equivalent characterizations. The one we use here is Brin's characterization [2005] in terms of equivalence classes of labeled tree pair diagrams. The properties of the groups nV relevant here are as follows.

We denote the unit interval [0, 1] by I, and the n-dimensional unit cube by I^n.

An element of V is given by two finite-rooted binary trees with the same number of leaves and a permutation, which represents a bijection between the leaves of the two trees. Hence an element of V can be seen as a triple (T_+, π, T_-), where T_+ and T_- are trees with k leaves, and $\pi \in \mathcal{S}_k$.

Each binary tree can be seen as a way of subdividing the interval I into dyadic subintervals of the type $[i/2^r, i+1/2^r]$, where $r > 0$ and $0 \le i < 2^r$. A rooted binary tree gives instructions for successive halvings of subintervals to obtain a particular dyadic subdivision. Given two binary trees with n leaves and a permutation in $\mathcal{S}_n$, an element in V is represented as a left-continuous map of the interval into itself, sending each interval in the first subdivision to a corresponding interval in the second subdivision, as specified by the given permutation. Though not every dyadic subdivision of I can be obtained by a successive halving process described by trees, every dyadic subdivision has a refinement that can be obtained by a successive halving process.

For more details on the group V, including presentations and a proof of its simplicity, see Cannon, Floyd and Parry [1996] as well as Brin [2004].

To obtain an element in $2V$, we will define a partition of I^2 into dyadic rectangles of the type

$$\left[\frac{i}{2^r}, \frac{i+1}{2^r}\right] \times \left[\frac{j}{2^s}, \frac{j+1}{2^s}\right].$$

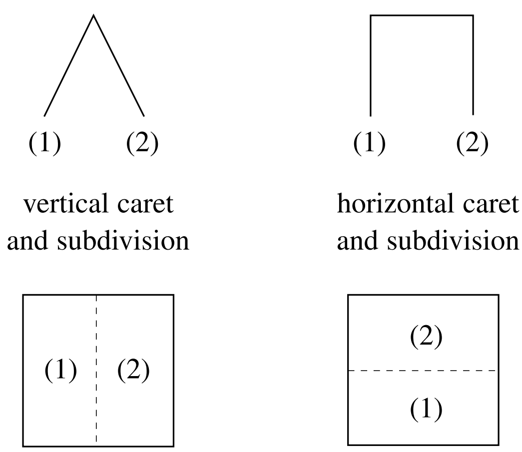
Though again, not every such partition can be realized by successive halving processes, every such partition has a refinement that can be realized in that way. For the two-dimensional successive halving process, there are two possible halvings that can occur at each stage to a specified dyadic rectangle — a horizontal or a vertical subdivision. We can obtain a refinement of any dyadic partition of I^2 via iterated horizontal and vertical subdivisions. This gives a natural means of describing elements of $2V$ as pairs of dyadic subdivisions, each with m rectangles, and a permutation in S_m giving the bijection between the rectangles.

We can describe elements of $2V$ also with pairs of binary trees, with the nodes representing steps in successive halving processes to represent the refined subdivisions. Since there are two types of subdivisions, vertical and horizontal, we will consider binary trees that contain two types of carets, *vertical* and *horizontal* carets, represented by *triangular* and *square* carets, respectively.

As with other tree-based definitions of Thompson's groups, there are numerous representatives of a given group element, with natural equivalence relations coming from expansions and contractions of leaves of trees. To multiply two elements, we typically find representatives for each of them where the trees are compatible for a natural multiplication via composition.

We define the group $2V$ as the set of equivalence classes of triples (T_+, π, T_-) in which the trees are labeled to indicate the successive halving process and π is a permutation between the leaves of the trees (where each leaf represents a dyadic rectangle).

To fix a labeling convention, we will represent a vertical subdivision with a traditional triangular caret, where the left and right leaves naturally represent the left and right rectangles. A horizontal subdivision will be represented by a square caret; in it, the left leaf represents the bottom rectangle, and the right leaf represents the top rectangle, as shown below.



We say a representative of an element is *reduced* if for each caret c in T_- of with two leaves, the permutation π does not map both leaves of c to leaves of the same caret of the same type in T_+ in the same order. We say a representative

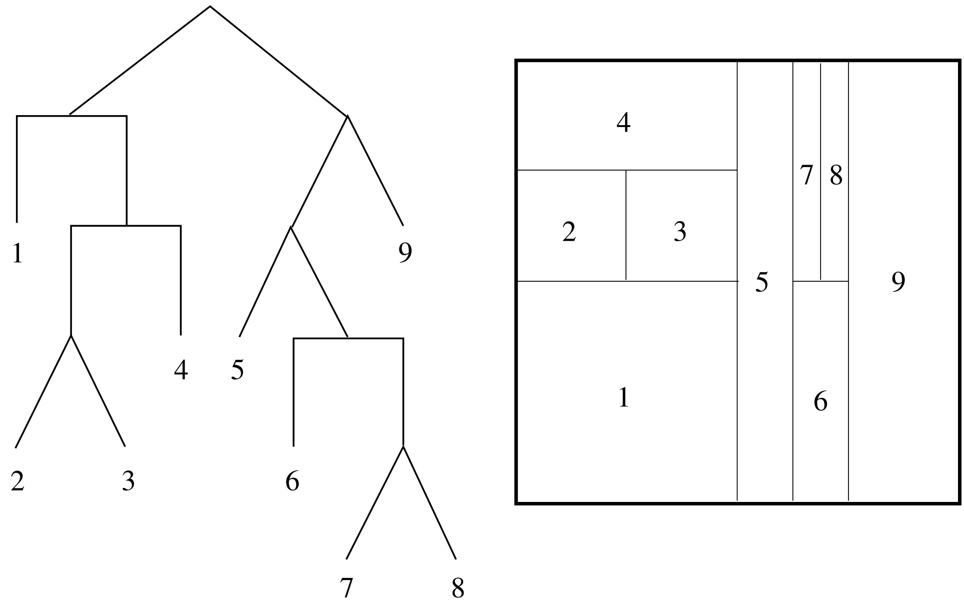


Figure 1. An example of a binary tree with the two types of carets, and its corresponding partition of I^2.

of an element is of *minimal size* if it has the minimal number of carets among all representatives in its equivalence class, and we denote by $N(x)$ the number of carets in a minimal size representative of an element x. Unlike the case of V, but similar to that of $F(n, m)$ (see Wladis [2007]), each element of minimal size may have multiple distinct representatives.

See Figure 1 for an example of a partition of I^2 and its corresponding binary tree. Every dyadic partition of the unit square has a refinement that can be obtained with a binary tree with these two types of carets.

For the general groups nV, the partitions are divisions of the unit n-cube in n-rectangles of dyadic lengths, and the corresponding binary trees have n types of carets. For simplicity, we state our results for $n = 2$, but all of the results below extend naturally to $n > 2$.

The fact that we have vertical and horizontal subdivisions brings new relations to the group. These relations arise when both types of subdivisions are combined in different orders to obtain different descriptions of the same dyadic partition of I^2. The obvious relation (and the one from which all other relations are deducible) is the combination of one subdivision of each type in the two possible orders, as illustrated in Figure 2.

Subdividing in both the vertical and horizontal directions once, but in the two possible orders, gives the same dyadic partitions, but according to our convention

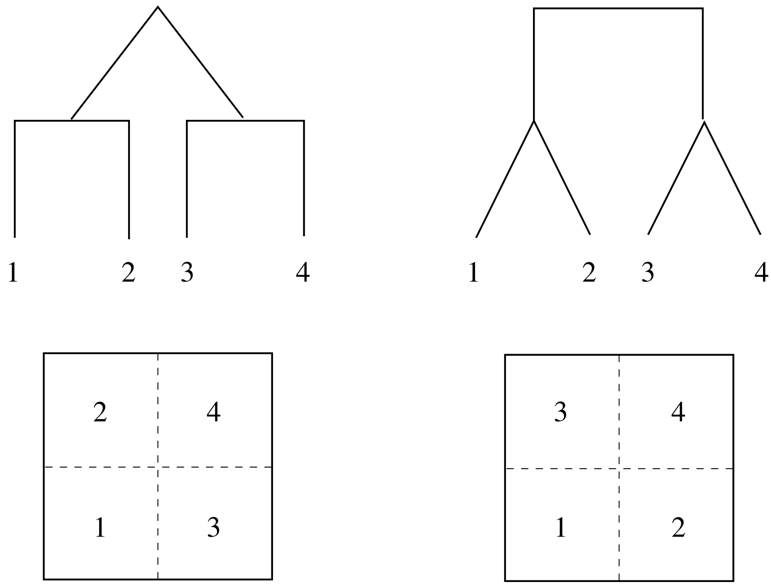


Figure 2. The relation obtained when performing vertical and horizontal subdivisions.

the resulting rectangles are numbered in different way. This makes using tree diagrams to express the multiplication of elements tricky and quite unwieldy for large elements. Even though the leaves of the two-caret-type tree diagrams are ordered in a natural way, this order is not apparent in the square, and it is not preserved when different diagrams represent the same partition. Thus, there are (nonminimal) diagrams representing the identity whose leaves are ordered in different ways — an example is illustrated in Figure 3.

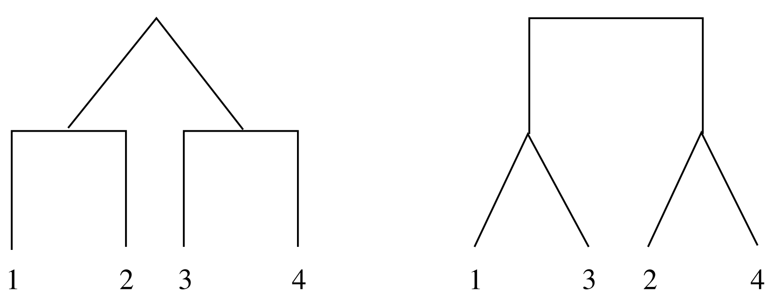


Figure 3. A tree-pair diagram for the identity element that has a nonidentity permutation.

2. Families of generators

Brin [2004; 2005] showed that the groups nV are finitely generated and gave several generating sets for $2V$. As is common with groups of the Thompson family, there is an infinite presentation, which is useful for its symmetry and regularity and which has a finite subpresentation. For the purposes of word length, we work with a standard finite generating set as described by Brin. We next define the families of generators for $2V$, following the usual procedure in Thompson's groups; see Cannon, Floyd and Parry [1996]. We build the generators on a backbone of an all-right tree with only vertical carets.

(1) The generators A_n involve only vertical subdivisions, and they are the traditional generators of Thompson's group F.

(2) The generators B_n have the one nonright caret replaced by a horizontal one.

(3) The generators C_n have the last caret of the all-right tree of horizontal type, and they are used to build horizontal carets on the right-hand-side of the tree, as will be seen later.

(4) The generators π_n and $\overline{\pi}_n$ are permutations built on an all-right tree with only vertical carets. These generators are exactly equal to those for the subgroup V appearing as the purely vertical elements.

Theorem 2.1 [Brin 2004; 2005]. *The families A_n, B_n, C_n, π_n and $\overline{\pi}_n$ generate the group $2V$.*

Proof. Since our proof is quite different from Brin's, we will include it here and use aspects of it later for metric considerations.

An element of $2V$ is given by a triple (T_+, π, T_-), where the two trees are composed of the two types of carets. First, we subdivide a given element into three elements: (T_+, id, R_k), (R_k, π, R_k), and (R_k, id, T_-), where R_k is the all-right tree with k leaves and only vertical carets. Clearly (R_k, π, R_k) is product of the permutation generators, as is the case in V already.

To obtain the element (T_+, id, R_k), we will concentrate first on the backbone of the tree T_+, that is, the sequence of carets in the right hand side. If this sequence of carets has horizontal carets in the positions $m_1, m_2, \ldots, m_p$, then the product of the generators $C_{m_1} C_{m_2} \cdots C_{m_p}$ produces exactly a backbone with horizontal carets in the desired positions. We denote this backbone by K; that is, K is the subtree of T_+ that consists only of right carets. Thus, at this stage, we have constructed the element (K, id, R_{m_p}).

Once the backbone is constructed, each new caret is obtained by a generator of the type A_i or B_i. To attach a vertical caret from the leaf labeled i, we only need to multiply by A_i on the right. Similarly, to attach a horizontal caret to leaf i, we

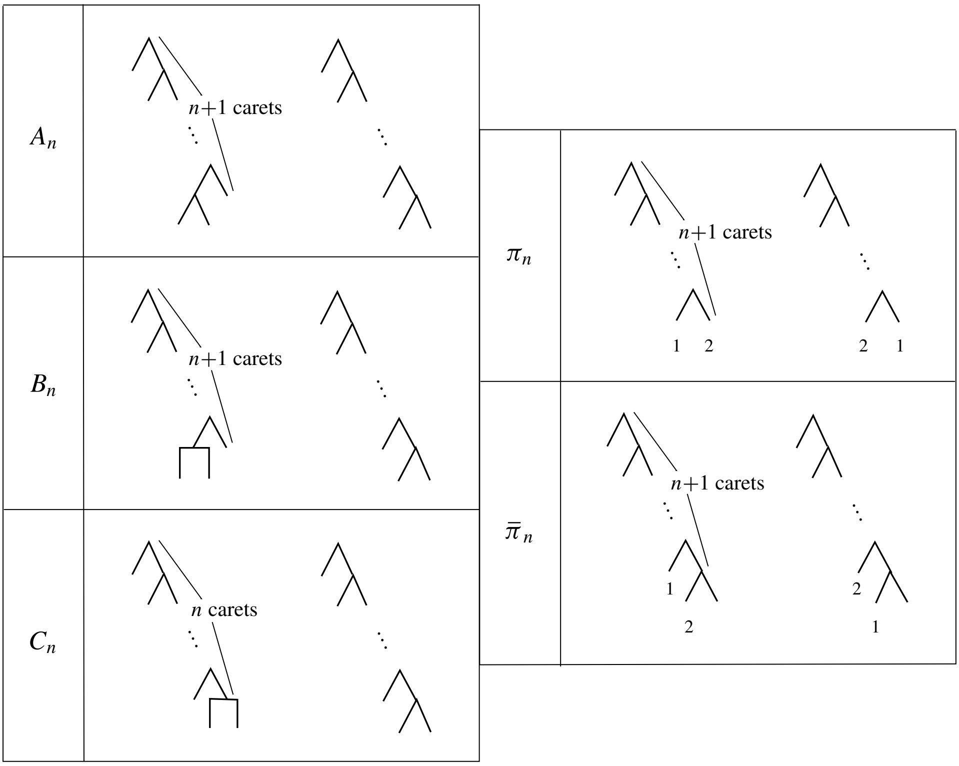


Figure 4. Generators of $2V$.

multiply by B_i on the right. This multiplication process is shown in Figure 5, with an example illustrated in Figure 6.

This proves that the element (T_+, id, R_k) is product of the generators A_i, B_i and C_i, and using inverses, we see that the entire group is generated by the full family of generators. □

An element of the type (T_+, id, R_k) is called a *positive* element of $2V$. Positive elements can always be written as products of the generators A_i, B_i and C_i without using their inverses.

In the process of proving Brin's theorem, we have obtained this:

Theorem 2.2. *Elements of* $2V$, *with respect to the standard infinite generating set* $\{A_i, B_i, C_i, \pi_i, \overline{\pi}_n\}$, *have these properties*:

- *Any positive element always admits an expression of the type*

$$C_{m_1} \cdots C_{m_p} W_1(A_{i_1}, B_{i_1}) \cdots W_r(A_{i_r}, B_{i_r}).$$

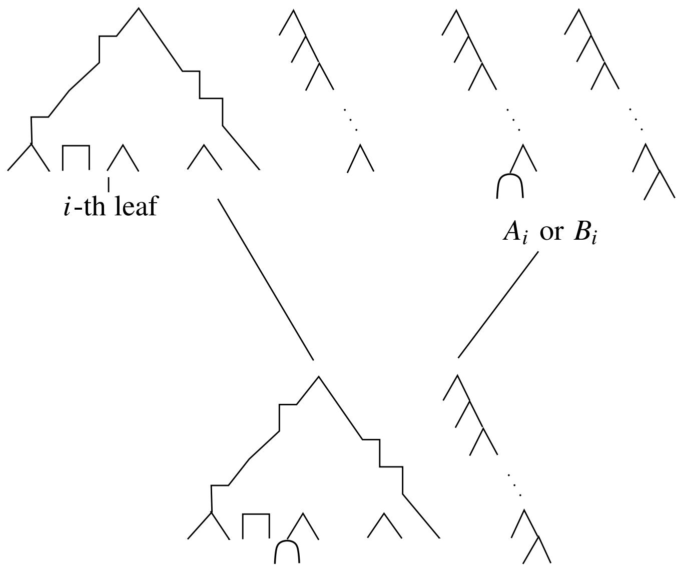


Figure 5. The building process: attaching a caret to the i-th leaf by multiplying by A_i or B_i.

where the W_i are words on the positive generators A_i and B_i and never their inverses, and where $m_1 < m_2 < \cdots < m_p$ and $i_1 < i_2 < \cdots < i_r$.

- *Each element admits an expression $P \Pi Q^{-1}$, where P and Q are positive elements and Π is a permutation on an all-right vertical tree, and thus a word in the π_i and $\bar{\pi}_i$.*

This expression will be used as a seminormal form for elements of $2V$.

Brin also shows that several finite generating sets suffice to generate each nV. We will use the finite generating set $\{A_0, A_1, B_0, B_1, C_0, C_1, \pi_0, \pi_1, \bar{\pi}_0, \bar{\pi}_1\}$. This set is larger than the smallest ones used by Brin, but is more convenient for our methods below.

3. Metric properties

We will be interested in metric properties up to the standard affine equivalence, defined here.

Definition 3.1. Given two functions $f, g : G \to \mathbb{R}$, we say that $f \prec g$ if there exists a constant $C > 0$ such that $f(x) \leq Cg(x)$ for all $x \in G$. We say that f and g are equivalent, written $f \sim g$, if $f \prec g$ and $g \prec f$.

Elements of Thompson's groups admit representations as diagrams with binary trees in one form or another, perhaps with associated permutations or braids between the leaves. For several of these groups, the distance $|x|$ from an element x

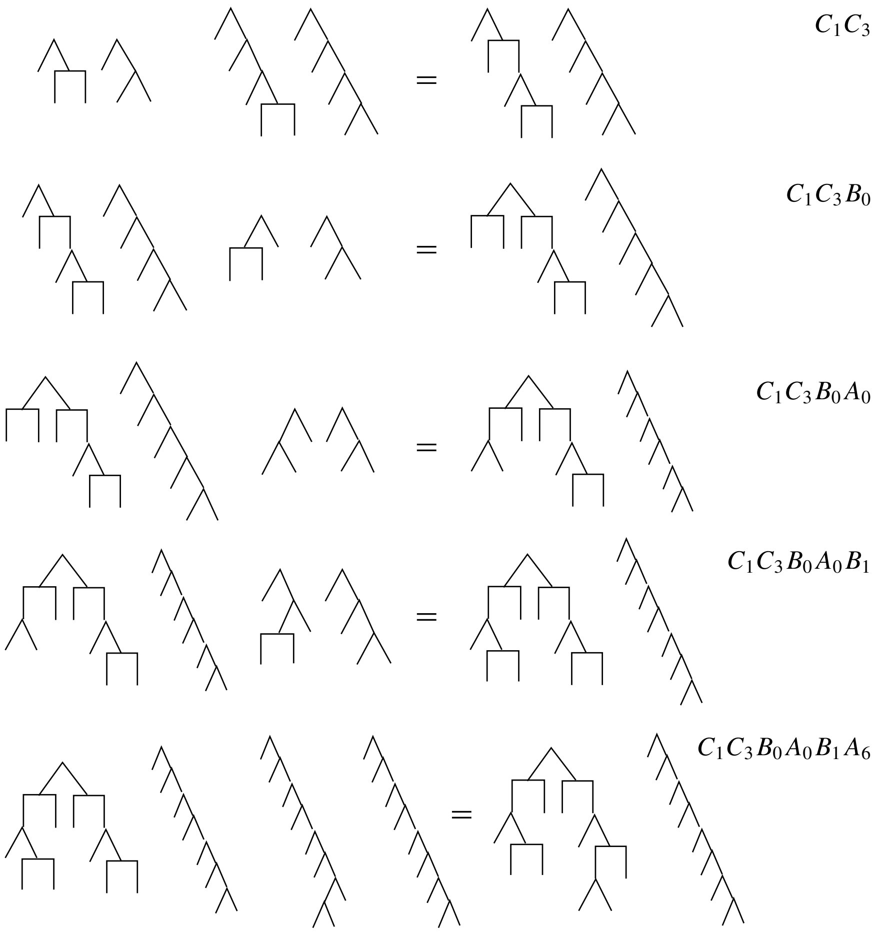


Figure 6. Constructing an element generator-by-generator.

to the identity is closely related with the number of carets $N(x)$ of the minimal diagram, indicating that the complexity of the minimal (reduced) diagram is a good indication of the size of the element.

In Thompson’s groups F and T, the two functions are equivalent: for both groups we have $|x| \sim N(x)$; see Burillo, Cleary and Stein [2001] and Burillo, Cleary, Stein and Taback [2009]. This result cannot hold in V, for counting reasons. The number of elements of V whose trees have N carets is at least the order of $N!$ (due to the permutations), while the number of distinct elements of length N can be at most exponential in N in any group.

The best possible bound for the number of carets in terms of word length in the group V is an inequality proved by Birget as [2004, Theorem 3.8], which says that

$N(x) \prec |x| \prec N(x) \log N(x)$ and that these bounds are optimal, in a sense similar to that which will be made precise in Section 4.

In the case of V, the lower bounds on word length are linear in the number of carets. Therefore, the standard inclusions $F \subset T \subset V$ are all quasiisometric embeddings; that is, F is undistorted in T, and T is undistorted in V. This property will be no longer true for inclusions in $2V$, as we will show in Section 5.

In this section we will prove the analog for $2V$ of the inequality above. In the next section, we address optimality of the bounds.

Theorem 3.2. *In the group $2V$, the word length of an element $|x|$ and number of carets $N(x)$ in a minimal size tree-pair representative satisfy*

$$\log N(x) \prec |x| \prec N(x) \log N(x),$$

and the bounds cannot be improved.

Proof. We defer the proof of its optimality to the following section.

The upper bound is proved the standard way. We take a positive word P, and according to Theorem 2.2, write it as

$$C_{m_1} \cdots C_{m_p} W_1(A_{i_1}, B_{i_1}) \cdots W_r(A_{i_r}, B_{i_r}).$$

Then, we rewrite each generator using the relations

$$A_{i+1} = A_0^{-i+1} A_1 A_0^{i-1}, \qquad B_{i+1} = A_0^{-i+1} B_1 A_0^{i-1}, \qquad C_{i+1} = A_0^{-i+1} C_1 A_0^{i-1}$$

for all $i > 1$. Since the conjugating element is always A_0, cancellations ensure that the length stays approximately the same. The bound $N \log N$ appears because of the permutations, since $N \log N$ is the diameter of $\mathcal{S}_N$ with respect to the relevant transpositions.

For the lower bound, we note that if an element has k carets, when it is multiplied by a generator it is possible that the multiplication could have up to $4k$ carets. If an element has, for instance, only horizontal subdivisions, when multiplied by A_1 the number of subdivisions (and hence the number of carets) becomes $4k$, since each one of the previous subdivision is divided in four, as illustrated in Figure 7.

If each multiplication by a generator can multiply the number of carets by 4, an element of length ℓ could have up to 4^ℓ carets. Other generators may increase word length by additive factors, but the worst case is exactly the increase by a factor of 4. Hence $N(x) \le 4^\ell$, so $\log N(x) \le |x|$. □

4. Optimality of bounds

To show the optimality of the bounds in Theorem 3.2, we first describe some properties of the lower bound for word length in terms of the number of carets.

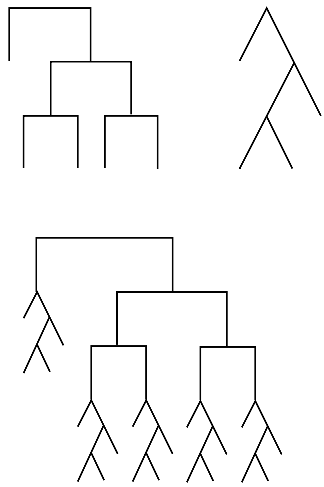


Figure 7. An illustration of why when two types of carets are involved, the number of carets gets multiplied. The tree on the top left has only horizontal subdivisions, so when joined with a tree with only vertical ones, these have to be put at every leaf.

Lemma 4.1. *The word length of an element of* $2V$ *that is represented by a tree-pair diagram of depth* D *is at least* $D/3$.

Proof. Right multiplying by a generator in the finite generating set can add levels to the tree-pair diagram. When we right multiply by a generator, we may need to add carets to T_- (and thus to T_+) if the carets in the generator are not already present in T_-. This can add at most 3 levels, which happens only in the case that we right-multiply an element (T_+, π, T_-) with T_- having no right subtree by any of A_1, B_1 or C_1 from the finite generating set. In that case, the right child of the root of T_- is a leaf labeled n and we will need to add carets to it and the corresponding leaf from T_+ to obtain a representative compatible with multiplication by the generator. If leaf $\pi^{-1}(n)$ in T_+ is at maximal level, then the level could increase but only by the number of carets added to leaf $\pi^{-1}(n)$ in T_-. Thus, any product of less than $D/3$ generators cannot have depth D. □

From this it follows that the cyclic subgroups generated by any single generator from the A_n or B_n families are undistorted in $2V$, for example.

As seen in the proof of Theorem 3.2, multiplying by a generator can increase the number of carets by a multiplicative factor. The powers of the element C_0 have exponentially many carets. In fact, the minimal number of carets needed to represent C_0^n is 2^n. We give further related specific examples later in Section 5. Any of these examples show that the lower bound of Theorem 3.2 is optimal.

To analyze the genericity of the upper bound in Theorem 3.2, we use arguments analogous to those for V. Birget [2004] showed not only that $n \log n$ is an upper bound on the growth of elements in V with respect to the number of carets n in minimal length representatives, but also that the fraction of elements close to this bound converges exponentially fast to 1. In 2V, we note analogous behavior:

Theorem 4.2. *Let H_n be the set of elements of $2V$ representable with n carets and having no representatives with fewer than n carets. The fraction of elements of H_n that have word length greater than $n \log n$ converges exponentially fast to* 1.

Proof. Here we use a counting argument, analogous to that used by Birget [2004] to count elements in V. In $2V$, we consider the subset of representatives of elements that have only horizontal subdivisions in the domain and only vertical subdivisions in the range. Their tree-pair diagrams are guaranteed to be reduced and of minimal size by a straightforward argument. So to count the set of elements of this type of diagram size n, we have C_n choices (where C_n is the n-th Catalan number) for the first all-square tree of size n and C_n choices for the all-triangular tree of size n, and $n!$ choices for the permutation. This set of elements has size $C_n^2 n!$, which is, by the Stirling formula,

$$(C_n)^2 n! = \left(\frac{(2n)!}{(n+1)!n!}\right)^2 n! = \sqrt{\frac{2}{\pi}}\,\frac{16^n}{e^n} n^{(n-2)}(1+o(1)),$$

where the $o(1)$ term goes to zero as n increases.

The number of elements of word length n in any finitely generated group with d generators is no more than $(2d)^n$. Thus we see that the ratio of elements that have word length less than $n \log_{2d} n$ out of elements that have tree-pair diagrams of size n is less than

$$\frac{n^n}{\sqrt{\frac{2}{\pi}}\,\frac{16^n}{e^n} n^{(n-2)}(1+o(1))} \sim c\frac{n^2 e^n}{16^n},$$

which converges to 0 exponentially fast, as desired. So by complementing we have the result. □

5. Distortion

Elements of $2V$ can be represented with pairs of binary trees, where the carets have been subdivided in two types. This fact makes it more difficult to multiply elements whose caret types disagree, as illustrated already in Theorem 3.2, and the number of carets can grow faster because of this situation. This phenomenon has been described already by Wladis [2007] for the group $F(2,3)$, which has also carets of two types (binary and ternary). This feature of $2V$ implies now that

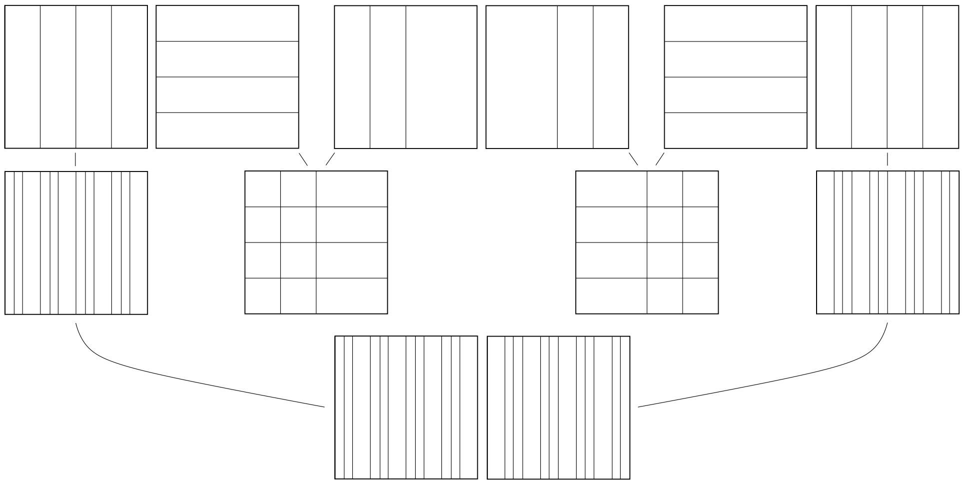


Figure 8. The process of multiplication for $C_0^{-2}A_0C_0^2$, illustrating the exponential number of subdivisions.

the groups F, T and V, when seen as subgroups of $2V$ (by doing only vertical subdivisions, for instance), are exponentially distorted.

Theorem 5.1. *The groups F, T and V are at least exponentially distorted in $2V$.*

Proof. We consider the specific element $C_0^{-n}A_0C_0^n$, illustrated in Figure 8. This element lies in a copy of F in $2V$ obtained by putting F into $2V$ using only vertical subdivisions.

The element C_0^n has 2^n carets, as seen with an easy induction. Its two trees are balanced trees of depth n, one with only vertical subdivisions, and one with only horizontal subdivisions. By matching the horizontal subdivisions with the vertical ones in A_0, we see the element $C_0^{-n}A_0C_0^n$ has a number of carets of order 2^n, and all the carets are of vertical type, so the element is in F. This element has length in V no more than $2n+1$, but the number of carets is exponential in n and thus its word length as an element of the vertical copy of F in $2V$ is also exponential.

Thus F is at least exponentially distorted in $2V$. Since F is undistorted in T and V, we see that T and V (as subgroups using only vertical subdivisions) are also at least exponentially distorted in $2V$. □

Acknowledgments

The authors are grateful for helpful conversations with Matt Brin, Melanie Stein and Claire Wladis.

References

[Birget 2004] J.-C. Birget, "The groups of Richard Thompson and complexity", *Internat. J. Algebra Comput.* **14**:5-6 (2004), 569–626. MR 2006g:20048 Zbl 1088.20016

[Brin 2004] M. G. Brin, "Higher dimensional Thompson groups", *Geom. Dedicata* **108** (2004), 163–192. MR 2005m:20008 Zbl 1136.20025

[Brin 2005] M. G. Brin, "Presentations of higher dimensional Thompson groups", *J. Algebra* **284**:2 (2005), 520–558. MR 2007e:20062 Zbl 1135.20022

[Burillo et al. 2001] J. Burillo, S. Cleary, and M. Stein, "Metrics and embeddings of generalizations of Thompson's group F", *Trans. Amer. Math. Soc.* **353**:4 (2001), 1677–1689. MR 2001k:20087 Zbl 0989.20030

[Burillo et al. 2009] J. Burillo, S. Cleary, M. Stein, and J. Taback, "Combinatorial and metric properties of Thompson's group T", *Trans. Amer. Math. Soc.* **361**:2 (2009), 631–652. MR 2010e:20061 Zbl 05518616

[Cannon et al. 1996] J. W. Cannon, W. J. Floyd, and W. R. Parry, "Introductory notes on Richard Thompson's groups", *Enseign. Math.* (2) **42**:3-4 (1996), 215–256. MR 98g:20058 Zbl 0880.20027

[Wladis 2007] C. Wladis, *Metric Properties of Thompson's Groups $F(n)$ and $F(n, m)$*, thesis, The City University of New York, 2007.

Received October 21, 2008. Revised October 6, 2009.

JOSÉ BURILLO
DEPARTAMENT DE MATEMÀTICA APLICADA IV
UNIVERSITAT POLITÈCNICA DE CATALUNYA
ESCOLA POLITÈCNICA, SUPERIOR DE CASTELLDEFELS
AVDA. DEL CANAL OLÍMPIC 15
08860 CASTELLDEFELS
SPAIN

burillo@mat.upc.es
http://www-ma4.upc.edu/~burillo/

SEAN CLEARY
DEPARTMENT OF MATHEMATICS
THE CITY COLLEGE OF NEW YORK
CONVENT AVE AT 138TH ST.
NEW YORK, NY 10031
UNITED STATES

cleary@sci.ccny.cuny.edu
http://www.sci.ccny.cuny.edu/~cleary

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

SOLITARY WAVES FOR THE HARTREE EQUATION WITH A SLOWLY VARYING POTENTIAL

KIRIL DATCHEV AND IVAN VENTURA

We study the Hartree equation with a slowly varying smooth potential, $V(x) = W(hx)$, and with an initial condition that is $\varepsilon \le \sqrt{h}$ away in H^1 from a soliton. We show that up to time $|\log h|/h$ and errors of size $\varepsilon + h^2$ in H^1, the solution is a soliton evolving according to the classical dynamics of a natural effective Hamiltonian. This result is based on methods of Holmer and Zworski, who prove a similar theorem for the Gross–Pitaevskii equation, and on spectral estimates for the linearized Hartree operator recently obtained by Lenzmann. We also provide an extension of the result of Holmer and Zworski to more general initial conditions.

1. Introduction

In this paper we study the Hartree equation with an external potential:

$$\begin{aligned} i\partial_t u &= -\tfrac{1}{2}\Delta u + V(x)u - (|x|^{-1} * |u|^2)u, \\ u(x,0) &= u_0(x) \in H^1(\mathbb{R}^3;\mathbb{C}). \end{aligned} \tag{1-1}$$

In the case $V \equiv 0$, solving the associated nonlinear eigenvalue equation,

$$-\tfrac{1}{2}\Delta\eta - (|\eta|^2 * |x|^{-1})\eta = -\lambda\eta, \tag{1-2}$$

gives solutions to (1-1) with evolution $u(t,x) = e^{i\lambda t}\eta(x)$. It is known that (1-2) has a unique radial, positive solution $\eta \in H^1(\mathbb{R}^3)$ for a given $\lambda > 0$; see [Lieb 1977] and [Lenzmann 2009, Appendix A], as well as Appendix A. For convenience of exposition we take λ so that $\|\eta\|_{L^2}^2 = 2$, but this is not essential. Using the symmetries of (1-1), we can construct from this η the following family of *soliton solutions* to (1-1) in the case $V \equiv 0$:

$$u(x,t) = e^{ix\cdot v}e^{i|v|^2t/2}e^{i\gamma}e^{i\lambda t}\mu^2\eta(\mu(x-a-vt)) \qquad \text{for } (a,v,\gamma,\mu) \in \mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}\times\mathbb{R}_+.$$

MSC2000: primary 35Q55; secondary 35Q51.
Keywords: solitons, nonlinear Schrödinger equation, effective dynamics, Hartree equation.
Supported in part by NSF grant DMS-0654436.

If V is not identically zero but is slowly varying, there exist approximate soliton solutions in a sense made precise by the following theorem.

Theorem 1. *Let $V(x) = W(hx)$, where $W \in C^3(\mathbb{R}^3;\mathbb{R})$ is bounded together with all derivatives up to order 3. Fix a constant $0 < c_1$, and fix $(v_0, a_0) \in \mathbb{R}^3 \times \mathbb{R}^3$. Suppose $0 < \delta \le 1/2$, $0 < h \le h_0$, and $u_0 \in H^1(\mathbb{R}^3)$ satisfies*

$$\|u_0 - e^{iv_0\cdot(x-a_0)}\eta(x-a_0)\|_{H^1} \le c_1h^2.$$

Then if $u(t,x)$ solves (1-1) and

$$0 \le t \le \frac{c_1}{h} + \frac{\delta|\log h|}{c_2h},$$

we have

$$\big\|u(t,x) - e^{v(t)\cdot(x-a(t))}e^{i\gamma(t)}\eta((x-a(t)))\big\|_{H^1_x(\mathbb{R}^3)} \le c_2h^{2-\delta}.$$

Here (a, v, γ) solve the system

$$\begin{aligned} \dot a &= v,\\ \dot v &= -\frac{1}{2}\int \nabla V\,(x+a)\,\eta^2(x)dx, \qquad\qquad (1\text{-}3)\\ \dot\gamma &= \tfrac{1}{2}|v|^2 + \lambda - \frac{1}{2}\int V\,(x+a)\,\eta^2(x)dx + \frac{1}{2}\int x\cdot\nabla V\,(x+a)\,\eta^2(x)dx \end{aligned}$$

with initial data $(a_0, v_0, 0)$. The constants h_0 and c_2, depend only on c_1, $|v_0|$, and $\|W\|_{C^3(\mathbb{R}^3)}$. They are in particular independent of δ.

Note that in (1-3), the equation of motion of the center of mass a of the soliton is given by Newton's equation, $\ddot a = -\nabla\overline{V}(a)$, where $\overline{V} := V * \eta^2/2$. Observe also that because η is exponentially localized (see Appendix A), $\eta^2/2$ is an approximation of a delta function and hence the effective potential $\overline{V}$ that governs the motion of the soliton is an approximation of V. The evolution of γ is more complicated and explained by the Hamiltonian formulation of the problem developed in Section 2.

Our next theorem gives a slightly weaker result in the case of a more general initial condition.

Theorem 2. *Let $V(x) = W(hx)$, where $W \in C^3(\mathbb{R}^3;\mathbb{R})$ is bounded together with all of its derivatives up to order 3. Fix constants $0 < c_1$, and $0 \le 2\delta \le \delta_0 < 3/4$, and fix $(v_0, a_0) \in \mathbb{R}^3 \times \mathbb{R}^3$. Suppose $0 < h \le h_0$, and $u_0 \in H^1(\mathbb{R}^3)$ satisfies*

$$\|u_0 - e^{iv_0\cdot(x-a_0)}\eta(x-a_0)\|_{H^1} =: \varepsilon \le c_1h^{1/2+\delta_0}.$$

Then for

$$0 \le t \le \frac{c_1}{h} + \frac{\delta|\log h|}{c_2h},$$

we have

$$\left\| u(t,x) - e^{v(t)\cdot(x-a(t))} e^{i\gamma(t)} \mu(t)^2 \eta(\mu(t)(x-a(t))) \right\|_{H^1_x(\mathbb{R}^3)} \le c_2 h^{-\delta} \tilde{\varepsilon},$$

where $\tilde{\varepsilon} := \varepsilon + h^2$. *Here* (a, v, μ, γ) *solve the system*

$$\begin{aligned}
\dot{a} &= v + \mathbb{O}(\tilde{\varepsilon}^2),\\
\dot{v} &= -\frac{\mu}{2}\int \nabla V(x/\mu + a)\eta^2(x)dx + \mathbb{O}(\tilde{\varepsilon}^2),\\
\dot{\mu} &= \mathbb{O}(\tilde{\varepsilon}^2),\\
\dot{\gamma} &= \tfrac{1}{2}|v|^2 + \lambda\mu^2 - \frac{1}{2}\int V(x/\mu + a)\eta^2(x)dx\\
&\qquad\qquad - \frac{1}{2\mu}\int x\cdot\nabla V(x/\mu + a)\eta^2(x)dx + \mathbb{O}(\tilde{\varepsilon}^2),
\end{aligned}$$

with initial data $(a_0, v_0, 1, 0)$. *The constants* h_0 *and* c_2, *as well as the implicit constants in the* $\mathbb{O}$ *error terms, depend only on* c_1, $|v_0|$, *and* $\|W\|_{C^3(\mathbb{R}^3)}$. *They are in particular independent of* δ.

This phenomenon was studied in the physics literature by Éboli and Marques [1983], who show for various (not necessarily slowly varying) potentials V that soliton solutions obeying Newtonian equations of motion exist. Later Bronski and Jerrard [2000] proved a similar theorem in the case of a power nonlinearity, and then more general nonlinearities were treated by Fröhlich, Tsai, and Yau [2002] and by Fröhlich, Gustafson, Jonsson, and Sigal [2004]. More recently Jonsson, Fröhlich, Gustafson, and Sigal [2006] have extended the validity of the effective dynamics to longer time in the case of a confining potential V, and Abou Salem [2008] has treated the case of a potential V that is permitted to vary in time. The case of the cubic nonlinear Schrödinger equation in dimension one was also studied by Holmer and Zworski [2007; 2008]. Other papers have established effective classical dynamics in quantum equations of motion in a wide variety of settings: see [Fröhlich, Gustafson, Jonsson and Sigal 2004] and [Abou Salem 2008] for many references.

Our result improves the results of [Fröhlich, Tsai and Yau 2002; Fröhlich, Gustafson, Jonsson and Sigal 2004] and [Abou Salem 2008] in the case of (1-1) in several respects. First, we provide a more precise error bound, improving $\tilde{\varepsilon}$ from $h + \varepsilon$ to $h^2 + \varepsilon$. Second, we remove the errors in the equations of motion when $\varepsilon = \mathbb{O}(h^{2-\delta})$. Finally, we establish the effective dynamics for longer time: The result in the first two papers was valid only up to time $c(\varepsilon^2 + h)^{-1}$ for a small constant c, while in the third the result was valid only up to time $\delta|\log h|/h$ and required the assumption $\varepsilon = \mathbb{O}(h)$.

Fröhlich, Gustafson, Jonsson, and Sigal [2004] consider more general initial data: ε is assumed to be small but not necessarily $\mathbb{O}(h^{1/2+})$, although in this case the result is obtained only for time ε^{-2}. In that situation the methods of this paper, although applicable, do not improve that result, so for ease of exposition we consider only the special case $\varepsilon = \mathbb{O}(h^{1/2+})$ where we have an improvement.

In this paper we follow most closely [Holmer and Zworski 2008], which in turn builds on [Holmer and Zworski 2007] and on earlier work on soliton stability going back to Weinstein [1986]. We adapt those arguments to a higher-dimensional setting where in particular there is no longer an explicit form for η, and to the nonlocal Hartree nonlinearity. For this last task we make use of the classical Hardy–Littlewood–Sobolev inequality and of Lenzmann's [2009] spectral estimates for the linearized Hartree operator

$$\mathscr{L}w := -\tfrac{1}{2}\Delta w - (|x|^{-1} * \eta(w+\overline{w}))\eta - (|x|^{-1} * \eta^2)w + \lambda w.$$

In Section 4, we also extend the methods of [Holmer and Zworski 2008] by adapting them to more general initial data. It is at this point that our proofs depart most significantly from theirs. The crucial addition is a closer analysis of the differential equation for the error studied in Lemmas 4.3 and 4.4. This analysis applies also to the Gross–Pitaevskii equation studied in [Holmer and Zworski 2008], giving us Theorem 3.

To state this theorem, we suppose $u\colon \mathbb{R}\times\mathbb{R} \to \mathbb{C}$ solves

$$\begin{aligned} i\partial_t u &= -\tfrac{1}{2}\partial_x^2 u + V(x)u - |u|^2 u,\\ u(x,0) &= u_0(x) \in H^1(\mathbb{R};\mathbb{C}). \end{aligned} \tag{1-4}$$

In this case the ground state soliton solution of the corresponding elliptic nonlinear eigenvalue equation

$$-\tfrac{1}{2}\eta = -\tfrac{1}{2}\eta'' - \eta^3$$

is given by

$$\eta(x) = \operatorname{sech}(x).$$

Theorem 3. *Let* $V(x) = W(hx)$, *where* $W \in C^3(\mathbb{R};\mathbb{R})$ *is bounded together with all derivatives up to order* 3. *Fix constants* $0 < c_1$ *and* $0 < \delta_0 < 3/4$ *and fix* $(v_0, a_0) \in \mathbb{R}\times\mathbb{R}$. *Suppose* $0 \le 2\delta \le \delta_0$ *and* $0 < h \le h_0$. *For* $u_0 \in H^1(\mathbb{R})$, *put*

$$\big\|u_0 - e^{iv_0\cdot(x-a_0)}\operatorname{sech}(x-a_0)\big\|_{H^1} := \varepsilon \le c_1 h^{1/2+\delta_0}.$$

Then for

$$0 \le t \le \frac{c_1}{h} + \frac{\delta|\log h|}{c_2 h},$$

we have

$$\big\|u(t,x) - e^{v(t)\cdot(x-a(t))}e^{i\gamma(t)}\mu(t)\operatorname{sech}(\mu(t)(x-a(t)))\big\|_{H^1_x(\mathbb{R}^3)} \le c_2 h^{-\delta}\tilde{\varepsilon},$$

where u solves (1-4) *and* $\tilde{\varepsilon} := \varepsilon + h^2$. *Here* (a, v, μ, γ) *solve the system*

$$
\begin{aligned}
\dot{a} &= v + \mathbb{O}(\tilde{\varepsilon}^2),\\
\dot{v} &= -\frac{\mu^2}{2}\int V'(x+a)\,\mathrm{sech}^2(\mu x)\,dx + \mathbb{O}(\tilde{\varepsilon}^2),\\
\dot{\mu} &= \mathbb{O}(\tilde{\varepsilon}^2),\\
\dot{\gamma} &= \tfrac{1}{2}\mu^2 + \tfrac{1}{2}v^2 - \mu\int V(x+a)\,\mathrm{sech}^2(\mu x)\,dx\\
&\qquad\qquad + \mu^2\int xV(x+a)\,\mathrm{sech}^2(\mu x)\tanh(\mu x)\,dx + \mathbb{O}(\tilde{\varepsilon}^2)
\end{aligned}
$$

with initial data $(a_0, v_0, 1, 0)$. *The constants h_0 and c_2, as well as the implicit constants in the $\mathbb{O}$ error terms, depend only on c_1, δ_0, $|v_0|$, and $\|W\|_{C^3(\mathbb{R}^3)}$. They are in particular independent of δ.*

This is proved by replacing [Holmer and Zworski 2008, Lemmas 5.1 and 5.2] with our Lemmas 4.3 and 4.4. Because the details are very similar to the ones given in Section 4, we omit them.

The methods of this paper can be extended to more general nonlinearities under additional spectral nondegeneracy assumptions: see [Fröhlich, Gustafson, Jonsson and Sigal 2004] for examples. That paper, and also [Fröhlich, Tsai and Yau 2002], considers more general classes of equations under such assumptions. Here we restrict our attention to two physical nonlinearities for which the necessary spectral results are known.

The outline of the proof and of this paper are as follows.

In Section 2, we recast (1-1) as a Hamiltonian evolution equation in $H^1(\mathbb{R}^3)$, with the Hamiltonian given by (2-14). We define the manifold of solitons to be the set of functions of the form $e^{v\cdot(x-a)}e^{i\gamma}\mu^2\eta(\mu(x-a))$ for some (a, v, γ, μ) in $\mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}\times\mathbb{R}^+$, and we show that the equations (1-3) come from the restriction of the Hamiltonian (2-14) to this manifold.

In Section 3, we review and extend slightly the relevant spectral results from [Lenzmann 2009].

In Section 4, we compute the differential equation for the difference between the true solution u and the "closest point" on the manifold of solitons. We then estimate this difference, proving Theorem 2.

In Section 5, we show how the additional assumption on the initial condition in Theorem 1 gives the exact equations of motion (1-3).

In Appendix A we collect the properties of η that we need for our proofs, and in Appendix B we review a standard proof of the global well-posedness of (1-1).

2. Hamiltonian equations of motion

This section has four subsections. In the first, we define a symplectic structure on H^1 and recall a few basic lemmas from symplectic geometry. In the second, we define the manifold of solitons, which has a natural action on it by the group of symmetries of (1-1). We compute the Lie algebra associated to this group of symmetries and from that deduce a formula for the derivative of a curve in the group in terms of the Lie algebra. In the third, we prove that the manifold of solitons is a symplectic submanifold and compute the restriction of the symplectic form to it. In the fourth, we compute the Hartree Hamiltonian and its restriction to the manifold of solitons, and derive the equations (1-3) as the equations of motion associated to the restricted Hamiltonian. Most of the ideas in this section are present in [Holmer and Zworski 2007, Section 2].

Symplectic structure. We work over the vector space

$$\mathcal{V} := H^1(\mathbb{R}, \mathbb{C}) \subset L^2(\mathbb{R}, \mathbb{C}),$$

viewed as a *real* Hilbert space. The inner product and the symplectic form are given by

$$\langle u, v\rangle := \operatorname{Re} \int u\bar{v} \quad \text{and} \quad \omega(u, v) := \operatorname{Im} \int u\bar{v}. \tag{2-1}$$

Let $H : \mathcal{V} \to \mathbb{R}$ be a function, a Hamiltonian. The associated Hamiltonian vector field is a map $\Xi_H : \mathcal{V} \to T\mathcal{V}$. The vector field Ξ_H is defined by the relation

$$\omega(v, (\Xi_H)_u) = d_u H(v), \tag{2-2}$$

where $v \in T_u\mathcal{V}$, and $d_u H : T_u\mathcal{V} \to \mathbb{R}$ is defined by

$$d_u H(v) = \frac{d}{ds}\Big|_{s=0} H(u + sv).$$

In the notation above, we have

$$d_u H(v) = \langle dH_u, v\rangle \quad \text{and} \quad (\Xi_H)_u = -i\,dH_u, \tag{2-3}$$

where the first equation provides a definition of dH_u, and the second a formula for computing Ξ_H.

For reference we present two simple lemmas from symplectic geometry. The proofs can be found in [Holmer and Zworski 2007, Section 2].

Lemma 2.1. *Suppose that $g : \mathcal{V} \to \mathcal{V}$ is a diffeomorphism such that $g^*\omega = \mu(g)\omega$, where $\mu(g) \in C^\infty(\mathcal{V}, \mathbb{R})$. Then for $f \in C^\infty(\mathcal{V}, \mathbb{R})$*

$$(g^{-1})_*\big((\Xi_f)_{g(\rho)}\big) = \frac{1}{\mu(g)}\,\Xi_{g^* f}(\rho) \quad \textit{for } \rho \in \mathcal{V}. \tag{2-4}$$

Suppose that $f \in C^\infty(\mathcal{V}, \mathbb{R})$ and that $df(\rho_0) = 0$. Then the Hessian of f at ρ_0, $f''(\rho_0) : T_\rho\mathcal{V} \to T^*_\rho\mathcal{V}$, is well-defined. We can identify $T_\rho\mathcal{V}$ with $T^*_\rho\mathcal{V}$ using the inner product, and define the Hamiltonian map $F : T_\rho\mathcal{V} \to T_\rho\mathcal{V}$ by

$$F = -if''(\rho_0) \quad \text{and} \quad \langle f''(\rho_0)X, Y\rangle = \omega(Y, FX). \tag{2-5}$$

Lemma 2.2. *Suppose that N is a finite-dimensional symplectic submanifold of $\mathcal{V}$ and $f \in C^\infty(V, \mathbb{R})$ satisfies*

$$\Xi_f(\rho) \in T_\rho N \subset T_\rho\mathcal{V} \quad \textit{for } \rho \in N.$$

If $df(\rho_0) = 0$ at $\rho_0 \in N$, then the Hamiltonian map defined by (2-5) *satisfies*

$$F(T_\rho N) \subset T_\rho N.$$

Manifold of solitons as group orbit. For $g = (a, v, \gamma, \mu) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R} \times \mathbb{R}_+$, we define the map

$$H^1 \ni u \mapsto g \cdot u \in H^1, \quad (g \cdot u)(x) := e^{i\gamma} e^{iv\cdot(x-a)} \mu^2 u(\mu(x-a)). \tag{2-6}$$

This action gives the group structure

$$(a, v, \gamma, \mu) \cdot (a', v', \gamma', \mu') = (a'', v'', \gamma'', \mu'')$$

on $\mathbb{R}^7 \times \mathbb{R}_+$, where

$$v'' = v + \mu v', \quad a'' = a + a'/\mu, \quad \gamma'' = \gamma + \gamma' + va'/\mu, \quad \mu'' = \mu\mu'.$$

The action of G is conformally symplectic in that

$$g^*\omega = \mu\omega \quad \text{and} \quad g = (a, v, \gamma, \mu), \tag{2-7}$$

as is easily seen from (2-1).

The Lie algebra of G, denoted $\mathfrak{g}$, is generated by the eight elements

$$\begin{aligned} &e_1 = -\partial_{x_1}, \quad e_4 = ix_1 \quad\ e_7 = i, \\ &e_2 = -\partial_{x_2}, \quad e_5 = ix_2, \quad e_8 = 2 + x \cdot \nabla, \\ &e_3 = -\partial_{x_3}, \quad e_6 = ix_3. \end{aligned} \tag{2-8}$$

These are simply the partial derivatives at the identity of $(g \cdot u)(x)$ with respect to each of the eight parameters (a, v, γ, μ). The following computation gives the derivative of a curve in G in terms of this basis.

Lemma 2.3. *Let $g \in C^1(\mathbb{R}, G)$ and $u \in \mathcal{S}(\mathbb{R})$. Then, in the notation of* (2-6),

$$\frac{d}{dt} g(t) \cdot u = g(t) \cdot (Y(t)u),$$

where $Y(t) \in \mathfrak{g}$ is given by

$$Y(t) = \mu(t) \sum_{j=1}^{3} \dot{a}_j(t) e_j + \mu(t) \sum_{j=1}^{3} \frac{\dot{v}_j(t)}{\mu(t)} e_{3+j} + (\dot{\gamma}(t) - \dot{a}(t) \cdot v(t)) e_7 + \frac{\dot{\mu}(t)}{\mu(t)} e_8, \tag{2-9}$$

where

$$\begin{aligned} g(t) &= (a(t), v(t), \gamma(t), \mu(t)) \\ &= (a_1(t), a_2(t), a_3(t), v_1(t), v_2(t), v_3(t), \gamma(t), \mu(t)). \end{aligned}$$

We define the submanifold $M \subset H^1$ of solitons as the orbit of η under G, where η is the function described in Appendix A:

$$M = G \cdot \eta \simeq G/\mathbb{Z} \quad \text{and} \quad T_\eta M = \mathfrak{g} \cdot \eta \simeq \mathfrak{g}. \tag{2-10}$$

The quotient corresponds to the $\mathbb{Z}$-action

$$(a, v, \gamma, \mu) \mapsto (a, v, \gamma + 2\pi k, \mu) \quad \text{for } k \in \mathbb{Z}.$$

The following is a simple consequence of the implicit function theorem and of the nondegeneracy of ω. The proof can be found, for example, in [Holmer and Zworski 2007, Lemma 3.1].

Lemma 2.4. *For Σ and compact subset of $G/\mathbb{Z}$, let*

$$U_{\Sigma,\delta} = \{u \in H^1 : \inf_{g \in \Sigma} \|u - g \cdot \eta\|_{H^1} < \delta\}.$$

If $\delta \le \delta_0 = \delta_0(\Sigma)$, then for any $u \in U_{\Sigma,\delta}$, there exists a unique $g(u) \in \Sigma$ such that

$$\omega(g(u)^{-1} \cdot u - \eta, X \cdot \eta) = 0 \quad \textit{for all } X \in \mathfrak{g}.$$

Moreover, the map $u \mapsto g(u)$ is in $C^1(U_{\Sigma,\delta}, \Sigma)$.

Symplectic structure on the manifold of solitons. We compute the symplectic form $\omega|_M$ on $T_\eta M$ by using

$$(\omega|_M)_\eta(e_i, e_j) = \operatorname{Im} \int (e_i \cdot \eta)(x) (\overline{e_j \cdot \eta})(x).$$

We remind the reader (as mentioned in Appendix A) that $\|\eta\|_{L^2}^2 = 2$. Using (2-8) we compute all these forms.

Lemma 2.5. *The evaluation at η of the restriction of the symplectic form to M is given by*

$$(\omega|_M)_\eta = (dv \wedge da + d\gamma \wedge d\mu)_{(0,0,0,1)} = (d(v da + \gamma d\mu))_{(0,0,0,1)}.$$

Proof. If j, k are both taken from $\{1, 2, 3, 8\}$ or both taken from $\{4, 5, 6, 7\}$, then the integrand $(e_j \cdot \eta)(x)(\overline{e_k \cdot \eta})(x)$ is a real function, implying that $(\omega|_M)_\eta(e_j, e_k) = 0$.

If $j \in \{1, 2, 3\}$ and $k \in \{4, 5, 6\}$, we have $e_j = -\partial_j$ and $e_k = ix_{k-3}$.

If $j \neq k-3$, integrating by parts gives

$$\begin{aligned}(\omega|_M)_\eta(e_j, e_k) &= \operatorname{Im} \int (e_j \cdot \eta)(x)(\overline{e_k \cdot \eta})(x) \\ &= \operatorname{Im} \int (-\partial_j \eta)(\overline{i x_{k-3} \eta}) = -\int (\eta)(x_{k-3} \partial_j \eta).\end{aligned}$$

This implies that $(\omega|_M)_\eta(e_j, e_k) = 0$.

If $j = k-3$, integrating by parts gives

$$(\omega|_M)_\eta(e_j, e_k) = \operatorname{Im} \int (e_j \cdot \eta)(x)(\overline{e_k \cdot \eta})(x) = \int (\partial_j \eta)(x_j \eta) = -\int (\eta(\eta + x_j \partial_j \eta)).$$

Solving this yields $(\omega|_M)_\eta(e_j, e_k) = -1$.

If $j \in \{1, 2, 3\}$ and $k = 7$, integrating by parts gives

$$\begin{aligned}(\omega|_M)_\eta(e_j, e_k) &= \operatorname{Im} \int (e_j \cdot \eta)(x)(\overline{e_k \cdot \eta})(x) \\ &= \operatorname{Im} \int (-\partial_j \eta)(\overline{i\eta}) = \int (\partial_j \eta)(\eta) = -\int (\eta)(\partial_j \eta),\end{aligned}$$

implying $(\omega|_M)_\eta(e_j, e_k) = 0$.

If $j \in \{4, 5, 6\}$ and $k = 8$, we get

$$\begin{aligned}(\omega|_M)_\eta(e_j, e_k) &= \operatorname{Im} \int (e_j \cdot \eta)(x)(\overline{e_k \cdot \eta})(x) = \operatorname{Im} \int i x_j \eta (2 + x \cdot \nabla)\eta \\ &= 2 \int x_j \eta^2 + \int x_j \eta x \cdot \nabla \eta \\ &= 2 \int x_j \eta^2 + \int x_j \eta (x_1 \partial_1 \eta + x_2 \partial_2 \eta + x_3 \partial_3 \eta).\end{aligned}$$

Now $\int x_j \eta^2$ is zero since it is odd in the x_j variable. Since all the terms in this last expression can be reduced to this by integrating by parts, we see that $(\omega|_M)_\eta\, (e_j, e_k) = 0$.

If $j = 7$ and $k = 8$, we observe that since by integration by parts we have $\int \eta x \cdot \nabla \eta = -\frac{3}{2} \|\eta\|_{L^2}^2$, we also have

$$(\omega|_M)_\eta(e_j, e_k) = \operatorname{Im} \int (e_j \cdot \eta)(x)(\overline{e_k \cdot \eta})(x) = \int \eta (2 + x \cdot \nabla)\eta = 2\|\eta\|_{L^2}^2 - \tfrac{3}{2} \|\eta\|_{L^2}^2,$$

giving $(\omega|_M)_\eta(e_j, e_k) = 1$.

Putting all this together gives the result. □

We now observe from (2-10) and (2-7) that

$$\omega|_M = \mu dv \wedge da + v d\mu \wedge da + d\gamma \wedge d\mu. \tag{2-11}$$

Now let f be a function defined on M, that is, $f = f(a, v, \gamma, \mu)$. The associated Hamiltonian vector field Ξ_f is given by

$$\omega(\cdot, \Xi_f) = df = f_a da + f_v dv + f_\mu d\mu + f_\gamma d\gamma.$$

Using (2-11), we obtain

$$\Xi_f = \frac{1}{\mu}\nabla_v f \cdot \nabla_a + \frac{1}{\mu}\big(-\nabla_a f - (\partial_\gamma f)v\big)\cdot \nabla_v + \frac{\partial}{\partial\gamma} f \partial_\mu + \Big(\frac{1}{\mu} v \cdot \nabla_v f - -\partial_\mu f\Big)\partial_\gamma. \tag{2-12}$$

The Hamiltonian flow is obtained by solving

$$\dot v = -\nabla_a f - (\partial_\gamma f)v, \quad \dot a = \frac{1}{\mu}\nabla_v f, \quad \dot\mu = \partial_\gamma f, \quad \dot\gamma = \frac{1}{\mu} v \cdot \nabla_v f - \partial_\mu f.$$

The Hartree Hamiltonian restricted to the manifold of solitons. Using the symplectic form given in (2-1), and

$$H(u) := \int \tfrac{1}{4}|\nabla u|^2 - \tfrac{1}{4}|u|^2(|u|^2 * |x|^{-1}),$$

we find that

$$d_u H(v) = \operatorname{Re}\int \big(-\tfrac{1}{2}\Delta u - (|u|^2 * |x|^{-1})u\big)\bar v.$$

The Hamiltonian flow associated to this vector field is

$$\dot u = (\Xi_H)_u = -i\big(-\tfrac{1}{2}\Delta u - (|u|^2 * |x|^{-1})u\big). \tag{2-13}$$

The restriction of

$$H(u) = \int \tfrac{1}{4}|\nabla u|^2 - \tfrac{1}{4}|u|^2(|u|^2 * |x|^{-1}),$$

to M is given by computing

$$H(g\cdot\eta) = \tfrac{1}{4}|v|^2\mu\|\eta\|_{L^2}^2 + \mu^3 H(\eta) = \tfrac{1}{2}|v|^2\mu + \mu^3 H(\eta) \quad \text{for } g = (a, v, \gamma, \mu).$$

The flow of (2-12) for this f describes the evolution of a soliton. We have in particular $\dot\gamma = \frac{1}{2}|v|^2 - 3\mu^2 H(\eta)$, and because we know that $e^{i\lambda t}\eta(x)$ solves (1-1), we can compute that $H(\eta) = -\lambda/3$.

We now consider the Hartree Hamiltonian,

$$H_V(u) = \frac{1}{4}\int |\nabla u|^2 - \frac{1}{4}\int |u|^2(|u|^2 * |x|^{-1}) + \frac{1}{2}\int V(x)|u|^2, \tag{2-14}$$

and its restriction to $M = G \cdot \eta$ given by

$$H_V|_M = \tfrac{1}{2}|v|^2\mu + \lambda\tfrac{1}{3}\mu^3 + \tfrac{1}{2}\mu^4 \int V(x)\eta^2(\mu(x-a)). \tag{2-15}$$

The flow of $H_V|_M$ can be read off from (2-12):

$$\dot{v} = -\tfrac{1}{2}\mu \int \nabla V(x/\mu + a)\eta^2(x)dx, \qquad \dot{a} = v, \qquad \dot{\mu} = 0,$$
$$\dot{\gamma} = \tfrac{1}{2}|v|^2 + \lambda\mu^2 - \frac{1}{2}\int V(x/\mu+a)\eta^2(x)dx + \frac{1}{2\mu}\int x\cdot\nabla V(x/\mu+a)\eta^2(x)dx.$$

These are the same as the ones given in (1-3). The evolution of a and v is simply the Hamiltonian evolution of $\frac{1}{2}|v|^2 + \frac{1}{2}\mu^3 \int \nabla V(x+a)\eta^2(\mu x)$ when μ is held constant. As a result the evolution of the phase is explained by (2-15).

Finally we give an important application of Lemma 2.2. We put

$$H_\lambda(u) = \int \tfrac{1}{4}|\nabla u|^2 - \tfrac{1}{4}|u|^2(|u|^2 * |x|^{-1}) + \tfrac{1}{2}\lambda \int |u|^2,$$

and observe that η is a critical point of this functional, while the Hessian of H_λ at η is given by

$$\mathcal{L}w := -\tfrac{1}{2}\Delta u - (|x|^{-1} * \eta(w+\overline{w}))\eta - (|x|^{-1} * \eta^2)w + \lambda w. \tag{2-16}$$

Now in Lemma 2.2 take H_λ to be f, take N to be the eight-dimensional manifold of solitons M, and take $\rho = \eta$. We find that

$$i\mathcal{L}(T_\eta M) \subset T_\eta M. \tag{2-17}$$

3. Spectral estimates

In this section we recall crucial spectral estimates for the operator $\mathcal{L}$ from (2-16), which is the linearization of $-\frac{1}{2}\Delta u - (|u|^2 * |x|^{-1})u + \lambda u$. We observe that this operator can be decomposed as

$$\mathcal{L}w = \begin{bmatrix} L_+ & 0 \\ 0 & L_- \end{bmatrix} \begin{bmatrix} \operatorname{Re} w \\ \operatorname{Im} w \end{bmatrix},$$

with

$$L_+ \operatorname{Re} w = -\tfrac{1}{2}\Delta \operatorname{Re} w - 2(|x|^{-1} * \eta \operatorname{Re} w)\eta - (|x|^{-1} * \eta^2)\operatorname{Re} w + \lambda \operatorname{Re} w,$$
$$L_- \operatorname{Im} w = -\tfrac{1}{2}\Delta \operatorname{Im} w - (|x|^{-1} * \eta^2)\operatorname{Im} w + \lambda \operatorname{Im} w.$$

From the second remark following [Lenzmann 2009, Theorem 4] we have the following proposition:

Proposition 3.1. *Let $w \in H^1(\mathbb{R}, \mathbb{C})$ and suppose that $\omega(w, X\eta) = 0$ for any $X \in \mathfrak{g}$. Then*

$$\langle \mathscr{L}w, w\rangle \geq c\|w\|_{H^1}^2, \tag{3-1}$$

where c is an absolute constant.

Now we consider solutions f of the equation

$$L_+ f = Q(x)\eta(x), \tag{3-2}$$

where $Q(x)$ is real-valued and of the form $Q(x) = a_0(t) + \sum a_{ij}(t)x_i x_j$, with $Q(x)\eta$ symplectically orthogonal to the generalized kernel of $i\mathscr{L}$, and with $a_{ij}(t)$ bounded in t.

Proposition 3.2. *Equation* (3-2) *has a unique solution in* $(\ker(L^+))^\perp \subset L^2(\mathbb{R}^3)$. *This solution is also in $C^\infty(\mathbb{R}^3)$ with the property*

$$e^{(\sqrt{2\lambda}-\epsilon)|x|/2}\partial^\alpha f \in L^\infty(\mathbb{R}^3) \tag{3-3}$$

for all $\epsilon > 0$ and for any multiindex $\alpha \in \mathbb{N}^3$. Furthermore

$$\omega(f, X\eta) = 0 \quad \text{for all } X \in \mathfrak{g}. \tag{3-4}$$

Proof. We first use $Q(x)\eta \in (\ker L_+)^\perp$ to show that a unique solution exists. Indeed, it is suffices to show this result for any $Q_{ij}(x) = x_i x_j$ or $Q_0 = 1$. By [Lenzmann 2009, Theorem 4], we know that $\ker L_+ = \operatorname{span}\{\partial_1\eta, \partial_2\eta, \partial_3\eta\}$. Clearly $\langle \partial_j\eta, \eta\rangle = 0$ for all $j \in \{1, 2, 3\}$. It remains only to show for all $i, j, k \in \{1, 2, 3\}$ that

$$\langle -\partial_i\eta, x_j x_k\eta\rangle = 0. \tag{3-5}$$

If $i \neq j$ and $i \neq k$, then (3-5) is clear because the integrand is odd in the x_i direction. So we assume $i = j$. If $j \neq k$, then

$$\langle -\partial_i\eta, x_i x_k\eta\rangle = -\int \partial_i\eta(x_i x_k)\eta = \int x_k\eta^2 + \int \partial_i\eta(x_i x_k)\eta.$$

But $x_k\eta^2$ is odd in the x_k direction, leading to (3-5). A similar argument gives (3-5) for $j = k$.

It follows from the PDE solved by f that if $f \in H^s(\mathbb{R}^3)$ then $f \in H^{s+2}(\mathbb{R}^3)$, implying that $f \in C^\infty(\mathbb{R}^3)$. The proof of (3-3) now follows closely the proof of Proposition A.2, and we give it only in outline. We put $w = e^\phi f$ and introduce

$$L_+^\phi w := e^\phi L_+ e^{-\phi} w = (P_\phi + \lambda)w - 2e^\phi\eta(|x|^{-1} * (\eta e^{-\phi}w)).$$

We now have

$$\langle L_+^\phi w, w\rangle = \frac{1}{2}\int |\nabla w|^2 + \int (\widetilde{V} - \tfrac{1}{2}|\nabla\phi|^2 + \lambda)w^2 \\ - 2\int e^\phi \eta(|x|^{-1} * (\eta f))w + \int e^\phi Q(x)\eta w.$$

Then

$$\varepsilon \int w^2 \le \int (\lambda - \tfrac{1}{2}|\nabla\phi|^2)w^2 \\ \le -\int \widetilde{V} w^2 - 2\int e^\phi \eta(|x|^{-1} * (\eta f))w + \int e^\phi P(x)\eta w.$$

The $\widetilde{V}$ term is handled as before. The two e^ϕ factors in the last term can be absorbed by the η factor provided the exponential growth in ϕ is no more than $\exp((\sqrt{2\lambda} - \varepsilon|x|)/2)$. For the middle term, observe that, as in the case of $\widetilde{V}$, the convolution $|x|^{-1} * (\eta f)$ is continuous and decaying to zero at infinity. Then, the two e^ϕ factors can be absorbed by the η factor just as in the case of the last term. In this way we show that

$$\int w^2 \le C,$$

and proceed as in the proof of Proposition A.2.

We now prove (3-4). First of all, since f is real, $\omega(f, e_j\eta) = \operatorname{Im}\int f e_j \eta = 0$ for $j \in \{1, 2, 3, 8\}$ since then $e_j\eta$ is real. Next write

$$f = f_0 + \sum_{j,k=1}^{3} f_{jk}, \quad \text{where } L_+ f = a_0 \text{ and } L_+ f_{jk} = a_{jk} x_j x_k.$$

Since L_+ preserves symmetry in x_k for all k, we observe that if $j \in \{4, 5, 6\}$, then

$$\omega(f_{k\ell}, e_j\eta) = \int f_{k\ell} x_{j-1}\eta = 0,$$

as the integrand will be odd in some x_i direction. Finally a calculation shows that $L_+((2 + x\cdot\nabla)\eta) = \eta$, from which it follows that

$$\omega(f, e_7\eta) = \int f\eta = \int L_+(f)(2 + x\cdot\nabla)\eta = \int (Q(x)\eta)(2 + x\cdot\nabla)\eta = 0. \quad \square$$

4. Reparametrized evolution and proof of Theorem 2

We write

$$u(t) = g(t)\cdot(\eta + w(t)) \quad \text{and} \quad \omega(w(t), X\eta) = 0 \quad \text{for all } X \in \mathfrak{g}.$$

To see that this decomposition is possible, initially for small times, we apply Lemma 2.4, which allows us to define

$$g(t) := g(u(t)), \quad \tilde{u} := g(t)^{-1}u(t), \quad w(t) := \tilde{u} - \eta,$$

and derive an equation for $w(t)$. Before doing so, however, we introduce some abbreviated notation. For $g(t)$, we write $g = (a, v, \gamma, \mu)$, and observe that as a result of $\mathrm{Re}\langle w, \eta\rangle = 0$ and the L^2 conservation of the original equation, we have

$$2 + \|w\|_{L^2}^2 = \|\eta + w\|_{L^2}^2 = \|g^{-1}u\|_{L^2}^2 = \mu^{-1}\|u_0\|_{L^2}^2,$$

and hence

$$\frac{2-\varepsilon}{2+\|w\|_{L^2}^2} \le \mu \le \frac{2+\varepsilon}{2+\|w\|_{L^2}^2}, \tag{4-1}$$

with ε as in the statement of Theorem 2. This gives a precise sense in which $\mu \approx 1$. For the remainder of the section we will assume $0 \le \varepsilon \le 1$, although in our theorems ε is required to be much smaller than 1.

Next we define

$$\begin{aligned}
\alpha = \alpha(a, \mu) &:= \frac{1}{2}\int V(x/\mu + a)\eta^2(x)dx - \frac{1}{2\mu}\int x \cdot \nabla V(x/\mu + a)\eta^2(x)dx, \\
\beta = \beta(a, \mu) &:= \frac{1}{2\mu}\int \nabla V(x/\mu + a)\eta^2(x)dx, \\
X = \mu \sum_{j=1}^{3}(-\dot{a}_j + v_j)e_j &+ \sum_{j=1}^{3}(\dot{v}_j/\mu - \beta_j)e_{j+3} \\
&\qquad + (-\dot{\gamma} + \dot{a}\cdot v - \tfrac{1}{2}|v|^2 + \lambda\mu^2 - \alpha)e_7 - (\dot{\mu}/\mu)e_8.
\end{aligned}$$

Observe that $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^3$, and $X \in \mathfrak{g}$. Set further

$$\begin{aligned}
\mathcal{L}w &:= -\tfrac{1}{2}\Delta w - (|x|^{-1} * \eta^2)w - (|x|^{-1} * (\eta(w + \bar{w})))\eta + \lambda w, \\
\mathcal{N}w &:= (|x|^{-1} * |w|^2)\eta + (|x|^{-1} * \eta(w + \bar{w}))w + (|x|^{-1} * |w|^2)w.
\end{aligned}$$

These terms come from writing out $i\,\Xi_H(\eta + w)$. The operator $\mathcal{L}$ collects the linear terms, and $\mathcal{N}$ the nonlinear terms.

Lemma 4.1. *In the notation above, the equation for w is*

$$\begin{aligned}
\partial_t w = &\quad X\eta + i(-V(x/\mu + a) + \alpha + \beta \cdot x)\eta \\
&+ Xw + i(-V(x/\mu + a) + \alpha + \beta \cdot x)w + i\mu^2(-\mathcal{L} + \mathcal{N})w.
\end{aligned}$$

Proof. The proof is a straightforward calculation that follows nearly the same lines as that of [Holmer and Zworski 2008, Lemma 3.2], and here we give only a sketch.

We first use the definition of w and the chain rule to write

$$\partial_t w = -Y(\eta + w) + g^{-1}\Xi_H g(\eta + w),$$

with Y taken from Lemma 2.3. We use Lemma 2.1 to write $g^{-1}\Xi_H g = \mu^{-1}\Xi_{g^*H}$, and compute Ξ_{g^*H} from formula (2-3). Finally, using the soliton equation

$$-\lambda\eta + \tfrac{1}{2}\Delta\eta + (|x|^{-1} * \eta^2)\eta = 0$$

gives the desired formula. □

We now explain the reasons for this notation. Note that if $X = 0$, then

$$\dot{a} = \dot{v}, \qquad \dot{v} = -\mu\beta, \qquad \dot{\gamma} = \tfrac{1}{2}|v|^2 + \lambda\mu^2 - \alpha, \qquad \dot{\mu} = 0.$$

giving the equations of motion in (1-3). In this section and the next we prove that $|X|$ and $\|w\|_{H^1_x}$ are small, giving Theorem 2. Then in Section 5 we give the improvement to Theorem 1 under the necessary additional assumptions on the initial data.

To understand the other crucial features of the notation in Lemma 4.1, we introduce the symplectic projection P, characterized by

$$\omega(u, Y\eta) = \omega(P(u)\eta, Y\eta) \quad \text{for all } Y \in \mathfrak{g}.$$

This is given explicitly by

$$P = \sum_{j=1}^{8} e_j P_j, \quad P_j : \mathcal{S}' \to \mathbb{R},$$

$$P_j(u) = -\frac{2}{\|\eta\|_{L^2}^2}\omega(u, e_{j+3}\eta) = \operatorname{Re}\int u(x)x_j\eta(x)\,dx \qquad \text{for } j \in \{1, 2, 3\},$$

$$P_j(u) = \frac{2}{\|\eta\|_{L^2}^2}\omega(u, e_{j-3}\eta) = -\operatorname{Im}\int u(x)\partial_{j-3}\eta(x)\,dx \qquad \text{for } j \in \{4, 5, 6\},$$

$$P_7(u) = \frac{2}{\|\eta\|_{L^2}^2}\omega(u, e_8\eta) = \operatorname{Im}\int u(x)(2 + x\cdot\nabla)\eta(x)\,dx,$$

$$P_8(u) = -\frac{2}{\|\eta\|_{L^2}^2}\omega(u, e_7\eta) = \operatorname{Re}\int u(x)\eta(x)\,dx.$$

We now compute

$$P(if(x)\eta(x)) = \sum_{j=4}^{6} P_j(if(x)\eta(x))e_j + P_7(if(x)\eta(x))e_7$$

$$= -\sum_{j=4}^{6}\Big(\int f(x)\eta(x)\partial_{j-3}\eta(x)\,dx\Big)e_j + \Big(\int f(x)\eta(x)(2 + x\cdot\nabla)\eta(x)\,dx\Big)e_7$$

$$
\begin{aligned}
&= \frac{1}{2}\Big(-\sum_{j=4}^{6}\Big(\int f(x)\partial_{j-3}\eta^2(x)dx\Big)e_j + \Big(\int f(x)\Big(4\eta^2(x) + x\cdot\nabla\eta^2(x)\Big)dx\Big)e_7\Big)\\
&= \frac{1}{2}\Big(\sum_{j=4}^{6}\Big(\int \partial_{j-3}f(x)\eta^2(x)dx\Big)e_j + \Big(\int \Big(f(x) - x\cdot\nabla f(x)\Big)\eta^2(x)dx\Big)e_7\Big)\\
&:= i\alpha + i\beta\cdot x.
\end{aligned}
$$

Observe that in the case that $f(x) = V(x/\mu + a)$ these α and β agree with those defined previously.

We have the following Taylor expansions, where δ_{jk} is the Kronecker delta:

$$
V(x/\mu + a) = V(a) + \nabla V(a)\cdot(x/\mu) + \frac{1}{\mu^2}\sum_{j,k=1}^{3}(1 - \tfrac{1}{2}\delta_{jk})x_jx_k\partial_j\partial_k V(a) + \mathcal{O}(h^3),
$$

$$
\begin{aligned}
\alpha &= V(a) + \frac{3}{4\mu^2}\int\Big(\sum_{j=1}^{3}x_j^2\partial_j^2 V(a)\Big)\eta^2(x)dx + \mathcal{O}(h^3),\\
\beta &= \frac{\nabla V(a)}{\mu} + \mathcal{O}(h^3),
\end{aligned}
$$

and thus

$$
\begin{aligned}
&-V(x/\mu + a) + \alpha + \beta\cdot x\\
&= -\frac{1}{\mu^2}\sum_{j,k=1}^{3}\big(1 - \tfrac{1}{2}\delta_{jk}\big)x_jx_k\partial_j\partial_k V(a) + \frac{3}{4\mu^2}\int\Big(\sum_{j=1}^{3}x_j^2\partial_j^2 V(a)\Big)\eta^2(x)dx + \mathcal{O}(h^3),\\
&:= \sum_{j,k=1}^{3} a_{jk}x_jx_k + a_0 + \mathcal{O}(h^3) := Q(x) + \mathcal{O}(h^3).
\end{aligned}
$$

where all the errors are polynomially bounded in x. In the sequel we will apply Proposition 3.2 using this $Q(x)$. It satisfies the necessary orthogonality condition because $\omega(i(V(x/\mu + a), X\eta)) = 0$, and $Q(x)$ is of order h^2.

We now study w by writing $w = \tilde{w} + w_1$, where $\tilde{w}$ solves away the principal forcing terms of the equation of w. More precisely, we put

$$
\begin{aligned}
\tilde{w} &:= \sum_{j,k=1}^{3}\tilde{w}_{jk}, \quad \tilde{w}_{jk} := -\frac{\partial_j\partial_k V(a)}{\mu^4} f_{jk},\\
f_{jk} &:= L_+^{-1}\Big(-\sum_{j,k=1}^{3}\big(1 - \tfrac{1}{2}\delta_{jk}\big)x_jx_k + \delta_{jk}\frac{3}{4}\int x_j^2\eta^2(x)dx\Big)\eta.
\end{aligned}
$$

Then $\tilde{w}$ satisfies the PDE

$$\partial_t \tilde{w} = -i\mu^2 \mathscr{L}\tilde{w} - \frac{i}{\mu^2}\Big(-\sum_{j,k=1}^{3} \big(1 - \tfrac{1}{2}\delta_{jk}\big) x_j x_k \partial_j \partial_k V(a) + \frac{3}{4}\int \Big(\sum_{j=1}^{3} x_j^2 \partial_j^2 V(a)\Big)\eta^2(x)dx\Big)\eta + \sum_{j,k=1}^{3} \theta_{jk} f_{jk},$$

where

$$\theta_{jk}(t) := \frac{d}{dt}\Big(\frac{-\partial_j\partial_k V(a)}{\mu^4}\Big) = \frac{-\partial_j\partial_k \nabla V(a)\cdot \dot{a}}{\mu^4} + \frac{4\partial_j\partial_k V(a)\dot{\mu}}{\mu^5}.$$

Lemma 4.2. *There exists an absolute constant c such that if* $\|w\|_{H^1} \le 1/c$, *then*

$$|X| \le c(h^2\|w\|_{H^1} + \|w\|_{H^1}^2 + \|w\|_{H^1}^3).$$

Proof. Since $Pw_t = \partial_t P w = 0$, Lemma 4.1 gives

$$X = P(i(V(x/\mu + a) - \alpha - \beta\cdot x)\eta) + P(i(V(x/\mu+a) - \alpha - \beta\cdot x)w) - P(Xw) - \mu^2 P(i\mathscr{N}w) - \mu^2 P(i\mathscr{L}w).$$

We've already seen that the first term vanishes. The estimate $|P(Yw)| \le c|Y|\|w\|_{H^1}$ shows that

$$|P(i(V(x/\mu+a) - \alpha - \beta\cdot x)w)| \le ch^2\|w\|_{H^1} \quad \text{and} \quad |P(Xw)| \le c|X|\|w\|_{H^1}.$$

For the $P(i\mathscr{N}w)$ term we must estimate the following integral, where ψ_k are taken from $w, \eta, e_j\eta$:

$$\begin{aligned} \int |(x^{-1} * (\psi_1\psi_2))\psi_3\psi_4| &\le \||x|^{-1} * (\psi_1\psi_2)\|_{L^3}\|\psi_3\|_{L^6}\|\psi_4\|_{L^2} \\ &\le c\|\psi_1\psi_2\|_{L^1}\|\psi_3\|_{L^6}\|\psi_4\|_{L^2} \\ &\le c\|\psi_1\|_{L^2}\|\psi_2\|_{L^2}\|\psi_3\|_{H^1}\|\psi_4\|_{L^2}. \end{aligned} \tag{4-2}$$

For this we used Hölder's inequality, the Hardy–Littlewood–Sobolev inequality, and Sobolev embedding. This results in $|P(i\mathscr{N}w)| \le c(\|w\|_{H^1}^2 + \|w\|_{H^1}^3)$.

Finally, from (2-17) we have $P(i\mathscr{L}w) = 0$, which combines with the previous estimates to give

$$|X| \le ch^2\|w\|_{H^1} + c|X|\|w\|_{H^1} + c(\|w\|_{H^1}^2 + \|w\|_{H^1}^3).$$

Here we have removed the factors of μ using (4-1). If $\|w\|_{H^1}$ is sufficiently small, this implies the desired inequality. □

Lemma 4.3. *Suppose there are positive constants* c_1, *and* h_0 *such that*

$$\|w\|_{L^\infty_{[t_1,t_2]}H^1_x} \le c_1 h^{1/2+\delta}, \quad h^{2+2\delta}(t_2 - t_1)\langle t_2 - t_1\rangle \le c_1 \quad \textit{if } 0 < h \le h_0,$$

for some $t_1 < t_2$ *and nonnegative* δ. *Then*

$$\sup_{t_1<t<t_2} |\theta(t)| \le ch^3 \quad \textit{and} \quad \sup_{t_1<t<t_2} |v(t)| \le c$$

for a constant c *depending only on* c_1, h_0, $\|W\|_{C^3(\mathbb{R}^3)}$ *and* $|v(t_1)|$.

Proof. The conclusion concerning θ will follow from $|\dot{\mu}| \le ch^{1+2\delta}$ and $|\dot{a}| \le c$. Our assumption on w implies that the bounds for μ in (4-1) can be improved to

$$1 - ch^{1/2+\delta} \le \mu \le 1 + ch^{1/2+\delta}.$$

By the definition of X and the Taylor expansions and the bound on X, we have

$$\left|\frac{\dot{v}}{\mu} + \nabla V(a)\right| + \left|\frac{\dot{\mu}}{\mu}\right| + |\mu(-\dot{a}+v)| \le c|X| \le c(h^2\|w\|_{H^1} + \|w\|_{H^1}^2 + \|w\|_{H^1}^3),$$

which immediately gives the desired bound on $|\dot{\mu}|$. For the bound on $|\dot{a}|$, it suffices to prove $|v| \le c$, which we do by first integrating the inequality above to obtain

$$\sup_{t_1<t<t_2} |v(t)| \le |v(t_1)| + ch\|\nabla W\|_{L^\infty}(t_2-t_1) + c|X|(t_2-t_1).$$

Next we prove a near conservation of classical energy:

$$\begin{aligned}
\sup_{t_1\le t\le t_2} &\left|\left(\tfrac{1}{2}|v|^2 + V(a)\right) - \left(\tfrac{1}{2}|v(t_1)|^2 + V(a(t_1))\right)\right| \\
&\le (t_2-t_1) \sup_{t_1\le t\le t_2} |\dot{v}\cdot v + \nabla V \cdot a| \\
&\le (t_2-t_1) \sup_{t_1\le t\le t_2} \left(|\dot{v} + \nabla V(a)||v| + |\nabla V(a)||\dot{a}-v|\right) \\
&\le c(t_2-t_1)\Big(|X| \sup_{t_1\le t\le t_2} |v| + h\|\nabla W\|_{L^\infty}|X|\Big) \\
&\le c|X|(t_2-t_1)\big(|v(t_1)| + ch\|\nabla W\|_{L^\infty}\langle t_2-t_1\rangle + c|X|(t_2-t_1)\big).
\end{aligned}$$

From this it follows that $\sup_{t_1\le t\le t_2}|v(t)| \le c$, which concludes the proof. □

This will be crucial for the estimate of the true error w.

Lemma 4.4 (Lyapounov energy estimate). *Suppose that, for some constants* c_1 *and* h_0,

$$\|w\|_{L^\infty_{[t_1,t_2]}H^1_x} \le c_1 h^{1/2} \quad \textit{if } 0 < h \le h_0.$$

Then, provided

$$|t_2 - t_1| \le c_2/h,$$

we have

$$\|w\|_{L^\infty_{[t_1,t_2]}H^1_x} \le c_3\|w_1(t_1)\|_{H^1} + c_4 h^2.$$

The constants c_2 *and* c_4 *depend only upon* c_1, h_0, $\|W\|_{C^3(\mathbb{R}^3)}$ *and* $|v(t_1)|$. *The constant* c_3 *is an absolute constant.*

We postpone the proof of this lemma to the end of the section, first demonstrating how it is applied in the bootstrap argument. We prove the following proposition, from which Theorem 2 follows.

Proposition 4.5. *Let $w_0 = w(0)$ and fix constants $\tilde{c}_1 > 0$ and $\delta_0 \in (0, 3/4)$. Then there exist constants h_0 and c such that if*

$$0 \le \delta \le \delta_0, \quad 0 < h \le h_0, \quad \|w_0\|_{H^1} \le \tilde{c}_1 h^{1/2+3\delta_0}, \quad 0 < T \le \frac{\tilde{c}_1}{h} + \frac{\delta|\log h|}{ch},$$

then

$$\|w\|_{L^\infty_{[0,T]} H^1_x} \le ch^{-\delta}(\|w_0\|_{H^1} + h^2).$$

The constants h_0 and c depend only on $\tilde{c}_1$, δ_0, $|v(0)|$, and $\|W\|_{C^3(\mathbb{R}^3)}$.

Proof. To apply Lemma 4.4, we observe that by the continuity in t of $\|w\|_{L^\infty_{[0,t]} H^1_x}$ we know immediately that the hypotheses are satisfied on $[0, t]$ for sufficiently small t. At this point the conclusion of the lemma tells us that at the end of this interval the error is still small enough that we may proceed for larger t, until we reach $t = c_2/h$. In this way we apply Lemma 4.4 k times on successive intervals of length c_2/h, where c_2 and k will be fixed later, giving the bound

$$\|w\|_{L^\infty_{[0,c_2k/h]} H^1_x} \le c_3^k \|w_0\|_{H^1} + \left(\sum_{j=0}^{k-1} c_3^j\right) c_4 h^2.$$

This is only valid provided that the hypotheses of Lemmas 4.3 and 4.4 are satisfied over the whole collection of time intervals. We must use Lemma 4.3 to control $|v|$ uniformly over the full time interval $[0, c_2k/h]$, and to apply this we need

$$c_3^k \|w_0\|_{H^1} + \left(\sum_{j=0}^{k-1} c_3^j\right) c_4 h^2 \le c_1 h^{1/2+\delta} \quad \text{and} \quad c_2^2 k^2 h^{2\delta} \le c_1$$

for some constant c_1. We will determine c_1 momentarily, and at that point c_2 will be the constant that emerges from Lemma 4.4. If

$$k = \frac{\tilde{c}_1}{c_2} + \delta \frac{|\log h|}{\log c_3},$$

it suffices to have

$$\text{(4-3)} \quad c_3^{\tilde{c}_1/c_2} \tilde{c}_1 h^{1/2+3\delta_0-\delta} + c_3^{\tilde{c}_1/c_2} c_4 h^{2-\delta} \le c_1 h^{1/2+\delta} \quad \text{and} \quad \tilde{c}_1^2 \left\langle \delta \frac{|\log h|}{\log c_3} \right\rangle^2 h^{2\delta} \le c_1.$$

We can now choose our constants. We first take c_1 so that the second inequality of (4-3) holds. Then c_2 is given by Lemma 4.4, and we take h_0 so that the first inequality of (4-3) holds. The hypotheses of Lemma 4.3 are satisfied a fortiori. □

It now remains only to prove Lemma 4.4.

Proof of Lemma 4.4. In this proof, unless otherwise mentioned, all constants depend only upon c_1, $\|W\|_{W^{\infty,3}}$ and $|v(t_1)|$.

Let $w_1 := w - \tilde{w}$. Now

$$\partial_t w_1 = -i\mu^2 \mathscr{L} w_1 + X\eta - \theta f + i\Big(-V\Big(\frac{x}{\mu}+a\Big) + \alpha + \beta\cdot x - \frac{x}{2\mu^2}\cdot\nabla^2 V(a)x \\ + \frac{3}{2\mu^2\|\eta\|_{L^2}^2}\int x\cdot\nabla^2 V(a)x\eta^2(x)\,dx\Big)\eta \\ + Xw + i\Big(-V\Big(\frac{x}{\mu}+a\Big)+\alpha+\beta\cdot x\Big)w + i\mu^2\mathscr{N}w.$$

By grouping forcing terms into f_1, we rewrite this as

$$\partial_t w_1 = -i\mu^2\mathscr{L}w_1 + X\eta + f_1 + Xw + i\Big(-V\Big(\frac{x}{\mu}+a\Big)+\alpha+\beta\cdot x\Big)w + i\mu^2\mathscr{N}w,$$

observing that, using Lemma 4.3, we have $\|f_1\|_{H^1} \le ch^3$.

We recall that $\mathscr{L}$ is self-adjoint with respect to $\langle u, v\rangle = \operatorname{Re}\int u\bar{v}$, and hence

$$\begin{aligned}\tfrac{1}{2}\partial_t\langle\mathscr{L}w_1, w_1\rangle &= \langle\mathscr{L}w_1, \partial_t w_1\rangle \\ &= -\mu^2\langle\mathscr{L}w_1, i\mathscr{L}w_1\rangle + \langle\mathscr{L}w_1, X\eta\rangle + \langle\mathscr{L}w_1, f_1\rangle + \langle\mathscr{L}w_1, Xw_1\rangle + \langle\mathscr{L}w_1, X\tilde{w}\rangle \\ &\quad + \langle\mathscr{L}w_1, i\Big(-V\Big(\frac{x}{\mu}+a\Big)+\alpha+\beta\cdot x\Big)w_1\rangle \\ &\quad + \langle\mathscr{L}w_1, i\Big(-V\Big(\frac{x}{\mu}+a\Big)+\alpha+\beta\cdot x\Big)\tilde{w}\rangle + \langle\mathscr{L}w_1, i\mu^2\mathscr{N}w\rangle \\ &= \mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV}+\mathrm{V}+\mathrm{VI}+\mathrm{VII}+\mathrm{VIII}.\end{aligned}$$

Now we analyze these terms one by one. First

$$\mathrm{I} = \mathrm{II} = 0.$$

In the case of I this follows from (2-1), the definition of $\langle\cdot,\cdot\rangle$. In the case of II, we recall that $\omega(w, X\eta) = 0$ by construction of w, and that $\omega(\tilde{w}, X\eta) = 0$ from (3-4), as a result of which we have $\omega(w_1, X\eta) = 0$. Finally $\omega(i\mathscr{L}w_1, X\eta) = 0$ by (2-17), and then we use (2-1) to relate $\langle\cdot,\cdot\rangle$ and $\omega(\cdot,\cdot)$.

Next we show that

$$|\mathrm{III}| \le c\|w_1\|_{H^1}\|f_1\|_{H^1} \le ch^3\|w_1\|_{H^1}.$$

This estimate is straightforward in the case of the convolution-free terms of $\mathscr{L}$. For the terms with convolutions, we apply (4-2) with f_1 in place of ψ_4 and the other ψ_k chosen appropriately from among η, w and $\bar{w}$.

Next we look at $\mathrm{IV} = \langle\mathscr{L}w_1, Xw_1\rangle$. We first recall that $X = \sum_{j=1}^{8} a_j e_j$ with $|a_j| \le c(h^2\|w\| + \|w\|_{H^1}^2 + \|w\|_{H^1}^3)$. We the proceed term by term according to

$\mathscr{L}w_1 = \frac{1}{2}w_1 - \frac{1}{2}\Delta w_1 - (|x|^{-1} * \eta^2)w_1 - \eta(|x|^{-1} * (\eta(w_1 + \bar{w}_1)))$:

$$\langle w_1, Xw_1\rangle = a_8\langle w_1, 2w_1 + x\cdot\nabla w_1\rangle = \tfrac{1}{2}a_8\langle w_1, w_1\rangle,$$

$$\begin{aligned}\langle \Delta w_1, Xw_1\rangle &= \sum_{j=1}^{3} a_{j+3}\langle \Delta w_1, ix_j w_1\rangle + a_8\langle \Delta w_1, 2w_1 + x\cdot\nabla w_1\rangle \\ &= \sum_{j=1}^{3} a_{j+3}\langle \partial_j w_1, iw_1\rangle + \tfrac{1}{2}a_8\langle \nabla w_1, \nabla w_1\rangle,\end{aligned}$$

and thus these two terms are bounded by $c|X|\|w_1\|_{H^1}^2$. For the terms involving η we use (4-2) to obtain the same bound, giving

$$|\mathrm{IV}| \le c(h^2 + \|w\|_{H^1} + \|w\|_{H^1}^2)\|w_1\|_{H^1}^3.$$

Next $\mathrm{V} = \langle \mathscr{L}w_1, X\tilde{w}\rangle$ has a similar expansion, but includes more nonzero terms. We estimate these terms as before in (4-2). We use Hölder's inequality, Hardy–Littlewood–Sobolev, and Sobolev embedding to obtain

$$|\mathrm{V}| \le c|X|\|w_1\|_{H^1}\|\langle x\rangle\tilde{w}\|_{H^2}.$$

However, $\|\langle x\rangle\tilde{w}\|_{H^2} \le ch^2$, giving

$$|\mathrm{V}| \le ch^2(h^2 + \|w\|_{H^1} + \|w\|_{H^1}^2)\|w_1\|_{H^1}.$$

For VI, once again we obtain a number of vanishing terms:

$$\begin{aligned}\mathrm{VI} &= \langle \mathscr{L}w_1, i(-V(x/\mu + a) + \alpha + \beta\cdot x)w_1\rangle \\ &= \big\langle -\tfrac{1}{2}\Delta w_1 - \eta(|x|^{-1} * (\eta(w_1 + \bar{w}_1))), i(-V(x/\mu + a) + \alpha + \beta\cdot x)w_1\big\rangle.\end{aligned}$$

To estimate the first term, we integrate by parts as before and use

$$|-(1/\mu)\nabla V(x/\mu + a) + \beta| \le ch.$$

For the second term, we use (4-2) together with

$$|(-V(x/\mu + a) + \alpha + \beta\cdot x)\eta| \le ch^2.$$

This gives the bound $|\mathrm{VI}| \le ch\|w_1\|_{H^1}^2$.

For VII, we proceed in the same way, without the vanishing terms but also without the restriction that only H^1 norms may be used. We obtain

$$\begin{aligned}|\mathrm{VII}| &\le c\|w_1\|_{H^1}\|(-V(x/\mu + a) + \alpha + \beta\cdot x)\tilde{w}\|_{H^1} \\ &\le ch^2\|w_1\|_{H^1}\|\langle x\rangle^2\tilde{w}\|_{H^1} \le ch^4\|w_1\|_{H^1}.\end{aligned}$$

Finally, for $\mathrm{VIII} = \langle \mathscr{L}w_1, i\mu^2\mathscr{N}w\rangle$ we write $w = w_1 + \tilde{w}$ and expand. We integrate by parts for the Δ term, and use (4-2), twice as needed for the terms with two convolutions. This allows us to put all factors in an H^1 norm, giving a bound of

$$|\mathrm{VIII}| \le c(h^6\|w_1\|_{H^1} + h^4\|w_1\|_{H^1}^2 + h^2\|w_1\|_{H^1}^3 + \|w_1\|_{H^1}^4).$$

Combining all this gives

$$|\partial_t\langle \mathscr{L}w_1, w_1\rangle| \le c(h^3\|w\|_{H^1} + h\|w\|_{H^1}^2 + h^2\|w\|_{H^1}^3 + \|w\|_{H^1}^4 + \|w\|_{H^1}^5).$$

From (B-1) we have uniform boundedness of $\|u\|_{H^1}$, while from Lemma 4.3 we have uniform boundedness of $|v|$ over our time interval, from which we conclude that $\|w\|_{H^1} \le c$, and hence

$$|\partial_t\langle \mathscr{L}w_1, w_1\rangle| \le c(h^3\|w\|_{H^1} + h\|w\|_{H^1}^2 + \|w\|_{H^1}^4).$$

Now we use $w = w_1 + \tilde{w}$ to write $\|w\|_{H^1} \le c(\|w_1\|_{H^1} + h^2)$ and hence

$$|\partial_t\langle \mathscr{L}w_1, w_1\rangle| \le c(h^5 + h\|w_1\|_{H^1}^2 + \|w_1\|_{H^1}^4).$$

Integrating in time gives

$$\langle \mathscr{L}w_1(t), w_1(t)\rangle \le \langle \mathscr{L}w_1(t_1), w_1(t_1)\rangle + c(t-t_1)(h^5 + h\|w_1\|_{H^1}^2 + \|w_1\|_{H^1}^4).$$

From (3-1), we have

$$\|w_1(t)\|_{H^1}^2 \le c\langle \mathscr{L}w_1(t), w_1(t)\rangle,$$

and by direct estimation we have

$$|\langle \mathscr{L}w_1(t), w_1(t)\rangle| \le c\|w_1(t)\|_{H^1}^2.$$

This leads to

$$\|w_1\|_{L^\infty_{[t_1,t]}H^1_x}^2 \le \tilde{c}\|w_1(t_1)\|_{H^1}^2 + c(t-t_1)\big(h^5 + h\|w_1\|_{L^\infty_{[t_1,t]}H^1_x}^2 + \|w_1\|_{L^\infty_{[t_1,t]}H^1_x}^4\big),$$

with $\tilde{c}$ an absolute constant. Requiring that $t_2 - t_1 \le c_2/h$ for a small constant c_2 and subtracting the quadratic term to the left hand side implies

$$\|w_1\|_{L^\infty_{[t_1,t]}H^1_x}^2 \le 2\tilde{c}\|w_1(t_1)\|_{H^1}^2 + c(t_2-t_1)\big(h^5 + h\|w_1\|_{L^\infty_{[t_1,t]}H^1_x}^4\big).$$

This is a quadratic inequality in $\|w_1\|_{L^\infty_{[t_1,t]}H^1_x}^2$. In general,

$$A > 0, \quad B > 0, \quad X \in \mathbb{R}, \quad BX^2 - X + A \ge 0, \quad X \le (2B)^{-1}, \quad 4AB < 1$$

implies $X \le 2A$. In our case, assuming that

$$(t_2-t_1)h\|w_1\|_{L^\infty_{[t_1,t]}H^1_x}^2 + (t_2-t_1)^2h^6 \le c_2,$$

we have

$$\|w_1\|_{L^\infty_{[t_1,t_2]}H^1_x}^2 \le 4\tilde{c}\|w_1(t_1)\|_{H^1}^2 + ch^5(t_2-t_1).$$

From this, together with $w = w_1 + \tilde{w}$ the desired result follows. □

5. Proof of Theorem 1

Lemma 5.1. *Suppose that* $0 < h \ll 1$, *and* $a = a(t)$, $v = v(t)$, $\epsilon_1 = \epsilon_1(t)$, $\epsilon_2 = \epsilon_2(t)$ *are* C^1 *real-valued functions. Suppose* $f : \mathbb{R}^3 \to \mathbb{R}$ *is* C^2 *mapping such that* $|f|$ *and* $|f'|$ *are uniformly bounded. Suppose that on* $[0, T]$,

$$\dot{a} = v + \epsilon_1, \qquad a(0) = a_0,$$
$$\dot{v} = hf(ha) + \epsilon_2, \quad v(0) = v_0.$$

Let $\bar{a} = \bar{a}(t)$ *and* $\bar{v} = \bar{v}(t)$ *be the* C^1 *real-valued functions satisfying the exact equations*

$$\dot{\bar{a}} = \bar{v} + \epsilon_1, \qquad \bar{a}(0) = a_0,$$
$$\dot{\bar{v}} = hf(h\bar{a}) + \epsilon_2, \quad \bar{v}(0) = v_0$$

with the same initial data. Suppose that on $[0, T]$, *we have* $|\epsilon_j| \le h^{4-\delta}$ *for* $j = 1, 2$. *Then provided* $T \le ch^{-1} + \delta h^{-1} \log(1/h)$, *we have on* $[0, T]$ *the estimates*

$$|a - \bar{a}| \le \tilde{c}h^{2-2\delta} \log(1/h) \quad \textit{and} \quad |v - \bar{v}| \le \tilde{c}h^{3-2\delta} \log(1/h).$$

The statement and proof of this lemma is almost identical to those of [Holmer and Zworski 2008, Lemma 6.1]. The only change in this proof is that we use $g = \int_0^1 \nabla f(h\bar{a} + t(ha - h\bar{a}))dt$.

For Theorem 1, we assume $\varepsilon = \mathbb{O}(h^2)$, in which case a and v satisfy the ODEs

$$\dot{a} = v + \mathbb{O}(h^{4-4\delta}) \quad \text{and} \quad \dot{v} = -\tfrac{1}{2} \int \nabla V(x + a)\eta^2(x)dx + \mathbb{O}(h^{4-4\delta}).$$

Lemma 5.1 allows us to replace these with

$$\dot{a} = v \quad \text{and} \quad \dot{v} = -\tfrac{1}{2} \int \nabla V(x + a)\eta^2(x)dx.$$

Direct integration of the error terms in the equations for μ and γ allows them to be dropped as well, giving Theorem 1. □

Appendix A: Properties of η

In this appendix we review the properties of the function η used in this paper. This material is essentially well known, and further information and references may be found in [Lenzmann 2009].

Lemma A.1 [Lenzmann 2009, Appendix A]. *For each* $\lambda > 0$, *the equation*

$$-\tfrac{1}{2}\Delta\eta + \widetilde{V}\eta = -\lambda\eta \tag{A-1}$$

with $\widetilde{V} = -|x|^{-1} * \eta^2$, *has a unique radial, nonnegative solution* $\eta \in H^1(\mathbb{R}^3)$ *with* $\eta \not\equiv 0$. *Moreover*, $\eta(r)$ *is strictly positive.*

In this paper we choose λ so that $\|\eta\|_{L^2}^2 = 2$.

We will also need the following exponential decay result.

Proposition A.2. *Let* $\eta \in H^1(\mathbb{R}^3; \mathbb{R})$ *satisfy* (A-1). *Then* $\eta \in C^\infty(\mathbb{R}^3)$, *and for any multiindex* α *and* $\epsilon > 0$ *there exists* C *such that*

$$|\partial^\alpha \eta(x)| \le C e^{-(\sqrt{2\lambda}-\epsilon)|x|}.$$

Proof. Observe first that $\widetilde{V}$ is continuous and obeys $\lim_{|x|\to\infty} \widetilde{V} = 0$. Indeed, write $|x|^{-1} = \chi_1 + \chi_2$, where χ_1 is smooth and agrees with $|x|^{-1}$ near infinity, and χ_2 is compactly supported and in L^p for $p < 3$. The χ_1 term is clearly smooth, and we prove the decay by treating it in two pieces:

$$\int_{|y|\le|x|/2} \chi_1(x-y)\eta^2(y)dy \le \int_{|y|\le|x|/2} \frac{C}{\langle x-y\rangle}\eta^2(y)dy \le \frac{C}{|x|}\|\eta\|_{L^2}^2,$$
$$\int_{|y|\ge|x|/2} \chi_1(x-y)\eta^2(y)dy \le \|\chi_1\|_{L^\infty} \int_{|y|\ge|x|/2} \eta^2(y)dy.$$

On the other hand, note that since $\eta \in H^1(\mathbb{R}^3)$, the Gagliardo–Nirenberg inequality implies that $\eta \in L^6(\mathbb{R}^3)$, and in particular $\eta^2 \in L^2$. Thus $\chi_2 * \eta^2$ has a Fourier transform in L^1, giving the desired regularity and decay.

Now it follows from (A-1) that $\eta \in H^2$. Differentiating the equation and applying the previous argument shows that $\eta \in H^3$. By induction we find that $\eta \in H^s$, and in particular $\eta \in C^\infty$.

We now prove the exponential decay as follows. Let $P = -\frac{1}{2}\Delta + \widetilde{V}$, let $\phi \in C^\infty$ be bounded together with its first derivatives, and let

$$P_\phi := e^\phi P e^{-\phi} = -\tfrac{1}{2}\Delta + \nabla\phi\cdot\nabla - \tfrac{1}{2}|\nabla\phi|^2 + \tfrac{1}{2}\Delta\phi + \widetilde{V}.$$

Let $w = e^\phi \eta$ and, observing that integrating by parts gives $\int(\nabla\phi\cdot\nabla w)w = -\int(\nabla\phi\cdot\nabla w)w - \int(\Delta\phi)w^2$, write

$$0 = \langle (P_\phi + \lambda)w, w\rangle_{L^2} = \frac{1}{2}\int |\nabla w|^2 + \int (\widetilde{V} + \lambda - \tfrac{1}{2}|\nabla\phi|^2)w^2.$$

Now, provided $|\nabla\phi|^2 \le 2\lambda - 2\epsilon$, we have

$$\begin{aligned}\epsilon \int w^2 &\le \int (\lambda - \tfrac{1}{2}|\nabla\phi|^2)w^2 \le -\int \widetilde{V}w^2 \\ &\le \tfrac{1}{2}\epsilon \int_{\{x:\widetilde{V}(x)\ge -\epsilon/2\}} w^2 - \int_{\{x:\widetilde{V}(x)< -\epsilon/2\}} \widetilde{V}w^2.\end{aligned}$$

The integral over $\{x : \widetilde{V}(x) \ge -\epsilon/2\}$ can now be subtracted to the other side of the inequality, while $\{x : \widetilde{V}(x) < -\epsilon/2\}$ is a bounded set since $\lim_{|x|\to\infty} \widetilde{V}(x) = 0$. We may then write $\int w^2 \le C$, where C depends on η, $\sup|\phi|$, and ϵ. If we apply this result with a sequence of functions ϕ_n such that ϕ_n is equal to $(\sqrt{2\lambda - 2\epsilon})x_1$

on the ball of radius n and is modified outside that ball to be smooth with bounded derivatives, we find that $e^{\sqrt{2\lambda-2\epsilon}x_1}\eta \in L^2$, and similarly

$$e^{\sqrt{2\lambda-2\epsilon}|x|}\eta(x) \in L^2.$$

Differentiating (A-1) and applying the same argument proves that

$$e^{\sqrt{2\lambda-2\epsilon}|x|}\partial^\alpha \eta(x) \in L^2,$$

from which the desired result follows. □

Appendix B: Well-posedness

In this appendix we prove well-posedness for Equation (1-1) in $H^1(\mathbb{R}^3)$. This result is known (see for example [Cazenave 1996]), but for the reader's convenience we review the result in the special case that we study here. We adopt the notation $\|u\|_{W^{k,p}} = \sum_{|\alpha|\le k} \|\partial^\alpha u\|_{L^p}$.

We will use these Strichartz estimates (see for example [Keel and Tao 1998]):

Lemma B.1. *Suppose* $q, r, \tilde{q}', \tilde{r}' \in [1, \infty]$ *satisfy*

$$\frac{2}{q} + \frac{n}{r} = \frac{n}{2} \quad \text{and} \quad \frac{2}{\tilde{q}'} + \frac{n}{\tilde{r}'} = \frac{4+n}{2}.$$

Then

$$\|e^{it\Delta}u_0\|_{L^q_{[0,T]}L^r_x} \le c\|u_0\|_{L^2} \quad \text{and} \quad \left\| \int_0^t e^{i(t-s)\Delta} f(s)ds \right\|_{L^q_{[0,T]}L^r_x} \le c\|f\|_{L^{\tilde{q}'}_{[0,T]}L^{\tilde{r}'}_x}$$

for all $u_0 \in L^2(\mathbb{R}^n)$ *and* $f \in L^{\tilde{q}'}([0, T], L^{\tilde{r}'}(\mathbb{R}^n))$.

In the remainder of this section only, c denotes a constant that may vary from line to line, but is absolute and independent of all parameters in the problem. Let $V \in W^{1,\infty}(\mathbb{R}^3, \mathbb{R})$, let $u_0 \in H^1(\mathbb{R}^3)$ be given, and define

$$N(u) = -(|x|^{-1} * |u|^2)u,$$
$$F(u)(t) = e^{it\Delta}u_0 - i \int_0^t e^{i(t-s)\Delta}(N(u(s)) + Vu(s))ds.$$

A function u solves the Hartree equation if and only if it is a fixed point of F.

Lemma B.2. *For any* $T > 0$, *we have*

$$\|N(u)\|_{H^1(\mathbb{R}^3)} \le c\|u\|_{L^2(\mathbb{R}^3)}\|\nabla u\|_{H^1(\mathbb{R}^3)},$$
$$\|F(u)\|_{L^\infty([0,T],H^1(\mathbb{R}^3))} \le \|u_0\|_{H^1(\mathbb{R}^3)} + T^{1/2}(c\|u\|^3_{H^1(\mathbb{R}^3)} + \|V\|_{W^{1,\infty}(\mathbb{R}^3)}\|u\|_{H^1(\mathbb{R}^3)}),$$

where c is an absolute constant.

Proof. We first compute

$$\text{(B-1)}\qquad \begin{aligned}\|(|x|^{-1}*|u|^2)u\|_{L^2} &\le \|(|x|^{-1}*|u|^2)\|_{L^3}\|u\|_{L^6}\\ &\le c\||u|^2\|_{L^1}\|u\|_{L^6}\le c\|\nabla u\|_{L^2}\|u\|_{L^2}^2,\end{aligned}$$

where we have used in the first inequality Hölder, in the second Hardy–Littlewood–Sobolev, and in the third Hölder followed by the Sobolev inclusion of $\dot{H}^1(\mathbb{R}^3)$ into $L^6(\mathbb{R}^3)$. From this the result concerning N follows.

We now look at F. We have $\|e^{it\Delta}u_0\|_{L^\infty([0,T],H^1(\mathbb{R}^3))} = \|u_0\|_{H^1(\mathbb{R}^3)}$ because the Schrödinger propagator is unitary on all Sobolev spaces. We then compute using Strichartz estimates that

$$\begin{aligned}\Big\|\int_0^t e^{i(t-s)\Delta}N(u(s))ds\Big\|_{L^\infty([0,T],L^2(\mathbb{R}^3))} &\le c\|N(u)\|_{L^2_{[0,T]}L^{6/5}_x}\\ &\le cT^{1/2}\|N(u)\|_{L^\infty_{[0,T]}L^{6/5}_x}.\end{aligned}$$

Using the same sequence of inequalities as in (B-1), we get

$$\big\|(|x|^{-1}*|u|^2)u\big\|_{L^{6/5}} \le \big\||x|^{-1}*|u|^2\big\|_{L^3}\|u\|_{L^2} \le c\||u|^2\|_{L^1}\|u\|_{L^2} = c\|u\|^3_{L^2}.$$

The same arguments show that

$$\Big\|\nabla\int_0^t e^{i(t-s)\Delta}N(u(s))ds\Big\|_{L^\infty([0,T],L^2(\mathbb{R}^3))} \le T^{1/2}\|u\|^2_{L^2}\|\nabla u\|_{L^2}. \qquad \square$$

Proposition B.3. *For each $u_0 \in H^1(\mathbb{R}^3;\mathbb{C})$, there exists $T\in\mathbb{R}$ such that* (1-1) *has a solution $u(x,t)\in L^\infty([0,T],H^1(\mathbb{R}^3))$. This T depends only on $\|u_0\|_{H^1}$.*

Proof. We prove this using a standard contraction argument. We adopt the notation $\|\cdot\| = \|\cdot\|_{L^\infty([0,T]H^1(\mathbb{R}^3))}$:

$$\begin{aligned}&\|F(u)-F(v)\|\\ &\quad\le \Big\|\int_0^t e^{i(t-s)\Delta}(N(u(s))-N(v(s)))ds\Big\| + \Big\|\int_0^t e^{i(t-s)\Delta}[Vu(s)-Vv(s)]ds\Big\|\\ &\quad\le c\big(\|N(u(t))-N(v(t))\|_{L^2_{[0,T]}W^{1,6/5}_x} + T\|Vu(t)-Vv(t)\|\big).\end{aligned}$$

But then

$$\begin{aligned}&c\|N(u(t))-N(v(t))\|_{L^2_{[0,T]}W^{1,6/5}_x}\\ &\quad\le cT^{1/2}\|N(u)-N(v)\|_{L^\infty_{[0,T]}W^{1,6/5}_x}\\ &\quad\le cT^{1/2}\big(\|(|x|^{-1}*|u|^2)(u-v)\|_{L^\infty_{[0,T]}W^{1,6/5}_x} + \|(|x|^{-1}*u(\bar u-\bar v))v\|_{L^\infty_{[0,T]}W^{1,6/5}_x}\\ &\qquad\qquad + \|(|x|^{-1}*(u-v)\bar v)v\|_{L^\infty_{[0,T]}W^{1,6/5}_x}\big)\\ &\quad\le cT^{1/2}\|u-v\|(\|u\|^2+\|u\|\|v\|+\|v\|^2).\end{aligned}$$

Thus taking

$$T^{1/2} \leq \frac{1}{c\left(\|u\|^2+\|u\|\,\|v\|+\|v\|^2+\|V\|_{W^{1,\infty}(\mathbb{R}^3)}\right)},$$

we find that F is a contraction on a closed ball of $L^\infty([0,T], H^1(\mathbb{R}^3))$, implying there exists a solution to (1-1). □

We use almost conservation of energy to extend this to global well-posedness.

Proposition B.4. *Equation* (1-1) *has a solution in* $L^\infty(\mathbb{R}, H^1(\mathbb{R}^3))$.

Proof. Because of Proposition B.3, it is sufficient to prove that the H^1 norm of u is bounded. Clearly $\|u\|_{L^2}$ is preserved so it suffices to bound $\|\nabla u\|_{L^2}$. To do this, we study the energy

$$E(t) = \|\nabla u\| - \int_{\mathbb{R}^3} N(u)\bar{u}.$$

An argument as above shows that

$$\begin{aligned}\int (|x|^{-1} * |u|^2)|u|^2 &\leq \||x|^{-1} * |u|^2\|_{L^3}\|u^2\|_{L^{3/2}} \leq c\|u\|_{L^2}^3\|\nabla u\|_{L^2}\\ &\leq \frac{c}{\epsilon}\|u\|_{L^2}^3 + c\epsilon\|\nabla u\|_{L^2}.\end{aligned}$$

From this we deduce that

$$\|\nabla u\|_{L^2}^2 \leq c\left(E(t) + \|u\|_{L^2}^3 + \|V\|_{W^{1,\infty}}\right).$$

This bounds $\|u\|_{H^1_x}$ uniformly in time, giving the desired conclusion. □

Acknowledgments

We would like to thank Maciej Zworski for suggesting this problem and for his generous guidance and advice. Thanks also to Justin Holmer and Sebastian Herr for helpful discussions regarding Appendix B.

References

[Abou Salem 2008] W. K. Abou Salem, "Solitary wave dynamics in time-dependent potentials", *J. Math. Phys.* **49**:3 (2008), Art. ID 032101. MR 2009c:35424 Zbl 1153.81428

[Bronski and Jerrard 2000] J. C. Bronski and R. L. Jerrard, "Soliton dynamics in a potential", *Math. Res. Lett.* **7**:2-3 (2000), 329–342. MR 2001g:37139 Zbl 0955.35067

[Cazenave 1996] T. Cazenave, *An introduction to nonlinear Schrödinger equations*, 3rd ed., Textos de Métodos Matemáticos **26**, Instituto de Matematica - UFRJ, Rio de Janeiro, 1996.

[Éboli and Marques 1983] O. J. P. Éboli and G. C. Marques, "Solitons as Newtonian particles", *Phys. Rev. B* (3) **28**:2 (1983), 689–696. MR 85e:78005

[Fröhlich, Gustafson, Jonsson and Sigal 2004] J. Fröhlich, S. Gustafson, B. L. G. Jonsson, and I. M. Sigal, "Solitary wave dynamics in an external potential", *Comm. Math. Phys.* **250**:3 (2004), 613–642. MR 2005h:35320 Zbl 1075.35075

[Fröhlich, Tsai and Yau 2002] J. Fröhlich, T.-P. Tsai, and H.-T. Yau, "On the point-particle (Newtonian) limit of the non-linear Hartree equation", *Comm. Math. Phys.* **225**:2 (2002), 223–274. MR 2003e:81047

[Holmer and Zworski 2007] J. Holmer and M. Zworski, "Slow soliton interaction with delta impurities", *J. Mod. Dyn.* **1**:4 (2007), 689–718. MR 2008k:35446

[Holmer and Zworski 2008] J. Holmer and M. Zworski, "Soliton interaction with slowly varying potentials", *Int. Math. Res. Not.* **2008**:10 (2008), Art. ID rnn026. MR 2009i:37183 Zbl 1147.35084

[Jonsson et al. 2006] B. L. G. Jonsson, J. Fröhlich, S. Gustafson, and I. M. Sigal, "Long time motion of NLS solitary waves in a confining potential", *Ann. Henri Poincaré* **7**:4 (2006), 621–660. MR 2007f:35269 Zbl 1100.81019

[Keel and Tao 1998] M. Keel and T. Tao, "Endpoint Strichartz estimates", *Amer. J. Math.* **120**:5 (1998), 955–980. MR 2000d:35018 Zbl 0922.35028

[Lenzmann 2009] E. Lenzmann, "Uniqueness of ground states for pseudorelativistic Hartree equations", *Anal. PDE* **2**:1 (2009), 1–27. MR 2010j:35423 Zbl 1183.35266

[Lieb 1977] E. H. Lieb, "Existence and uniqueness of the minimizing solution of Choquard's nonlinear equation", *Studies in Appl. Math.* **57**:2 (1977), 93–105. MR 57 #11508 Zbl 0369.35022

[Weinstein 1986] M. I. Weinstein, "Lyapunov stability of ground states of nonlinear dispersive evolution equations", *Comm. Pure Appl. Math.* **39**:1 (1986), 51–67. MR 87f:35023 Zbl 0594.35005

Received July 27, 2009.

KIRIL DATCHEV
MATHEMATICS DEPARTMENT
UNIVERSITY OF CALIFORNIA
BERKELEY, CA 94720
UNITED STATES
datchev@math.berkeley.edu

IVAN VENTURA
MATHEMATICS DEPARTMENT
UNIVERSITY OF CALIFORNIA
BERKELEY, CA 94720
UNITED STATES
iventura@math.berkeley.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

UNIQUELY PRESENTED FINITELY GENERATED COMMUTATIVE MONOIDS

PEDRO A. GARCÍA-SÁNCHEZ AND IGNACIO OJEDA

A finitely generated commutative monoid is uniquely presented if it has a unique minimal presentation. We give necessary and sufficient conditions for finitely generated, combinatorially finite, cancellative, commutative monoids to be uniquely presented. We use the concept of gluing to construct commutative monoids with this property. Finally, for some relevant families of numerical semigroups we describe the elements that are uniquely presented.

Introduction

Rédei [1965] proved that every finitely generated commutative monoid is finitely presented. Since then, the proof has been shortened drastically, and much progress has been made on the study and computation of minimal presentations of monoids, more specifically, of finitely generated subsemigroups of $\mathbb{N}^n$, known usually as affine semigroups; see for instance [Rosales 1997] and [Briales et al. 1998] or [Rosales and García-Sánchez 1999a, Chapter 9] and the references therein. For affine semigroups, the concepts of minimal presentations with respect to cardinality or set inclusion coincide, that is, any two minimal presentations have the same cardinality. This even occurs in a more general setting; see [Rosales et al. 1999].

Interest of the study of such kind of monoids and their presentations was partially motivated by their application in commutative algebra and algebraic geometry [Bruns and Herzog 1993, Chapter 6; Fulton 1993].

Recently, new applications of affine semigroups have been found in the so-called algebraic statistic. In this context, an interesting problem is to decide under which conditions such monoids have a unique minimal presentation. Roughly speaking, convenient algebraic techniques for the study of some statistical models seem to be

MSC2000: primary 20M14; secondary 20M05.
Keywords: commutative monoid, affine semigroup, numerical semigroup, congruences, minimal presentation, Betti numbers, indispensability, gluing of semigroups.
García-Sánchez is supported by the project MTM2007-62346 and FEDER funds. Ojeda is partially supported by the project MTM2007-64704, National Plan I+D+I, by Junta de Extremadura and FEDER funds.

more interesting for statisticians when a certain semigroup associated to the model is uniquely presented; see [Takemura and Aoki 2004].

Efforts to understand the problem of the uniqueness come from an algebraic setting and consist essentially in identifying particular minimal generators in a presentation as R-module of the semigroup algebra, where R is a polynomial ring over a field; see [Charalambous et al. 2007; Ojeda and Vigneron-Tenorio 2009]. So, whole families of uniquely presented monoids have not been determined (with the exception of some previously known cases [Ojeda 2008]) and techniques for constructing uniquely presented monoids have not been developed.

Here, we approach the uniqueness of the minimal presentations from a semigroup theoretic point of view. To begin, we recall the basic definitions and how to obtain minimal presentations of finitely generated, combinatorially finite, cancellative and commutative monoids (which include affine semigroups). In Section 2, we focus on the elements of the monoid whose factorizations yield these presentations, which we call Betti elements. Section 3 provides a necessary and sufficient condition for a monoid to be uniquely presented (Corollary 6). Some results in these sections may be also stated in combinatorial terms by using the simplicial complexes introduced by S. Eliahou in his unpublished PhD thesis (1983); see [Charalambous et al. 2007; Ojeda and Vigneron-Tenorio 2010].

In Section 4, we make extensive use of the gluing of affine semigroups, a concept defined by J. C. Rosales [1997] and used later by different authors to characterize complete intersection affine semigroup rings. In that section, given a gluing S of two affine semigroups S_1 and S_2, we show that S is uniquely presented if and only if S_1 and S_2 are uniquely presented and some extra natural condition holds where S_1 and S_2 are glued (Theorem 12). To reach this result, we need Theorem 10, which shows that the Betti elements of S are the union of the Betti elements of S_1, S_2 and the element in which S_1 and S_2 glue to produce S. We consider these two theorems to be our main results. Furthermore, Theorem 12 may be used to systematically produce uniquely presented monoids, as we show in Example 14.

Finally Section 5 identifies all uniquely presented monoids in some classical families of numerical semigroups (submonoids of $\mathbb{N}$ with finite complement in $\mathbb{N}$).

1. Preliminaries

We summarize some definitions, notations and results that will be useful later in the paper. See [Rosales and García-Sánchez 1999a] for further information.

Let S denote a *commutative monoid*, that is, a set with a binary operation that is associative, commutative and has an identity element $\mathbf{0}$. Since S is commutative, we will use additive notation. Assume that S is *cancellative*, that is, $\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{a}+\mathbf{c}$ in S implies $\boldsymbol{b}=\mathbf{c}$. The monoids we study here are also *free of units*: $S\cap(-S)=\{\mathbf{0}\}$.

Some authors call these monoids *reduced* [Rosales and García-Sánchez 1999a]; others refer to this property as *positivity* [Bruns and Herzog 1993, Chapter 6]. Regardless of what we call them, their most important property is that they are *combinatorially finite*, that is, every element $\boldsymbol{a} \in S$ can be expressed only in finitely many ways as a sum $\boldsymbol{a} = \boldsymbol{a}_1 + \cdots + \boldsymbol{a}_q$, with $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_q \in S \setminus \{0\}$. See [Briales et al. 1998; Rosales et al. 1999] for a wider class of monoids where this condition still holds true. Monoids with this property are also known as FF-monoids. In [Geroldinger and Halter-Koch 2006] it is proved that mutiplicative monoids of all Krull monoids, all Dedekind domains, all orders in number fields are FF-monoids. Moreover, the binary relation on S defined by $\boldsymbol{b} \prec_S \boldsymbol{a}$ if $\boldsymbol{a} - \boldsymbol{b} \in S$ is a well-defined order on S that satisfies the descending chain condition.

All monoids considered here are finitely generated, commutative, cancellative and free of units, and thus we will omit these adjectives in what follows. Examples of monoids fulfilling these conditions are *affine semigroups*, that is, monoids isomorphic to finitely generated submonoids of $\mathbb{N}^r$ with r a positive integer ($\mathbb{N}$ denotes here the set of nonnegative integers), and in particular, *numerical semigroups* that are submonoids of the set of nonnegative integers with finite complement in $\mathbb{N}$.

We will write $S = \langle \boldsymbol{a}_1, \ldots, \boldsymbol{a}_r \rangle$ for the monoid generated by $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r\}$, that is, $S = \boldsymbol{a}_1\mathbb{N} + \cdots + \boldsymbol{a}_r\mathbb{N}$. In such a case, $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r\}$ will be said to be a *system of generators* of S. If no proper subset of $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r\}$ generates S, the set $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r\}$ is a *minimal* system of generators of S. In our context, every monoid has a unique minimal system of generators: If $S^* = S \setminus \{0\}$, then the minimal system of generators of S is $S^* \setminus (S^* + S^*)$; see [Rosales and García-Sánchez 1999a, Chapter 3]. In particular, if S is the set of solutions of a system of linear Diophantine equations and/or inequalities, the minimal system of generators of S coincides with the so-called Hilbert basis; see for example [Sturmfels 1996, Chapter 13].

If S is a numerical semigroup minimally generated by $\{a_1 < \cdots < a_r\} \subset \mathbb{N}$, the number r is usually called the *embedding dimension* of S, and the number a_1 is the *multiplicity* of S. It is easy to show (and well known) that $\boldsymbol{a}_1 \geq r$; see [Rosales and García-Sánchez 2009, Proposition 2.10]. When $a_1 = r$, we say S is of *maximal embedding dimension*.

Given the minimal system $A = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r\}$ of generators of a monoid S, consider the monoid map

$$\varphi_A : \mathbb{N}^r \to S, \quad \boldsymbol{u} = (u_1, \ldots, u_r) \mapsto \textstyle\sum_{i=1}^r u_i \boldsymbol{a}_i.$$

This map is sometimes known as the *factorization homomorphism* associated to S.

Notice that each $\boldsymbol{u} = (u_1, \ldots, u_r) \in \varphi_A^{-1}(\boldsymbol{a})$ gives a *factorization* of $\boldsymbol{a} \in S$, say $\boldsymbol{a} = \sum_{i=1}^r u_i \boldsymbol{a}_i$. Thus, $\#\varphi_A^{-1}(\boldsymbol{a})$ is the number of factorizations of $\boldsymbol{a} \in S$. This number is finite because of the combinatorial finiteness of S; see also [Rosales and García-Sánchez 1999a, Lemma 9.1].

Let $\sim_A$ be the kernel congruence of φ_A, that is, $\boldsymbol{u} \sim_A \boldsymbol{v}$ if $\varphi_A(\boldsymbol{u}) = \varphi_A(\boldsymbol{v})$ (the kernel congruence is actually a congruence, an equivalence relation compatible with addition). It follows easily that S is isomorphic to the monoid $\mathbb{N}^r/{\sim_A}$.

Given $\rho \subseteq \mathbb{N}^r \times \mathbb{N}^r$, the congruence generated by ρ is the least congruence containing ρ, that is, the intersection of all congruences containing ρ. If $\sim$ is the congruence generated by ρ, we say that ρ is a *system of generators* of $\sim$. Rédei's theorem [1965] precisely states that every congruence on $\mathbb{N}^r$ is finitely generated. A *presentation* for S is a system of generators of $\sim_A$, and a *minimal presentation* is a minimal system of generators of $\sim_A$ (in the sense that none of its proper subsets generates $\sim_A$). In our setting, all minimal presentations have the same cardinality; see for instance [Rosales et al. 1999; Rosales and García-Sánchez 1999a]. This is not the case for finitely generated monoids in general.

Next we briefly describe a procedure for finding all minimal presentations for S as presented in [Rosales et al. 1999]; in our context this description is given in [Rosales and García-Sánchez 1999a, Chapter 9].

For $\boldsymbol{u} = (u_1, \dots, u_r)$ and $\boldsymbol{v} = (v_1, \dots, v_r) \in \mathbb{N}^r$, we write $\boldsymbol{u} \cdot \boldsymbol{v}$ for $\sum_{i=1}^r u_i v_i$ (the dot product).

Given $\boldsymbol{a} \in S$, we define a binary relation on $\varphi_A^{-1}(\boldsymbol{a})$: For $\boldsymbol{u}, \boldsymbol{u}' \in \varphi_A^{-1}(\boldsymbol{a})$, we say $\boldsymbol{u} \,\mathcal{R}\, \boldsymbol{u}'$ if there exists a chain $\boldsymbol{u}_0, \dots, \boldsymbol{u}_k \in \varphi_A^{-1}(\boldsymbol{a})$ such that

(a) $\boldsymbol{u}_0 = \boldsymbol{u},\ \boldsymbol{u}_k = \boldsymbol{u}'$, and

(b) $\boldsymbol{u}_i \cdot \boldsymbol{u}_{i+1} \neq 0$ for $i \in \{0, \dots, k-1\}$.

For every $\boldsymbol{a} \in S$, define $\rho_{\boldsymbol{a}}$ in the following way.

- If $\varphi_A^{-1}(\boldsymbol{a})$ has one $\mathcal{R}$-class, set $\rho_{\boldsymbol{a}} = \varnothing$.
- Otherwise, let $\mathcal{R}_1, \dots, \mathcal{R}_k$ be the different $\mathcal{R}$-classes of $\varphi_A^{-1}(\boldsymbol{a})$. Choose $\boldsymbol{v}_i \in \mathcal{R}_i$ for all $i \in \{1, \dots, k\}$ and set $\rho_{\boldsymbol{a}}$ to be any set of $k-1$ pairs of elements in $V = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ such that any two elements in V are connected by a sequence of pairs in $\rho_{\boldsymbol{a}}$ (or their symmetrics). For instance, we can choose $\rho_{\boldsymbol{a}} = \{(\boldsymbol{v}_1, \boldsymbol{v}_2), \dots, (\boldsymbol{v}_1, \boldsymbol{v}_k)\}$ or $\rho_{\boldsymbol{a}} = \{(\boldsymbol{v}_1, \boldsymbol{v}_2), (\boldsymbol{v}_2, \boldsymbol{v}_3), \dots, (\boldsymbol{v}_{k-1}, \boldsymbol{v}_k)\}$.

Then $\rho = \bigcup_{\boldsymbol{a} \in S} \rho_{\boldsymbol{a}}$ is a minimal presentation of S. In this way one can construct all minimal presentations for S. Because S is finitely presented, there are finitely many elements $\boldsymbol{a}$ in S for which $\varphi_A^{-1}(\boldsymbol{a})$ has more than one $\mathcal{R}$-class.

2. Betti elements

As we have seen above, a minimal presentation of S is a set of pairs of factorizations of some elements in S, namely, those having more than one $\mathcal{R}$-class. We say that $\boldsymbol{a} \in S$ is a *Betti element* if $\varphi_A^{-1}(\boldsymbol{a})$ has more than one $\mathcal{R}$-class.

We will say the $\boldsymbol{a} \in S$ is *Betti-minimal* if it is minimal among all the Betti elements in S with respect to $\prec_S$. Of course, Betti elements in S are not necessarily Betti-minimal. Consider, for instance, $S = \langle 4, 6, 21 \rangle$ and $\boldsymbol{a} = 42$.

We will write $\mathrm{Betti}(S)$ and $\text{Betti-minimal}(S)$ for the sets of Betti elements and Betti minimal elements of the monoid S, respectively.

Lemma 1. *Let $S = \langle \boldsymbol{a}_1, \ldots, \boldsymbol{a}_r \rangle$. If $\boldsymbol{a} \notin \mathrm{Betti}(S)$ and $\#\varphi_A^{-1}(\boldsymbol{a}) \geq 2$, there exists $\boldsymbol{a}' \in \mathrm{Betti}(S)$ such that $\boldsymbol{a}' \prec_S \boldsymbol{a}$.*

Proof. We proceed by induction on $\#\varphi_A^{-1}(\boldsymbol{a})$. If $\varphi_A^{-1}(\boldsymbol{a}) = \{\boldsymbol{u}, \boldsymbol{v}\}$ with $\boldsymbol{u} \cdot \boldsymbol{v} > 0$, consider $\boldsymbol{a}' = \boldsymbol{a} - \sum_{i=1}^r \min(u_i, v_i)\boldsymbol{a}_i$. Then, putting $u_i' = u_i - \min(u_i, v_i)$ and $v_i' = v_i - \min(u_i, v_i)$ for $i \in \{1, \ldots, r\}$, we have $\varphi_A^{-1}(\boldsymbol{a}') = \{\boldsymbol{u}', \boldsymbol{v}'\}$, and $\boldsymbol{u}' \cdot \boldsymbol{v}' = 0$. So, $\boldsymbol{a}' \prec \boldsymbol{a}$ is Betti. Assume now that the result is true for every $\boldsymbol{a}' \in S$ such that

$$2 \leq \#\varphi_A^{-1}(\boldsymbol{a}') < \#\varphi_A^{-1}(\boldsymbol{a}).$$

Since $\boldsymbol{a}$ is not Betti, there exist unequal $\boldsymbol{u}, \boldsymbol{v} \in \varphi_A^{-1}(\boldsymbol{a})$ such that $\boldsymbol{u} \cdot \boldsymbol{v} > 0$. If $\boldsymbol{a}' = \boldsymbol{a} - \sum_{i=1}^r \min(u_i, v_i)\boldsymbol{a}_i$, then $2 \leq \#\varphi_A^{-1}(\boldsymbol{a}') \leq \#\varphi_A^{-1}(\boldsymbol{a})$. If the second inequality is strict, we conclude by induction hypothesis. Otherwise, if $\boldsymbol{a}'$ is not Betti, we may repeat the previous argument to produce $\boldsymbol{a}'' \prec_S \boldsymbol{a}' \prec_S \boldsymbol{a}$. The descending chain condition for $\prec_S$ guarantees that this process cannot continue indefinitely. □

Remark 2. When $S \not\cong \mathbb{N}^r$, this lemma implies the existence of Betti elements in S. Otherwise, $\mathrm{Betti}(S) = \varnothing$ because φ_A is an isomorphism.

Betti-minimal elements are characterized in the following result. As we will see later, they play an important role in the study of monoids with unique presentations.

Proposition 3. *Let S be a monoid. The element $\boldsymbol{a} \in \text{Betti-minimal}(S)$ if and only $\varphi_A^{-1}(\boldsymbol{a})$ has more than one $\mathcal{R}$-class and each $\mathcal{R}$-class is a singleton.*

Proof. First, observe that $\varphi_A^{-1}(\boldsymbol{a})$ has more than one $\mathcal{R}$-class and each $\mathcal{R}$-class is a singleton if and only if $\#\varphi_A^{-1}(\boldsymbol{a}) \geq 2$ and $\boldsymbol{u} \cdot \boldsymbol{v} = 0$ for every unequal $\boldsymbol{u}, \boldsymbol{v} \in \varphi_A^{-1}(\boldsymbol{a})$.

If $\boldsymbol{a} \in \text{Betti-minimal}(S)$ and there exist unequal $\boldsymbol{u}, \boldsymbol{v} \in \varphi_A^{-1}(\boldsymbol{a})$ such that $\boldsymbol{u} \cdot \boldsymbol{v} > 0$, we consider $\boldsymbol{a}' = \boldsymbol{a} - \sum_{i=1}^r \min(u_i, v_i)\boldsymbol{a}_i$. Since $\#\varphi_A^{-1}(\boldsymbol{a}') \geq 2$, either $\boldsymbol{a}' \prec_S \boldsymbol{a}$ is Betti or, by Lemma 1, there exist $\boldsymbol{a}'' \in \mathrm{Betti}(S)$ such that $\boldsymbol{a}'' \prec_S \boldsymbol{a}' \prec_S \boldsymbol{a}$, contradicting in both cases the Betti-minimality of $\boldsymbol{a}$. Conversely, we suppose that

$$\varphi_A^{-1}(\boldsymbol{a}) = \bigcup_{i=1}^{\#\varphi_A^{-1}(\boldsymbol{a})} \{\boldsymbol{u}^{(i)}\},$$

with $\boldsymbol{u}^{(i)} \cdot \boldsymbol{u}^{(j)} = 0$ for $i \neq j$. In particular, $\boldsymbol{a} \in \mathrm{Betti}(S)$. If $\boldsymbol{a}' \prec_S \boldsymbol{a}$, then $\#\varphi_A^{-1}(\boldsymbol{a}') = 1$; otherwise, we will find unequal i, j with $\boldsymbol{u}^{(i)} \cdot \boldsymbol{u}^{(j)} \neq 0$. Thus we conclude that $\boldsymbol{a} \in \text{Betti-minimal}(S)$. □

The notion of Betti-minimal is stronger than the notion of minimal multielement given in [Aoki et al. 2008, Definition 3.2]. Concretely, $\boldsymbol{a} \in S$ is a minimal multi-element if and only if $\varphi_A^{-1}(\boldsymbol{a})$ has more than one $\mathcal{R}$-class and at least one of them is a singleton.

3. Monoids having a unique minimal presentation

According to what we have recalled and defined so far, a monoid S has a unique minimal presentation if and only if the set of factorizations of all its Betti elements have just two $\mathcal{R}$-classes, and each of which is a singleton. Moreover, if $\boldsymbol{a}$ is a Betti element of S and $\varphi_A^{-1}(\boldsymbol{a}) = \{\boldsymbol{u}, \boldsymbol{v}\}$, then either the pair $(\boldsymbol{u}, \boldsymbol{v})$ or the pair $(\boldsymbol{v}, \boldsymbol{u})$ is in any minimal presentation of S. Hence we will say that $(\boldsymbol{u}, \boldsymbol{v}) \in \mathbb{N}^r \times \mathbb{N}^r$ is *indispensable*, and that $\boldsymbol{a}$ has *unique presentation*.

Example 4. The numerical semigroup $S = \langle 6, 10, 15\rangle$ has no indispensable elements. Using the techniques explained in [Rosales and García-Sánchez 2009], one can easily see that $\mathrm{Betti}(S) = \{30\}$, and that the factorizations of 30 are $\{(0, 0, 2), (0, 3, 0), (5, 0, 0)\}$. One can also use the `GAP` package `numericalsgps` to perform this computation [Delgado et al. 2008].

Clearly, S admits a unique minimal presentation if and only if either it is isomorphic to $\mathbb{N}^r$ for some positive integer r (and thus the empty set is its unique minimal presentation) or every element in any of its minimal presentations is indispensable. If this is the case, we say that S has a *unique presentation*.

The following results are straightforward consequences of Proposition 3.

Corollary 5. *Let $\boldsymbol{a} \in S$. The following are equivalent.*

(a) *$\boldsymbol{a}$ has unique presentation.*

(b) *$\boldsymbol{a} \in \mathrm{Betti}(S)$ and $\#\varphi_A^{-1}(\boldsymbol{a}) = 2$.*

(c) *$\boldsymbol{a} \in \mathrm{Betti\text{-}minimal}(S)$ and $\#\varphi_A^{-1}(\boldsymbol{a}) = 2$.*

Corollary 6. *A monoid S is uniquely presented if and only if either $\mathrm{Betti}(S) = \varnothing$ or the number of Betti-minimal elements in S equals the cardinality of a minimal presentation of S. In particular all Betti elements of S are Betti-minimal.*

By using the close relationship between toric ideals and semigroups, one can obtain necessary and sufficient conditions for a semigroup to be uniquely presented from the results in [Charalambous et al. 2007; Ojeda and Vigneron-Tenorio 2009; Takemura and Aoki 2004].

Example 7. Corollary 6 does not hold if we remove the minimal condition. For instance, one can use `numericalsgps` to compute that $S = \langle 4, 6, 21\rangle$ has a minimal presentation with cardinality 2, and $\mathrm{Betti}(S) = \{12, 42\}$. However, 42 admits 5 different factorizations in S.

Example 8. Let $S \subset \mathbb{Z}^r$ be a monoid minimally generated by $A = \{\boldsymbol{a}_1, \boldsymbol{a}_2\}$ for some positive integer r. If the rank of the group spanned by S is one, there exist u and $v \in \mathbb{N}$ such that $u\boldsymbol{a}_1 = v\boldsymbol{a}_2$. So, there is only one Betti element $\boldsymbol{a} = u\boldsymbol{a}_1 = v\boldsymbol{a}_2$ and $\varphi_A^{-1}(\boldsymbol{a}) = \{(u, 0), (0, v)\}$. Therefore, S is uniquely presented. In particular, embedding dimension 2 numerical semigroups are uniquely presented (the group generated by any numerical semigroup is $\mathbb{Z}$).

4. Gluings

We first fix the notation of this section. Let S be an affine semigroup generated by $A = \{\boldsymbol{a}_1, \dots, \boldsymbol{a}_r\} \subseteq \mathbb{Z}^n$. Let A_1 and A_2 be two proper subsets of A such that $A = A_1 \cup A_2$ and $A_1 \cap A_2 = \varnothing$. Let S_1 and S_2 be the affine semigroups generated by A_1 and A_2, respectively.

Set r_1 and r_2 to be the cardinality of A_1 and A_2, respectively. After rearranging the elements of A if necessary, we may assume that $A_1 = \{\boldsymbol{a}_1, \dots, \boldsymbol{a}_{r_1}\}$ and $A_2 = \{\boldsymbol{a}_{r_1+1}, \dots, \boldsymbol{a}_r\}$.

Since $\mathbb{N}^r = \mathbb{N}^{r_1} \oplus \mathbb{N}^{r_2}$, elements in $\mathbb{N}^{r_1}$ and $\mathbb{N}^{r_2}$ may be regarded as elements in $\mathbb{N}^r$ of the form $(\cdot, 0)$ and $(0, \cdot)$, respectively. With this in mind, subsets of $\mathbb{N}^{r_i}$ will be considered as subsets of $\mathbb{N}^r$ for $i \in \{1, 2\}$, and the elements of $\sim_{A_1}$ and $\sim_{A_2}$ are viewed inside $\sim_A$.

The monoid S is said to be the *gluing* of S_1 and S_2 if $G(S_1) \cap G(S_2) = \boldsymbol{d}\mathbb{Z}$, with $\boldsymbol{d} \in S_1 \cap S_2 \setminus \{0\}$, where G denotes the group generated by its argument.

According to [Rosales 1997, Theorem 1.4], S admits a presentation of the form $\rho_1 \cup \rho_2 \cup \{((\boldsymbol{u}, 0), (0, \boldsymbol{v}))\}$, where ρ_1 and ρ_2 are presentations of S_1 and S_2, respectively, and $\boldsymbol{u} \in \varphi_{A_1}^{-1}(\boldsymbol{d})$ and $\boldsymbol{v} \in \varphi_{A_2}^{-1}(\boldsymbol{d})$. We next explore the conditions that we must impose on S_1, S_2 and $\boldsymbol{d}$ to ensure that S has a unique minimal presentation. We start by describing the Betti elements of S, and for this we need a lemma describing the factorizations of $\boldsymbol{d}$.

Lemma 9. *Let S be the gluing of S_1 and S_2 with $G(S_1) \cap G(S_2) = \boldsymbol{d}\mathbb{Z}$. Every factorization of $\boldsymbol{d}$ in S is either a factorization of $\boldsymbol{d}$ in S_1 or a factorization of $\boldsymbol{d}$ in S_2. In particular $\boldsymbol{d} \in$* Betti(S).

Proof. By definition $\boldsymbol{d} \in S_1 \cap S_2 \setminus \{0\}$, so there exist $\boldsymbol{u} \in \mathbb{N}^{r_1}$ and $\boldsymbol{v} \in \mathbb{N}^{r_2}$ such that $\boldsymbol{d} = \sum_{i=1}^{r_1} u_i \boldsymbol{a}_i = \sum_{i=r_1+1}^{r} v_i \boldsymbol{a}_i$. If $\boldsymbol{d} = \sum_{i=1}^{r} w_i \boldsymbol{a}_i = \sum_{i=1}^{r_1} w_i \boldsymbol{a}_i + \sum_{i=r_1+1}^{r} w_i \boldsymbol{a}_i$, then

$$\boldsymbol{d} - \sum_{i=1}^{r_1} w_i \boldsymbol{a}_i = \sum_{i=1}^{r_1} u_i \boldsymbol{a}_i - \sum_{i=1}^{r_1} w_i \boldsymbol{a}_i = \sum_{i=r_1+1}^{r} w_i \boldsymbol{a}_i \in G(S_1) \cap G(S_2),$$

that is, $\boldsymbol{d} - \sum_{i=1}^{r_1} w_i \boldsymbol{a}_i = z\boldsymbol{d}$. Hence either $z = 1$ and then $w_i = 0$ for $i \in \{1, \dots, r_1\}$, or $z = 0$ and then $w_i = 0$ for $i \in \{r_1 + 1, \dots, r\}$, as claimed.

Also, we have $\varphi_A^{-1}(\boldsymbol{d}) = \varphi_{A_1}^{-1}(\boldsymbol{d}) \cup \varphi_{A_2}^{-1}(\boldsymbol{d})$ with $(\boldsymbol{u}, 0) \cdot (0, \boldsymbol{v}) = 0$ for every $\boldsymbol{u} \in \varphi_{A_1}^{-1}(\boldsymbol{d})$ and $\boldsymbol{v} \in \varphi_{A_2}^{-1}(\boldsymbol{d})$, which means that $\varphi_A^{-1}(\boldsymbol{d})$ has at least two $\mathcal{R}$-classes. Hence $\boldsymbol{d} \in \mathrm{Betti}(S)$. □

Theorem 10. *Let S be the gluing of S_1 and S_2, and $G(S_1) \cap G(S_2) = \boldsymbol{d}\mathbb{Z}$. Then*

$$\mathrm{Betti}(S) = \mathrm{Betti}(S_1) \cup \mathrm{Betti}(S_2) \cup \{\boldsymbol{d}\}.$$

Proof. By [Rosales 1997, Theorem 1.4], S admits a presentation of the form $\rho = \rho_1 \cup \rho_2 \cup \{((\boldsymbol{u}, 0), (0, \boldsymbol{v}))\}$, where ρ_1 and ρ_2 are sets of generators for $\sim_{A_1}$ and $\sim_{A_2}$, respectively, and $\varphi_{A_1}(\boldsymbol{u}) = \varphi_{A_2}(\boldsymbol{v}) = \boldsymbol{d}$. Since every system of generators of $\sim_A$ can be refined to a minimal system of generators [Rosales and García-Sánchez 1999a, Chapter 9], from the shape of ρ we deduce that the Betti elements of S are either a Betti element of S_1, a Betti element of S_2, or $\boldsymbol{d}$ itself, that is, $\mathrm{Betti}(S) \subseteq \mathrm{Betti}(S_1) \cup \mathrm{Betti}(S_2) \cup \{\boldsymbol{d}\}$.

Recall that $\boldsymbol{d} \in \mathrm{Betti}(S)$ by Lemma 9. Therefore, to demonstrate the inclusion $\mathrm{Betti}(S) \supseteq \mathrm{Betti}(S_1) \cup \mathrm{Betti}(S_2) \cup \{\boldsymbol{d}\}$, it suffices to prove $\mathrm{Betti}(S_1) \cup \mathrm{Betti}(S_2) \subseteq \mathrm{Betti}(S)$. Suppose by way of contradiction that there is a $\boldsymbol{b}$ in $\mathrm{Betti}(S_1) \setminus \mathrm{Betti}(S)$ (the case where $\boldsymbol{b}$ is in $\mathrm{Betti}(S_2) \setminus \mathrm{Betti}(S)$ is argued similarly).

Since $\boldsymbol{b} \in \mathrm{Betti}(S_1)$, there exist two $\mathcal{R}$-classes in $\varphi_{A_1}^{-1}(\boldsymbol{b})$, say $\mathscr{C}_1$ and $\mathscr{C}_2$. We know because $\boldsymbol{b} \notin \mathrm{Betti}(S)$ that $\varphi_A^{-1}(\boldsymbol{b})$ has only one $\mathcal{R}$-class. Hence:

- There exist $\boldsymbol{w} \in \mathscr{C}_1$ and $\overline{\boldsymbol{w}} \in \varphi_A^{-1}(\boldsymbol{b})$ such that $\overline{\boldsymbol{w}} \cdot (\boldsymbol{w}, 0) \neq 0$ and $\boldsymbol{b} = \sum_{i=1}^{r_1} \overline{w}_i \boldsymbol{a}_i + \sum_{i=r_1+1}^{r} \overline{w}_i \boldsymbol{a}$, where $\overline{w}_i$ for $1 \leq i \leq r$ are the coordinates of $\overline{\boldsymbol{w}}$ and $\overline{w}_i \neq 0$ for some $r_1 + 1 \leq i \leq r$.
- There exist $\boldsymbol{w}' \in \mathscr{C}_2$ and $\overline{\boldsymbol{w}}' \in \varphi_A^{-1}(\boldsymbol{b})$ such that $\overline{\boldsymbol{w}}' \cdot (\boldsymbol{w}', 0) \neq 0$ and $\boldsymbol{b} = \sum_{i=1}^{r_1} \overline{w}'_i \boldsymbol{a}_i + \sum_{i=r_1+1}^{r} \overline{w}'_i \boldsymbol{a}_i$, where $\overline{w}'_i$ for $1 \leq i \leq r$ are the coordinates of $\overline{\boldsymbol{w}}'$ and $\overline{w}'_i \neq 0$ for some $r_1 + 1 \leq i \leq r$.

Since $0 \neq \boldsymbol{b} - \sum_{i=1}^{r_1} \overline{w}_i \boldsymbol{a}_i = \sum_{i=r_1+1}^{r} \overline{w}_i \boldsymbol{a}_i \in G(S_1) \cap G(S_2) = \boldsymbol{d}\mathbb{Z}$, we have

$$\boldsymbol{b} = \sum_{i=1}^{r_1} \overline{w}_i \boldsymbol{a}_i + \sum_{i=1}^{r_1} z u_i \boldsymbol{a}_i = \sum_{i=1}^{r_1} (\overline{w}_i + z u_i) \boldsymbol{a}_i \quad \text{for some } z > 0.$$

Analogously, $\boldsymbol{b} = \sum_{i=1}^{r_1} (\overline{w}'_i + z' u_i) \boldsymbol{a}_i$ for some $z' > 0$.

Let $\tilde{\boldsymbol{w}}$ and $\tilde{\boldsymbol{w}}' \in \varphi_{A_1}^{-1}(\boldsymbol{b})$ be the corresponding vectors of coordinates $\overline{w}_i + z u_i$ for $1 \leq i \leq r_1$ and $\overline{w}'_i + z' u_i$ for $1 \leq i \leq r_1$, respectively. This yields a contradiction, since $\boldsymbol{w}$ and $\boldsymbol{w}'$ are not $\mathcal{R}$-related; however $\boldsymbol{w} \cdot \tilde{\boldsymbol{w}} \neq 0$, $\tilde{\boldsymbol{w}} \cdot \tilde{\boldsymbol{w}}' \neq 0$ and $\tilde{\boldsymbol{w}}' \cdot \boldsymbol{w}' \neq 0$. □

Observe that $\varphi_A^{-1}(\boldsymbol{d}) \supseteq \{(\boldsymbol{u}, 0), (0, \boldsymbol{v})\}$, with $\varphi_{A_1}(\boldsymbol{u}) = \varphi_{A_2}(\boldsymbol{v}) = \boldsymbol{d}$, and that the equality holds if and only if $\boldsymbol{d}$ has unique presentation as an element of S.

Corollary 11. *Let S be the gluing of S_1 and S_2, and let $G(S_1) \cap G(S_2) = \boldsymbol{d}\mathbb{Z}$. Then the element $\boldsymbol{d}$ in S has unique presentation if and only if $\boldsymbol{d} - \boldsymbol{a} \notin S$ for every $\boldsymbol{a} \in \mathrm{Betti}(S_1) \cup \mathrm{Betti}(S_2)$.*

Proof. If $\boldsymbol{d}$ has unique presentation, $\boldsymbol{d}$ belongs to Betti-minimal(S) by Corollary 5. So $\boldsymbol{d}-\boldsymbol{a} \notin S$ for every $\boldsymbol{a} \in \text{Betti}(S)\setminus\{\boldsymbol{d}\}$. Now $\boldsymbol{d} \notin \text{Betti}(S_1)\cup\text{Betti}(S_2)$ since $\boldsymbol{d}$ has unique factorization in S_i for $i \in \{1, 2\}$. Hence $\text{Betti}(S)\setminus\{\boldsymbol{d}\} = \text{Betti}(S_1)\cup\text{Betti}(S_2)$ by Theorem 10. We conclude that $\boldsymbol{d} - \boldsymbol{a} \notin S$ for every $\boldsymbol{a} \in \text{Betti}(S_1) \cup \text{Betti}(S_2)$.

Conversely, in view of Lemma 1, we deduce that $\boldsymbol{d}$ admits a unique factorization in S_i for $i \in \{1, 2\}$, that is, $\varphi_{A_1}^{-1}(\boldsymbol{d}) = \{\boldsymbol{u}\}$ and $\varphi_{A_2}^{-1}(\boldsymbol{d}) = \{\boldsymbol{v}\}$. Since $\boldsymbol{d}$ is a Betti element by Lemma 9, we conclude that $\varphi_{A}^{-1}(\boldsymbol{d}) = \{(\boldsymbol{u}, 0), (0, \boldsymbol{v})\}$. □

Theorem 12. *Let S be the gluing of S_1 and S_2, and $G(S_1) \cap G(S_2) = \boldsymbol{d}\mathbb{Z}$. Then S is uniquely presented if and only if*

(a) *S_1 and S_2 are uniquely presented, and*

(b) *$\pm(\boldsymbol{d} - \boldsymbol{a}) \notin S$ for every $\boldsymbol{a} \in \text{Betti}(S_1) \cup \text{Betti}(S_2)$,*

Proof. By Theorem 10, $\text{Betti}(S) = \text{Betti}(S_1) \cup \text{Betti}(S_2) \cup \{\boldsymbol{d}\}$. So, if S is uniquely presented, then every $\boldsymbol{a} \in \text{Betti}(S_1)\cup\text{Betti}(S_2)\cup\{\boldsymbol{d}\}$ has unique presentation. Thus, S_1 and S_2 are uniquely presented and $\boldsymbol{d} - \boldsymbol{a} \notin S$ for every $\boldsymbol{a} \in \text{Betti}(S_1) \cup \text{Betti}(S_2)$ by Corollary 11. Finally, since, by Corollary 5, every $\boldsymbol{a} \in \text{Betti}(S)$ is Betti-minimal, we conclude that $\boldsymbol{a}-\boldsymbol{d} \notin S$, for every $\boldsymbol{a} \in \text{Betti}(S_1)\cup\text{Betti}(S_2)$. (Note that $\boldsymbol{d}-\mathbf{m} \notin S$ implies $\boldsymbol{d} \neq \mathbf{m}$ for every $\mathbf{m} \in \text{Betti}(S_1) \cup \text{Betti}(S_2)$.)

Conversely, suppose that (a) and (b) hold. In particular, every $\boldsymbol{a} \in \text{Betti}(S_i)$ has only two factorizations as element of S_i for $i \in \{1, 2\}$ and, by Corollary 11, $\boldsymbol{d}$ has only two factorizations in S, say $\boldsymbol{d} = \sum_{i=1}^{r_1} u_i \boldsymbol{a}_i = \sum_{i=r_1+1}^{r} v_i \boldsymbol{a}_i$. So, if $\boldsymbol{a} \in \text{Betti}(S)$ has more than two factorizations in S, then $\boldsymbol{d} \neq \boldsymbol{a} \in \text{Betti}(S_1) \cup \text{Betti}(S_2)$. If $\boldsymbol{a} \in \text{Betti}(S_1)$, then $\boldsymbol{a} = \sum_{i=1}^{r_1} w_i \boldsymbol{a}_i + \sum_{i=r_1+1}^{r} w_i \boldsymbol{a}_i$, with $w_i \neq 0$ for some i such that $r_1 + 1 \leq i \leq r$. Thus, $\boldsymbol{a} - \sum_{i=1}^{r_1} w_i \boldsymbol{a}_i = \sum_{i=r_1+1}^{r} w_i \boldsymbol{a}_i \in G(S_1) \cap G(S_2) = \boldsymbol{d}\mathbb{Z}$ and thus $\boldsymbol{a} - \boldsymbol{d} \in S$, which is impossible by hypothesis. □

The affine semigroup in the next example is borrowed from [Rosales and García-Sánchez 1999b], where the authors use it to illustrate their algorithm for checking freeness of simplicial semigroups. We use $\mathbf{e}_i \in \mathbb{N}^r$ to denote the i-th row of the identity $r \times r$ matrix.

Example 13. Let us see that $S = \langle(2, 0), (0, 3), (2, 1), (1, 2)\rangle$ is uniquely presented. On the one hand, by taking $A_1 = \{(2, 0), (0, 3), (2, 1)\}$, $A_2 = \{(1, 2)\}$, $S_1 = \langle A_1\rangle$ and $S_2 = \langle A_2\rangle$, we have $G(S_1)\cap G(S_2) = 2(1, 2)\mathbb{Z}$. On the other hand, by taking $A_{11} = \{(2, 0), (0, 3)\}$, $A_{12} = \{(2, 1)\}$, $S_{11} = \langle A_{11}\rangle$ and $S_{12} = \langle A_{12}\rangle$, we have $G(S_{11}) \cap G(S_{12}) = 3(2, 1)\mathbb{Z}$. Since $S_{11} \cong \mathbb{N}^2$ and $S_{12} \cong \mathbb{N}$ are uniquely presented (because, their corresponding presentations are the empty set) and condition (b) in Theorem 12 is trivially satisfied, we are assured that S_1 is uniquely presented by $\{(3\mathbf{e}_3, 3\mathbf{e}_1 + \mathbf{e}_2)\}$. Finally, since S_1 and $S_2 \cong \mathbb{N}$ are uniquely presented and the element $2(1, 2) - 3(2, 1)$ is not in S, we conclude that S is uniquely presented by $\{(3\mathbf{e}_3, 3\mathbf{e}_1 + \mathbf{e}_2),\ (2\mathbf{e}_4, \mathbf{e}_2 + \mathbf{e}_3)\}$.

Example 14. We may construct an infinite sequence of uniquely presented numerical semigroups. Let us start with $S_1 = \langle 2, 3\rangle$, and given S_i minimally generated by $\{a_1, \dots, a_{i+1}\}$, $i \geq 2$, set $S_{i+1} = \langle 2a_1, a_1 + a_2, 2a_2, \dots, 2a_{i+1}\rangle$. We prove by induction on i that S_{i+1} is uniquely presented by

$$\rho_{i+1} = \{(2\mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_3), (2\mathbf{e}_3, \mathbf{e}_1 + \mathbf{e}_4), \dots, (2\mathbf{e}_i, \mathbf{e}_1 + \mathbf{e}_i), (2\mathbf{e}_{i+1}, 3\mathbf{e}_1)\}.$$

For $i = 1$, the result follows easily. Assume that $i \geq 2$ and that the result holds for S_i, and let us show it holds for S_{i+1}. Observe that S_{i+1} is the gluing of $\langle 2a_1, \dots, 2a_{i+1}\rangle = 2S_i$ and $\langle a_1 + a_2\rangle$, with $d = 2a_1 + 2a_2$, and consequently S_{i+1} is minimally generated by $\{2a_1, a_1 + a_2, 2a_2, \dots, 2a_{i+1}\}$; apply [Rosales and García-Sánchez 2009, Lemma 9.8] with $\lambda = 2$ and $\mu = a_1 + a_2$. Note that $\mathrm{Betti}(\langle a_1 + a_2\rangle) = \varnothing$ and, by induction hypothesis, $\mathrm{Betti}(2S_i) = 2\mathrm{Betti}(S_i) = \{2(2a_2), \dots, 2(2a_{i+1})\}$. Thus, by Theorem 10,

$$\mathrm{Betti}(S_{i+1}) = \{d\} \cup \mathrm{Betti}(2S_i) = \{2a_1 + 2a_2, 2(2a_2), \dots, 2(2a_{i+1})\}.$$

Now, a direct computation shows that ρ_{i+1} is a minimal presentation of S_{i+1}.

In view of Theorem 12, it suffices to prove the uniqueness of the presentation to check that, for $b = 2(2a_j) - (2a_1 + 2a_2)$, neither b nor $-b$ belongs to S_{i+1}. Observe that $-b < 0$ since $j \geq 2$, and thus it is not in S_{i+1}. Also, if $j \neq i$, then $2(2a_j) - (2a_1 + 2a_2) = 2a_1 + 2a_{j+1} - 2a_1 - 2a_2 = 2a_{j+1} - 2a_2$. This element cannot be in S_{i+1} because $2a_{j+1}$ is one of its minimal generators. For $j = i$, we get $2(2a_{i+1}) - (2a_1 + 2a_2) = 2(3a_1) - 2a_1 - 2a_2 = 2(2a_1) - 2a_2$. If this integer belongs to S_{i+1}, then by the minimality of $2a_2$, there exists $a \in S_{i+1} \setminus \{0\}$ such that $2(2a_1) = 2a_2 + a$. But then $a \geq 2a_1$, and since $2a_2 > 2a_1$, we get a contradiction.

For every positive integer i, the numerical semigroup S_{i+1} is a free numerical semigroup in the sense of [Bertin and Carbonne 1977], and thus it is a complete intersection, that is, a numerical semigroup with minimal presentations with the least possible cardinality, the embedding dimension minus one. Some authors call these semigroups telescopic. Not all free numerical semigroups have unique minimal presentation; $\langle 4, 6, 21\rangle$ illustrates this fact (see Example 7).

5. Uniquely presented numerical semigroups

In some sense, only a few numerical semigroups have unique minimal presentation. The following sequences have been computed with the `numericalsgps GAP` package [Delgado et al. 2008]. The first contains in the i-th position the number of numerical semigroups with *Frobenius number* $i \in \{1, \dots, 20\}$, meaning that i is the largest integer not in the semigroup. The second contains those with the same condition having a unique minimal presentation.

$$(1, 1, 2, 2, 5, 4, 11, 10, 21, 22, 51, 40, 106, 103, 200, 205, 465, 405, 961, 900),$$
$$(1, 1, 1, 1, 3, 1, 5, 2, 5, 4, 8, 2, 12, 8, 6, 9, 17, 8, 20, 12).$$

Next we explore three big families of numerical semigroups and determine the elements having unique minimal presentations.

5.1. ***Numerical semigroups generated by intervals.*** Let a and x be two positive integers, and let $S = \langle a, a+1, \dots, a+x\rangle$. Since $\mathbb{N}$ is uniquely presented, we may assume that $2 \le a$. In order that $\{a, \dots, a+x\}$ becomes a minimal system of generators for S, we suppose that $x < a$.

Theorem 15. *$S = \langle a, a+1, \dots, a+x\rangle$ is uniquely presented if and only if either*

$a = 1$ (*that is*, $S = \mathbb{N}$) *or* $x = 1$ *or* $x = 2$ *or* $x = 3$ *and* $(a-1) \bmod x \neq 0$.

Proof. The Betti elements in S are fully described in [García-Sánchez and Rosales 1999, Theorem 8]. If $x \ge 4$, then $m = 2(a+2)$ is a Betti element and $\#\varphi_A^{-1}(m) = 3$. Thus S is not uniquely presented for $x \ge 4$. Hence we focus on $x \in \{1, 2, 3\}$. For simplicity in the forthcoming notation, let q and r be the quotient and the remainder in the division of $a-1$ by x, that is, $a = xq + r + 1$ with $0 \le r \le x-1$. Notice that $x < a$ implies $q \ge 1$.

For $x = 1$, we get an embedding dimension two numerical semigroup that is uniquely presented; see Example 8.

For $x = 2$,

$$\mathrm{Betti}(S) = \begin{cases} \{2(a+1), qa + 2(q-1) + 1, qa + 2(q-1) + 2\} & \text{if } r = 0, \\ \{2(a+1), qa + 2(q-1) + 2\} & \text{if } r = 1. \end{cases}$$

Since the cardinality of a minimal presentation of S is $3 - r$ [ibid., Theorem 8], by Corollary 6 we only must check whether they are incomparable with respect to $\prec_S$. If $r = 0$, clearly $qa + 2(q-1) + 1$ and $qa + 2(q-1) + 2$ are incomparable, since $1 \notin S$. Also,

$$qa + 2(q-1) + 1 - 2(a+1) = (q-1)a + 2q - 1 \notin S$$

in view of [ibid., Lemma 1] (since $2q - 1 > 2(q-1)$). The same argument applies to $qa + 2(q-1) + 2 - 2(a+1) = (q-1)a + 2q$. If $r = 1$, then

$$qa + 2(q-1) + 2 - 2(a+1) = (q-2)a + 2(q-1) \notin S$$

(we use again [ibid., Lemma 1]), so we also get a (complete intersection) uniquely presented numerical semigroup. Hence every numerical semigroup of the form $\langle a, a+1, a+2\rangle$ with $a \ge 3$ is uniquely presented.

Assume that $x = 3$ (and thus $a \ge 4$).

Case: $r = 0$. In this setting, both $(q+1)(a+3)$ and $2(a+1)$ are Betti elements. However, $(q+1)(a+3)-2(a+1) = (q-1)a+q3+1 = (q-1)a+(a-1)+1 = qa \in S$. Hence $(q+1)(a+3) \notin \text{Betti-minimal}(S)$ and so, by Corollary 6, it is not uniquely presented.

Case: $r \neq 0$. In this case,

$$\text{Betti}(S) = \begin{cases} \{2(a+1),\ (a+1)+(a+2), 2(a+2), \\ \quad qa+3(q-1)+2,\ qa+3(q-1)+3\} & \text{if } r=1, \\ \{2(a+1),\ (a+1)+(a+2), \\ \quad 2(a+2),\ qa+3(q-1)+3\} & \text{if } r=2. \end{cases}$$

Since the cardinality of a minimal presentation of S is $6-r$ [García-Sánchez and Rosales 1999, Theorem 8], by Corollary 6 we only must check whether they are incomparable with respect to $\prec_S$. Observe that

$$qa+(q-1)3+j-2a-i = (q-2)a+(q-1)3+j-i \notin S$$

if and only if $q+j+1 > i$ [ibid., Lemma 1]. Since in our case $i \in \{2,3,4\}$, $j \in \{2,3\}$ and $q \geq 1$, we obtain that these elements are incomparable. Thus, S is uniquely presented. □

5.2. *Embedding dimension three numerical semigroups.* As we have pointed out above, the Frobenius number of a numerical semigroup is the largest integer not belonging to it. A numerical semigroup S with Frobenius number f is *symmetric* if $f-x \in S$ for every $x \in \mathbb{Z} \setminus S$. For embedding dimension three numerical semigroups, it is well known that the concepts of symmetric and complete intersection numerical semigroups coincide. (In the embedding dimension three case, the concept of free also coincides with that of symmetric and complete intersection; see for instance [Rosales and García-Sánchez 2009, Chapter 9] or [Herzog 1970].) Nonsymmetric numerical semigroups with embedded dimension three are uniquely presented [Herzog 1970]. Thus, we will focus on the symmetric case, which is the free case, and as Delorme [1976] proved, these semigroups are the gluing of an embedding dimension two numerical semigroup and $\mathbb{N}$; see [Rosales 1997] for a proof using the concept of gluing. So every symmetric numerical semigroup with embedding dimension three can be described as follows.

Proposition 16 [Rosales and García-Sánchez 2009, Theorem 10.6]. *Let m_1 and m_2 be two relatively prime integers greater than one. Let a, b and c be nonnegative integers with $a \geq 2$, $b+c \geq 2$ and $\gcd(a, bm_1+cm_2) = 1$. Then $S = \langle am_1, am_2, bm_1+cm_2\rangle$ is a symmetric numerical semigroup with embedding dimension three. Every embedding dimension three symmetric numerical semigroup is of this form.*

Our main result is now just a special case of what we have seen in Section 4.

Theorem 17. *In the notation of Proposition 16, S is a symmetric numerical semigroup uniquely presented with embedding dimension three if and only if* $0 < b < m_2$ *and* $0 < c < m_1$.

Lemma 18. *Let m_1 and m_2 be two relatively prime integers greater than one. Then, $m_1 m_2 = \alpha m_1 + \beta m_2$ for some $\alpha \geq 0$ and $\beta \geq 0$ if and only if $\alpha = m_2$ and $\beta = 0$, or $\alpha = 0$ and $\beta = m_1$.*

Proof. We have $m_1 m_2 = \alpha m_1 + \beta m_2$ for some $\alpha \geq 0$ and $\beta \geq 0$ if and only if $(m_2 - \alpha) m_1 = \beta m_2$ for some $\alpha \geq 0$ and $\beta \geq 0$. Since $\gcd(m_1, m_2) = 1$, it follows that $(m_2 - \alpha) m_1 = \beta m_2$ for some $\alpha \geq 0$ and $\beta \geq 0$, if and only if $m_2 - \alpha = \gamma m_2$ and $\beta = \gamma m_1$ for some $\gamma \geq 0$, if and only if $\alpha = (1 - \gamma) m_2$ and $\beta = \gamma m_1$ for some $0 \leq \gamma \leq 1$, if and only if $\alpha = m_2$ and $\beta = 0$ or $\alpha = 0$ and $\beta = m_1$. □

Proof of Theorem 17. S is the gluing of $S_1 = \langle a m_1, a m_2 \rangle$ and $S_2 = \langle b m_1 + c m_2 \rangle$ with $d = a(b m_1 + c m_2)$. Also $\mathrm{Betti}(S_1) = a m_1 m_2$ and $\mathrm{Betti}(S_2) = \varnothing$. Therefore, by Theorem 10, $\mathrm{Betti}(S) = \{a m_1 m_2, a(b m_1 + c m_2)\}$. Thus, by Theorem 12, S is uniquely presented if and only if $\pm(a m_1 m_2 - a(b m_1 + c m_2)) \notin S$.

By direct computation, one can check that $a(b m_1 + c m_2) - a m_1 m_2 \in S$ if and only if $b \geq m_2$ or $c \geq m_1$. Also, $a m_1 m_2 - a(b m_1 + c m_2) \in S$ if and only if

$$m_1 m_2 = ((\alpha_3 + 1) b + \alpha_1) m_1 + ((\alpha_3 + c) c + \alpha_1) m_2$$

for some $\alpha_i \geq 0$, with $i \in \{1, 2, 3\}$. In view of Lemma 18, this is equivalent to $((\alpha_3 + 1) b + \alpha_1) = 0$ and $((\alpha_3 + c) c + \alpha_1) = m_1$ or $((\alpha_3 + 1) b + \alpha_1) = m_2$ and $((\alpha_3 + c) c + \alpha_1) = 0$ for some $\alpha_i \geq 0$ with $i \in \{1, 2, 3\}$. This holds if and only if $b = 0$ and $c \leq m_1$ or $b \leq m_2$ and $c = 0$.

Therefore, $\pm(a m_1 m_2 - a(b m_1 + c m_2)) \notin S$ if and only if $0 < b < m_2$ and $0 < c < m_1$. □

5.3. *Maximal embedding dimension numerical semigroups.*

Theorem 19. *A numerical semigroup S minimally generated by $a_1 < a_2 < \cdots < a_r$ with $a_1 = r$ is uniquely presented if only if $r = 3$.*

Proof. For $r = 3$, we obtain numerical semigroups of the form $\langle 3, a, b \rangle$, with a and b not multiples of 3 and thus coprime with 3. It follows easily that these semigroups do not have the shape given in Proposition 16, and thus are not symmetric. Consequently, they are uniquely presented.

We now prove that S cannot be uniquely presented if $a_1 = r \geq 4$. According to [Rosales 1996], $\mathrm{Betti}(S) = \{a_i + a_j \mid i, j \in \{2, \ldots, r\}\}$. All the elements in $\{0, a_2, \ldots, a_r\}$ belong to different classes modulo a_1, and there are precisely a_1 of them. Thus $2a_r$ can be uniquely be written as $b a_1 + a_i$ for some $i \in \{2, \ldots, r - 1\}$ and b a positive integer.

Let f be the Frobenius number of S. It is well known that $f = a_r - a_1$ in this setting; see for instance [Rosales and García-Sánchez 2009]. Since $2a_r - a_i = a_r + (a_r - a_i) > a_r - a_1 = f$ for all i, it follows that $2a_r - a_i \in S$. Hence $2a_r = a_i + m_i$ for some $m_i \in S$ for every $i \in \{1, \ldots, r\}$. Take $i \neq k$. Then $2a_r$ admits at least three expressions: $2a_r$, $ba_1 + a_k$ and $a_i + m_i$. By Corollary 5, S cannot have a unique minimal presentation. □

Acknowledgments

We gratefully thank Anargyros Katsabekis for a careful reading and correcting some misprints; we thank the referee for valuable remarks.

Part of this work was done during a visit of García-Sánchez to the University of Extremadura financed by the Plan Propio 2009 of the University of Extremadura.

References

[Aoki et al. 2008] S. Aoki, A. Takemura, and R. Yoshida, "Indispensable monomials of toric ideals and Markov bases", *J. Symbolic Comput.* **43** (2008), 490–507. MR 2009c:13065 Zbl 1170.13008

[Bertin and Carbonne 1977] J. Bertin and P. Carbonne, "Semi-groupes d'entiers et application aux branches", *J. Algebra* **49**:1 (1977), 81–95. MR 58 #27957 Zbl 0498.14016

[Briales et al. 1998] E. Briales, A. Campillo, C. Marijuán, and P. Pisón, "Minimal systems of generators for ideals of semigroups", *J. Pure Appl. Algebra* **124**:1-3 (1998), 7–30. MR 98k:20105 Zbl 0913.20036

[Bruns and Herzog 1993] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, Cambridge Studies in Advanced Mathematics **39**, Cambridge University Press, 1993. MR 95h:13020 Zbl 0788.13005

[Charalambous et al. 2007] H. Charalambous, A. Katsabekis, and A. Thoma, "Minimal systems of binomial generators and the indispensable complex of a toric ideal", *Proc. Amer. Math. Soc.* **135**:11 (2007), 3443–3451. MR 2009a:13033 Zbl 1127.13018

[Delgado et al. 2008] M. Delgado, P. A. García-Sánchez, and J. Morais, *GAP package NumericalSgps: A package for numerical semigroups*, 2008, available at http://www.gap-system.org/Packages/numericalsgps.html.

[Delorme 1976] C. Delorme, "Sous-monoïdes d'intersection complète de N", *Ann. Sci. École Norm. Sup.* (4) **9**:1 (1976), 145–154. MR 53 #10821 Zbl 0325.20065

[Fulton 1993] W. Fulton, *Introduction to toric varieties*, Annals of Mathematics Studies **131**, Princeton University Press, 1993. MR 94g:14028 Zbl 0813.14039

[García-Sánchez and Rosales 1999] P. A. García-Sánchez and J. C. Rosales, "Numerical semigroups generated by intervals", *Pacific J. Math.* **191**:1 (1999), 75–83. MR 2000i:20095 Zbl 1009.20069

[Geroldinger and Halter-Koch 2006] A. Geroldinger and F. Halter-Koch, *Non-unique factorizations: Algebraic, combinatorial and analytic theory*, Pure and Applied Mathematics (Boca Raton) **278**, Chapman & Hall/CRC, 2006. MR 2006k:20001 Zbl 1113.11002

[Herzog 1970] J. Herzog, "Generators and relations of abelian semigroups and semigroup rings", *Manuscripta Math.* **3** (1970), 175–193. MR 42 #4657 Zbl 0211.33801

[Ojeda 2008] I. Ojeda Martínez de Castilla, "Examples of generic lattice ideals of codimension 3", *Comm. Algebra* **36**:1 (2008), 279–287. MR 2008j:13027 Zbl 1133.13014

[Ojeda and Vigneron-Tenorio 2009] I. Ojeda and A. Vigneron-Tenorio, "Indispensable binomials in semigroups ideals", *Proc. Amer. Math. Soc.* **138** (2009), 4205–4216. arXiv 0903.1030

[Ojeda and Vigneron-Tenorio 2010] I. Ojeda and A. Vigneron-Tenorio, "Simplicial complexes and minimal free resolution of monomial algebras", *J. Pure Appl. Algebra* **214**:6 (2010), 850–861. MR 2580663 Zbl 05673207

[Rédei 1965] L. Rédei, *The theory of finitely generated commutative semigroups*, Pergamon, Oxford, 1965. MR 32 #5761 Zbl 0133.27904

[Rosales 1996] J. C. Rosales, "On numerical semigroups", *Semigroup Forum* **52**:3 (1996), 307–318. MR 96m:20092 Zbl 0853.20041

[Rosales 1997] J. C. Rosales, "On presentations of subsemigroups of $\mathbb{N}^n$", *Semigroup Forum* **55**:2 (1997), 152–159. MR 98h:20104 Zbl 0951.20042

[Rosales and García-Sánchez 1999a] J. C. Rosales and P. A. García-Sánchez, *Finitely generated commutative monoids*, Nova Science Publishers, Commack, NY, 1999. MR 2000d:20074 Zbl 0966.20028

[Rosales and García-Sánchez 1999b] J. C. Rosales and P. A. García-Sánchez, "On free affine semigroups", *Semigroup Forum* **58**:3 (1999), 367–385. MR 2000j:13020 Zbl 0934.20041

[Rosales and García-Sánchez 2009] J. C. Rosales and P. A. García-Sánchez, *Numerical semigroups*, Developments in Mathematics **20**, Springer, New York, 2009. MR 2549780 Zbl 05623301

[Rosales et al. 1999] J. C. Rosales, P. A. García-Sánchez, and J. M. Urbano-Blanco, "On presentations of commutative monoids", *Internat. J. Algebra Comput.* **9** (1999), 539–553. MR 2000h:20102 Zbl 1028.20037

[Sturmfels 1996] B. Sturmfels, *Gröbner bases and convex polytopes*, University Lecture Series **8**, American Mathematical Society, Providence, RI, 1996. MR 97b:13034 Zbl 0856.13020

[Takemura and Aoki 2004] A. Takemura and S. Aoki, "Some characterizations of minimal Markov basis for sampling from discrete conditional distributions", *Ann. Inst. Statist. Math.* **56**:1 (2004), 1–17. MR 2005g:62103 Zbl 1049.62068

Received July 24, 2009. Revised May 6, 2010.

PEDRO A. GARCÍA-SÁNCHEZ
DEPARTAMENTO DE ÁLGEBRA
UNIVERSIDAD DE GRANADA
AV. FUENTENUEVA, S/N
E18071 GRANADA
SPAIN
pedro@ugr.es
http://www.ugr.es/~pedro/

IGNACIO OJEDA
DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD DE EXTREMADURA
AV. DE ELVAS, S/N
E06071 BADAJOZ
SPAIN
ojedamc@unex.es
http://matematicas.unex.es/~ojedamc/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

THE UNITARY DUAL OF p-ADIC $\widetilde{\mathrm{Sp}(2)}$

MARCELA HANZER AND IVAN MATIĆ

We investigate the composition series of the induced admissible representations of the metaplectic group $\widetilde{\mathrm{Sp}(2)}$ over a p-adic field F. In this way, we determine the nonunitary and unitary duals of $\widetilde{\mathrm{Sp}(2)}$ modulo cuspidal representations.

1. Introduction

The admissible representations of reductive groups over p-adic fields have been studied intensively by many authors, but knowledge about the unitary dual of such groups is still incomplete. Besides some results concerning specific parts of the unitary dual of some classical and exceptional groups (that is, spherical, generic [Lapid et al. 2004] and so on), there are some situations where, for some low rank groups, the complete unitary dual is described [Sally and Tadić 1993; Muić 1997; Hanzer 2006; Matić 2010].

In this paper, we completely describe the noncuspidal unitary dual of the double cover of the symplectic group of split rank two. Although this is not an algebraic group, some recent results enabled us to study this group in the same spirit as the classical split groups. More concretely, Hanzer and Muić [2009] related reducibilities of the induced representations of metaplectic groups with those of the odd orthogonal groups (using theta correspondence), while their paper [2010] describes the extension of the Jacquet module techniques of Tadić for classical groups to metaplectic groups. More specifically, Tadić's structure formula for symplectic and odd-orthogonal groups [1995] (which is a version of a geometric lemma of [Bernstein and Zelevinsky 1977]) is extended to metaplectic groups. These ingredients made the determination of the irreducible subquotients of the principal series for $\widetilde{\mathrm{Sp}(2)}$ very similar to the one obtained in [Matić $\geq$ 2010] for SO(5), but this happens to be insufficient tool in some cases. In these cases, we will use the theta correspondence to again obtain the formal similarity to the SO(5) case. This similarity was expected; see for example [Zorn 2010]. After determining complete nonunitary dual, modulo cuspidal representations, the unitary dual follows in the almost the same way as in [Matić 2010], but after discussion of some exceptional

MSC2000: primary 22E35, 22E50; secondary 11F70.
Keywords: metaplectic groups, unitary dual, theta correspondence.

cases (for example, the discussion of the unitary principal series): In the case of the odd orthogonal group SO(5), the irreducibility of the unitary principal series follows from the considerations about R-groups, and in the case of $\widetilde{\mathrm{Sp}(2)}$, since the R-group theory for metaplectic groups is not available in its full generality, irreducibility is obtained using theta correspondence. In the forthcoming paper [Hanzer and Matić 2010], we extend the methods used here to prove for general n the irreducibility of unitary principal series for $\widetilde{\mathrm{Sp}(n)}$. We hope these results will have applications in the theory of automorphic forms.

We now describe the content of the paper. In Section 2, we recall the definition of the metaplectic double cover $\widetilde{\mathrm{Sp}(n)}$. We also recall the notions of parabolic subgroups, Jacquet functor, and parabolic induction in the context of metaplectic groups. We then recall the notion of the dual pair, and the lifts of an irreducible representations of one member of the pair dual to the Weil representation of the ambient metaplectic group. We recall the criteria for the square integrability and temperedness of the irreducible representations of metaplectic groups, due to Ban and Jantzen [2009] and recall the classification of the irreducible genuine representations of $\widetilde{\mathrm{Sp}(n)}$ obtained in [Hanzer and Muić 2010]. In Section 3, we analyze the principal series for $\widetilde{\mathrm{Sp}(2)}$, using both theta correspondence and Tadić's methods applied to metaplectic groups. In Section 4, we determine the unitary dual of $\widetilde{\mathrm{Sp}(2)}$ supported in the minimal parabolic subgroup. In Section 5, we describe irreducible representations of $\widetilde{\mathrm{Sp}(2)}$ supported on maximal parabolic subgroups, and the unitary dual of $\widetilde{\mathrm{Sp}(2)}$ supported on maximal parabolic subgroups.

2. Preliminaries

Let $\widetilde{\mathrm{Sp}(2)}$ be the unique nontrivial two-fold central extension of symplectic group $\mathrm{Sp}(2, F)$, where F is a non-Archimedean local field of characteristic different from two. In other words, we have maps

$$1 \to \mu_2 \to \widetilde{\mathrm{Sp}(2)} \to \mathrm{Sp}(2, F) \to 1.$$

The multiplication in $\widetilde{\mathrm{Sp}(2)}$, which is as a set given by $\mathrm{Sp}(2, F) \times \mu_2$, is given the cocycle of [Ranga Rao 1993]. The topology of $\widetilde{\mathrm{Sp}(n)}$ is explained in detail in [Hanzer and Muić 2010, Section 3.3]. There exist compact open subgroups of $\mathrm{Sp}(n)$ that split in $\widetilde{\mathrm{Sp}(n)}$. Recall that a maximal good compact subgroup $\mathrm{Sp}(O_F)$ splits if the residual characteristic of F is odd (here O_F denotes the ring of integers on F). In [Hanzer and Muić 2010], the metaplectic group $\widetilde{\mathrm{Sp}(2)}$ was denoted by $\widetilde{\mathrm{Sp}(W_2)}$. We say that the representation of $\widetilde{\mathrm{Sp}(2)}$ (or, more generally, $\widetilde{\mathrm{Sp}(n)}$) is smooth if, for every vector v in the representation space V, there exists a compact open subgroup K_1 of $\mathrm{Sp}(2)$ that splits in $\widetilde{\mathrm{Sp}(2)}$ and fixes v. The representation is admissible if for every K_1 as above, the space V^{K_1} is finite-dimensional.

Lemma 2.1. *Irreducible smooth representations of $\widetilde{\mathrm{Sp}(n)}$ are admissible.*

Proof. First, we prove that an irreducible, smooth, cuspidal representation of $\widetilde{\mathrm{Sp}(n)}$ is admissible. We can proceed as in the corollary on [Bernstein 1992, page 36], so we have to prove that an irreducible cuspidal representation of $\widetilde{\mathrm{Sp}(n)}$ is compact, and that was proved in proving [Hanzer and Muić 2009, Lemma 3.1]. Then, on this compact irreducible representation we can apply [Bernstein 1992, Proposition 11], which is formulated for a general totally disconnected group (so the metaplectic groups satisfy the conditions), and says that finitely generated compact representations are admissible. The claim follows since every irreducible smooth representation can be embedded in the representations parabolically induced from the cuspidal representations of Levi subgroups (and for the representations of Levi subgroups, the same reasoning as above shows that these representations are also admissible), which was proved in [Hanzer and Muić 2008, Proposition 4.4]. □

In this paper we are interested only in genuine representations of $\widetilde{\mathrm{Sp}(n)}$ (that is, those that do not factor through μ_2). So, let $R(n)$ be the Grothendieck group of the category of all admissible genuine representations of finite length of $\widetilde{\mathrm{Sp}(n)}$ (that is, a free abelian group over the set of all irreducible genuine representations of $\widetilde{\mathrm{Sp}(n)}$), and define $R = \bigoplus_{n\geq 0} R(n)$. By ν we denote a character of $\mathrm{GL}(k, F)$ defined by $|\det|_F$. Further, for an ordered partition $s = (n_1, n_2, \ldots, n_j)$ of some $m \leq n$, we denote by P_s a standard parabolic subgroup of $\mathrm{Sp}(n, F)$ (consisting of block upper-triangular matrices), whose Levi factor equals $\mathrm{GL}(n_1)\times\mathrm{GL}(n_2)\times\cdots\times \mathrm{GL}(n_j) \times \mathrm{Sp}(n - |s|, F)$, where $|s| = \sum_{i=1}^{j} n_i$. By a standard parabolic subgroup $\widetilde{P}_s$ of $\widetilde{\mathrm{Sp}(n)}$ we mean the preimage of P_s in $\widetilde{\mathrm{Sp}(n)}$. We have the analogous notation for the Levi subgroups of the metaplectic groups, and, for the completeness, we explicitly describe the structure of the parabolic and Levi subgroups, as explained in [Hanzer and Muić 2010, Section 2.2]. There is a natural splitting from the unipotent radical of N_s of the corresponding standard parabolic subgroup P_s of $\mathrm{Sp}(n, F)$ to its cover [Mœglin et al. 1987, Lemma 2.9 on page 43]; let N'_s be the image of that homomorphism. We then have $\widetilde{P}_s \cong \widetilde{M}_s \ltimes N'_s$.

We can explicitly describe $\widetilde{M}_s$ as follows. There is a natural epimorphism

$$\phi : \widetilde{\mathrm{GL}(n_1, F)} \times \cdots \times \widetilde{\mathrm{GL}(n_k, F)} \times \widetilde{\mathrm{Sp}(W_{n-|s|})} \to \widetilde{M}_s \tag{1}$$

given by

$$([g_1, \epsilon_1], \ldots, [g_k, \epsilon_k], [h, \epsilon]) \mapsto [(g_1, g_2, \ldots, g_k, h), \epsilon_1 \cdots \epsilon_k \epsilon \beta], \tag{2}$$

with $\beta = \prod_{i<j}(\det g_i, \det g_j)_F \big(\prod_{i=1}^{k}(\det g_i, x(h))_F\big)$, where $x(h)$ is defined in [Ranga Rao 1993, Lemma 5.1] and $(\cdot,\cdot)_F$ denotes the Hilbert symbol of the field F. Although $\widetilde{M}$ is not exactly the product at left in (1), it differs from it by a finite subgroup that enables us to write every irreducible representation π

of $\widetilde{M}$ in the form $\pi_1 \otimes \cdots \otimes \pi_k \otimes \pi'$, where the representations $\pi_1, \ldots, \pi_k, \pi'$ are either all genuine or none genuine. This simple property enables us to set up Tadić's machinery [Tadić 1995; Hanzer and Muić 2008] of parabolic induction and Jacquet functors. Recall that the irreducible representations in this paper, unless mentioned otherwise, are assumed to be genuine (that is, nontrivial on μ_2). Also, the cuspidality of representations is defined in the same way as for the reductive groups (because of the splitting of the unipotent radical) and characterized in terms if the support of the matrix coefficients also as for the reductive groups.

Let σ be a representation of $\widetilde{\mathrm{Sp}(2)}$. Following the notation introduced in [Hanzer and Muić 2010], we denote by $R_{\widetilde{P_{(1,1)}}}(\sigma)$ the normalized Jacquet module with respect to $\widetilde{M}_{(1,1)}$; by $R_{\widetilde{P}_1}(\sigma)$ the normalized Jacquet module with respect to $\widetilde{M}_{(1)}$; and by $R_{\widetilde{P}_2}(\sigma)$ the normalized Jacquet module with respect to $\widetilde{M}_{(2)}$.

We fix a nontrivial additive character ψ of F and let $\omega_{n,r}$ be the pullback of the Weil representation $\omega_{n(2r+1),\psi}$ of the group $\widetilde{\mathrm{Sp}(n(2r+1))}$, restricted to the dual pair $\widetilde{\mathrm{Sp}(n)} \times O(2r+1)$ [Kudla 1996, Chapter II]. Here $O(2r+1)$ denotes the split odd-orthogonal group of the split rank r, with the one-dimensional anisotropic space sitting at the bottom of the orthogonal tower [Kudla 1996, Chapter III.1]. The standard parabolic subgroups (containing the upper triangular Borel subgroup) of $O(2r+1)$ have the analogous description as the standard parabolic subgroups of $\mathrm{Sp}(n, F)$; we use the analogous notation for the normalized Jacquet functors.

Let σ be an irreducible smooth genuine representation of $\widetilde{\mathrm{Sp}(n)}$. We write $\Theta(\sigma, r)$ for the smooth isotypic component of σ in $\omega_{n,r}$ (we view it as a representation of $O(2r+1)$). Denote by r_0 the smallest r such that $\Theta(\sigma, r) \neq 0$. When σ is cuspidal, we know that $\Theta(\sigma, r_0)$ is an irreducible cuspidal representation of $O(2r+1)$.

Let $\widetilde{\mathrm{GL}(n, F)}$ be a double cover of $\mathrm{GL}(n, F)$, where the multiplication is given by

$$(g_1, \epsilon_1)(g_2, \epsilon_2) = (g_1 g_2, \epsilon_1 \epsilon_2 (\det g_1, \det g_2)_F).$$

Here $\epsilon_i \in \mu_2$ for $i = 1, 2$ and $(\cdot, \cdot)_F$ denotes the Hilbert symbol of the field F, and this cocycle on $\mathrm{GL}(n, F)$ is actually a restriction of Ranga Rao's cocyle on $\mathrm{Sp}(n, F)$ to $\mathrm{GL}(n, F)$, if we view this group as the Siegel Levi subgroup of $\mathrm{Sp}(n, F)$ [Kudla 1986, page 235]. Now we fix a character $\chi_{V,\psi}(g, \epsilon) = \chi_V(\det g)\epsilon\gamma(\det g, \psi_{1/2})^{-1}$ of $\widetilde{\mathrm{GL}(n, F)}$. Here γ denotes the Weil invariant, while χ_V is a character related to the quadratic form on $O(2r+1)$ [Kudla 1996, pages 17 and 37], and $\psi_a(x) = \psi(ax)$ for $a \in F^*$. We may suppose $\chi_V \equiv 1$ (but the arguments that follow are valid without this assumption). We write $\alpha = \chi_{V,\psi}^2$ and observe that α is a quadratic character on $\mathrm{GL}(n, F)$.

The following fact, follows directly from [Hanzer and Muić 2010], and we use it frequently while determining composition series of induced representations: For an irreducible genuine representation π of $\widetilde{\mathrm{GL}(k, F)}$ and an irreducible genuine representation σ of $\widetilde{\mathrm{Sp}(n)}$ we have $\pi \rtimes \sigma = \widetilde{\pi}\alpha \rtimes \sigma$ (in R), where $\pi \rtimes \sigma$ denotes the representation of the group $\widetilde{\mathrm{Sp}(n+k)}$ parabolically induced from the representation $\pi \otimes \sigma$ of the maximal Levi subgroup $\widetilde{M}_{(k)}$. We follow here the usual notation for parabolic induction for classical groups, adapted to the metaplectic case [Tadić 1994; Hanzer and Muić 2010]. We also freely use Zelevinsky's notation [1980] for the parabolic induction for general linear groups. We denote the Steinberg representation of the reductive algebraic group G by St_G and the trivial representation of that group by 1_G. Following [Kudla 1996], we let $\omega^{+}_{\psi_a,n}$ denote the even part of the Weil representation of $\widetilde{\mathrm{Sp}(n)}$ determined by the additive character ψ_a. The nontrivial character of μ_2, when we view it as a representation of $\widetilde{\mathrm{Sp}(0)}$, is denoted by ω_0.

If ζ is a quadratic character of $F^{\times}$, we can write $\zeta(x) = (xa)_F$ for some $a \in F^{\times}$. Let $sp_{\zeta,1}$ be an irreducible (square-integrable, according to the criterion for the square-integrability which we recall below) subrepresentation of $\chi_{V,\psi}\zeta\nu^{1/2} \rtimes \omega_0$. Then, as in [Kudla 1996, page 89], we have the exact sequence

$$0 \to sp_{\zeta,1} \longrightarrow \chi_{V,\psi}\zeta\nu^{1/2} \rtimes \omega_0 \longrightarrow \omega^{+}_{\psi_a,1} \to 0.$$

The results of [Ban and Jantzen 2009] imply that Casselman's criteria for square-integrability and temperedness hold for metaplectic groups in a similar form as for the classical groups (for example symplectic). We now recall these criteria.

Let π be an admissible irreducible (genuine) representation of $\widetilde{\mathrm{Sp}(n)}$ and let $\widetilde{P}_s$ be any standard parabolic subgroup minimal with respect to the property that $R_{\widetilde{P}_s}(\pi) \neq 0$. Write $s = (n_1, \dots, n_k)$ and let σ be any irreducible subquotient of $R_{\widetilde{P}_s}(\pi)$. As we saw above, we can write $\sigma = \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_k \otimes \rho$, where ρ_i is an irreducible genuine cuspidal representation of some $\widetilde{\mathrm{GL}(n_i, F)}$ for $i = 1, \dots, k$ and ρ is an irreducible genuine cuspidal representation of some $\widetilde{\mathrm{Sp}(n-l)}$. Define $e(\rho_i)$ by $\rho_i = \nu^{e(\rho_i)}\rho_i^u$, where ρ_i^u is unitary for $1 \le i \le n$.

Assume that the inequalities

$$\begin{aligned} n_1e(\rho_1) &> 0,\\ n_1e(\rho_1) + n_2e(\rho_2) &> 0,\\ &\vdots\\ n_1e(\rho_1) + n_2e(\rho_2) + \cdots + n_ke(\rho_k) &> 0. \end{aligned}$$

hold for every s and σ as above. Then π is a square integrable representation. For such s and σ, these inequalities also hold if π is a square integrable representation.

The criterion for tempered representations is given by replacing every $>$ with $\geq$.

We recall the definition of a negative representation [Hanzer and Muić 2010, Definition 4.1].

Let σ be an admissible irreducible genuine representation of $\widetilde{\mathrm{Sp}(n)}$. Then σ is a strongly negative (respectively, negative) representation if and only if for every embedding $\sigma \hookrightarrow \rho_1 \times \rho_2 \times \cdots \times \rho_k \rtimes \rho$, where ρ_i for $1 \le i \le k$ and ρ are irreducible genuine supercuspidal representations of some of the $\widetilde{\mathrm{GL}}$ and of some $\widetilde{\mathrm{Sp}(n-l)}$, we have

$$\begin{aligned} n_1 e(\rho_1) &< 0 \quad \text{(respectively, } \le 0), \\ n_1 e(\rho_1) + n_2 e(\rho_2) &< 0 \quad \text{(respectively, } \le 0), \\ &\vdots \\ n_1 e(\rho_1) + n_2 e(\rho_2) + \cdots + n_k e(\rho_k) &< 0 \quad \text{(respectively, } \le 0). \end{aligned}$$

As soon as σ as above is genuine, the ρ_i and ρ are also necessarily genuine. For notation, we recall [Hanzer and Muić 2010, Theorems 4.5 and 4.6]. Recall that, for a cuspidal representation ρ of some $\mathrm{GL}(m_\rho, F)$, a segment Δ is a set of cuspidal representations $\Delta = \{\rho, \nu\rho, \dots, \nu^{k-1}\rho\}$ and $\langle \Delta \rangle$ is a unique irreducible subrepresentation of $\rho \times \nu\rho \times \cdots \times \nu^{k-1}\rho$. We use the same notation for genuine cuspidal representations of $\widetilde{\mathrm{GL}(m_\rho, F)}$ since the transfer from nongenuine to genuine representations in the case of $\widetilde{\mathrm{GL}(m_\rho, F)}$ is particularly simple (obtained by multiplication with the character $\chi_{V,\psi}(g, \epsilon)$ defined above). Now the two theorems above follow from the analogous results in the case of classical reductive groups of [Hanzer and Muić 2008], since the analogous calculations with Jacquet modules are possible, due to results in [Hanzer and Muić 2010].

- Suppose that $\Delta_1, \dots, \Delta_k$ is a sequence of segments (of genuine representations) such that $e(\Delta_1) \ge \cdots \ge e(\Delta_k) > 0$ (we also allow $k = 0$). Let σ_{neg} be a negative (genuine) representation. Then the induced representation $\langle \Delta_1 \rangle \times \langle \Delta_2 \rangle \times \cdots \times \langle \Delta_k \rangle \rtimes \sigma_{\mathrm{neg}}$ has a unique irreducible subrepresentation; we denote it by $\langle \Delta_1, \dots, \Delta_k; \sigma_{\mathrm{neg}} \rangle$.
- If σ is an irreducible admissible genuine representation of $\widetilde{\mathrm{Sp}(n)}$, then there exist a sequence of segments (of genuine representations) $\Delta_1, \dots, \Delta_k$ such that $e(\Delta_1) \ge \cdots \ge e(\Delta_k) > 0$ and a negative (genuine) representation σ_{neg} such that $\sigma \simeq \langle \Delta_1, \dots, \Delta_k; \sigma_{\mathrm{neg}} \rangle$.

We can carry over Tadić's structure formula for classical groups to the metaplectic case [Hanzer and Muić 2010, Proposition 4.5], which enables us to calculate Jacquet modules of the induced representations. In more detail, let

$$R^{\mathrm{gen}} = \bigoplus_n R(\widetilde{\mathrm{GL}(n, F)})_{\mathrm{gen}},$$

where $R(\widetilde{\mathrm{GL}(n, F)})_{\mathrm{gen}}$ denotes the Grothendieck group of finite length, smooth, genuine representations of $\widetilde{\mathrm{GL}(n, F)}$. We denote by $\times$ the linear extension to

$R^{\mathrm{gen}} \otimes R^{\mathrm{gen}}$ of the parabolic induction (from a maximal parabolic subgroup). We can easily check that if σ is an irreducible genuine representation of $\widetilde{\mathrm{Sp}(W_n)}$, then $r_k(\sigma)$, the normalized Jacquet module of σ with respect to the standard maximal parabolic $\widetilde{P}_k$, is a genuine representation of $\widetilde{M_{(k)}}$ and as such can be interpreted as a (genuine) representation of $\widetilde{\mathrm{GL}(k, F)} \times \widetilde{\mathrm{Sp}(W_{n-k})}$, that is, as an element of $R^{\mathrm{gen}} \otimes R$, with R defined as above. So for irreducible genuine σ, we can introduce $\mu^*(\sigma) \in R^{\mathrm{gen}} \otimes R$ by

$$\mu^*(\sigma) = \sum_{k=0}^{n} \mathrm{s.s.}(r_k(\sigma)),$$

where s.s. stands for semisimplification. We can extend μ^* linearly to the whole R. Using Jacquet modules for the maximal parabolic subgroups of $\widetilde{\mathrm{GL}(n, F)}$ we can analogously define

$$m^*(\pi) = \sum_{k=0}^{n} \mathrm{s.s.}(r_k(\pi)) \in R^{\mathrm{gen}} \otimes R^{\mathrm{gen}}$$

for a genuine, irreducible representation π of $\widetilde{\mathrm{GL}(n, F)}$ and then extend m^* linearly to the whole R^{gen}. Let $\kappa : R^{\mathrm{gen}} \otimes R^{\mathrm{gen}} \to R^{\mathrm{gen}} \otimes R^{\mathrm{gen}}$ be defined by $\kappa(x \otimes y) = y \otimes x$. We extend the contragredient $\tilde{}$ to an automorphism of R^{gen} naturally. Finally, we define

$$M^* = (m \otimes \mathrm{id}) \circ (\tilde{}\alpha \otimes m^*) \circ \kappa \circ m^*.$$

Here $\tilde{}\alpha$ means taking contragredient of a representation, and then multiplying by the character α, acting on the general linear group as $\alpha(g) = (\det g, -1)_F$.

For π in R^{gen} and σ from R, we have

$$\mu^*(\pi \rtimes \sigma) = M^*(\pi) \rtimes \mu^*(\sigma).$$

Using this formula for the induced representations of $\widetilde{\mathrm{Sp}(2)}$, we get the following:

- Fix an admissible representation π of $\widetilde{\mathrm{GL}(2)}$, and suppose that π is of finite length. Let $m^*(\pi) = 1 \otimes \pi + \sum_i \pi_i^1 \otimes \pi_i^2 + \pi \otimes 1$, where $\sum_i \pi_i^1 \otimes \pi_i^2$ is a decomposition into a sum of irreducible representations. Now we have

$$\mu^*(\pi \rtimes \omega_0) = 1 \otimes \pi \rtimes \omega_0 + \sum_i \pi_i^1 \otimes \pi_i^2 \rtimes \omega_0 + \sum_i \alpha\widetilde{\pi}_i^2 \otimes \pi_i^1 \rtimes \omega_0 \\ + \pi \otimes \omega_0 + \alpha\widetilde{\pi} \otimes \omega_0 + \sum_i \pi_i^1 \times \alpha\widetilde{\pi}_i^2 \otimes \omega_0.$$

- Fix an admissible representation π of $\widetilde{\mathrm{GL}(1)}$ and an admissible representation σ of $\widetilde{\mathrm{Sp}(1)}$. If we have

$$\mu^*(\sigma) = 1 \otimes \sigma + \sum_i \sigma_i^1 \otimes \sigma_i^2,$$

where σ_i^1 and σ_i^2 are irreducible representations, then

$$\mu^*(\pi \rtimes \sigma) = 1 \otimes \pi \rtimes \sigma + \pi \otimes \sigma + \alpha\widetilde{\pi} \otimes \sigma + \sum_i \sigma_i^1 \otimes \pi \rtimes \sigma_i^2 + \sum_i \pi \times \sigma_i^1 \otimes \sigma_i^2 + \sum_i \sigma_i^1 \times \alpha\widetilde{\pi} \otimes \sigma_i^2.$$

From now on, $\widehat{F^\times}$ denotes the set of the unitary characters of $F^\times$, while $\widetilde{F^\times}$ denotes those that are not necessarily unitary.

3. Principal series

We first state an important reducibility result that follows directly from [Hanzer and Muić 2009, Theorems 3.5. and 4.2].

Proposition 3.1. *Let $\chi \in \widehat{F^\times}$ and let $s \in \mathbb{R}$ be nonnegative. The representation $\chi_{V,\psi} \nu^s \chi \rtimes \omega_0$ of $\widetilde{\mathrm{Sp}(1)}$ reduces if and only if $\chi^2 = 1_{F^\times}$ and $s = 1/2$.*

Let $\zeta \in \widehat{F^\times}$ such that $\zeta^2 = 1_{F^\times}$. In R we have (see [Kudla 1996, page 89]

$$\chi_{V,\psi} \nu^{1/2} \zeta \rtimes \omega_0 = sp_{\zeta,1} + \omega^+_{\psi_a,1}.$$

The following proposition is well known and follows easily from the analogous results for the split SO(3) and SO(5).

Proposition 3.2. (1) *Let $\chi \in \widehat{F^\times}$ and suppose $s \in \mathbb{R}$ is nonnegative. The representation $\nu^s \chi \rtimes 1$ of $O(3)$ reduces if and only if $\chi^2 = 1_{F^\times}$ and $s = 1/2$. In that situation, the length of $\nu^{1/2}\chi \rtimes 1$ is two, and this representation has the unique subrepresentation that is square integrable.*

(2) *Let $\zeta_1, \zeta_2 \in \widehat{F^\times}$. Then, the unitary principal series $\zeta_1 \times \zeta_2 \rtimes 1$ of $O(5)$ is irreducible.*

We use these two propositions in the sequel without explicitly mentioning them.

3.1. ***Unitary principal series.*** In this subsection we prove irreducibility of the unitary principal series $\chi_{V,\psi}\chi_1 \times \chi_{V,\psi}\chi_2 \rtimes \omega_0$, where $\chi_i \in \widehat{F^\times}$ for $i = 1, 2$.

Let Π denote the representation $\chi_{V,\psi}\chi_1 \times \chi_{V,\psi}\chi_2 \rtimes \omega_0$. Using the structure formula for $\mu^*(\Pi)$ from the end of the previous section, we get

$$R_{\widetilde{P}_1}(\Pi) = \chi_{V,\psi}\chi_1^{-1} \otimes \chi_{V,\psi}\chi_2 \rtimes \omega_0 + \chi_{V,\psi}\chi_1 \otimes \chi_{V,\psi}\chi_2 \rtimes \omega_0 + \chi_{V,\psi}\chi_2^{-1} \otimes \chi_{V,\psi}\chi_1 \rtimes \omega_0 + \chi_{V,\psi}\chi_2 \otimes \chi_{V,\psi}\chi_1 \rtimes \omega_0.$$

Remark. Let π be an irreducible subrepresentation of Π. Because of irreducibility of the representations $\chi_{V,\psi}\chi_1 \times \chi_{V,\psi}\chi_2$ and $\chi_{V,\psi}\chi_i \rtimes \omega_0$ for $i = 1, 2$, we get

$$\pi \hookrightarrow \Pi \simeq \chi_{V,\psi}\chi_1^{-1} \times \chi_{V,\psi}\chi_2 \rtimes \omega_0 \simeq \chi_{V,\psi}\chi_2^{-1} \times \chi_{V,\psi}\chi_1 \rtimes \omega_0.$$

If $\chi_i \neq \chi_i^{-1}$ holds for both $i = 1, 2$ and $\chi_1 \neq \chi_2^{\pm 1}$, then Frobenius reciprocity implies that $R_{\widetilde{P}_1}(\pi) = R_{\widetilde{P}_1}(\Pi)$, so $\pi = \Pi$ and the representation Π is irreducible.

Now we prove the irreducibility of the unitary principal series for general unitary characters. Let ζ_1, ζ_2 be the unitary characters of $F^\times$. We prove irreducibility of the representation $\chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0$ using theta correspondence, beginning with this lemma:

Lemma 3.3. *Let π_1 be an irreducible subrepresentation of $\chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0$. Then $\Theta(\pi_1, 2) = \zeta_1 \times \zeta_2 \rtimes 1$.*

Proof. According to the stable range condition [Kudla 1996, page 48], $\Theta(\pi_1, 4) \neq 0$ (observe that $\Theta(\pi_1, 4)$ is a smooth representation of $O(9)$). We have epimorphisms $\omega_{2,4} \to \pi_1 \otimes \Theta(\pi_1, 4)$ and $R_{P_1}(\omega_{2,4}) \to \pi_1 \otimes R_{P_1}(\Theta(\pi_1, 4))$. If τ is an irreducible quotient of $\Theta(\pi_1, 4)$, then [Kudla 1986, Corollary 2.6] implies $[\tau] = [\nu^{-3/2}, \nu^{-1/2}, \zeta_1, \zeta_2; 1]$, where $[\tau]$ denotes the cuspidal support of τ. Clearly, $R_{P_{(1,1,1,1)}}(\tau) \geq \nu^{l_1/2} \otimes \nu^{l_2/2} \otimes \zeta_1^{\pm 1} \otimes \zeta_2^{\pm 1}$ or $R_{P_{(1,1,1,1)}}(\tau) \geq \zeta_1^{\pm 1} \otimes \nu^{l_1/2} \otimes \zeta_2^{\pm 1} \otimes \nu^{l_2/2}$ (or we have some order of factors) for some $l_1, l_2 \in \{\pm 1, \pm 3\}$. If we assume that in the Jacquet module $R_{P_{(1,1,1,1)}}(\tau)$ there is an irreducible subquotient as above whose first factor consists of a unitary character, then, using [Bernstein 1992, Lemma 26] together with Frobenius reciprocity, easily follows that

$$\mathrm{Hom}(\tau, \zeta_1^{\pm 1} \times \nu^{l_1/2} \times \zeta_2^{\pm 1} \times \nu^{l_2/2} \rtimes 1) \neq 0.$$

But since $\zeta_i^{\pm 1} \times \nu^{l_i/2} \cong \nu^{l_i/2} \times \zeta_i^{\pm 1}$, we have $\mathrm{Hom}(\tau, \nu^{l_1/2} \times \zeta_1^{\pm 1} \times \zeta_2^{\pm 1} \times \nu^{l_2/2} \rtimes 1) \neq 0$. So, there is an irreducible subquotient τ' of $\zeta_1 \times \zeta_2 \times \nu^{l_2/2} \rtimes 1$ such that τ is a subrepresentation of $\nu^{l_1/2} \rtimes \tau'$. This implies that $R_{P_1}(\tau)(\nu^{l_1/2})$, the isotypic component of $R_{P_1}(\tau)$ along the generalized character $\nu^{l_1/2}$, is nonzero, as is $R_{P_1}(\Theta(\pi_1, 4))(\nu^{l_1/2})$.

Observations above imply that there is an irreducible representation τ_1 of $O(3)$ such that the mappings $R_{P_1}(\omega_{2,4}) \to \pi_1 \otimes R_{P_1}(\Theta(\pi_1, 4)) \to \pi_1 \otimes \nu^{l_1/2} \otimes \tau_1$ are epimorphisms. We denote the epimorphism $R_{P_1}(\omega_{2,4})) \to \pi_1 \otimes \nu^{l_1/2} \otimes \tau_1$ by T. Now $R_{P_1}(\omega_{2,4})$ has the filtration in which

- $I_{10} = \nu^{-3/2} \otimes \omega_{2,3}$ is the quotient and
- $I_{11} = \mathrm{Ind}_{\mathrm{GL}(1) \times \widetilde{P}_1 \times O(3)}^{M_1 \times \widetilde{\mathrm{Sp}(2)}}(\chi_{V,\psi}\Sigma_1' \otimes \omega_{1,3})$ is the subrepresentation.

See [Kudla 1996, page 57] and [Hanzer and Muić 2009, Proposition 3.3], where the notation is explained in detail.

Suppose $T|_{I_{11}} \neq 0$. Because $\chi_{V,\psi}\nu^{-l_1/2}$ is the isotypic component of $\nu^{l_1/2}$ in the $\widetilde{\mathrm{GL}(1, F)} \times \mathrm{GL}(1, F)$-module $\chi_{V,\psi}\Sigma_1'$, by applying the second Frobenius we get a nonzero $\widetilde{\mathrm{GL}(1, F)} \times \mathrm{GL}(1, F) \times \widetilde{\mathrm{Sp}(1)} \times O(3)$-homomorphism

$$\nu^{l_1/2} \otimes \chi_{V,\psi}\nu^{-l_1/2} \otimes \omega_{1,3} \to \nu^{l_1/2} \otimes \tau_1 \otimes \widetilde{R_{\widetilde{P}_1}(\widetilde{\pi}_1)},$$

which implies that $\widetilde{R_{\widetilde{P}_1}(\widetilde{\pi}_1)}(\chi_{V,\psi}\nu^{-l_1/2}) \neq 0$. Because $l_1 \neq 0$, this contradicts our assumption $\pi_1 \hookrightarrow \chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0$; hence $T|_{I_{11}} = 0$. Therefore, we can consider T as an epimorphism $I_{10} \to \pi_1 \otimes \nu^{l_1/2} \otimes \tau_1$. Consequently, $l_1 = -3$ and there is an epimorphism $\omega_{2,3} \to \pi \otimes \tau_1$. Obviously, $\Theta(\pi_1, 3) \neq 0$.

Repeating the same procedure once again, we obtain $\Theta(\pi_1, 2) \neq 0$. Since the cuspidal support of each irreducible quotient of $\Theta(\pi_1, 2)$ equals $[\zeta_1, \zeta_2; 1]$, all of the irreducible quotients of $\Theta(\pi_1, 2)$ are equal to $\zeta_1 \times \zeta_2 \rtimes 1$. □

Proposition 3.4. *Let $\zeta_1, \zeta_2 \in \widehat{F^\times}$. Then the unitary principal series representation $\chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0$ is irreducible.*

We present two proofs of this proposition, both based on the previous lemma. The first proof is much simpler than the second — it also uses some known results about Whittaker models for the principal series for metaplectic groups, but we have to assume that the residue characteristic of F is odd. The second proof is more technical, but it doesn't depend on the residue characteristic of F. We feel that presenting both proofs may be useful.

First proof of Proposition 3.4. We denote the representation $\chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0$ by Π. Suppose that the residue characteristic of F is not 2. Howe's duality conjecture and lemma then implies that the representation $\Theta(\zeta_1 \times \zeta_2 \rtimes 1, 2)$ has a unique irreducible quotient, so, by Lemma 3.3, all the irreducible subrepresentations of Π are isomorphic, that is,

$$\Pi = \pi \oplus \cdots \oplus \pi. \tag{3}$$

Now, observe that the representation Π has a unique Whittaker model [Banks 1998; Szpruch 2007]. In more words, for a nondegenerate character θ of the unipotent radical U of Borel subgroup of $\mathrm{Sp}(n)$ (observe that $\widetilde{\mathrm{Sp}(n)}$ splits over U, and the mapping $n \mapsto (n, 1)$ is the splitting) and a genuine character $\chi_{V,\psi}\zeta_1 \otimes \cdots \otimes \chi_{V,\psi}\zeta_n$ of $\widetilde{T}$ (where $\widetilde{T}$ denotes the preimage of maximal diagonal torus in $\mathrm{Sp}(n)$), we have

$$\dim_{\mathbb{C}} \mathrm{Hom}_{\widetilde{\mathrm{Sp}_n}}(\chi_{V,\psi}\zeta_1 \times \cdots \times \chi_{V,\psi}\zeta_n \rtimes \omega_0, \mathrm{Ind}_U^{\widetilde{\mathrm{Sp}(n)}}(\theta)) = 1.$$

This forces that the number of copies of π in (3) to be one, and this finishes the first proof. □

Second proof of Proposition 3.4. We have already seen that there is an epimorphism $R_{\widetilde{P}_1}(\omega_{2,2}) \to \chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi} \rtimes \omega_0 \otimes \zeta_1 \times \zeta_2 \rtimes 1$, so

$$\Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes \zeta_1 \times \zeta_2 \rtimes 1, R_{\widetilde{P}_1}(\omega_{2,2})) \neq 0.$$

$R_{\widetilde{P}_1}(\omega_{2,2})$ has the filtration in which

- $J_{10} = \chi_{V,\psi}\nu^{1/2} \otimes \omega_{1,2}$ is the quotient and
- $J_{11} = \mathrm{Ind}_{\widetilde{\mathrm{GL}(1)} \times P_1 \times \widetilde{\mathrm{Sp}(1)}}^{\widetilde{M_1} \times O(2)}(\chi_{V,\psi}\Sigma_1' \otimes \omega_{1,1})$ is the subrepresentation.

Lemma 3.5. *There is an isomorphism*

$$\mathrm{Hom}_{\widetilde{M}_1}(R_{\widetilde{P}_1}(\omega_{2,2}),\, \chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0) \cong \mathrm{Hom}_{\widetilde{M}_1}(J_{11},\, \chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0)$$

of vector spaces that is given by restriction (that is, $T \mapsto T|_{J_{11}}$*).*

Proof of Lemma 3.5. The map obtained by the restriction is obviously a homomorphism, while the injectivity follows directly. Surjectivity is proved as follows:

We consider the filtration $0 \subseteq W_2 \subseteq W_1 \subseteq R_{\widetilde{P}_1}(\omega_{2,2})$, where W_1 is the representation J_{11}, and $W_1/W_2 \cong \chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes \Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0,\, J_{11})$. Observe that

$$(R_{\widetilde{P}_1}(\omega_{2,2})/W_2)/(W_1/W_2) \cong R_{\widetilde{P}_1}(\omega_{2,2})/W_1 \cong J_{10}.$$

Using standard argument, it can be proved that the representation $R_{\widetilde{P}_1}(\omega_{2,2})/W_2$ is $\widetilde{\mathrm{GL}(1)}$-finite. Then, using the decomposition along the generalized central characters, which in this case coincide with the central characters because W_1/W_2 and J_{10} have different central characters, we obtain

$$R_{\widetilde{P}_1}(\omega_{2,2})/W_2 \cong W_1/W_2 \oplus J_{10}.$$

Now an element of $\mathrm{Hom}_{\widetilde{M}_1}(J_{11},\, \chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0)$ is trivial on W_2, so it can be extended to $R_{\widetilde{P}_1}(\omega_{2,2})$ in an obvious way and surjectivity is proved. □

Using a standard relation between taking a smooth part of the isotypic component of a representation and the homomorphism functor [Hanzer and Muić 2009, page 10], it follows from Lemma 3.5 that

$$\Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0,\, R_{\widetilde{P}_1}(\omega_{2,2})) \cong \Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0,\, J_{11}),$$

if we can prove that $\Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0,\, J_{11})$ is admissible.

Lemma 3.6. *We have* $\Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0,\, J_{11}) = \zeta_1 \times \zeta_2 \rtimes 1$.

Proof of Lemma 3.6. Since $\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes \zeta_1 \times \zeta_2 \rtimes 1$ is a quotient of J_{11}, there is an epimorphism $\Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0,\, J_{11}) \to \zeta_1 \times \zeta_2 \rtimes 1$.

Applying [Hanzer and Muić 2009, Lemma 3.2], we have

$$\begin{aligned}&\mathrm{Hom}_{\widetilde{M}_1 \times O(2)}(J_{11},\, \chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes \Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0,\, J_{11}))\\ &\cong \mathrm{Hom}_{\widetilde{M}_1 \times M(1)}(\chi_{V,\psi}\Sigma_1' \otimes \omega_{1,1},\\ &\qquad \chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes R_{\overline{P}_1}(\Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0,\, J_{11}))).\end{aligned}$$

For every intertwining map T from the first space, let T_0 be the corresponding intertwining map from the second space. Let φ be a natural homomorphism belonging to the first space.

Since $\chi_{V,\psi}\zeta_1 \otimes \zeta_1^{-1}$ (respectively, $\chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes \zeta_2 \rtimes 1$) are the corresponding isotypic components in the $\widetilde{\mathrm{GL}(1, F)} \times \mathrm{GL}(1, F)$-module $\chi_{V,\psi}\Sigma_1'$(respectively, in the $\widetilde{\mathrm{Sp}(1)} \times O(3)$-module $\omega_{1,1}$), irreducibility of these isotypic components implies that the image of φ_0 is isomorphic to $\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes \zeta_1^{-1} \otimes \zeta_2 \rtimes 1$. Now, we write $\varphi_0 = \varphi'' \circ \varphi'$, where φ' is a canonical epimorphism

$$\chi_{V,\psi}\Sigma_1' \otimes \omega_{1,1} \to \chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes \zeta_1^{-1} \otimes \zeta_2 \rtimes 1$$

and φ'' is an inclusion of the representation

$$\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes \zeta_1^{-1} \otimes \zeta_2 \rtimes 1$$

in $\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes R_{\overline{P}_1}(\Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0, J_{11}))$. Observe that $\mathrm{Ind}(\varphi')$ is a homomorphism

$$\mathrm{Ind}_{\widetilde{\mathrm{GL}(1)} \times P_1 \times \widetilde{\mathrm{Sp}(1)}}^{\widetilde{M_1} \times O(2)}(\chi_{V,\psi}\Sigma_1' \otimes \omega_{1,1}) \to \chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes \zeta_1^{-1} \times \zeta_2 \rtimes 1.$$

Let φ_1 be an operator belonging to

$$\mathrm{Hom}(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes \zeta_1 \times \zeta_2 \rtimes 1, \\ \chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes \Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0, J_{11}))$$

such that $(\varphi_1)_0 = \varphi''$.

Lemma 3.7. *Under the assumptions above*, $(\varphi_1 \circ \mathrm{Ind}(\varphi'))_0 = \varphi_0$.

Proof of Lemma 3.7. We prove it much more generally. Let (π, V) be a smooth representation of some Levi subgroup M' in the parabolic P' and the opposite parabolic $\overline{P'}$ of the group G' (which is one of the groups we are considering, that is, metaplectic or odd orthogonal) and let (Π, W) be a smooth representation of G'. Then the second Frobenius isomorphism asserts

$$\mathrm{Hom}_{G'}(\mathrm{Ind}_{M'}^{G'}(\pi), \Pi) \cong \mathrm{Hom}_{M'}(\pi, R_{\overline{P'}}(\Pi)).$$

Let $\psi \hookrightarrow R_{\overline{P'}}(\mathrm{Ind}_{M'}^{G'}(\pi))$ be an embedding corresponding to the open cell $P'\overline{P'}$ in G' given in the following way:

For an open compact subgroup K of G' that has Iwahori decomposition with respect to both P' and $\overline{P'}$, and for $v \in V^{K \cap M'}$, we define

$$f_{v,K}(g) = \frac{1}{\mathrm{meas}_{\overline{N'}}(K \cap \overline{N'})} \begin{cases} 0 & \text{if } g \notin P'K \\ \delta_P^{1/2}(m)\pi(m)v & \text{if } g = mnk \\ & \quad \text{for } m \in M', n \in N', k \in K. \end{cases}$$

Then $\psi : v \mapsto f_{v,K} + \mathrm{Ind}_{M'}^{G'}(\pi)(\overline{N'})$ is independent on the choice of K.

For $\varphi \in \mathrm{Hom}_{G'}(\mathrm{Ind}_{M'}^{G'}(\pi), \Pi)$, we take φ_0 to be the corresponding element of $\mathrm{Hom}_{M'}(\pi, R_{\overline{P}}(\Pi))$. It follows that $\varphi_0(v) = \varphi(f_{v,K}) + \Pi(\overline{N'})$. Write $\varphi_0 = \varphi'' \circ \varphi'$, where φ' denotes the canonical epimorphism $\pi \to \pi / \mathrm{Ker}\,\varphi_0$ and φ'' denotes the embedding $\pi/\operatorname{Ker}\varphi_0 \hookrightarrow R_{\overline{P'}}$. So, we are able to construct the mapping $\operatorname{Ind}(\varphi')$: $\operatorname{Ind}_{M'}^{G'}(\pi) \to \operatorname{Ind}_{M'}^{G'}(\pi/\operatorname{Ker}\varphi_0)$. Since

$$\operatorname{Hom}_{M'}(\pi/\operatorname{Ker}\varphi_0, R_{\overline{P'}}(\Pi)) \cong \operatorname{Hom}_{G'}(\operatorname{Ind}_{M'}^{G'}(\pi/\operatorname{Ker}\varphi_0), \Pi),$$

analogously as above, we see there is an element $\varphi_1 \in \operatorname{Hom}_{G'}(\operatorname{Ind}_{M'}^{G'}(\pi/\operatorname{Ker}\varphi_0), \Pi)$ such that $(\varphi_1)_0 = \varphi''$.

To prove $(\varphi_1 \circ \operatorname{Ind}(\varphi'))_0 = \varphi_0$, it is enough to prove $(\varphi_1 \circ \operatorname{Ind}(\varphi'))_0 = (\varphi_1)_0 \circ \varphi'$. Let $v \in V$. Clearly, $\varphi'(v) = v + \operatorname{Ker}\varphi_0$. Further,

$$(\varphi_1)_0(\varphi'(v)) = \varphi_1(f_{v+\operatorname{Ker}\varphi_0, K}) + \Pi(\overline{N'}),$$
$$(\varphi_1 \circ \operatorname{Ind}(\varphi'))_0(v) = \varphi_1(\operatorname{Ind}(\varphi') f_{v,K}) + \Pi(\overline{N'}).$$

It follows easily that $f_{v+\operatorname{Ker}\varphi_0, K} = f_{v,K} + \operatorname{Ker}\varphi_0$ and $\operatorname{Ind}(\varphi') f_{v,K} = f_{v,K} + \operatorname{Ker}\varphi_0$, and the lemma follows. □

We can complete the proof of Lemma 3.6. Lemma 3.7 gives $\varphi_1 \circ \operatorname{Ind}(\varphi') = \varphi$, so the image of φ is a quotient of $\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes \zeta_1^{-1} \times \zeta_2 \rtimes 1$. This implies that $\Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0, J_{11})$ is a quotient of $\zeta_1^{-1} \times \zeta_2 \rtimes 1$. Since $\zeta_1^{-1} \times \zeta_2 \rtimes 1 \simeq \zeta_1 \times \zeta_2 \rtimes 1$ is an irreducible representation,

$$\Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0, J_{11}) = \zeta_1 \times \zeta_2 \rtimes 1. \qquad \square$$

Lemma 3.8. *There is an epimorphism* $\Theta(\zeta_1 \times \zeta_2 \rtimes 1, 2) \to \chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0$.

Proof of Lemma 3.8. We have an isomorphism

$$\operatorname{Hom}_{O(2)}(\omega_{2,2}, \zeta_1^{-1} \times \zeta_2 \rtimes 1) \cong \operatorname{Hom}(R_{P_1}(\omega_{2,2}), \zeta_1^{-1} \otimes \zeta_2 \rtimes 1)$$

of vector spaces, which also an isomorphism of $\widetilde{\mathrm{Sp}(2)}$ modules. By taking the smooth parts, we obtain

$$\operatorname{Hom}_{\widetilde{\mathrm{Sp}(2)} \times O(2)}(\omega_{2,2}, \zeta_1^{-1} \times \zeta_2 \rtimes 1)_\infty \cong \operatorname{Hom}(R_{P_1}(\omega_{2,2}), \zeta_1^{-1} \otimes \zeta_2 \rtimes 1)_\infty,$$

so that $\Theta(\zeta_1^{-1} \times \zeta_2 \rtimes 1, 2)^\sim \cong \Theta(\zeta_1^{-1} \otimes \zeta_2 \rtimes 1, R_{P_1}(\omega_{2,2}))^\sim$.

In the same way as before, we get

$$\Theta(\zeta_1^{-1} \otimes \zeta_2 \rtimes 1, R_{P_1}(\omega_{2,2}))^\sim \cong \Theta(\zeta_1^{-1} \otimes \zeta_2 \rtimes 1, I_{11})^\sim.$$

Now, the epimorphism $I_{11} \to \zeta_1^{-1} \otimes \zeta_2 \rtimes 1 \otimes \chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0$ gives an epimorphism $\Theta(\zeta_1^{-1} \otimes \zeta_2 \rtimes 1, I_{11})) \to \chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0$. Since the representations $\zeta_1^{-1} \times \zeta_2 \rtimes 1$ and $\zeta_1 \times \zeta_2 \rtimes 1$ are isomorphic, we obtain the epimorphism $\Theta(\zeta_1 \times \zeta_2 \rtimes 1, 2) \to \chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0$, which proves the lemma. □

Now we finish the second proof of Proposition 3.4. Suppose that the representation $\chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0$ reduces. Suppose also that it is the representation of length 2 and write $\chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0 = \pi_1 \oplus \pi_2$. Obviously, $R_{\widetilde{P_1}}(\chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0) = R_{\widetilde{P_1}}(\pi_1) \oplus R_{\widetilde{P_1}}(\pi_2)$.

We have, by Lemma 3.8, an epimorphism

$$\omega_{2,2} \to \zeta_1 \times \zeta_2 \rtimes 1 \otimes \chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0,$$

which leads to the epimorphisms $R_{\widetilde{P_1}}(\omega_{2,2}) \to \zeta_1 \times \zeta_2 \rtimes 1 \otimes (R_{\widetilde{P_1}}(\pi_1) \oplus R_{\widetilde{P_1}}(\pi_2))$ and $R_{\widetilde{P_1}}(\omega_{2,2}) \to \chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0 \otimes (\zeta_1 \times \zeta_2 \rtimes 1 \oplus \zeta_1 \times \zeta_2 \rtimes 1)$.

Finally, we obtain an epimorphism

$$\Theta(\chi_{V,\psi}\zeta_1 \otimes \chi_{V,\psi}\zeta_2 \rtimes \omega_0, R_{\widetilde{P_1}}(\omega_{2,2})) \to \zeta_1 \times \zeta_2 \rtimes 1 \oplus \zeta_1 \times \zeta_2 \rtimes 1,$$

which contradicts Lemmas 3.5 and 3.6.

The same proof remains valid if we suppose that $\chi_{V,\psi}\zeta_1 \times \chi_{V,\psi}\zeta_2 \rtimes \omega_0$ is the representation of the length 4. □

3.2. ***Nonunitary principal series.*** First we determine the reducibility points of the representations with cuspidal support in the minimal parabolic subgroup $\widetilde{P_{(1,1)}}$.

Let $\chi_1, \chi_2 \in \widehat{F^\times}$ and $s_i \geq 0$ for $i = 1, 2$, such that $s_i > 0$ for at least one i. Define $\Pi = \chi_{V,\psi}\nu^{s_1}\chi_1 \times \chi_{V,\psi}\nu^{s_2}\chi_2 \rtimes \omega_0$. We have

$$\begin{aligned}\mu^*(\Pi) = {} & \chi_{V,\psi}\nu^{s_1}\chi_1 \otimes \chi_{V,\psi}\nu^{s_2}\chi_2 \rtimes \omega_0 + \chi_{V,\psi}\nu^{-s_1}\chi_1^{-1} \otimes \chi_{V,\psi}\nu^{s_2}\chi_2 \rtimes \omega_0 \\ & + \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \chi_{V,\psi}\nu^{s_1}\chi_1 \rtimes \omega_0 + \chi_{V,\psi}\nu^{-s_2}\chi_2^{-1} \otimes \chi_{V,\psi}\nu^{s_1}\chi_1 \rtimes \omega_0 \\ & + \chi_{V,\psi}\nu^{-s_1}\chi_1^{-1} \times \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \omega_0 + \chi_{V,\psi}\nu^{s_1}\chi_1 \times \chi_{V,\psi}\nu^{-s_2}\chi_2^{-1} \otimes \omega_0 \\ & + \chi_{V,\psi}\nu^{s_1}\chi_1 \times \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \omega_0 + \chi_{V,\psi}\nu^{-s_1}\chi_1^{-1} \times \chi_{V,\psi}\nu^{-s_2}\chi_2^{-1} \otimes \omega_0 \\ & + 1 \otimes \chi_{V,\psi}\nu^{s_1}\chi_1 \times \chi_{V,\psi}\nu^{s_2}\chi_2 \rtimes \omega_0.\end{aligned}$$

We prove that irreducibility of all the representations above implies irreducibility of the representation Π. We keep this assumption throughout this subsection.

First, suppose that $\nu^{s_1}\chi_1 \neq \nu^{-s_1}\chi_1^{-1}$, $\nu^{s_2}\chi_2 \neq \nu^{-s_2}\chi_2^{-1}$ and $\nu^{s_1}\chi_1 \neq \nu^{\pm s_2}\chi_2^{\pm 1}$ (that is, Jacquet modules of Π are multiplicity one).

Let τ be an irreducible subquotient of Π such that

$$\chi_{V,\psi}\nu^{-s_1}\chi_1^{-1} \times \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \omega_0 \leq R_{\widetilde{P_2}}(\tau).$$

From transitivity of Jacquet modules we get

$$\chi_{V,\psi}\nu^{-s_1}\chi_1^{-1} \otimes \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \omega_0 + \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \chi_{V,\psi}\nu^{-s_1}\chi_1^{-1} \otimes \omega_0 \leq R_{\widetilde{P_{(1,1)}}}(\tau).$$

This implies

$$\chi_{V,\psi}\nu^{-s_1}\chi_1^{-1} \otimes \chi_{V,\psi}\nu^{s_2}\chi_2 \rtimes \omega_0 + \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \chi_{V,\psi}\nu^{s_1}\chi_1 \rtimes \omega_0 \leq R_{\widetilde{P_1}}(\tau).$$

We get directly that

$$R_{\widetilde{P}_2}(\tau) = \chi_{V,\psi}\nu^{-s_1}\chi_1^{-1} \times \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \omega_0 + \chi_{V,\psi}\nu^{s_1}\chi_1 \times \chi_{V,\psi}\nu^{-s_2}\chi_2^{-1} \otimes \omega_0 + \chi_{V,\psi}\nu^{s_1}\chi_1 \times \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \omega_0 + \chi_{V,\psi}\nu^{-s_1}\chi_1^{-1} \times \chi_{V,\psi}\nu^{-s_2}\chi_2^{-1} \otimes \omega_0,$$

so $\tau = \Pi$ and Π is irreducible.

Now we assume that there is some i such that $\nu^{s_i}\chi_i \neq \nu^{-s_i}\chi_i^{-1}$. Without loss of generality, let $i = 1$. So, $s_1 = 0$ and $\chi_1 = \chi_1^{-1}$, that is, $\chi_1^2 = 1_{F^\times}$. We prove that in this case Π is also irreducible. Again, we start by writing corresponding Jacquet modules:

$$\begin{aligned} R_{\widetilde{P}_1}(\Pi) &= 2\chi_{V,\psi}\chi_1 \otimes \chi_{V,\psi}\nu^{s_2}\chi_2 \rtimes \omega_0 + \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \chi_{V,\psi}\chi_1 \rtimes \omega_0 \\ &\quad + \chi_{V,\psi}\nu^{-s_2}\chi_2^{-1} \otimes \chi_{V,\psi}\chi_1 \rtimes \omega_0, \\ R_{\widetilde{P}_2}(\Pi) &= 2\chi_{V,\psi}\chi_1 \times \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \omega_0 + \chi_{V,\psi}\chi_1 \times \chi_{V,\psi}\nu^{-s_2}\chi_2^{-1} \otimes \omega_0. \end{aligned}$$

Let τ be an irreducible subquotient of Π such that

$$R_{\widetilde{P}_1}(\tau) \geq \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \chi_{V,\psi}\chi_1 \rtimes \omega_0.$$

Of course, $R_{\widetilde{P_{(1,1)}}}(\tau) \geq 2\chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \chi_{V,\psi}\chi_1 \otimes \omega_0$, so

$$R_{\widetilde{P}_2}(\tau) \geq 2\chi_{V,\psi}\chi_1 \times \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \omega_0.$$

Continuing in the same way, we get

$$\begin{aligned} R_{\widetilde{P_{(1,1)}}}(\tau) &\geq 2\chi_{V,\psi}\chi_1 \otimes \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \omega_0 + 2\chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \chi_{V,\psi}\chi_1 \otimes \omega_0, \\ R_{\widetilde{P}_1}(\tau) &\geq 2\chi_{V,\psi}\chi_1 \otimes \chi_{V,\psi}\nu^{s_2}\chi_2 \rtimes \omega_0 + \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \chi_{V,\psi}\chi_1 \rtimes \omega_0. \end{aligned}$$

Finally,

$$\begin{aligned} R_{\widetilde{P_{(1,1)}}}(\tau) &\geq 2\chi_{V,\psi}\chi_1 \otimes \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \omega_0 + 2\chi_{V,\psi}\chi_1 \otimes \chi_{V,\psi}\nu^{-s_2}\chi_2^{-1} \otimes \omega_0 \\ &\quad + 2\chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \chi_{V,\psi}\chi_1 \otimes \omega_0, \\ R_{\widetilde{P}_2}(\tau) &\geq 2\chi_{V,\psi}\chi_1 \times \chi_{V,\psi}\nu^{s_2}\chi_2 \otimes \omega_0 + \chi_{V,\psi}\chi_1 \times \chi_{V,\psi}\nu^{-s_2}\chi_2^{-1} \otimes \omega_0 = R_{\widetilde{P}_2}(\Pi). \end{aligned}$$

So, $\Pi = \tau$ and Π is irreducible.

If $\nu^{s_1}\chi_1 = \nu^{s_2}\chi_2$ or $\nu^{s_1}\chi_1 = \nu^{-s_2}\chi_2^{-1}$, then the irreducibility of Π follows in the same way as above. Observe that equalities $\nu^{s_1}\chi_1 = \nu^{-s_1}\chi_1^{-1}$ and $\nu^{s_2}\chi_2 = \nu^{-s_2}\chi_2^{-1}$ lead to unitary principal series.

In this way we have proved irreducibility of the principal series, with these exceptions:

- Some of the representations $\chi_{V,\psi}\nu^{s_1}\chi_1 \rtimes \omega_0$ or $\chi_{V,\psi}\nu^{s_2}\chi_2 \rtimes \omega_0$ reduce (the so-called $\widetilde{\mathrm{Sp}(1)}$ reducibility).

- Some of the representations

$$\chi_{V,\psi}\nu^{-s_1}\chi_1^{-1}\times\chi_{V,\psi}\nu^{s_2}\chi_2,\qquad \chi_{V,\psi}\nu^{s_1}\chi_1\times\chi_{V,\psi}\nu^{-s_2}\chi_2^{-1},$$
$$\chi_{V,\psi}\nu^{s_1}\chi_1\times\chi_{V,\psi}\nu^{s_2}\chi_2,\qquad \chi_{V,\psi}\nu^{-s_1}\chi_1^{-1}\times\chi_{V,\psi}\nu^{-s_2}\chi_2^{-1}$$

  reduce (the so-called $\widetilde{\mathrm{GL}(2)}$ reducibility).

3.2.1. $\widetilde{\mathrm{Sp}(1)}$ *reducibility.* Let $\chi,\zeta\in\widehat{F^\times}$, $\zeta^2=1_{F^\times}$, and $s\geq 0$. It is well known that, in R,

$$\chi_{V,\psi}\nu^s\chi\times\chi_{V,\psi}\nu^{1/2}\zeta\rtimes\omega_0=\chi_{V,\psi}\nu^s\rtimes sp_{\zeta,1}+\chi_{V,\psi}\nu^s\rtimes\omega^+_{\psi_a,1}.$$

Let Π denote $\chi_{V,\psi}\nu^s\rtimes sp_{\zeta,1}$.

Calculating Jacquet modules, we find

$$R_{\widetilde{P_1}}(\Pi)=\chi_{V,\psi}\nu^{-s}\chi^{-1}\otimes sp_{\zeta,1}+\chi_{V,\psi}\nu^s\chi\otimes sp_{\zeta,1}+\chi_{V,\psi}\nu^{1/2}\zeta\otimes\chi_{V,\psi}\nu^s\chi\rtimes\omega_0,$$
$$R_{\widetilde{P_2}}(\Pi)=\chi_{V,\psi}\nu^{-s}\chi^{-1}\times\chi_{V,\psi}\nu^{1/2}\zeta\otimes\omega_0+\chi_{V,\psi}\nu^s\chi\times\chi_{V,\psi}\nu^{1/2}\zeta\otimes\omega_0.$$

If the representation $\chi_{V,\psi}\nu^s\chi\rtimes\omega_0$ is irreducible (that is, when $\nu^s\chi\neq\nu^{\pm1/2}\zeta_2$, where $\zeta_2^2=1_{F^\times}$), we proceed in the following way:

Let ρ be an irreducible subquotient of Π such that

$$\chi_{V,\psi}\nu^{1/2}\zeta\otimes\chi_{V,\psi}\nu^s\chi\rtimes\omega_0\leq s_1(\rho).$$

We directly get that

$$\chi_{V,\psi}\nu^{1/2}\zeta\otimes\chi_{V,\psi}\nu^s\chi\otimes\omega_0+\chi_{V,\psi}\nu^{1/2}\zeta\otimes\chi_{V,\psi}\nu^{-s}\chi^{-1}\otimes\omega_0\leq R_{\widetilde{P_{(1,1)}}}(\rho).$$

If both $\chi_{V,\psi}\nu^{-s}\chi^{-1}\times\chi_{V,\psi}\nu^{1/2}\zeta$ and $\chi_{V,\psi}\nu^s\chi\times\chi_{V,\psi}\nu^{1/2}\zeta$ are irreducible, Π is also irreducible.

For the reducibility of the $\widetilde{\mathrm{Sp}(1)}$ part we still have to determine the composition factors of the representations

(i) $\chi_{V,\psi}\nu^{1/2}\zeta_1\times\chi_{V,\psi}\nu^{1/2}\zeta_2\rtimes\omega_0$,

(ii) $\chi_{V,\psi}\nu^{1/2}\zeta\times\chi_{V,\psi}\nu^{1/2}\zeta\rtimes\omega_0$, and

(iii) $\chi_{V,\psi}\nu^{3/2}\zeta\times\chi_{V,\psi}\nu^{1/2}\zeta\rtimes\omega_0$, where $\zeta^2=\zeta_1^2=\zeta_2^2=1_{F^*}$.

Thus, we have proved the following result:

Proposition 3.9. *Let $\chi\in\widehat{F^\times}$, a nonnegative $s\in\mathbb{R}$, and $\zeta\in\widehat{F^\times}$ with $\zeta^2=1_{F^\times}$. The representations $\chi_{V,\psi}\nu^s\chi\rtimes sp_{\zeta,1}$ and $\chi_{V,\psi}\nu^s\chi\rtimes\omega^+_{\psi_a,1}$ are irreducible unless $(s,\chi)=(3/2,\zeta)$ or $(1/2,\zeta_1)$, where $\zeta_1^2=1_{F^\times}$. In R, we have*

$$\chi_{V,\psi}\nu^s\chi\times\chi_{V,\psi}\nu^{1/2}\zeta\rtimes\omega_0=\chi_{V,\psi}\nu^s\chi\rtimes sp_{\zeta,1}+\chi_{V,\psi}\nu^s\chi\rtimes\omega^+_{\psi_a,1}.$$

Also, if $(s,\chi)\neq(3/2,\zeta)$ *and* $(s,\chi)\neq(1/2,\zeta_1)$, *then*

$$\chi_{V,\psi}\nu^s\chi\rtimes sp_{\zeta,1}=\begin{cases}\langle\chi_{V,\psi}\nu^{1/2}\zeta;\chi_{V,\psi}\chi\rtimes\omega_0\rangle & \text{if } s=0,\\ \langle\chi_{V,\psi}\nu^{1/2}\zeta,\chi_{V,\psi}\nu^s\chi;\omega_0\rangle & \text{if } 0<s\le 1/2,\\ \langle\chi_{V,\psi}\nu^s\chi,\chi_{V,\psi}\nu^{1/2}\zeta;\omega_0\rangle & \text{if } s>1/2,\end{cases}$$

$$\chi_{V,\psi}\nu^s\chi\rtimes\omega^+_{\psi_a,1}=\begin{cases}\chi_{V,\psi}\chi\rtimes\omega^+_{\psi_a,1} & \text{if } s=0,\\ \langle\chi_{V,\psi}\nu^s\chi;\omega^+_{\psi_a,1}\rangle & \text{if } s>0.\end{cases}$$

3.2.2. $\widetilde{\mathrm{GL}(2)}$ *reducibility.* Let $\chi\in\widehat{F^\times}$ and $s\in\mathbb{R}$ be nonnegative. In R, we have

$$\chi_{V,\psi}\nu^{s+1/2}\chi\times\chi_{V,\psi}\nu^{s-1/2}\chi\rtimes\omega_0=\chi_{V,\psi}\nu^s\chi\,\mathrm{St}_{\mathrm{GL}(2)}\rtimes\omega_0+\chi_{V,\psi}\nu^s\chi\,1_{\mathrm{GL}(2)}\rtimes\omega_0.$$

Let Π denote the representation $\chi_{V,\psi}\nu^s\chi\,\mathrm{St}_{\mathrm{GL}(2)}\rtimes\omega_0$. Calculation of $\mu^*(\Pi)$ gives

$$\begin{aligned}R_{\widetilde{P}_1}(\Pi)&=\chi_{V,\psi}\nu^{s+1/2}\chi\otimes\chi_{V,\psi}\nu^{s-1/2}\chi\rtimes\omega_0\\&\quad+\chi_{V,\psi}\nu^{1/2-s}\chi^{-1}\otimes\chi_{V,\psi}\nu^{s+1/2}\chi\rtimes\omega_0,\\ R_{\widetilde{P}_2}(\Pi)&=\chi_{V,\psi}\nu^s\chi\,\mathrm{St}_{\mathrm{GL}(2)}\otimes\omega_0+\chi_{V,\psi}\nu^{-s}\chi^{-1}\,\mathrm{St}_{\mathrm{GL}(2)}\otimes\omega_0\\&\quad+\chi_{V,\psi}\nu^{s+1/2}\chi\times\chi_{V,\psi}\nu^{1/2-s}\chi^{-1}\otimes\omega_0.\end{aligned}$$

Looking at Jacquet modules with respect to different parabolic subgroups we can conclude, in the same way as in the $\widetilde{\mathrm{Sp}(1)}$ reducibility case, that if

$$\chi_{V,\psi}\nu^{s-1/2}\chi\rtimes\omega_0,\quad \chi_{V,\psi}\nu^{s+1/2}\chi\rtimes\omega_0,\quad \chi_{V,\psi}\nu^{s+1/2}\chi\times\chi_{V,\psi}\nu^{1/2-s}\chi^{-1}$$

are irreducible representations, then the representation Π is also irreducible.

Observe that the representation $\chi_{V,\psi}\nu^{s+1/2}\chi\rtimes\omega_0$ reduces for $(\chi,s)=(\zeta,0)$, while $\chi_{V,\psi}\nu^{s-1/2}\chi\rtimes\omega_0$ reduces for $(\chi,s)=(\zeta,1)$, where $\zeta^2=1_{F^\times}$.

The representation $\chi_{V,\psi}\nu^{s+1/2}\chi\times\chi_{V,\psi}\nu^{1/2-s}\chi^{-1}$ reduces for $(\chi,s)=(\zeta,1/2)$, where $\zeta^2=1_{F^\times}$. These observations imply this:

Proposition 3.10. *Let* $\chi\in\widehat{F^\times}$ *and* $s\in\mathbb{R}$ *be nonnegative. The representations* $\chi_{V,\psi}\nu^s\chi\,\mathrm{St}_{\mathrm{GL}(2)}\rtimes 1$ *and* $\chi_{V,\psi}\nu^s\chi\,1_{\mathrm{GL}(2)}\rtimes 1$ *are irreducible except in the cases that* $(s,\chi)=(1/2,\zeta)$, $(s,\chi)=(1,\zeta)$ *or* $(s,\chi)=(0,\zeta)$, *where* $\zeta^2=1_{F^\times}$. *In* R, *we have*

$$\chi_{V,\psi}\nu^{s+1/2}\chi\times\chi_{V,\psi}\nu^{s-1/2}\chi\rtimes\omega_0=\chi_{V,\psi}\nu^s\chi\,\mathrm{St}_{\mathrm{GL}(2)}\rtimes\omega_0+\chi_{V,\psi}\nu^s\chi\,1_{\mathrm{GL}(2)}\rtimes\omega_0.$$

Also, if $\chi_{V,\psi}\nu^{s+1/2}\chi\times\chi_{V,\psi}\nu^{s-1/2}\chi\rtimes\omega_0$ *is a representation of length* 2, *then* $\chi_{V,\psi}\nu^s\chi\,1_{\mathrm{GL}(2)}\rtimes\omega_0=\langle\chi_{V,\psi}\nu^s\chi\,1_{\mathrm{GL}(2)};\omega_0\rangle$ *and*

$$\chi_{V,\psi}\nu^s\chi\,\mathrm{St}_{\mathrm{GL}(2)}\rtimes\omega_0=\begin{cases}\langle\chi_{V,\psi}\nu^{s+1/2}\chi,\chi_{V,\psi}\nu^{1/2-s}\chi;\omega_0\rangle & \text{if } s<1/2,\\ \langle\chi_{V,\psi}\nu^{s+1/2}\chi;\chi_{V,\psi}\nu^{s-1/2}\chi\rtimes\omega_0\rangle & \text{if } s=1/2,\\ \langle\chi_{V,\psi}\nu^{s+1/2}\chi,\chi_{V,\psi}\nu^{s-1/2}\chi;\omega_0\rangle & \text{if } s>1/2.\end{cases}$$

For the reducibility of the $\widetilde{\mathrm{GL}(2)}$ part, we still have to determine the composition factors of the representations

(i) $\chi_{V,\psi}\nu^{1/2}\zeta \times \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega_0$,

(ii) $\chi_{V,\psi}\nu^{3/2}\zeta \times \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega_0$, and

(iii) $\chi_{V,\psi}\nu\zeta \times \chi_{V,\psi}\zeta \rtimes \omega_0$, where $\zeta^2 = 1_{F^\times}$.

Altogether, this leaves us four exceptional cases of the representations whose composition series we have to determine:

(a) $\chi_{V,\psi}\nu^{1/2}\zeta_1 \times \chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes \omega_0$,

(b) $\chi_{V,\psi}\nu^{1/2}\zeta \times \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega_0$,

(c) $\chi_{V,\psi}\nu^{3/2}\zeta \times \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega_0$,

(d) $\chi_{V,\psi}\nu\zeta \times \chi_{V,\psi}\zeta \rtimes \omega_0$, where $\zeta^2 = \zeta_1^2 = \zeta_2^2 = 1_{F^\times}$ and $\zeta_1 \neq \zeta_2$.

These cases are treated in the following subsection.

3.2.3. *Exceptional cases.* All the equalities that follow are given in semisimplifications. We obtain desired composition series using case-by-case examination:
(a) Write $\chi_{V,\psi}\nu^{1/2}\zeta_i \rtimes \omega_0 = \chi_{V,\psi}\, sp_{\zeta_i,1} + \omega^+_{\psi_{a_i},1}$ for $i = 1, 2$. In R, we have

$$\begin{aligned}\chi_{V,\psi}\nu^{1/2}\zeta_1 \times \chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes \omega_0 &= \chi_{V,\psi}\nu^{1/2}\zeta_2 \times \chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes \omega_0\\ &= \chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes sp_{\zeta_2,1} + \chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes \omega^+_{\psi_{a_2},1}\\ &= \chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes sp_{\zeta_1,1} + \chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes \omega^+_{\psi_{a_1},1}.\end{aligned}$$

Using standard calculations, we obtain

$$\begin{aligned}R_{\widetilde{P_1}}(\chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes sp_{\zeta_2,1}) = \chi_{V,\psi}\nu^{-1/2}\zeta_1 \otimes sp_{\zeta_1,1} &+ \chi_{V,\psi}\nu^{1/2}\zeta_1 \otimes sp_{\zeta_2,1}\\ &+ \chi_{V,\psi}\nu^{1/2}\zeta_2 \otimes sp_{\zeta_1,1} + \chi_{V,\psi}\nu^{1/2}\zeta_2 \otimes \omega^+_{\psi_{a_1},1}\end{aligned}$$

and

$$\begin{aligned}R_{\widetilde{P_2}}(\chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes sp_{\zeta_2,1}) = \chi_{V,\psi}\nu^{-1/2}\zeta_1 &\times \chi_{V,\psi}\nu^{1/2}\zeta_2 \otimes \omega_0\\ &+ \chi_{V,\psi}\nu^{1/2}\zeta_1 \times \chi_{V,\psi}\nu^{1/2}\zeta_2 \otimes \omega_0.\end{aligned}$$

The last equality implies that the length of $\chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes sp_{\zeta_2,1}$ is no more than 2. If $\chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes sp_{\zeta_2,1}$ were an irreducible representation, then it would have to be equal either to $\chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes sp_{\zeta_1,1}$ or to $\chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes \omega^+_{\psi_{a_1},1}$, but Jacquet modules of those two representations show that this is not the case. So, we write $\chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes sp_{\zeta_2,1} = \rho_1 + \rho_2$, where ρ_1 and ρ_2 are irreducible representations such that

$$R_{\widetilde{P}_2}(\rho_1) = \chi_{V,\psi}\nu^{-1/2}\zeta_1 \times \chi_{V,\psi}\nu^{1/2}\zeta_2 \otimes \omega_0,$$
$$R_{\widetilde{P}_2}(\rho_2) = \chi_{V,\psi}\nu^{1/2}\zeta_1 \times \chi_{V,\psi}\nu^{1/2}\zeta_2 \otimes \omega_0.$$

Clearly, ρ_2 is square-integrable (since $\rho_2 = \langle \chi_{V,\psi}\nu^{1/2}\zeta_1, \chi_{V,\psi}\nu^{1/2}\zeta_2; \omega_0\rangle$) and $\rho_1 = \langle \chi_{V,\psi}\nu^{1/2}\zeta_2; \omega^+_{\psi_{a_1},1}\rangle$.

Reasoning in the same way, we obtain that $\chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes \omega^+_{\psi_{a_2},1} = \rho_3 + \rho_4$, where ρ_3 and ρ_4 are irreducible representations such that

$$R_{\widetilde{P}_2}(\rho_3) = \chi_{V,\psi}\nu^{-1/2}\zeta_1 \times \chi_{V,\psi}\nu^{-1/2}\zeta_2 \otimes \omega_0,$$
$$R_{\widetilde{P}_2}(\rho_4) = \chi_{V,\psi}\nu^{1/2}\zeta_1 \times \chi_{V,\psi}\nu^{-1/2}\zeta_2 \otimes \omega_0.$$

So, ρ_3 is a strongly negative representation, while $\rho_4 = \langle \chi_{V,\psi}\nu^{1/2}\zeta_1; \omega^+_{\psi_{a_2},1}\rangle$. Using Jacquet modules again, we easily obtain the composition factors of the representations above. Thus, we conclude:

Proposition 3.11. *Let $\zeta_1, \zeta_2 \in \widehat{F^\times}$ such that $\zeta_i^2 = 1_{F^\times}$ for $i = 1, 2$ (with $\zeta_1 \neq \zeta_2$). Than the representations*

$$\chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes \omega^+_{\psi_{a_1},1}, \quad \chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes sp_{\zeta_1,1},$$
$$\chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes \omega^+_{\psi_{a_2},1}, \quad \chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes sp_{\zeta_2,1}$$

are reducible and $\chi_{V,\psi}\nu^{1/2}\zeta_1 \times \chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes \omega_0$ is a representation of length 4. The representations $\chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes sp_{\zeta_2,1}$ and $\chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes sp_{\zeta_1,1}$ have exactly one irreducible subquotient in common; that subquotient is square-integrable, and we denote it with σ (that is, $\sigma = \langle \chi_{V,\psi}\nu^{1/2}\zeta_1, \chi_{V,\psi}\nu^{1/2}\zeta_2; \omega_0\rangle$). Also, the unique irreducible common subquotient of

$$\chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes \omega^+_{\psi_{a_2},1} \quad \textit{and} \quad \chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes \omega^+_{\psi_{a_1},1}$$

is a strongly negative representation; we denote it by ρ_{sneg}. In R, we have

$$\chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes sp_{\zeta_2,1} = \sigma + \langle \chi_{V,\psi}\nu^{1/2}\zeta_2; \omega^+_{\psi_{a_1},1}\rangle,$$
$$\chi_{V,\psi}\nu^{1/2}\zeta_1 \rtimes \omega^+_{\psi_{a_2},1} = \langle \chi_{V,\psi}\nu^{1/2}\zeta_1; \omega^+_{\psi_{a_2},1}\rangle + \rho_{\mathrm{sneg}},$$
$$\chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes sp_{\zeta_1,1} = \sigma + \langle \chi_{V,\psi}\nu^{1/2}\zeta_1; \omega^+_{\psi_{a_2},1}\rangle,$$
$$\chi_{V,\psi}\nu^{1/2}\zeta_2 \rtimes \omega^+_{\psi_{a_1},1} = \langle \chi_{V,\psi}\nu^{1/2}\zeta_2; \omega^+_{\psi_{a_1},1}\rangle + \rho_{\mathrm{sneg}}.$$

(b) In this case, we have

$$\begin{aligned}\chi_{V,\psi}\nu^{1/2}\zeta \times \chi_{V,\psi}\nu^{-1/2}\zeta \rtimes \omega_0 &= \chi_{V,\psi}\nu^{1/2}\zeta \rtimes sp_{\zeta,1} + \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega^+_{\psi_a,1}\\ &= \chi_{V,\psi}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0 + \chi_{V,\psi}\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0.\end{aligned}$$

From Jacquet modules, we get

$$\begin{aligned}
R_{\widetilde{P}_1}(\chi_{V,\psi}\nu^{1/2}\zeta \rtimes sp_{\zeta,1}) &= 2\chi_{V,\psi}\nu^{1/2}\zeta \otimes sp_{\zeta,1} + \chi_{V,\psi}\nu^{-1/2}\zeta \otimes sp_{\zeta,1} \\
&\quad + \chi_{V,\psi}\nu^{1/2}\zeta \otimes \omega^+_{\psi_a,1}, \\
R_{\widetilde{P}_1}(\chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega^+_{\psi_1}) &= 2\chi_{V,\psi}\nu^{-1/2}\zeta \otimes \omega^+_{\psi_a,1} + \chi_{V,\psi}\nu^{1/2}\zeta \otimes \omega^+_{\psi_a,1} \\
&\quad + \chi_{V,\psi}\nu^{-1/2}\zeta \otimes sp_{\zeta,1}, \\
R_{\widetilde{P}_1}(\chi_{V,\psi}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0) &= 2\chi_{V,\psi}\nu^{1/2}\zeta \otimes sp_{\zeta,1} + 2\chi_{V,\psi}\nu^{1/2}\zeta \otimes \omega^+_{\psi_a,1}, \\
R_{\widetilde{P}_1}(\chi_{V,\psi}\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0) &= 2\chi_{V,\psi}\nu^{-1/2}\zeta \otimes sp_{\zeta,1} + 2\chi_{V,\psi}\nu^{-1/2}\zeta \otimes \omega^+_{\psi_a,1}.
\end{aligned}$$

From preceding Jacquet modules we conclude, as in [Tadić 1994, Chapter 3], that $\chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega^+_{\psi_a,1}$ and $\chi_{V,\psi}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0$ have an irreducible subquotient in common, which is different from both $\chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega^+_{\psi_a,1}$ and $\chi_{V,\psi}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0$. For simplicity of the notation, we let ρ_1 stand for this subquotient. Thus $R_{\widetilde{P}_1}(\rho_1) = \chi_{V,\psi}\nu^{1/2}\zeta \otimes \omega^+_{\psi_a,1}$.

In the same way, let ρ_2 be an irreducible common subquotient that

$$\chi_{V,\psi}\nu^{1/2}\zeta \rtimes sp_{\zeta,1} \quad \text{and} \quad \chi_{V,\psi}\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0$$

have in common. Then $R_{\widetilde{P}_1}(\rho_2) = \chi_{V,\psi}\nu^{-1/2}\zeta \otimes sp_{\zeta,1}$.

The representations $\chi_{V,\psi}\zeta\, 1_{\mathrm{GL}(2)} \otimes \omega_0$ and $\chi_{V,\psi}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \otimes \omega_0$ are irreducible and unitary. The multiplicity of $\chi_{V,\psi}\zeta\, 1_{\mathrm{GL}(2)} \otimes \omega_0$ in $R_{\widetilde{P}_2}(\chi_{V,\psi}\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0)$ is equal to 2, which implies that length of $\chi_{V,\psi}\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0$ is 2. Analogously, the length of $\chi_{V,\psi}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0$ also equals 2.

Now we write $\chi_{V,\psi}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0 = \rho_1 + \rho_3$ and $\chi_{V,\psi}\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0 = \rho_2 + \rho_4$. Observe that

$$\begin{aligned}
R_{\widetilde{P}_1}(\rho_3) &= 2\chi_{V,\psi}\nu^{1/2}\zeta \otimes sp_{\zeta,1} + \chi_{V,\psi}\nu^{1/2}\zeta \otimes \omega^+_{\psi_a,1}, \\
R_{\widetilde{P}_1}(\rho_4) &= \chi_{V,\psi}\nu^{-1/2}\zeta \otimes sp_{\zeta,1} + 2\chi_{V,\psi}\nu^{-1/2}\zeta \otimes \omega^+_{\psi_a,1}.
\end{aligned}$$

We immediately get this:

Proposition 3.12. *Let $\zeta \in \widehat{F^\times}$ such that $\zeta^2 = 1_{F^\times}$. Then the representations*

$$\zeta\, 1_{\widetilde{\mathrm{GL}(2)}} \rtimes \omega_0, \quad \chi_{V,\psi}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0, \quad \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega^+_{\psi_a,1} \quad \chi_{V,\psi}\nu^{1/2}\zeta \rtimes sp_{\zeta,1}$$

are reducible and $\chi_{V,\psi}\nu^{1/2}\zeta \times \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega_0$ is a representation of length 4. *The representations*

$$\chi_{V,\psi}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0 \quad \textit{and} \quad \nu^{1/2}\chi_{V,\psi}\zeta \rtimes \omega^+_{\psi_a,1}$$

(respectively $\chi_{V,\psi}\nu^{1/2}\zeta \rtimes sp_{\zeta,1}$) have exactly one irreducible subquotient in common, which is tempered and denoted by τ_1 (respectively τ_2). Observe that

$$\tau_1 = \langle \chi_{V,\psi}\nu^{1/2}\zeta; \omega^+_{\psi_a,1} \rangle \quad \textit{and} \quad \tau_2 = \langle \chi_{V,\psi}\nu^{1/2}\zeta, \chi_{V,\psi}\nu^{1/2}\zeta; \omega_0 \rangle.$$

Also, the unique irreducible common subquotient of

$$\chi_{V,\psi}\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0 \quad \textit{and} \quad \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega^+_{\psi_a,1}$$

is a negative representation, which we denote by ρ_{neg}. *In* R, *we have*

$$\chi_{V,\psi}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0 = \tau_1 + \tau_2, \qquad \chi_{V,\psi}\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0 = \rho_{\mathrm{neg}} + \langle \chi_{V,\psi}\zeta\, 1_{\mathrm{GL}(2)}; \omega_0\rangle$$
$$\chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega^+_{\psi_a,1} = \tau_1 + \rho_{\mathrm{neg}}, \quad \chi_{V,\psi}\nu^{1/2}\zeta \rtimes sp_{\zeta,1} = \tau_2 + \langle \chi_{V,\psi}\zeta\, 1_{\mathrm{GL}(2)}; \omega_0\rangle.$$

(c) In this case we have

$$\begin{aligned}\chi_{V,\psi}\nu^{3/2}\zeta \times \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega_0 &= \chi_{V,\psi}\nu^{3/2}\zeta \rtimes sp_{\zeta,1} + \chi_{V,\psi}\nu^{3/2}\zeta \rtimes \omega^+_{\psi_a,1}\\ &= \chi_{V,\psi}\nu\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0 + \chi_{V,\psi}\nu\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0.\end{aligned}$$

Observe that $\chi_{V,\psi}\nu^{3/2}\zeta \times \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega_0$ is a regular representation. So, it is a representation of the length $2^2 = 4$ by [Tadić 1998b] (there only the techniques of Jacquet modules were used, so they can be applied in our case). Since the irreducible subquotients of $\chi_{V,\psi}\nu^{3/2}\zeta \times \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega_0$ are

$$\langle \chi_{V,\psi}\nu^{3/2}\zeta, \chi_{V,\psi}\nu^{1/2}\zeta; \omega_0\rangle, \quad \langle \chi_{V,\psi}\nu\zeta\, 1_{\mathrm{GL}(2)}; \omega_0\rangle, \quad \omega^+_{\psi_a,2}, \quad \langle \chi_{V,\psi}\nu^{3/2}\zeta; \omega^+_{\psi_a,1}\rangle,$$

using Jacquet modules we easily obtain the following proposition:

Proposition 3.13. *Let* $\zeta \in \widehat{F^\times}$ *such that* $\zeta^2 = 1_{F^\times}$. *Then the representations*

$$\chi_{V,\psi}\nu^{3/2}\zeta \rtimes sp_{\zeta,1}, \qquad \chi_{V,\psi}\nu^{3/2}\zeta \rtimes \omega^+_{\psi_a,1},$$
$$\chi_{V,\psi}\nu\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0, \qquad \chi_{V,\psi}\nu\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0$$

are reducible and $\chi_{V,\psi}\nu^{3/2}\zeta \times \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega_0$ *is a representation of length* 4. *The unique irreducible common subquotient of the representations* $\chi_{V,\psi}\nu^{3/2}\zeta \rtimes sp_{\zeta,1}$ *and* $\chi_{V,\psi}\nu\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0$ *is square-integrable. In* R, *we have*

$$\begin{aligned}\chi_{V,\psi}\nu^{3/2}\zeta \rtimes sp_{\zeta,1} &= \langle \chi_{V,\psi}\nu^{3/2}\zeta, \chi_{V,\psi}\nu^{1/2}\zeta; \omega_0\rangle + \langle \chi_{V,\psi}\nu\zeta\, 1_{\mathrm{GL}(2)}; \omega_0\rangle,\\ \chi_{V,\psi}\nu^{3/2}\zeta \rtimes \omega^+_{\psi_a,1} &= \omega^+_{\psi_a,2} + \langle \chi_{V,\psi}\nu^{3/2}\zeta; \omega^+_{\psi_a,1}\rangle,\\ \chi_{V,\psi}\nu\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0 &= \langle \chi_{V,\psi}\nu^{3/2}\zeta, \chi_{V,\psi}\nu^{1/2}\zeta; \omega_0\rangle + \langle \chi_{V,\psi}\nu^{3/2}\zeta; \omega^+_{\psi_a,1}\rangle,\\ \chi_{V,\psi}\nu\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0 &= \langle \chi_{V,\psi}\nu\zeta\, 1_{\mathrm{GL}(2)}; \omega_0\rangle + \omega^+_{\psi_a,2}.\end{aligned}$$

(d) In this case,

$$\chi_{V,\psi}\nu\zeta \times \chi_{V,\psi}\zeta \rtimes \omega_0 = \chi_{V,\psi}\nu^{1/2}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0 + \chi_{V,\psi}\nu^{1/2}\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0.$$

Since it isn't known yet if the results related to the R-groups [Goldberg 1994] also hold for metaplectic groups, this case will not be solved using only the method Jacquet modules. Tadić [1998a] used a combination of Jacquet modules techniques and knowledge about R-groups for symplectic groups to determine the composition series of the representations similar to this one (for symplectic groups). We resolve this case using theta correspondence.

Lemma 3.14. *The following equalities hold*:

(1) $\Theta(\zeta\nu\otimes\zeta\rtimes 1, R_{P_1}(\omega_{2,2})) = \chi_{V,\psi}\nu^{-1}\zeta\times\chi_{V,\psi}\zeta\rtimes\omega_0$,

(2) $\Theta(\zeta\nu^{-1}\otimes\zeta\rtimes 1, R_{P_1}(\omega_{2,2})) = \chi_{V,\psi}\nu\zeta\times\chi_{V,\psi}\zeta\rtimes\omega_0$,

(3) $\Theta(\zeta\otimes\zeta\nu\rtimes 1, R_{P_1}(\omega_{2,2})) = \chi_{V,\psi}\zeta\times\chi_{V,\psi}\nu\zeta\rtimes\omega_0$,

(4) $\Theta(\chi_{V,\psi}\nu\zeta\otimes\chi_{V,\psi}\zeta\rtimes\omega_0, R_{\widetilde{P}_1}(\omega_{2,2})) = \zeta\nu^{-1}\times\zeta\rtimes 1$,

(5) $\Theta(\chi_{V,\psi}\nu^{-1}\zeta\otimes\chi_{V,\psi}\zeta\rtimes\omega_0, R_{\widetilde{P}_1}(\omega_{2,2})) = \zeta\nu\times\zeta\rtimes 1$,

(6) $\Theta(\chi_{V,\psi}\zeta\otimes\chi_{V,\psi}\nu\zeta\rtimes\omega_0, R_{\widetilde{P}_1}(\omega_{2,2})) = \zeta\times\zeta\nu\rtimes 1$.

Proof. Recall that $R_{P_1}(\omega_{2,2})$ has the filtration in which

- $I_{10}=\nu^{1/2}\otimes\omega_{2,1}$ is the quotient, and
- $I_{11}=\mathrm{Ind}_{\mathrm{GL}(1)\times\widetilde{P}_1\times O(3)}^{M_1\times\widetilde{\mathrm{Sp}(2)}}(\chi_{V,\psi}\Sigma_1'\otimes\omega_{1,1})$ is the subrepresentation.

We will prove (1); the proofs of (2)–(6) are analogous. In the same way as in the second proof of Proposition 3.4, we get

$$\Theta(\zeta\nu\otimes\zeta\rtimes 1, R_{P_1}(\omega_{2,2})) = \Theta(\zeta\nu\otimes\zeta\rtimes 1, I_{11}),$$

so it is sufficient to show $\Theta(\zeta\nu\otimes\zeta\rtimes 1, I_{11}) = \chi_{V,\psi}\nu^{-1}\zeta\times\chi_{V,\psi}\zeta\rtimes\omega_0$. It can be seen easily that there is an $\mathrm{GL}(1)\times\widetilde{M_1}\times O(3)$-invariant epimorphism

$$\chi_{V,\psi}\Sigma_1'\otimes\omega_{1,1}\to\chi_{V,\psi}\zeta\nu^{-1}\otimes\chi_{V,\psi}\zeta\rtimes\omega_0\otimes\zeta\nu\otimes\zeta\rtimes 1.$$

Consequently, we get an $M_1\times\widetilde{\mathrm{Sp}(2)}$-invariant epimorphism

$$I_{11}=\mathrm{Ind}_{\mathrm{GL}(1)\times\widetilde{P}_1\times O(3)}^{M_1\times\widetilde{\mathrm{Sp}(2)}}(\chi_{V,\psi}\Sigma_1'\otimes\omega_{1,1})\to\zeta\nu\otimes\zeta\rtimes 1\otimes\chi_{V,\psi}\zeta\nu^{-1}\times\chi_{V,\psi}\zeta\rtimes\omega_0,$$

so we conclude that $\chi_{V,\psi}\zeta\nu^{-1}\times\chi_{V,\psi}\zeta\rtimes\omega_0$ is a quotient of $\Theta(\zeta\nu\otimes\zeta\rtimes 1, I_{11})$.

We prove that $\Theta(\zeta\nu\otimes\zeta\rtimes 1, I_{11})$ is also a quotient of $\chi_{V,\psi}\zeta\nu^{-1}\times\chi_{V,\psi}\zeta\rtimes\omega_0$. Let $\varphi\in\mathrm{Hom}(I_{11},\zeta\nu\otimes\zeta\rtimes 1\otimes\Theta(\zeta\nu\otimes\zeta\rtimes 1, I_{11}))$. Using the second Frobenius reciprocity, as before, we get

$$\begin{aligned}&\mathrm{Hom}(I_{11},\zeta\nu\otimes\zeta\rtimes 1\otimes\Theta(\zeta\nu\otimes\zeta\rtimes 1, I_{11}))\\&\qquad\cong\mathrm{Hom}(\chi_{V,\psi}\Sigma_1'\otimes\omega_{1,1},\zeta\nu\otimes\zeta\rtimes 1\otimes R_{\overline{\widetilde{P}}_1}(\Theta(\zeta\nu\otimes\zeta\rtimes 1, I_{11})));\end{aligned}$$

let φ_0 be an element corresponding to φ. Since the representations $\zeta\rtimes 1$ and $\chi_{V,\psi}\zeta\rtimes\omega_0$ are irreducible, the image of φ_0 equals $\zeta\nu\otimes\zeta\rtimes 1\otimes\chi_{V,\psi}\zeta\nu^{-1}\otimes\chi_{V,\psi}\zeta\rtimes\omega_0$. Reasoning as before, we get that the image of φ is a quotient of $\zeta\nu\otimes\zeta\rtimes 1\otimes\chi_{V,\psi}\zeta\nu^{-1}\times\chi_{V,\psi}\zeta\rtimes\omega_0$. Finally, $\Theta(\zeta\nu\otimes\zeta\rtimes 1, I_{11})$ is a quotient of $\chi_{V,\psi}\zeta\nu^{-1}\times\chi_{V,\psi}\zeta\rtimes\omega_0$. Hence $\Theta(\zeta\nu\otimes\zeta\rtimes 1, I_{11})=\chi_{V,\psi}\zeta\nu^{-1}\times\chi_{V,\psi}\zeta\rtimes\omega_0$. □

Proposition 3.15. *Let $\zeta \in \widehat{F^\times}$ such that $\zeta^2 = 1_{F^\times}$. Then the representations*

$$\chi_{V,\psi}\nu^{1/2}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0 \quad \text{and} \quad \chi_{V,\psi}\nu^{1/2}\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0$$

are irreducible and $\chi_{V,\psi}\nu\zeta \times \chi_{V,\psi}\zeta \rtimes \omega_0$ is a representation of length 2. Also

$$\chi_{V,\psi}\nu^{1/2}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0 = \langle \chi_{V,\psi}\nu\zeta; \chi_{V,\psi}\zeta \rtimes \omega_0\rangle,$$
$$\chi_{V,\psi}\nu^{1/2}\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0 = \langle \chi_{V,\psi}\nu^{1/2}\zeta\, 1_{\mathrm{GL}(2)}; \omega_0\rangle.$$

Proof. Suppose on the contrary that the representation $\chi_{V,\psi}\nu^{1/2}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0$ reduces. Jacquet modules imply that length of this representation is at most 2. Choose π_1 and π_2 so the equality $\chi_{V,\psi}\nu^{1/2}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0 = \pi_1 + \pi_2$ holds in R. Also suppose $R_{\widetilde{P}_1}(\pi_1) = \chi_{V,\psi}\zeta \otimes \chi_{V,\psi}\zeta\nu \rtimes \omega_0$ and $R_{\widetilde{P}_1}(\pi_2) = \chi_{V,\psi}\zeta\nu \otimes \chi_{V,\psi}\zeta \rtimes \omega_0$. Frobenius reciprocity implies

$$\mathrm{Hom}(\omega_{2,2}, \pi_1 \otimes \zeta \times \zeta\nu^{-1} \rtimes 1) \cong \mathrm{Hom}(R_{P_1}(\omega_{2,2}), \pi_1 \otimes \zeta \otimes \zeta\nu^{-1} \rtimes 1).$$

Using Lemma 3.14 we obtain

$$\mathrm{Hom}(R_{P_1}(\omega_{2,2}), \pi_1 \otimes \zeta \otimes \zeta\nu^{-1} \rtimes 1) \cong \mathrm{Hom}(\chi_{V,\psi}\zeta \times \chi_{V,\psi}\nu^{-1}\zeta \rtimes \omega_0, \pi_1) \neq 0,$$

because π_1 is a quotient of $\chi_{V,\psi}\zeta \times \chi_{V,\psi}\nu\zeta \rtimes \omega_0$. So, $\Theta(\pi_1, 2) \neq 0$.

The representation $\chi_{V,\psi}\zeta \otimes \chi_{V,\psi}\zeta\nu \rtimes \omega_0 \otimes \Theta(\pi_1, 2)$ is a quotient of $R_{\widetilde{P}_1}(\omega_{2,2})$. Lemma 3.14 implies that $\Theta(\pi_1, 2)$ is a quotient of $\zeta \times \zeta\nu \rtimes 1$. Listing quotients of $\zeta \times \zeta\nu \rtimes 1$ we get the possibilities

(a) $\Theta(\pi_1, 2) = \zeta \times \zeta\nu \rtimes 1$,

(b) $\Theta(\pi_1, 2) = \nu^{1/2}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes 1$,

(c) $\Theta(\pi_1, 2) = \nu^{-1/2}\zeta\, 1_{\mathrm{GL}(2)} \rtimes 1$.

Suppose that (a) holds. Obviously, $\pi_1 \otimes \zeta\nu^{-1} \otimes \zeta \rtimes 1$ is then a quotient of $R_{P_1}(\omega_{2,2})$, since it is a quotient of $\pi_1 \otimes R_{P_1}(\zeta \times \zeta\nu \rtimes 1)$. This implies that π_1 is a quotient of $\chi_{V,\psi}\zeta\nu \times \chi_{V,\psi}\zeta \rtimes \omega_0$ and $R_{\widetilde{P}_1}(\pi_1)$ contains $\chi_{V,\psi}\zeta\nu^{-1} \otimes \chi_{V,\psi}\zeta \rtimes \omega_0$. This contradicts our assumption on π_1.

Similarly, using Jacquet modules, we obtain contradiction with (b) and (c). So, $\chi_{V,\psi}\nu^{1/2}\zeta\, \mathrm{St}_{\mathrm{GL}(2)} \rtimes \omega_0$ is irreducible.

Irreducibility of $\chi_{V,\psi}\nu^{1/2}\zeta\, 1_{\mathrm{GL}(2)} \rtimes \omega_0$ can be proved in the same way. □

4. Unitary dual supported in minimal parabolic subgroup

Let π be an irreducible genuine admissible representation of $\widetilde{\mathrm{Sp}(n)}$. We recall that the contragredient representation is denoted by $\widetilde{\pi}$. We write $\overline{\pi}$ for the complex conjugate representation of the representation π. The representation π is called Hermitian if $\pi \simeq \widetilde{\overline{\pi}}$. It is well known that every unitary representation is Hermitian. For a deeper discussion, we refer the reader to [Muić and Tadić 2007].

Suppose that $\Delta_1, \dots, \Delta_k$ is a sequence of segments such that $e(\Delta_1) \geq \cdots \geq e(\Delta_k) > 0$, let σ_{neg} be a negative representation of some $\widetilde{\mathrm{Sp}(n')}$. From [Hanzer and Muić 2010, Theorem 4.5(v)], we directly get

$$\overline{\langle \Delta_1, \dots, \Delta_k; \sigma_{\text{neg}} \rangle} = \langle \overline{\Delta}_1, \dots, \overline{\Delta}_k; \overline{\sigma}_{\text{neg}} \rangle.$$

Also, we have an epimorphism $\langle \widetilde{\Delta_1} \rangle \times \cdots \times \langle \widetilde{\Delta_k} \rangle \rtimes \widetilde{\sigma}_{\text{neg}} \to \langle \Delta_1, \dots, \Delta_k; \sigma_{\text{neg}} \rangle^{\sim}$. We know that the group $\mathrm{GSp}(n)$ acts on $\widetilde{\mathrm{Sp}(n)}$, by [Mœglin et al. 1987, II.1(3)]. Moreover, by [ibid., page 92], this action extends to the action on irreducible representations, which is equivalent to taking contragredients. We choose an element $\eta' = (1, \eta) \in \mathrm{GSp}(n)$, where $\eta \in \mathrm{GSp}(n')$ is an element with similitude equal to -1, and 1 denotes the identity acting on the GL part. Thus, we obtain an epimorphism

$$\alpha \langle \widetilde{\Delta_1} \rangle \times \cdots \times \alpha \langle \widetilde{\Delta_k} \rangle \rtimes \widetilde{\sigma}_{\text{neg}}^{\eta} \to \langle \Delta_1, \dots, \Delta_k; \sigma_{\text{neg}} \rangle^{\sim \eta'}.$$

Since $\widetilde{\sigma}_{\text{neg}}^{\eta} \simeq \sigma_{\text{neg}}$, we have $\langle \Delta_1, \dots, \Delta_k; \sigma_{\text{neg}} \rangle^{\sim} = \langle \alpha \Delta_1, \dots, \alpha \Delta_k; \widetilde{\sigma}_{\text{neg}} \rangle$.

Remark. When we are dealing with the action of the group of similitudes on the symplectic groups, α does not appear in the situation similar to the one above. However, since the action of $\mathrm{GSp}(n)$ on the metaplectic group is not trivial on its center μ_2, one has to compare the action η on the metaplectic part of the Levi subgroup with the action of η' on the whole Levi subgroup. The calculation is not very complicated and resembles the calculations in [Hanzer and Muić 2010, Lemma 3.2].

First, we classify Hermitian irreducible genuine representations.

Proposition 4.1. *Let $\chi, \zeta, \zeta_1, \zeta_2 \in \widehat{F^{\times}}$ such that $\zeta^2 = \zeta_i^2 = 1_{F^{\times}}$ for $i = 1, 2$, with ζ_1 and ζ_2 not necessarily different. Let $s, s_1, s_2 > 0$. The following families of representations are Hermitian and exhaust all irreducible Hermitian genuine representations of $\widetilde{\mathrm{Sp}(2)}$ supported in the minimal parabolic subgroup $\widetilde{P_{(1,1)}}$:*

(1) *irreducible tempered representations supported in $\widetilde{P_{(1,1)}}$,*
(2) $\langle \chi_{V,\psi} \nu^s \chi, \chi_{V,\psi} \nu^s \chi^{-1}; \omega_0 \rangle$,
(3) $\langle \chi_{V,\psi} \nu^{s_1} \zeta_1, \chi_{V,\psi} \nu^{s_2} \zeta_2; \omega_0 \rangle$,
(4) $\langle \chi_{V,\psi} \nu^s \zeta\, 1_{\mathrm{GL}(2)}; \omega_0 \rangle$,
(5) $\langle \chi_{V,\psi} \nu^s \zeta; \chi_{V,\psi} \chi \rtimes \omega_0 \rangle$,
(6) $\langle \chi_{V,\psi} \nu^s \zeta; \omega^{+}_{\psi_a,1} \rangle$,
(7) $\omega^{+}_{\psi_a,2}$.

Proof. Using the reasoning before this proposition, we see that a representation $\langle \Delta_1, \dots, \Delta_k; \sigma_{\text{neg}} \rangle$ is Hermitian, if and only if

$$\langle \Delta_1, \dots, \Delta_k; \sigma_{\text{neg}} \rangle = \langle \alpha \overline{\Delta}_1, \dots, \alpha \overline{\Delta}_k; \overline{\widetilde{\sigma}}_{\text{neg}} \rangle.$$

The representation σ_{neg} also has to be Hermitian. Now we just check this requirement on the set of all irreducible representations of $\widetilde{\mathrm{Sp}(2)}$ with the support in the minimal parabolic subgroup; we have classified them in the previous section. For example, if we analyze the representation $\Pi = \langle \chi_{V,\psi}\nu^{s_1}\chi_1, \chi_{V,\psi}\nu^{s_2}\chi_2; \omega_0\rangle$ with $s_2 \geq s_1 > 0$, we have

$$\overline{\overline{\Pi}} = \langle \alpha\overline{\chi_{V,\psi}\nu^{s_1}\chi_1}, \alpha\overline{\chi_{V,\psi}\nu^{s_2}\chi_2}; \omega_0\rangle.$$

Now we see that this representation is isomorphic to Π if and only if $\chi_1^2 = 1 = \chi_2^2$ or $s_1 = s_2$ and $\chi_1^{-1} = \chi_2$. This gives us the second and the third case from the proposition. All other cases are dealt with analogously. □

Theorem 4.2. *Let $\chi, \zeta, \zeta_1, \zeta_2 \in (\widehat{F^\times})$ such that $\zeta^2 = \zeta_i^2 = 1_{F^\times}$ for $i = 1, 2$, with ζ_1 and ζ_2 not necessarily different. The following families of representations are unitary and exhaust all irreducible unitary genuine representations of $\widetilde{\mathrm{Sp}(2)}$ that are supported in the minimal parabolic subgroup $\widetilde{P_{(1,1)}}$:*

(1) *irreducible tempered representations supported in $\widetilde{P_{(1,1)}}$.*

(2) $\langle \chi_{V,\psi}\nu^s\chi, \chi_{V,\psi}\nu^s\chi^{-1}; \omega_0\rangle$ *for* $0 < s \leq 1/2$,

(3) $\langle \chi_{V,\psi}\nu^{s_1}\zeta_1, \chi_{V,\psi}\nu^{s_2}\zeta_2; \omega_0\rangle$ *for* $s_2 \leq s_1$ *and* $0 < s_1 \leq 1/2$,

(4) $\langle \chi_{V,\psi}\nu^s\zeta; \chi_{V,\psi}\chi \rtimes \omega_0\rangle$ *for* $0 < s \leq 1/2$,

(5) $\langle \chi_{V,\psi}\nu^s\zeta; \omega^+_{\psi_a,1}\rangle$ *for* $s \leq 1/2$,

(6) $\omega^+_{\psi_a,2}$.

Proof. We first review some basic facts of representation theory of reductive groups, which directly carry over to the case of metaplectic groups.

Unitarizability of the complementary series. As explained in detail in [Tadić 1993, Section 3], it is enough to have a continuous family of $\widetilde{\mathrm{Sp}(2)}$-invariant hermitian forms (and the representations should be realized on one space — a compact picture). Then, a linear algebra argument (involving finite-dimensional representations of a compact subgroup) ensures that if this family of hermitian forms is positive definite at one point, it has to be positive definite everywhere, and this finishes the argument. So, we first have to show that, the restriction of an irreducible admissible (hermitian) representation π of $\widetilde{\mathrm{Sp}(2)}$ to the inverse image $\widetilde{K}$ in $\widetilde{\mathrm{Sp}(2)}$ of a maximal, good compact subgroup of Sp(2) (for example, $K = \mathrm{Sp}(2, O_F)$, where O_F is the ring of integers of F) decomposes into a direct sum of irreducible representations of $\widetilde{K}$ with finite multiplicities. But this follows directly from the admissibility of the representation π. Second, we have to have a way to form a continuous families of hermitian forms. This is obtained using intertwining operators, in the same way as for the algebraic groups. To define them, we note that the unipotent radicals of the standard parabolic subgroups of Sp(2) are split in $\widetilde{\mathrm{Sp}(2)}$.

Then we can define standard intertwining operators for the complex argument deep enough in the Weyl chamber in the same way as in [Shahidi 1981]. These operators can be meromorphically continued, using the results on filtration via Bruhat cells; see [Casselman 1995, Section 6 and 7] or [Muić 2008]. The arguments in the last reference carry over to the metaplectic groups without change, through the splitting of the unipotent radicals and Frobenius reciprocity, also valid for the metaplectic group. This passage in the construction of the intertwining operators from the linear case to the case of metaplectic groups is explained in detail also in [Zorn 2007; 2010]. We now illustrate how the hermitian form is defined. For example, suppose that $\chi_1, \chi_2 \in \hat{F^\times}$ such that $\chi_1^2 = \chi_2^2 = 1$, so that, for the longest element w_0 of the Weyl group, we have a map $A(s_1, s_2, \chi_1, \chi_2, w_0)$ from

$$\chi_{V,\psi}\chi_1\nu^{s_1} \times \chi_{V,\psi}\chi_2\nu^{s_2} \rtimes \omega_0 \quad \text{to} \quad \chi_{V,\psi}\nu^{-s_1}\chi_1 \times \chi_{V,\psi}\nu^{-s_2}\chi_2 \rtimes \omega_0.$$

Let f_{s_1,s_2}, g_{s_1,s_2} be sections from the compact picture of the induced representation $\chi_{V,\psi}\chi_1\nu^{s_1} \times \chi_{V,\psi}\chi_2\nu^{s_2} \rtimes \omega_0$. Then, a hermitian form indexed by (s_1, s_2) is defined by

$$(f, g)_{(s_1,s_2)} = \int_{\widetilde{K}} A(s_1, s_2, \chi_1, \chi_2, w_0) f_{s_1,s_2}(\widetilde{k})\overline{g_{s_1,s_2}(\widetilde{k})}\, d\widetilde{k}.$$

The $\widetilde{\mathrm{Sp}(2)}$-invariance of this form follows from [Casselman 1981, Theorem 2.4.2] (in the context of totally disconnected groups) and then from Proposition 3.1.3 therein, after normalizing the measure on $\widetilde{P}$ so that $\widetilde{P} \cap \widetilde{K}$ is of volume one (since $\widetilde{P}\widetilde{K} = \widetilde{G}$).

Unitarizability of the ends of the complementary series. For the reductive algebraic groups, the unitarizability of the ends of the complementary series is proved by Miličić [1973] using C^*-algebra arguments. To avoid that (although this argument may also apply in the case of metaplectic groups), we use a similar result, that is, [Tadić 1986, Theorem 2.5]. The proof of this result relies on calculations of the limits of the operators acting on the finite-dimensional complex vector spaces, and the only requirements are admissibility of the irreducible smooth representations in question (our Lemma 2.1) and a result of Bernstein about uniform admissibility. But, since we do not require the generality in which that theorem is posed, we actually do not need Bernstein's argument, since we are dealing with the family of representations in the complementary series — all of them have the same restriction to the compact open subgroup K_1 (which splits in $\widetilde{\mathrm{Sp}}$), and the requirement labeled $(*)$ there is automatically fulfilled. Hecke algebra $H(\widetilde{\mathrm{Sp}(2)}, K_1)$ is defined in the same way as in the case of reductive groups.

The asymptotics of the matrix coefficients of the representations of the metaplectic group. These can also be estimated in terms of Jacquet modules of the representations [Casselman 1981, Section 4]. Indeed, the arguments there rely on the calculation of the spaces of the coinvariants for the unipotent subgroups (which split in $\widetilde{\mathrm{Sp}(2)}$) and spaces of vectors fixed by some small compact subgroups of $\widetilde{\mathrm{Sp}(2)}$. These subgroups can always be taken to belong to the maximal compact subgroup of Sp(2) that splits in $\widetilde{\mathrm{Sp}(2)}$ (if the residual characteristic is odd) or to some smaller open compact subgroup that splits, so we actually take the fixed vectors by these splittings of compact subgroups.

On the other hand, the reducibility points of the principal series for SO(5) are analogous to those for $\widetilde{\mathrm{Sp}(2)}$, so the unbounded areas of [Matić 2010, Figure 1 of Theorem 3.5], through the asymptotics explained above, give rise to the representations with unbounded matrix coefficients. Thus none of these representations are unitarizable (because of the continuity of the hermitian forms on these unbounded parts).

The arguments above (plus the irreducibility of the unitary principal series) were the main tools in the proof of [Matić 2010, Theorem 3.5]; there was only the problem of how to deal with certain isolated representations.

Recall that in R we have $\chi_{V,\psi}\nu^{3/2}\zeta \times \chi_{V,\psi}\nu^{1/2}\zeta \rtimes \omega_0$ is equal to

$$\langle \chi_{V,\psi}\nu^{3/2}\zeta, \chi_{V,\psi}\nu^{1/2}\zeta; \omega_0\rangle + \omega^+_{\psi_a,2} + \langle \chi_{V,\psi}\nu^{3/2}\zeta; \omega^+_{\psi_a,1}\rangle + \langle \chi_{V,\psi}\nu\zeta\, 1_{\mathrm{GL}(2)}; \omega_0\rangle,$$

where $\langle \chi_{V,\psi}\nu^{3/2}\zeta, \chi_{V,\psi}\nu^{1/2}\zeta; \omega_0\rangle$ and $\omega^+_{\psi_a,2}$ are unitarizable. Observe that the representation $\langle \chi_{V,\psi}\nu^{3/2}\zeta; \omega^+_{\psi_a,1}\rangle$ (respectively, $\langle \chi_{V,\psi}\nu\zeta\, 1_{\mathrm{GL}(2)}; \omega_0\rangle$) has Jacquet modules analogous to those of the representation $L(\delta([\nu^{1/2}, \nu^{3/2}]), 1)$ (respectively, $L(\nu^{3/2}, \mathrm{St}_{\mathrm{SO}(3)})$) of the group SO(5). Hence, nonunitarizability of these two representations can be proved analagously to the nonunitarizability of the representations $L(\delta([\nu^{1/2}, \nu^{3/2}]), 1)$ and $L(\nu^{3/2}, \mathrm{St}_{\mathrm{SO}(3)})$, which is a special case of [Hanzer and Tadić 2010, Propositions 4.1 and 4.6]. The arguments used there rely on the Jacquet modules method, which also applies to group $\widetilde{\mathrm{Sp}(2)}$, and the simple fact that every unitary representation is also semisimple. □

5. Unitary dual supported in maximal parabolic subgroups

5.1. ***The Siegel case.*** Using [Hanzer and Muić 2009], [Matić 2010, Proposition 4.1] and previously discussed issues of complementary series and nonunitarizability of the representations indexed by the (geometrically) unbounded pieces of the plane, we directly get the following:

Proposition 5.1. *Let ρ be an irreducible cuspidal representation of* GL(2, F)*. There is at most one $s \geq 0$ such that $\chi_{V,\psi}\nu^s\rho \rtimes \omega_0$ reduces. One of the following holds*:

(1) *If ρ is not self-dual, then $\chi_{V,\psi}\rho \rtimes \omega_0$ is irreducible and unitarizable. Also, the representations $\chi_{V,\psi}\nu^s\rho \rtimes \omega_0$ are irreducible and nonunitarizable for $s > 0$.*

(2) *If ρ is self-dual and $\omega_\rho = 1$, where ω_ρ denotes the central character of ρ, then the representation $\chi_{V,\psi}\rho \rtimes \omega_0$ reduces, while all of the representations $\chi_{V,\psi}\nu^s\rho \rtimes \omega_0$ are nonunitarizable for $s > 0$.*

(3) *If ρ is self-dual and $\omega_\rho \neq 1$, then the unique $s \geq 0$ such that $\chi_{V,\psi}\nu^s\rho \rtimes \omega_0$ reduces is equal to $1/2$. For $0 \leq s \leq 1/2$, the representations $\chi_{V,\psi}\nu^s\rho \rtimes \omega_0$ are all unitarizable; for $s > 1/2$, the representations $\chi_{V,\psi}\nu^s\rho \rtimes \omega_0$ are all nonunitarizable. All irreducible subquotients of $\chi_{V,\psi}\nu^{1/2}\rho \rtimes \omega_0$ are unitarizable.*

5.2. ***The non-Siegel case.*** Hanzer and Muić [2009, Section 5.2] determine the reducibility points of the representations $\chi_{V,\psi}\nu^s\zeta \rtimes \pi$, where $s \in \mathbb{R}$, $\zeta \in \widehat{F^\times}$ and π is an irreducible cuspidal representation of $\widetilde{\mathrm{Sp}(1)}$. After determining the reducibility points, the unitarizability of the induced representations and irreducible subquotients follow in the same way as in Proposition 5.1. For the convenience of the reader, we write down all the results.

To the fixed quadratic character χ_V we attach, as in [Kudla 1996, Chapter V], two odd-orthogonal towers, the $+$-tower and the $-$-tower. We denote by $\Theta^{\pm}(\pi)$ the first appearance of the representation $\Theta(\pi)$ in the respective $\pm$-tower. Analogously, for $r \geq 0$, we denote by $\Theta^{\pm}(\pi, r)$ the lift of the representation π to the r-th level of the respective $\pm$-tower.

Since the representation $\chi_{V,\psi}\nu^s\zeta \rtimes \pi$ is irreducible for $\zeta^2 \neq 1$, we suppose $\zeta^2 = 1$ and consider two cases:

(a) $\zeta \neq 1$. Applying [Hanzer and Muić 2009, Theorem 3.5] we see that $\chi_{V,\psi}\nu^s\zeta \rtimes \pi$ reduces if and only if $\zeta\nu^s \rtimes \Theta^+(\pi)$ reduces (in the $+$-tower) if and only if $\zeta\nu^s \rtimes \Theta^-(\pi)$ reduces (in the $-$-tower).

Now $\Theta^+(\pi)$ is an irreducible cuspidal representation of some of the groups $O(1)$, $O(3)$ or $O(5)$. Let r denote the first occurrence of a nonzero lift of π in the odd orthogonal $+$-tower. We have several cases depending on r:

- If $r = 0$, that is, if π equals $\omega_{\psi_a^-,1}$, which is an odd part of the Weil representation attached to additive character ψ, then $\Theta^+(\pi, 0) = \mathrm{sgn}_{O(1)}$, so the representation $\zeta\nu^s \rtimes \mathrm{sgn}_{O(1)}$ reduces if and only if $\zeta\nu^s \rtimes 1$ reduces in SO(3). It is well known that this representation reduces when $s = 1/2$.
- If $r = 1$, the representation $\zeta\nu^s \rtimes \pi$ reduces if and only if the representation $\zeta\nu^s \rtimes \Theta^+(\pi, 1)|_{\mathrm{SO}(3)}$ reduces. As in [Matić 2010], we obtain that the unique s such that $\zeta\nu^s \rtimes \pi$ reduces is equal to $1/2$.
- If $r = 2$, the representation $\zeta\nu^s \rtimes \pi$ reduces if and only if $\zeta\nu^s \rtimes \Theta^+(\pi, 2)$ reduces, and that is if and only if $\zeta\nu^s \rtimes \Theta^+(\pi, 2)|_{\mathrm{SO}(5)}$ reduces. We do

not know if the representation $\Theta^+(\pi, 2)$ is generic, so we turn our attention to the representation $\zeta\nu^s \rtimes \Theta^-(\pi, 0)$, because we know that $\Theta^-(\pi, 0)$ is a nonzero representation of $O(1)$ (since π is cuspidal, the dichotomy conjecture holds). Recall that $\zeta\nu^s \rtimes \Theta^-(\pi, 0)$ reduces for $s = 0$ if and only if $\mu(s, \zeta \otimes \Theta^-(\pi, 0)) \neq 0$ for $s = 0$ and that $\zeta\nu^s \rtimes \Theta^-(\pi, 0)$ reduces for $s_0 > 0$ if and only if $\mu(s, \zeta \otimes \Theta^-(\pi, 0))$ has a pole for $s = s_0$. In the same way as in [Hanzer and Muić 2009, Section 5.2, case 3], we obtain $\mu(s, \zeta\otimes\Theta^-(\pi, 0)) \cong \mu(s, \zeta\otimes JL(\Theta^-(\pi, 0)))$, where $JL(\Theta^-(\pi, 0)))$ denotes the Jacquet–Langlands lift of $\Theta^-(\pi, 0))$. Now we consider two possibilities:

(I) $\Theta^-(\pi, 0)$ is not one-dimensional. In this case, $JL(\Theta^-(\pi, 0))$ is a cuspidal generic representation of $\mathrm{SO}(3, F)$ and the reducibility point is $s = 1/2$.

(II) $\Theta^-(\pi, 0) = \zeta_1 \circ \nu_D$, where ζ_1 is a quadratic character of $F^\times$, while ν_D is a reduced norm on $D^\times$ (here D is a nonsplit quaternion algebra over F). We have $JL(\Theta^-(\pi, 0)) = \zeta_1 \mathrm{St}_{\mathrm{GL}(2,F)}$. If $\zeta_1 = \zeta$, then the reducibility point is $s = 1/2$, otherwise the reducibility point is $s = 3/2$.

(b) $\zeta = 1$. This case can be completely solved using [Hanzer and Muić 2009, Theorem 4.2]. We again denote by r the first occurrence of nonzero lift of representation π in the odd orthogonal $+$-tower and consider all the possible cases:

- If $r = 0$, then π equals $\omega_{\psi_a^-,1}$ and the representation $\chi_{V,\psi}\nu^s \rtimes \omega_{\psi_a^-,1}$ reduces for $s = \pm 3/2$.
- If $r = 1$, the representation $\chi_{V,\psi}\nu^s \rtimes \pi$ reduces for $s = 1/2$.
- If $r = 2$, the representation $\chi_{V,\psi}\nu^s \rtimes \pi$ reduces for $s = 1/2$.

Acknowledgments

We thank the Erwin Schrödinger Institute in Vienna and J. Schwermer for their hospitality during the workshop “Representation theory of reductive groups — local and global aspects”, where work on this paper began. We also thank M. Tadić for showing us his proof of the unitarizability of the ends of complementary series in [Tadić 1986].

References

[Ban and Jantzen 2009] D. Ban and C. Jantzen, “Langlands quotient theorem for covering groups”, preprint, 2009.

[Banks 1998] W. D. Banks, “Heredity of Whittaker models on the metaplectic group”, *Pacific J. Math.* **185**:1 (1998), 89–96. MR 99m:22016 Zbl 0980.22021

[Bernstein 1992] J. Bernstein, “Representations of p-adic groups”, lecture notes, 1992, available at http://www.math.ubc.ca/~cass/research/pdf/p-adic-book.pdf.

[Bernstein and Zelevinsky 1977] I. N. Bernstein and A. V. Zelevinsky, "Induced representations of reductive $\mathfrak{p}$-adic groups, I", *Ann. Sci. École Norm. Sup.* (4) **10**:4 (1977), 441–472. MR 58 #28310 Zbl 0412.22015

[Casselman 1981] W. Casselman, "A new nonunitarity argument for p-adic representations", *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **28**:3 (1981), 907–928 (1982). MR 84e:22018

[Casselman 1995] W. Casselman, "Introduction to the theory of admissible representations of p-adic reductive groups", preprint, 1995, available at http://www.math.ubc.ca/~cass/research/pdf/p-adic-book.pdf.

[Goldberg 1994] D. Goldberg, "Reducibility of induced representations for $\mathrm{Sp}(2n)$ and $\mathrm{SO}(n)$", *American J. Math.* **116**:5 (1994), 1101–1151. MR 95g:22016 Zbl 0851.22021

[Hanzer 2006] M. Hanzer, "The unitary dual of the Hermitian quaternionic group of split rank 2", *Pacific J. Math.* **226**:2 (2006), 353–388. MR 2007e:22005 Zbl 1143.22009

[Hanzer and Matić 2010] M. Hanzer and I. Matić, "Irreducibility of the unitary principal series of p-adic $\widetilde{\mathrm{Sp}(n)}$", *Manuscr. Math.* **132**:3-4 (2010), 539–547. Zbl 05729742

[Hanzer and Muić 2008] M. Hanzer and G. Muić, "On an algebraic approach to the Zelevinsky classification for classical p-adic groups", *J. Algebra* **320**:8 (2008), 3206–3231. MR 2009k:20107 Zbl 1166.22011

[Hanzer and Muić 2009] M. Hanzer and G. Muić, "Rank one reducibility for metapletic groups via theta correspondence", preprint, 2009, available at http://tinyurl.com/338maxk. To appear in *Canad. J. Math.*

[Hanzer and Muić 2010] M. Hanzer and G. Muić, "Parabolic induction and Jacquet functors for metaplectic groups", *J. Algebra* **323**:1 (2010), 241–260. MR 2564837 Zbl 1185.22013

[Hanzer and Tadić 2010] M. Hanzer and M. Tadić, "A method of proving non-unitarity of representations of p-adic groups", *Math. Z.* **265**:4 (2010), 799–816.

[Kudla 1986] S. S. Kudla, "On the local theta-correspondence", *Invent. Math.* **83**:2 (1986), 229–255. MR 87e:22037 Zbl 0583.22010

[Kudla 1996] S. S. Kudla, "Notes on the local theta correspondence", lectures at the European School in Group Theory, 1996, available at http://www.math.toronto.edu/~skudla/castle.pdf.

[Lapid et al. 2004] E. Lapid, G. Muić, and M. Tadić, "On the generic unitary dual of quasisplit classical groups", *Int. Math. Res. Not.* **2004**:26 (2004), 1335–1354. MR 2005b:22021 Zbl 1079.22015

[Matić 2010] I. Matić, "The unitary dual of p-adic SO(5)", *Proc. Amer. Math. Soc.* **138**:2 (2010), 759–767. MR 2557193 Zbl 05678540

[Matić $\geq$ 2010] I. Matić, "Composition series of the induced representations of $\mathrm{SO}(5, F)$ using intertwining operators", *Glasnik Mat. Ser. III* **45**, 93–107. Zbl 05722023

[Miličić 1973] D. Miličić, "On C^*-algebras with bounded trace", *Glasnik Mat. Ser. III* **8**:28 (1973), 7–22. MR 48 #2781 Zbl 0265.46072

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

[Muić 1997] G. Muić, "The unitary dual of p-adic G_2", *Duke Math. J.* **90**:3 (1997), 465–493. MR 98k:22073 Zbl 0896.22006

[Muić 2008] G. Muić, "A geometric construction of intertwining operators for reductive p-adic groups", *Manuscripta Math.* **125**:2 (2008), 241–272. MR 2009i:22024 Zbl 1145.22010

[Muić and Tadić 2007] G. Muić and M. Tadić, "Unramified unitary duals for split classical p-adic groups: The topology and isolated representations", preprint, 2007, available at http://web.math.hr/~gmuic/vita/LATEX-spherical-unit-clas.pdf. To appear in Clay Mathematics Proceedings (Shahidi volume).

[Ranga Rao 1993] R. Ranga Rao, "On some explicit formulas in the theory of Weil representation", *Pacific J. Math.* **157**:2 (1993), 335–371. MR 94a:22037 Zbl 0794.58017

[Sally and Tadić 1993] P. J. Sally, Jr. and M. Tadić, "Induced representations and classifications for GSp(2, F) and Sp(2, F)", pp. 75–133 Mém. Soc. Math. France (N.S.) **52**, Math. France (N.S.), Paris, 1993. MR 94e:22030 Zbl 0784.22008

[Shahidi 1981] F. Shahidi, "On certain L-functions", *American J. Math.* **103**:2 (1981), 297–355. MR 82i:10030 Zbl 0467.12013

[Szpruch 2007] D. Szpruch, "Uniqueness of Whittaker model for the metaplectic group", *Pacific J. Math.* **232**:2 (2007), 453–469. MR 2009b:11094 Zbl 05366274

[Tadić 1986] M. Tadić, "Classification of unitary representations in irreducible representations of general linear group (non-Archimedean case)", *Ann. Sci. École Norm. Sup.* (4) **19**:3 (1986), 335–382. MR 88b:22021

[Tadić 1993] M. Tadić, "An external approach to unitary representations", *Bull. Amer. Math. Soc. (N.S.)* **28**:2 (1993), 215–252. MR 93g:22020 Zbl 0799.22010

[Tadić 1994] M. Tadić, "Representations of p-adic symplectic groups", *Compositio Math.* **90**:2 (1994), 123–181. MR 95a:22025 Zbl 0797.22008

[Tadić 1995] M. Tadić, "Structure arising from induction and Jacquet modules of representations of classical p-adic groups", *J. Algebra* **177**:1 (1995), 1–33. MR 97b:22023 Zbl 0874.22014

[Tadić 1998a] M. Tadić, "On reducibility of parabolic induction", *Israel J. Math.* **107** (1998), 29–91. MR 2001d:22012 Zbl 0914.22019

[Tadić 1998b] M. Tadić, "On regular square integrable representations of p-adic groups", *American J. Math.* **120**:1 (1998), 159–210. MR 99h:22026 Zbl 0903.22008

[Zelevinsky 1980] A. V. Zelevinsky, "Induced representations of reductive $\mathfrak{p}$-adic groups, II: On irreducible representations of GL(n)", *Ann. Sci. École Norm. Sup.* (4) **13** (1980), 165–210. MR 83g:22012 Zbl 0441.22014

[Zorn 2007] C. Zorn, *Computing local L-factors for the unramified principal series of* $\mathrm{Sp}_2(F)$ *and its metaplectic cover*, thesis, University of Maryland, 2007, available at http://tinyurl.com/379axpc.

[Zorn 2010] C. Zorn, "Theta dichotomy for the genuine unramified principal series of $\widetilde{\mathrm{Sp}(2)}$", preprint, 2010, available at http://tinyurl.com/3yoj5lp.

Received July 16, 2009. Revised November 10, 2009.

MARCELA HANZER
DEPARTMENT OF MATHEMATICS,
UNIVERSITY OF ZAGREB
ZAGREB
CROATIA
hanmar@math.hr
http://web.math.hr/~hanmar

IVAN MATIĆ
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OSIJEK
TRG LJUDEVITA GAJA 6
OSIJEK
CROATIA
imatic@mathos.hr
http://www.mathos.hr/~imatic

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

A CASSON–LIN TYPE INVARIANT FOR LINKS

ERIC HARPER AND NIKOLAI SAVELIEV

In 1992, Xiao-Song Lin constructed an invariant $h(K)$ of knots $K \subset S^3$ via a signed count of conjugacy classes of irreducible SU(2) representations of $\pi_1(S^3 - K)$ with trace-free meridians. Lin showed that $h(K)$ equals one half times the knot signature of K. Using methods similar to Lin's, we construct an invariant $h(L)$ of two-component links $L \subset S^3$. Our invariant is a signed count of conjugacy classes of projective SU(2) representations of $\pi_1(S^3 - L)$ with a fixed 2-cocycle and corresponding nontrivial w_2. We show that $h(L)$ is, up to a sign, the linking number of L.

1. Introduction

One of the characteristic features of the fundamental group of a closed 3-manifold is that its representation variety in a compact Lie group tends to be finite, in a properly understood sense. This has been a guiding principle for defining invariants of 3-manifolds ever since Casson defined his λ-invariant for integral homology 3-spheres via a signed count of the SU(2) representations of the fundamental group, where signs were determined using Heegaard splittings.

Among numerous generalizations of Casson's construction, we will single out the invariant of knots in S^3 defined by Xiao-Song Lin [1992] via a signed count of SU(2) representations of the fundamental group of the knot exterior. The latter is a 3-manifold with nonempty boundary so the finiteness principle above only applies after one imposes a proper boundary condition. Lin's choice of boundary condition, namely, that all of the knot meridians are represented by trace-free SU(2) matrices, resulted in an invariant $h(K)$ of knots $K \subset S^3$. Lin further showed that $h(K)$ equals half the knot signature of K.

The signs in Lin's construction were determined using braid representations for knots. Austin (unpublished) and Heusener and Kroll [1998] extended Lin's construction by letting the meridians of the knot be represented by SU(2) matrices with

MSC2000: primary 57M25; secondary 57M05.
Keywords: link group, braid, projective representation.
Saveliev was partially supported by NSF grant 0305946 and the Max-Planck-Institut für Mathematik in Bonn, Germany.

a fixed trace that need not be zero. Their construction gives a knot invariant that equals, for most choices of the trace, one half times the equivariant knot signature.

In this paper, we extend Lin's construction to two-component links L in S^3. In essence, we replace the count of SU(2) representations with a count of *projective* SU(2) representations of $\pi_1(S^3 - L)$, in the sense of [Ruberman and Saveliev 2004], with a fixed 2-cocycle representing a nontrivial element in the second group cohomology of $\pi_1(S^3 - L)$. The resulting signed count is denoted by $h(L)$. The two main results of this paper are then as follows.

Theorem 1. *For any two-component link $L \subset S^3$, the integer $h(L)$ is a well-defined invariant of L.*

Theorem 2. *For any two-component link $L = \ell_1 \cup \ell_2$ in S^3, one has*

$$h(L) = \pm \operatorname{lk}(\ell_1, \ell_2).$$

Our choice of the 2-cocycle imposes Lin's trace-free condition on us. This is in contrast to Lin's construction, where the choice of boundary condition seemed somewhat arbitrary. This also means one should not expect to extend our construction to SU(2) representations with nonzero trace boundary condition.

Shortly after Casson introduced his invariant for homology 3-spheres, Taubes [1990] gave a gauge theoretic description of it in terms of a signed count of flat SU(2) connections. After Lin's work, but before Heusener and Kroll, a gauge theoretic interpretation of the Lin invariant was given by Herald [1997]. He used this interpretation to define an extension of the Lin invariant, now known as the Herald–Lin invariant, to knots in arbitrary homology spheres, with arbitrary fixed-trace (possibly nonzero) boundary condition.

Another attractive feature of the gauge theoretic approach is that it can be used to produce ramified versions of the above invariants. Floer [1988] introduced the instanton homology theory whose Euler characteristic is twice the Casson invariant. We expect that our invariant will have a similar interpretation, perhaps along the lines of the knot instanton homology theory of Kronheimer and Mrowka [2008], which in turn is a variant of the orbifold Floer homology of Collin and Steer [1999]. We hope to discuss this elsewhere, together with possible extensions to links in homology spheres and to links of more than two components.

2. Braids and representations

Let F_n be a free group of rank $n \geq 2$, with a fixed generating set $x_1, \dots, x_n$. We will follow the conventions of [Long 1989] and define the n-string braid group $\mathcal{B}_n$ to be the subgroup of $\operatorname{Aut}(F_n)$ generated by the automorphisms $\sigma_1, \dots, \sigma_{n-1}$, where the action of σ_i is given by

$$\begin{aligned}\sigma_i : x_i &\mapsto x_{i+1},\\ x_{i+1} &\mapsto (x_{i+1})^{-1} x_i x_{i+1},\\ x_j &\mapsto x_j \quad \text{for } j \neq i, i+1.\end{aligned}$$

The natural homomorphism $\mathcal{B}_n \to S_n$, $\sigma \mapsto \bar{\sigma}$, onto the symmetric group on n letters maps each generator σ_i to the transposition $\bar{\sigma}_i = (i, i+1)$. A useful observation is that, for any $\sigma \in \mathrm{Aut}(F_n)$, one has

$$\sigma(x_i) = w x_{\bar{\sigma}^{-1}(i)} w^{-1} \tag{1}$$

for some word $w \in F_n$. Also σ preserves the product $x_1 \cdots x_n$, that is,

$$\sigma(x_1 \cdots x_n) = x_1 \cdots x_n. \tag{2}$$

2a. SU(2) *representations.* Consider the Lie group SU(2) of unitary two-by-two matrices with determinant one, that is, complex matrices

$$\begin{pmatrix} u & v \\ -\bar{v} & \bar{u} \end{pmatrix}$$

such that $u\bar{u} + v\bar{v} = 1$. We will often identify SU(2) with the group Sp(1) of unit quaternions via

$$\begin{pmatrix} u & v \\ -\bar{v} & \bar{u} \end{pmatrix} \mapsto u + vj \in \mathbb{H}.$$

Let $R_n = \mathrm{Hom}(F_n, \mathrm{SU}(2))$ be the space of SU(2) representations of F_n, and identify it with $\mathrm{SU}(2)^n$ by sending a representation $\alpha : F_n \to \mathrm{SU}(2)$ to the vector $(\alpha(x_1), \ldots, \alpha(x_n))$ of SU(2) matrices. The above representation $\mathcal{B}_n \to \mathrm{Aut}(F_n)$ then gives rise to the representation

$$\rho : \mathcal{B}_n \to \mathrm{Diff}(R_n) \tag{3}$$

via $\rho(\sigma)(\alpha) = \alpha \circ \sigma^{-1}$. We will abbreviate $\rho(\sigma)$ to σ. We will also denote $X = (X_1, \ldots, X_n) \in R_n$ and write $\sigma(X) = (\sigma(X)_1, \ldots, \sigma(X)_n)$.

Example. For any $(X_1, \ldots, X_n) \in R_n$, we have

$$\sigma_1(X_1, X_2, X_3, \ldots, X_n) = (X_1 X_2 X_1^{-1}, X_1, X_3, \ldots, X_n).$$

2b. *Extension to the wreath product* $\mathbb{Z}_2 \wr \mathcal{B}_n$. The wreath product $\mathbb{Z}_2 \wr \mathcal{B}_n$ is the semidirect product of $\mathcal{B}_n$ with $(\mathbb{Z}_2)^n$, where $\mathcal{B}_n$ acts on $(\mathbb{Z}_2)^n$ by permuting the coordinates, $\sigma(\varepsilon_1, \ldots, \varepsilon_n) = (\varepsilon_{\bar{\sigma}(1)}, \ldots, \varepsilon_{\bar{\sigma}(n)})$. Thus the elements of $\mathbb{Z}_2 \wr \mathcal{B}_n$ are the pairs $(\varepsilon, \sigma) \in (\mathbb{Z}_2)^n \times \mathcal{B}_n$, with the group multiplication law

$$(\varepsilon, \sigma) \cdot (\varepsilon', \sigma') = (\varepsilon\sigma(\varepsilon'), \sigma\sigma').$$

The representation (3) can be extended to a representation

(4) $$\rho : \mathbb{Z}_2 \wr \mathcal{B}_n \to \mathrm{Diff}(R_n)$$

by defining

$$\rho(\varepsilon, \sigma)(X) = \varepsilon \cdot \sigma(X) = (\varepsilon_1 \sigma(X)_1, \dots, \varepsilon_n \sigma(X)_n),$$

where the ε_i are viewed as elements of the center $\mathbb{Z}_2 = \{\pm 1\}$ of SU(2). That (4) is a representation follows by a direct calculation after one observes that, by (1),

(5) $$\sigma(X)_i = A X_{\bar{\sigma}(i)} A^{-1} \quad \text{for some } A \in \mathrm{SU}(2).$$

Again, we will abuse notation and write simply $\varepsilon\sigma$ for both (ε, σ) and $\rho(\varepsilon, \sigma)$.

Example. For any $(X_1, \dots, X_n) \in R_n$ and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \in (\mathbb{Z}_2)^n$, we have

$$\begin{aligned}(\varepsilon\sigma_1)(X_1, X_2, X_3, \dots, X_n) &= (\varepsilon_1 X_1 X_2 X_1^{-1}, \varepsilon_2 X_1, \varepsilon_3 X_3, \dots, \varepsilon_n X_n),\\ \sigma_1(\varepsilon X) = \sigma_1(\varepsilon)\sigma_1(X) &= (\varepsilon_2 X_1 X_2 X_1^{-1}, \varepsilon_1 X_1, \varepsilon_3 X_3, \dots, \varepsilon_n X_n).\end{aligned}$$

2c. *Braids and link groups.* The closure $\hat{\sigma}$ of a braid $\sigma \in \mathcal{B}_n$ is a link in S^3 with link group

$$\pi_1(S^3 - \hat{\sigma}) = \langle x_1, \dots, x_n \mid x_i = \sigma(x_i) \text{ for } i = 1, \dots, n\rangle,$$

where each x_i represents a meridian of $\hat{\sigma}$. One can easily see that the fixed points of the diffeomorphism $\sigma : R_n \to R_n$ are representations $\pi_1(S^3 - \hat{\sigma}) \to \mathrm{SU}(2)$. This paper grew out of the observation that a fixed point $\alpha = (\alpha(x_1), \dots, \alpha(x_n))$ of the map $\varepsilon\sigma : R_n \to R_n$ gives rise to a representation $\mathrm{ad}\,\alpha : \pi_1(S^3 - \hat{\sigma}) \to \mathrm{SO}(3)$ by composing with the adjoint representation $\mathrm{ad} : \mathrm{SU}(2) \to \mathrm{SO}(3)$. Depending on ε, the representation $\mathrm{ad}\,\alpha$ may or may not lift to an SU(2) representation, the obstruction being the second Stiefel–Whitney class $w_2(\mathrm{ad}\,\alpha) \in H^2(\pi_1(S^3 - \hat{\sigma}); \mathbb{Z}_2)$.

3. Definition of $h(\varepsilon\sigma)$

Every link in S^3 is the closure $\hat{\sigma}$ of a braid σ; see [Alexander 1923]. Let σ be a braid whose closure $\hat{\sigma}$ has two components. We will associate with it, for a carefully chosen ε, an integer $h(\varepsilon\sigma)$. We will prove in Section 4 that h is an invariant of the link $\hat{\sigma}$.

3a. *Choice of ε.* The number of components of the link $\hat{\sigma}$ is exactly the number of cycles in the permutation $\bar{\sigma}$. We will be interested in two component links, that is, the closures of braids σ with

(6) $$\bar{\sigma} = (i_1 \dots i_m)(i_{m+1} \dots i_n) \quad \text{for some } 1 \le m \le n-1.$$

Given such a braid σ, choose a vector $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \in (\mathbb{Z}_2)^n$ such that

$$\varepsilon_{i_1} \cdots \varepsilon_{i_m} = \varepsilon_{i_{m+1}} \cdots \varepsilon_{i_n} = -1. \tag{7}$$

This choice of ε is dictated by the following two considerations. First, we wish to preserve condition (2) in the form

$$(\varepsilon\sigma)(X)_1 \cdots (\varepsilon\sigma)(X)_n = X_1 \cdots X_n, \tag{8}$$

and second, we want the fixed points α of the diffeomorphism $\varepsilon\sigma : R_n \to R_n$ to have nonzero $w_2(\operatorname{ad}\alpha)$.

Lemma 3.1. *If α is a fixed point of $\varepsilon\sigma : R_n \to R_n$ with ε as in* (7), *then $w_2(\operatorname{ad}\alpha) \neq 0$.*

Proof. The class $w_2(\operatorname{ad}\alpha)$ is the obstruction to lifting $\operatorname{ad}\alpha$ to an SU(2) representation. Extend α arbitrarily to a function $\alpha : \pi_1(S^3 - \hat\sigma) \to \mathrm{SU}(2)$ lifting $\operatorname{ad}\alpha$. Then $w_2(\operatorname{ad}\alpha)$ will vanish if and only if there is a function $\eta : \pi_1(S^3 - \hat\sigma) \to \mathbb{Z}_2 = \{\pm 1\}$ such that $\eta \cdot \alpha$ is a representation. Suppose such a function exists, and write $\eta(x_i) = \eta_i = \pm 1$. Also, assume without loss of generality that $\bar\sigma = (1 \dots m)(m+1 \dots n)$. It follows from (5) that to satisfy the relations $X_i = (\varepsilon\sigma)(X)_i$ we must have $\eta_1 = \varepsilon_1\eta_2 = \varepsilon_1\varepsilon_2\eta_3 = \cdots = \varepsilon_1 \cdots \varepsilon_m \eta_1 = -\eta_1$, a contradiction with $\eta_1 = \pm 1$. □

This result for $w_2(\operatorname{ad}\alpha)$ can be sharpened using an algebraic topology lemma.

Lemma 3.2. *Let $\hat\sigma$ be a link of two components. Then $H^2(\pi_1(S^3 - \hat\sigma); \mathbb{Z}_2)$ is equal to $\mathbb{Z}_2$ if $\hat\sigma$ is nonsplit and is zero otherwise.*

Proof. If $\hat\sigma$ is nonsplit, then $S^3 - \hat\sigma$ is a $K(\pi, 1)$ by the sphere theorem; hence $H^2(\pi_1(S^3 - \hat\sigma); \mathbb{Z}_2) = H^2(S^3 - \hat\sigma; \mathbb{Z}_2) = \mathbb{Z}_2$. If $\hat\sigma$ is split, then $K(\pi_1(S^3 - \hat\sigma), 1)$ has the homotopy type of a one-point union of two circles and the result again follows. □

Corollary 3.3. *Let $\hat\sigma$ be a split link of two components*, *and let ε be chosen as in* (7). *Then the diffeomorphism $\varepsilon\sigma : R_n \to R_n$ has no fixed points.*

3b. ***The zero-trace condition.*** A naive way to define $h(\varepsilon\sigma)$ would be as the intersection number of the graph of $\varepsilon\sigma : R_n \to R_n$ with the diagonal in the product $R_n \times R_n$. One can observe though that, in addition to this intersection not being transversal, its points $(X, X) = (X_1, \dots, X_n, X_1, \dots, X_n)$ have the property that $\operatorname{tr} X_1 = \cdots = \operatorname{tr} X_n = 0$. This can be seen as follows.

Assume without loss of generality that $\bar\sigma = (1 \dots m)(m+1 \dots n)$. Then the relations $X = \varepsilon\sigma(X)$ together with (5) imply that

$$\begin{aligned} X_1 &= \varepsilon_1\sigma(X)_1 = \varepsilon_1 A_1 \cdot X_{\bar\sigma(1)} \cdot A_1^{-1} = \varepsilon_1 A_1 X_2 A_1^{-1} \\ &= \varepsilon_1 A_1 \cdot \varepsilon_2 \sigma(X)_2 \cdot A_1^{-1} = \varepsilon_1\varepsilon_2 A_1 A_2 \cdot X_{\bar\sigma(2)} \cdot A_2^{-1} A_1^{-1} = \cdots \\ &= \varepsilon_1 \cdots \varepsilon_m (A_1 \cdots A_m) \cdot X_1 \cdot (A_1 \dots A_m)^{-1}. \end{aligned}$$

Since trace is conjugation invariant and $\varepsilon_1 \cdots \varepsilon_m = -1$, we conclude that $\operatorname{tr} X_1 = \cdots = \operatorname{tr} X_m = 0$. Similarly, $\operatorname{tr} X_{m+1} = \cdots = \operatorname{tr} X_n = 0$.

Hence, in our definition we will restrict ourselves to the subset of R_n consisting of $X = (X_1, \ldots, X_n)$ with $\operatorname{tr} X_1 = \cdots = \operatorname{tr} X_n = 0$. The nontransversality problem will be addressed below by factoring out the conjugation symmetry and lowering the dimension of the ambient manifold.

3c. *The definition.* The subset of SU(2) consisting of the matrices with zero trace is a conjugacy class in SU(2) diffeomorphic to S^2. Define

$$Q_n = \{(X_1, \ldots, X_n) \in R_n \mid \operatorname{tr} X_i = 0\} \subset R_n,$$

so that Q_n is a manifold diffeomorphic to $(S^2)^n$. Also define

$$H_n = \{(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \in Q_n \times Q_n \mid X_1 \cdots X_n = Y_1 \cdots Y_n\}.$$

This is no longer a manifold due to the presence of *reducibles*. We call a point $(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \in Q_n \times Q_n$ reducible if all X_i and Y_j commute with each other or, equivalently, if there is a matrix $A \in \mathrm{SU}(2)$ such that the AX_iA^{-1} and AY_iA^{-1} are diagonal matrices for $i = 1, \ldots, n$. The subset $S_n \subset Q_n \times Q_n$ of reducibles is closed.

Lemma 3.4. *$H_n^* = H_n - S_n$ is an open manifold of dimension* $4n - 3$.

Proof. Consider the open manifold $(Q_n \times Q_n)^* = Q_n \times Q_n - S_n$ of dimension $4n$ and the map $f : (Q_n \times Q_n)^* \to \mathrm{SU}(2)$ given by

$$f(X_1, \ldots, X_n, Y_1, \ldots, Y_n) = X_1 \cdots X_n Y_n^{-1} \cdots Y_1^{-1}. \tag{9}$$

According to [Lin 1992, Lemma 1.5], this map has $1 \in \mathrm{SU}(2)$ as a regular value. Since $H_n^* = f^{-1}(1)$, the result follows. □

Because of (5) and the fact that multiplication by $-1 \in \mathrm{SU}(2)$ preserves the zero trace condition, the representation (4) gives rise to a representation

$$\rho : \mathbb{Z}_2 \wr \mathcal{B}_n \to \operatorname{Diff}(Q_n). \tag{10}$$

Given $\varepsilon\sigma \in \mathbb{Z}_2 \wr \mathcal{B}_n$ such that (6) and (7) are satisfied, consider two submanifolds of $Q_n \times Q_n$: the graph $\Gamma_{\varepsilon\sigma} = \{(X, \varepsilon\sigma(X)) \mid X \in Q_n\}$ of $\varepsilon\sigma : Q_n \to Q_n$, and the diagonal $\Delta_n = \{(X, X) \mid X \in Q_n\}$. Note that both $\Gamma_{\varepsilon\sigma}$ and Δ_n are subsets of H_n; this is obvious for Δ_n and follows from (8) for $\Gamma_{\varepsilon\sigma}$.

Proposition 3.5. *The intersection $\Gamma_{\varepsilon\sigma} \cap \Delta_n \subset H_n$ consists of irreducible representations.*

Proof. Assume without loss of generality that $\bar\sigma = (1 \ldots m)(m + 1 \ldots n)$, and suppose that $(X, X) = (X_1, \ldots, X_n, X_1, \ldots, X_n) \in \Gamma_{\varepsilon\sigma} \cap \Delta_n$ is reducible. Then all of the X_i commute with each other, and in particular $\sigma(X) = (X_{\bar\sigma(1)}, \ldots, X_{\bar\sigma(n)})$.

The equality $X = \varepsilon\sigma(X)$ then implies that $X_1 = \varepsilon_1 X_{\bar\sigma(1)} = \varepsilon_1 X_2 = \varepsilon_1\varepsilon_2 X_{\bar\sigma(2)} = \cdots = \varepsilon_1 \cdots \varepsilon_m X_1 = -X_1$, which contradicts that $X_1 \in \mathrm{SU}(2)$. □

Let $\Gamma^*_{\varepsilon\sigma} = \Gamma_{\varepsilon\sigma} \cap H^*_n$ and $\Delta^*_n = \Delta_n \cap H^*_n$ be the irreducible parts of $\Gamma_{\varepsilon\sigma}$ and Δ_n, respectively. They are both open submanifolds of H^*_n of dimension $2n$.

Corollary 3.6. *The intersection* $\Delta^*_n \cap \Gamma^*_{\varepsilon\sigma} \subset H^*_n$ *is compact.*

Proof. Proposition 3.5 implies that $\Delta^*_n \cap \Gamma^*_{\varepsilon\sigma} = \Delta_n \cap \Gamma_{\varepsilon\sigma}$, and the latter intersection is obviously compact since it is the intersection of two compact subsets of H_n. □

The group $\mathrm{SO}(3) = \mathrm{SU}(2)/\{\pm 1\}$ acts freely by conjugation on H^*_n, Δ^*_n, and $\Gamma^*_{\varepsilon\sigma}$. Denote the resulting quotient manifolds by

$$\hat H_n = H^*_n / \mathrm{SO}(3), \quad \hat\Delta_n = \Delta^*_n / \mathrm{SO}(3), \quad \text{and} \quad \hat\Gamma_{\varepsilon\sigma} = \Gamma^*_{\varepsilon\sigma} / \mathrm{SO}(3).$$

The dimension of $\hat H_n$ is $4n-6$, and $\hat\Delta_n$ and $\hat\Gamma_{\varepsilon\sigma}$ are $(2n-3)$-dimensional submanifolds. Since the intersection $\hat\Delta_n \cap \hat\Gamma_{\varepsilon\sigma}$ is compact, one can isotope $\hat\Gamma_{\varepsilon\sigma}$ into a submanifold $\widetilde\Gamma_{\varepsilon\sigma}$ using an isotopy with compact support so that $\hat\Delta_n \cap \widetilde\Gamma_{\varepsilon\sigma}$ consists of finitely many points. Define

$$h(\varepsilon\sigma) = \#_{\hat H_n}(\hat\Delta_n \cap \widetilde\Gamma_{\varepsilon\sigma})$$

as the algebraic intersection number, where the orientations of $\hat H_n$, $\hat\Delta_n$, and $\widetilde\Gamma_{\varepsilon\sigma}$ are described in the next subsection. It is obvious that $h(\varepsilon\sigma)$ does not depend on the perturbation of $\hat\Gamma_{\varepsilon\sigma}$, so we will simply write

$$h(\varepsilon\sigma) = \langle \hat\Delta_n, \hat\Gamma_{\varepsilon\sigma} \rangle_{\hat H_n}.$$

3d. *Orientations.* Choose an arbitrary orientation of the copy of $S^2 \subset \mathrm{SU}(2)$ cut out by the trace zero condition, and endow $Q_n = (S^2)^n$ and $Q_n \times Q_n$ with product orientations. The diagonal Δ_n and the graph $\Gamma_{\varepsilon\sigma}$ are naturally diffeomorphic to Q_n via projection onto the first factor, and they are given the induced orientations. If we reverse the orientation of S^2, then the orientation of Q_n is reversed if n is odd. Hence the orientations of both Δ_n and $\Gamma_{\varepsilon\sigma}$ are reversed if n is odd, while the orientation of $Q_n \times Q_n = (S^2)^{2n}$ is preserved regardless of the parity of n.

Orient $\mathrm{SU}(2)$ by the standard basis $\{i, j, k\}$ in its Lie algebra $\mathfrak{su}(2)$, and orient $H^*_n = f^{-1}(1)$ by applying the base-fiber rule to the map (9). The adjoint action of $\mathrm{SO}(3)$ on $S^2 \subset \mathrm{SU}(2)$ is orientation preserving; hence the $\mathrm{SO}(3)$ quotients $\hat H_n$, $\hat\Delta_n$, and $\hat\Gamma_{\varepsilon\sigma}$ are orientable. We orient them using the base-fiber rule. The discussion in the previous paragraph shows that reversing orientation on S^2 may reverse the orientations of $\hat\Delta_n$ and $\hat\Gamma_{\varepsilon\sigma}$ but that it does not affect the intersection number $\langle \hat\Delta_n, \hat\Gamma_{\varepsilon\sigma} \rangle_{\hat H_n}$.

4. The link invariant *h*

In this section, we will prove Theorem 1. This will be accomplished by proving that $h(\varepsilon\sigma)$ is independent first of ε and then of σ.

4a. *Independence of ε.* We will first show that, for a fixed σ whose closure $\hat{\sigma}$ is a link of two components, $h(\varepsilon\sigma)$ is independent of the choice of ε as long as ε satisfies (7).

Proposition 4.1. *Let ε and ε' be such that* (7) *is satisfied. Then* $h(\varepsilon\sigma) = h(\varepsilon'\sigma)$.

Proof. Assume without loss of generality that $\bar{\sigma} = (1\dots m)(m+1\dots n)$ and let $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ and $\varepsilon' = (\varepsilon'_1, \dots, \varepsilon'_n)$. Define $\delta = (\delta_1, \dots, \delta_n)$ as the vector in $(\mathbb{Z}_2)^n$ with coordinates

$$\delta_1 = 1 \quad \text{and} \quad \delta_{k+1} = \delta_k \varepsilon_k \varepsilon'_k \quad \text{for } k = 1, \dots, n-1,$$

and define the involution $\tau : Q_n \to Q_n$ by the formula

$$\tau(X) = \delta X = (\delta_1 X_1, \delta_2 X_2, \dots, \delta_n X_n).$$

Recall that $Q_n = (S^2)^n$ so that τ is a diffeomorphism that restricts to each of the factors S^2 as either the identity or the antipodal map. In particular, τ need not be orientation preserving.

The map $\tau \times \tau : Q_n \times Q_n \to Q_n \times Q_n$ obviously preserves the irreducibility condition and commutes with the SO(3) action. It gives rise to an orientation preserving automorphism of $\hat{H}_n$, which will again be called $\tau \times \tau$. It is clear that $(\tau \times \tau)(\hat{\Delta}_n) = \hat{\Delta}_n$. It is also true that $(\tau \times \tau)(\hat{\Gamma}_{\varepsilon\sigma}) = \hat{\Gamma}_{\varepsilon'\sigma}$, which can be seen as follows. Write a pair $(\delta X, \delta\varepsilon\sigma(X))$ whose conjugacy class belongs to $(\tau \times \tau)(\hat{\Gamma}_{\varepsilon\sigma})$ as

$$(\delta X, \delta\varepsilon\sigma(X)) = (\delta X, \delta\varepsilon\sigma(\delta\delta X)) = (\delta X, \delta\varepsilon\sigma(\delta)\sigma(\delta X))$$

using the multiplication law in the group $\mathbb{Z}_2 \wr \mathcal{B}_n$. The conjugacy class of this pair belongs to $\Gamma_{\varepsilon'\sigma}$ if and only if $\delta\varepsilon\sigma(\delta) = \varepsilon'$. That this condition holds can be verified directly from the definition of δ.

Recall that the orientations of $\hat{\Delta}_n$, $\hat{\Gamma}_{\varepsilon\sigma}$, and $\hat{\Gamma}_{\varepsilon'\sigma}$ are induced by the orientation of Q_n. Therefore, the maps $\tau \times \tau : \hat{\Delta}_n \to \hat{\Delta}_n$ and $\tau \times \tau : \hat{\Gamma}_{\varepsilon\sigma} \to \hat{\Gamma}_{\varepsilon'\sigma}$ are either both orientation preserving or both orientation reversing depending on whether $\tau : Q_n \to Q_n$ preserves or reverses orientation. Hence we have

$$\begin{aligned} h(\varepsilon\sigma) = \langle \hat{\Delta}_n, \hat{\Gamma}_{\varepsilon\sigma} \rangle_{\hat{H}_n} &= \langle (\tau \times \tau)(\hat{\Delta}_n), (\tau \times \tau)(\hat{\Gamma}_{\varepsilon\sigma}) \rangle_{(\tau\times\tau)(\hat{H}_n)} \\ &= \langle \hat{\Delta}_n, \hat{\Gamma}_{\varepsilon'\sigma} \rangle_{\hat{H}_n} = h(\varepsilon'\sigma). \end{aligned}$$ □

From now on, we will drop ε from the notation and simply write $h(\sigma)$ for $h(\varepsilon\sigma)$ assuming that a choice of ε satisfying (7) has been made.

4b. ***Independence of σ.*** In this section, we will show that $h(\sigma)$ only depends on the link $\hat{\sigma}$, not on a particular choice of braid σ, by verifying that h is preserved under Markov moves. We will follow the proof of [Lin 1992, Theorem 1.8], which goes through with little change once the right ε are chosen.

Recall that two braids $\alpha \in \mathcal{B}_n$ and $\beta \in \mathcal{B}_m$ have isotopic closures $\hat{\alpha}$ and $\hat{\beta}$ if and only if one braid can be obtained from the other by a finite sequence of Markov moves; see for instance [Birman 1974]. A type 1 Markov move replaces $\sigma \in \mathcal{B}_n$ by $\xi^{-1}\sigma\xi \in \mathcal{B}_n$ for any $\xi \in \mathcal{B}_n$. A type 2 Markov move means replacing $\sigma \in \mathcal{B}_n$ by $\sigma_n^{\pm 1}\sigma \in \mathcal{B}_{n+1}$, or the inverse of this operation.

Proposition 4.2. *The invariant $h(\sigma)$ is preserved by type* 1 *Markov moves.*

Proof. Let $\xi, \sigma \in \mathcal{B}_n$ and assume as usual that $\bar{\sigma} = (1 \dots m)(m+1 \dots n)$. Then

$$\overline{\xi^{-1}\sigma\xi} = (\bar{\xi}(1) \dots \bar{\xi}(m))(\bar{\xi}(m+1) \dots \bar{\xi}(n))$$

has the same cycle structure as $\bar{\sigma}$. To compute $h(\xi^{-1}\sigma\xi)$, we will make a choice of $\varepsilon \in (\mathbb{Z}_2)^n$ that satisfies condition (7) with respect to the braid $\xi^{-1}\sigma\xi$, that is, $\varepsilon_{\bar{\xi}(1)} \cdots \varepsilon_{\bar{\xi}(m)} = \varepsilon_{\bar{\xi}(m+1)} \cdots \varepsilon_{\bar{\xi}(n)} = -1$.

The braid ξ gives rise to the map $\xi : Q_n \to Q_n$. It acts by permutation and conjugation on the S^2 factors in Q_n; hence it is orientation preserving (we use the fact that S^2 is even-dimensional). It induces an orientation-preserving map $\xi \times \xi : Q_n \times Q_n \to Q_n \times Q_n$, which preserves the irreducibility condition and commutes with the SO(3) action. Equation (2) then ensures that we have a well-defined orientation-preserving automorphism $\xi \times \xi : \hat{H}_n \to \hat{H}_n$.

That this automorphism preserves the diagonal, $(\xi \times \xi)(\hat{\Delta}_n) = \hat{\Delta}_n$, is obvious. Concerning the graphs, let $(X, \varepsilon\xi^{-1}\sigma\xi(X)) \in \hat{\Gamma}_{\varepsilon\xi^{-1}\sigma\xi}$; then

$$(\xi \times \xi)(X, \varepsilon\xi^{-1}\sigma\xi(X)) = (\xi(X), \xi(\varepsilon\xi^{-1}\sigma\xi(X))) = (\xi(X), \xi(\varepsilon)\sigma(\xi(X))) \in \hat{\Gamma}_{\xi(\varepsilon)\sigma}.$$

Therefore, $(\xi \times \xi)(\hat{\Gamma}_{\varepsilon\xi^{-1}\sigma\xi}) = \hat{\Gamma}_{\xi(\varepsilon)\sigma}$. Since $\xi : Q_n \to Q_n$ is orientation preserving, the identifications above of the diagonals and graphs via $\xi \times \xi$ are also orientation preserving.

Observe that $\xi(\varepsilon)_i = \varepsilon_{\bar{\xi}(i)}$. Hence $\xi(\varepsilon)$ satisfies (7) with respect to σ and thus can be used to compute $h(\sigma)$. The argument is completed by the calculation

$$\begin{aligned} h(\xi^{-1}\sigma\xi) = \langle \hat{\Delta}_n, \hat{\Gamma}_{\varepsilon\xi^{-1}\sigma\xi} \rangle_{\hat{H}_n} &= \langle (\xi \times \xi)(\hat{\Delta}_n), (\xi \times \xi)(\hat{\Gamma}_{\varepsilon\xi^{-1}\sigma\xi}) \rangle_{(\xi \times \xi)(\hat{H}_n)} \\ &= \langle \hat{\Delta}_n, \hat{\Gamma}_{\xi(\varepsilon)\sigma} \rangle_{\hat{H}_n} = h(\sigma). \end{aligned}$$

□

Proposition 4.3. *The invariant $h(\sigma)$ is preserved by type* 2 *Markov moves.*

Proof. Given $\sigma \in \mathcal{B}_n$ and ε satisfying (7), change σ to $\sigma_n\sigma \in \mathcal{B}_{n+1}$ and let $\varepsilon' = \sigma_n(\varepsilon, 1)$. If $X = (X_1, \dots, X_n)$ and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$, then

$$\begin{aligned}(\sigma_n\sigma)(X, X_{n+1}) &= \sigma_n(\sigma(X), X_{n+1}) \\ &= (\sigma(X)_1, \dots, \sigma(X)_{n-1}, \sigma(X)_n X_{n+1} \sigma(X)_n^{-1}, \sigma(X)_n)\end{aligned}$$

and $\varepsilon' = (\varepsilon_1, \dots, \varepsilon_{n-1}, 1, \varepsilon_n)$. In particular, ε' satisfies (7) with respect to $\sigma_n\sigma$. Consider the embedding $g : Q_n \times Q_n \to Q_{n+1} \times Q_{n+1}$ given by

$$g(X_1, \dots, X_n, Y_1, \dots, Y_n) = (X_1, \dots, X_n, Y_n, Y_1, \dots, Y_n, Y_n).$$

One can easily see that $g(H_n) \subset H_{n+1}$ and that g commutes with the conjugation, thus giving rise to an embedding $\hat{g} : \hat{H}_n \to \hat{H}_{n+1}$. A straightforward calculation using the formulas above for $\sigma_n\sigma$ and ε' then shows that

$$\hat{g}(\hat{\Delta}_n) \subset \hat{\Delta}_{n+1}, \quad \hat{g}(\hat{\Gamma}_{\varepsilon\sigma}) \subset \hat{\Gamma}_{\varepsilon'\sigma_n\sigma}, \quad \text{and} \quad \hat{g}(\hat{\Delta}_n \cap \hat{\Gamma}_{\varepsilon\sigma}) = \hat{\Delta}_{n+1} \cap \hat{\Gamma}_{\varepsilon'\sigma_n\sigma}.$$

Now one can achieve all the necessary transversalities and match the orientations in exactly the same way as in the second half of the proof of [Lin 1992, Theorem 1.8]. This shows that $h(\sigma_n\sigma) = h(\sigma)$. The proof that $h(\sigma_n^{-1}\sigma) = h(\sigma)$ is similar. □

5. The invariant $h(\sigma)$ as the linking number

In this section we will prove Theorem 2, that is, show that for any link $\hat{\sigma} = \ell_1 \cup \ell_2$ of two components, one has $h(\sigma) = \pm\,\mathrm{lk}(\ell_1, \ell_2)$. Our strategy will be to show that the invariant $h(\sigma)$ and the linking number $\mathrm{lk}(\ell_1, \ell_2)$ change according to the same rule as we change a crossing between two strands from two different components of $\hat{\sigma} = \ell_1 \cup \ell_2$ (the link $\hat{\sigma}$ will need to be oriented for that, although a particular choice of orientation will not matter). After changing finitely many such crossings, we will arrive at a split link, for which both the invariant $h(\sigma)$ and the linking number $\mathrm{lk}(\ell_1, \ell_2)$ vanish; see Corollary 3.3. The change of crossing as above obviously changes the linking number by ± 1. To calculate the effect of the crossing change on $h(\sigma)$, we will follow [Lin 1992] and reduce the problem to a calculation in the pillowcase $\hat{H}_2$.

5a. ***The pillowcase.*** We begin with a geometric description of $\hat{H}_2$ as a pillowcase; compare with [Lin 1992, Lemma 1.2]. Remember that

$$H_2 = \{(X_1, X_2, Y_1, Y_2) \in Q_2 \times Q_2 \mid X_1 X_2 = Y_1 Y_2\}.$$

We will use the identification of SU(2) with Sp(1) when convenient. Since X_2 is trace free, we may assume that $X_2 = i$ after conjugation. Conjugating by $e^{i\varphi}$ will not change X_2 but, for an appropriate choice of φ, will make X_1 into

$$X_1 = \begin{pmatrix} ir & u \\ -u & -ir \end{pmatrix},$$

where both r and u are real, and u is also nonnegative. Since $r^2 + u^2 = 1$, we can write $r = \cos\theta$ and $u = \sin\theta$ for a unique θ such that $0 \le \theta \le \pi$. In quaternionic language, $X_1 = ie^{-k\theta}$ with $0 \le \theta \le \pi$. Similarly, the condition $\operatorname{tr}(Y_2) = \operatorname{tr}(Y_1^{-1}X_1X_2) = 0$ implies that $Y_1 = ie^{-k\psi}$, this time with $-\pi \le \psi \le \pi$. To summarize,

$$X_1 = ie^{-k\theta}, \quad X_2 = i, \quad Y_1 = ie^{-k\psi}, \quad Y_2 = ie^{-k(\psi-\theta)}.$$

Thus $\hat{H}_2$ is parametrized by the rectangle $[0,\pi] \times [-\pi,\pi]$, with proper identifications along the edges and with the reducibles removed. The reducibles occur when both θ and ψ are multiples of π, and hence $\hat{H}_2$ is a 2-sphere with the points $A = (0,0)$, $B = (\pi,0)$, $A' = (0,\pi)$, and $B' = (\pi,\pi)$ removed; see Figure 1. According to [Lin 1992], the orientation on the front sheet of $\hat{H}_2$ coincides with the standard orientation on the (θ, ψ) plane.

Example. Let $\sigma = \sigma_1^2$, so that $\hat{\sigma} = \ell_1 \cup \ell_2$ is the Hopf link with $\operatorname{lk}(\ell_1, \ell_2) = \pm 1$. To calculate $h(\sigma)$, we let $\varepsilon = (-1,-1)$, the only available choice satisfying (7), and consider the submanifolds $\hat{\Delta}_2$ and $\hat{\Gamma}_{\varepsilon\sigma}$ of $\hat{H}_2$. We have, in quaternionic notation, $\hat{\Delta}_2 = \{(ie^{-k\theta}, i, ie^{-k\theta}, i)\}$, which is the diagonal $\psi = \theta$ in the pillowcase. A straightforward calculation shows that $\hat{\Gamma}_{\varepsilon\sigma} \subset \hat{H}_2$ is given by $\psi = 3\theta - \pi$. As can be seen in Figure 1, the intersection $\hat{\Delta}_2 \cap \hat{\Gamma}_{\varepsilon\sigma}$ consists of one point coming with a sign. Hence $h(\sigma_1^2) = \pm 1$, which is consistent with the fact that $\operatorname{lk}(\ell_1, \ell_2) = \pm 1$.

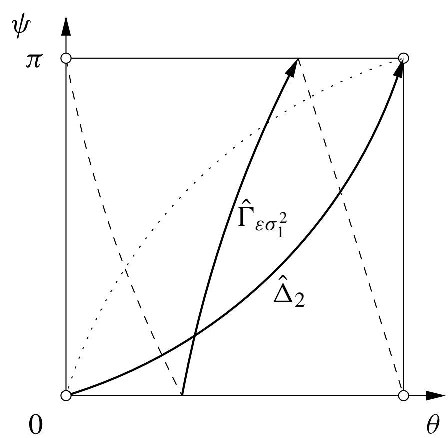


Figure 1. The pillowcase.

Example. Let $\sigma = \sigma_1^{2n}$. Then arguing as above one can show that $\hat{\Gamma}_{\varepsilon\sigma} \subset \hat{H}_2$ is given by $\psi = (2n+1)\theta - \pi$. In this case there are n intersection points all of which come with the same sign. This shows that $h(\sigma_1^{2n}) = \pm n$, which is again consistent with the fact that $\operatorname{lk}(\ell_1, \ell_2) = \pm n$.

5b. ***The difference cycle.*** Fix an orientation on a given two component link $\hat{\sigma}$. A particular choice of orientation will not matter because we are only interested in identifying $h(\sigma)$ with the linking number $\operatorname{lk}(\ell_1, \ell_2)$ up to sign. We wish to change one of the crossings between the two components of $\hat{\sigma}$. Using a sequence of first Markov moves, we may assume that the first two strands of σ belong to two different components of $\hat{\sigma}$, and that the crossing change occurs between these two strands. Furthermore, we may assume that the crossing change makes σ into $\sigma_1^{\pm 2}\sigma$, where the sign depends on the type of the crossing we change. Note that the braid $\sigma_1^{\pm 2}\sigma$ has the same permutation type as σ; in particular, its closure is a link of two components. In fact, if we let $\sigma' = \sigma_1^{-2}\sigma$, then

$$h(\sigma_1^{-2}\sigma) - h(\sigma) = h(\sigma') - h(\sigma_1^2\sigma') = -(h(\sigma_1^2\sigma') - h(\sigma'));$$

hence we only need to understand the difference $h(\sigma_1^2\sigma) - h(\sigma)$. Let us fix $\varepsilon = (-1, -1, 1, \dots, 1)$. Since σ_1^2 and ε commute, we have

$$\begin{aligned} h(\sigma_1^2\sigma) - h(\sigma) &= \langle \hat{\Delta}_n, \hat{\Gamma}_{\varepsilon\sigma_1^2\sigma} \rangle - \langle \hat{\Delta}_n, \hat{\Gamma}_{\varepsilon\sigma} \rangle \\ &= \langle \hat{\Gamma}_{\sigma_1^{-2}}, \hat{\Gamma}_{\varepsilon\sigma} \rangle - \langle \hat{\Delta}_n, \hat{\Gamma}_{\varepsilon\sigma} \rangle = \langle \hat{\Gamma}_{\sigma_1^{-2}} - \hat{\Delta}_n, \hat{\Gamma}_{\varepsilon\sigma} \rangle, \end{aligned}$$

where all intersection numbers are taken in $\hat{H}_n$. This leads us to consider the *difference cycle* $\hat{\Gamma}_{\sigma_1^{-2}} - \hat{\Delta}_n$ that is carried by $\hat{H}_n$. The next step in our argument will be to reduce the analysis of the intersection above to an intersection theory in the pillowcase $\hat{H}_2$.

5c. ***The pillowcase reduction.*** We consider the subset $V_n \subset H_n$ consisting of all $(X_1, \dots, X_n, Y_1, \dots, Y_n) \in H_n$ such that $X_k = Y_k$ for $k = 3, \dots, n$. Equivalently, V_n consists of all $(X_1, \dots, X_n, Y_1, \dots, Y_n) \in Q_n \times Q_n$ such that $(X_1, X_2, Y_1, Y_2) \in H_2$ and $X_k = Y_k$ for all $k = 3, \dots, n$. Therefore, V_n can be identified as

$$V_n = H_2 \times \Delta_{n-2} \subset (Q_2 \times Q_2) \times (Q_{n-2} \times Q_{n-2}).$$

Lemma 5.1. *The quotient* $\hat{V}_n = (H_2^* \times \Delta_{n-2})/\operatorname{SO}(3)$ *is a submanifold of* $\hat{H}_n$ *of dimension* $2n - 2$.

Proof. Since H_2^* and Δ_{n-2} are smooth manifolds of dimensions 5 and $2n - 4$, respectively, and their product contains no reducibles, the statement follows. □

Lemma 5.2. *The difference cycle* $\hat{\Gamma}_{\sigma_1^{-2}} - \hat{\Delta}_n$ *is carried by* $\hat{V}_n$.

Proof. Observe that neither $\hat{\Gamma}_{\sigma_1^{-2}}$ nor $\hat{\Delta}_n$ are subsets of $\hat{V}_n$. However, their portions that do not fit in $\hat{V}_n$,

$$\hat{\Gamma}_{\sigma_1^{-2}} - (\hat{\Gamma}_{\sigma_1^{-2}} \cap \hat{V}_n) \quad \text{and} \quad \hat{\Delta}_n - (\hat{\Delta}_n \cap \hat{V}_n),$$

are exactly the same. Namely, they consist of the equivalence classes of $2n$-tuples $(X_1, \dots, X_n, X_1, \dots, X_n)$ such that X_1 commutes with X_2. These cancel in the difference cycle $\hat{\Gamma}_{\sigma_1^{-2}} - \hat{\Delta}_n$, thus making it belong to $\hat{V}_n$. □

One can isotope $\hat{\Gamma}_{\varepsilon\sigma}$ into $\widetilde{\Gamma}_{\varepsilon\sigma}$ using an isotopy with compact support so that $\widetilde{\Gamma}_{\varepsilon\sigma}$ is transverse to $\hat{\Gamma}_{\sigma_1^{-2}} - \hat{\Delta}_n$. The latter means precisely that $\widetilde{\Gamma}_{\varepsilon\sigma}$ stays away from $(S_2 \times \Delta_{n-2})/\,\mathrm{SO}(3)$ and is transverse to both $\hat{\Gamma}_{\sigma_1^{-2}}$ and $\hat{\Delta}_n$; a precise argument can be found in [Heusener and Kroll 1998, page 491]. We further extend this isotopy to make $\widetilde{\Gamma}_{\varepsilon\sigma}$ transverse to $\hat{V}_n$ so that their intersection is a naturally oriented 1-dimensional submanifold of $\hat{H}_n$.

The natural projection $p : V_n \to H_2$ induces a map $\hat{p} : \hat{V}_n \to \hat{H}_2$. Use a further small compactly supported isotopy of $\widetilde{\Gamma}_{\varepsilon\sigma}$, if necessary, to make $\hat{p}(\hat{V}_n \cap \widetilde{\Gamma}_{\varepsilon\sigma})$ into a 1-submanifold of $\hat{H}_2$. The proofs of [Lin 1992, Lemmas 2.2 and 2.3] then go through with little change to give us the identity

$$\langle \hat{\Gamma}_{\sigma_1^{-2}} - \hat{\Delta}_n, \hat{\Gamma}_{\varepsilon\sigma} \rangle_{\hat{H}_n} = \langle \hat{\Gamma}_{\sigma_1^{-2}} - \hat{\Delta}_2, \hat{p}(\hat{V}_n \cap \tilde{\Gamma}_{\varepsilon\sigma}) \rangle_{\hat{H}_2}.$$

5d. ***Computation in the pillowcase.*** We first study the behavior of $\hat{p}(\hat{V}_n \cap \hat{\Gamma}_{\varepsilon\sigma})$ near the corners of $\hat{H}_2$.

Proposition 5.3. *There is a neighborhood around A' in the pillowcase $\hat{H}_2$ inside which $\hat{p}(\hat{V}_n \cap \hat{\Gamma}_{\varepsilon\sigma})$ is a curve approaching A'.*

Proof. Let us consider the submanifold

$$\Delta_n' = \{(X_1, X_2, X_3, \dots, X_n; Y_1, Y_2, X_3, \dots, X_n)\} \subset Q_n \times Q_n$$

and observe that $V_n \cap \Gamma_{\varepsilon\sigma} = \Delta_n' \cap \Gamma_{\varepsilon\sigma}$. We will show that the intersection of Δ_n' with $\Gamma_{\varepsilon\sigma}$ is transversal at $(\boldsymbol{i}, \varepsilon\boldsymbol{i}) = (i, \dots, i; -i, -i, i, \dots, i)$. This will imply that $\Delta_n' \cap \Gamma_{\varepsilon\sigma}$ is a submanifold of dimension four in a neighborhood of $(\boldsymbol{i}, \varepsilon\boldsymbol{i})$ and, after factoring out the SO(3) symmetry, that $\hat{p}(\hat{V}_n \cap \hat{\Gamma}_{\varepsilon\sigma})$ is a curve approaching $A' = p(\boldsymbol{i}, \varepsilon\boldsymbol{i})$.

Since $\dim \Delta_n' = 2n+4$, the dimension of $T_{(\boldsymbol{i},\varepsilon\boldsymbol{i})}(\Delta_n' \cap \Gamma_{\varepsilon\sigma}) = T_{(\boldsymbol{i},\varepsilon\boldsymbol{i})}\Delta_n' \cap T_{(\boldsymbol{i},\varepsilon\boldsymbol{i})}\Gamma_{\varepsilon\sigma}$ is at least four. Therefore, checking the transversality amounts to showing that this dimension is exactly four. Write

$$T_{(\boldsymbol{i},\varepsilon\boldsymbol{i})}(\Delta_n') = \{(u_1, \dots u_n; v_1, v_2, u_3, \dots, u_n)\} \subset T_{(\boldsymbol{i},\varepsilon\boldsymbol{i})}(Q_n \times Q_n),$$
$$T_{(\boldsymbol{i},\varepsilon\boldsymbol{i})}(\Gamma_{\varepsilon\sigma}) = \{(u_1, \dots, u_n; d_{\boldsymbol{i}}(\varepsilon\sigma)(u_1, \dots, u_n))\} \subset T_{(\boldsymbol{i},\varepsilon\boldsymbol{i})}(Q_n \times Q_n).$$

Then $T_{(\boldsymbol{i},\varepsilon\boldsymbol{i})}(\Delta'_n) \cap T_{(\boldsymbol{i},\varepsilon\boldsymbol{i})}(\Gamma_{\varepsilon\sigma})$ consists of the vectors $(u_1, \dots, u_n) \in T_{\boldsymbol{i}} Q_n = T_i S^2 \oplus \cdots \oplus T_i S^2$ that solve the matrix equation

$$\begin{bmatrix} d_{\boldsymbol{i}}(\sigma) \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} * \\ * \\ u_3 \\ \vdots \\ u_n \end{bmatrix}; \tag{11}$$

since $\varepsilon = (-1, -1, 1, \dots, 1)$, we can safely replace $[d_{\boldsymbol{i}}(\varepsilon\sigma)]$ by $[d_{\boldsymbol{i}}(\sigma)]$. It is shown in [Long 1989] that $[d_{\boldsymbol{i}}(\sigma)]$ is the Burau matrix of σ with parameter equal to -1. It is a real matrix acting on $T_{\boldsymbol{i}} Q_n = \mathbb{C}^n$; hence all we need to show is that the space of $(u_1, \dots, u_n) \in \mathbb{R}^n$ solving (11) has real dimension two. Let us write

$$\begin{bmatrix} d_{\boldsymbol{i}}(\sigma) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

where A is a 2×2 matrix and D is an $(n-2) \times (n-2)$ matrix. Equation (11) is equivalent to

$$\begin{bmatrix} C \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 1 - D \end{bmatrix} \begin{bmatrix} u_3 \\ \vdots \\ u_n \end{bmatrix},$$

so the proposition will follow as soon as we show that $1 - D$ is invertible. The invertibility of $1 - D$ is a consequence of the following two lemmas. □

Lemma 5.4. *Let $\sigma \in \mathcal{B}_n$. Then the Burau matrix of σ with parameter -1 and the permutation matrix of $\bar{\sigma}^{-1}$ are the same modulo 2.*

Proof. According to [Birman 1974], the Burau matrix of σ with parameter t is the matrix

$$\left.\frac{\partial \sigma(x_i)}{\partial x_j}\right|_{x_i = t},$$

where the x_i are generators of the free group and ∂ is the derivative in the Fox free differential calculus; see [Fox 1962]. Applying the Fox calculus we obtain

$$\begin{aligned} \frac{\partial \sigma(x_i)}{\partial x_j} = \frac{\partial (w x_{\bar{\sigma}^{-1}(i)} w^{-1})}{\partial x_j} &= \frac{\partial w}{\partial x_j} + w \left(\frac{\partial (x_{\bar{\sigma}^{-1}(i)} w^{-1})}{\partial x_j} \right) \\ &= \frac{\partial w}{\partial x_j} + w \left(\frac{\partial x_{\bar{\sigma}^{-1}(i)}}{\partial x_j} + x_{\bar{\sigma}^{-1}(i)} \frac{\partial w^{-1}}{\partial x_j} \right) \\ &= \frac{\partial w}{\partial x_j} + w \frac{\partial x_{\bar{\sigma}^{-1}(i)}}{\partial x_j} - w x_{\bar{\sigma}^{-1}(i)} w^{-1} \frac{\partial w}{\partial x_j}, \end{aligned}$$

where w is a word in the x_i. After evaluating at $t = -1$ and reducing modulo 2, the above becomes simply $\partial x_{\bar\sigma^{-1}(i)}/\partial x_j$, which is the permutation matrix of $\bar\sigma^{-1}$. □

Lemma 5.5. *Let $\sigma \in \mathcal{B}_n$ be such that $\hat\sigma$ is a two component link. Then $1 - D$ is invertible.*

Proof. Our assumption in this section has been that $\bar\sigma = (1, \dots)(2, \dots)$. We may further assume that

$$\bar\sigma = (1, 3, 4, \dots, k)(2, k+1, k+2, \dots, n)$$

by applying a sequence of first Markov moves fixing the first two strands of σ. The matrix $D \pmod 2$ is obtained by crossing out the first two rows and first two columns in the permutation matrix of $\bar\sigma$; see Lemma 5.4. This description implies that $D \pmod 2$ is upper diagonal, and hence so is $(1 - D) \pmod 2$. The diagonal elements of the latter matrix are all equal to one; therefore, $\det(1 - D) = 1 \pmod 2$ so $1 - D$ is invertible. □

Remark 5.6. The orientation of the component of $\hat p(\hat V_n \cap \hat\Gamma_{\varepsilon\sigma})$ limiting to A' can be read off its description near A' given in the proof of Proposition 5.3. In particular, this orientation is independent of the choice of σ.

Proposition 5.7. *There are neighborhoods around A and B' in the pillowcase $\hat H_2$ that are disjoint from $\hat p(\hat V_n \cap \hat\Gamma_{\varepsilon\sigma})$.*

Proof. The arguments for A and B' are essentially the same so we will only give the proof for A. Assuming the contrary we have a curve in $\hat V_n \cap \hat\Gamma_{\varepsilon\sigma}$ limiting to a reducible representation in $V_n \cap \Gamma_{\varepsilon\sigma}$. After conjugating if necessary, this representation must have the form

$$(i, i, e^{i\varphi_3}, \dots, e^{i\varphi_n}, i, i, e^{i\varphi_3}, \dots, e^{i\varphi_n}).$$

Using that $\varepsilon = (-1, -1, 1, \dots, 1)$ and arguing as in the proof of Proposition 3.5, we arrive at the contradiction $i = -i$. □

Proof of Theorem 2. According to Proposition 5.3, the 1-submanifold $\hat p(\hat V_n \cap \widetilde\Gamma_{\varepsilon\sigma})$, near A', is a curve approaching A'. According to Proposition 5.7, the other end of this curve must approach B. Therefore

$$h(\sigma_1^2\sigma) - h(\sigma) = \langle \hat\Gamma_{\sigma_1^{-2}} - \hat\Delta_2, \hat p(\hat V_n \cap \widetilde\Gamma_{\varepsilon\sigma})\rangle_{\hat H_2}$$

is the same as the intersection number of an arc going from A' to B with the difference cycle $\hat\Gamma_{\sigma_1^{-2}} - \hat\Delta_2$. This number is either 1 or -1 but is the same for all σ; see Remark 5.6. This is sufficient to prove that $h(\sigma)$ is the linking number up to an overall sign. □

References

[Alexander 1923] J. Alexander, "A lemma on systems of knotted curves", *Proc. Nath. Acad. Sci.* **9** (1923), 93–95. JFM 49.0408.03

[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

[Collin and Steer 1999] O. Collin and B. Steer, "Instanton Floer homology for knots via 3-orbifolds", *J. Differential Geom.* **51**:1 (1999), 149–202. MR 2000j:57069 Zbl 1025.57034

[Floer 1988] A. Floer, "An instanton-invariant for 3-manifolds", *Comm. Math. Phys.* **118**:2 (1988), 215–240. MR 89k:57028 Zbl 0684.53027

[Fox 1962] R. H. Fox, "A quick trip through knot theory", pp. 120–167 in *Topology of 3-manifolds and related topics* (Athens, GA, 1961), edited by M. K. Fort, Jr., Prentice-Hall, Englewood Cliffs, NJ, 1962. MR 25 #3522

[Herald 1997] C. M. Herald, "Flat connections, the Alexander invariant, and Casson's invariant", *Comm. Anal. Geom.* **5**:1 (1997), 93–120. MR 98i:57062 Zbl 0893.57003

[Heusener and Kroll 1998] M. Heusener and J. Kroll, "Deforming abelian SU(2)-representations of knot groups", *Comment. Math. Helv.* **73**:3 (1998), 480–498. MR 99g:57008 Zbl 0910.57004

[Kronheimer and Mrowka 2008] P. Kronheimer and T. Mrowka, "Knot homology groups from instantons", preprint, version 2, 2008. arXiv 0806.1053v2

[Lin 1992] X.-S. Lin, "A knot invariant via representation spaces", *J. Differential Geom.* **35**:2 (1992), 337–357. MR 93e:57012 Zbl 0774.57007

[Long 1989] D. D. Long, "On the linear representation of braid groups", *Trans. Amer. Math. Soc.* **311**:2 (1989), 535–560. MR 89h:20054 Zbl 0714.20019

[Ruberman and Saveliev 2004] D. Ruberman and N. Saveliev, "Rohlin's invariant and gauge theory, I: Homology 3-tori", *Comment. Math. Helv.* **79**:3 (2004), 618–646. MR 2006e:57020

[Taubes 1990] C. H. Taubes, "Casson's invariant and gauge theory", *J. Differential Geom.* **31**:2 (1990), 547–599. MR 91m:57025 Zbl 0702.53017

Received July 20, 2009.

ERIC HARPER
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MIAMI
PO BOX 249085
CORAL GABLES, FL 33124
UNITED STATES

harper@math.miami.edu

NIKOLAI SAVELIEV
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MIAMI
PO BOX 249085
CORAL GABLES, FL 33124
UNITED STATES

saveliev@math.miami.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

SEMIQUANDLES AND FLAT VIRTUAL KNOTS

ALLISON HENRICH AND SAM NELSON

We define an algebraic structure we call a *semiquandle*, whose axioms are derived from the flat Reidemeister moves. Finite semiquandles have associated counting invariants and enhanced invariants, defined for flat virtual knots and links. We also introduce *singular semiquandles* and *virtual singular semiquandles*, which define invariants of flat singular virtual knots and links. As an application, we use semiquandle invariants to compare two Vassiliev invariants.

1. Introduction

Recent works, such as [Kauffman 1999], take a combinatorial approach to knot theory, in which knots and links are regarded as equivalence classes of knot and link diagrams. New types of combinatorial knots and links can then be defined by introducing new types of crossings and Reidemeister-style moves that govern their interactions. These new combinatorial classes of knots and links have various topological and geometric interpretations relating to simple closed curves in 3-manifolds, rigid vertex isotopy of graphs, and so on.

A *flat crossing* is a classical crossing in which we ignore the over/under information. A flat knot or link is a planar projection (or shadow) of a knot or link, on the surface on which the knot or link diagram is drawn. Every classical knot diagram may be regarded as a *lift* of a flat knot and, conversely, every classical knot diagram has a corresponding flat *shadow*.

Flat knots might seem uninteresting since flattening classical crossings apparently throws away the information which defines the knotting. However, a little thought reveals potential applications of flat crossings; for example, invariants of links with classical intercomponent crossings and flat intracomponent crossings are related to link homotopy and the Milnor invariants of ordinary classical links.

Flat crossings also prove useful in virtual knot theory. Every purely flat knot is trivial, that is, reducible by flat Reidemeister moves to the unknot. However, flat virtual knots and links (that is, diagrams with virtual and flat crossings) are

MSC2000: 57M25, 57M27.
Keywords: flat knots and links, virtual knots and links, singular knots and links, semiquandles, Vassiliev invariants.

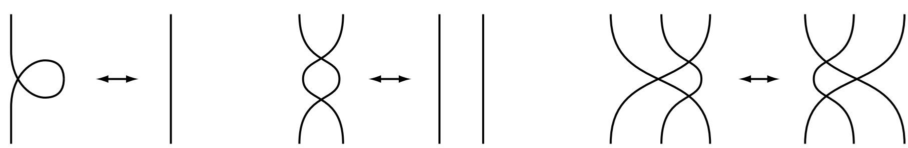
generally nontrivial. The nontriviality of a flat virtual knot says that no choice of classical crossing information for the flat crossings would yield a classical knot. Hence, flat crossings are useful in the study of nonclassicality for virtual knots.

A *singular crossing* is a crossing where two strands are fused together. Singular knots and links may be understood as rigid vertex isotopy classes of knotted and linked graphs, and they play a role in the study of Vassiliev invariants of classical knots and links.

In this paper, we define an algebraic structure we call a *semiquandle*, which yields counting invariants for flat virtual knots. The paper is organized as follows. In Section 2, we define flat, singular, and virtual knots and links. In Section 3, we define semiquandles and give some examples. In Section 4, we define singular semiquandles by including operations at singular crossings. In Section 5, we introduce virtual semiquandles and virtual singular semiquandles by including an operation at virtual crossings. In Section 6, we give examples to show that the counting invariants with respect to finite semiquandles can distinguish flat virtual knots and links. In Section 7, we give an application to Vassiliev invariants of virtual knots. In Section 8, we collect some questions for further research.

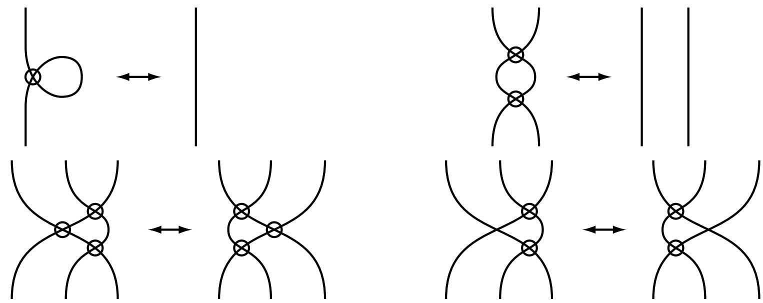
2. Flat knots, virtual knots and singular knots

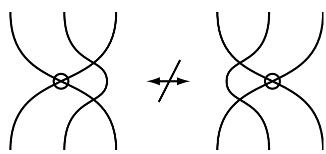
We introduce several types of knots discussed in this paper. We assume that all knots are oriented, unless otherwise specified. The simplest type of knot among those we consider, a *flat knot*, is an immersion of S^1 in $\mathbb{R}^2$. Alternatively, a flat knot can be described as an equivalence class of knot diagrams where under/over strand information at each crossing is unspecified. Their equivalence relation is given by flat versions of the Reidemeister moves, illustrated here:



It is an easy exercise to show that any flat knot is related by a sequence of flat Reidemeister moves to the trivial flat knot (that is, the flat knot with no crossings). While the theory of flat knots appears uninteresting, if we consider the analogous theory of flat virtual knots, we enter a highly nontrivial category.

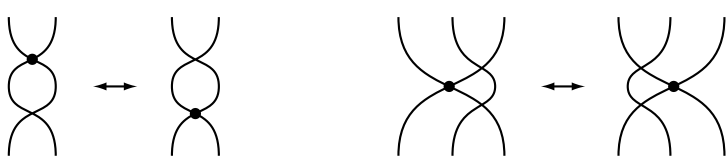
A *flat virtual knot* is a decorated immersion of S^1 in $\mathbb{R}^2$, where each crossing is decorated to indicate whether it is flat or virtual. (Virtual crossings are pictured by an encircled flat crossing.) Once again, we may also describe a flat virtual knot as an equivalence class of virtual knot diagrams where under/over strand information at each classical crossing is unspecified. The corresponding equivalence relation is given by the flat versions of the virtual Reidemeister moves, in addition to the flat versions of the ordinary Reidemeister moves.



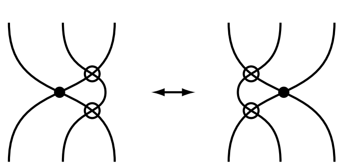
The move  is forbidden.

As with ordinary virtual knots, flat virtual knot diagrams have a geometric interpretation as flat knot diagrams on surfaces. In this case, the virtual crossings are interpreted as artifacts of a projection of the knot diagram on the surface to a knot diagram in the plane [Kauffman 1999].

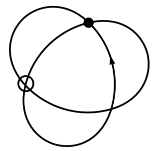
Finally, we consider flat knots and flat virtual knots that have singularities. The singularities should be thought of as rigid vertices, or places where the knot is actually glued to itself. Thus, *flat singular knots* are simply equivalence classes of flat knots where some crossings are decorated to indicate that they are singular. The Reidemeister moves corresponding to flat singular equivalence are the ordinary flat equivalence moves together with these two moves:



Similarly, *flat virtual singular knots* are equivalence classes of flat virtual knots where some of the crossings may be designated as singular. Hence, there are three types of crossings that may be contained in a diagram of a flat virtual singular knot. The equivalence relation is given by all of the previous flat, virtual, and singular moves, together with this:



The simplest nontrivial flat virtual singular knot that contains all three types of crossings is the *triple trefoil*:



3. Semiquandles

In recent years, a number of algebraic structures have been defined with axioms derived from variations on the oriented Reidemeister moves. The earliest of these is the *quandle* [Joyce 1982; Matveev 1982], in which we have generators corresponding to arcs in a link diagram, and an invertible binary operation at crossings.

Subsequent papers have generalized this idea in various ways. Fenn and Rourke [1992] replace ambient isotopy with framed isotopy to define *racks*. In [Kauffman and Radford 2003; Fenn et al. 1995], arcs in an oriented knot diagram are replaced with *semiarcs*, to define *biquandles*. Kauffman and Manturov [2005] include an operation at virtual crossings in the biquandle definition to yield *virtual biquandles*.

Definition 1. A *semiquandle* is a set X with two binary operations $(x, y) \mapsto x^y$ and $(x, y) \mapsto x_y$ such that, for all $x, y, z \in X$,

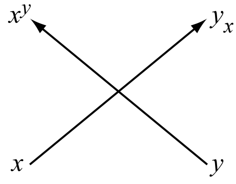
(0) there are *unique* w and $z \in X$ with $x = w^y$ and $x = z_y$;

(i) $x_y = y$ if and only if $y^x = x$;

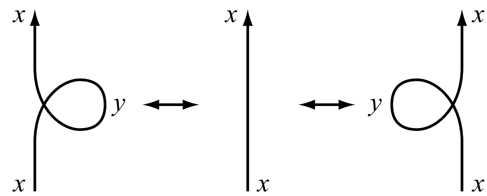
(ii) $(x_y)^{(y^x)} = x$ and $(x^y)_{(y_x)} = x$;

(iii) $(x^y)^z = (x^{z_y})^{y^z}$, $(y_x)^{z_{x^y}} = (y^z)_{x^{z_y}}$ and $(z_{x^y})_{y_x} = (z_y)_x$.

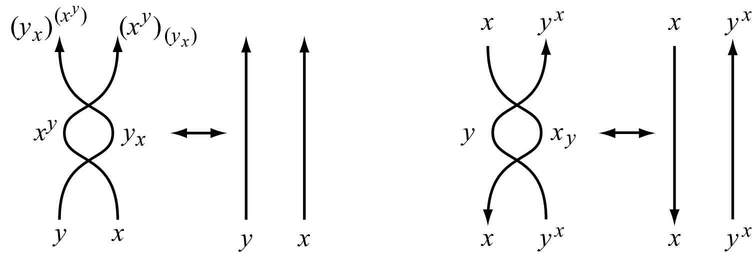
Axiom (0) says that the actions $x \mapsto x^y$ and $x \mapsto x_y$ are invertible; these unique z and w will be denoted by $z = x_{y^{-1}}$ and $w = x^{y^{-1}}$. The axioms come from dividing an oriented flat knot into semiarcs, that is, edges between vertices in the flat diagram regarded as a graph, and then translating the flat Reidemeister moves into algebraic axioms.



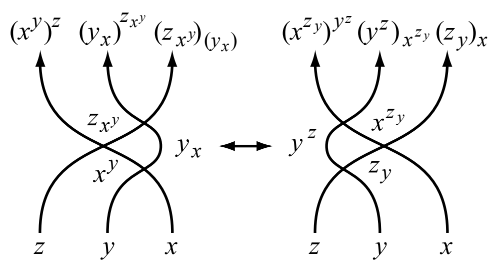
In the first Reidemeister move, right invertibility guarantees the uniqueness of y given x, and the relationship between x and y becomes axiom (i).



The direct II move, in which both strands are oriented in the same direction, gives us axiom (ii). Given axiom (0), the reverse II move yields the same relationship between x and y, where the uncrossed strands are labeled by x and y^x.

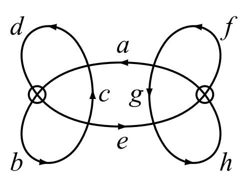


Reidemeister move III yields the three equations in axiom (iii).



Definition 2. For any flat virtual link L, the *fundamental semiquandle* FSQ(L) of L is the set of equivalence classes of semiquandle words in a set of generators corresponding to semiarcs in a diagram D of L, that is, edges in the graph obtained from D by considering flat crossings as vertices under the equivalence relation generated by the semiquandle axioms and the relations at the crossings. As with the knot quandle, fundamental rack, and knot biquandle, we can express the fundamental semiquandle with a presentation read from a diagram.

Example 3. The *flat Kishino knot* is



and has the fundamental semiquandle presentation

$$\mathrm{FSQ}(K) = \left\langle a, b, c, d, e, f, g, h \;\middle|\; \begin{array}{l} a^c = b,\ c_a = d,\ b^d = e,\ d_b = c, \\ e^g = f,\ g_e = h,\ f^h = g,\ h_f = a \end{array} \right\rangle.$$

Remark 4. An alternative definition for the fundamental semiquandle of a flat virtual knot is that $\mathrm{FSQ}(L)$ is the quotient of the (strong) knot biquandle of any lift of L (that is, of any choice of classical crossing type for the flat crossings of L), under the equivalence relation generated by setting $a^{\bar{b}} \sim a_b$ and $a_{\bar{b}} \sim a^b$ for all $a, b \in B(L)$. Indeed, this operation yields a "flattening" functor $\mathrm{SQ} : \mathcal{B} \to \mathcal{S}$ from the category of strong biquandles to the category of semiquandles.

Example 5. For any set X and bijection $\sigma : X \to X$, the operations $x^y = \sigma(x)$ and $x_y = \sigma^{-1}(x)$ define a semiquandle structure on X. We call this type of semiquandle a *constant-action semiquandle*, since the action of y on x is constant as y varies.

As is the case with quandles and biquandles [Ho and Nelson 2005; Nelson and Vo 2006], for a finite semiquandle $X = \{x_1, \ldots, x_n\}$ we can conveniently express the semiquandle structure with a block matrix $M_X = [U \mid L]$, where $U_{i,j} = k$ and $L_{i,j} = l$ for $x_k = (x_i)^{(x_j)}$ and $x_l = (x_i)_{(x_j)}$. This matrix notation enables us to do computations with semiquandles without the need for formulas for x^y and x_y.

Example 6. The constant-action semiquandle on $X = \{1, 2, 3\}$ with $\sigma = (132)$ has semiquandle matrix

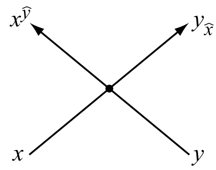
$$M_X = \left[\begin{array}{ccc|ccc} 3 & 3 & 3 & 2 & 2 & 2 \\ 1 & 1 & 1 & 3 & 3 & 3 \\ 2 & 2 & 2 & 1 & 1 & 1 \end{array}\right].$$

Example 7. Any (strong) biquandle in which $a^{\bar{b}} = a_b$ and $a_{\bar{b}} = a^b$ is a semiquandle. Indeed, an alternative name for semiquandles might be *symmetric biquandles*. An example of a nonconstant action semiquandle, found in [Nelson and Vo 2006], is

$$M_T = \left[\begin{array}{cccc|cccc} 1 & 4 & 2 & 3 & 1 & 3 & 4 & 2 \\ 2 & 3 & 1 & 4 & 3 & 1 & 2 & 4 \\ 4 & 1 & 3 & 2 & 2 & 4 & 3 & 1 \\ 3 & 2 & 4 & 1 & 4 & 2 & 1 & 3 \end{array}\right].$$

4. Singular semiquandles

We now consider what happens to our algebraic structure when we allow singular crossings in an oriented flat virtual knot. As with flat crossings, we define two binary operations at a singular crossing. One notable difference is that, unlike flat crossings, singular crossings are permanent — there are no moves that either introduce or remove singular crossings. Indeed, the number of singular crossings is an invariant of the singular knot type. In particular, at singular crossings, we do not need right invertibility for our operations.

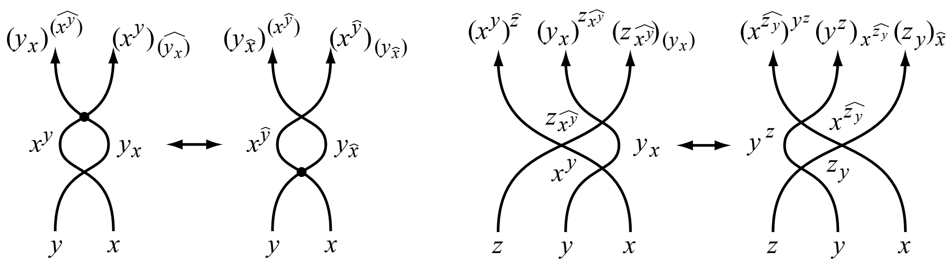


Definition 8. Let X be a semiquandle. A *singular semiquandle* structure on X is a pair $(x, y) \mapsto x^{\hat y}$ and $(x, y) \mapsto x_{\hat y}$ of binary operations on X that satisfy, for all $x, y, z \in X$,

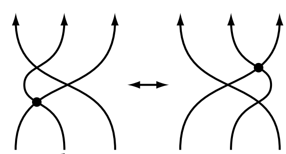
(hi) $(y_x)^{(\widehat{x^y})} = (y_{\hat x})^{(x^{\hat y})}$ and $(x^y)_{(\widehat{y_x})} = (x^{\hat y})_{(y_{\hat x})}$;

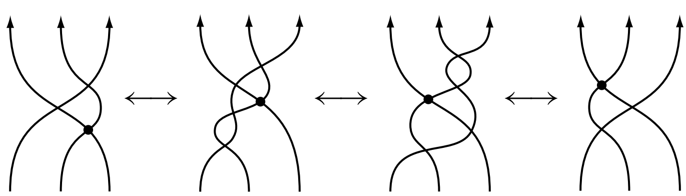
(hii) $(x^y)^{\hat z} = (x^{\widehat{z_y}})^{y^z}$, $(y_x)^{z_{\widehat{x^y}}} = (y^z)_{x^{\widehat{z_y}}}$, and $(z_{\widehat{x^y}})_{y_x} = (z_y)_{\hat x}$.

We call axioms (hi) and (hii) the *hat axioms*. These axioms come from the subset of the oriented singular flat Reidemeister moves pictured here:



To see that the two pictured oriented singular moves are sufficient to give us all of the oriented flat singular moves, we note the following key lemmas.

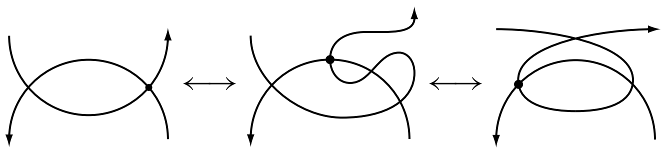
Lemma 9. *The move*  *follows from the flat Reidemeister moves and the two pictured moves.*

Proof.  □

Similar move sequences yield the other oriented flat/singular type III moves.

Lemma 10. *The reverse oriented singular II move follows from the flat moves and the moves pictured above.*

Proof. Starting with one side of a reverse singular II move, we can use flat moves to get a symmetric diagram in which we can apply a direct singular II move.



Reversing the process gives the other side of the reverse singular II move. □

Example 11. Let X be a semiquandle. Clearly, setting $x^{\hat{y}} = x^y$ and $x_{\hat{y}} = x_y$ for all $x, y \in X$ defines a compatible singular structure (X, X), which we call the *flat* singular structure.

Example 12. If X is a semiquandle, then $a^{\hat{b}} = a_{\hat{b}} = b$ is a compatible singular structure, since we have

$$(y_x)^{(\widehat{x^y})} = x^y = (y_{\hat{x}})^{(x^{\hat{y}})}, \qquad (y_x)^{z_{\widehat{x^y}}} = (y_x)^{x^y} = y = (y^z)_{\widehat{z_y}} = (y^z)_{x^{\widehat{z_y}}},$$

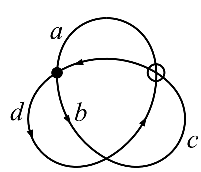
$$(x^y)_{(\widehat{y_x})} = y_x = (x^{\hat{y}})_{(y_{\hat{x}})}, \qquad (x^y)^{\hat{z}} = z = (z_y)^{y^z} = (x^{\widehat{z_y}})^{y^z},$$

$$(z_{\widehat{x^y}})_{y_x} = (x^y)_{y_x} = x = (z_y)_{\hat{x}}.$$

We call this singular structure the *operator* singular structure on X and denote it by (X, O).

As with the flat virtual case, for any flat singular virtual link L there is an associated *fundamental singular semiquandle* $\mathrm{FSSQ}(L)$, with presentation readable from the diagram. It elements are equivalence classes of singular semiquandle words, in generators corresponding to semiarcs in the diagram (here, we divide the diagram at both flat and singular crossing points, but not at virtual crossings) under the equivalence relation generated by the axioms (0), (i), (ii), (iii), (hi), and (hii).

Example 13. The *triple trefoil* below has the fundamental singular semiquandle presentation $\langle a, b, c, d \mid a^{\hat{c}} = b, c_{\hat{a}} = d, d^b = a,\ b_d = c \rangle$.



As with the semiquandle structure, we can represent the singular operations in a finite singular semiquandle by matrices encoding the operation tables. Indeed, it seems convenient to combine these matrices with the semiquandle operation matrices, into a single block matrix of the form

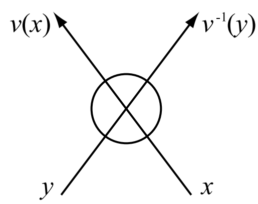
$$M_T = \left[\begin{array}{c|c} i^j & i_j \\ \hline i^{\hat{j}} & i_{\hat{j}} \end{array}\right].$$

Example 14. The constant action semiquandle $X = \{1, 2, 3\}$ with $\sigma = (132)$ and operator singular structure $a^{\hat{b}} = a_{\hat{b}} = b$ has block matrix

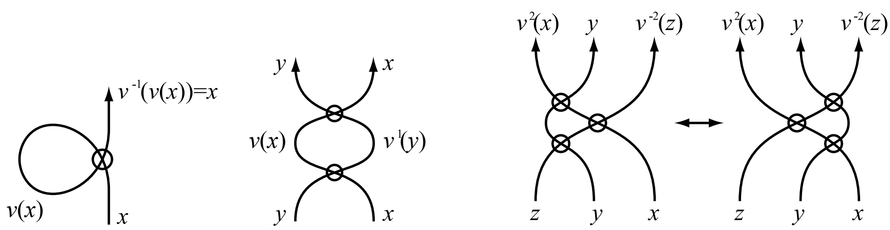
$$M_{(X,O)} = \left[\begin{array}{ccc|ccc} 3 & 3 & 3 & 2 & 2 & 2 \\ 1 & 1 & 1 & 3 & 3 & 3 \\ 2 & 2 & 2 & 1 & 1 & 1 \\ \hline 1 & 2 & 3 & 1 & 2 & 3 \\ 1 & 2 & 3 & 1 & 2 & 3 \\ 1 & 2 & 3 & 1 & 2 & 3 \end{array}\right].$$

5. Virtual semiquandles and virtual singular semiquandles

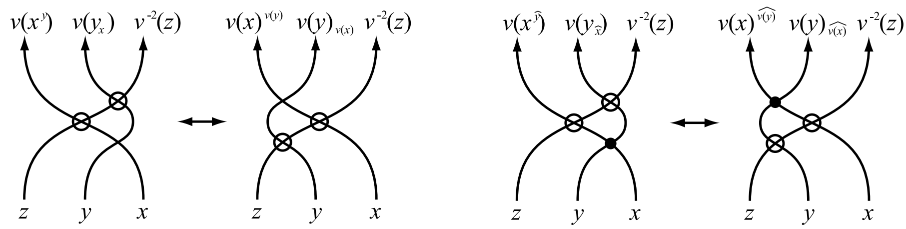
As with singular crossings, we can further generalize semiquandles by adding an operation at virtual crossings. The simplest way to do this is to use a unary operation at each virtual crossing, defined by applying a bijection v when going through a virtual crossing from right to left (looking in the direction of the strand being crossed) and applying v^{-1} when going through a virtual crossing from left to right (looking in the direction of the strand being crossed).



As noted in [Kauffman and Manturov 2005], this setup ensures that the virtual I, II and III moves are respected by the virtual operation.



The interaction of the virtual crossings with the flat and singular crossings given by the Reidemeister moves tells us how the virtual operation should interact with the semiquandle and singular semiquandle structures; namely, v must be an automorphism of both structures.



Definition 15. A *virtual semiquandle* is a semiquandle S with a choice of automorphism $v : S \to S$. A *virtual singular semiquandle* is a singular semiquandle with a semiquandle automorphism $v : S \to S$ that is also an automorphism of the singular structure. That is, $v : S \to S$ is a bijection satisfying

$$v(x^y) = v(x)^{v(y)}, \quad v(x_y) = v(x)_{v(y)}, \quad v(x^{\hat{y}}) = v(x)^{\widehat{v(y)}}, \quad v(x_{\hat{y}}) = v(x)_{\widehat{v(y)}}.$$

Example 16. Every semiquandle is a virtual semiquandle with $v = \mathrm{Id}_S$. More generally, the set of virtual semiquandle structures on a semiquandle S corresponds to the set of conjugacy classes in the automorphism group $\mathrm{Aut}(S)$ of the semiquandle S: if $v, v', \varphi \in \mathrm{Aut}(S)$ with $v' = \varphi^{-1} v \varphi$, then $\varphi(S, v) \to (S, v')$ is an isomorphism of virtual semiquandles.

Every flat singular virtual knot or link has a *fundamental virtual singular semiquandle*, obtained by dividing the knot or link into semiarcs at flat, singular and virtual crossings; then $\mathrm{FVSSQ}(L)$ has generators corresponding to semiarcs, and relations at the crossings as determined by crossing type, in addition to relations coming from the virtual singular semiquandle axioms.

6. Counting invariants of flat singular virtuals

As with finite groups, quandles, and biquandles, finite semiquandles can be used to define computable invariants of flat virtual knots and links by counting homomorphisms.

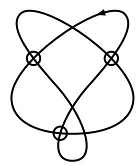
Definition 17. Let L be a flat virtual link and T a finite semiquandle. The *semiquandle counting invariant* of L with respect to T is the cardinality

$$\mathrm{sc}(L, T) = |\mathrm{Hom}(\mathrm{FSQ}(L), T)|$$

of the set of semiquandle homomorphisms $f : \mathrm{FSQ}(L) \to T$ from the fundamental semiquandle of L to T (that is, of maps such that $f(x_y) = f(x)_{f(y)}$ and $f(x^y) = f(x)_{f(y)}$ for all $x, y \in \mathrm{FSQ}(L)$).

Remark 18. A semiquandle homomorphism $f : \mathrm{FSQ}(L) \to T$ can be pictured as a "coloring" of a diagram D of L by T, that is, as an assignment of an element of T to each semiarc in D such that the colors satisfy the semiquandle operation conditions at every crossing.

Example 19. The semiquandle counting invariant with respect to the semiquandle T in Example 7 distinguishes the flat Kishino knot FK from the flat unknot FU, with $\mathrm{sc}(\mathrm{FK}, T) = 16$ and $\mathrm{sc}(\mathrm{FU}, T) = 4$. This same semiquandle also distinguishes the flat virtual knot K below [Kauffman 1999] from both the unknot and the flat Kishino, with $\mathrm{sc}(K, T) = 2$.



We can enhance the semiquandle counting invariant by taking note of the cardinality of the image subsemiquandles $\mathrm{Im}(f)$ for each homomorphism to obtain a multiset-valued invariant, which we can also express in a polynomial form by converting multiset elements to exponents of a dummy variable z and multiplicities to coefficients. Note that specializing $z = 1$ in the enhanced invariant yields the original counting invariant.

Definition 20. Let L be a flat virtual link and T a finite semiquandle. The *enhanced semiquandle counting multiset* is the multiset

$$\mathrm{sqcm}(L, T) = \{\mathrm{Im}(f) \mid f \in \mathrm{Hom}(\mathrm{FSQ}(L), T))\},$$

and the *enhanced semiquandle polynomial* is

$$\mathrm{sqp}(L, T) = \sum_{f \in \mathrm{Hom}(\mathrm{FSQ}(L), T)} z^{|\mathrm{Im}(f)|}.$$

For singular semiquandles, we also have counting invariants and polynomial enhanced invariants.

Definition 21. Let L be a flat singular virtual link and (T, S) a finite singular semiquandle. We have the *singular semiquandle counting invariant*

$$\mathrm{ssc}(L, (T, S)) = |\mathrm{Hom}(\mathrm{FSSQ}(L), (T, S))|,$$

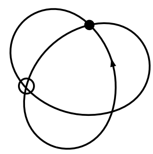
the *enhanced singular semiquandle counting multiset*

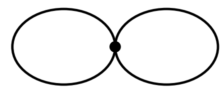
$$\mathrm{ssqcm}(L, (T, S)) = \{\, \mathrm{Im}(f) \mid f \in \mathrm{Hom}(\mathrm{FSSQ}(L), (T, S)))\,\},$$

and the *enhanced singular semiquandle polynomial*

$$\mathrm{ssqp}(L, (T, S)) = \sum_{f \in \mathrm{Hom}(\mathrm{FSSQ}(L), (T,S))} z^{|\mathrm{Im}(f)|}.$$

Example 22. The constant-action semiquandle $(X_{(132)}, O)$ with operator singular structure distinguishes the triple trefoil TT from the singular knot SU_1 with one singular crossing and no other crossings:





$\mathrm{ssqp}(\mathrm{TT}, (X_{(132)}, O)) = 0 \qquad \mathrm{ssqp}(\mathrm{SU}_1, (X_{(132)}, O)) = 9z^3$

Finally, we have counting invariants for flat singular virtual knots and links, defined analogously using finite virtual singular semiquandles.

Definition 23. Let L be a flat singular virtual link and (T, S, v) a finite virtual singular semiquandle. Then we have the *virtual singular semiquandle counting invariant* $\mathrm{vssc}(L, (T, S, v)) = |\mathrm{Hom}(\mathrm{FVSSQ}(L), (T, S, v))|$, the *enhanced virtual singular semiquandle counting multiset*

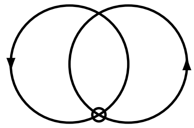
$$\mathrm{vssqcm}(L, (T, S, v)) = \{\mathrm{Im}(f) \mid f \in \mathrm{Hom}(\mathrm{FVSSQ}(L), (T, S, v))\},$$

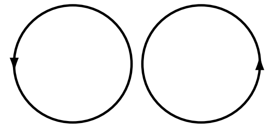
and the *enhanced virtual singular semiquandle polynomial*

$$\mathrm{vssqp}(L, (T, S, v)) = \sum_{f \in \mathrm{Hom}(\mathrm{FVSSQ}(L), (T,S,v))} z^{|\mathrm{Im}(f)|}.$$

Example 24. The *flat virtual Hopf link* fvH below is distinguished from the flat unlink of two components by the counting invariants with respect to the listed virtual semiquandle. Note that we can regard T as a flat singular virtual semiquandle with trivial singular operations $x^{\hat{y}} = x = x_{\hat{y}}$.

$$M_{T,S} = \left[\begin{array}{ccc|ccc} 1 & 3 & 1 & 1 & 3 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 \\ 3 & 1 & 3 & 3 & 1 & 3 \end{array}\right], \qquad v = (13)$$





$\mathrm{vsqp}(\mathrm{fvH}, (T, S, v)) = q + 4z^2 \qquad \mathrm{vsqp}(U_2, (T, S, v)) = q + 4z^2 + 4z^3$

Remark 25. A virtual semiquandle is a virtual singular semiquandle with trivial singular structure, that is, $x^{\hat{y}} = x_{\hat{y}} = x$; a singular semiquandle is a virtual singular semiquandle with trivial virtual operation, that is, $v = \mathrm{Id}$; and a semiquandle is a virtual singular semiquandle with trivial virtual and singular structures.

7. Application to Vassiliev invariants

In [Henrich 2010], one finds several degree-one Vassiliev invariants for virtual knots. One invariant, **S**, takes its values in the free abelian group on the set of two-component flat virtual links. Another invariant, **G**, takes its values in the free abelian group on the set of flat virtual singular knots with one singularity. It is easy to show that **G** is at least as strong as **S**, but somewhat difficult to show that **G** is strictly stronger than **S**. Here, we give the definitions of these invariants and provide an alternative proof that **G** is strictly stronger than **S**.

Definition 26. Let K be a virtual knot with diagram $\widetilde{K}$. Let $\widetilde{K}^d_{\text{smooth}}$ be the flat virtual link obtained by smoothing $\widetilde{K}$ at the crossing d and projecting onto the associated flat virtual link. Furthermore, let $\widetilde{K}^0_{\text{link}}$ be the flat virtual link obtained by taking the disjoint union of the unknot with the flat projection of $\widetilde{K}$. We let the bracket $[\,\cdot\,]$ denote the associated generator of the free abelian group on the set of (two-component) flat virtual links. Then **S** is given by the following element of this free abelian group:

$$\mathbf{S}(K)=\sum_d \operatorname{sign}(d)([\widetilde{K}^d_{\text{smooth}}]-[\widetilde{K}^0_{\text{link}}]).$$

The sum ranges over all classical crossings in $\widetilde{K}$, and $\operatorname{sign}(d)$ is the local writhe.

Since this “smoothing” invariant has values involving flat virtual links, it is clear that semiquandles may be of use in computing **S** for pairs of virtual knots. Moreover, singular semiquandles can be put to use when computing the next invariant:

Definition 27. Let K be a virtual knot with diagram $\widetilde{K}$. Let $\widetilde{K}^d_{\text{glue}}$ be the flat virtual singular knot obtained by gluing $\widetilde{K}$ at the crossing d and projecting onto the associated flat virtual singular knot. Let $\widetilde{K}^0_{\text{sing}}$ be the flat virtual singular knot obtained by taking the flat projection of $\widetilde{K}$, introducing a kink via the flat Reidemeister I move, and gluing at the resulting crossing. Here, we let the bracket $[\,\cdot\,]$ denote the associated generator of the free abelian group on the set of flat virtual singular knots (with one singularity). Using this notation, we define **G** by $\mathbf{G}(K)=\sum_d \operatorname{sign}(d)([\widetilde{K}^d_{\text{glue}}]-[\widetilde{K}^0_{\text{sing}}])$. Again, the sum ranges over all classical crossings in $\widetilde{K}$ and $\operatorname{sign}(d)$ is the local writhe.

Henrich [2010] proved that both **S** and **G** are degree-one Vassiliev invariants of virtual knots, and **G** is at least as strong as **S**. To show that **G** is stronger than **S**, consider this pair of virtual knots in Figure 1. Call the first of these knots K_1, and the second K_2. Since the only difference between the two knots is the signs of the crossings labeled a and b, we see that

$$\mathbf{S}(K_2)-\mathbf{S}(K_1)=2([\widetilde{K}^a_{\text{smooth}}]-[\widetilde{K}^b_{\text{smooth}}]),\quad \text{and}$$
$$\mathbf{G}(K_2)-\mathbf{G}(K_1)=2([\widetilde{K}^a_{\text{glue}}]-[\widetilde{K}^b_{\text{glue}}]).$$

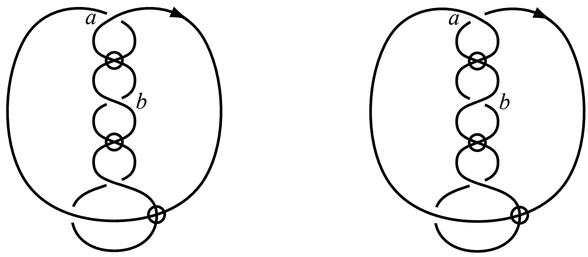
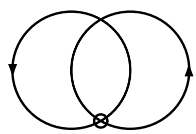
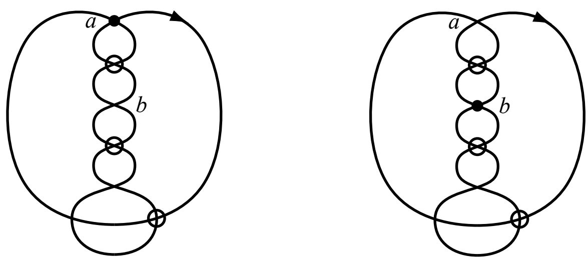


Figure 1

Now, $\widetilde{K}^a_{\text{smooth}}$ is the same as $\widetilde{K}^b_{\text{smooth}}$; they are both this flat virtual link:



It follows that $\mathbf{S}(K_1) = \mathbf{S}(K_2)$. On the other hand, we can show, by using singular semiquandles, that $\widetilde{K}^a_{\text{glue}}$ and $\widetilde{K}^b_{\text{glue}}$, as pictured next, are distinct.



Consider the following singular semiquandle, T, given in terms of its matrix M.

$$M = \left[\begin{array}{cccc|cccc} 1 & 4 & 2 & 3 & 1 & 3 & 4 & 2 \\ 2 & 3 & 1 & 4 & 3 & 1 & 2 & 4 \\ 4 & 1 & 3 & 2 & 2 & 4 & 3 & 1 \\ 3 & 2 & 4 & 1 & 4 & 2 & 1 & 3 \\ \hline 1 & 1 & 4 & 4 & 1 & 2 & 2 & 1 \\ 1 & 1 & 4 & 4 & 4 & 3 & 3 & 4 \\ 2 & 2 & 3 & 3 & 4 & 3 & 3 & 4 \\ 2 & 2 & 3 & 3 & 1 & 2 & 2 & 1 \end{array}\right]$$

The enhanced singular semiquandle polynomials for $\widetilde{K}^a_{\text{glue}}$ and $\widetilde{K}^b_{\text{glue}}$ are

$$\text{ssqp}(\widetilde{K}^a_{\text{glue}}, T) = 2z \quad \text{and} \quad \text{ssqp}(\widetilde{K}^b_{\text{glue}}, T) = 2z + 2z^4.$$

Hence, the two flat virtual singular knots are distinct, and thus $\mathbf{G}(K_1) \neq \mathbf{G}(K_2)$.

8. Questions

In this section, we collect questions for future research.

Singular semiquandles bear a certain resemblance to virtual biquandles, in which a biquandle is augmented with operations at virtual crossings. Given a biquandle B, the set of virtual biquandle structures on B forms a group isomorphic to the automorphism group of B. What is the structure of the set of singular semiquandle structures on a semiquandle X?

Our algebra-agnostic approach to the computation of our various semiquandle-based invariants works well for small-cardinality semiquandles and for link diagrams with small crossing numbers. However, for links with higher crossing numbers and for larger coloring semiquandles, we will need more algebraic descriptions. We have given a few examples of classes of semiquandle structures, such as constant-action semiquandles and operator singular structures. What are some examples of group-based or module-based semiquandle and singular semiquandle structures, akin to Alexander biquandles? (Note that the only Alexander biquandles which are semiquandles are constant-action Alexander biquandles.)

Enhancement techniques for biquandle counting invariants that should extend to semiquandles include *semiquandle cohomology*, which is the special case of Yang–Baxter cohomology described in [Carter et al. 2004], and the flattened case of S-cohomology as described in [Ceniceros and Nelson 2009]. Similarly, we might define *semiquandle polynomials* and the resulting enhancements of the counting invariants, as in [Nelson 2008]. What other enhancements of semiquandle counting invariants are there?

What is the relationship, if any, between semiquandle invariants and the quaternionic biquandle invariants described in [Bartholomew and Fenn 2008]?

Our Python code for computing semiquandle-based invariants is available from the second author's website at http://www.esotericka.org.

References

[Bartholomew and Fenn 2008] A. Bartholomew and R. Fenn, "Quaternionic invariants of virtual knots and links", *J. Knot Theory Ramifications* **17**:2 (2008), 231–251. MR 2009b:57006 Zbl 1149.57301

[Carter et al. 2004] J. S. Carter, M. Elhamdadi, and M. Saito, "Homology theory for the set-theoretic Yang–Baxter equation and knot invariants from generalizations of quandles", *Fund. Math.* **184** (2004), 31–54. MR 2005k:57009 Zbl 1067.57006

[Ceniceros and Nelson 2009] J. Ceniceros and S. Nelson, "Virtual Yang–Baxter cocycle invariants", *Trans. Amer. Math. Soc.* **361**:10 (2009), 5263–5283. MR 2010k:57021 Zbl 1182.57006

[Fenn and Rourke 1992] R. Fenn and C. Rourke, "Racks and links in codimension two", *J. Knot Theory Ramifications* **1**:4 (1992), 343–406. MR 94e:57006 Zbl 0787.57003

[Fenn et al. 1995] R. Fenn, C. Rourke, and B. Sanderson, "Trunks and classifying spaces", *Appl. Categ. Structures* **3**:4 (1995), 321–356. MR 96i:57023 Zbl 0853.55021

[Henrich 2010] A. Henrich, "A sequence of degree one Vassiliev invariants for virtual knots", *J. Knot Theory Ramifications* **19**:4 (2010), 461–487. MR 2646641 Zbl 05709835

[Ho and Nelson 2005] B. Ho and S. Nelson, "Matrices and finite quandles", *Homology Homotopy Appl.* **7**:1 (2005), 197–208. MR 2006h:20094 Zbl 1102.57008

[Joyce 1982] D. Joyce, "A classifying invariant of knots, the knot quandle", *J. Pure Appl. Algebra* **23**:1 (1982), 37–65. MR 83m:57007 Zbl 0474.57003

[Kauffman 1999] L. H. Kauffman, "Virtual knot theory", *European J. Combin.* **20**:7 (1999), 663–690. MR 2000i:57011 Zbl 0938.57006

[Kauffman and Manturov 2005] L. H. Kauffman and V. O. Manturov, "Virtual biquandles", *Fund. Math.* **188** (2005), 103–146. MR 2006k:57015 Zbl 1088.57006

[Kauffman and Radford 2003] L. H. Kauffman and D. Radford, "Bi-oriented quantum algebras, and a generalized Alexander polynomial for virtual links", pp. 113–140 in *Diagrammatic morphisms and applications* (San Francisco, 2000), edited by D. E. Radford et al., Contemp. Math. **318**, Amer. Math. Soc., Providence, RI, 2003. MR 2004c:57013 Zbl 1031.57009

[Matveev 1982] S. V. Matveev, "Distributive groupoids in knot theory", *Mat. Sb.* (*N.S.*) **119 (161)**:1 (1982), 78–88, 160. MR 84e:57008 Zbl 0523.57006

[Nelson 2008] S. Nelson, "Generalized quandle polynomials", preprint, 2008. To appear in *Can. Bull. Math.* arXiv 0801.2979

[Nelson and Vo 2006] S. Nelson and J. Vo, "Matrices and finite biquandles", *Homology, Homotopy Appl.* **8**:2 (2006), 51–73. MR 2007j:57009 Zbl 1111.57009

Received April 28, 2009.

ALLISON HENRICH
SEATTLE UNIVERSITY
MATHEMATICS DEPARTMENT
901 12TH AVENUE
PO BOX 22200
SEATTLE, WA 98122
UNITED STATES
henricha@seattleu.edu

SAM NELSON
CLAREMONT MCKENNA COLLEGE
DEPARTMENT OF MATHEMATICAL SCIENCES
850 COLUMBIA AVE
CLAREMONT CA, 91711
UNITED STATES
knots@esotericka.org

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

INFINITESIMAL RIGIDITY OF POLYHEDRA WITH VERTICES IN CONVEX POSITION

IVAN IZMESTIEV AND JEAN-MARC SCHLENKER

Let $P \subset \mathbb{R}^3$ be a polyhedron. It was conjectured that if P is weakly convex (that is, its vertices lie on the boundary of a strictly convex domain) and decomposable (that is, P can be triangulated without adding new vertices), then it is infinitesimally rigid. We prove this conjecture under a weak additional assumption of codecomposability.

The proof relies on a result of independent interest about the Hilbert–Einstein function of a triangulated convex polyhedron. We determine the signature of the Hessian of that function with respect to deformations of the interior edges. In particular, if there are no interior vertices, then the Hessian is negative definite.

1. Introduction

The rigidity of convex polyhedra. The rigidity of convex polyhedra is a classical result in geometry, first proved by Cauchy [1813] using ideas going back to Legendre [1794, note XII, pages 321–334].

Theorem 1.1 [Cauchy 1813; Legendre 1794]. *Let P, $Q \subset \mathbb{R}^3$ be two convex polyhedra with the same combinatorics whose corresponding faces are isometric. Then P and Q are congruent.*

This result had a profound influence on geometry over the last two centuries. It led for instance to the discovery of the rigidity of smooth convex surfaces in $\mathbb{R}^3$, to Alexandrov's rigidity, and to his results on the realization of positively curved cone-metrics on the boundary of polyhedra; see [Alexandrov 2005].

From a practical viewpoint, global rigidity is perhaps not as relevant as infinitesimal rigidity (see Definition 1.8). Although the infinitesimal rigidity of convex polyhedra can be proved using Cauchy's argument, the first proof was given much later, and is completely different from Cauchy's.

Theorem 1.2 [Dehn 1916]. *Any convex polyhedron is infinitesimally rigid.*

MSC2000: 52B10, 52C25.
Keywords: rigidity, polyhedra, nonconvex, Hilbert–Einstein functional.
Izmestiev was supported by the DFG Research Unit "Polyhedral Surfaces".

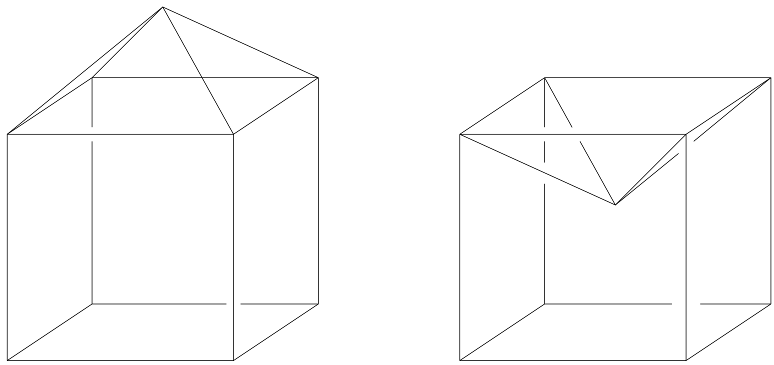


Figure 1. Cauchy's rigidity theorem fails for nonconvex polyhedra.

Neither Theorem 1.1 nor Theorem 1.2 extends to nonconvex polyhedra. It is easy to find a counterexample to the extension of Theorem 1.1 to nonconvex polyhedra; see Figure 1. Counterexamples to the extension of Theorem 1.2 are more complicated; see Figure 2.

In this paper we deal with a generalization of Theorem 1.2 to a vast class of nonconvex polyhedra. The main idea is that it is not necessary to consider convex polyhedra; what is important is that the vertices are in convex position. Additional assumptions are necessary, but are automatically satisfied for convex polyhedra.

Main result. By a polyhedron we mean a body in $\mathbb{R}^3$ bounded by a closed polyhedral surface.

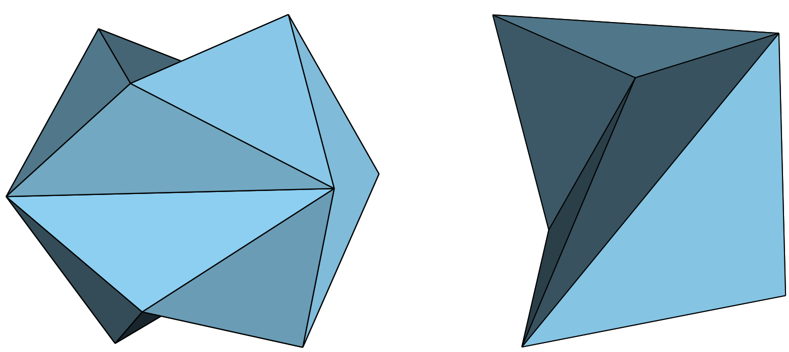


Figure 2. Examples of infinitesimally flexible polyhedra: Jessen's orthogonal icosahedron [Jessen 1967; Weisstein] and Schönhardt's polyhedron [Schönhardt 1928; Wunderlich 1965]. Both examples are weakly convex (Definition 1.3) but not decomposable (Definition 1.4).

Definition 1.3. A polyhedron $P \subset \mathbb{R}^3$ is called *weakly convex* if its vertices are in convex position in $\mathbb{R}^3$.

In other words, P is weakly convex if its vertices are the vertices of a strictly convex polyhedron.

Definition 1.4. A polyhedron P is called *decomposable* if it can be triangulated without adding new vertices.

In other words, every simplex of the triangulation must have vertices among those of P.

Our work was motivated by the following conjecture.

Conjecture 1.5. *Every weakly convex decomposable polyhedron is infinitesimally rigid.*

Any infinitesimally flexible polyhedron known to us fails to satisfy one of the assumptions of Conjecture 1.5. Thus, both polyhedra on Figure 2 are weakly convex but not decomposable. The infinitesimally flexible nonconvex octahedron pictured in [Gluck 1975] is decomposable but not weakly convex.

The main result of this paper is the proof of a weakening of Conjecture 1.5. We state it for polyhedra with triangular faces for simplicity, and explain in the next subsection how it extends to polyhedra with nontriangular faces. To state it, we need another, simple definition.

Definition 1.6. We call a polyhedron P *codecomposable* if its complement in $\operatorname{conv} P$ can be triangulated without adding new vertices. We call P *weakly codecomposable* if P is contained in a convex polyhedron Q, such that all vertices of P are vertices of Q and that the complement of P in Q can be triangulated without adding new vertices.

Theorem 1.7. *Let P be a weakly convex, decomposable, and weakly codecomposable polyhedron with triangular faces. Then P is infinitesimally rigid.*

Note that P is not required to be homeomorphic to a ball. The hypothesis that P is weakly codecomposable, however, appears to be quite weak for polyhedra homeomorphic to a ball. In the appendix we describe a simple example of a polyhedron that is not weakly codecomposable, but however is not homeomorphic to a ball; it's quite possible that this example can be modified fairly simply to make it contractible.

It is easy to come up with many examples of polyhedra to which Theorem 1.7 applies. Consider a convex polyhedron Q, and select an edge e of Q adjacent to two triangular faces f and f'. Cut out from Q the simplex that has f and f' as two of its faces, and let Q_1 be the nonconvex polyhedron obtained. Combinatorially, Q_1 is the same as Q, except that the edge e has been removed and replaced by the other diagonal of the quadrilateral made of the two triangular faces adjacent to e. By construction, Q_1 is weakly convex and weakly codecomposable; it is easy to check that it is decomposable (actually, star-shaped with respect to at least 2 of its vertices). This operation of cutting out a simplex can then be repeated, to obtain polyhedra $Q_2, Q_3, \ldots$, which are always weakly convex and weakly codecomposable. It is not guaranteed, however, that they remain decomposable, and indeed the Schönhardt polyhedron depicted above shows that they might cease to be decomposable.

On the other hand, examples of noncodecomposable weakly convex polyhedra homeomorphic to a ball are quite complicated [Aichholzer et al. 2002], so a counterexample to Conjecture 1.5 would be difficult to construct. On the other hand, the codecomposability assumption is used in our proof of Theorem 1.7 in a very essential way. Thus the question whether the codecomposability assumption may be omitted remains wide open. (However, this assumption does not appear in [Schlenker 2005].)

There is another, clearly equivalent way to state Theorem 1.7. Let Q be a convex polyhedron, with a triangulation T with no interior vertex (all vertices of the simplices of T are vertices of Q). Let Σ be a subcomplex of T, homeomorphic to a closed surface. Then Σ, considered as a polyhedral surface, is infinitesimally rigid. This statement is also, in an obvious way, an extension of the Cauchy–Dehn rigidity result, Theorem 1.2.

Polyhedra with nontriangular faces. It is well known that to prove the infinitesimal rigidity of a polyhedron with nontriangular faces, it suffices to prove it is so after triangulating the faces; see for example [Alexandrov 2005]. In Theorem 1.7 the fact that two triangular faces are coplanar makes no difference, so it basically extends as is to polyhedra with some nontriangular faces. There is however a slightly subtle point that should be mentioned.

If a polyhedron P with some nontriangular faces is decomposable, then there is a triangulation of the faces that is compatible with its triangulation. Similarly, if P is codecomposable, there is a triangulation of its faces that is compatible with a decomposition of the complement of P in its convex hull. However, it is conceivable that P could be decomposable and codecomposable, but such that there is no triangulation of its nontriangular faces that is compatible both with a triangulation of P and with a triangulation of the complement of P in its convex hull. In this case, Theorem 1.7 would not apply to P. We have no example of such a polyhedron, and do not know whether any such example exists.

From this point on, we only consider polyhedra with triangular faces.

Earlier results. Conjecture 1.5 originated as a question in [Schlenker 2005], where a related result was proved: If P is a decomposable polyhedron such that there exists an ellipsoid that intersects all edges of P but contains none of its vertices, then P is infinitesimally rigid. The proof relies on hyperbolic geometry, more precisely the properties of the volume of hyperideal hyperbolic polyhedra.

Connelly and Schlenker [2010] then proved two special cases of the conjecture: when P is a weakly convex suspension containing its north-south axis, and when P has only one concave edge, or two concave edges adjacent to a vertex. The proof for suspensions was based on stress arguments, while the proof of the other result used a refinement of Cauchy's argument.

More recently, Schlenker [2009] proved that the conjecture holds when P is star-shaped with respect to one of its vertices. This implies the two results in [Connelly and Schlenker 2010]. The proof was based on recent results of [Izmestiev 2008] concerning convex caps. Schlenker's result — and therefore the two results of Connelly and Schlenker — are consequences of Theorem 1.7, since it is not difficult to show that a polyhedron that is star-shaped with respect to one of its vertices is codecomposable (the proof actually appears as a step in [Schlenker 2009]).

Definitions. Every polyhedron has faces, edges, and vertices. As mentioned above we only consider polyhedra with triangular faces.

Definition 1.8. A polyhedron P is called *infinitesimally rigid* if every infinitesimal flex of its 1-skeleton is trivial.

Definition 1.9. By an *infinitesimal flex* of a graph in $\mathbb{R}^3$, we mean an assignment of vectors to the vertices of the graph such that the displacements of the vertices in the assigned directions induce a zero first-order change of the edge lengths:

$$(p_i - p_j) \cdot (q_i - q_j) = 0 \quad \text{for every edge } p_i p_j,$$

where q_i is the vector associated to the vertex p_i. An infinitesimal flex is called trivial if it is the restriction of an infinitesimal rigid motion of $\mathbb{R}^3$.

Polyhedra with flat vertices. We will need to deal with triangulations of the boundaries of polyhedra that contain additional vertices. Infinitesimal flexes and infinitesimal rigidity for such *triangulated spheres* are defined in the same way. Note that a triangulated sphere may be infinitesimally flexible even if it bounds a convex polyhedron, see Figure 3.

Definition 1.10. Let S be a triangulation of the boundary of a polyhedron P. A vertex p of S is called a *flat vertex* if it lies in the interior of a face of P.

The following statement is an easy generalization of Dehn's theorem.

Theorem 1.11. *Let S be a triangulation of the boundary of a convex polyhedron P. Then every infinitesimal flex of S is the sum of an infinitesimal rigid motion and of displacements of flat vertices in the directions orthogonal to their ambient faces.*

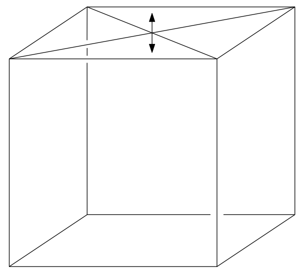


Figure 3. Convex triangulated sphere with a flat vertex. Moving this vertex in the direction orthogonal to the face produces a nontrivial infinitesimal flex.

The Hilbert–Einstein function. The proof of Theorem 1.7 is based on some striking properties of the discrete Hilbert–Einstein function, also known in the physics community as the Regge function [1961]. First we have to define a space of deformations of a triangulated polyhedron.

Definition 1.12. Let T be a triangulation of a polyhedron P, and let $e_1, \ldots, e_n$ be the interior edges of T. We denote by $\mathcal{D}_{P,T}$ the space of n-tuples $(l_1, \ldots, l_n) \in \mathbb{R}^n_{>0}$ such that for every simplex σ of T, replacing the lengths of the edges of σ that are interior edges of T by the corresponding l_j produces a nondegenerate simplex.

For every element $l \in \mathcal{D}_{P,T}$, there is an associated metric structure on P obtained by gluing the simplices with changed edge lengths. The resulting metric space is locally Euclidean except that it has cone singularities along the interior edges of T. For every $i \in \{1, \ldots, n\}$, denote by ω_i the total angle around e_i and by $\kappa_i := 2\pi - \omega_i$ the singular curvature along e_i. Let $e'_1, \ldots, e'_r$ be the boundary edges of P; for every $j \in \{1, \ldots, r\}$ denote by α_j the dihedral angle of P at e'_j, and by l'_j the length of e'_j.

Definition 1.13. The *Hilbert–Einstein function* on $\mathcal{D}_{P,T}$ is given by the formula

$$\mathcal{S}(l) := \sum_{i=1}^{n} l_i \kappa_i + \sum_{j=1}^{r} l'_j (\pi - \alpha_j) .$$

The Schläfli formula. A key tool in polyhedral geometry, this formula has several generalizations. The 3-dimensional Euclidean version states simply that, under a first-order deformation of any Euclidean polyhedron,

$$\sum_e l_e d\alpha_e = 0 , \tag{1}$$

where the sum is taken over all edges e, with l_e denoting the length of the edge e, and α_e the dihedral angle at e. This equality is also known as the Regge formula.

It follows directly from the Schläfli formula that, under any first-order variation of the lengths of the interior edges of a triangulation T of the polyhedron P — that is, for any vector tangent to $\mathcal{D}_{P,T}$ — the first-order variation of $\mathcal{S}$ is simply

$$d\mathcal{S} = \sum_{i=1}^{n} \kappa_i dl_i \ . \tag{2}$$

Therefore, the Hessian of $\mathcal{S}$ equals the Jacobian of the map $(l_i)_{i=1}^n \mapsto (\kappa_i)_{i=1}^n$.

Definition 1.14. Let T be a triangulation of a polyhedron P with n interior edges. Define the $n \times n$ matrix M_T as

$$M_T = \left(\frac{\partial \omega_i}{\partial l_j}\right) = -\left(\frac{\partial^2 \mathcal{S}}{\partial l_i \partial l_j}\right).$$

The derivatives are taken at the point $l \in \mathcal{D}_{P,T}$ that corresponds to the actual edge lengths in T.

The arguments in this paper use only M_T, and not directly the Hilbert–Einstein function $\mathcal{S}$. The fact that M_T is minus the Hessian of $\mathcal{S}$ does imply, however, that M_T is symmetric.

The matrix M_T is directly related to the infinitesimal rigidity of P, an idea that, in the smooth rather than the polyhedral context, goes back to Blaschke and Herglotz.[1]

Lemma 1.15. *Let T be a triangulation of a polyhedron P without interior vertices. Then P is infinitesimally rigid if and only if M_T is nondegenerate.*

The proof can be found in [Bobenko and Izmestiev 2008; Schlenker 2009] and is based on the observation that an isometric deformation of P induces a first-order variation of the interior edge lengths but a zero variation of the angles around them.

The second-order behavior of $\mathcal{S}$. The following is the key technical statement of the paper.

Theorem 1.16. *Let P be a convex polyhedron, and let T be a triangulation of P with* $\mathrm{Vert}(T) = \mathrm{Vert}(P)$. *Then M_T is positive definite.*

Theorem 1.16 is actually a special case of the following theorem that describes the signature of M_T for T *any* triangulation of P.

Theorem 1.17. *Let P be a convex polyhedron, and let T be a triangulation of P with m interior and k flat vertices. Then the dimension of the kernel of M_T is $3m+k$, and M_T has m negative eigenvalues.*

[1] Blaschke and Herglotz suggested that the critical points of the Hilbert–Einstein function on a manifold with boundary (in the smooth case), with fixed boundary metric, correspond to Einstein metrics, that is, to constant curvature metrics in dimension 3. The analog of M_T in this context is the Hessian of the Hilbert–Einstein function.

From Theorem 1.16 to Theorem 1.7. Let P be a polyhedron satisfies the assumptions of Theorem 1.7. Since P is decomposable and weakly codecomposable, there exists a convex polyhedron Q such that all vertices of P are vertices of Q, and a triangulation $\overline{T}$ of Q that contains a triangulation T of P and whose vertices are only the vertices of Q. It is easy to see that the matrix M_T is then a principal minor of the matrix $M_{\overline{T}}$. By Theorem 1.16, $M_{\overline{T}}$ is positive definite; thus so is M_T. In particular, M_T is nondegenerate. Lemma 1.15 implies that the polyhedron P is infinitesimally rigid.

Since Theorem 1.16 is a special case of Theorem 1.17, the rest of this paper deals with proving Theorem 1.17.

Plan 1.18 (for proving Theorem 1.17). The proof is based on a standard procedure. To show that the matrix M_T has the desired property for every triangulation T, we prove three points:

- any two triangulations can be connected by a sequence of moves;
- the moves don't affect the desired property;
- the property holds for a special triangulation.

These points are dealt with in the given order in the next three sections.

2. Connectedness of the set of triangulations

Moves on simplicial complexes are well studied; see [Lickorish 1999] for an overview. Several theorems state that any two triangulations of a given manifold can be connected by certain kinds of simplicial moves. However, we are in a different situation here, since we deal with triangulations of a fixed geometric object. Taking a closer look, one sees that a simplicial move is defined as a geometric move preceded and followed by a simplicial isomorphism. Performing an isomorphism is the possibility that is missing in our case.

To emphasize the difference between the combinatorial and the geometric situation, let us cite a negative result concerning geometric moves. Santos [2005] exhibited two triangulations with the same set of vertices in $\mathbb{R}^5$ that cannot be connected via $2 \leftrightarrow 5$ and $3 \leftrightarrow 4$ bistellar moves. For an overview on geometric bistellar moves, see [Santos 2006].

Geometric stellar moves: the Morelli–Włodarczyk theorem. Morelli [1996] and Włodarczyk [1997] obtained a positive result on geometric simplicial moves. As a crucial step in the proof of the weak Oda conjecture, they showed that any two triangulations of a convex polyhedron can be connected by a sequence of *geometric stellar moves*.

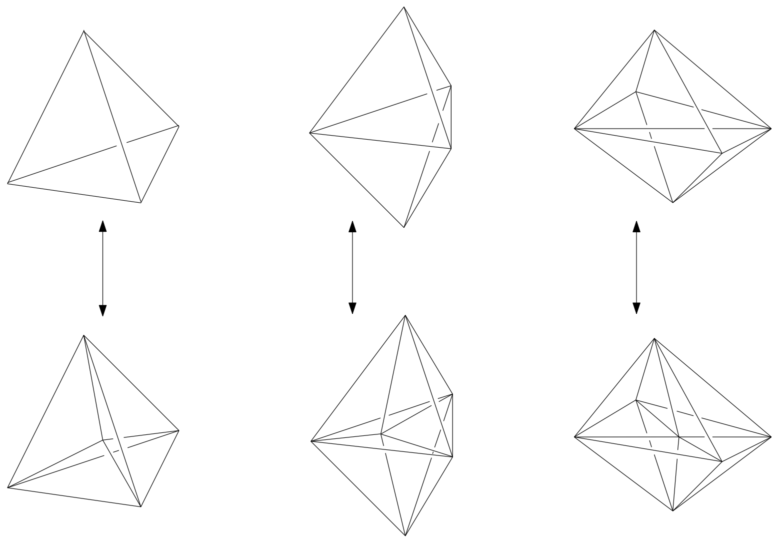


Figure 4. Interior stellar moves in dimension 3.

Definition 2.1. Let p be an interior point of a simplex $\sigma \subset \mathbb{R}^n$. The *starring* of σ at p is an operation that replaces σ by the cone with the apex p over the boundary of σ.

Let T be a triangulation of a subset of $\mathbb{R}^n$, let σ be a simplex of T, and let p be a point in the relative interior of σ. The operation of *starring* of T at p consists of replacing the star $\operatorname{st}\sigma$ of σ by the cone with apex p over the boundary of $\operatorname{st}\sigma$. The operation inverse to starring is called *welding*.

Starrings and weldings are called *stellar moves*.

See Figures 4 and 5 for stellar moves in dimension 3. Figure 4 depicts starrings and weldings at interior points of T, while Figure 5 shows starrings and weldings at boundary points. In the case of a boundary point our definition is not completely correct: A starring replaces $\operatorname{st}\sigma$ by the cone over $\partial \operatorname{st}\sigma \setminus \partial T$.

Theorem 2.2 [Morelli 1996; Włodarczyk 1997]. *Any two triangulations of a convex polyhedron $P \subset \mathbb{R}^n$ can be connected by a sequence of geometric stellar moves.*

We outline of Morelli's proof using more elementary language and tools.

Outline of the proof. Let T and T' be two triangulations of P. A triangulation Σ of $P \times [0, 1]$ with $\Sigma_{P\times\{1\}} = T$ and $\Sigma|_{P\times\{0\}} = T'$ is called a *simplicial cobordism* between T and T'.

Definition 2.3. Let pr denote the orthogonal projection $P \times [0, 1] \to P$. A simplex $\sigma \in \Sigma$ is called a *circuit* if $\dim \operatorname{pr}(\sigma) < \dim \sigma$ and σ is inclusion minimal with this property.

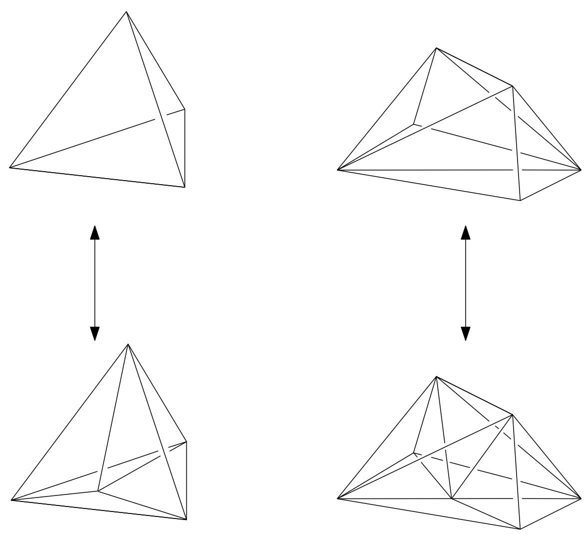


Figure 5. Boundary stellar moves in dimension 3.

Clearly, the stars of the circuits are simplicial balls with no vertical faces and $\Sigma = \bigcup_{\text{circuits } \sigma} \operatorname{st} \sigma$ with disjoint interiors.

Definition 2.4. We call a simplicial cobordism Σ *collapsible* if there is a sequence of triangulations $\Sigma = \Sigma_0, \Sigma_1, \dots, \Sigma_N = \Sigma_{P \times \{0\}}$ such that

- $\Sigma_{i+1} = \Sigma_i \setminus \operatorname{st} \sigma_i$ for a circuit σ_i;
- the upper boundary of Σ_i projects one-to-one on P for every i.

In other words, Σ is collapsible if it can be "dismantled with a crane".

Lemma 2.5. *The triangulation* $\operatorname{pr}(\partial^+ \Sigma_{i+1})$ *can be obtained from* $\operatorname{pr}(\partial^+ \Sigma_i)$ *by a starring with a subsequent welding. Here* ∂^+ *denotes the upper boundary.*

Proof. For every circuit σ, the transformation $\operatorname{pr}(\partial^+ \sigma) \rightsquigarrow \operatorname{pr}(\partial^- \sigma)$ is a bistellar move and can be realized by a starring and a welding. These extend to a starring and a welding in $\operatorname{st} \sigma$. □

Thus, a collapsible simplicial cobordism between two triangulations gives rise to a sequence of stellar moves joining the triangulations.

Definition 2.6. We call a triangulation Σ *coherent* if there is a function $h : |\Sigma| \to \mathbb{R}$ that is piecewise linear with respect to Σ and strictly convex across every facet of Σ. (Here $|\Sigma| = \bigcup_{\sigma \in \Sigma} \sigma$ is the support of Σ.)

The barycentric subdivision of any convex polytope Q is coherent — one can choose values of h at the barycenters of faces of Q consecutively, along with increasing dimension. Each time the value must be sufficiently large.

Lemma 2.7. *A coherent simplicial cobordism is collapsible.*

Proof. Let σ and σ' be two circuits of Σ such that some point of $\operatorname{st}\sigma$ lies directly above a point of $\operatorname{st}\sigma'$. It follows that $\partial h/\partial x_{n+1}|_{\sigma} > \partial h/\partial x_{n+1}|_{\sigma'}$, where $\partial/\partial x_{n+1}$ denotes the derivative in the vertical direction. Thus the stars of the circuits can be lifted up in nondecreasing order of the vertical derivative of h on the circuits. □

To prove the theorem, we construct a coherent cobordism between stellar subdivisions of T and T'. (By a stellar subdivision, we mean the result of a sequence of starrings.)

Lemma 2.8. *Let Σ and Σ' be two triangulations with the same support. Then Σ can be stellarly subdivided to a triangulation Σ'' that refines Σ'.*

The reader can find a proof of this classical statement in [Glaser 1970].

Lemma 2.9. *Let Σ be an arbitrary triangulation of a convex polytope. Then Σ can be stellarly subdivided to a coherent triangulation.*

Proof. By Lemma 2.8, the barycentric triangulation of the convex polytope $|\Sigma|$ can be stellarly subdivided to a triangulation Σ' that refines Σ. But the barycentric subdivision of any polytope is coherent. Since starring a coherent triangulation produces a coherent triangulation, Σ' is coherent.

Again by Lemma 2.8, the triangulation Σ can be stellarly subdivided to a triangulation Σ'' that refines Σ'. We claim that Σ'' is coherent.

Let us show that there exists a function $s'' : |\Sigma| \to \mathbb{R}$ that is piecewise linear with respect to Σ'' and strictly convex across all facets of Σ'' that are not contained in the codimension 1 skeleton of Σ. We construct s'' by induction on the number of stellar subdivisions that transform Σ to Σ''. As the induction base, we take the zero function. When a stellar subdivision is done, we redefine the function at the center of the subdivision by increasing it a little. Then we get strict convexity across appearing facets and don't destroy convexity where it already takes place. Note that s'' can be concave across facets of Σ'' that are contained in facets of Σ.

Now, since Σ' is coherent, there exists a function h' piecewise linear and strictly convex with respect to Σ'. Since Σ' refines Σ, the function h' is strictly convex across the codimension 1 skeleton of Σ. Therefore the function $h'' = h' + \epsilon s''$ is strictly convex on Σ'' for a sufficiently small positive ϵ. □

Outline of the proof of Theorem 2.2. Applying Lemma 2.9 to an arbitrary simplicial cobordism Σ between T and T', we get a coherent simplicial cobordism Σ''. Since $\Sigma''|_{P\times 1}$ and $\Sigma''|_{P\times 0}$ are stellar subdivisions of T and T' respectively, this yields a sequence of stellar moves connecting T and T'.

An alternative way to derive Theorem 2.2 from Lemma 2.9 was suggested to us by Francisco Santos and is as follows. By Lemma 2.9, the triangulations T and T' can be stellarly subdivided to coherent triangulations S and S', respectively. Let

$h : P \times \{1\} \to \mathbb{R}$ and $h' : P \times \{0\} \to \mathbb{R}$ be corresponding convex piecewise linear functions. Then their lower envelope $\widetilde{h} : P \times [0, 1]$ is a convex function whose linearity domains determine a polyhedral subdivision of $P \times [0, 1]$. If h and h' are in general position, then this subdivision is a coherent triangulation, and thus a collapsible simplicial cobordism between T and T'. □

Realizing interior stellar moves by bistellar moves. To simplify our task in the next section, we show that instead of interior stellar moves one can use *bistellar* or *Pachner moves* and continuous displacements of the vertices of the triangulation.

Definition 2.10. Let T be a triangulation of a subset of $\mathbb{R}^3$.

- Let σ be a 3-dimensional simplex of T. A $1 \to 4$ Pachner move replaces σ by four smaller simplices sharing a vertex that is an interior point of σ.
- Let σ and τ be two 3-simplices of T such that the union $\sigma \cup \tau$ is a strictly convex bipyramid. A $2 \to 3$ Pachner move replaces σ and τ by three simplices sharing the edge that joins the opposite vertices of σ and τ.
- A $3 \to 2$ Pachner move is the inverse of a $2 \to 3$ Pachner move.
- A $4 \to 1$ Pachner move is the inverse of a $1 \to 4$ Pachner move.

The Pachner moves are depicted on Figure 6.

Lemma 2.11. *Any two triangulations of a convex polyhedron P can be connected by a sequence of Pachner moves, boundary stellar moves and continuous displacements of the interior vertices.*

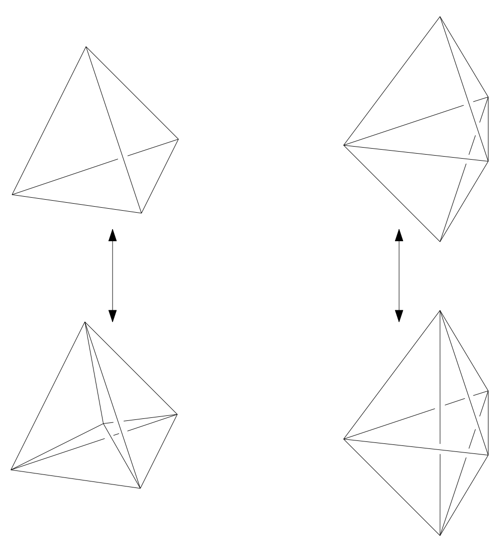


Figure 6. The $1 \leftrightarrow 4$ and $2 \leftrightarrow 3$ Pachner moves.

Proof. By Theorem 2.2, it suffices to show that every interior stellar move can be realized as a sequence of Pachner moves and vertex displacements. Since Pachner moves are invertible, we realize only interior starrings.

The starring in a 3-simplex is a $1 \to 4$ Pachner move.

Consider the starring in a triangle (middle of Figure 4). Denote the vertices of the triangle to be starred by $1, 2, 3$, and the two remaining vertices by a and b. Perform a $1 \to 4$ move on the tetrahedron $a123$ and denote the new vertex by p. Then perform a $2 \to 3$ move on the tetrahedra $p123$ and $b123$. Finally move the vertex p so that it lies in the triangle 123.

To realize a starring of an edge, we also first perform a sequence of Pachner moves to obtain a triangulation combinatorially equivalent to the starring, and then move the new vertex. Denote by a and b the vertices of the edge to be starred, and denote the vertices in the link of the edge ab by $1, 2, \ldots, n$ in the cyclic order. Perform a $1 \to 4$ move on the tetrahedron $ab1n$. The new vertex p should be chosen so that the plane of the triangle abp does not pass through any other vertex. Let $(k, k+1)$ be the edge intersected by this plane. Perform a $2 \to 3$ move on the tetrahedra $ab1p$ and $ab12$, then a $2 \to 3$ move on the tetrahedra $ab2p$ and $ab23$, and so on. This sequence finishes with a $2 \to 3$ move on $ab(k-1)p$ and $ab(k-1)k$. After that apply a similar sequence of $2 \to 3$ moves on the other side starting with the tetrahedra $abnp$ and $abn(n-1)$ and finishing with $ab(k+2)p$ and $ab(k+2)(k+1)$. Finally perform a $3 \to 2$ move over the tetrahedra $abpk$, $abp(k+1)$ and $abk(k+1)$. It remains to move the vertex p so that it lies on the edge ab. □

3. The effect of the elementary moves on the signature of M_T

In this section we realize the second point of Plan 1.18. Namely, we show that if Theorem 1.17 holds for some triangulation T, then it holds for a triangulation T' that is obtained from T by an elementary move. An elementary move is either a Pachner move or a boundary stellar move or a continuous displacement of the interior vertices of T.

The rank of the matrix M_T. Here we prove a part of Theorem 1.17:

Lemma 3.1. *The corank of the matrix M_T equals $3m + k$, with m the number of interior vertices and k the number of flat boundary vertices in the triangulation T*:

$$\dim \ker M_T = 3m + k.$$

Proof. If $m > 0$ or $k > 0$, then it is easy to find a whole bunch of vectors in the kernel of M_T. Any continuous displacement of the interior vertices of T changes the lengths of the interior edges, but doesn't change the angles around them, which stay equal to 2π. Similarly, moving a flat boundary vertex in the direction orthogonal to its ambient face doesn't change any of the angles ω_i. It does change the lengths of the boundary edges incident to this vertex, but only in the second order. It follows that the variations of interior edge lengths induced by the orthogonal displacement of a flat boundary vertex belong to the kernel of M_T.

Being formal, let $Q : \mathcal{V}(T) \to \mathbb{R}^3$ be an assignment to every vertex p_i of T of a vector q_i such that

(1) $q_i = 0$ if p_i is a nonflat boundary vertex of T;

(2) $q_i \perp F_i$ if p_i is a flat boundary vertex lying in the face F_i of P.

For every edge ij of T put

$$\ell_{ij}^Q = \frac{p_i - p_j}{\|p_i - p_j\|} \cdot (q_i - q_j).$$

It is easy to see that this formula gives the infinitesimal change of ℓ_{ij} that results from the infinitesimal displacements of the vertices p_i, p_j by the vectors q_i, q_j. By the previous paragraph, $\ell_{ij}^Q \in \ker M_T$.

Let us show that the span of the vectors ℓ^Q has dimension $3m+k$. The correspondence between Q and ℓ^Q is linear, and the space of assignments Q with properties (1) and (2) has dimension $3m+k$, so it suffices to show that $\ell^Q = 0$ implies $Q = 0$. Indeed, $\ell^Q = 0$ means that Q is an infinitesimal flex of the 1-skeleton of T; see Definition 1.9. But T is infinitesimally rigid, since every simplex is. Thus $\ell^Q = 0$ implies that Q is trivial. Since $q_i = 0$ on the vertices of P, we have $Q = 0$.

It remains to show that any vector $\dot{\ell} \in \ker M_T$ has the form ℓ^Q for some Q. Let $p_1 p_2 p_3$ be a triangle of T. Choose q_1, q_2, and q_3 arbitrarily. Let p_4 be a vertex such that there is a simplex $p_1 p_2 p_3 p_4$ in T. The values of $\dot{\ell}_{i4}$ for $i = 1, 2, 3$ determine uniquely a vector q_4 such that $\dot{\ell}_{i4} = \ell_{i4}^Q$ for $i = 1, 2, 3$. If ij is a boundary edge of T, we put $\dot{\ell}_{ij} = 0$. Similarly, we define q_5 for the vertex p_5 of a simplex that shares a face with $p_1 p_2 p_3 p_4$. Proceeding in this manner, we can assign a vector q_i to every vertex p_i if we show that this is well-defined (we extend our assignment along paths in the dual graph of T, and it needs to be shown that the extension does not depend on the choice of a path). It is not hard to see that this is ensured by the property $M_T \dot{\ell} = 0$. Thus we have constructed an assignment $Q : \mathcal{V}(T) \to \mathbb{R}^3$ such that $\dot{\ell} = \ell^Q$. Since $\dot{\ell}_{ij} = 0$ for every boundary edge ij of T, the vectors $(q_i)|_{p_i \in \partial P}$ define an infinitesimal flex of the boundary of P. Due to Theorem 1.11, Q satisfies properties (1) and (2) above, after subtracting an infinitesimal motion. Thus the kernel of M_T consists of the vectors of the form ℓ^Q. □

Corollary 3.2. *Let T be a triangulation of a convex polyhedron P. Consider a continuous displacement of the vertices of T such that no simplex of the triangulation degenerates, the underlying space of T remains a convex polyhedron, all flat* *boundary vertices remain flat, and nonflat remain nonflat. Then the signature of the matrix M_T does not change during this deformation.*

Proof. Due to Lemma 3.1, the rank of M_T does not change during the deformation. Hence no eigenvalue changes its sign. □

The effect of the Pachner moves.

Lemma 3.3. *Let P be a convex polyhedron, and let T and T' be two triangulations of P such that T' is obtained from T by a $2 \to 3$ Pachner move. Then the statement of Theorem 1.17 applies to T if and only if it applies to T'.*

Proof. Since triangulations T and T' have the same number of interior and flat boundary vertices, the matrices M_T and $M_{T'}$ have the same corank by Lemma 3.1. It remains to show that M_T and $M_{T'}$ have the same number of negative eigenvalues.

Matrices M_T and $M_{T'}$ define symmetric bilinear forms (that are denoted by the same letters) on the spaces $\mathbb{R}^{\mathscr{E}_{\text{int}}(T)}$ and $\mathbb{R}^{\mathscr{E}_{\text{int}}(T')}$, respectively. Here $\mathscr{E}_{\text{int}}(T)$ denotes the set of interior edges of the triangulation T. Note that $\mathscr{E}_{\text{int}}(T') = \mathscr{E}_{\text{int}}(T) \cup \{e_0\}$, where e_0 is the vertical edge on the lower right of Figure 6. Extend M_T to a symmetric bilinear form on $\mathbb{R}^{\mathscr{E}_{\text{int}}(T')}$ by augmenting the matrix M_T with a zero row and a zero column, and put $\Phi = M_{T'} - M_T$. By Definition 1.14, we have

$$\Phi = \left(\frac{\partial(\omega_i' - \omega_i)}{\partial \ell_j} \right)_{i,j \in \mathscr{E}_{\text{int}}(T')},$$

where we put $\partial\omega_0/\partial\ell_j = 0$ for all j.

Denote those edges on the upper right of Figure 6 that are interior edges of T by $e_1, \dots, e_s$. Note that $\omega_i = \omega_i'$ as functions of the edge lengths for all $i \notin \{0, \dots, s\}$. Thus, the matrix Φ reduces to an $(s+1) \times (s+1)$ matrix with rows corresponding to the edges $e_0, \dots, e_s$.

We claim that the matrix Φ is positive semidefinite of rank 1. To construct a vector in the kernel of Φ, note that during any continuous deformation of the bipyramid on Figure 6 we have $\omega_i = \omega_i'$ as functions of edge lengths for $i = 1, \dots, s$, while ω_0' is identically 2π. Thus if we choose $\dot\ell_1, \dots, \dot\ell_s$ arbitrarily and define $\dot\ell_0$ as the infinitesimal change of the length of e_0 under the corresponding infinitesimal deformation of the bipyramid, then we have $\Phi\dot\ell = 0$. Therefore $\operatorname{rank}\Phi \le 1$. The infinitesimal rigidity of the bipyramid implies $\partial\omega_0'/\partial\ell_0 \ne 0$; thus $\operatorname{rank}\Phi = 1$. Since the space of convex bipyramids is connected, it suffices to prove the positive semidefiniteness of Φ in some special case. In the case when all edges of the bipyramid have equal length, one can easily see that $\partial\omega_0'/\partial\ell_0 > 0$, which implies the positivity of the unique eigenvalue of Φ.

The equation

$$\operatorname{rank} M_{T'} = \operatorname{rank} M_T + 1 = \operatorname{rank} M_T + \operatorname{rank}\Phi$$

implies that $\ker M_T$ and $\ker \Phi$ intersect transversally and $\ker M_{T'} = \ker M_T \cap \ker \Phi$. Therefore

$$\operatorname{rank}(M_T + t\Phi) = \operatorname{rank} M_T + 1 \quad \text{for all } t \neq 0.$$

The Courant minimax principle [Courant and Hilbert 1953, Chapter I, Section 4] implies that the eigenvalues of $M_T + \epsilon\Phi$ are larger than or equal to the corresponding eigenvalues of M_T. It follows that when M_T is deformed into $M_{T'}$ via $\{M_T + t\Phi\}_{t\in[0,1]}$, exactly one of the zero eigenvalues of M_T becomes positive, and all of the nonzero eigenvalues preserve their sign. Thus $M_{T'}$ has the same number of negative eigenvalues as M_T and the lemma is proved. □

Lemma 3.4. *Let P be a convex polyhedron, and let T and T' be two triangulations of P such that T' is obtained from T by a $1 \to 4$ Pachner move. Then the statement of Theorem 1.17 applies to T if and only if it applies to T'.*

Proof. The same arguments as in the proof of Lemma 3.3 work. The triangulation T' has one interior vertex more that the triangulation T and four interior edges more than T. By Lemma 3.1, we have $\operatorname{rank} M_{T'} = \operatorname{rank} M_T + 1$, and we have to prove that $M_{T'}$ has the same number of positive eigenvalues as M_T and one negative eigenvalue more. For this it suffices to show that the quadratic form $\Phi = M_{T'} - M_T$ is negative semidefinite of rank 1. In the same way as in the proof of Lemma 3.3, one shows that $\operatorname{rank}\Phi \leq 1$. After that, it suffices to show that the restriction of Φ to the space spanned by the variations of lengths of the four interior edges on the lower left of Figure 6 is nontrivial and negative semidefinite. The nontriviality follows from the infinitesimal rigidity of the simplex, and it suffices to check the negative semidefiniteness in some convenient special case. □

The effect of the boundary stellar moves.

Lemma 3.5. *Let P be a convex polyhedron, and let T and T' be two triangulations of P such that T' is obtained from T by the starring of a boundary 2-simplex. Then the statement of Theorem 1.17 applies to T if and only if it applies to T'.*

Proof. We have $\operatorname{rank} M_{T'} = \operatorname{rank} M_T$ and need to show that $M_{T'}$ has the same signature as M_T. This is true because in fact $M_{T'} = M_T$ — more precisely, $M_{T'}$ is obtained from M_T by adding a column and a row, each with all elements equal to zero. This can be shown using the explicit formulas for $\partial\omega_i/\partial\ell_j$ and $\partial\omega_i'/\partial\ell_j$ from [Bobenko and Izmestiev 2008, Section 3.1] and [Korepanov 2000]. □

Lemma 3.6. *Let P be a convex polyhedron, and let T and T' be two triangulations of P such that T' is obtained from T by the starring of a boundary 1-simplex. Then the statement of Theorem 1.17 applies to T if and only if it applies to T'.*

Proof. The strategy is the same as in the proofs of Lemmas 3.3 and 3.4. Put $\Phi = M_{T'} - M_T$ and note that by Lemma 3.1

$$\operatorname{rank} M_{T'} = \operatorname{rank} M_T + i, \tag{3}$$

where i is one less than the number of simplices incident to the starred edge (for example, $i = 2$ in the right column of Figure 5). As in the proof of Lemma 3.3, one shows that $\operatorname{rank} \Phi \le i$. Then (3) implies $\operatorname{rank} \Phi = i$. Since we aim to show that $M_{T'}$ has the same number of negative eigenvalues as M_T, it suffices to show that Φ is positively semidefinite.

Let Ψ be the $i \times i$ principal minor of Φ formed by the rows and columns that correspond to the interior edges of the triangulation on the lower right of Figure 5. We claim that Ψ is positively definite, which implies the nonnegativity of Φ. The proof is by continuity argument as in Lemma 3.3. To prove the nondegeneracy of Ψ, it suffices to show that the framework of the boundary edges on the lower right of Figure 5 is infinitesimally rigid. The framework on the upper right of Figure 5 is infinitesimally rigid, since it is formed by skeleta of 3-simplices that are rigid. This implies the infinitesimal rigidity of the boundary framework on the lower right (an easy exercise in applying the definition of an infinitesimal flex). Now consider a deformation of the triangulation on the upper right that makes the underlying polyhedron convex. This deformation can be extended to a deformation of the triangulation on the lower right. Since the matrix Ψ remains nondegenerate during the deformation, its signature is preserved. After the polyhedron is made convex, push the starring vertex off the starred edge so that the vertices of the triangulation are in the convex position. This also preserves the signature of Ψ. In the final position, Ψ is positive due to Theorem 4.1. □

4. Investigating M_T for a special triangulation T.

Let P be a convex polyhedron. Let S be a triangulation of ∂P such that $\operatorname{Vert}(S) = \operatorname{Vert}(P)$, and let p be a vertex of P. Consider the triangulation T consisting of simplices with a common vertex p and opposite faces the triangles of S disjoint from p.

Theorem 4.1. *The matrix M_T is positive definite.*

Proof. Formally, this is a special case of [Schlenker 2009, Theorem 1.5] that claims that M_T is positive if P is weakly convex and star-shaped with respect to the vertex p. The proof uses the positivity of the corresponding matrix for convex caps [Izmestiev 2008, Lemma 6 and Theorem 5] and the projective invariance of infinitesimal rigidity [Schlenker 2009, Section 5]. □

Theorem 4.1 accomplishes the plan outlined in Plan 1.18. Theorem 1.17 is proved, and therewith Theorem 1.7.

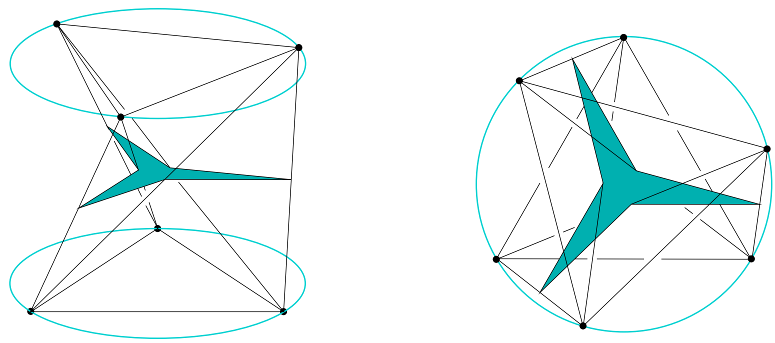


Figure 7. The twisted octahedron.

Appendix: A polyhedron that is not weakly codecomposable

Definition A.1. Let $\theta \in (-2\pi/3, 2\pi/3)$. The twisted octahedron Oct_θ of Figure 7 is the polyhedron with vertices A, B, C, A', B', C' of coordinates respectively

$$(1, 0, 1), \quad (\cos(2\pi/3), \sin(2\pi/3), 1),$$
$$(\cos(4\pi/3), \sin(4\pi/3), 1), \quad (\cos(-\pi+\theta), \sin(-\pi+\theta), -1),$$
$$(\cos(-\pi/3+\theta), \sin(-\pi/3+\theta), -1), \quad (\cos(\pi/3+\theta), \sin(\pi/3+\theta), -1).$$

The edges are the segments joining A to B' and C', B to A' and C', C to A' and B', and the faces are the triangles (ABC), $(A'B'C')$, $(AB'C')$, $(A'BC')$, $(A'B'C)$, (ABC'), $(AB'C)$, $(A'BC)$.

Note that $\mathrm{Oct}_{\pm\pi/2}$ is a Schönhardt polyhedron; see Figure 2, right.

Proposition A.2. Oct_θ *is embedded for all* $\theta \in (-2\pi/3, 2\pi/3)$.

For $\theta \in (-2\pi/3, 2\pi/3)$, we call $A_t(\theta)$ the area of the intersection of Oct_θ with the horizontal plane $\{z = t\}$.

Proposition A.3. $\lim_{\theta\to 2\pi/3} A_0(\theta) = 0$.

Let K be a large enough convex polygon in the plane Oxy (it suffices to require that the interior of K contains the disk $x^2 + y^2 \le 1$). Consider the polyhedron $P_\theta = \mathrm{conv}(A, B, C, A', B', C', K) \setminus \mathrm{Oct}_\theta$ homeomorphic to a solid torus.

Lemma A.4. *For* θ *close enough to* $2\pi/3$, P_θ *is not weakly codecomposable.*

Proof. Suppose that P_θ is weakly codecomposable. Then there exists a convex polyhedron $Q_\theta \supset P_\theta$ such that $Q_\theta \setminus P_\theta$ can be triangulated without an interior vertex. Let $S_1, \dots, S_n$ be the simplices in this triangulation that intersect $\mathrm{Oct}_\theta \cap (Oxy)$. For each $i \in \{1, \dots, n\}$, let $a_i(t)$ be the area of the intersection of S_i with the horizontal plane $\{z = t\}$.

Each of the S_i can have either:

- Two vertices with $z \geq 1$ and two vertices with $z \leq -1$. Then the restriction of a_i to $(-1, 1)$ is a concave quadratic function, so that $2a_i(0) \geq a_i(-1) + a_i(1)$.
- One vertex with $z \geq 1$ and three vertices with $z \leq -1$. Then a_i is of the form $a_i(t) = c_i(t + b_i)^2$ with $b_i \geq 1$. It easily implies that $4a_i(0) \geq a_i(-1) + a_i(1)$.
- One vertex with $z \leq -1$ and three vertices with $z \geq 1$. The same argument then shows the same result.

So $4a_i(0) \geq a_i(-1) + a_i(1)$ for all i and the union of the S_i contains Oct_θ. It follows that $4A_0(\theta) \geq A_{-1}(\theta) + A_1(\theta)$. But $A_1(\theta)$ and $A_{-1}(\theta)$ are equal to the area of an equilateral triangle of fixed side length, while $A_0(\theta)$ goes to 0 as $\theta \to 2\pi/3$. This is a contradiction, and the claim follows. □

Acknowledgements

We are grateful to Bob Connelly, Günter Rote and Igor Pak for useful conversations and remarks concerning the subject considered here. We thank Francisco Santos for pointing out some mistakes in the proof of Lemma 2.9 and helping to correct them.

References

[Aichholzer et al. 2002] O. Aichholzer, L. S. Alboul, and F. Hurtado, "On flips in polyhedral surfaces", *Internat. J. Found. Comput. Sci.* **13**:2 (2002), 303–311. MR 1895784 Zbl 1066.68136

[Alexandrov 2005] A. D. Alexandrov, *Convex polyhedra*, Springer, Berlin, 2005. MR 2005j:52002 Zbl 1067.52011

[Bobenko and Izmestiev 2008] A. I. Bobenko and I. Izmestiev, "Alexandrov's theorem, weighted Delaunay triangulations, and mixed volumes", *Ann. Inst. Fourier* (*Grenoble*) **58**:2 (2008), 447–505. MR 2009j:52016 Zbl 1154.52005

[Cauchy 1813] A. L. Cauchy, "Sur les polygones et polyèdres, second mémoire", *Journal de l'Ecole Polytechnique* **19** (1813), 87–98.

[Connelly and Schlenker 2010] R. Connelly and J.-M. Schlenker, "On the infinitesimal rigidity of weakly convex polyhedra", *European J. Combin.* **31**:4 (2010), 1080–1090. MR 2607252 Zbl 1191.52022

[Courant and Hilbert 1953] R. Courant and D. Hilbert, *Methods of mathematical physics*, vol. 1, Interscience, New York, 1953. MR 16,426a Zbl 0053.02805

[Dehn 1916] M. Dehn, "Über die Starrheit konvexer Polyeder", *Math. Ann.* **77**:4 (1916), 466–473. MR 1511873 JFM 46.1115.01

[Glaser 1970] L. C. Glaser, *Geometrical combinatorial topology*, vol. 1, Mathematical Studies **27**, Van Nostrand Reinhold, New York, 1970.

[Gluck 1975] H. Gluck, "Almost all simply connected closed surfaces are rigid", pp. 225–239 in *Geometric topology* (Park City, UT), edited by L. C. Glaser and T. B. Rushing, Lecture Notes in Math. **438**, Springer, Berlin, 1975. MR 53 #4074 Zbl 0315.50002

[Izmestiev 2008] I. Izmestiev, "A variational proof of Alexandrov's convex cap theorem", *Discrete Comput. Geom.* **40**:4 (2008), 561–585. MR 2009j:52004 Zbl 1162.52004

[Jessen 1967] B. Jessen, "Orthogonal icosahedra", *Nordisk Mat. Tidskr* **15** (1967), 90–96. MR 37 #2083 Zbl 0156.41001

[Korepanov 2000] I. G. Korepanov, "A formula with volumes of five tetrahedra and discrete curvature", preprint, 2000. arXiv nlin/0003001

[Legendre 1794] A.-M. Legendre, *Eléments de géométrie*, Firmin Didot, Paris, 1794.

[Lickorish 1999] W. B. R. Lickorish, "Simplicial moves on complexes and manifolds", pp. 299–320 in *Proceedings of the Kirbyfest* (Berkeley, CA, 1998), Geom. Topol. Monogr. **2**, Geom. Topol. Publ., Coventry, 1999. MR 2000j:57057 Zbl 0963.57013

[Morelli 1996] R. Morelli, "The birational geometry of toric varieties", *J. Algebraic Geom.* **5**:4 (1996), 751–782. MR 99b:14056 Zbl 0871.14041

[Regge 1961] T. Regge, "General relativity without coordinates", *Nuovo Cimento* (10) **19** (1961), 558–571. MR 23 #B418

[Santos 2005] F. Santos, "Non-connected toric Hilbert schemes", *Math. Ann.* **332**:3 (2005), 645–665. MR 2006g:52021 Zbl 1078.52010

[Santos 2006] F. Santos, "Geometric bistellar flips: The setting, the context and a construction", pp. 931–962 in *International Congress of Mathematicians*, vol. 3, Eur. Math. Soc., Zürich, 2006. MR 2008a:52011 Zbl 1107.52007

[Schlenker 2005] J.-M. Schlenker, "A rigidity criterion for nonconvex polyhedra", *Discrete Comput. Geom.* **33**:2 (2005), 207–221. MR 2005k:52042 Zbl 1083.52006

[Schlenker 2009] J.-M. Schlenker, "On weakly convex star-shaped polyhedra", *Discrete Math.* **309** (2009), 6139–6145. MR 2552649 Zbl 1193.52013

[Schönhardt 1928] E. Schönhardt, "Über die Zerlegung von Dreieckspolyedern in Tetraeder", *Math. Ann.* **98**:1 (1928), 309–312. MR 1512408

[Weisstein] E. W. Weisstein, "Jessen's Orthogonal Icosahedron", MathWorld — A Wolfram Web Resource, available at http://mathworld.wolfram.com/JessensOrthogonalIcosahedron.html.

[Włodarczyk 1997] J. Włodarczyk, "Decomposition of birational toric maps in blow-ups & blow-downs", *Trans. Amer. Math. Soc.* **349**:1 (1997), 373–411. MR 97d:14021 Zbl 0867.14005

[Wunderlich 1965] W. Wunderlich, "Starre, kippende, wackelige und bewegliche Achtfläche", *Elem. Math.* **20** (1965), 25–32. MR 32 #4579 Zbl 0126.36906

Received August 22, 2009.

IVAN IZMESTIEV
INSTITUT FÜR MATHEMATIK, MA 8-3
TECHNISCHE UNIVERSITÄT BERLIN
STRASSE DES 17. JUNI 136
10623 BERLIN
GERMANY

izmestiev@math.tu-berlin.de
http://www.math.tu-berlin.de/~izmestie/

JEAN-MARC SCHLENKER
INSTITUT DE MATHÉMATIQUES DE TOULOUSE, UMR CNRS 5219
UNIVERSITÉ TOULOUSE III
31062 TOULOUSE CEDEX 9
FRANCE

jmschlenker@gmail.com
http://www.math.univ-toulouse.fr/~schlenker/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

ROBUST FOUR-MANIFOLDS AND ROBUST EMBEDDINGS

VYACHESLAV S. KRUSHKAL

A link in the 3-sphere is homotopically trivial, according to Milnor, if its components bound disjoint maps of disks in the 4-ball. This paper is concerned with the question of what spaces, when used in place of disks in an analogous definition, give rise to the same class of homotopically trivial links. We show that there are 4-manifolds for which this property depends on their embedding in the 4-ball. This work is motivated by the A-B slice problem, a reformulation of the 4-dimensional topological surgery conjecture. As a corollary, this provides a new, secondary obstruction in the A-B slice problem for a certain class of decompositions of D^4.

1. Introduction

The classification of knots and links up to concordance, in both the smooth and topological categories, is an important and difficult problem in 4-dimensional topology. Recall that two links in the 3-sphere are concordant if they bound disjoint embeddings of (smooth or locally flat, depending on the category) annuli in $S^3 \times [0, 1]$. Milnor [1954] introduced the notion of *link homotopy*, often referred to as the "theory of links modulo knots", which turned out to be much more tractable. In particular, there is an elegant characterization of homotopically trivial links using the *Milnor group*, a certain rather natural nilpotent quotient of the fundamental group of the link. Two links are link homotopic if they bound disjoint maps of annuli in $S^3 \times [0, 1]$, so the annuli are disjoint from each other, but (unlike the definition of concordance) they are allowed to have self-intersections. (Strictly speaking, this defines the notion of a singular concordance of links; however, it is known [Giffen 1979; Goldsmith 1979] to be equivalent to Milnor's original notion [Milnor 1954] of link homotopy.)

The subject of link homotopy brings together 4-dimensional geometric topology and the classical techniques of nilpotent group theory. An area where both approaches are important, and which is a motivation for the results in this paper, is the

MSC2000: 57M25, 57N13.
Keywords: robust 4-manifolds, the A-B slice problem, 4-dimensional topological surgery, link homotopy.
Research supported in part by NSF grant DMS-0605280.

A-B slice problem, a reformulation of the topological 4-dimensional surgery conjecture [Freedman and Lin 1989; Freedman and Quinn 1990]. Roughly speaking, this paper is concerned with the problem of characterizing spaces (in interesting cases, 4-manifolds with a specified curve in the boundary) which, when used in place of disks in a definition analogous to Milnor's, give rise to the same class of homotopically trivial links.

A specific question about the A-B slice problem, and related to link-homotopy theory, is the following: Suppose M is a codimension-zero submanifold of the 4-ball, $i : M \hookrightarrow D^4$, with a specified curve $\gamma \subset \partial M$ forming a knot in the 3-sphere: $i(\gamma) \subset S^3 = \partial D^4$. Given such a pair (M, γ), does there exist a homotopically essential link $\big(i_1(\gamma), \dots, i_n(\gamma)\big)$ in the 3-sphere, formed by disjoint embeddings $i_1, \dots, i_n$ of (M, γ) into (D^4, S^3)? If the answer is negative, the pair (M, γ) is called *robust*. The analysis of this problem is substantially more involved than the classical link homotopy case where one considers disks with self-intersections: in general the 1- and 2-handles of 4-manifolds embedded in D^4 may link, and the relations in the fundamental group of the complement do not have the "standard" form implied by the Clifford tori in Milnor's theory.

The main result of this paper is the existence of 4-manifolds for which this property depends on their embedding in the 4-ball:

Theorem 1. *There exist submanifolds $i : (M, \gamma) \hookrightarrow (D^4, S^3)$ such that*

(i) *there are disjoint embeddings of copies (M_i, γ_i) of (M, γ) into (D^4, S^3) forming a homotopically essential link $(\gamma_1, \dots, \gamma_n)$ in the 3-sphere, and*

(ii) *given any disjoint embeddings of (M, γ) into (D^4, S^3), each one isotopic to the original embedding i, the link $(\gamma_1, \dots, \gamma_n)$ formed by their attaching curves in the 3-sphere is homotopically trivial.*

In the A-B slice problem, one considers *decompositions* of the 4-ball, $D^4 = A \cup B$, where the specified curves α, β of the two parts form the Hopf link in $S^3 = \partial D^4$ (see [Freedman and Lin 1989; Krushkal 2008], and Section 2 below). It is shown in [Krushkal 2008] that there exist decompositions where neither of the two sides is robust. This result left open the question of whether, in fact, these decompositions may be used to solve the general 4-dimensional topological surgery conjecture.

This paper provides a detailed analysis of the construction in [Krushkal 2008]. The proof of Theorem 1 shows that, in this case, one of the two parts of the decomposition is robust provided that the re-embeddings forming the link $(\gamma_1, \dots, \gamma_n)$ are topologically equivalent to the original embedding into the 4-ball. This provides a new obstruction in the A-B slice problem for this class of decompositions of the 4-ball, exhibiting a new phenomenon where the obstruction depends not just on the submanifold but also on its specific embedding into D^4. In particular, this shows that the construction in [Krushkal 2008] does not satisfy the equivariance condition which is necessary for solving the canonical 4-dimensional surgery problems. An important open question is whether, given *any* decomposition $D^4 = A \cup B$, the conclusion (ii) of Theorem 1 holds for either A or B.

The main tool in the proof of part (ii) is the Milnor group, in the context of the *relative-slice problem*. This context is substantially different from Milnor's original work, since we use it to analyze embeddings of more general submanifolds in the 4-ball. The topology of these spaces is richer than the setting of disks with self-intersections in the 4-ball, considered in classical link homotopy. To find an obstruction, one has to consider in detail the structure of the graded Lie algebra associated to the lower central series of the link group. The strategy of the proof should be useful for further study of the A-B slice problem.

Section 2 gives a detailed definition of robust 4-manifolds and robust embeddings, and discusses its relation with the A-B slice problem. The construction [Krushkal 2008] of the submanifolds, used in the proof of Theorem 1, is recalled in Section 3. In Section 4, we review the Milnor group in the 4-dimensional setting and complete the proof of Theorem 1.

2. Robust 4-manifolds and the A-B slice problem

This section states the definition of robust 4-manifolds and robust embeddings, notions which provide a convenient setting for the results of this paper and are important for the A-B slice problem. Let M be a 4-manifold with a specified curve γ in its boundary. Let $i : M \hookrightarrow D^4$ be an embedding into the 4-ball with $i(\gamma) \subset S^3 = \partial D^4$.

Definition 2.1. The pair (M, γ) is called *robust* if, given any $n \geq 2$ and disjoint embeddings $i_1, \ldots, i_n$ of (M, γ) into (D^4, S^3), the link formed by the curves $i_1(\gamma), \ldots, i_n(\gamma)$ in the 3-sphere is homotopically trivial.

An *embedding* $i : (M, \gamma) \hookrightarrow (D^4, S^3)$ is *robust* if, given any $n \geq 2$ and disjoint embeddings $i_1, \ldots, i_n$ of (M, γ) into (D^4, S^3), each isotopic to the original embedding i, the link formed by the curves $i_1(\gamma), \ldots, i_n(\gamma)$ in the 3-sphere is homotopically trivial. (In this case, we say that the re-embeddings are *standard*.)

In these terms, Theorem 1 states that there exist pairs (M, γ) that are not robust but admit robust embeddings; these are the first examples of this phenomenon.

It follows easily from the definition that the 2-handle $(D^2 \times D^2, \{0\} \times \partial D^2)$, and more generally any kinky handle (a regular neighborhood in the 4-ball of a disk with self-intersections) is robust.

It is not difficult to give further examples: it follows from the link composition lemma [Freedman and Lin 1989; Krushkal and Teichner 1997] that the 4-manifold (B_0, β) in Figure 1, obtained from the collar $\beta \times D^2 \times [0, 1]$ by attaching 2-handles to the Bing double of the core of the solid torus, is robust. This example illustrates the important point that the disjoint copies $i_j(M)$ in the definition above are *embedded*: it is easy to see that, if the 2-handles H_1, H_2 in Figure 1 were allowed to intersect, this 4-manifold may be mapped to the collar on its attaching curve, and therefore there exist disjoint *singular* maps of copies of this manifold such that their attaching curves $\{\gamma_i\}$ form a homotopically essential link in the 3-sphere.

The complement in D^4 of the standard embedding of the 4-manifold in Figure 1 is the 4-manifold $A =$ (genus-one surface with one boundary component α) $\times D^2$. It is easy to see that (A, α) is not robust: for example, the Borromean rings form a homotopically essential link bounding disjoint standard genus-one surfaces in the 4-ball.

To review the relation of these results to the 4-dimensional topological surgery conjecture, recall the definition of an A-B slice link (see [Freedman and Lin 1989; Krushkal 2008] for a more detailed discussion.)

Definition 2.2. A *decomposition* of D^4 is a pair of compact codimension-zero submanifolds with boundary $A, B \subset D^4$, satisfying conditions (1)–(3) below, where

$$\partial^+A = \partial A \cap \partial D^4, \quad \partial^+B = \partial B \cap \partial D^4, \quad \partial^-A = \overline{\partial A \setminus \partial^+A}, \quad \partial^-B = \overline{\partial B \setminus \partial^+B}.$$

(1) $A \cup B = D^4$.

(2) $A \cap B = \partial^-A = \partial^-B$.

(3) $S^3 = \partial^+A \cup \partial^+B$ is the standard genus-one Heegaard decomposition of S^3.

Given an n-component link $L = (l_1, \dots, l_n) \subset S^3$, let $D(L) = (l_1, l_1', \dots, l_n, l_n')$ denote the $2n$-component link obtained by adding an untwisted parallel copy L' to L. The link L is A-B *slice* if there exist decompositions (A_i, B_i), $i = 1, \dots, n$, of D^4, and self-homeomorphisms φ_i, ψ_i of D^4, for $i = 1, \dots, n$, such that all sets in the collection $\varphi_1 A_1, \dots, \varphi_n A_n, \psi_1 B_1, \dots, \psi_n B_n$ are disjoint, and the following boundary data is satisfied: $\varphi_i(\partial^+A_i)$ is a tubular neighborhood of l_i and $\psi_i(\partial^+B_i)$ is a tubular neighborhood of l_i', for each i.

The surgery conjecture is equivalent to the statement that the Borromean rings, and a certain family of their generalizations, are A-B slice. In [Krushkal 2008], we constructed a decomposition $D^4 = A \cup B$ and disjoint embeddings A_i, B_i into D^4 so that the attaching curves $\{\alpha_i\}$ of the A_i formed the Borromean rings (or, more generally, any given link with trivial linking numbers) and the curves $\{\beta_i\}$ formed an untwisted parallel copy. The validity of one of the conditions necessary for solving the canonical surgery problems was unknown at the time of that construction, namely the equivariance (the existence of the homeomorphisms φ_i, ψ_i); phrased differently, it was not known whether there exist disjoint re-embeddings of the submanifolds A, B which are *standard*. It follows from Theorem 1 that standard disjoint embeddings for these decompositions do not exist. Therefore, an open question (important in the search for an obstruction to surgery in the context of the A-B slice problem) is: *Given any decomposition $D^4 = A \cup B$, is one of the two embeddings $A \hookrightarrow D^4$, $B \hookrightarrow D^4$ necessarily robust?*

3. Construction of the submanifolds

This section reviews the construction [Krushkal 2008] of the submanifolds of D^4, which will be used in the proof of Theorem 1 in Section 4. The construction consists of a series of modifications of the handle structures, starting with a standard surface and its complement in the 4-ball. Consider the genus-one surface S with a single boundary component α, and set $A_0 = S \times D^2$. Consider the standard embedding $(S, \alpha) \subset (D^4, S^3)$ (take an embedding of the surface in S^3, push it into the 4-ball and take a regular neighborhood). Then, A_0 is identified with a regular neighborhood of S in D^4. The complement B_0 of A_0 in the 4-ball is obtained from the collar on its attaching curve, $S^1 \times D^2 \times I$, by attaching a pair of zero-framed 2-handles to the Bing double of the core of the solid torus $S^1 \times D^2 \times \{1\}$, as in Figure 1. (See for example [Freedman and Lin 1989] for a proof of this statement.)

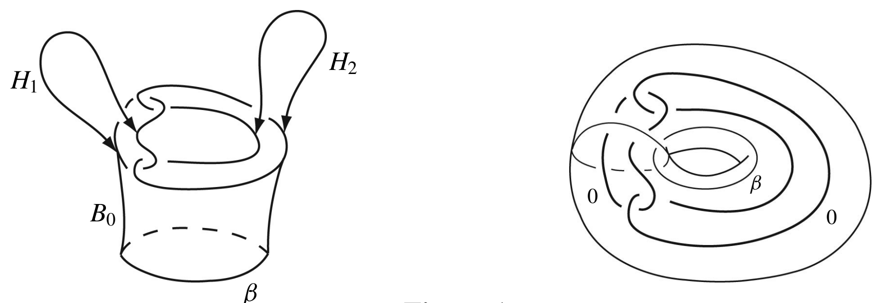


Figure 1

Note that a distinguished pair of curves α_1, α_2, forming a symplectic basis in the surface S, is determined as the meridians (linking circles) to the cores of the 2-handles H_1, H_2 of B_0 in D^4. In other words, α_1, α_2 are fibers of the circle normal bundles over the cores of H_1, H_2 in D^4.

An important observation [Freedman and Lin 1989] is that this construction may be iterated: Consider the 2-handle H_1 in place of the original 4-ball. The pair of curves (α_1 and the attaching circle β_1 of H_1) form the Hopf link in the boundary of H_1. In H_1, consider the standard genus-one surface T bounded by β_1. As discussed above, its complement is given by two zero-framed 2-handles attached to the Bing double of α_1. Assembling this data, consider the new decomposition $D^4 = A_1 \cup B_1$ (in this paper we need only the B-side of the decomposition, shown in Figure 2.) As above, the diagrams are drawn in solid tori (complements in S^3 of the unknotted circles drawn dashed in the figures.)

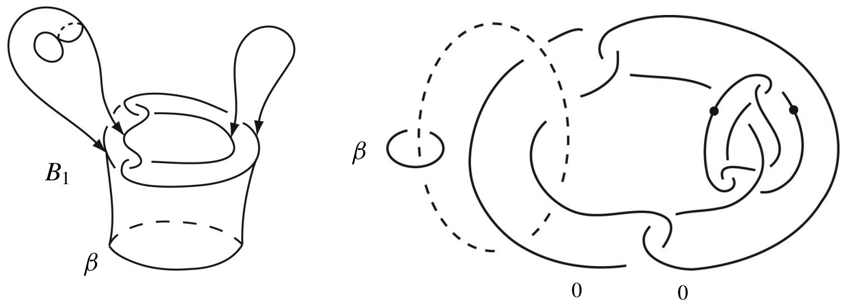


Figure 2

The handlebodies A_1, B_1 are examples of *model decompositions* [Freedman and Lin 1989] obtained by iterated applications of the construction above. It is known that such model handlebodies are robust and, in particular, the Borromean rings are not weakly A-B slice when restricted to the class of model decompositions. (A link L is *weakly* A-B slice if the submanifolds $\{A_i, B_i\}$ in Definition 2.2 may be embedded into D^4 disjointly, but the equivariance condition encoded by the existence of the homeomorphisms φ_i, ψ_i is omitted; see [Krushkal 2008].)

We are now in a position to define the decomposition $D^4 = A \cup B$ used in the proof of Theorem 1.

Definition 3.1. Consider $B = (B_1 \cup \text{zero-framed 2-handle})$, attached as shown in the Kirby diagram in Figure 3.

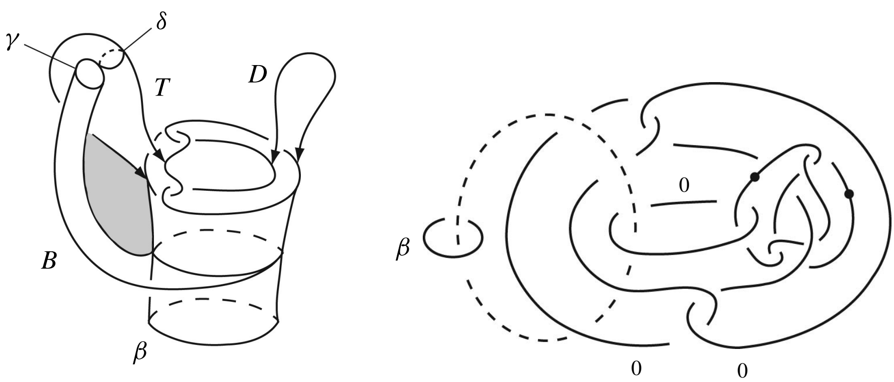


Figure 3

Imprecisely (up to homotopy, on the level of spines) B may be viewed as the union of B_1 with a 2-cell, attached along the composition of the attaching circle β of B_1 and a curve representing a generator of H_1(the second-stage surface of B_1). This 2-cell is schematically shown in the spine picture of B in Figure 3, left, as a cylinder connecting the two curves. The shading indicates that the new generator of π_1 created by adding the cylinder is filled-in with a disk. The figure showing

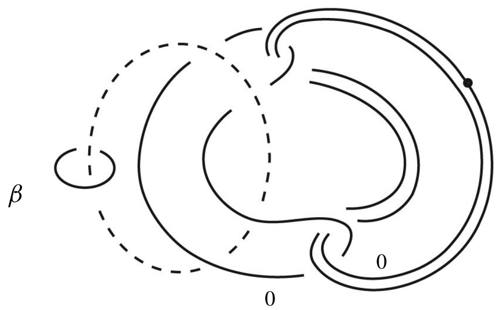


Figure 4

the spine is provided only as motivation for the construction; a precise description is given by the handle diagram.

Note that, by canceling a (1-handle, 2-handle) pair, one gets the diagram for B shown in Figure 4; this fact will be used in the proof of Theorem 1 in the next section. (Observe that the handle diagram in Figure 4 may also be obtained from the handle diagram of its complement, [Krushkal 2008, Figure 12].)

4. The Milnor group and the proof of Theorem 1

We start this section by summarizing the relevant information about the Milnor group, which will be used in the proof of Theorem 1. The reader is referred to the original [Milnor 1954] for a more complete introduction to the Milnor group of links in the 3-sphere; see also [Freedman and Teichner 1995, Section 2] for a discussion of the Milnor group in the more general 4-dimensional context.

Definition 4.1. Given a group G, normally generated by elements $g_1, \dots, g_n$, the *Milnor group* of G relative to the given normal generating set $\{g_i\}$ is defined as

$$MG = G \,/\, \langle\!\langle\, [g_i^x, g_i^y] \mid x, y \in G, i = 1, \dots, n \,\rangle\!\rangle. \tag{4-1}$$

The Milnor group is a finitely presented nilpotent group of class $\le n$, where n is the number of normal generators in the previous definition. In this paper, an example of interest is $G = \pi_1(D^4 \setminus \Sigma)$, where Σ is a collection of surfaces with boundary, properly and disjointly embedded in (D^4, S^3). In this case, a choice of normal generators is provided by the meridians m_i to the components Σ_i of Σ. Here, a meridian m_i is an element of G which is obtained by following a path α_i in $D^4 \setminus \Sigma$ from the basepoint to the boundary of a regular neighborhood of Σ_i, followed by a small circle (a fiber of the circle normal bundle) linking Σ_i, then followed by α_i^{-1}.

Denote by $F_{g_1,\dots,g_n}$ the free group generated by the g_i, and consider the Magnus expansion

$$M : F_{g_1,\dots,g_n} \to \mathbb{Z}[x_1, \dots, x_n] \tag{4-2}$$

into the ring of formal power series in noncommuting variables $\{x_i\}$, defined by

$$M(g_i) = 1 + x_i, \qquad M(g_i^{-1}) = 1 - x_i + x_i^2 \mp \dots$$

We will keep the same notation for the homomorphism

$$M : MF_{g_1,\dots,g_n} \to R_{x_1,\dots,x_n}, \tag{4-3}$$

induced by the Magnus expansion, into the quotient $R_{x_1,\dots,x_n}$ of $\mathbb{Z}[x_1, \dots, x_n]$ by the ideal generated by all monomials $x_{i_1} \cdots x_{i_k}$ with some index occurring at least twice. It is established in [Milnor 1954] that the homomorphism (4-3) is well defined and injective.

We now turn to the proof of Theorem 1. Consider the submanifold $i : (B, \beta) \subset (D^4, S^3)$ that was constructed in Definition 3.1. Part (i) of the theorem follows from [Krushkal 2008, Theorem 1], which showed that there exist disjoint embeddings of three copies (B_i, β_i), such that the link formed by the curves $\beta_1, \beta_2, \beta_3$ in the 3-sphere is the Borromean rings. It is convenient to introduce the next definition.

Definition 4.2. An embedding $j : (B, \beta) \hookrightarrow (D^4, S^3)$ is *standard* if there exists an ambient isotopy between j and the original embedding i from Definition 3.1.

Examining the proof of [Krushkal 2008, Theorem 1], one may check that the embeddings of the B_i, constructed there and giving rise to the Borromean rings on the boundary, are not standard. (In terms of the spine picture of B in Figure 3, for the standard embedding, the curve δ bounds a disk in D^4 which is disjoint from the 2-sphere formed by the core of the 2-handle D capped off with a null-homotopy for its attaching curve; on the other hand, for the embedding constructed in [Krushkal 2008], δ has linking number 1 with this 2-sphere.)

We will now show that given disjoint *standard* embeddings of several copies (B_i, β_i) into the 4-ball, the link formed by the curves $\beta_1, \dots, \beta_n$ in the 3-sphere is necessarily homotopically trivial. We will show that the Borromean rings do not bound disjoint standard embeddings of three copies of (B, β). The Borromean rings case is the most interesting example from the perspective of the A-B slice problem, while the case of other homotopically essential links is proved analogously.

Suppose to the contrary that the Borromean rings bound disjoint standard embeddings B_1, B_2, B_3. We will consider the *relative-slice* reformulation of the problem; see [Freedman and Lin 1989] and also [Krushkal 2008] for a more detailed introduction. Using the handle diagram in Figure 4, one then observes that there is a solution to the relative-slice problem shown in Figure 5. This means that the six components $l_1, \dots, l_6$ (drawn solid in the figure) bound disjoint embedded disks in the handlebody $D^4 \cup_{a,b,c}$ 2-handles, where the 2-handles are attached to the 4-ball with zero framings along the curves a, b, c (drawn dashed) from Figure 5. The fact that the embeddings $B_i \hookrightarrow D^4$ are standard is reflected by the fact that the slices bounded by the "solid" curves of each B_i do not go over the 2-handles (dashed curves) corresponding to the same B_i. This means that the slices for l_1, l_2 do not go over a and, similarly, l_3, l_4 do not go over b, nor l_5, l_6 over c. (Note that, without this restriction, there is a rather straightforward solution to this relative-slice problem.)

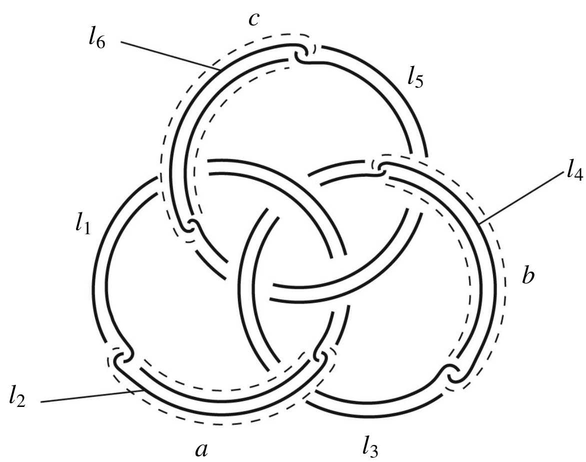


Figure 5

Therefore, assume the link in Figure 5 is relatively slice, subject to the "standard" condition discussed previously. Denote by D_i the slice bounded by l_i, $i = 1, \dots, 6$, and let $D = \bigcup_{i=2,\dots,6} D_i$. Consider

$$X := (D^4 \cup_{a,b,c} \text{2-handles}) \setminus D.$$

Denote by m_i the meridians to the components l_i, and by m_a, m_b, m_c the meridians to the curves a, b, c, respectively. The first homology $H_1(X)$ is generated by $m_2, \dots, m_6$. In fact, we view $\{m_i\}$ as based loops in X normally generating $\pi_1(X)$.

If we omit the first component l_1, the remaining link $(l_2, \dots, l_6, a, b, c)$ in Figure 5 is the unlink. This implies that the second homology $H_2(X)$ is spherical. Indeed, its generators may be represented by parallel copies of the cores of the 2-handles attached to a, b, c, capped off by disks in the complement of a neighborhood of the link $(l_2, \dots, l_6, a, b, c)$ in the 3-sphere. Therefore, the Milnor group $M\pi_1(X)$ with respect to the normal generators m_i is isomorphic to the free

Milnor group:

$$M\pi_1(X) \cong MF_{m_2,\dots,m_6}.$$

Indeed, since the Milnor group is nilpotent, it is obtained from the quotient by a term of the lower central series, $\pi_1(X)/(\pi_1(X))^n$, by adding the Milnor relations (4-1). The relations in the nilpotent group $\pi_1(X)/(\pi_1(X))^n$ may be read off from surfaces representing generators of $H_2(X)$; see [Krushkal 1998, Lemma 13]. In particular, the relations corresponding to spherical classes are trivial. Therefore, all relations in $M\pi_1(X)$ are the standard relations (4-1) or, in other words, $M\pi_1(X)$ is the free Milnor group.

It follows that the Magnus expansion (4-3)

$$M : M\pi_1(X) \cong MF_{m_2,\dots,m_6} \to R_{x_2,\dots,x_6} \tag{4-4}$$

is well defined. Connecting the first component l_1 to the basepoint, consider it as an element of $M\pi_1(X)$. From the assumption that the link in Figure 5 is relatively slice, it follows that l_1 bounds a disk in X and, in particular, that it is trivial in $M\pi_1(X)$. We will find a nontrivial term in the Magnus expansion (4-4) $M(l_1) \in R_{x_2,\dots,x_6}$, giving a contradiction with the relative-slice assumption.

Consider the meridians m_i to the components l_i in S^3, for $i = 2, \dots, 6$, and also the meridians m_a, m_b, m_c to a, b, c. The meridians m_i will also serve as meridians to the slices D_i bounded by l_i, for $i = 2, \dots, 6$, that were discussed above. Consider

$$l_1 = \big[m_a m_2, \big[[m_3, m_b m_4], [m_5, m_6 m_c]\big]\big] \in M\pi_1(X). \tag{4-5}$$

In this expression, m_a, m_b, m_c are elements of $M\pi_1(X)$ that depend on how the hypothetical slices D_i go over the 2-handles attached to a, b, c. The expression (4-5) may be read from the capped grope (see Figure 6) bounded by l_1 in the complement of the other components in the 3-sphere. (Note that the components $l_2, \dots, l_6, a, b, c$ intersect only the caps and not the body of the grope.)

Recall a basic commutator identity: any three elements f, g, h in a group satisfy

$$[fg, h] = [f, h]^g \, [g, h]. \tag{4-6}$$

Suppose two elements $s, t \in M\pi_1(X)$ have Magnus expansions

$$M(s) = 1 + \boldsymbol{x} \qquad \text{and} \qquad M(t) = 1 + \boldsymbol{y},$$

where $\boldsymbol{x}, \boldsymbol{y}$ denote the sum of all monomials of nonzero degree in the expansions of s, t. Then, the expansion of the conjugate tst^{-1} is of the form

$$1 + \boldsymbol{x} + \boldsymbol{xy} + \boldsymbol{yx} + \cdots$$

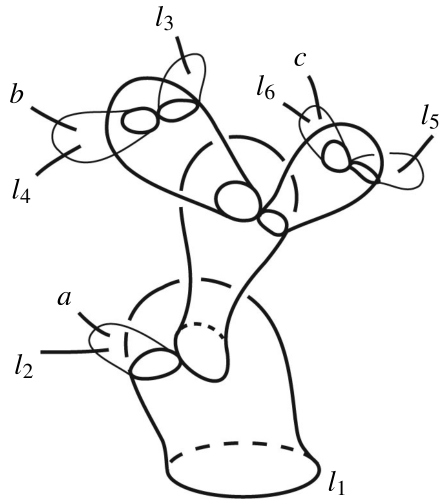


Figure 6

In particular, any first nontrivial term in the expansion $M(s)$ also appears in the expansion of any conjugate of s, $M(s^t)$. The expression (4-5) for l_1 is a 5-fold commutator and, since the ring $R_{x_2,\dots,x_6}$ is defined in terms of *nonrepeating* variables, this implies that any monomial of nonzero degree in the expansion $M(l_1)$ contains all the variables $x_2, \dots x_6$. This means that, to read off the Magnus expansion, any conjugation coming up while using (4-6) to simplify (4-5) may be omitted. These observations imply that the Magnus expansion $M(l_1)$ equals

$$M\big(\big[m_2, \big[[m_3, m_4], [m_5, m_6]\big]\big]\big), \tag{4-7}$$

times the Magnus expansion of seven other terms where some (or all) of m_2, m_4, m_6 are replaced with m_a, m_b, m_c. Moreover, recall that the "standard" embedding assumption implies that the slice D_2 bounded by l_2 does not go over the 2-handle attached to a. Therefore, the Magnus expansion of m_a is of the form

$$M(m_a) = 1 + \sum_{i=3}^{6} \alpha_i x_i + \text{higher terms},$$

for some coefficients α_i. Since the meridians m_3, m_5 are present in each commutator obtained by simplifying (4-5), the only terms in the Magnus expansion of m_a that may contribute to a nontrivial monomial in $M(l_1)$ are x_4 and x_6. Similarly, the only possibly nontrivial contributions to $M(l_1)$ of m_b are x_2, x_6, and of m_c are x_2, x_4.

Using the fact that

$$M([s,t]) = 1 + \boldsymbol{xy} - \boldsymbol{yx} \pm \cdots,$$

where $M(s)=1+\boldsymbol{x}$, $M(t)=1+\boldsymbol{y}$, note that the expansion (4-7) contains the monomial $x_2x_3x_4x_6x_5$. We claim that this monomial does not cancel with any other term in the expansion $M(l_1)$. This claim is proved by a direct inspection: any monomial in the Magnus expansion of a commutator of the form (4-7) with m_2 replaced by m_a has x_4 or x_6 as either the first or last variable. The only other possibility is the expansion of the commutator $\big[m_2, \big[[m_3, m_b], [m_5, m_c]\big]\big]$ with $M(m_b)$ contributing x_6, and $M(m_c)$ contributing x_4. The monomial $x_2x_3x_4x_6x_5$ does not appear in this expansion either. Therefore, we found a nontrivial term in the Magnus expansion $M(l_1) \in M\pi_1(X)$, contradicting the relative-slice assumption. This contradiction completes the proof.

References

[Freedman and Lin 1989] M. H. Freedman and X.-S. Lin, "On the (A, B)-slice problem", *Topology* **28**:1 (1989), 91–110. MR 90b:57002 Zbl 0845.57016

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

[Freedman and Teichner 1995] M. H. Freedman and P. Teichner, "4-manifold topology, II: Dwyer's filtration and surgery kernels", *Invent. Math.* **122**:3 (1995), 531–557. MR 96k:57016 Zbl 0857.57018

[Giffen 1979] C. H. Giffen, "Link concordance implies link homotopy", *Math. Scand.* **45**:2 (1979), 243–254. MR 82a:57005 Zbl 0453.57002

[Goldsmith 1979] D. L. Goldsmith, "Concordance implies homotopy for classical links in M^3", *Comment. Math. Helv.* **54**:3 (1979), 347–355. MR 80h:57006 Zbl 0441.57005

[Krushkal 1998] V. S. Krushkal, "Additivity properties of Milnor's $\overline{\mu}$-invariants", *J. Knot Theory Ramifications* **7**:5 (1998), 625–637. MR 2000a:57011 Zbl 0931.57005

[Krushkal 2008] V. S. Krushkal, "A counterexample to the strong version of Freedman's conjecture", *Ann. of Math.* (2) **168**:2 (2008), 675–693. MR 2009f:57029 Zbl 1176.57025

[Krushkal and Teichner 1997] V. S. Krushkal and P. Teichner, "Alexander duality, gropes and link homotopy", *Geom. Topol.* **1** (1997), 51–69. MR 98i:57013 Zbl 0885.55001

[Milnor 1954] J. Milnor, "Link groups", *Ann. of Math.* (2) **59** (1954), 177–195. MR 17,70e Zbl 0055.16901

Received August 19, 2009.

VYACHESLAV S. KRUSHKAL
UNIVERSITY OF VIRGINIA
DEPARTMENT OF MATHEMATICS
CHARLOTTESVILLE, VA 22904-4137
UNITED STATES

krushkal@virginia.edu
http://www.math.virginia.edu/~vk6e/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

ON SECTIONS OF GENUS TWO LEFSCHETZ FIBRATIONS

SINEM ÇELIK ONARAN

We find new relations in the mapping class group of a genus 2 surface with n boundary components for $n = 1, \dots, 8$ that induce a genus 2 Lefschetz fibration $\mathbb{CP}^2 \# 13\overline{\mathbb{CP}}^2 \to S^2$ with n disjoint sections. As a consequence, we show any holomorphic genus 2 Lefschetz fibration without separating singular fibers admits a section.

1. Introduction

The study of Lefschetz fibrations is important in low-dimensional topology because of a close relationship between symplectic 4-manifolds and Lefschetz fibrations, [Donaldson 1999; Gompf and Stipsicz 1999]. Sections of Lefschetz fibrations play an important role in the theory. For example, in the presence of a section, the fundamental group and the signature of a Lefschetz fibration can be easily computed.

Here, we provide sections for genus 2 Lefschetz fibrations $\mathbb{CP}^2 \# 13\overline{\mathbb{CP}}^2 \to S^2$, with global monodromy given by the relation $(t_{c_1}t_{c_2}t_{c_3}t_{c_4}t_{c_5}^2t_{c_4}t_{c_3}t_{c_2}t_{c_1})^2 = 1$ in the mapping class group Γ_2 of a closed genus 2 surface, where each c_i is a simple closed curve as in Figure 3, and t_{c_i} is a right-handed Dehn twist about c_i for $i = 1, \dots, 5$. In [Korkmaz and Ozbagci 2008], similar relations were found in the mapping class group $\Gamma_{1,n}$ of a genus 1 surface with n boundary components for $n = 4, \dots, 9$, giving an elliptic Lefschetz fibration $\mathbb{CP}^2 \# 9\overline{\mathbb{CP}}^2 \to S^2$ with n disjoint sections.

In Section 2, we recall definitions and relations in the mapping class group to be used in our computations, and we fix notation. In Section 3, we give brief background information on Lefschetz fibrations. In Section 4, we provide the necessary relations in the mapping class group $\Gamma_{2,n}$ for a genus 2 Lefschetz fibration $\mathbb{CP}^2 \# 13\overline{\mathbb{CP}}^2 \to S^2$ with n disjoint sections for $n = 1, \dots, 6$. In Section 5 we list several observations and open problems related to sections of Lefschetz fibrations. We show that a genus 2 Lefschetz fibration $\mathbb{CP}^2 \# 13\overline{\mathbb{CP}}^2 \to S^2$ may admit at most 12 disjoint sections. We provide relations in the corresponding mapping class

MSC2000: primary 57M50; secondary 57R17.
Keywords: Lefschetz fibrations, sections of Lefschetz fibrations.
The author is supported in part by TUBITAK-2214.

group that give $n = 7$ and $n = 8$ disjoint sections for genus 2 Lefschetz fibrations $\mathbb{CP}^2\#13\overline{\mathbb{CP}}^2 \to S^2$. We conclude that any holomorphic genus 2 Lefschetz fibration without separating singular fibers admits a section.

2. Mapping class groups

Let $\Sigma^k_{g,n}$ denote an oriented, connected, genus g surface, with n boundary components and k marked points. *The mapping class group* of $\Sigma^k_{g,n}$ is defined as the isotopy classes of orientation-preserving self-diffeomorphisms of $\Sigma^k_{g,n}$ that fix the marked points and the points on the boundary. Denote the mapping class group of $\Sigma^k_{g,n}$ by $\Gamma^k_{g,n}$. When $k = 0$, denote the mapping class group of $\Sigma_{g,n}$ by $\Gamma_{g,n}$.

Let a be a simple closed curve on $\Sigma^k_{g,n}$. A *right-handed Dehn twist* t_a about a is the isotopy class of a self-diffeomorphism of $\Sigma^k_{g,n}$ obtained by cutting $\Sigma^k_{g,n}$ along a and gluing it back after twisting one side by 2π to the right. The inverse of a right-handed Dehn twist is a *left-handed Dehn twist*, denoted by t_a^{-1}.

We now briefly mention the facts and relations to be used in our computations; for the proofs, see [Farb and Margalit 2005; Ivanov 2002]. If $f : \Sigma^k_{g,n} \to \Sigma^k_{g,n}$ is an orientation-preserving diffeomorphism, then $f t_a f^{-1} = t_{f(a)}$ for a a simple closed curve on $\Sigma^k_{g,n}$.

For simplicity, we will denote a right-handed Dehn twist t_a along a by a, and a left-handed Dehn twist t_a^{-1} by $\bar{a}$. The product ab means that we first apply the Dehn twist b, then the Dehn twist a. A simple closed curve parallel to a boundary component of a given surface will be called a *boundary curve* of the surface.

The following relations will be useful:

The *commutativity relation*. If a and b are two disjoint simple closed curves on $\Sigma^k_{g,n}$, then the Dehn twists along a and b commute: $ab = ba$.

The *braid relation*. If a and b are two simple closed curves on $\Sigma^k_{g,n}$ intersecting transversely at a single point, then their Dehn twists satisfy $aba = bab$.

The *lantern relation*. Consider a sphere with four holes, the boundary curves $\delta_1, \delta_2, \delta_3, \delta_4$, and the simple closed curves α, γ, σ, as shown in Figure 1. We have $\delta_1\delta_2\delta_3\delta_4 = \gamma\sigma\alpha$. Dehn discovered this relation; it was then rediscovered and named by D. Johnson.

The *star relation* [Gervais 2001]. Let $\Sigma_{1,3}$ be a torus with three boundary curves $\delta_1, \delta_2, \delta_3$. In $\Gamma_{1,3}$ we have $\delta_1\delta_2\delta_3 = (a_1a_2a_3b)^3$ for the simple closed curves a_1, a_2, a_3, b from Figure 1.

The *chain relation* for a two-holed torus. Consider a torus $\Sigma_{1,2}$ with two boundary curves δ_1, δ_2, and the simple closed curves c_1, c_2, b, as shown in Figure 1. We have $\delta_1\delta_2 = (c_1bc_2)^4$.

The *chain relations* for the genus 2 case: If c_1, c_2, c_3, c_4, c_5 is the chain of curves

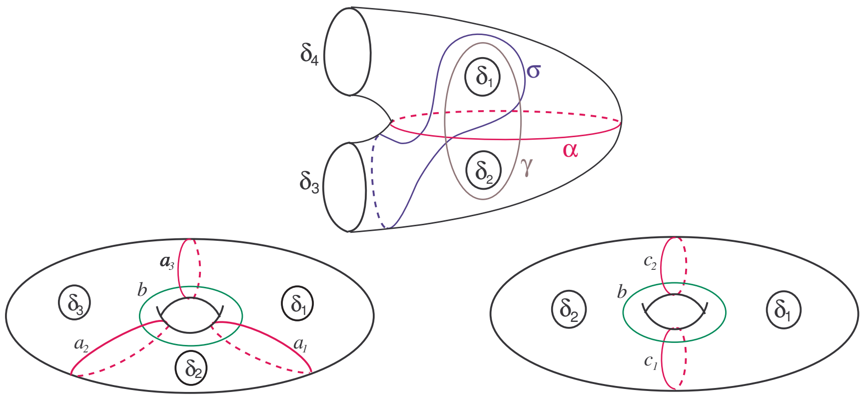


Figure 1. Counterclockwise from the top: The lantern relation, $\delta_1\delta_2\delta_3\delta_4 = \gamma\sigma\alpha$, the star relation, $\delta_1\delta_2\delta_3 = (a_1a_2a_3b)^3$, and the two-holed torus relation, $\delta_1\delta_2 = (c_1bc_2)^4$.

shown in Figure 2, then for a genus 2 surface $\Sigma_{2,1}$ with one boundary curve δ_1 we have $\delta_1 = (c_1c_2c_3c_4)^{10}$, while for a genus 2 surface $\Sigma_{2,2}$ with two boundary curves δ_1, δ_2 we have $\delta_1\delta_2 = (c_1c_2c_3c_4c_5)^6$.

3. Lefschetz fibrations

A *Lefschetz fibration* on a closed, connected, oriented smooth 4-manifold X is a map $f : X \to \Sigma$, where Σ is a closed, connected, oriented smooth surface, such that f is surjective, has isolated critical points, and for each critical point p there is an orientation-preserving local complex coordinate chart on which f takes the form $f(z_1, z_2) = z_1^2 + z_2^2$.

The Lefschetz fibration f is a smooth fiber bundle away from critical points. A regular fiber of f is diffeomorphic to a closed, oriented smooth genus g surface. We define the *genus* of the Lefschetz fibration to be the genus of a regular fiber.

A *singular fiber* is a fiber containing a critical point. We assume that each singular fiber contains only one critical point. The singular fiber can be obtained by taking a simple closed curve on a regular fiber and shrinking it to a point. This

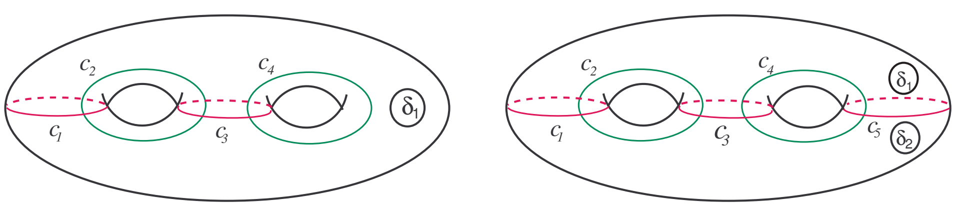


Figure 2. The chain relations: $\delta_1 = (c_1c_2c_3c_4)^{10}$ and $\delta_1\delta_2 = (c_1c_2c_3c_4c_5)^6$.

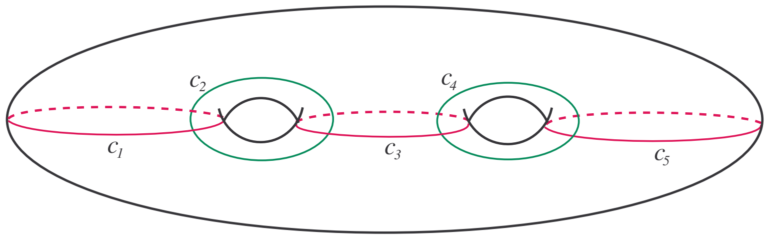


Figure 3. Σ_2.

simple closed curve describing the singular fiber is called a *vanishing cycle*. If this curve is a nonseparating curve, then the singular fiber is called *nonseparating*; otherwise it is called *separating*. For a genus g Lefschetz fibration over S^2, the product of Dehn twists along the vanishing cycles gives us the *global monodromy* of the Lefschetz fibration.

On a closed surface Σ_g, the right-handed Dehn twists c_i along the simple closed curves c_i for $i = 1, \dots, s$, with the relation $c_1 c_2 \cdots c_s = 1$, define a genus g Lefschetz fibration over S^2 with vanishing cycles $c_1, \dots, c_s$. In particular, in Γ_2 we have

$$(c_1 c_2 c_3 c_4 c_5^2 c_4 c_3 c_2 c_1)^2 = 1,$$
$$(c_1 c_2 c_3 c_4 c_5)^6 = 1,$$
$$(c_1 c_2 c_3 c_4)^{10} = 1,$$

where $c_1, \dots, c_5$ are simple closed curves as in Figure 3. For each relation above, we have genus 2 Lefschetz fibrations over S^2 with total spaces $\mathbb{CP}^2 \# 13\overline{\mathbb{CP}}^2$, $K3 \# 2\overline{\mathbb{CP}}^2$, and the Horikawa surface H, respectively. For more on Lefschetz fibrations, see [Auroux 2003; Gompf and Stipsicz 1999].

A section of a Lefschetz fibration is a map $\sigma : \Sigma \to X$ such that $f\sigma = \mathrm{id}_\Sigma$. Consider a collection of simple closed curves $c_1, \dots, c_s$ on a genus g surface $\Sigma_{g,n}$ with the relation

$$c_1 \cdots c_s = \delta_1^{k_1} \cdots \delta_n^{k_n}$$

in $\Gamma_{g,n}$, where $\delta_1, \dots, \delta_n$ are boundary curves, and $k_1, \dots, k_n$ are positive integers. This relation defines a genus g Lefschetz fibration over S^2 admitting n disjoint sections, with global monodromy $c_1 \cdots c_s = 1$. Moreover, the self-intersection of the i-th section is $-k_i$. To see this, note that after gluing a disk along each boundary curve one gets the relation $c_1 \cdots c_s = 1$ in Γ_g. Thus, this relation will give us a genus g Lefschetz fibration over S^2 as before. One can then use the centers of the capping disks to construct sections. For details, see [Gompf and Stipsicz 1999]; for more on self-intersection of sections, see [Smith 1998].

In the following sections, we will find relations of the above type, $c_1 \cdots c_s = \delta_1 \cdots \delta_n$, in the mapping class group $\Gamma_{2,n}$ for $n = 1, \dots, 8$.

4. Relations in $\Gamma_{2,n}$

For each $n = 1, \dots, 6$, we write the product of right-handed Dehn twists along the boundary curves $\delta_1, \dots, \delta_n$ as a product of twenty right-handed Dehn twists along nonboundary parallel simple closed curves on a genus 2 surface $\Sigma_{2,n}$. Namely, we provide relations of the form $\delta_1 \cdots \delta_n = \beta_1 \cdots \beta_{20}$, where $\beta_1, \dots, \beta_{20}$ are nonboundary parallel simple closed curves on $\Sigma_{2,n}$. After gluing disks along the boundary curves $\delta_1, \dots, \delta_n$, we get the relation $1 = \beta_1 \cdots \beta_{20}$ in the mapping class group Γ_2. By using the commutativity relation and the braid relation, one can simplify the right-hand side of the equation $1 = \beta_1 \cdots \beta_{20}$ so that it gives us the global monodromy $(c_1c_2c_3c_4c_5^2c_4c_3c_2c_1)^2 = 1$ of a Lefschetz fibration $\mathbb{CP}^2\#13\overline{\mathbb{CP}}^2 \to S^2$. For the simple closed curves $c_1, \dots, c_5$, see Figure 3. Here, twenty right-handed Dehn twists along nonboundary parallel, nonseparating simple closed curves c_i correspond to twenty nonseparating singular fibers. The consecutive simple closed curves c_i are the vanishing cycles.

In the following subsections, the relations we find in $\Gamma_{2,n}$ will give us a genus 2 Lefschetz fibration $\mathbb{CP}^2 \# 13\overline{\mathbb{CP}}^2 \to S^2$ admitting n disjoint sections of self-intersection -1, for $n = 1, \dots, 6$.

4.1. ***Genus two surface with one hole.*** Consider the genus 2 surface $\Sigma_{2,1}$ with one boundary curve δ_1, as in Figure 4. We have

$$\delta_1 = (a_1b_1a_2b_2)^{10} = (a_1b_1a_2b_2)^5(a_1b_1a_2b_2)^5.$$

Using the commutativity and braid relations, one can show

$$(a_1b_1a_2b_2)^5 = (a_1b_1a_2)^4(b_2a_2b_1a_1^2b_1a_2b_2).$$

(For the proof, see the appendix.) Notice the two-holed torus embedded in $\Sigma_{2,1}$ with two boundary curves a_3, a_4; then, by the chain relation for this torus, we have $(a_1b_1a_2)^4 = a_3a_4$.

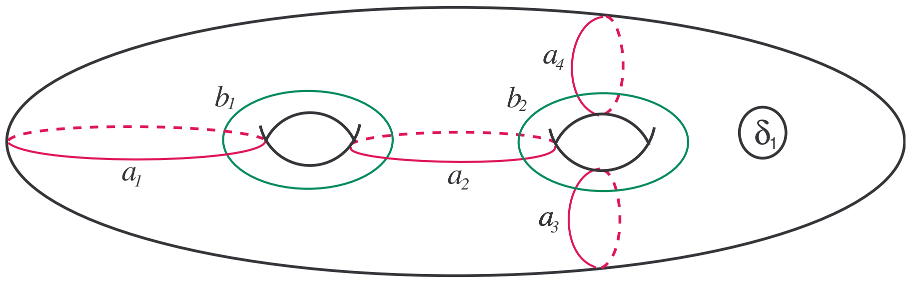


Figure 4. $\Sigma_{2,1}$.

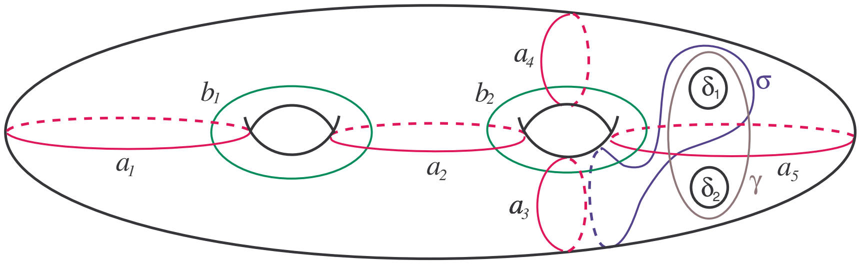


Figure 5. $\Sigma_{2,2}$.

By combining the relations above, we get

$$\begin{aligned}\delta_1 &= (a_1b_1a_2)^4(b_2a_2b_1a_1^2b_1a_2b_2)(a_1b_1a_2)^4(b_2a_2b_1a_1^2b_1a_2b_2)\\ &= a_3a_4(b_2a_2b_1a_1^2b_1a_2b_2)a_3a_4(b_2a_2b_1a_1^2b_1a_2b_2)\\ &= (a_3a_4b_2a_2b_1a_1^2b_1a_2b_2)^2.\end{aligned}$$

4.2. ***Genus two surface with two holes.*** Consider the genus 2 surface $\Sigma_{2,2}$ with two boundary curves δ_1, δ_2 from Figure 5, and notice the embedded sphere embedded in $\Sigma_{2,2}$ with four boundary curves $\delta_1, \delta_2, a_3, a_4$. Then, using the lantern relation, we have

$$a_3a_4\delta_1\delta_2 = \gamma\sigma a_5.$$

Notice in Figure 5 the genus 2 surface with one boundary curve γ; we thus have the chain relation $\gamma = (a_1b_1a_2b_2)^{10}$. Substituting γ in the lantern relation and then using the two-holed torus relation $(a_1b_1a_2)^4 = a_3a_4$ in the equation, we get

$$\begin{aligned}a_3a_4\delta_1\delta_2 &= \gamma\sigma a_5\\ &= (a_1b_1a_2b_2)^{10}\sigma a_5\\ &= (a_1b_1a_2b_2)^5(a_1b_1a_2b_2)^5\sigma a_5\\ &= (a_1b_1a_2)^4(b_2a_2b_1a_1^2b_1a_2b_2)(a_1b_1a_2)^4(b_2a_2b_1a_1^2b_1a_2b_2)\sigma a_5\\ &= a_3a_4(b_2a_2b_1a_1^2b_1a_2b_2)a_3a_4(b_2a_2b_1a_1^2b_1a_2b_2)\sigma a_5.\end{aligned}$$

We simplify this equation to

$$\delta_1\delta_2 = b_2a_2b_1a_1^2b_1a_2b_2a_3a_4(b_2a_2b_1a_1^2b_1a_2b_2)\sigma a_5.$$

4.3. ***Genus two surface with three holes.*** First, the lantern relation for the sphere with four boundary curves $\delta_1, \delta_2, a_4, a_5$ in $\Sigma_{2,3}$ from Figure 6 is $a_4a_5\delta_1\delta_2 = \gamma\sigma a_6$. For the three-holed torus with boundary curves γ, a_1, a_2, we have the star relation $\gamma a_1a_2 = (a_4a_5a_3b_2)^3$, while for the three-holed torus with boundary curves δ_3, a_4, a_5, we have $\delta_3a_4a_5 = (a_1a_2a_3b_1)^3$.

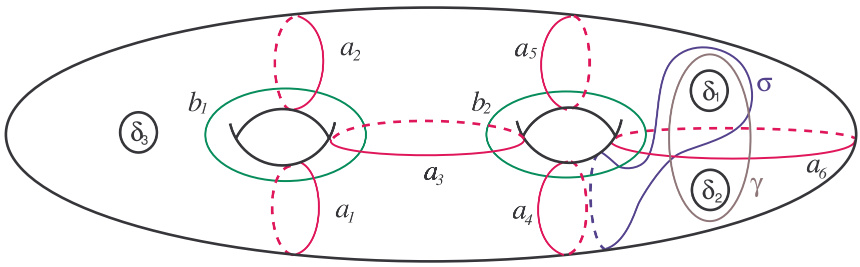


Figure 6. $\Sigma_{2,3}$.

Now, combine these relations, substitute δ_3 and γ, then simplify by using the commutativity and braid relations. Note that all the a_i commute for $i = 1, \dots, 6$. The simple closed curves a_1, a_2, a_3 intersect b_1 transversely at a single point, and a_4, a_5, a_3 intersect b_2 transversely at a single point. Thus, with $\beta = \bar{a}_5\bar{a}_4 b_2 a_4 a_5$,

$$\begin{aligned}
\delta_1\delta_2 &= \bar{a}_4\bar{a}_5\gamma\sigma a_6,\\
\delta_1\delta_2\delta_3 &= \delta_3\bar{a}_4\bar{a}_5\gamma\sigma a_6\\
&= \delta_3\bar{a}_4\bar{a}_5\bar{a}_1\bar{a}_2(a_4a_5a_3b_2)^3\sigma a_6\\
&= \bar{a}_1\bar{a}_2(\delta_3)\bar{a}_4\bar{a}_5(a_4a_5a_3b_2)(a_4a_5a_3b_2)^2\sigma a_6\\
&= \bar{a}_1\bar{a}_2((a_1a_2a_3b_1)^3\bar{a}_4\bar{a}_5)a_3b_2(a_4a_5a_3b_2)^2\sigma a_6\\
&= a_3b_1(a_1a_2a_3b_1)^2a_3(\bar{a}_4\bar{a}_5b_2a_4a_5)a_3b_2a_4a_5a_3b_2\sigma a_6\\
&= a_3b_1(a_1a_2a_3b_1)^2a_3\beta a_3b_2a_4a_5a_3b_2\sigma a_6.
\end{aligned}$$

4.4. ***Genus two surface with four holes.*** We will use the three-holed genus two relation we found in Section 4.3. Notice in Figure 7 the genus 2 surface with three boundary curves $\delta_3, \delta_4, \gamma$. Then

$$\begin{aligned}
\delta_3\delta_4\gamma &= a_3b_2\,(a_5a_4a_3b_2)^2a_3\beta_1a_3b_1a_2a_1a_3b_1\sigma_1a_7\\
&= a_3\beta_1a_3b_1a_2a_1a_3b_1\sigma_1a_7a_3b_2\,(a_5a_4a_3b_2)^2.
\end{aligned}$$

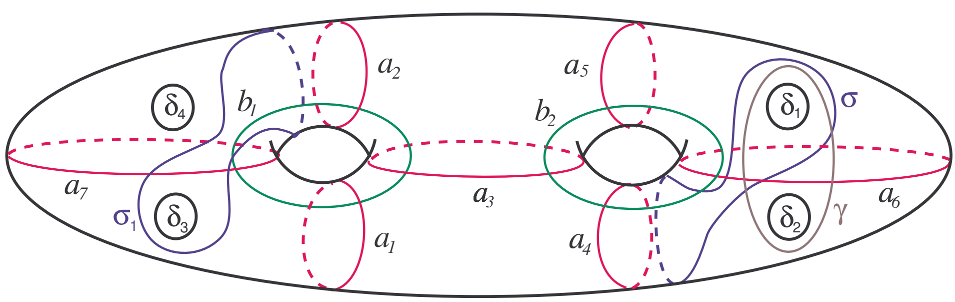


Figure 7. $\Sigma_{2,4}$.

where $\beta_1 = \bar{a}_1\bar{a}_2 b_1 a_2 a_1$. Note that we identify the curves $(a_1, a_2, a_3, a_4, a_5, a_6)$ in $\Sigma_{2,3}$ from Figure 6 with the curves $(a_5, a_4, a_3, a_2, a_1, a_7)$ in $\Sigma_{2,4}$ from Figure 7.

By the lantern relation for the sphere with four boundary curves $\delta_1, \delta_2, a_4, a_5$, we have

$$a_4 a_5 \delta_1 \delta_2 = \gamma \sigma a_6.$$

Now, combine the above relations to get

$$\begin{aligned}\delta_1\delta_2\delta_3\delta_4 &= a_3\beta_1 a_3 b_1 a_2 a_1 a_3 b_1 \sigma_1 a_7 a_3 b_2 (a_5 a_4 a_3 b_2)^2 \bar{\gamma}\bar{a}_4\bar{a}_5 \gamma\sigma a_6\\ &= a_3\beta_1 a_3 b_1 a_2 a_1 a_3 b_1 \sigma_1 a_7 a_3 b_2 (a_5 a_4 a_3 b_2) a_3 a_5 a_4 b_2 (\bar{a}_4\bar{a}_5 \sigma a_6)\\ &= a_3\beta_1 a_3 b_1 a_2 a_1 a_3 b_1 \sigma_1 a_7 a_3 b_2 (a_5 a_4 a_3 b_2) a_3 \beta_2 \sigma a_6,\end{aligned}$$

where $\beta_1 = \bar{a}_1\bar{a}_2 b_1 a_2 a_1$ and $\beta_2 = a_5 a_4 b_2 \bar{a}_4 \bar{a}_5$.

4.5. ***Genus two surface with five holes.*** The lantern relation for the sphere with four boundary curves $\delta_1, \delta_2, a_4, a_5$ in $\Sigma_{2,5}$ from Figure 8 is $a_4 a_5 \delta_1 \delta_2 = \gamma\sigma a_6$. In Figure 8, notice the genus 2 surface with four boundary curves $\delta_3, \delta_4, \delta_5, \gamma$. Identify the curves $(\delta_1, \delta_2, a_5, a_6, \sigma)$ in $\Sigma_{2,4}$ from Figure 7 with $(\delta_5, \gamma, a_8, a_5, \sigma_2)$ in $\Sigma_{2,5}$ from Figure 8. Then, by the relation given in Section 4.4, we have

$$\delta_3\delta_4\delta_5\gamma = a_3\beta_1 a_3 b_1 a_2 a_1 a_3 b_1 \sigma_1 a_7 a_3 b_2 (a_8 a_4 a_3 b_2) a_3 \beta_2 \sigma_2 a_5,$$

where $\beta_1 = \bar{a}_1\bar{a}_2 b_1 a_2 a_1$ and $\beta_2 = a_8 a_4 b_2 \bar{a}_4 \bar{a}_8$.

Now, combine the above relations and simplify the equation as

$$\begin{aligned}\delta_1\delta_2\delta_3\delta_4\delta_5 &= a_3\beta_1 a_3 b_1 a_2 a_1 a_3 b_1 \sigma_1 a_7 a_3 b_2 (a_8 a_4 a_3 b_2) a_3 \beta_2 \sigma_2 a_5 \bar{\gamma}\bar{a}_4\bar{a}_5 \gamma\sigma a_6\\ &= a_3\beta_1 a_3 b_1 a_2 a_1 a_3 b_1 \sigma_1 a_7 a_3 b_2 (a_8 a_4 a_3 b_2) a_3 \beta_2 \sigma_2 \bar{a}_4 \sigma a_6\\ &= a_3\beta_1 a_3 b_1 a_2 a_1 a_3 b_1 \sigma_1 a_7 a_3 (\bar{a}_4) b_2 (a_4 a_8 a_3 b_2) a_3 \beta_2 \sigma_2 \sigma a_6\\ &= a_3\beta_1 a_3 b_1 a_2 a_1 a_3 b_1 \sigma_1 a_7 a_3 \beta_3 (a_8 a_3 b_2) a_3 \beta_2 \sigma_2 \sigma a_6,\end{aligned}$$

where $\beta_1 = \bar{a}_1\bar{a}_2 b_1 a_1 a_2$, $\beta_2 = a_8 a_4 b_2 \bar{a}_4 \bar{a}_8$, and $\beta_3 = \bar{a}_4 b_2 a_4$.

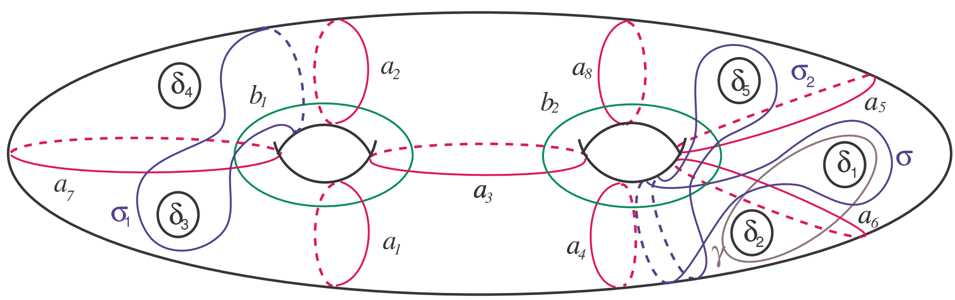


Figure 8. $\Sigma_{2,5}$.

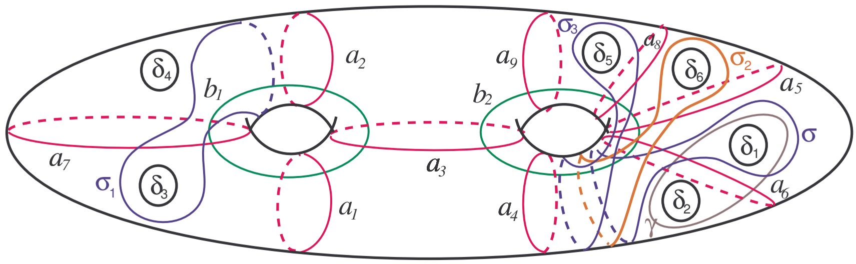


Figure 9. $\Sigma_{2,6}$.

4.6. ***Genus two surface with six holes.*** The lantern relation for the sphere with four boundary curves δ_1, δ_2, a_4, a_5 in $\Sigma_{2,6}$ from Figure 9 is $a_4a_5\delta_1\delta_2 = \gamma\sigma a_6$. Now, identify the curves $(\delta_1, \delta_2, a_6, a_8, \sigma, \sigma_2)$ in $\Sigma_{2,5}$ from Figure 8 with the curves $(\delta_6, \gamma, a_5, a_9, \sigma_2, \sigma_3)$ in $\Sigma_{2,6}$ from Figure 9. By the relation given in Section 4.5 for the genus 2 surface with five boundary curves δ_3, δ_4, δ_5, δ_6, γ, we have

$$\delta_6\gamma\delta_3\delta_4\delta_5 = a_3\beta_1a_3b_1a_2a_1a_3b_1\sigma_1a_7a_3\beta_3a_9a_3b_2a_3\beta_2\sigma_3\sigma_2a_5.$$

where $\beta_1 = \bar{a}_1\bar{a}_2b_1a_2a_1$, $\beta_2 = a_9a_4b_2\bar{a}_4\bar{a}_9$, and $\beta_3 = \bar{a}_4b_2a_4$.

Now, combine the above relations to get

$$\begin{aligned}\delta_1\delta_2\delta_3\delta_4\delta_5\delta_6 &= a_3\beta_1a_3b_1a_2a_1a_3b_1\sigma_1a_7a_3\beta_3a_9a_3b_2a_3\beta_2\sigma_3\sigma_2a_5\bar{\gamma}\bar{a}_4\bar{a}_5\gamma\sigma a_6\\ &= a_3\beta_1a_3b_1a_2a_1a_3b_1\sigma_1a_7a_3\beta_3a_9a_3b_2a_3\beta_2\sigma_3\sigma_2(\bar{a}_4)\sigma a_6\\ &= a_3\beta_1a_3b_1a_2a_1a_3b_1\sigma_1a_7a_3(\bar{a}_4)(\beta_3)a_9a_3b_2a_3\beta_2\sigma_3\sigma_2\sigma a_6\\ &= a_3\beta_1a_3b_1a_2a_1a_3b_1\sigma_1a_7a_3(\bar{a}_4)(\bar{a}_4b_2a_4)a_9(a_3b_2a_3)\beta_2\sigma_3\sigma_2\sigma a_6\\ &= a_3\beta_1a_3b_1a_2a_1a_3b_1\sigma_1a_7a_3(\bar{a}_4)(b_2a_4\bar{b}_2)a_9(b_2a_3b_2)\beta_2\sigma_3\sigma_2\sigma a_6\\ &= a_3\beta_1a_3b_1a_2a_1a_3b_1\sigma_1a_7a_3\beta_3\beta_4a_3b_2\beta_2\sigma_3\sigma_2\sigma a_6\end{aligned}$$

where $\beta_1 = \bar{a}_1\bar{a}_2b_1a_2a_1$, $\beta_2 = a_5a_4b_2\bar{a}_4\bar{a}_5$, $\beta_3 = \bar{a}_4b_2a_4 = b_2a_4\bar{b}_2$, and $\beta_4 = \bar{b}_2a_9b_2$.

5. Final remarks

Lemma 5.1. *A genus* 2 *Lefschetz fibration* $\mathbb{CP}^2\,\#\,13\overline{\mathbb{CP}}^2 \to S^2$ *admits at most* 12 *disjoint sections.*

Proof. Suppose that $\mathbb{CP}^2\,\#\,13\overline{\mathbb{CP}}^2 \to S^2$ admits 13 disjoint sections. Each section is a sphere with self-intersection -1. Furthermore, each section intersects a regular fiber, a genus 2 surface Σ_2 with self-intersection 0, at one point. Now, by blowing down all -1 spheres, we get a genus 2 surface $\widetilde{\Sigma}_2$ with self-intersection 13, which cannot exist in a manifold with second homology $\mathbb{Z}$. □

In Section 4, we found relations giving n disjoint sections for genus 2 Lefschetz fibration $\mathbb{CP}^2 \# 13\overline{\mathbb{CP}}^2 \to S^2$ for $n = 1, \dots, 6$. The technique applied in Section 4 stops at $n = 6$. However, by using results from [Korkmaz and Ozbagci 2008], we can find relations in the corresponding mapping class group that give $n = 7$ and $n = 8$ disjoint sections for genus 2 Lefschetz fibration $\mathbb{CP}^2 \# 13\overline{\mathbb{CP}}^2 \to S^2$. We will next show how to derive these relations. This method does not go further, and it remains unknown whether there are more than eight sections.

The seven-holed torus relation from [Korkmaz and Ozbagci 2008] sits in $\Sigma_{2,7}$:

$$\delta_1\delta_2\delta_3\delta_4\delta_5\delta_6\delta_7 = \alpha_3\alpha_4\alpha_1 b_1\sigma_5\alpha_2\beta_5\sigma_3\sigma_6\alpha_6\beta_3\sigma_4,$$

where $\beta_3 = \alpha_3 b_1\overline{\alpha}_3$ and $\beta_5 = \alpha_5 b_1\overline{\alpha}_5$ in $\Sigma_{2,7}$; see Figure 10. We identify the boundary curves $(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7)$ in $\Sigma_{1,7}$ from Figure 10 with the curves $(\delta_6, \delta_5, a_2, a_1, \delta_2, \delta_1, \delta_7)$ in $\Sigma_{2,7}$. The seven-holed torus relation gives

$$a_1 a_2\delta_1\delta_2\delta_5\delta_6\delta_7 = a_3 a_4 a_9 b_2\sigma_5 a_{10}\beta_5\sigma_3\sigma_6 a_5\beta_3\sigma_4$$

where $\beta_3 = a_3 b_2\overline{a}_3$ and $\beta_5 = a_6 b_2\overline{a}_6$.

Next, we combine this with the lantern relation $a_1 a_2\delta_3\delta_4 = \gamma\sigma a_7$ for the sphere with four boundary curves $\delta_3, \delta_4, a_1, a_2$ in $\Sigma_{2,7}$. The star relation for the torus with three boundary curves γ, a_4, a_{10} is $a_4 a_{10}\gamma = (a_1 a_2 a_3 b_1)^3$. We now substitute

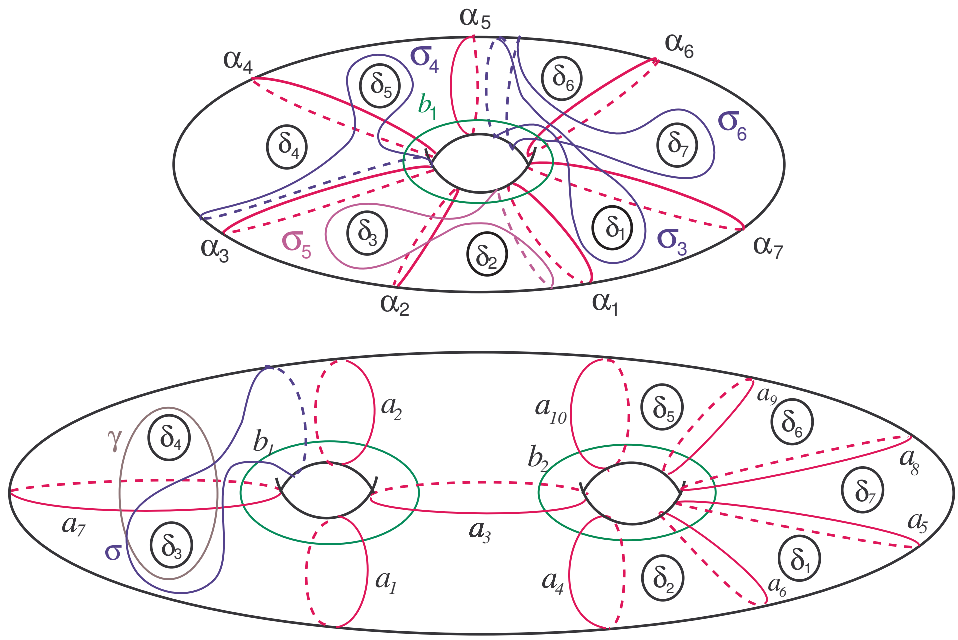


Figure 10. Seven-holed torus relation.

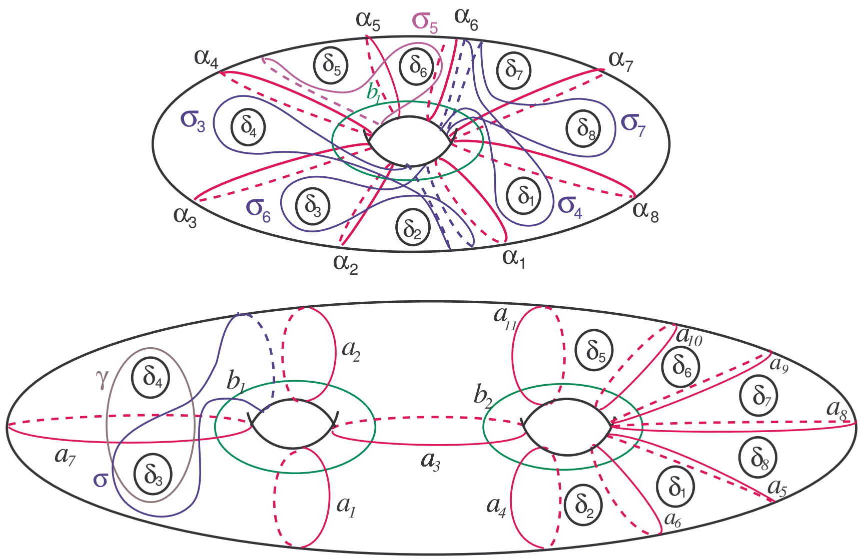


Figure 11. Eight-holed torus relation.

$\gamma = \bar{a}_4\bar{a}_{10}(a_1a_2a_3b_1)^3$ into the lantern relation; then we simplify the equation and write the product of right-handed Dehn twists along the boundary curves $\delta_1, \dots, \delta_7$ as a product of twenty right-handed Dehn twists along nonboundary parallel simple closed curves on $\Sigma_{2,7}$:

$$
\begin{aligned}
&\delta_1\delta_2\delta_3\delta_4\delta_5\delta_6\delta_7\\
&\qquad = a_3a_4a_9b_2\sigma_5a_{10}\beta_5\sigma_3\sigma_6a_5\beta_3\sigma_4\bar{a}_1\bar{a}_2\bar{a}_1\bar{a}_2\gamma\sigma a_7\\
&\qquad = a_3a_9b_2\sigma_5a_{10}\beta_5\sigma_3\sigma_6a_5\beta_3\sigma_4(a_4)\bar{a}_1\bar{a}_2\bar{a}_1\bar{a}_2\bar{a}_4\bar{a}_{10}(a_1a_2a_3b_1)^3\sigma a_7\\
&\qquad = a_3a_9b_2\sigma_5a_{10}\beta_5\sigma_3\sigma_6a_5\beta_3\sigma_4\bar{a}_1\bar{a}_2a_3b_1(a_1a_2a_3b_1)^2(\bar{a}_{10})\sigma a_7\\
&\qquad = a_3a_9(\bar{a}_{10})b_2\sigma_5a_{10}\beta_5\sigma_3\sigma_6a_5\beta_3\sigma_4a_3(\bar{a}_1\bar{a}_2b_1a_1a_2)a_3b_1(a_1a_2a_3b_1)\sigma a_7\\
&\qquad = a_3a_9(\bar{a}_{10}b_2a_{10})(\bar{a}_{10}\sigma_5a_{10})\beta_5\sigma_3\sigma_6a_5\beta_3\sigma_4a_3\tilde{\beta}a_3b_1(a_1a_2a_3b_1)\sigma a_7\\
&\qquad = a_3a_9\tilde{\beta}_1\tilde{\beta}_2\beta_5\sigma_3\sigma_6a_5\beta_3\sigma_4a_3\tilde{\beta}a_3b_1(a_1a_2a_3b_1)\sigma a_7,
\end{aligned}
$$

where $\tilde{\beta}_1 = \bar{a}_{10}b_2a_{10}$, $\tilde{\beta}_2 = \bar{a}_{10}\sigma_5a_{10}$, $\beta_5 = a_6b_2\bar{a}_6$, $\beta_3 = a_3b_2\bar{a}_3$, and $\tilde{\beta} = \bar{a}_1\bar{a}_2b_1a_1a_2$. Note that the simple closed curves σ_5 and a_{10} intersect at 2 points.

Similarly, the eight-holed torus relation from [Korkmaz and Ozbagci 2008] sits in $\Sigma_{2,8}$ (see Figure 11):

$$\delta_1\delta_2\delta_3\delta_4\delta_5\delta_6\delta_7\delta_8 = \alpha_4\alpha_5\beta_1\sigma_3\sigma_6\alpha_2\beta_6\sigma_4\sigma_7\alpha_7\beta_4\sigma_5 = \alpha_4b_1\sigma_5\alpha_5\beta_1\sigma_3\sigma_6\alpha_2\beta_6\sigma_4\sigma_7\alpha_7,$$

where $\beta_1 = \alpha_1b_1\bar{\alpha}_1$, $\beta_4 = \alpha_4b_1\bar{\alpha}_4$, and $\beta_6 = \alpha_6b_1\bar{\alpha}_6$ in $\Sigma_{2,8}$. We identify the curves

$(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7, \delta_8)$ in $\Sigma_{1,8}$ with $(\delta_1, \delta_8, \delta_7, \delta_6, \delta_5, a_2, a_1, \delta_2)$ in $\Sigma_{2,8}$.

By applying the same technique, we also get the necessary relation for $n = 8$. The eight-holed torus relation gives

$$a_1 a_2 \delta_1 \delta_2 \delta_5 \delta_6 \delta_7 \delta_8 = a_{10} b_2 \sigma_5 a_{11} \beta_1 \sigma_3 \sigma_6 a_8 \beta_6 \sigma_4 \sigma_7 a_4,$$

where $\beta_1 = a_5 b_2 \bar{a}_5$ and $\beta_6 = a_3 b_2 \bar{a}_3$. We combine this with the lantern relation $a_1 a_2 \delta_3 \delta_4 = \gamma \sigma a_7$ for the sphere with four boundary curves $\delta_3, \delta_4, a_1, a_2$ in $\Sigma_{2,8}$. Using the star relation $a_4 a_{11} \gamma = (a_1 a_2 a_3 b_1)^3$ for the torus with three boundary curves γ, a_4, a_{11}, we substitute $\gamma = \bar{a}_4 \bar{a}_{11} (a_1 a_2 a_3 b_1)^3$ in the lantern relation. We simplify the equation as

$$\begin{aligned}
&\delta_1 \delta_2 \delta_3 \delta_4 \delta_5 \delta_6 \delta_7 \delta_8 \\
&\quad = a_{10} b_2 \sigma_5 a_{11} \beta_1 \sigma_3 \sigma_6 a_8 \beta_6 \sigma_4 \sigma_7 a_4 \bar{a}_1 \bar{a}_2 \bar{a}_1 \bar{a}_2 \gamma \sigma a_7 \\
&\quad = a_{10} b_2 \sigma_5 a_{11} \beta_1 \sigma_3 \sigma_6 a_8 \beta_6 \sigma_4 \sigma_7 a_4 \bar{a}_1 \bar{a}_2 \bar{a}_1 \bar{a}_2 \bar{a}_4 \bar{a}_{11} (a_1 a_2 a_3 b_1)^3 \sigma a_7 \\
&\quad = a_{10} b_2 \sigma_5 a_{11} \beta_1 \sigma_3 \sigma_6 a_8 \beta_6 \sigma_4 \sigma_7 \bar{a}_1 \bar{a}_2 a_3 b_1 (a_1 a_2 a_3 b_1)^2 (\bar{a}_{11}) \sigma a_7 \\
&\quad = a_{10} (\bar{a}_{11}) b_2 \sigma_5 a_{11} \beta_1 \sigma_3 \sigma_6 a_8 \beta_6 \sigma_4 \sigma_7 a_3 (\bar{a}_1 \bar{a}_2 b_1 a_1 a_2) a_3 b_1 (a_1 a_2 a_3 b_1) \sigma a_7 \\
&\quad = a_{10} (\bar{a}_{11} b_2 a_{11}) (\bar{a}_{11} \sigma_5 a_{11}) \beta_1 \sigma_3 \sigma_6 a_8 \beta_6 \sigma_4 \sigma_7 a_3 \tilde{\beta} a_3 b_1 (a_1 a_2 a_3 b_1) \sigma a_7 \\
&\quad = a_{10} \tilde{\beta}_1 \tilde{\beta}_2 \beta_1 \sigma_3 \sigma_6 a_8 \beta_6 \sigma_4 \sigma_7 a_3 \tilde{\beta} a_3 b_1 (a_1 a_2 a_3 b_1) \sigma a_7,
\end{aligned}$$

where $\beta_1 = a_5 b_2 \bar{a}_5$, $\beta_6 = a_3 b_2 \bar{a}_3$, $\tilde{\beta} = \bar{a}_1 \bar{a}_2 b_1 a_1 a_2$, $\tilde{\beta}_1 = \bar{a}_{11} \beta_2 a_{11}$ and $\tilde{\beta}_2 = \bar{a}_{11} \sigma_5 a_{11}$. Note that the simple closed curves σ_5 and a_{11} intersect at 2 points.

By Lemma 5.1, for $n > 12$ there is no relation in the mapping class group $\Gamma_{2,n}$ inducing a genus 2 Lefschetz fibration $\mathbb{CP}^2 \# 13\overline{\mathbb{CP}}^2 \to S^2$ with n disjoint sections. For $n = 1, \dots, 8$, we did find relations giving n disjoint sections for genus 2 Lefschetz fibration $\mathbb{CP}^2 \# 13\overline{\mathbb{CP}}^2 \to S^2$. As a consequence, by using a result of K. Chakiris, we observe:

Corollary 5.2. *Any genus 2 holomorphic Lefschetz fibration without separating singular fibers admits a section.*

Proof. Chakiris [1978] showed that any genus 2 holomorphic Lefschetz fibration without separating singular fibers is obtained by fiber-summing the three genus 2 Lefschetz fibrations given by the relations

$$(c_1 c_2 c_3 c_4 c_5^2 c_4 c_3 c_2 c_1)^2 = 1, \quad (c_1 c_2 c_3 c_4 c_5)^6 = 1, \quad \text{and} \quad (c_1 c_2 c_3 c_4)^{10} = 1$$

in Γ_2, where $c_1, \dots, c_5$ are the simple closed curves shown in Figure 3. As noted in Section 2, each relation gives us a genus 2 Lefschetz fibration with total spaces $\mathbb{CP}^2 \# 13\overline{\mathbb{CP}}^2$, $\mathrm{K3} \# 2\overline{\mathbb{CP}}^2$, and the Horikawa surface H, respectively.

For the Lefschetz fibrations with total spaces $\mathrm{K3} \# 2\overline{\mathbb{CP}}^2$ and H, it is known that they have sections: The relation $(c_1 c_2 c_3 c_4 c_5)^6 = \delta_1 \delta_2$ in $\Gamma_{2,2}$ gives 2 disjoint sections for the Lefschetz fibration $K3\#2\overline{\mathbb{CP}}^2 \to S^2$, and $(c_1c_2c_3c_4)^{10} = \delta_1$ in $\Gamma_{2,1}$ gives a section for the Lefschetz fibration $H \to S^2$; see Figure 2.

Earlier, we found sections for the genus 2 Lefschetz fibration $\mathbb{CP}^2\#13\overline{\mathbb{CP}}^2 \to S^2$. By sewing the sections during the fiber-sum operation, we get a section for any genus 2 holomorphic Lefschetz fibration without separating singular fibers. □

Remark 5.3. One may continue and try to write similar relations for $9 \le n \le 12$ to see the exact number of disjoint sections that $\mathbb{CP}^2\#13\overline{\mathbb{CP}}^2 \to S^2$ can admit. One can also try to find the exact number of disjoint sections of the genus 2 Lefschetz fibrations with total spaces $K3\#2\overline{\mathbb{CP}}^2$ and H, respectively. It is still not known whether every genus g Lefschetz fibration over S^2 admits a section.

Appendix

In this appendix, we deduce the relation

$$(a_1b_1a_2b_2)^5 = (a_1b_1a_2)^4(b_2a_2b_1a_1^2b_1a_2b_2),$$

used in Section 4.1; for the corresponding curves, see Figure 4.

Note that a_1 intersects b_1 transversely at a single point, and commutes with a_2 and b_2. Also note that a_2 intersects b_1 and b_2 transversely at a single point, and the simple closed curves b_1 and b_2 commute. By the commutativity and braid relations, we have

$$\begin{aligned}
(a_1b_1a_2b_2)^5 &= (a_1b_1a_2(b_2))(a_1b_1a_2b_2)(a_1b_1a_2b_2)^3\\
&= (a_1b_1a_2)a_1b_1(b_2a_2b_2)(a_1b_1a_2b_2)^3\\
&= (a_1b_1a_2)a_1b_1(a_2b_2a_2)(a_1b_1a_2b_2)^3\\
&= (a_1b_1a_2)^2b_2a_2((a_1)b_1a_2b_2)(a_1b_1a_2b_2)^2\\
&= (a_1b_1a_2)^2a_1b_2(a_2b_1a_2)b_2(a_1b_1a_2b_2)^2\\
&= (a_1b_1a_2)^2a_1b_2((b_1)a_2(b_1))b_2(a_1b_1a_2b_2)^2\\
&= (a_1b_1a_2)^2a_1b_1(b_2a_2b_2)(b_1a_1b_1)a_2b_2(a_1b_1a_2b_2)\\
&= (a_1b_1a_2)^2a_1b_1(a_2b_2a_2)(a_1b_1a_1)a_2b_2(a_1b_1a_2b_2)\\
&= (a_1b_1a_2)^3b_2a_2((a_1)b_1a_1)(a_2)b_2(a_1b_1a_2b_2)\\
&= (a_1b_1a_2)^3a_1b_2(a_2b_1a_2)a_1b_2(a_1b_1a_2b_2)\\
&= (a_1b_1a_2)^3a_1b_2((b_1)a_2b_1)a_1(b_2)(a_1b_1a_2b_2)\\
&= (a_1b_1a_2)^3a_1b_1(b_2a_2b_2)b_1a_1(a_1b_1a_2b_2)\\
&= (a_1b_1a_2)^3a_1b_1(a_2b_2a_2)b_1a_1(a_1b_1a_2b_2)\\
&= (a_1b_1a_2)^4(b_2a_2b_1a_1^2b_1a_2b_2).
\end{aligned}$$

Acknowledgments

I am grateful to Anar Akhmedov for his suggestions and interest in this work. I would like to thank Kenneth L. Baker, John B. Etnyre, and Mustafa Korkmaz for helpful comments on the paper.

References

[Auroux 2003] D. Auroux, "Monodromy invariants in symplectic topology", preprint, 2003. arXiv math.SG/0304113

[Chakiris 1978] K. Chakiris, *The monodromy of genus 2 pencils*, thesis, Columbia University, 1978.

[Donaldson 1999] S. K. Donaldson, "Lefschetz pencils on symplectic manifolds", *J. Differential Geom.* **53**:2 (1999), 205–236. MR 2002g:53154 Zbl 1040.53094

[Farb and Margalit 2005] B. Farb and D. Margalit, "A primer on mapping class groups", preprint, 2005, available at http://www.math.utah.edu/~margalit/primer/.

[Gervais 2001] S. Gervais, "A finite presentation of the mapping class group of a punctured surface", *Topology* **40**:4 (2001), 703–725. MR 2002m:57025 Zbl 0992.57013

[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

[Ivanov 2002] N. V. Ivanov, "Mapping class groups", pp. 523–633 in *Handbook of geometric topology*, edited by R. J. Daverman and R. B. Sher, North-Holland, Amsterdam, 2002. MR 2003h:57022 Zbl 1002.57001

[Korkmaz and Ozbagci 2008] M. Korkmaz and B. Ozbagci, "On sections of elliptic fibrations", *Michigan Math. J.* **56**:1 (2008), 77–87. MR 2009f:57043 Zbl 1158.57033

[Smith 1998] I. Smith, *Symplectic geometry of Lefschetz fibrations*, thesis, University of Oxford, 1998.

Received October 4, 2009.

SINEM ÇELIK ONARAN
DEPARTMENT OF MATHEMATICS
MIDDLE EAST TECHNICAL UNIVERSITY
ANKARA
TURKEY

e114485@metu.edu.tr

and

SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GA
UNITED STATES

sinemo@math.gatech.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

BIHARMONIC HYPERSURFACES IN RIEMANNIAN MANIFOLDS

YE-LIN OU

We study biharmonic hypersurfaces in a generic Riemannian manifold. We first derive an invariant equation for such hypersurfaces generalizing the biharmonic hypersurface equation in space forms studied by Jiang, Chen, Caddeo, Montaldo, and Oniciuc. We then apply the equation to show that the generalized Chen conjecture is true for totally umbilical biharmonic hypersurfaces in an Einstein space, and construct a 2-parameter family of conformally flat metrics and a 4-parameter family of multiply warped product metrics, each of which turns the foliation of an upper-half space of $\mathbb{R}^m$ by parallel hyperplanes into a foliation with each leaf a proper biharmonic hypersurface. We also study the biharmonicity of Hopf cylinders of a Riemannian submersion.

1. Biharmonic maps and submanifolds

All manifolds, maps, and tensor fields that appear in this paper are assumed to be smooth unless stated otherwise.

A biharmonic map is a map $\varphi:(M,g)\to(N,h)$ between Riemannian manifolds that is a critical point of the bienergy functional

$$E^2(\varphi,\Omega)=\frac{1}{2}\int_\Omega|\tau(\varphi)|^2\mathrm{d}x$$

for every compact subset Ω of M, where $\tau(\varphi)=\mathrm{Trace}_g\nabla\mathrm{d}\varphi$ is the tension field of φ. The Euler–Lagrange equation of this functional gives the biharmonic map equation [Jiang 1986b]

$$\tau^2(\varphi):=\mathrm{Trace}_g(\nabla^\varphi\nabla^\varphi-\nabla^\varphi_{\nabla^M})\tau(\varphi)-\mathrm{Trace}_g R^N(\mathrm{d}\varphi,\tau(\varphi))\mathrm{d}\varphi=0, \tag{1}$$

which states that φ is biharmonic if and only if its bitension field $\tau^2(\varphi)$ vanishes identically. In this equation we used R^N to denote the curvature operator of (N,h)

MSC2000: 53C12, 58E20, 53C42.
Keywords: biharmonic maps, biharmonic hypersurfaces, biharmonic foliations, conformally flat space, Einstein space.
Supported by the Texas A&M University-Commerce Faculty Development Program, 2008–2009.

defined by

$$R^N(X,Y)Z = [\nabla^N_X, \nabla^N_Y]Z - \nabla^N_{[X,Y]}Z.$$

Clearly, it follows from (1) that any harmonic map is biharmonic and we call the nonharmonic biharmonic maps *proper biharmonic maps*.

Let M^m be a submanifold of Euclidean space $\mathbb{R}^n$ with the mean curvature vector H viewed as a map $H : M \to \mathbb{R}^n$. B. Y. Chen [1991] called M^m a biharmonic submanifold if $\triangle H = (\Delta H^1, \dots, \Delta H^n) = 0$, where Δ is the Beltrami–Laplace operator of the induced metric on M^m. If we use $i : M \to \mathbb{R}^n$ to denote the inclusion map of the submanifold, then the tension field of the inclusion map i is given by $\tau(i) = \Delta i = mH$, and hence the submanifold $M^n \subset \mathbb{R}^n$ is biharmonic if and only if $\triangle H = \triangle(\frac{1}{m}\Delta i) = \frac{1}{m}\triangle^2 i = \frac{1}{m}\tau^2(i) = 0$, that is, the inclusion map is a biharmonic map. In general, a submanifold M of (N, h) is called a *biharmonic submanifold* if the inclusion map $i : (M, i^*h) \to (N, h)$ is a biharmonic isometric immersion. It is well known that an isometric immersion is minimal if and only if it is harmonic. So a minimal submanifold is trivially biharmonic, and we call a nonminimal biharmonic submanifold a *proper biharmonic submanifold*.

Here are some known facts about biharmonic submanifolds:

Biharmonic submanifolds in Euclidean spaces. Jiang [1987] and then Chen and Ishikawa [1998] proved that any biharmonic submanifold in $\mathbb{R}^3$ is minimal. In [1992], Dimitrić showed that any biharmonic curve in $\mathbb{R}^n$ is a part of a straight line, any biharmonic submanifold of finite type in $\mathbb{R}^n$ is minimal, any pseudoumbilical submanifolds $M^m \subset \mathbb{R}^n$ with $m \neq 4$ is minimal, and any biharmonic hypersurface in $\mathbb{R}^n$ with at most two distinct principal curvatures is minimal. Hasanis and Vlachos [1995] proved that any biharmonic hypersurface in $\mathbb{R}^4$ is minimal. Based on these results, B. Y. Chen [1991] made the still-open conjecture that any biharmonic submanifold of Euclidean space is minimal.

Biharmonic submanifolds in hyperbolic space forms. Caddeo, Montaldo and Oniciuc [2002] showed that any biharmonic submanifold in hyperbolic 3-space is minimal, and that any m-dimensional pseudoumbilical biharmonic submanifold of hyperbolic n-space is minimal if $m \neq 4$. It is shown in [Balmuş et al. 2008] that any biharmonic hypersurface of hyperbolic n-space with at most two distinct principal curvatures is minimal. Based on these, Caddeo, Montaldo and Oniciuc [2001] extended Chen's conjecture to the *generalized Chen conjecture*: any biharmonic submanifold in (N, h) is minimal if $\text{Riem}^N \leq 0$.

Biharmonic submanifolds in spheres. The first example of a proper biharmonic submanifold in S^{n+1} was found in [Jiang 1986a] to be the generalized Clifford torus $S^p(1/\sqrt{2}) \times S^q(1/\sqrt{2})$ with $p \neq q$ and $p + q = n$. Caddeo, Montaldo, and Oniciuc [2001] found a second type of proper biharmonic submanifolds in S^{n+1}

to be the hypersphere $S^n(1/\sqrt{2})$, and also gave a complete classification of biharmonic submanifolds in S^3. Balmuş, Montaldo, and Oniciuc [2008] proved that any pseudoumbilical biharmonic submanifold $M^m \subset S^{n+1}$ with $m \neq 4$ has constant mean curvature, and also showed that if a hypersurface $M^n \subset S^{n+1}$ with at most two distinct principal curvatures (which by [Nishikawa and Maeda 1974] is equivalent for $n > 3$ to saying that M is a quasiumbilical or conformally flat hypersurface in S^{n+1}) is biharmonic, then M is an open part of the hypersphere $S^n(1/\sqrt{2})$ or the generalized Clifford torus $S^p(1/\sqrt{2}) \times S^q(1/\sqrt{2})$ with $p \neq q$ and $p+q=n$. Zhang [2008] found some examples of proper biharmonic real hypersurfaces in $\mathbb{C}P^n$ and determined all proper biharmonic tori $T^{n+1} = S^1(r_1) \times S^1(r_2) \times \cdots \times S^1(r_{n+1})$ in S^{2n+1}. All known examples of biharmonic submanifolds in spheres are consistent with the conjecture of [Balmuş et al. 2008] that any biharmonic submanifold in sphere has constant mean curvature, and any proper biharmonic hypersurface in S^{n+1} is an open part of the hypersphere $S^n(1/\sqrt{2})$ or the generalized Clifford torus $S^p(1/\sqrt{2}) \times S^q(1/\sqrt{2})$ with $p \neq q$ and $p+q=n$.

Biharmonic submanifolds in other model spaces. The survey article [Montaldo and Oniciuc 2006] contains an account of the study of biharmonic curves in various models. See [Arslan et al. 2005; Inoguchi 2004; Fetcu and Oniciuc 2009a; 2009b; Sasahara 2005; 2008] for special biharmonic submanifolds in contact manifolds or Sasakian space forms.

Biharmonic submanifolds in other senses. Some authors, for example, Javaloyes and Meroño [2003], use the condition $\Delta H = 0$ to define a biharmonic submanifold of a Riemannian manifold; this definition agrees with ours only if the ambient space is flat. For conformal biharmonic submanifolds (that is, conformal biharmonic immersions), see [Ou 2009].

This paper studies biharmonic hypersurfaces in a generic Riemannian manifold. In Section 2, we derive an invariant equation for such hypersurfaces that involves the mean curvature function, the norm of the second fundamental form, the shape operator of the hypersurface, and the Ricci curvature of the ambient space. We prove that the generalized Chen conjecture holds for totally umbilical hypersurfaces in an Einstein space. Section 3 is devoted to constructing a family of conformally flat metrics and a family of multiply warped product metrics, each of which turns the foliation of an upper-half space of $\mathbb{R}^m$ by parallel hyperplanes into a foliation with each leaf a proper biharmonic hypersurface. We accomplish these by starting with hyperplanes in Euclidean space and then looking for a type of conformally flat or multiply warped product metric on the ambient space that will reduce the biharmonic hypersurface equation into ordinary differential equations whose solutions give the metrics that render the inclusion maps proper biharmonic isometric immersions. In Section 4, we study biharmonicity of Hopf cylinders given by a Riemannian submersion from a complete 3-manifold. Our method shows that there is no proper biharmonic Hopf cylinder in S^3, thus recovering [Inoguchi 2004, Proposition 3.1].

2. The equations of biharmonic hypersurfaces

Recall that if $\varphi : M \to (N, h)$ is the inclusion map of a submanifold, or more generally, an isometric immersion, then we have an orthogonal decomposition of the vector bundle $\varphi^{-1}TN = \tau M \oplus \nu M$ into the tangent and normal bundles. We use $\mathrm{d}\varphi$ to identify TM with its image τM in $\varphi^{-1}TN$. Then, for any $X, Y \in \Gamma(TM)$ we have $\nabla^{\varphi}_X(\mathrm{d}\varphi(Y)) = \nabla^N_X Y$, whereas $\mathrm{d}\varphi(\nabla^M_X Y)$ equals the tangential component of $\nabla^N_X Y$. It follows that

$$\nabla \mathrm{d}\varphi(X, Y) = \nabla^{\varphi}_X(\mathrm{d}\varphi(Y)) - \mathrm{d}\varphi(\nabla^M_X Y) = B(X, Y), \tag{2}$$

that is, the second fundamental form $\nabla \mathrm{d}\varphi(X, Y)$ of the isometric immersion φ agrees with the second fundamental form $B(X, Y)$ of the immersed submanifold $\varphi(M)$ in N; for details see [Kobayashi and Nomizu 1969, Chapter 7] and [Baird and Wood 2003, Example 3.2.3]. From (2) we see that the tension field $\tau(\varphi)$ of an isometric immersion and the mean curvature vector field η of the submanifold are related by

$$\tau(\varphi) = m\eta.$$

For a hypersurface, that is, a codimension one isometric immersion φ of M^m into N^{m+1}, we can choose a local unit vector field ξ normal to $\varphi(M) \subset N$. Then $\eta = H\xi$ for H the mean curvature function, and we can write $B(X, Y) = b(X, Y)\xi$, where $b : TM \times TM \to C^{\infty}(M)$ is the second fundamental form. The relationship between the shape operator A of the hypersurface with respect to the unit normal vector field ξ and the second fundamental form is given by

$$B(X, Y) = \langle \nabla^N_X Y, \xi\rangle\xi = -\langle Y, \nabla^N_X \xi\rangle\xi = \langle AX, Y\rangle\xi, \tag{3}$$

$$\langle AX, Y\rangle = \langle B(X, Y), \xi\rangle = \langle b(X, Y)\xi, \xi\rangle = b(X, Y). \tag{4}$$

Theorem 2.1. *Let $\varphi : M^m \to N^{m+1}$ be an isometric immersion of codimension one with mean curvature vector $\eta = H\xi$. Then φ is biharmonic if and only if*

$$\begin{aligned} \Delta H - H|A|^2 + H\,\mathrm{Ric}^N(\xi, \xi) &= 0,\\ 2A(\mathrm{grad}\, H) + \tfrac{1}{2}m\,\mathrm{grad}\, H^2 - 2H(\mathrm{Ric}^N(\xi))^{\top} &= 0, \end{aligned} \tag{5}$$

where $\mathrm{Ric}^N : T_qN \to T_qN$ denotes the Ricci operator of the ambient space, defined by $\langle \mathrm{Ric}^N(Z), W\rangle = \mathrm{Ric}^N(Z, W)$, and A is the shape operator of the hypersurface with respect to the unit normal vector ξ.

Proof. First choose a local orthonormal frame $\{e_i\}_{i=1,\dots,m}$ on M such that the orthonormal frame $\{d\varphi(e_1), \dots, d\varphi(e_m), \xi\}$ is adapted to the ambient space defined on the hypersurface. Identifying $d\varphi(X) = X$ and $\nabla^{\varphi}_X W = \nabla^N_X W$ and noting that the tension field of φ is $\tau(\varphi) = mH\xi$, we can compute the bitension field of φ as

$$\begin{aligned}
\tau^2(\varphi) &= \sum_{i=1}^m \{\nabla^{\varphi}_{e_i}\nabla^{\varphi}_{e_i}(mH\xi) - \nabla^{\varphi}_{\nabla_{e_i}e_i}(mH\xi) - R^N(d\varphi(e_i), mH\xi)d\varphi(e_i)\} \\
&= m\sum_{i=1}^m (e_ie_i(H)\xi + 2e_i(H)\nabla^N_{e_i}\xi + H\nabla^N_{e_i}\nabla^N_{e_i}\xi - (\nabla_{e_i}e_i)(H)\xi - H\nabla^N_{\nabla_{e_i}e_i}\xi) \\
&\qquad\qquad - mH\sum_{i=1}^m R^N(d\varphi(e_i), \xi)d\varphi(e_i) \\
&= m(\Delta H)\xi - 2mA(\operatorname{grad} H) - mH\Delta^{\varphi}\xi - mH\sum_{i=1}^m R^N(d\varphi(e_i), \xi)d\varphi(e_i).
\end{aligned}$$

To find the tangential and normal parts of the bitension field, we first compute the tangential and normal components of the curvature term, getting

$$\sum_{i,k=1}^m \langle R^N(d\varphi(e_i), \xi)d\varphi(e_i), e_k\rangle e_k = -(\operatorname{Ric}^N(\xi, e_k))e_k = -(\operatorname{Ric}(\xi))^{\top},$$

$$\sum_{i=1}^m \langle R^N(d\varphi(e_i), \tau(\varphi))d\varphi(e_i), \xi\rangle = -mH\operatorname{Ric}^N(\xi, \xi).$$

To find the normal part of $\Delta^{\varphi}\xi$, we compute

$$\langle \Delta^{\varphi}\xi, \xi\rangle = \sum_{i=1}^m \langle -\nabla^N_{e_i}\nabla^N_{e_i}\xi + \nabla^N_{\nabla_{e_i}e_i}\xi, \xi\rangle = \sum_{i=1}^m \langle \nabla^N_{e_i}\xi, \nabla^N_{e_i}\xi\rangle. \tag{6}$$

On the other hand, using (3) and (4), we have

$$\begin{aligned}
|A|^2 &= \sum_{i,j=1}^m \langle Ae_i, e_j\rangle^2 = \sum_{i,j=1}^m \langle \nabla^N_{e_i}\xi, e_j\rangle^2 = \sum_{i=1}^m \langle \nabla^N_{e_i}\xi, \sum_{j=1}^m \langle \nabla^N_{e_i}\xi, e_j\rangle e_j\rangle \\
&= \sum_{i=1}^m \langle \nabla^N_{e_i}\xi, \nabla^N_{e_i}\xi\rangle,
\end{aligned}$$

which, together with (6), implies that

$$(\Delta^{\varphi}\xi)^{\perp} = \langle \Delta^{\varphi}\xi, \xi\rangle\xi = \sum_{i=1}^m \langle \nabla^N_{e_i}\xi, \nabla^N_{e_i}\xi\rangle\xi = |A|^2\xi.$$

A straightforward computation gives the tangential part of $\Delta^{\varphi}\xi$ as

$$
(7)\quad
\begin{aligned}
(\Delta^{\varphi}\xi)^{\top} &= \sum_{i,k=1}^{m} \langle -\nabla^{N}_{e_i}\nabla^{N}_{e_i}\xi + \nabla^{N}_{\nabla_{e_i}e_i}\xi, e_k\rangle e_k \\
&= \sum_{i,k=1}^{m} \langle \nabla^{N}_{e_i} A e_i - A(\nabla_{e_i}e_i),\ e_k\rangle e_k = \sum_{i,k=1}^{m} ((\nabla_{e_i} b)(e_k, e_i))e_k.
\end{aligned}
$$

Substituting the Codazzi–Mainardi equation for a hypersurface, namely,

$$(\nabla_{e_i} b)(e_k, e_i) - (\nabla_{e_k} b)(e_i, e_i) = (R^{N}(e_i, e_k)e_i)^{\perp} = \langle R^{N}(e_i, e_k)e_i, \xi\rangle,$$

into (7) and using normal coordinates at a point, we have

$$
\begin{aligned}
(\Delta^{\varphi}\xi)^{\top} &= \sum_{i,k=1}^{m} ((\nabla_{e_i} b)(e_k e_i))e_k \\
&= \sum_{k=1}^{m}\Big(\sum_{i=1}^{m}(\nabla_{e_k} b)(e_i, e_i) - \mathrm{Ric}(\xi, e_k)\Big)e_k = m\,\mathrm{grad}(H) - (\mathrm{Ric}(\xi, e_k))e_k.
\end{aligned}
$$

Therefore, by collecting all the tangent and normal parts of the bitension field separately, we finally have

$$
\begin{aligned}
(\tau^2(\varphi))^{\perp} &= \langle \tau^2(\varphi), \xi\rangle\xi = m(\Delta H - H|A|^2 + H\,\mathrm{Ric}^{N}(\xi, \xi))\xi, \\
(\tau^2(\varphi))^{\top} &= \sum_{k=1}^{m}\langle \tau^2(\varphi), e_k\rangle e_k \\
&= -m\big(2A(\mathrm{grad}\, H) + \tfrac{1}{2}m(\mathrm{grad}\, H^2) - 2H(\mathrm{Ric}(\xi))^{\top}\big). \qquad \square
\end{aligned}
$$

As an immediate consequence of Theorem 2.1 is this:

Corollary 2.2. *A constant mean curvature hypersurface in a Riemannian manifold is biharmonic if and only if it is minimal or* $\mathrm{Ric}^{N}(\xi, \xi) = |A|^2$ *and* $(\mathrm{Ric}^{N}(\xi))^{\top} = 0$. *In particular, we recover* [Oniciuc 2002, Proposition 2.4], *which states that a constant mean curvature hypersurface in a Riemannian manifold* (N^{m+1}, h) *with nonpositive Ricci curvature is biharmonic if and only if it is minimal.*

Corollary 2.3. *A hypersurface in an Einstein space* (N^{m+1}, h) *is biharmonic if and only if its mean curvature function* H *is a solution of the PDEs*

$$
(8)\quad
\begin{aligned}
\Delta H - H|A|^2 + \frac{rH}{m+1} &= 0, \\
2A(\mathrm{grad}\, H) + \tfrac{1}{2}m\,\mathrm{grad}\, H^2 &= 0,
\end{aligned}
$$

where r *is the scalar curvature of the ambient space. In particular, a hypersurface* $\varphi : (M^m, g) \to (N^{m+1}(C), h)$ *in a space of constant sectional curvature* C *is* *biharmonic if and only if its mean curvature function H is a solution of the PDEs* [Jiang 1987; Chen 1991; Caddeo et al. 2002]

$$\begin{aligned} &\Delta H - H|A|^2 + mCH = 0, \\ &2A(\operatorname{grad} H) + \tfrac{1}{2} m \operatorname{grad} H^2 = 0. \end{aligned} \tag{9}$$

Proof. It is well known that if (N^{m+1}, h) is an Einstein manifold, then

$$\mathrm{Ric}^N(Z, W) = \frac{r}{m+1} h(Z, W) \quad \text{for any } Z, W \in TN$$

and hence $(\mathrm{Ric}^N(\xi))^\top = 0$ and $\mathrm{Ric}^N(\xi, \xi) = r/(m+1)$. From these and (5) we obtain (8). When $(N^{m+1}(C), h)$ is a space of constant sectional curvature C, it is an Einstein space with scalar curvature $r = m(m+1)C$. Substituting this into (8) we obtain (9). □

Theorem 2.4. *A totally umbilical hypersurface in an Einstein space with nonpositive scalar curvature is biharmonic if and only if it is minimal.*

Proof. Take an orthonormal frame $\{e_1, \dots, e_m, \xi\}$ of (N^{m+1}, h) adapted to the hypersurface M so that $Ae_i = \lambda_i e_i$, where A is the Weingarten map of the hypersurface and λ_i is the principal curvature in the direction e_i. Since M is assumed to be totally umbilical, all principal normal curvatures at any point $p \in M$ are equal to the same number $\lambda(p)$. It follows that

$$H = \frac{1}{m} \sum_{i=1}^{m} \langle Ae_i, e_i \rangle = \lambda, \quad |A|^2 = m\lambda^2,$$

$$A(\operatorname{grad} H) = A\Big(\sum_{i=1}^{m} (e_i\lambda) e_i\Big) = \tfrac{1}{2} \operatorname{grad} \lambda^2,$$

The biharmonic hypersurface equations (8) become

$$\Delta\lambda - m\lambda^3 + \frac{r\lambda}{m+1} = 0 \quad \text{and} \quad (2+m) \operatorname{grad} \lambda^2 = 0.$$

Solving these, we have either $\lambda = 0$ and hence $H = 0$, or $\lambda = \pm\sqrt{r/(m(m+1))}$ is a constant. The latter happens only if the scalar curvature is nonnegative, from which we obtain the theorem. □

Remark 2.5. Theorem 2.4 generalizes the results of [Balmuş et al. 2008; Caddeo et al. 2002; Dimitrić 1992] about totally umbilical biharmonic hypersurfaces in a space form. It also implies that the generalized Chen conjecture is true for totally umbilical hypersurfaces in an Einstein space with nonpositive scalar curvature. Note that nonpositive scalar curvature is a much weaker condition than nonpositive sectional curvature.

Corollary 2.6. *Any totally umbilical biharmonic hypersurface in a Ricci flat manifold is minimal.*

Proof. This follows from Theorem 2.4 and the fact that a Ricci flat manifold is an Einstein space with zero scalar curvature. □

3. Proper biharmonic foliations of codimension one

In general, proper biharmonic maps as local solutions of a system of fourth order PDEs are extremely difficult to unearth. Even in the case of biharmonic submanifolds (viewed as biharmonic maps with geometric constraints), few examples have been found. In this section, we construct families of metrics that turn some foliations of hypersurfaces into proper biharmonic foliations, thus providing infinitely many proper biharmonic hypersurfaces.

Theorem 3.1. *For any constant C, let $N = \{(x_1, \dots, x_m, z) \in \mathbb{R}^{m+1} \mid z > -C\}$ denote the upper half space. Then, the conformally flat space*

$$\Big(N, h = f^{-2}(z)\Big(\sum_{i=1}^{m} \mathrm{d}x_i{}^2 + \mathrm{d}z^2\Big)\Big)$$

is foliated by proper biharmonic hyperplanes $z = k$, where $k \in \mathbb{R}$ and $k > -C$, if and only if $f(z) = D/(z + E)$, where $E \geq C$ and $D \in \mathbb{R} \setminus \{0\}$.

Proof. Consider the isometric immersion

$$\varphi : (\mathbb{R}^m, g) \to \Big(\mathbb{R}^{m+1}, h = f^{-2}(z)\Big(\sum_{i=1}^{m} \mathrm{d}x_i{}^2 + \mathrm{d}z^2\Big)\Big)$$

with $\varphi(x_1, \dots, x_m) = (x_1, \dots, x_m, k)$ and k being a constant, where the induced metric g with respect to the natural frame $\partial_i = \partial/\partial x_i$ for $i = 1, 2, \dots, m$ and $\partial_{m+1} = \partial/\partial z$ has components

$$g_{ij} = g(\partial_i, \partial_j) = h(d\varphi(\partial_i), d\varphi(\partial_j)) \circ \varphi = \begin{cases} f^{-2}(k) & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

One can check that $e_A = f(z)\partial_A$, where $A = 1, 2, \dots, m, m + 1$, constitutes a local orthonormal frame on $\mathbb{R}^{m+1}$ adapted to the hypersurface $z = k$ with $\xi = e_{m+1}$ being the unit normal vector field. A straightforward computation using Koszul's formula gives the coefficients of the Levi-Civita connection of the ambient space:

$$(\nabla_{e_A} e_B) = \begin{pmatrix} f'e_{m+1} & 0 & \cdots & 0 & -f'e_1 \\ 0 & f'e_{m+1} & \cdots & 0 & -f'e_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & f'e_{m+1} & -f'e_m \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}_{(m+1)\times(m+1)}$$

Noting that $\xi = e_{m+1}$ is the unit normal vector field, we can easily compute the components of the second fundamental form as

$$h(e_i, e_j) = \langle \nabla_{e_i} e_j, e_{m+1} \rangle = \begin{cases} f' & \text{if } i = j = 1, 2, \dots, m, \\ 0 & \text{otherwise,} \end{cases}$$

from which we conclude that each of the hyperplanes $z = k$ is a totally umbilical hypersurface in the conformally flat space.

We compute the mean curvature of the hypersurface and the norm of the second fundamental form to be

$$H = \frac{1}{m} \sum_{i=1}^{m} h(e_i, e_i) = f' \quad \text{and} \quad |A|^2 = \sum_{i=1}^{m} |h(e_i, e_i)|^2 = m f'^2.$$

Since H depends only on z, we have $\operatorname{grad}_g H = \sum_{i=1}^{m} e_i(H) e_i = 0$ and hence $\Delta_g H = \operatorname{div}(\operatorname{grad}_g H) = 0$. Therefore, by Theorem 2.1, the biharmonic equation of the isometric immersion reduces to the system

$$(10) \qquad -|A|^2 + \operatorname{Ric}^N(\xi, \xi) = 0 \quad \text{and} \quad \sum_{i=1}^{m} (\operatorname{Ric}^N(\xi, e_i)) e_i = 0.$$

We can compute the Ricci curvature of the ambient space:

$$\operatorname{Ric}(e_i, \xi) = \operatorname{Ric}(e_i, e_{m+1}) = \sum_{j=1}^{m} \langle R(e_{m+1}, e_j) e_j, e_i \rangle = 0 \quad \text{for all } i = 1, 2, \dots, m,$$

$$\operatorname{Ric}(\xi, \xi) = \operatorname{Ric}(e_{m+1}, e_{m+1}) = \sum_{j=1}^{m} \langle R(e_{m+1}, e_j) e_j, e_{m+1} \rangle = m f f'' - m f'^2.$$

Substituting these into the system (10), we conclude that all isometric immersions $\varphi : \mathbb{R}^m \to (\mathbb{R}^{m+1}, h = f^{-2}(z)(\sum_{i=1}^{m} \mathrm{d}x_i^2 + \mathrm{d}z^2))$ with $\varphi(x_1, \dots, x_m) = (x_1, \dots, x_m, k)$ are biharmonic if and only if $f f'' - 2 f'^2 = 0$. This equation can be written as $(f'/f)' - (f'/f)^2 = 0$. This ordinary differential equation has solution $f(z) = D/(z + C)$, where C and D are constants. Since the mean curvature of the hypersurface $H = f'(k)$ is never zero, we conclude that each of the hyperplanes $z = k$ for $k \neq -C$ is a proper biharmonic hypersurface in the conformally flat space $(N, h = ((z + C)/D)^2 (\sum_{i=1}^{m} \mathrm{d}x_i^2 + \mathrm{d}z^2))$. □

Theorem 3.2. *The isometric immersion*

$$\varphi : \mathbb{R}^2 \to (\mathbb{R}^3, h = e^{2p(z)} \mathrm{d}x^2 + e^{2q(z)} \mathrm{d}y^2 + \mathrm{d}z^2)$$

with $\varphi(x, y) = (x, y, c)$ is biharmonic if and only if

$$(11) \qquad p'' + 2p'^2 + q'' + 2q'^2 = 0.$$

In particular, for any positive constants A, B, C, D, *the upper half space* $\mathbb{R}^3_+ = \{(x, y, z) \mid z > 0\}$ *with metric* $h = (Az + B)\mathrm{d}x^2 + (Cz + D)\mathrm{d}y^2 + \mathrm{d}z^2$ *is foliated by proper biharmonic planes* $z =$ constant.

Proof. Let φ be as stated, with c being a positive constant. Using the notation $\partial_1 = \partial/\partial x$, $\partial_2 = \partial/\partial y$ and $\partial_3 = \partial/\partial z$ we can easily check that the induced metric is given by

$$\begin{aligned} g_{11} &= g(\partial_1, \partial_1) = h(d\varphi(\partial_1), d\varphi(\partial_1)) \circ \varphi = e^{2p(c)}, \\ g_{12} &= g(\partial_1, \partial_2) = h(d\varphi(\partial_1), d\varphi(\partial_2)) \circ \varphi = 0, \\ g_{22} &= g(\partial_2, \partial_2) = h(d\varphi(\partial_2), d\varphi(\partial_2)) \circ \varphi = e^{2q(c)}. \end{aligned}$$

One can also check that $e_1 = e^{-p(z)}\partial_1$, $e_2 = e^{-q(z)}\partial_2$ and $e_3 = \partial_3$ constitute an orthonormal frame on $\mathbb{R}^3_+$ adapted to the surface $z = c$, with $\xi = e_3$ being the unit normal vector field. A further computation gives the Lie brackets

$$[e_1, e_2] = 0, \quad [e_1, e_3] = p' e_1, \quad [e_2, e_3] = q' e_2, \tag{12}$$

and the coefficients of the Levi-Civita connection:

$$\begin{aligned} &\nabla_{e_1} e_1 = -p' e_3, && \nabla_{e_1} e_2 = 0, && \nabla_{e_1} e_3 = p' e_1, \\ &\nabla_{e_2} e_1 = 0, && \nabla_{e_2} e_2 = -q' e_3, && \nabla_{e_2} e_3 = q' e_2, \\ &\nabla_{e_3} e_1 = 0, && \nabla_{e_3} e_2 = 0, && \nabla_{e_3} e_3 = 0. \end{aligned} \tag{13}$$

Since that $\xi = e_3$ is the unit normal vector field, the components of the second fundamental form are

$$\begin{aligned} h(e_1, e_1) &= \langle \nabla_{e_1} e_1, e_3 \rangle = -p', \\ h(e_1, e_2) &= \langle \nabla_{e_1} e_2, e_3 \rangle = 0, \\ h(e_2, e_2) &= \langle \nabla_{e_2} e_2, e_3 \rangle = -q'. \end{aligned}$$

From these, the mean curvature of the isometric immersion is

$$H = \tfrac{1}{2}(h(e_1, e_1) + h(e_2, e_2)) = -(p' + q')/2, \tag{14}$$

and the norm of the second fundamental form is

$$|A|^2 = \sum_{i=1}^{2} |h(e_i, e_i)|^2 = p'^2 + q'^2.$$

Since H depends only on z we have $\operatorname{grad}_g H = e_1(H)e_1 + e_2(H)e_2 = 0$ and hence $\Delta_g H = \operatorname{div}(\operatorname{grad}_g H) = 0$. Therefore, by Theorem 2.1, the biharmonic equation of the isometric immersion reduces to (10) with $m = 2$. To compute the Ricci curvature of the ambient space we can use (12) and (13) to get

$$\begin{aligned}\mathrm{Ric}(e_1,\xi)&=\mathrm{Ric}(e_1,e_3)=\langle R(e_3,e_2)e_2,e_1\rangle=0,\\ \mathrm{Ric}(e_2,\xi)&=\mathrm{Ric}(e_2,e_3)=\langle R(e_3,e_1)e_1,e_3\rangle=0,\\ \mathrm{Ric}(\xi,\xi)&=\mathrm{Ric}(e_3,e_3)=\langle R(e_3,e_1)e_1,e_3\rangle+\langle R(e_3,e_2)e_2,e_3\rangle\\ &=-p''-p'^2-q''-q'^2.\end{aligned}$$

Substituting these into (10) with $m=2$, we conclude that φ is biharmonic if and only if (11) holds, which gives the theorem's first statement. The second is obtained by looking for the solutions of (11) satisfying $p''+2p'^2=0$ and $q''+2q'^2=0$. In fact, we have special solutions $p(z)=\frac{1}{2}\ln(Az+B)$ and $q(z)=\frac{1}{2}\ln(Cz+D)$ with positive constants A, B, C, D. By (14) and the choice of these constants, we see that the mean curvature of the surface $z=c$ is

$$H=-\frac{2ACz+AD+BC}{2(Az+B)(Cz+D)}\neq 0,$$

and hence each such surface is a nonminimal biharmonic surface. □

Remark 3.3. Theorem 3.2 has a generalization to a higher dimensional space $\mathbb{R}^m_+$ for $m>3$.

Example 3.4. Let $\lambda(t)=\sqrt{At+B}$, where A and B are positive constants. Then the warped product space $N=(S^2\times\mathbb{R}^+, h=\lambda^2(t)g^{S^2}+\mathrm{d}t^2)$ is foliated by the spheres $(S^2\times\{t\}, \lambda^2(t)g^{S^2})$, each of which is a totally umbilical proper biharmonic surface.

To see what is claimed in Example 3.4, we parametrize the unit sphere S^2 by spherical polar coordinates:

$$\mathbb{R}\times\mathbb{R}\ni(\rho,\theta)\to(\cos\rho, \sin\rho\cos\theta, \sin\rho\sin\theta))\in\mathbb{R}^3.$$

Then, the standard metric can be written as $g^{S^2}=\mathrm{d}\rho^2+\sin^2\rho\,\mathrm{d}\theta^2$, and the warped product metric on N takes the form $h=\lambda^2(t)\mathrm{d}\rho^2+\lambda^2(t)\sin^2\rho\,\mathrm{d}\theta^2+\mathrm{d}t^2$. Consider the isometric immersion $\varphi: S^2\to(\mathbb{R}^+\times S^2, \mathrm{d}t^2+\lambda^2(t)g^{S^2})$ with $\varphi(\rho,\theta)=(\rho,\theta,c)$ and c being a positive constant. Using the notation $\partial_1=\partial/\partial\rho$, $\partial_2=\partial/\partial\theta$ and $\partial_3=\partial/\partial t$, we can easily check that the induced metric is given by

$$\begin{aligned}g_{11}&=g(\partial_1,\partial_1)=h(d\varphi(\partial_1), d\varphi(\partial_1))\circ\varphi=\lambda^2(c),\\ g_{12}&=g(\partial_1,\partial_2)=h(d\varphi(\partial_1), d\varphi(\partial_2))\circ\varphi=0,\\ g_{22}&=g(\partial_2,\partial_2)=h(d\varphi(\partial_2), d\varphi(\partial_2))\circ\varphi=\lambda^2(c)\sin^2\rho.\end{aligned}$$

Using the orthonormal frame $e_1=\lambda^{-1}(t)\partial_1$, $e_2=(\lambda(t)\sin\rho)^{-1}\partial_2$ and $e_3=\partial_3$, we have the Lie brackets

$$[e_1,e_2]=-(\cot\rho/\lambda)e_2,\quad [e_1,e_3]=fe_1,\quad [e_2,e_3]=fe_2,$$

where here and in the sequel we use the notation $f = (\ln \lambda)' = \lambda'/\lambda$. Clearly, e_1, e_2 and $\xi = e_3 = \partial_3$ constitute a local orthonormal frame of N adapted to the surface with ξ being the unit vector field normal to the surface. We can use the Kozsul formula to compute the components of the second fundamental form as

$$\begin{aligned} h(e_1, e_1) &= \langle \nabla_{e_1} e_1, \xi \rangle = \langle \nabla_{e_1} e_1, e_3 \rangle \\ &= \tfrac{1}{2}(-\langle e_1, [e_1, e_3]\rangle - \langle e_1, [e_1, e_3]\rangle + \langle e_3, [e_1, e_1]\rangle) = -f, \\ h(e_1, e_2) &= \langle \nabla_{e_1} e_2, \xi \rangle = \langle \nabla_{e_1} e_2, e_3 \rangle = 0, \\ h(e_2, e_2) &= \langle \nabla_{e_2} e_2, \xi \rangle = \langle \nabla_{e_2} e_2, e_3 \rangle = -f, \end{aligned}$$

from which we conclude that each such sphere is a totally umbilical surface in N.

The mean curvature of the isometric immersion and the norm of the second fundamental form are

$$H = \tfrac{1}{2}(h(e_1, e_1) + h(e_2, e_2)) = -f \quad \text{and} \quad |A|^2 = \sum_{i=1}^{2} |h(e_i, e_i)|^2 = 2f^2,$$

which depend only on t. It follows that $\operatorname{grad}_g H = 0$ and $\Delta_g H = 0$. Therefore, by Theorem 2.1, the proper biharmonic equation of φ reduces to (10) with $m = 2$.

On the other hand, using the Ricci curvature formula (for example [Besse 2008]) of the warped product $M = B \times_\lambda F$, we have

$$\begin{aligned} &\operatorname{Ric}(e_1, \xi) = \operatorname{Ric}(e_1, e_3) = 0, \qquad \operatorname{Ric}(e_2, \xi) = \operatorname{Ric}(e_2, e_3) = 0, \\ &\operatorname{Ric}(\xi, \xi) = \operatorname{Ric}(e_3, e_3) = \operatorname{Ric}^{\mathbb{R}}(e_3, e_3) - (2/\lambda)\operatorname{Hess}_\lambda(e_3, e_3) \\ &\qquad = -(2/\lambda)(e_3(e_3\lambda) - d\lambda(\nabla_{e_3} e_3)) = -2\lambda''/\lambda. \end{aligned}$$

Substituting these into (10) with $m = 2$ we conclude that the isometric immersion $\varphi : S^2 \to (S^2 \times \mathbb{R}^+, \lambda^2(t) g^{S^2} + dt^2)$ with $\varphi(\rho, \theta) = (\rho, \theta, c)$ is biharmonic if and only if $-2(\lambda'/\lambda)^2 - 2\lambda''/\lambda = 0$. Solving this final equation, we have $\lambda(t) = \sqrt{At + B}$, proving the claim in Example 3.4.

Remark 3.5. The referee points out that the biharmonicity of the inclusion maps in Example 3.4 is in fact a special case of [Balmuş et al. 2007, Corollary 3.4], which was proved by a different method.

4. Biharmonic cylinders of a Riemannian submersion

Let $\pi : (M^3, g) \to (N^2, h)$ be a Riemannian submersion with totally geodesic fibers from a complete manifold. Let $\alpha : I \to (N^2, h)$ be an immersed regular curve parametrized by arclength. Then $\Sigma = \bigcup_{t \in I} \pi^{-1}(\alpha(t))$ is a surface in M that can be viewed as a disjoint union of all horizontal lifts of the curve α. Let $\{\bar{X} = \alpha', \bar{\xi}\}$ be a Frenet frame along α, and let $\bar{\kappa}$ be the geodesic curvature of the curve. Then the Frenet formula for α is given by

$$\tilde{\nabla}_{\overline{X}}\overline{X} = \bar{\kappa}\bar{\xi},$$
$$\tilde{\nabla}_{\overline{X}}\bar{\xi} = -\bar{\kappa}\overline{X},$$

where $\tilde{\nabla}$ denotes the Levi-Civita connection of (N, h). Let $\beta : I \to (M^3, g)$ be a horizontal lift of α. Let X and ξ be the horizontal lifts of $\overline{X}$ and $\bar{\xi}$, respectively. Let V be the unit vector field tangent to the fibers of the submersion π. Then $\{X, \xi, V\}$ is an orthonormal frame of M adapted to the surface, with ξ the unit normal vector of the surface. The restriction of this frame to the curve β is the Frenet frame along β. Therefore, the Frenet formula along β is given by

$$\begin{aligned} \nabla_X X &= \kappa\xi, \\ \nabla_X \xi &= -\kappa X + \tau V, \\ \nabla_X V &= -\tau\xi, \end{aligned} \tag{15}$$

where ∇ denotes the Levi-Civita connection of (M, g). Since a Riemannian submersion preserves the inner product of horizontal vector fields, we can check that $\kappa = \bar{\kappa} \circ \pi$ and $\tau = \langle \nabla_X \xi, V \rangle = \langle A_X \xi, V \rangle$ is the torsion of the horizontal lift that vanishes if the Riemannian submersion has integrable horizontal distribution; here A is the A-tensor of the submersion [O'Neill 1966]. In what follows, we will use the frame $\{X, \xi, V\}$ to compute the mean curvature, second fundamental form, and other terms that appear in the biharmonic equation of the surface Σ.

Using (15) we have

$$\begin{aligned}
A(X) &= -\langle \nabla_X \xi, X \rangle X - \langle \nabla_X \xi, V \rangle V = \kappa X - \tau V, \\
A(V) &= -\langle \nabla_V \xi, X \rangle X - \langle \nabla_V \xi, V \rangle V = -\tau X, \\
b(X, X) &= \langle A(X), X \rangle = \kappa, \qquad b(X, V) = \langle A(X), V \rangle = -\tau, \\
b(V, X) &= \langle A(V), X \rangle = -\tau, \qquad b(V, V) = \langle A(V), V \rangle = 0; \\
H &= \tfrac{1}{2}(b(X, X) + b(V, V)) = \tfrac{1}{2}\kappa, \\
A(\operatorname{grad} H) &= A(X(\tfrac{1}{2}\kappa)X + V(\tfrac{1}{2}\kappa)V) = X(\tfrac{1}{2}\kappa)A(X) = \tfrac{1}{2}\kappa'(\kappa X - \tau V); \\
\Delta H &= XX(H) - (\nabla_X X)H + VV(H) - (\nabla_V V)H = \tfrac{1}{2}\kappa''; \\
|A|^2 &= (b(X, X))^2 + (b(X, V))^2 + (b(V, X))^2 + (b(V, V))^2 = \kappa^2 + 2\tau^2.
\end{aligned}$$

Substituting these into the biharmonic hypersurface equation (5), we conclude that the surface Σ is biharmonic in (M^3, g) if and only if

$$\begin{aligned}
\tfrac{1}{2}\kappa'' - \tfrac{1}{2}\kappa(\kappa^2 + 2\tau^2) + \tfrac{1}{2}\kappa \operatorname{Ric}^M(\xi, \xi) &= 0, \\
\kappa'(\kappa X - \tau V) + \tfrac{1}{2}\kappa\kappa' X - \kappa \operatorname{Ric}^M(\xi, X)X - \kappa \operatorname{Ric}^M(\xi, V)V &= 0.
\end{aligned}$$

These are equivalent to

$$
\begin{aligned}
&\kappa'' - \kappa(\kappa^2 + 2\tau^2) + \kappa \operatorname{Ric}^M(\xi, \xi) = 0,\\
(16)\qquad &3\kappa'\kappa - 2\kappa \operatorname{Ric}^M(\xi, X) = 0,\\
&\kappa'\tau + \kappa \operatorname{Ric}^M(\xi, V) = 0.
\end{aligned}
$$

Applying (16) to Hopf fibration $\pi : S^3 \to S^2$ we have the following corollary, which recovers [Inoguchi 2004, Proposition 3.1].

Corollary 4.1. *There is no proper biharmonic Hopf cylinder in S^3.*

Finally, applying (16) to the submersions $\pi : S^2 \times \mathbb{R} \to S^2$ and $\pi : H^2 \times \mathbb{R} \to H^2$ yields another corollary:

Corollary 4.2. (1) *The Hopf cylinder $\Sigma = \bigcup_{t\in I} \pi^{-1}(\alpha(t))$ is a proper biharmonic surface in $S^2 \times \mathbb{R}$ if and only if the directrix $\alpha : I \to (S^2, h)$ is a part of a circle in S^2 with radius $\sqrt{2}/2$;*

(2) *The Hopf cylinder $\Sigma = \bigcup_{t\in I} \pi^{-1}(\alpha(t))$ is biharmonic in $H^2 \times \mathbb{R}$ if and only if it is minimal.*

Acknowledgment

I would like to thank the referee for many invaluable comments and suggestions that helped to clarify and improve the original manuscript.

References

[Arslan et al. 2005] K. Arslan, R. Ezentas, C. Murathan, and T. Sasahara, "Biharmonic submanifolds in 3-dimensional (κ, μ)-manifolds", *Int. J. Math. Math. Sci.* 22 (2005), 3575–3586. MR 2006k:53108 Zbl 1096.53031

[Baird and Wood 2003] P. Baird and J. C. Wood, *Harmonic morphisms between Riemannian manifolds*, London Mathematical Society Monographs (N.S.) **29**, Oxford University Press, 2003. MR 2005b:53101 Zbl 1055.53049

[Balmuş et al. 2007] A. Balmuş, S. Montaldo, and C. Oniciuc, "Biharmonic maps between warped product manifolds", *J. Geom. Phys.* **57**:2 (2007), 449–466. MR 2007j:53066 Zbl 1108.58011

[Balmuş et al. 2008] A. Balmuş, S. Montaldo, and C. Oniciuc, "Classification results for biharmonic submanifolds in spheres", *Israel J. Math.* **168** (2008), 201–220. MR 2009j:53079 Zbl 1172.58004

[Besse 2008] A. L. Besse, *Einstein manifolds*, Springer, Berlin, 2008. MR 2008k:53084 Zbl 1147.53001

[Caddeo et al. 2001] R. Caddeo, S. Montaldo, and C. Oniciuc, "Biharmonic submanifolds of S^3", *Internat. J. Math.* **12**:8 (2001), 867–876. MR 2002k:53123 Zbl 1111.53302

[Caddeo et al. 2002] R. Caddeo, S. Montaldo, and C. Oniciuc, "Biharmonic submanifolds in spheres", *Israel J. Math.* **130** (2002), 109–123. MR 2003c:53090 Zbl 1038.58011

[Chen 1991] B.-Y. Chen, "Some open problems and conjectures on submanifolds of finite type", *Soochow J. Math.* **17**:2 (1991), 169–188. MR 92m:53091 Zbl 0749.53037

[Chen and Ishikawa 1998] B.-Y. Chen and S. Ishikawa, "Biharmonic pseudo-Riemannian submanifolds in pseudo-Euclidean spaces", *Kyushu J. Math.* **52**:1 (1998), 167–185. MR 99b:53078 Zbl 0892.53012

[Dimitrić 1992] I. Dimitrić, "Submanifolds of E^m with harmonic mean curvature vector", *Bull. Inst. Math. Acad. Sinica* **20**:1 (1992), 53–65. MR 93g:53087 Zbl 0778.53046

[Fetcu and Oniciuc 2009a] D. Fetcu and C. Oniciuc, "Explicit formulas for biharmonic submanifolds in Sasakian space forms", *Pacific J. Math.* **240**:1 (2009), 85–107. MR 2010c:53092 Zbl 1167.53024

[Fetcu and Oniciuc 2009b] D. Fetcu and C. Oniciuc, "On the geometry of biharmonic submanifolds in Sasakian space forms", *J. Geom. Symmetry Phys.* **14** (2009), 21–34. MR 2536497 Zbl 05580885

[Hasanis and Vlachos 1995] T. Hasanis and T. Vlachos, "Hypersurfaces in E^4 with harmonic mean curvature vector field", *Math. Nachr.* **172** (1995), 145–169. MR 96c:53085 Zbl 0839.53007

[Inoguchi 2004] J.-I. Inoguchi, "Submanifolds with harmonic mean curvature vector field in contact 3-manifolds", *Colloq. Math.* **100**:2 (2004), 163–179. MR 2005h:53105 Zbl 1076.53065

[Javaloyes and Meroño 2003] M. A. Javaloyes Victoria and M. A. Meroño Bayo, "Biharmonic lifts by means of pseudo-Riemannian submersions in dimension three", *Trans. Amer. Math. Soc.* **355**:1 (2003), 169–176. MR 2003f:53107 Zbl 1020.53033

[Jiang 1986a] G. Y. Jiang, "2-harmonic isometric immersions between Riemannian manifolds", *Chinese Ann. Math. Ser. A* **7**:2 (1986), 130–144. In Chinese; English summary in *Chinese Ann. Math. Ser. B* **7**:2 (1986), 255. MR 87k:53140 Zbl 0596.53046

[Jiang 1986b] G. Y. Jiang, "2-harmonic maps and their first and second variational formulas", *Chinese Ann. Math. Ser. A* **7**:4 (1986), 389–402. In Chinese; English summary in *Chinese Ann. Math. Ser. B* **7**:4 (1986), 523. MR 88i:58039 Zbl 0628.58008

[Jiang 1987] G. Y. Jiang, "Some nonexistence theorems on 2-harmonic and isometric immersions in Euclidean space", *Chinese Ann. Math. Ser. A* **8**:3 (1987), 377–383. In Chinese; English summary in *Chinese Ann. Math. Ser. B* **8**:3 (1987), 389. MR 89a:53071 Zbl 0637.53071

[Kobayashi and Nomizu 1969] S. Kobayashi and K. Nomizu, *Foundations of differential geometry*, vol. 2, Pure and Applied Mathematics **15**, Wiley, New York, 1969. MR MR0238225 (38 #6501) Zbl 0175.48504

[Montaldo and Oniciuc 2006] S. Montaldo and C. Oniciuc, "A short survey on biharmonic maps between Riemannian manifolds", *Rev. Un. Mat. Argentina* **47**:2 (2006), 1–22. MR 2008a:53063 Zbl 1140.58004

[Nishikawa and Maeda 1974] S. Nishikawa and Y. Maeda, "Conformally flat hypersurfaces in a conformally flat Riemannian manifold", *Tôhoku Math. J.* (2) **26** (1974), 159–168. MR 49 #3730 Zbl 0278.53018

[O'Neill 1966] B. O'Neill, "The fundamental equations of a submersion", *Michigan Math. J.* **13** (1966), 459–469. MR 34 #751 Zbl 0145.18602

[Oniciuc 2002] C. Oniciuc, "Biharmonic maps between Riemannian manifolds", *An. Ştiinţ. Univ. Al. I. Cuza Iaşi. Mat. (N.S.)* **48**:2 (2002), 237–248 (2003). MR 2004e:53097 Zbl 1061.58015

[Ou 2009] Y.-L. Ou, "On conformal biharmonic immersions", *Ann. Global Anal. Geom.* **36**:2 (2009), 133–142. MR 2010g:53118 Zbl 1178.53054

[Sasahara 2005] T. Sasahara, "Legendre surfaces in Sasakian space forms whose mean curvature vectors are eigenvectors", *Publ. Math. Debrecen* **67**:3-4 (2005), 285–303. MR 2006c:53064

[Sasahara 2008] T. Sasahara, "Stability of biharmonic Legendrian submanifolds in Sasakian space forms", *Canad. Math. Bull.* **51**:3 (2008), 448–459. MR 2009c:53078 Zbl 1147.53315

[Zhang 2008] W. Zhang, "New examples of biharmonic submanifolds in $\mathbb{C}P^n$ and S^{2n+1}", preprint, 2008. arXiv 0705.3961

Received July 23, 2009. Revised October 31, 2009.

YE-LIN OU
DEPARTMENT OF MATHEMATICS
TEXAS A&M UNIVERSITY-COMMERCE
COMMERCE TX, 75429
UNITED STATES
yelin_ou@tamu-commerce.edu
www.tamu-commerce.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 248, No. 1, 2010

SINGULAR FIBERS AND 4-DIMENSIONAL COBORDISM GROUP

OSAMU SAEKI

Using the technique of singular fibers of C^∞ stable maps, we give a new proof to the theorem, originally due to Rohlin, that the oriented cobordism group of 4-dimensional manifolds is infinite cyclic and is generated by the cobordism class of the complex projective plane. A byproduct is a new and transparent proof of the signature formula, originally due to T. Yamamoto and the author, for 4-dimensional manifolds in terms of singular fibers.

1. Introduction

Saeki [2004] developed the theory of singular fibers of generic differentiable maps $f : M \to N$ between manifolds with $\dim M > \dim N$. In particular, C^∞ stable maps with $(\dim M, \dim N) = (2, 1)$, $(3, 2)$ and $(4, 3)$ were thoroughly studied and their singular fibers were completely classified up to a natural equivalence. (For precise definitions, see [Golubitsky and Guillemin 1973] and Section 2. For the case of maps between nonorientable manifolds, see [Yamamoto 2006] as well.)

In this paper, we use these classifications of singular fibers to determine the structures of $\mathfrak{N}_2$, Ω_2, Ω_3 and Ω_4, where $\mathfrak{N}_n$ and Ω_n are the cobordism group and oriented cobordism group, respectively, of manifolds of dimension n. These groups were central objects of study in differential topology in the middle of 20th century, and their structures have been completely clarified. Our main objective here is to use classifications of singular fibers of C^∞ stable maps to show that Ω_4 is an infinite cyclic group generated by the cobordism class of the complex projective plane, and that the signature function $\Omega_4 \to \mathbb{Z}$ gives an isomorphism. This theorem is originally due to Rohlin [1952]; see also [Guillou and Marin 1986].

The idea of our proof, which is constructive, is as follows. If we have a smooth map f of a closed oriented 4-manifold M into a 3-manifold N, then a regular fiber is a finite disjoint union of circles. In particular, if f is nonsingular, then M is a circle bundle over a 3-manifold and therefore bounds the associated 2-disk bundle.

MSC2000: primary 57R45; secondary 57R75.
Keywords: singular fiber, stable map, cobordism group, signature, complex projective plane.
This work was supported in part by Grant-in-Aid for Scientific Research (B) Number 19340018, Japan Society for the Promotion of Science.

If f has singularities and is generic enough, then the 4-manifold M is decomposed into several pieces according to the classification of singular fibers. For example, regular fibers correspond to a circle bundle over a 3-manifold. Furthermore, we can find a "canonical" 5-manifold for each such piece, except for that corresponding to specific singular fibers, called singular fibers of type III^8 in [Saeki 2004]. By gluing these 5-manifold pieces to $M \times [0, 1]$, we get a cobordism between M and a finite disjoint union of copies of a certain 4-manifold, each corresponding to a singular fiber of type III^8. In this paper, we will show that this 4-manifold is in fact diffeomorphic to the complex projective plane $\mathbb{C}P^2$ up to orientation. This observation shows that $\mathbb{C}P^2$ is a natural representative of the generator of the 4-dimensional oriented cobordism group Ω_4, since our proof is natural and the appearance of $\mathbb{C}P^2$ is not artificial. We may also say that we give a modern proof of the classical Rohlin's theorem from a singularity theoretical viewpoint.

As a corollary to our argument, we get a new and constructive proof of the signature formula proved in [Saeki and Yamamoto 2006]: the signature of a 4-manifold M coincides with the number of singular fibers of type III^8 counted with signs for any C^∞ stable map $f : M \to N$ into a 3-manifold. The proof depended on the classification of singular fibers of C^∞ stable maps of n-dimensional manifolds into $(n-1)$-dimensional manifolds for $n \leq 5$, whereas our proof here needs only the classification of such singular fibers for $n \leq 4$.

The paper is organized as follows. In Section 2, we review some prerequisite notions about cobordisms of manifolds and singular fibers of differentiable maps. In Section 3, we show that $\mathfrak{N}_2 \cong \mathbb{Z}_2$ and $\Omega_2 = 0$ using the classification of singular fibers of Morse functions on surfaces. We will see that the real projective plane $\mathbb{R}P^2$ is a natural representative of the generator of $\mathfrak{N}_2$ and that the Euler characteristic modulo two gives an isomorphism $\mathfrak{N}_2 \to \mathbb{Z}_2$. Although the argument is quite elementary, the proof will turn out to be a good guideline for the 4-dimensional case. In Section 4, we will show that $\Omega_3 = 0$ using the classification of singular fibers of C^∞ stable maps of 3-manifolds into surfaces. A similar idea has been used by Costantino and D. Thurston [2008] in a proof that every 3-manifold efficiently bounds a 4-manifold. In fact, the idea of this paper is based on their work. In Section 5, we will show that $\Omega_4 \cong \mathbb{Z}$ by using the classification of singular fibers of C^∞ stable maps of 4-manifolds into 3-manifolds. As a corollary, we will also show that if $f : M \to N$ is a generic differentiable map between manifolds with $\dim M - \dim N = 1$ that has only singular fibers of codimension ≤ 3 and no singular fiber of type III^8, then M is null cobordant. We note that the results in this paper depend on a bundle structure theorem for singular fibers of stable maps due to Kalmár [2008, Section 5; 2009, Section 6].

Throughout the paper, all manifolds and maps are differentiable of class C^∞. The symbol $\cong$ denotes a diffeomorphism between manifolds or an appropriate isomorphism between algebraic objects. For a closed surface Σ and a positive integer m, we denote by $\Sigma_{(m)}$ the surface Σ with m open disks removed. We denote by $\mathrm{cl}(A)$ the closure of a subset A of a topological space.

2. Preliminaries

Let n be a nonnegative integer. Two closed oriented (possibly disconnected) n-dimensional manifolds M_0 and M_1 are *oriented cobordant* if there is a compact oriented $(n+1)$-dimensional manifold V such that $\partial V = (-M_0) \cup M_1$ as oriented manifolds, where $-M_0$ denotes the manifold obtained by reversing the orientation of M_0. This defines an equivalence relation on the set of all closed oriented manifolds of dimension n, and the oriented cobordism class of a closed oriented manifold M is denoted by $[M]$.

We denote by Ω_n the set of all oriented cobordism classes of closed oriented n-dimensional manifolds. This clearly forms an additive group under the operation given by $[M]+[M'] = [M \cup M']$. The abelian group Ω_n is called the *n-dimensional oriented cobordism group*.

If we ignore the orientations of the manifolds in these definitions above, then we get the usual notion of a *cobordism*, and the set of all cobordism classes of closed (possibly nonorientable) n-dimensional manifolds is denoted by $\mathfrak{N}_n$, which is called the (*unoriented*) *n-dimensional cobordism group*.

These groups were formulated and studied in the middle of 20th century, and their structures have been completely determined. For example, see [Milnor and Stasheff 1974; Pontryagin 1955; Thom 1954; Wall 1959]. In particular, Ω_n and $\mathfrak{N}_n$ are a finitely generated $\mathbb{Z}$-module and $\mathbb{Z}_2$-module, respectively. (Historically, Pontrjagin [1955] first introduced such groups to compute certain homotopy groups of spheres. Thom [1954] reduced the computation of the cobordism groups to the study of homotopy groups of certain spaces, and then the structures of the cobordism groups have been determined by several authors.)

We now recall some definitions about singular fibers. See [Saeki 2004].

Definition 2.1. Let $f_i : M_i \to N_i$ be maps between manifolds and take points $y_i \in N_i$ for $i = 0, 1$. We say that the fibers over y_0 and y_1 are *C^∞ equivalent* if for some open neighborhoods U_i of y_i in N_i, there exist diffeomorphisms $\tilde{\varphi} : (f_0)^{-1}(U_0) \to (f_1)^{-1}(U_1)$ and $\varphi : U_0 \to U_1$ with $\varphi(y_0) = y_1$ such that the following diagram is commutative:

$$\begin{array}{ccc} ((f_0)^{-1}(U_0), (f_0)^{-1}(y_0)) & \xrightarrow{\tilde{\varphi}} & ((f_1)^{-1}(U_1), (f_1)^{-1}(y_1)) \\ \downarrow f_0 & & \downarrow f_1 \\ (U_0, y_0) & \xrightarrow{\varphi} & (U_1, y_1). \end{array}$$

When $y \in N$ is a regular value of a map $f : M \to N$ between manifolds, we call the C^∞ equivalence class of the fiber over y (or the space $f^{-1}(y)$) a *regular fiber*; otherwise, we call it *singular*.

Given $f : M \to N$ and a point $y \in N$, consider the map $f \times \mathrm{id}_{\mathbb{R}} : M \times \mathbb{R} \to N \times \mathbb{R}$, where $\mathrm{id}_{\mathbb{R}}$ is the identity map of the real line $\mathbb{R}$. Then the fiber of $f \times \mathrm{id}_{\mathbb{R}}$ over the point $(y, 0) \in N \times \mathbb{R}$ is called the *suspension* of the fiber of f over y.

For certain dimension pairs $(\dim M, \dim N)$, singular fibers of C^∞ stable maps (defined below) of M into N have been classified up to C^∞ equivalence. For details, see [Saeki 2004; Yamamoto 2006; Yamamoto 2007].

Definition 2.2. For manifolds M and N, we denote by $C^\infty(M, N)$ the space of all smooth maps of M into N, endowed with the Whitney C^∞ topology. We say that a smooth map $f : M \to N$ is a C^∞ *stable map* if there exists a neighborhood U_f of f in $C^\infty(M, N)$ such that for each $g \in U_f$, the diagram

$$\begin{CD} M @>{\tilde{\psi}}>> M \\ @V{f}VV @VV{g}V \\ N @>{\psi}>> N \end{CD}$$

commutes for some diffeomorphisms $\tilde{\psi}$ and ψ; for details, see [Golubitsky and Guillemin 1973].

It is known that a smooth function $M \to \mathbb{R}$ on a closed manifold M is C^∞ stable if and only if it is a *Morse function*, that is, if and only if its critical points are all nondegenerate and have distinct critical values. Furthermore, if $\dim N \leq 5$, then the set of C^∞ stable maps is open and dense in $C^\infty(M, N)$; see [Mather 1971].

Let us recall the following notion of a Stein factorization, which will play an important role in this paper.

Definition 2.3. Let $f : M \to N$ be a smooth map between smooth manifolds. For two points $x, x' \in M$, we define $x \sim_f x'$ if $f(x) = f(x')$, and the points x and x' belong to the same connected component of a fiber of f. We define $W_f = M/\sim_f$ to be the quotient space with respect to this equivalence relation, and $q_f : M \to W_f$ denotes the quotient map. Then it is easy to see that there exists a unique continuous map $\bar{f} : W_f \to N$ such that the following diagram commutes:

$$\begin{array}{ccc} M & \xrightarrow{\quad f \quad} & N \\ & {\scriptstyle q_f}\searrow \quad \nearrow{\scriptstyle \bar{f}} & \\ & W_f & \end{array}$$

This diagram is called the *Stein factorization* of f; see [Levine 1985].

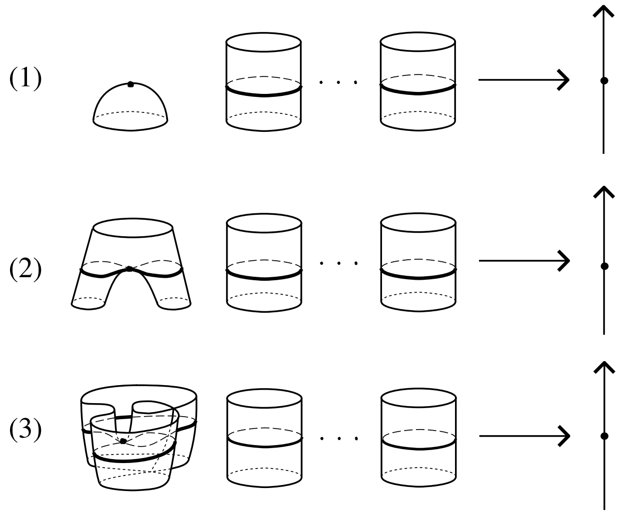


Figure 1. List of C^∞ equivalence classes of singular fibers for Morse functions on surfaces.

If f is a proper C^∞ stable map, then W_f is a polyhedron and all the maps appearing in this diagram are triangulable; for details, see [Hiratuka 2001].

The Stein factorization is a very useful tool for studying topological properties of C^∞ stable maps.

3. Two-dimensional cobordism group

In this section, we show that $\mathfrak{N}_2 \cong \mathbb{Z}_2$ and $\Omega_2 = 0$ using the classification of singular fibers of Morse functions on surfaces.

Let M be an arbitrary closed 2-dimensional manifold, possibly disconnected or nonorientable. It is known that there always exists a Morse function $f : M \to \mathbb{R}$. The singular fibers of such Morse functions are classified as in Figure 1, which may be folklore; for details, see [Saeki 2004].

Construct a compact 3-dimensional manifold V whose boundary includes M by attaching certain pieces to $M \times [0, 1]$, as follows.

Let W_f be the quotient space in the Stein factorization of $f = \bar{f} \circ q_f$. It is a graph whose vertices correspond to connected components of singular fibers: the degree of a vertex is equal to 1, 3 or 2 if it corresponds to the connected component of the singular fiber as (1), (2) or (3) of Figure 1, respectively, containing the critical point; see Figure 2 for an example.

For an edge e of W_f, set $e' = \mathrm{cl}(e \setminus N_0)$, where N_0 is a small regular neighborhood of the set of vertices in W_f and cl denotes the closure in W_f. Since the map

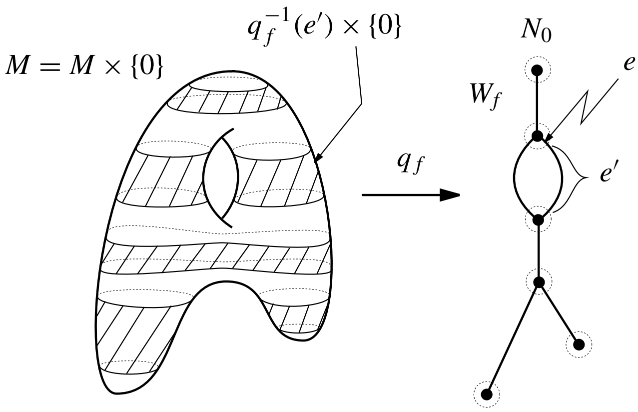


Figure 2. Constructing the 3-manifold V.

q_f restricted to $q_f^{-1}(e')$ is a locally trivial fiber bundle over e' with fiber S^1, it is diffeomorphic to $S^1 \times e'$. Let us glue a 2-handle $D^2 \times e'$ to $M \times [0, 1]$ by identifying $\partial D^2 \times e'$ and $q_f^{-1}(e') \times \{0\}$ by using the diffeomorphism above. Let us perform this operation for each edge of W_f and denote by V the resulting compact 3-dimensional manifold. For an example of the union of $M \times \{0\}$ and the 2-handles, see Figure 2.

We see that the boundary ∂V is a disjoint union of $M \times \{1\}$ and some closed surfaces F_j, where each F_j corresponds to a singular fiber of f. More precisely, let v be a vertex of W_f and $N_0(v)$ its small regular neighborhood in W_f, which is a component of N_0. Then, the corresponding surface F_j is diffeomorphic to the union of $q_f^{-1}(N_0(v))$ and some 2-disks attached to the regular fibers corresponding to $N_0(v) \cap \mathrm{cl}(W_f \setminus N_0)$.

Thus, according to the classification of singular fibers as in Figure 1, we see that each surface F_j is connected and is diffeomorphic to S^2 for the singular fibers as in (1) and (2) of Figure 1, and to $\mathbb{R}P^2$ for that in Figure 1(3). See Figure 3.

Since $S^2 = \partial D^3$ is null cobordant, we have proved the following.

Lemma 3.1. *Every closed surface is cobordant to the disjoint union of a finite number of copies of* $\mathbb{R}P^2$.

The cobordism class $[\mathbb{R}P^2 \cup \mathbb{R}P^2]$ is zero since $\mathbb{R}P^2 \cup \mathbb{R}P^2$ is the boundary of the compact 3-manifold $\mathbb{R}P^2 \times [0, 1]$. Let us consider the homomorphism

$$\varphi : \mathbb{Z}_2 \to \mathfrak{N}_2, \quad 1 \mapsto [\mathbb{R}P^2],$$

where $1 \in \mathbb{Z}_2$ is the generator. This is a well-defined homomorphism, and is surjective by Lemma 3.1.

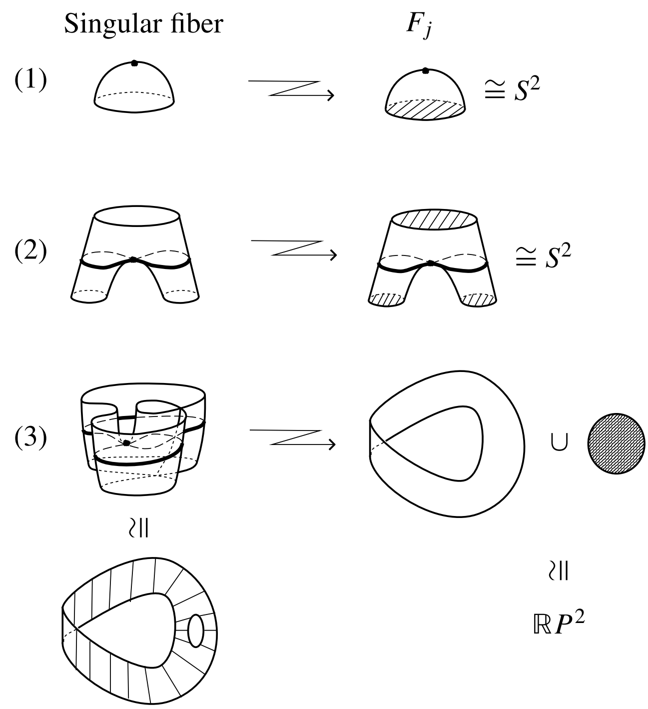


Figure 3. Surface F_j appearing around each singular fiber.

Let

$$\chi_2 : \mathfrak{N}_2 \to \mathbb{Z}_2 \tag{3-1}$$

be the homomorphism defined by associating to each cobordism class the Euler characteristic modulo two of its representative. Using standard techniques in algebraic topology, we can show that this defines a well-defined homomorphism; for example, see [Thom 1952].

Since the Euler characteristic of $\mathbb{R}P^2$ is equal to 1, we see that the composition $\chi_2 \circ \varphi : \mathbb{Z}_2 \to \mathbb{Z}_2$ is the identity. Therefore, φ must be injective. Thus, we have proved the following.

Theorem 3.2. *The* 2*-dimensional cobordism group* $\mathfrak{N}_2$ *is cyclic of order two and is generated by the cobordism class of* $\mathbb{R}P^2$. *In fact*, *the homomorphism* (3-1) *is an isomorphism.*

Our proof does not depend on the classification of closed surfaces.

As a corollary to the proof, we also get the following, which was originally obtained in [Saeki 2004].

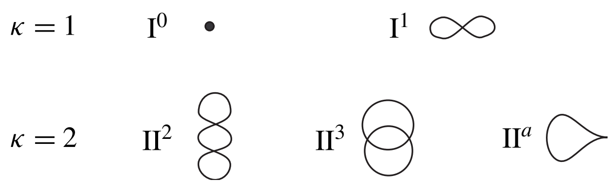


Figure 4. List of C^∞ equivalence classes of singular fibers of C^∞ stable maps of orientable 3-manifolds into surfaces.

Corollary 3.3. *Let $f : M \to \mathbb{R}$ be a Morse function on a closed surface M. Then, the number of singular fibers of f as in* Figure 1(3) *has the same parity as the Euler characteristic of M.*

If a closed surface M is oriented, then a singular fiber as in Figure 1(3) never appears, since its neighborhood is nonorientable. Furthermore, the 3-manifold V constructed above is orientable, since $M \times [0, 1]$ is orientable and attaching a 2-handle does not alter the orientability of a 3-manifold. Moreover, V can be oriented so that its oriented boundary consists of M and some 2-spheres. Thus:

Corollary 3.4. *The* 2*-dimensional oriented cobordism group* Ω_2 *vanishes.*

4. Three-dimensional oriented cobordism group

In this section, we show that $\Omega_3 = 0$ by using the classification of singular fibers of C^∞ stable maps of closed orientable 3-manifolds into surfaces.

For M a closed oriented 3-manifold, there always exists a C^∞ stable map f from M into any surface N; for example, see [Kushner et al. 1984; Levine 1985]. Singular fibers of such maps have been classified in [Saeki 2004] up to C^∞ equivalence; see also [Kushner et al. 1984; Levine 1985]. The connected components of singular fibers containing singular points are as depicted in Figure 4, where each figure represents the connected component of the inverse image of a point in the target. (In fact, it is known that two such singular fibers are C^∞ equivalent if and only if the corresponding inverse images are diffeomorphic to each other. For details, see [Saeki 2004].) Furthermore, in Figure 4, κ denotes the *codimension* of a singular fiber, that is, it is the codimension of the set of those points in the target over which lies a relevant singular fiber. Following the convention introduced in [Saeki 2004], we use the notation I^0, I^1 and so on for the C^∞ equivalence classes of singular fibers. The I^0-type (or I^1-type) singular fiber is the suspension of the singular fiber for Morse functions as in Figure 1(1) (respectively Figure 1(2)).

A C^∞ stable map of a 3-manifold into a surface has only fold and cusp singularities [Kushner et al. 1984; Levine 1985]. A singular fiber (or the corresponding inverse image) can be regarded as a graph: a fold point corresponds to an isolated vertex or to a vertex of degree four, and a cusp point corresponds to a cuspidal vertex of degree two (see Figure 4).

Let W_f be the quotient space in the Stein factorization of a C^∞ stable map f from a closed oriented 3-manifold M into a surface N. The space W_f is a compact 2-dimensional polyhedron and its local structure is completely determined [Kushner et al. 1984; Levine 1985]. Let $W^{(0)}$ denote the q_f-image of the singular fibers of $\kappa = 2$, and let $W^{(1)}$ denote the q_f-image of the singular fibers of $\kappa \geq 1$ (more precisely, they are the q_f-images of the components of the relevant singular fibers containing singular points). Note that $W^{(0)}$ is a finite set of points and $W^{(1)}$ is a 1-dimensional subcomplex of W_f whose complement is a nonsingular surface. For $i = 0, 1$, we denote by $N^{(i)}$ a small regular neighborhood of $W^{(i)}$ in W_f. We set $N_0 = N^{(0)}$, $N_1 = \mathrm{cl}(N^{(1)} \setminus N^{(0)})$ and $N_2 = \mathrm{cl}(W_f \setminus N^{(1)})$, where N_1 is regarded as a regular neighborhood of $\mathrm{cl}(W^{(1)} \setminus N^{(0)})$ in $\mathrm{cl}(W_f \setminus N^{(0)})$. Note that W_f is decomposed as

$$W_f = N_0 \cup N_1 \cup N_2.$$

Let us construct a compact 4-dimensional manifold V by attaching certain pieces to $M \times [0, 1]$ as follows. First note that q_f restricted to $q_f^{-1}(N_2)$ is a locally trivial fibration with fiber S^1 over the surface N_2. Thus, we can attach the total space of the associated D^2-bundle over N_2 to $M \times [0, 1]$ by identifying the associated (∂D^2)-bundle with $q_f^{-1}(N_2) \times \{0\}$. (Here, we use the well-known fact that the structure group of every smooth S^1-bundle can be reduced to the orthogonal group $O(2)$.) The resulting 4-manifold is denoted by V_1. Note that V_1 is orientable, since M and $q_f^{-1}(N_2)$ are orientable as 3-manifolds.

Let $V_1'(\subset V_1)$ denote the union of $M \times \{0\}$ and the D^2-bundle over N_2. There is a natural map $q_1 : V_1' \to W_f$, which on $M \times \{0\}$ is defined by $q_f : M \times \{0\} = M \to W_f$, and on the D^2-bundle is defined by the projection to N_2.

Let e be a connected component of $\mathrm{cl}(W^{(1)} \setminus N_0)$. Note that e is an arc or a circle. Let $N_1(e)$ denote the connected component of N_1 containing e. If the singular fiber lying over a point in e is of type I^0, then $q_f^{-1}(N_1(e))$ is diffeomorphic to the total space of a D^2-bundle over e [Kushner et al. 1984; Levine 1985]. In fact, $N_1(e)$ is diffeomorphic to $e \times [0, 1]$ by a diffeomorphism that induces the identity $e(\subset N_1(e)) \to e \times \{0\}$, and if J $(\cong [0, 1])$ is a fiber of the natural fibration $N_1(e) \to e$, then q_f restricted to $q_f^{-1}(J) \cong D^2$ is equivalent to the function

$$(x, y) \mapsto x^2 + y^2. \tag{4-1}$$

(In fact, the commutative diagram

$$\begin{array}{ccc} q_f^{-1}(N_1(e)) & \xrightarrow{\;q_f\;} & N_1(e) \\ & \searrow \quad \swarrow & \\ & e & \end{array}$$

can be regarded as a fiber bundle with the map $D^2 \to [0, 1]$ defined by (4-1) as fiber in an appropriate sense. For details, see [Kalmár 2008, Section 5; Kalmár 2009, Section 6].)

Therefore, q_1 restricted to $q_1^{-1}(N_1(e))$ followed by the natural projection from $N_1(e)$ to e is an S^2-bundle, where the fiber of this fibration can be identified with the 2-sphere as in Figure 3(1) in Section 3. Then, we can attach the associated D^3-bundle over e to V_1, where we identify the associated (∂D^3)-bundle with $q_1^{-1}(N_1(e))$. Here, we use the fact that the structure group of every smooth S^2-bundle can be reduced to $O(3)$; see [Smale 1959]. Note that the resulting 4-manifold is orientable, since so is V_1 and the orientability of $q_1^{-1}(N_1(e))$ coincides with that of $q_f^{-1}(N_1(e))$ $(\subset M)$.

If the singular fiber lying over a point in e is of type I^1, then the natural projection $N_1(e) \to e$ defines a Y-bundle, where

$$Y = \{r \exp(2\pi\sqrt{-1}k/3) \in \mathbb{C} \mid 0 \le r \le 1,\ k = 0, 1, 2\}.$$

Moreover, $q_f^{-1}(N_1(e))$ is diffeomorphic to the total space of an $S^2_{(3)}$-bundle over e, where $S^2_{(3)}$ denotes the 2-sphere with three open disks removed. Then, q_1 restricted to $q_1^{-1}(N_1(e))$ followed by the natural projection $N_1(e) \to e$ is again an S^2-bundle, but with fiber as in Figure 3(2), and we can attach the associated D^3-bundle.

We perform the operation described above for each e. The resulting 4-manifold, denoted by V, is a compact 4-dimensional manifold that is orientable. Furthermore, it can be oriented so that

$$\partial V = (M \times \{1\}) \cup \big(-\textstyle\bigcup_j F_j\big),$$

where each F_j is a closed oriented 3-manifold corresponding to a singular fiber of f of $\kappa = 2$.

Lemma 4.1. *The closed* 3*-manifold F_j is diffeomorphic to the* 3*-sphere S^3 for every singular fiber of $\kappa = 2$.*

In fact, for the singular fibers of types II^2 and II^3, this lemma has been essentially obtained in [Costantino and Thurston 2008]. Here we give a proof from a different viewpoint in a way that is useful in Section 5.

Proof. Let v be a point in W_f that is the q_f-image of a singular fiber of type II^2. Then, its regular neighborhood $N_0(v)$ in W_f is of the form depicted in Figure 5, where $N_0(v)$ is the component of N_0 containing v. Note that $\bar{f}(N_0(v)) \cong J_1 \times J_2$ with $J_1 = J_2 = [-1, 1]$.

Then, the map f restricted to $q_f^{-1}(N_0(v))$ can be regarded as a 1-parameter family of functions on $S^2_{(4)}$ with only nondegenerate critical points as depicted in Figure 6. This family is parametrized by J_1, where for each parameter value the relevant function is regarded as a height function $S^2_{(4)} \to J_2$.

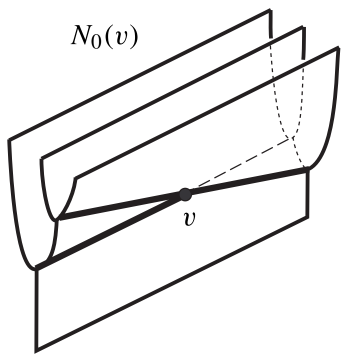


Figure 5. Neighborhood of the q_f-image of a II^2-type singular fiber.

In constructing V_1, we attached to each $S^2_{(4)}$ 2-disks along the four boundary circles so that we get a 2-sphere. Along the 2-spheres for $t = \pm 1$, we attached 3-disks to construct V. Thus, the relevant 3-manifold F_j is diffeomorphic to a manifold obtained by attaching two 3-disks to $S^2 \times [-1, 1]$ along the boundaries, and is therefore diffeomorphic to the 3-sphere S^3.

The same argument can be applied for the singular fiber of type II^3. The regular neighborhood of a corresponding point in W_f is shown in Figure 7.

Finally, for the singular fiber of type II^a, a similar argument can be applied as follows. The regular neighborhood of a corresponding point in W_f is shown in the top of Figure 8. The map f restricted to the inverse image of the neighborhood

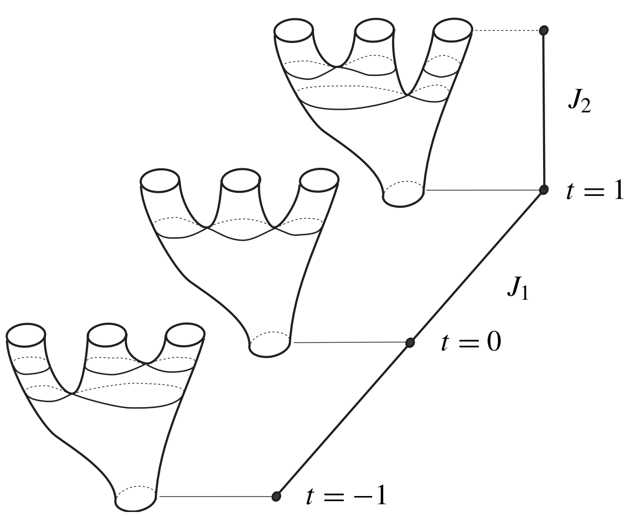


Figure 6. 1-parameter family of functions on $S^2_{(4)}$ corresponding to a II^2-type singular fiber.

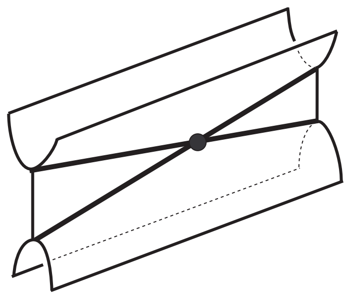


Figure 7. Neighborhood of the q_f-image of a II^3-type singular fiber.

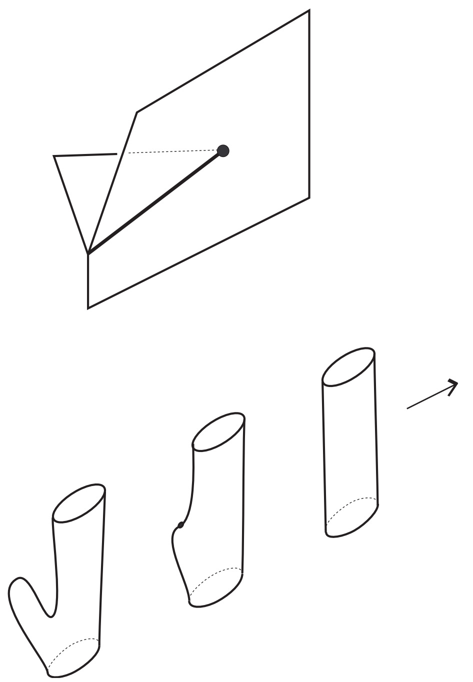


Figure 8. The case of a singular fiber of type II^a.

with respect to q_f can be regarded as a 1-parameter family of smooth functions on the annulus corresponding to a birth-death of a pair of nondegenerate critical points as shown in the bottom of Figure 8. Thus, the resulting 3-manifold F_j is again diffeomorphic to S^3.

This completes the proof of Lemma 4.1. □

Since S^3 is the oriented boundary of an oriented 4-disk, we see that M bounds a compact oriented 4-manifold. Therefore, we have proved the following.

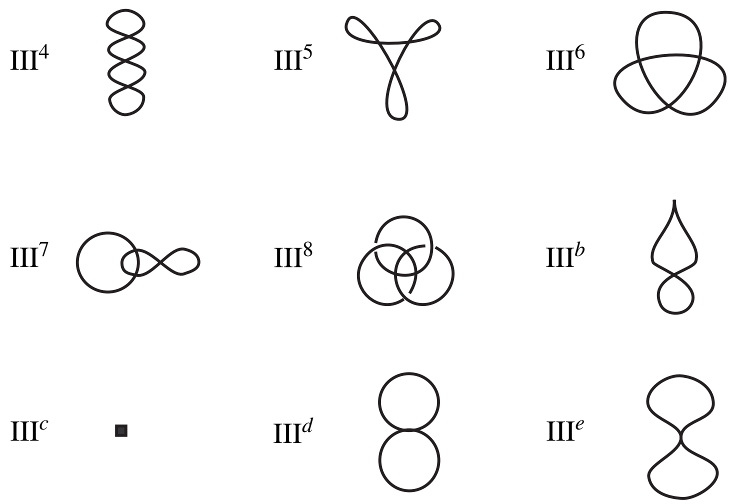


Figure 9. Singular fibers of C^∞ stable maps of orientable 4-manifolds into 3-manifolds of $\kappa = 3$.

Theorem 4.2. *The* 3*-dimensional oriented cobordism group* Ω_3 *vanishes.*

This result is originally due to Rohlin [1951] and Thom [1952]. In fact, our argument resembles that in [Costantino and Thurston 2008].

Remark 4.3. Kalmár [2009] showed that for every C^∞ stable map f of a closed orientable 3-manifold M into the plane, singular fibers of types II^2, II^3 and II^a can be eliminated by cobordism. Using this result, he showed that $\Omega_3 = 0$. For details, see [Kalmár 2009, Remark 2.8].

5. Four-dimensional oriented cobordism group

In this section, we show that the oriented cobordism group Ω_4 is infinite cyclic and is generated by the cobordism class of $\mathbb{C}P^2$, using the classification of singular fibers of C^∞ stable maps of orientable 4-manifolds into 3-manifolds.

For M a closed oriented 4-manifold, there always exists a C^∞ stable map f from M into any 3-manifold N; for example, see [Saeki 2004]. Singular fibers of such maps were classified up to C^∞ equivalence in [Saeki 2004]. The connected components of singular fibers containing singular points are shown in Figure 9 in addition to the singular fibers of codimension $\kappa = 1$ and 2 that are suspensions of the singular fibers as in Figure 4. (For the singular fibers of $\kappa \leq 2$, we continue to use the same notation I^0, I^1, and so on as in Figure 4 for the suspensions as well.)

A C^∞ stable map of a 4-manifold into a 3-manifold has only fold, cusp, and swallowtail singularities. A fold point corresponds to an isolated point or to a transverse crossing point of two line segments, a cusp point corresponds to a cuspidal vertex of degree two in a singular fiber, and a swallowtail point corresponds to an isolated point (depicted by a black square in Figure 9, III^c) or to a tangency point of two touching parabolas.

Let W_f be the quotient space in the Stein factorization of a C^∞ stable map f of a closed oriented 4-manifold M into a 3-manifold N. The space W_f is a compact 3-dimensional polyhedron, and its local structure has been completely determined. A complete list of local structures can be found in [Hiratuka 2001], although we do not need it here. Let $W^{(j)}$ denote the q_f-image of the components of singular fibers of $\kappa \geq 3 - j$ containing singular points for $j = 0, 1, 2$, and let $N^{(j)}$ denote a small regular neighborhood of $W^{(j)}$ in W_f. Then, set

$$N_0 = N^{(0)}, \quad N_1 = \mathrm{cl}(N^{(1)} \setminus N^{(0)}), \quad N_2 = \mathrm{cl}(N^{(2)} \setminus N^{(1)}), \quad N_3 = \mathrm{cl}(W_f \setminus N^{(2)}),$$

where N_1 is seen as a regular neighborhood of $\mathrm{cl}(W^{(1)} \setminus N^{(0)})$ in $\mathrm{cl}(W_f \setminus N^{(0)})$ and N_2 is seen as a regular neighborhood of $\mathrm{cl}(W^{(2)} \setminus N^{(1)})$ in $\mathrm{cl}(W_f \setminus N^{(1)})$.

Let us now construct a compact 5-dimensional manifold V by attaching certain pieces to $M \times [0, 1]$ as follows. First note that q_f restricted to $q_f^{-1}(N_3)$ is a locally trivial fibration with fiber S^1 over the 3-manifold possibly with boundary N_3. Thus, we can attach the total space of the associated D^2-bundle over N_3 to $M \times [0, 1]$ by identifying the associated (∂D^2)-bundle with $q_f^{-1}(N_3) \times \{0\}$. The resulting 5-manifold, denoted by V_1, is orientable.

Let V_1' (a subset of V_1) be the union of $M \times \{0\}$ and the D^2-bundle over N_3. There is a natural map $q_1 : V_1' \to W_f$ that on $M \times \{0\}$ is defined by $q_f : M \times \{0\} = M \to W_f$, and on the D^2-bundle is defined by the projection to N_3.

Let S be a connected component of $\mathrm{cl}(W^{(2)} \setminus N^{(1)})$. Note that S is a compact surface possibly with boundary. Let $N_2(S)$ denote the connected component of N_2 containing S. If the singular fiber lying over a point in S is of type I^0, then $q_f^{-1}(N_2(S))$ is diffeomorphic to the total space of a D^2-bundle over S. In fact, $N_2(S)$ is diffeomorphic to $S \times [0, 1]$ by a diffeomorphism that induces the identity S $(\subset N_2(S)) \to S \times \{0\}$, and if J $(\cong [0, 1])$ is a fiber of the natural fibration $N_2(S) \to S$, then q_f restricted to $q_f^{-1}(J) \cong D^2$ is equivalent to the function (4-1). (For this, we need the bundle structure theorem mentioned in [Kalmár 2008, Section 5; Kalmár 2009, Section 6].) Therefore, the map q_1 restricted to $q_1^{-1}(N_2(S))$ followed by the natural projection $N_2(S) \to S$ is an S^2-bundle whose fiber can be identified with the 2-sphere as in Figure 3(1). Then, we can attach the associated D^3-bundle over S to V_1, where we identify the associated (∂D^3)-bundle with $q_1^{-1}(N_2(S))$ $(\subset V_1')$. The resulting 5-manifold is orientable, since so is V_1 and the orientability of $q_1^{-1}(N_2(S))$ coincides with that of $q_f^{-1}(N_2(S))$ $(\subset M)$.

If the singular fiber lying over a point in S is of type I^1, then $q_f^{-1}(N_2(S))$ is diffeomorphic to the total space of an $S^2_{(3)}$-bundle over S. Then, q_1 restricted to $q_1^{-1}(N_2(S))$ followed by the natural projection $N_2(S) \to S$ is again an S^2-bundle,

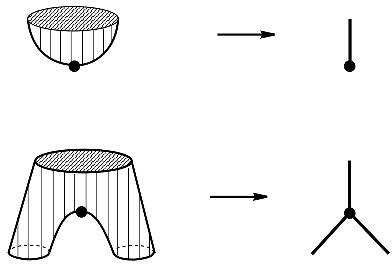


Figure 10. Function on each D^3-fiber.

but with fiber as in Figure 3(2), and we can attach the associated D^3-bundle. As before, the resulting 5-manifold is orientable.

We perform the operation described above for each connected component S of $\mathrm{cl}(W^{(2)} \setminus N^{(1)})$. The resulting 5-manifold, denoted V_2, is orientable. We denote by V_2' the union of V_1' and the D^3-bundle over $\mathrm{cl}(W^{(2)} \setminus N^{(1)})$. There is a natural map $q_2 : V_2' \to W_f$ that on V_1' is defined by $q_1 : V_1' \to W_f$, and on the D^3-bundle X over $\mathrm{cl}(W^{(2)} \setminus N^{(1)})$ is defined by the natural map $X \to N_2$ that makes the diagram

$$\begin{array}{ccc} X & \longrightarrow & N_2 \\ & \searrow \quad \swarrow & \\ & \mathrm{cl}(W^{(2)} \setminus N^{(1)}) & \end{array}$$

commutative and is given by the function on each fiber as shown in Figure 10. (For this, we again need Kalmár's bundle structure theorem.)

Let e be a connected component of $\mathrm{cl}(W^{(1)} \setminus N^{(0)})$ and we denote by $N_1(e)$ the connected component of N_1 containing e. If the singular fiber lying over a point in e is of type II^2, then $q_f^{-1}(N_1(e))$ is diffeomorphic to the total space of an $(S^2_{(4)} \times [0, 1])$-bundle over e; see Figure 6.

Therefore, q_2 restricted to $q_2^{-1}(N_1(e))$ followed by the natural projection from $N_1(e)$ to e is an S^3-bundle by Lemma 4.1. From [Hatcher 1983], the structure group of every smooth S^3-bundle can be reduced to the orthogonal group $O(4)$. Then, we can attach the associated D^4-bundle over e to V_2, where we identify the associated (∂D^4)-bundle with $q_2^{-1}(N_1(e))$ $(\subset V_2')$. The resulting 5-manifold is orientable, since so is V_2 and the orientability of $q_2^{-1}(N_1(e))$ coincides with that of $q_f^{-1}(N_1(e))$ $(\subset M)$.

If the singular fiber lying over a point in e is of another type (II^3 or II^a), then we can still perform the same operation by virtue of Lemma 4.1.

We perform such an operation for each e. The resulting 5-manifold V is compact and orientable. Note that V can be oriented so that

$$\partial V = (M \times \{1\}) \cup \bigl(-\textstyle\bigcup_j F_j\bigr),$$

where each F_j is a closed oriented 4-manifold corresponding to a singular fiber of f of $\kappa = 3$.

Lemma 5.1. *F_j is diffeomorphic to the 4-sphere S^4 for the singular fibers of types* III^4, III^5, III^6, III^7, III^b, III^c, III^d *and* III^e.

Proof. Let us first consider the case of the singular fiber of type III^4. Let $g : L \to D^3$ be a representative of the singular fiber; we assume that it has a singular fiber of type III^4 over the center of D^3. Then, we can regard g as a family of functions $\{h_s\}_{s\in\Delta}$ on $S^2_{(5)}$ with only nondegenerate critical points parametrized by $\Delta \cong D^2$ as depicted in Figure 11, where the target D^3 is identified with the product $[-1, 1]\times\Delta$, $\pi : [-1, 1] \times \Delta \to \Delta$ is the projection to the second factor, and the critical points of h_s are denoted by p_1, p_2 and p_3. More precisely, g can be identified with the map $L \cong S^2_{(5)} \times \Delta \to [-1, 1] \times \Delta \cong D^3$, $(x, s) \mapsto (h_s(x), s)$. The singular point set $S(g)$ of g consists of three 2-disks, and their images by g in D^3 intersect at the origin in general position.

Then, from the construction of V it follows that the 4-manifold F_j corresponding to $g^{-1}(0)$ is diffeomorphic to the boundary of a $\widetilde{D}$-bundle over Δ, where $\widetilde{D} \cong D^3$ is the 3-disk as in Figure 12, which is obtained by filling $S^2_{(5)} \times [0, 1]$ by 2- and 3-handles as in Section 3. Hence F_j is diffeomorphic to S^4.

For the singular fibers of types III^5, III^b, III^c and III^e, similar arguments show that $F_j \cong S^4$.

For the singular fiber of type III^6, we can again regard its representative $g : L \to D^3$ as a family of functions parametrized by a 2-disk Δ. Note that L is diffeomorphic to $S^2_{(5)} \times \Delta$. Set $F'_j = L \cup q_1^{-1}((\bar{g})^{-1}(\{\pm 1\}\times\Delta))$ and $F''_j = \mathrm{cl}(F_j \setminus F'_j)$, where $g = \bar{g} \circ q_g$ is the Stein factorization of $g : L \to [-1, 1] \times \Delta$ and we regard W_g as a subset of W_f.

Then, we see that F'_j is diffeomorphic to $S^2 \times D^2$. On the other hand, from the construction of V it follows that F''_j is the union of six copies of $D^3 \times [-1, 1]$ attached to each other along $D^3 \times \{\pm 1\}$ consecutively so that if forms the total space of a D^3-bundle over S^1. Hence, F''_j is diffeomorphic to $D^3 \times S^1$, since it is orientable. Therefore, F_j is diffeomorphic to the union of $S^2 \times D^2$ and $D^3 \times S^1$ attached along their boundaries, where $S^2 \times \{*\} \subset S^2 \times D^2$ and $\partial D^3 \times \{*\} \subset D^3 \times S^1$ are identified. Then a standard argument shows that F_j is diffeomorphic to S^4.

For the singular fibers of types III^7 and III^d, similar arguments give $F_j \cong S^4$.

This completes the proof of Lemma 5.1. □

Lemma 5.2. *For the singular fiber of type* III^8, *F_j is orientation-preservingly diffeomorphic to the complex projective plane $\mathbb{C}P^2$ or its orientation reversal $\overline{\mathbb{C}P^2}$.*

Proof. As in the proof of Lemma 5.1, a representative $g : L \to D^3$ of the singular fiber of type III^8 can be regarded as a family of functions $\{h_s\}_{s\in\Delta}$ on $T^2_{(3)}$ with only nondegenerate critical points parametrized by a small 2-disk Δ. (Here, the torus

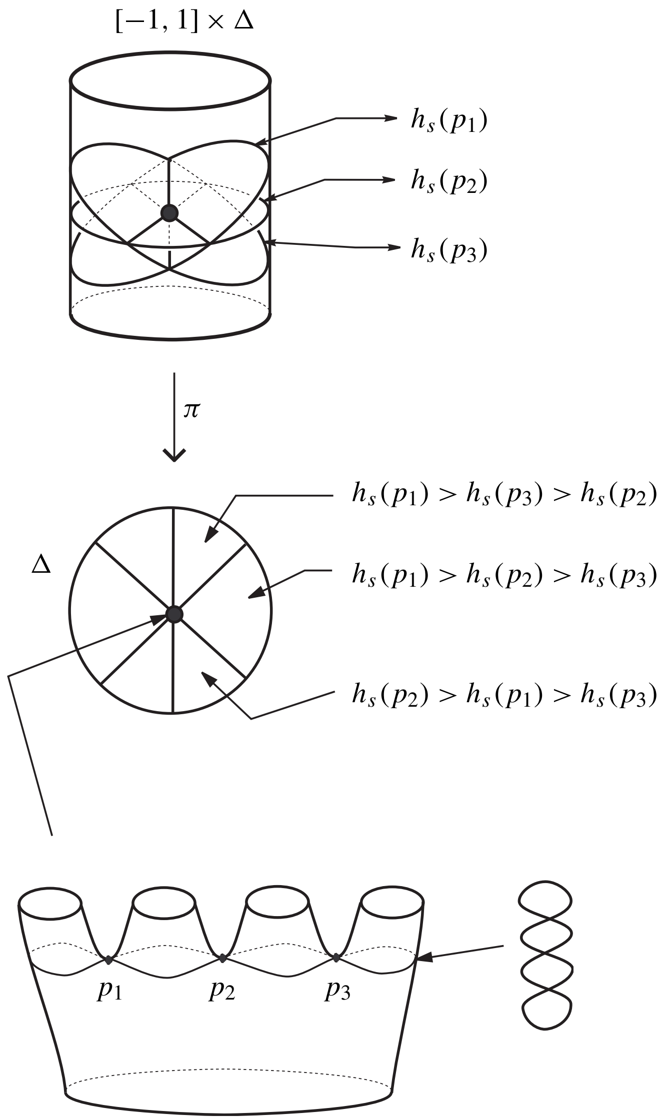


Figure 11. Family of functions corresponding to a singular fiber of type III^4.

with three holes appears, since the natural "thickening" of the III^8-type singular fiber is a compact orientable surface with three boundary circles and has Euler characteristic -3. See Figure 13.) See Figure 14, where $\pi : D^3 \cong [-1, 1] \times \Delta \to \Delta$ is the projection onto the second factor; see also [Saeki 2004, Figure 6.3].

Set $K = g^{-1}(\partial D^3) = \partial L$, which is a closed orientable 3-manifold. Then, $g|_K : K \to \partial D^3$ can be regarded as a stable map as in Section 4. Note that F_j is $L \cup W$

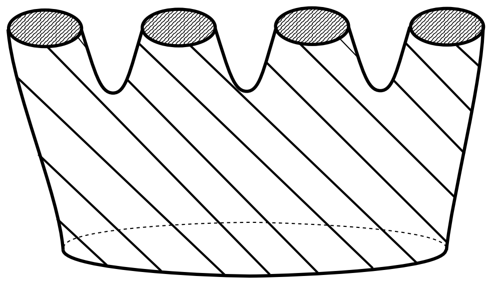


Figure 12. $\widetilde{D} \cong D^3$.

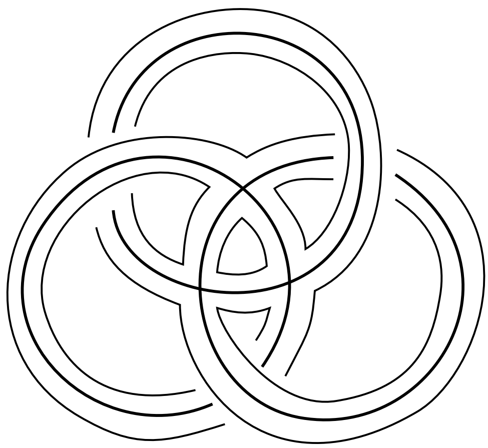


Figure 13. Natural "thickening" of the III^8-type singular fiber.

attached along K, where W is the compact orientable 4-manifold bounded by K constructed as in Section 4 from the stable map $g|_K$.

Set

$$F'_j = L \cup q_1^{-1}((\bar{g})^{-1}(\{\pm 1\} \times \Delta)) \quad \text{and} \quad F''_j = \mathrm{cl}(F_j \setminus F'_j)$$

as in the proof of the previous lemma. Note that F'_j is diffeomorphic to a T^2-bundle over Δ. (More precisely, the map $\pi \circ g : L \to \Delta$ is a smooth fiber bundle with $T^2_{(3)}$ as fibers, and F'_j is obtained from L by attaching three 2-disks to each of the fibers.)

Let us consider the piece P_i in F''_j corresponding to the arc α_i for $i = 1, 2, 3$, on $\partial\Delta$ as depicted in Figure 14. More precisely, P_i is the compact 4-manifold described as follows. First, note that $g|_{(\pi \circ g)^{-1}(\alpha_i)} : (\pi \circ g)^{-1}(\alpha_i) \to [-1, 1] \times \alpha_i$ can be regarded as a 1-parameter family of smooth functions on $T^2_{(3)}$ with exactly three nondegenerate critical points corresponding to interchanging the heights of the top two critical points. Furthermore, the singular fiber (of codimension two)

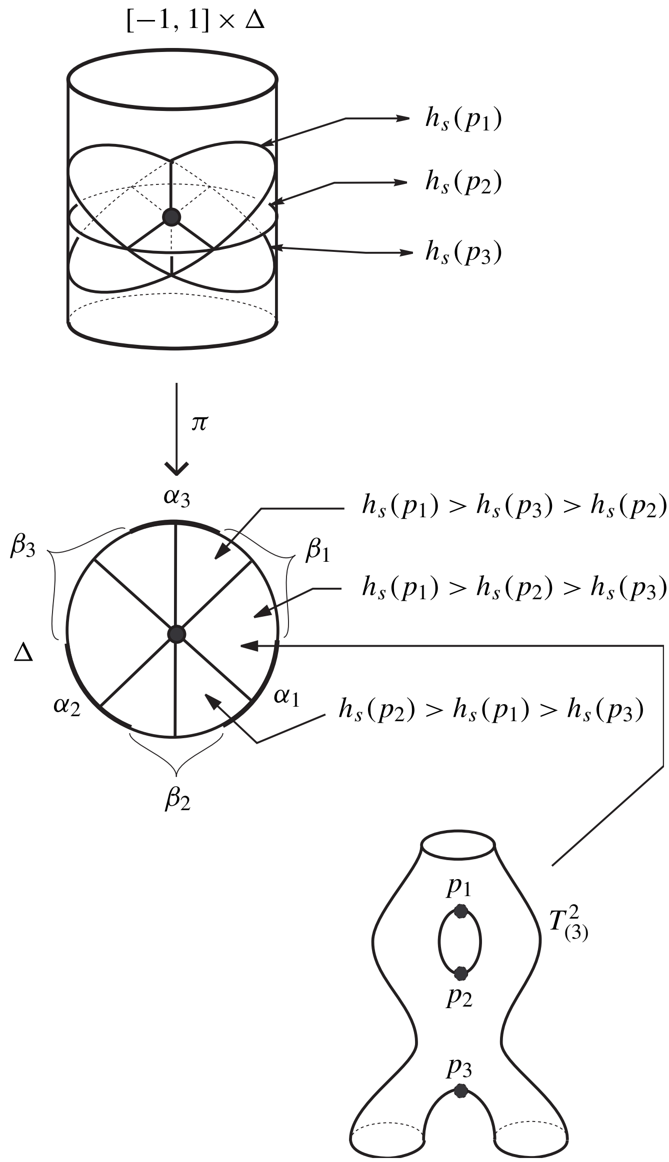


Figure 14. Family of functions corresponding to a singular fiber of type III^8.

over the middle point of α_i corresponds to the singular fiber of type II^3. The compact 4-manifold P_i is obtained from $(\pi \circ g)^{-1}(\alpha_i) \times [0, 1]$ by attaching D^2-bundles, D^3-bundles and a 4-disk as in Section 4, where the last 4-disk corresponds to the II^3-type singular fiber.

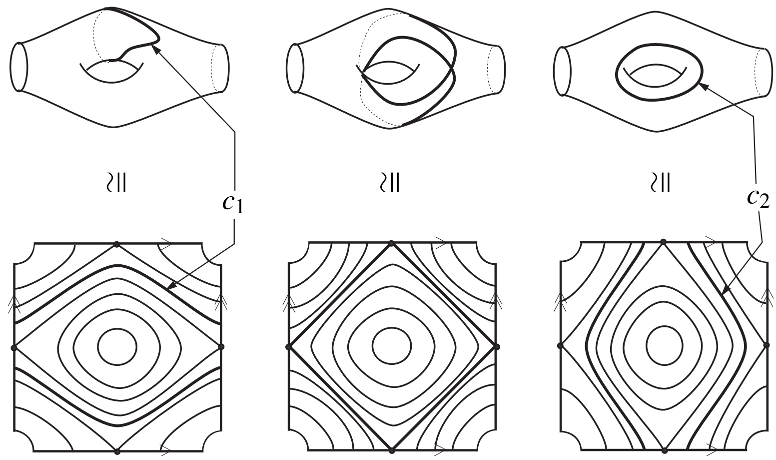


Figure 15. Family of functions corresponding to II^3.

By attaching D^2-bundles over arcs and three copies of D^3 to each component of $(\pi \circ g)^{-1}(\partial \alpha_i)$, we get two copies of the solid torus. Hence ∂P_i is diffeomorphic to the union of two solid tori attached along their boundaries. The attaching map sends the boundary c_1 of a meridian disk[1] to a simple closed curve on the boundary of the other solid torus that intersects with the boundary c_2 of its meridian disk transversely at one point. For details, see Figure 15, which shows how the fibers change around a singular fiber of type II^3. (Note that each component of a regular fiber bounds a disk by virtue of the construction of V.) Hence, ∂P_i is diffeomorphic to S^3. Since P_i is obtained from $\partial P_i \times [0, 1]$ by attaching a 4-disk, we see that P_i is diffeomorphic to D^4.

On the other hand, the piece Q_i corresponding to the arc β_i for $i = 1, 2, 3$ on $\partial \Delta$ as shown in Figure 14 is diffeomorphic to $(S^1 \times D^2) \times [-1, 1]$. This is because the singular fiber of type II^3 over the middle point of β_i corresponds to interchanging the heights of the bottom two critical points of the function $h_s : T^2_{(3)} \to [-1, 1]$, where $s \in \beta_i$. In terms of the quotient space, its behavior is similar to what is depicted in Figure 7. Therefore, Q_i is obtained from a 4-disk (corresponding to the II^3-type singular fiber) by attaching $D^3 \times \beta_i$ (corresponding to the top critical point of h_s)[2] along $(D^2_1 \cup D^2_2) \times \beta_i$, where D^2_1 and D^2_2 are disjoint 2-disks in ∂D^3. Therefore, Q_i is diffeomorphic to $(S^1 \times D^2) \times \beta_i \cong (S^1 \times D^2) \times [-1, 1]$, since over each end point of β_i we have a solid torus, which is orientable.

Consequently, F''_j is diffeomorphic to the compact 4-manifold obtained from three 4-disks P_1, P_2 and P_3 by attaching them appropriately along solid tori. Thus

[1] A properly embedded 2-disk in a solid torus is a *meridian disk* if its boundary is not null homotopic in the boundary torus.

[2] Each 3-disk $D^3 \times \{*\}$ corresponds to that in the lower-left figure of Figure 10.

P_1 can be regarded as a 0-handle and then P_2 is regarded as a 2-handle attached to P_1 along an unknotted circle on ∂P_1, since the exterior of the attaching circle in ∂P_1 is again a solid torus. Furthermore, F'_j is diffeomorphic to $T^2 \times D^2$, and $P_3 \cup F'_j$ is diffeomorphic to D^4 since $P_3 \cap F'_j$ is diffeomorphic to $T^2 \times \alpha_3$.

Therefore, F_j is diffeomorphic to the closed 4-manifold consisting of the 0-handle P_1, the 2-handle P_2 attached to P_1 along an unknotted circle on ∂P_1, and a 4-handle. In particular, the boundary of $P_1 \cup P_2$ must be diffeomorphic to S^3 so that the framing of the 2-handle P_2 must be equal to ± 1; see [Kirby 1989], for example. Therefore, F_j must be diffeomorphic to the complex projective plane $\mathbb{C}P^2$ up to orientation. □

Remark 5.3. It is easy to see that the singular fibers appearing in Figure 9 can be embedded in the 2-sphere, except for the III^8-type singular fiber. This fact implies that the corresponding singular fiber is associated with a 2-parameter family of smooth functions on a punctured 2-sphere, as pointed out in the proof of Lemma 5.1. The III^8-type singular fiber cannot be embedded in the 2-sphere, but can be embedded in the 2-dimensional torus (see the proof of Lemma 5.2). The proofs above show that this fact is essential in distinguishing the III^8-type singular fiber from the others.

We have proved that every closed oriented 4-manifold is oriented cobordant to the disjoint union of a finite number of copies of $\pm\mathbb{C}P^2$.

Let us consider the homomorphism $\varphi : \mathbb{Z} \to \Omega_4$ defined by $\varphi(1) = [\mathbb{C}P^2]$. This is a well-defined homomorphism, and is surjective by the argument above.

Let

$$\sigma : \Omega_4 \to \mathbb{Z} \tag{5-1}$$

be the homomorphism defined by associating to each oriented cobordism class the signature of its representative. Classical techniques in algebraic topology show that this defines a well-defined homomorphism; for example, see [Thom 1952].

Since $\sigma([\mathbb{C}P^2]) = 1$, we see that the composition $\sigma \circ \varphi : \mathbb{Z} \to \mathbb{Z}$ is the identity. Therefore, φ is injective. Thus, we get the following, which was originally proved by Rohlin [1952]; see also [Guillou and Marin 1986].

Theorem 5.4. *The* 4*-dimensional oriented cobordism group* Ω_4 *is infinite cyclic and is generated by the oriented cobordism class of* $\mathbb{C}P^2$*. In fact, the homomorphism* (5-1) *is an isomorphism.*

Our proof shows that the complex projective plane naturally appears around each singular fiber of type III^8, and therefore $\mathbb{C}P^2$ can be regarded as the genuinely natural representative of the generator of $\Omega_4 \cong \mathbb{Z}$.

As a corollary to our proof, we get the following, which was originally obtained in [Saeki and Yamamoto 2006]. The proof given there is somewhat complicated since it depends on the classification of singular fibers of C^∞ stable maps of n-dimensional manifolds into $(n-1)$-dimensional manifolds for $n \leq 5$, whereas our proof needs only the classification of such singular fibers for $n \leq 4$.

Corollary 5.5. *Let f be a C^∞ stable map of a closed oriented 4-manifold M into a 3-manifold N. Then the number of singular fibers of f of type* III^8 *counted with signs coincides with the signature of M.*

The sign of a singular fiber of type III^8 is $+1$ (or -1) if the corresponding manifold F_j is oriented diffeomorphic to $\mathbb{C}P^2$ (respectively $\overline{\mathbb{C}P^2}$). This sign convention must coincide with the one in [Saeki and Yamamoto 2006] since Corollary 5.5 determines the sign uniquely.

Gromov [2009] studied estimates for the number of self-intersections of the critical value set of a generic map from one manifold to another in terms of the topology of the source manifold. Corollary 5.5 gives a model case for such a study, as pointed out by Gromov.

Corollary 5.6. *Let f be a smooth map of a closed oriented n-dimensional manifold M into an $(n-1)$-dimensional manifold N for $n \geq 4$, and suppose its singular fibers are (iterated suspensions of) those of C^∞ stable maps of codimension ≤ 3 not of type* III^8 *(that is, the singular fibers of f are as in Figures 4 and 9 but without the* III^8*-type). Then the manifold M is oriented null cobordant.*

For the proof, we again need Kalmár's bundle structure theorem concerning the structure group since we need to deal with smooth fiber bundles with fiber S^4. Note that this corollary generalizes [Kalmár 2009, Corollary 2.7] about simple fold maps.

In particular, if M is not oriented null cobordant, then every generic map $M \to N$ has a singular fiber of codimension ≥ 4 or (an iterated suspension of) a singular fiber of type III^8.

Remark 5.7. Unfortunately, our technique in this paper does not directly apply for computing the 3-dimensional unoriented cobordism group $\mathfrak{N}_3$. This is because the 2-dimensional one is not trivial, and we cannot fill $\mathbb{R}P^2$-bundles over arcs and circles. For similar reasons, our method cannot directly be used for computating $\mathfrak{N}_m$ for $m \geq 4$ and Ω_n for $n \geq 5$.

Acknowledgments

I would like to express my sincere gratitude to Boldizsár Kalmár and Takahiro Yamamoto for stimulating discussions and invaluable comments. I also thank the referee for helpful comments that improved the presentation of the paper.

References

[Costantino and Thurston 2008] F. Costantino and D. Thurston, "3-manifolds efficiently bound 4-manifolds", *J. Topol.* **1**:3 (2008), 703–745. MR 2009g:57034 Zbl 1166.57016

[Golubitsky and Guillemin 1973] M. Golubitsky and V. Guillemin, *Stable mappings and their singularities*, Graduate Texts in Math. **14**, Springer, New York, 1973. MR 49 #6269 Zbl 0294.58004

[Gromov 2009] M. Gromov, "Singularities, expanders and topology of maps, I: Homology versus volume in the spaces of cycles", *Geom. Funct. Anal.* **19**:3 (2009), 743–841. MR 2563769 Zbl 05652310

[Guillou and Marin 1986] L. Guillou and A. Marin (editors), *À la recherche de la topologie perdue*, Progress in Mathematics **62**, Birkhäuser, Boston, 1986.

[Hatcher 1983] A. E. Hatcher, "A proof of the Smale conjecture, $\mathrm{Diff}(S^3) \simeq \mathrm{O}(4)$", *Ann. of Math.* (2) **117**:3 (1983), 553–607. MR 85c:57008 Zbl 0531.57028

[Hiratuka 2001] J. T. Hiratuka, *A fatorização de Stein e o número de singularidades de aplicações estáveis*, thesis, University of São Paulo, 2001.

[Kalmár 2008] B. Kalmár, "Pontryagin–Thom–Szűcs type construction for non-positive codimensional singular maps with prescribed singular fibers", pp. 66–79 in *The second Japanese-Australian Workshop on Real and Complex Singularities*, RIMS Kôkyûroku **1610**, Research Institute for Mathematical Sciences, Kyoto University, 2008.

[Kalmár 2009] B. Kalmár, "Fold maps and immersions from the viewpoint of cobordism", *Pacific J. Math.* **239**:2 (2009), 317–342. MR 2009i:57063 Zbl 1171.57028

[Kirby 1989] R. C. Kirby, *The topology of* 4*-manifolds*, Lecture Notes in Math. **1374**, Springer, Berlin, 1989. MR 90j:57012 Zbl 0668.57001

[Kushner et al. 1984] L. Kushner, H. Levine, and P. Porto, "Mapping three-manifolds into the plane, I", *Bol. Soc. Mat. Mexicana* (2) **29**:1 (1984), 11–33. MR 86j:58011 Zbl 0586.57018

[Levine 1985] H. Levine, *Classifying immersions into* $\mathbb{R}^4$ *over stable maps of* 3*-manifolds into* $\mathbb{R}^2$, Lecture Notes in Math. **1157**, Springer, Berlin, 1985. MR 88f:57056 Zbl 0567.57001

[Mather 1971] J. N. Mather, "Stability of C^∞ mappings, VI: The nice dimensions", pp. 207–253 in *Singularities Symposium* (Liverpool, 1969/70), vol. 1, Lecture Notes in Math. **192**, Springer, Berlin, 1971. MR 45 #2747

[Milnor and Stasheff 1974] J. W. Milnor and J. D. Stasheff, *Characteristic classes*, Annals of Mathematics Studies **76**, Princeton University Press, 1974. MR 55 #13428 Zbl 0298.57008

[Pontryagin 1955] L. S. Pontryagin, "Smooth manifolds and their applications in homotopy theory", Trudy Mat. Inst. im. Steklov. **45**, Izdat. Akad. Nauk SSSR, Moscow, 1955. In Russian; translated in pp. 1–114 of American Mathematical Society Translations (2) **11**, American Mathematical Society, Providence, RI, 1959. MR 22 #5980 Zbl 0064.17402

[Rohlin 1951] V. A. Rohlin, "A three-dimensional manifold is the boundary of a four-dimensional one", *Doklady Akad. Nauk SSSR* (*N.S.*) **81** (1951), 355–357. In Russian. MR 14,72g

[Rohlin 1952] V. A. Rohlin, "New results in the theory of four-dimensional manifolds", *Doklady Akad. Nauk SSSR* (*N.S.*) **84** (1952), 221–224. In Russian. MR 14,573b

[Saeki 2004] O. Saeki, *Topology of singular fibers of differentiable maps*, Lecture Notes in Mathematics **1854**, Springer, Berlin, 2004. MR 2005m:58085 Zbl 1072.57023

[Saeki and Yamamoto 2006] O. Saeki and T. Yamamoto, "Singular fibers of stable maps and signatures of 4-manifolds", *Geom. Topol.* **10** (2006), 359–399. MR 2007f:57058 Zbl 1107.57019

[Smale 1959] S. Smale, "Diffeomorphisms of the 2-sphere", *Proc. Amer. Math. Soc.* **10** (1959), 621–626. MR 22 #3004 Zbl 0118.39103

[Thom 1952] R. Thom, *Quelques propriétés des variétés-bords*, Colloque de Topologie de Strasbourg 1951 **5**, La Bibliothèque Nationale et Universitaire de Strasbourg, 1952. MR 14,492a

[Thom 1954] R. Thom, "Quelques propriétés globales des variétés différentiables", *Comment. Math. Helv.* **28** (1954), 17–86. MR 15,890a Zbl 0057.15502

[Wall 1959] C. T. C. Wall, "Note on the cobordism ring", *Bull. Amer. Math. Soc.* **65** (1959), 329–331. MR 21 #6586 Zbl 0128.16801

[Yamamoto 2006] T. Yamamoto, "Classification of singular fibres of stable maps of 4-manifolds into 3-manifolds and its applications", *J. Math. Soc. Japan* **58**:3 (2006), 721–742. MR 2007m:57035 Zbl 1105.57027

[Yamamoto 2007] T. Yamamoto, "Euler number formulas in terms of singular fibers of stable maps", pp. 427–457 in *Real and complex singularities* (Sydney, 2005), edited by Y. Paunescu et al., World Sci. Publ., Hackensack, NJ, 2007. MR 2008m:57071 Zbl 1143.57016

Received August 2, 2009. Revised October 19, 2009.

Osamu Saeki
Faculty of Mathematics
Kyushu University
Motooka 744, Nishi-ku
Fukuoka 819-0395
Japan

saeki@math.kyushu-u.ac.jp
http://www2.math.kyushu-u.ac.jp/∼saeki/index.html

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 248 No. 1 November 2010

0030-8730(201011)248:1;1-C